Tail-dependence, exceedance sets, and metric embeddings

Anja Janßen*, Sebastian Neblung†, Stilian Stoev‡

December 5, 2022

Abstract

There are many ways of measuring and modeling tail-dependence in random vectors: from the general framework of multivariate regular variation and the flexible class of max-stable vectors down to simple and concise summary measures like the matrix of bivariate tail-dependence coefficients. This paper starts by providing a review of existing results from a unifying perspective, which highlights connections between extreme value theory and the theory of cuts and metrics. Our approach leads to some new findings in both areas with some applications to current topics in risk management.

We begin by using the framework of multivariate regular variation to show that extremal coefficients, or equivalently, the higher-order tail-dependence coefficients of a random vector can simply be understood in terms of random exceedance sets, which allows us to extend the notion of Bernoulli compatibility. In the special but important case of bi-variate tail-dependence, we establish a correspondence between tail-dependence matrices and L^1- and ℓ_1-embeddable finite metric spaces via the spectral distance, which is a metric on the space of jointly 1-Fréchet random variables. Namely, the coefficients of the cut-decomposition of the spectral distance and of the Tawn-Molchanov max-stable model realizing the corresponding bi-variate extremal dependence coincide. We show that line metrics are rigid and if the spectral distance corresponds to a line metric, the higher order tail-dependence is determined by the bi-variate tail-dependence matrix.

Finally, the correspondence between ℓ_1-embeddable metric spaces and tail-dependence matrices allows us to revisit the realizability problem, i.e. checking whether a given matrix is a valid tail-dependence matrix. We confirm a conjecture of Shyamalkumar & Tao (2020) that this problem is NP-complete.

Keywords: Bernoulli compatibility, exceedance sets, line metrics, max-stable vectors, metric embedding, multivariate regular variation, NP completeness, tail-dependence (TD) matrix, tail-dependence coefficients, Tawn-Molchanov models, realization problem for TD matrices
MSC2020: primary 60G70; secondary 51K05, 60E05, 68R12, 68Q25

1 Introduction

Extreme events such as large portfolio losses in insurance and finance, spatial and environmental extremes such as heat-waves, floods, electric grid outages, and many other complex system failures are associated with tail-events. That is, the simultaneous occurrence of extreme values in

*Otto-von-Guericke University Magdeburg, Germany (anja.janssen@ovgu.de)
†University of Hamburg, Germany (sebastian.neblung@uni-hamburg.de)
‡University of Michigan, Ann Arbor, US (sstoev@umich.edu); This author was partially supported by the NSF Grant DMS-1916226.

the components of a possibly very high-dimensional vector $X = (X_i)_{1\le i\le p}$ of covariates. Such simultaneous extremes occur due to *dependence* among the extremes of the X_i's. This has motivated a large body of literature on modeling and quantifying tail-dependence, see, e.g. (Coles 2001, Finkenstädt & Rootzén 2003, Rachev 2003, Beirlant et al. 2004, Castillo 1988, Resnick 2007, de Haan & Ferreira 2007). One basic and popular measure is the bivariate (upper) tail-dependence coefficient

$$\lambda_X(i,j) := \lim_{u\uparrow 1} \mathbb{P}[X_i > F_i^{-1}(u) | X_j > F_j^{-1}(u)], \tag{1.1}$$

where $F_i^{-1}(u) := \inf\{x : \mathbb{P}[X_i \le x] \ge u\}$ is the generalized inverse of the cumulative distribution function F_i of X_i. Under weak conditions the above limit exists and is independent of the choice of the (continuous) marginal distributions of (X_i, X_j). The matrix $\Lambda := (\lambda_X(i,j))_{p\times p}$ of bivariate tail-dependence coefficients is necessarily positive (semi)definite and in fact, since $\lambda_X(i,i) = 1$, it is a correlation matrix of a random vector, see Schlather & Tawn (2003). We call Λ as defined in (1.1) a tail-dependence matrix or TD matrix for short.

The general theme of our paper is that we review and contribute to the unified treatment of tail-dependence using the powerful framework of multivariate regular variation. This leads to deep connections to existing results in the theory of cut (semi)metrics and ℓ_1-embeddable metrics (Deza & Laurent 1997), as well as to extensions to the Bernoulli compatibility characterization of tail-dependence matrices established in Embrechts et al. (2016) and Krause et al. (2018). What follows is an overview of our key ideas and contributions.

Since the marginal distributions of X are not important in quantifying tail-dependence, one may transform its marginals to be heavy-tailed. In fact, we make the additional and often very mild assumption that the vector X is regularly varying, i.e., that there exists a Radon measure μ on $\mathbb{R}^p \setminus \{0\}$ and a suitable positive sequence $a_n \uparrow \infty$ such that

$$n\mathbb{P}[X \in a_n A] \to \mu(A), \quad \text{as } n \to \infty,$$

for all Borel sets $A \subset \mathbb{R}^p$ that are bounded away from 0 and such that $\mu(\partial A) = 0$ (see Definition 2.1, below). This allows us to conclude that $n\mathbb{P}[h(X) > a_n] \to \mu\{h > 1\}$ for a large class of continuous and 1-homogeneous functions $h : \mathbb{R}^p \to [0,\infty)$ (Proposition 2.5). Therefore, if h is a certain risk functional, we readily obtain an asymptotic approximation of the probability of an extreme loss $\mathbb{P}[h(X) > a_n] \approx n^{-1}\mu\{h > 1\}$. By varying the risk functional h, one obtains different measures of tail-dependence, which may be of particular interest to practitioners. For example, if $L = \{i_1, \cdots, i_k\} \subset [p] := \{1, \cdots, p\}$ and taking $h_L(X) = (\min_{i\in L} X_i)_+ := \max\{0, \min_{i\in L} X_i\}$, the risk functional quantifies the joint exceedance probability

$$\mathbb{P}[h_L(X) > a_n] = \mathbb{P}[\min_{i\in L} X_i > a_n]$$

that all components of X with index in the set L are simultaneously extreme – an event with potentially devastating consequences. In practice, due to the limited horizon of historical data such extreme events especially for large sets L are rarely (if ever) observed. Thus, quantifying their probabilities is very challenging. Yet, as Emil Gumbel had eloquently put it *"It is not possible that the improbable will never occur."* This underscores the importance of the theoretical understanding, modeling, and inference of such functionals. Namely, one naturally arrives at the

higher order tail-dependence coefficients

$$\lambda_X(L) := \lim_{n\to\infty} n\mathbb{P}[\min_{i\in L} X_i > a_n].$$

It can be seen that if the marginals of the X_i's are identical and a_n is such that $n^{-1} \sim \mathbb{P}[X_i > a_n]$ (i.e. $\lim_{n\to\infty} n\mathbb{P}[X_i > a_n] = 1$), then $\lambda_X(\{i,j\}) = \lim_{n\to\infty} \mathbb{P}[X_i > a_n \mid X_j > a_n]$ recovers the classic bivariate tail-dependence coefficients $\lambda_X(i,j)$ in (1.1). Using the functionals $h(X) := \max_{j\in K} X_j$ for some $K \subset [p]$, one arrives at the popular extremal coefficients arising in the study of max-stable processes:

$$\theta_X(K) := \lim_{n\to\infty} n\mathbb{P}[\max_{j\in K} X_j > a_n].$$

Starting from the seminal works of Schlather & Tawn (2002, 2003), the structure of the extremal coefficients $\{\theta_X(K),\ K \subset [p]\}$ has been studied extensively, see Strokorb & Schlather (2015), Strokorb et al. (2015), Molchanov & Strokorb (2016), Fiebig et al. (2017), which address fundamental theoretical problems and develop stochastic process extensions. Our goal here is more modest. We want to study *both* the tail-dependence *and* extremal coefficients as risk functionals from the unifying perspective of regular variation. Interestingly, they can be succinctly understood in terms of exceedance sets. Namely, defining the random set

$$\Theta_n := \{i \in [p] \,:\, X_i > a_n\}$$

we show (Proposition 3.1 below)

$$\Theta_n | \{\Theta_n \neq \emptyset\} \xrightarrow{d} \Theta, \quad \text{as } n \to \infty,$$

where the limit Θ is a non-empty random subset of $[p]$ such that

$$\lambda_X(L) = a \cdot \mathbb{P}[L \subset \Theta] \quad \text{and} \quad \theta_X(K) = a \cdot \mathbb{P}[K \cap \Theta \neq \emptyset], \tag{1.2}$$

where $a = \theta_X([p])$. Thus, λ_X and θ_X (up to rescaling by a) are precisely the inclusion and hitting functionals characterizing the distribution of Θ (Molchanov 2017). Interestingly, the probability mass function of the random set Θ recovers (up to rescaling) the coefficients in a (generalized) Tawn-Molchanov max-stable model associated with X (see (3.6)).
The above probabilistic representation in (1.2) of the tail-dependence functionals leads to transparent proofs of seminal results from Embrechts et al. (2016) and Krause et al. (2018) on the characterization of TD matrices in terms of so-called Bernoulli-compatible matrices. In fact, we readily obtain a more general result on the characterization of higher-order tail-dependence coefficients via Bernoulli-compatible tensors (Proposition 3.4).

Associated to the bivariate tail-dependence coefficients $\lambda_X(\{i,j\})$ we introduce and discuss the so called spectral distance d_X given by

$$d_X(i,j) := \lambda_X(\{i\}) + \lambda_X(\{j\}) - 2\lambda_X(\{i,j\}).$$

This spectral distance defines a metric on the space of 1-Fréchet random variables (i.e. random variables with distribution function $F(x) = \exp\{-c/x\}, x \geq 0$, for some non-negative scale coefficient c, where we speak of a standard 1-Fréchet distribution if $c = 1$) living on a joint probability space, which metricizes convergence in probability and was considered in Davis & Resnick (1993),

Stoev & Taqqu (2005), Fiebig et al. (2017). In Section 4 we will establish the L^1-embeddability of this metric, which allows us to apply the rich theory about metric embeddings in the context of analyzing the tail-dependence coefficients.
In Section 4.2, utilizing the exceedence set representation of the bivariate tail-dependence coefficients and the L^1-embeddability of the spectral distance, we recover the equivalence of the L^1 and ℓ_1-embeddability as well as a probabilistic proof of the so-called cut-decomposition of ℓ_1-embeddable finite metric spaces. In this case, this decomposition turns out to be closely related to the Tawn-Molchanov model of an associated max-stable vector X (Proposition 4.5). When a given ℓ_1-embeddable metric has a unique cut-decomposition, it is called rigid (Deza & Laurent 1997). Rigidity of the spectral distance basically means that the bivariate tail-dependence coefficients Λ determine all higher order tail-dependence coefficients. In Theorem 4.11, we show that line metrics are rigid, which to the best of our knowledge is a new finding. In particular, we obtain that the bivariate tail-dependence coefficient matrices corresponding to line metrics determine the complete set of tail-dependence or, equivalently, extremal coefficients of X. Interestingly, the random set Θ corresponding to such line-metric tail-dependence is (after a suitable reordering of marginals) a random segment, more precisely a random set of the form $\{i, i+1, \dots, j-1, j\}$ for $1 \leq i \leq j \leq p$ with $i = 1$ or $j = p$. In general, the characterization of rigidity is computationally hard as it is equivalent to the characterization of the simplex faces of the cone of cut metrics (Deza & Laurent 1997).
The bivariate TD matrix Λ is a correlation matrix of a random vector. It is well-known, however, that not every correlation matrix with non-negative entries is a matrix of tail-dependence coefficients. The recent works of Fiebig et al. (2017), Embrechts et al. (2016), Krause et al. (2018), and Shyamalkumar & Tao (2020) among others have studied extensively various aspects of the class of TD matrices. One surprisingly difficult problem, referred to as the realizability problem, is checking whether a given matrix Λ is a valid TD matrix. The extensive study of Shyamalkumar & Tao (2020) proposed several practical and efficient algorithms for realizability. Moreover, Shyamalkumar & Tao (2020) conjectured that the realizability problem is NP-complete. In Section 5, we confirm their conjecture. We do so by exploiting the established connection to ℓ_1-embeddability, which allows us to utilize the rich theory on cuts and metrics outlined in the monograph of Deza & Laurent (1997). It is known that checking whether any given p-point metric space is ℓ_1-embeddable is a computationally hard problem in the NP-complete class.

The paper is structured as follows: In Section 2 we give an overview over several ways of modeling and measuring tail-dependence of a random vector, presented in a hierarchic fashion: First of all, multivariate regular variation allows for the most complete asymptotic description of the tail-behavior of (heavy-tailed) random vectors in terms of the tail measure, with a direct correspondence to the class of max-stable models as the natural representatives for each given tail measure. A more condensed description of tail-dependence is given by the values of special extremal dependence functionals like the extremal coefficients and tail-dependence coefficients. Finally, a rather coarse but popular description of the tail-dependence is given in form of those functions evaluated only at bivariate marginals, where the bivariate tail-dependence coefficients form the most prominent example.
In Section 3 we first discuss exceedance sets, as introduced above, and Bernoulli compatibility. Based on this interpretation we give a short introduction into generalized Tawn-Molchanov models. In Section 4 we explore the relationship between bivariate tail-dependence coefficients and the

spectral distance on the space of 1-Fréchet random variables. After a brief introduction into the concepts of metric embeddings of finite metric spaces we will show that the spectral distance is both L^1- and ℓ_1-embeddable, some consequences of which will be explored in Section 4.2 and Section 5. In Section 4.2 we introduce the concept of rigid metrics and prove that the building blocks of ℓ_1-embeddability, i.e. the line metrics, correspond to Tawn-Molchanov models with a special structure which is completely determine by this line metric.
Finally, in Section 5 we use known results about the computational complexity of embedding problems to show that the realization problem of a tail-dependence matrix is NP-complete. Some proofs are deferred to the Appendix A.

2 Regular variation, max-stability, and extremal dependence

In this section, we provide a concise overview of fundamental notions on multivariate regular variation and max-stable distributions, which underpin the study of tail-dependence.

2.1 Multivariate regular variation

The concept of multivariate regular variation is key to the unified treatment of the various tail-dependence notions we will consider. Much of this material is classic but we provide here a self-contained review tailored to our purposes. Many more details and insights can be found in Resnick (1987, 2007), Hult & Lindskog (2006), Basrak & Planinić (2019), Kulik & Soulier (2020) among other sources.

We start with a few notations. A set $A \subset \mathbb{R}^p$ is said to be bounded away from 0 if $0 \notin A^{\mathrm{cl}}$, i.e., $A \cap B(0,\varepsilon) = \emptyset$, for some $\varepsilon > 0$. Here A^{cl} is the closure of A and $B(x,r) := \{y \in \mathbb{R}^p : \|x-y\| < r\}$ is the ball of radius r centered at x in a given fixed norm $\|\cdot\|$. Furthermore, denote the Borel σ-Algebra on $\mathbb{R}^p$ by $\mathcal{B}(\mathbb{R}^p)$.
Consider the class $M_0(\mathbb{R}^p)$ of all Borel measures μ on $\mathcal{B}(\mathbb{R}^p)$ that are finite on sets bounded away from 0, i.e., such that $\mu(B(0,\varepsilon)^c) < \infty$, for all $\varepsilon > 0$. Such measures will be referred to as boundedly finite. For $\mu_n, \mu \in M_0(\mathbb{R}^p)$, we write

$$\mu_n \overset{\mathrm{M}_0}{\Longrightarrow} \mu, \quad \text{as } n \to \infty,$$

if $\int_{\mathbb{R}^p} f(x)\mu_n(dx) \to \int_{\mathbb{R}^p} f(x)\mu(dx)$, as $n \to \infty$, for all bounded and continuous f vanishing in a neighborhood of 0. The latter is equivalent to having

$$\mu_n(A) \to \mu(A), \quad \text{as } n \to \infty, \tag{2.1}$$

for all μ-continuity Borel sets A that are bounded away from 0 (Hult & Lindskog 2006, Theorems 2.1 and 2.4).

Definition 2.1. A random vector X in $\mathbb{R}^p$ is said to be regularly varying if there is a positive sequence $a_n \uparrow \infty$ and a non-zero $\mu \in M_0(\mathbb{R}^p)$ such that

$$n\mathbb{P}[X \in a_n \cdot] \overset{\mathrm{M}_0}{\Longrightarrow} \mu(\cdot), \quad \text{as } n \to \infty.$$

In this case, we write $X \in \mathrm{RV}(\{a_n\}, \mu)$ and call μ the tail measure of X.

If $X \in \mathrm{RV}(\{a_n\}, \mu)$, then it necessarily follows that there is a positive index $\alpha > 0$ such that

$$\mu(cA) = c^{-\alpha}\mu(A), \quad \text{for all } c > 0 \text{ and } A \in \mathcal{B}(\mathbb{R}^p), \tag{2.2}$$

and, moreover, $a_n \sim n^{1/\alpha}\ell(n)$, for some slowly varying function ℓ, see, e.g., Kulik & Soulier (2020), Section 2.1. We shall denote by $\mathrm{index}(X)$ the index of regular variation α and sometimes write $X \in \mathrm{RV}_\alpha(\{a_n\}, \mu)$ to specify that $\mathrm{index}(X) = \alpha$.
The measure μ is unique up to a multiplicative constant and the scaling property (2.2) implies that μ factors into a radial and an angular component. Namely, fix any norm $\|\cdot\|$ in $\mathbb{R}^p \setminus \{0\}$ and define the polar coordinates $r := \|x\|$ and $u := x/\|x\|,\ x \neq 0$. Then,

$$\mu(A) = \int_S \int_0^\infty 1_A(ru)\alpha r^{-\alpha-1} dr\sigma(du), \tag{2.3}$$

where $S := \{x : \|x\| = 1\}$ is the unit sphere and σ is a finite Borel measure on S referred to as the angular or spectral measure associated with μ, see, e.g., Kulik & Soulier (2020), Section 2.2. Given the norm $\|\cdot\|$, the measure σ is uniquely determined as

$$\sigma(B) = \mu(\{x : \|x\| > 1,\ x/\|x\| \in B\}), \quad B \in \mathcal{B}(S), \tag{2.4}$$

where $\mathcal{B}(A)$ for $A \subset \mathbb{R}^d$ denotes the d-dimensional Borel sets which are also subsets of A. The following is a useful characterization of regular variation sometimes taken as an equivalent definition, see again, e.g., Kulik & Soulier (2020), Section 2.2.

Proposition 2.2. *We have $X \in \mathrm{RV}_\alpha(\{a_n\}, \mu)$ if and only if for all $x > 0$*

$$n\mathbb{P}[\|X\| > a_n x] \to x^{-\alpha}, \quad \text{as } n \to \infty, \text{ and } \mathbb{P}\Big[\frac{X}{\|X\|} \in \cdot \mid \|X\| > r\Big] \Longrightarrow \sigma(\cdot), \quad \text{as } r \to \infty,$$

where $\Rightarrow$ denotes the weak convergence of probability distributions.

Proposition 2.2 characterizes regularly varying random vectors in terms of exceedances over a threshold. An equivalent charaterization is also possible in terms of maxima, see, e.g., Kulik & Soulier (2020), Section 2.1.

Proposition 2.3. *For a random vector $Y \in [0,\infty)^d$ we have $Y \in \mathrm{RV}_\alpha(\{a_n\}, \mu)$ if and only if there exists a non-degenerate random vector X such that for all $x \in [0,\infty)^d$*

$$\mathbb{P}\Big[a_n^{-1} \bigvee_{t=1}^n Y^{(t)} \le x\Big] \to \mathbb{P}[X \le x] = \exp\{-\mu[0,x]^c\}, \quad \text{as } n \to \infty,$$

where $[0,x]^c := \mathbb{R}^p_+ \setminus [0,x] = \mathbb{R}^p_+ \setminus ([0,x_1] \times \ldots \times [0,x_p])$ and $Y^{(t)},\ t = 1,\ldots,n$ are independent copies of Y and the operation $\vee$ denotes taking the component-wise maximum.

Multivariate regular variation provides an asymptotic framework and for given $\alpha, \{a_n\}$ and μ there exist several distributions of random vectors Y such that $Y \in \mathrm{RV}_\alpha(\{a_n\}, \mu)$, but according to Proposition 2.3 their maxima are all attracted to the same random vector X whose distribution depends only on μ. The class of limiting random variables in Proposition 2.3 will be inspected more closely in the next section.

2.2 Max-stable vectors

The homogeneity property (2.2) of μ implies that the limiting random vector in Proposition 2.3 has a certain stability property, namely that

$$\bigvee_{t=1}^{n} X^{(t)} \stackrel{d}{=} n^{1/\alpha} X \quad \text{for all } n \in \mathbb{N},$$

with the same notation as in Proposition 2.3 and where $\stackrel{d}{=}$ stands for equality in distribution, see Kulik & Soulier (2020), Section 2.1. We call such a random vector X max-stable and we call X non-degenerate max-stable if in addition $\mathbb{P}[X = (0, \dots, 0)] < 1$. For $\alpha = 1$ this simplifies to

$$\bigvee_{t=1}^{n} X^{(t)} \stackrel{d}{=} nX \quad \text{for all } n \in \mathbb{N}, \tag{2.5}$$

and we speak of a simple max-stable random vector X, which we will further analyze in the following.
The marginal distributions of simple max-stable distributions are necessarily 1-Fréchet, that is,

$$\mathbb{P}[X_i \le x] = e^{-\sigma_i/x}, \ x > 0,$$

for some non-negative scale coefficient σ_i. We shall write $\|X_i\|_1 := \sigma_i$ for the scale coefficient of the 1-Fréchet variable X_i. The next result characterizes all multivariate simple max-stable distributions. Here, we recall the so-called de Haan construction of a simple max-stable vector.

Proposition 2.4. *Let $(E, \mathcal{E}, \nu)$ be a measure space and let $L^1_+(E, \nu)$ denote the set of all non-negative ν-integrable functions on E. For every collection $f_i \in L^1_+(E, \nu)$, $1 \le i \le p$, there is a random vector $X = (X_i)_{1 \le i \le p}$, such that for all $x_i > 0, 1 \le i \le p$,*

$$\mathbb{P}[X_i \le x_i, \ 1 \le i \le p] = \exp\Big\{ - \int_E \max_{1 \le i \le p} \frac{f_i(u)}{x_i} \nu(du) \Big\}. \tag{2.6}$$

The random vector X is simple max-stable. Conversely, for every simple max-stable vector X, Equation (2.6) holds and $(E, \mathcal{E}, \nu)$ can be chosen as $([0,1], \mathcal{B}[0,1], \mathrm{Leb})$. In fact, we have the stochastic representation

$$(X_i)_{1 \le i \le p} \stackrel{d}{=} (I(f_i))_{1 \le i \le p}, \quad \text{with} \ \ I(f) := \bigvee_{j=1}^{\infty} \frac{f(U_j)}{\Gamma_j}, \ f \in L^1_+([0,1], \nu), \tag{2.7}$$

where $\{(\Gamma_j, U_j)\}$ is a Poisson point process on $(0, \infty) \times [0,1]$ with mean measure $dx \times \nu(du)$.

For a proof and more details, see e.g. de Haan (1984), Stoev & Taqqu (2005). The functions f_i in (2.6) and (2.7) are referred to as spectral functions associated with the vector X. From (2.6) and (2.7), one can readily see that for all $f \in L^1_+(E, \nu)$, the so-called extremal integral $I(f)$ in (2.7) is a well-defined 1-Fréchet random variable. More precisely, its cumulative distribution function is:

$$\mathbb{P}[I(f) \le x] = \exp\{-\|I(f)\|_1 / x\}, \ x > 0, \quad \text{where } \|I(f)\|_1 = \|f\|_{L^1} = \int_E f(u) \nu(du).$$

Moreover, the extremal integral functional $I(\cdot)$ is max-linear in the sense that for all $a_i \geq 0$ and $f_i \in L^1_+(E,\nu),\ 1 \leq i \leq n$, we have

$$I\Big(\bigvee_{t=1}^n a_t f_t\Big) = \bigvee_{t=1}^n a_t I(f_t).$$

Thus, every max-linear combination $\vee_{i=1}^n a_i X_i$ of X as above with coefficients $a_i \geq 0$ is a 1-Fréchet random variable with scale coefficient:

$$\Big\| \bigvee_{i=1}^n a_i X_i \Big\|_1 = \int_E \Big(\bigvee_{i=1}^n a_i f_i(u) \Big) \nu(du) = \Big\| \bigvee_{i=1}^n a_i f_i \Big\|_{L^1}.$$

We will further explore the asymptotic properties of simple max-stable random vectors and how they fit into the framework of multivariate regular variation in the following section.

2.3 Extremal dependence functionals and tail-dependence coefficients

The tail measure μ and the normalizing sequence $\{a_n\}$ from Section 2.1 provide a comprehensive description of the asymptotic behavior of a random vector X and allow to approximate probabilities of the form $\mathbb{P}[X \in a_n A]$ for all sets A bounded away from 0. Sometimes, however, one may be interested in those probabilities for certain simple sets A only and describe the asymptotic behavior of X by certain extremal dependence functions instead. In this section, we first derive a general result for such extremal dependence functions and then introduce two particularly popular families of them.

Proposition 2.5. *Let $X \in RV(\{a_n\}, \mu)$ in $\mathbb{R}^p$ with index $\alpha > 0$. Let also $h : \mathbb{R}^p \to [0,\infty)$ be a non-negative, continuous and 1-homogeneous function, i.e., $h(cx) = ch(x),\ c > 0,\ x \in \mathbb{R}^p$. Then,*

$$\lim_{n\to\infty} n\mathbb{P}[h(X) > a_n] = \int_S h(u)^\alpha \sigma(du) = \mathbb{E}[h(Y)^\alpha]\sigma(S), \tag{2.8}$$

where Y has probability distribution $\sigma(\cdot)/\sigma(S)$ and σ is as in (2.4).

Though this result is similar to Yuen et al. (2020), Lemma A.7, and also a special case to Dyszewski & Mikosch (2020), Theorem 2.1, its proof is given Section A.
We will apply the formula in (2.8) for homogeneous functionals of the form $h(x) = (\min_{i\in K} x_i)_+$ and $h(x) = (\max_{i\in K} x_i)_+$ for some subset $K \subset [p] = \{1,\dots,p\}$.
The next result shows that simple max-stable vectors are regularly varying and provides means to express their extremal dependence functionals *both* in terms of spectral functions *and* tail measures.

Proposition 2.6. *Let $X = (X_i)_{1\leq i\leq p}$ be a non-degenerate simple max-stable vector as in (2.6). Then, $X \in \mathrm{RV}_1(\{n\}, \mu)$, where μ is supported on $[0,\infty)^p$ and for all $x = (x_i)_{1\leq i\leq p} \in \mathbb{R}^p_+ \setminus \{0\}$*

$$\mathbb{P}[X \leq x] = \exp\{-\mu([0,x]^c)\}, \quad \text{with } \mu([0,x]^c) = \int_E \max_{1\leq i\leq p} \frac{f_i(u)}{x_i} \nu(du).$$

Moreover, for every non-negative, continuous 1-homogeneous function $h : \mathbb{R}^p \to [0,\infty)$, we have

$$\lim_{n\to\infty} n\mathbb{P}[h(X) > n] = \mu(\{h > 1\}) = \int_S h(u)\sigma(du) = \int_E h(\vec{f}(z))\nu(dz), \tag{2.9}$$

where $\vec{f}(z) = (f_1(z), \cdots, f_p(z))$. *In particular, the spectral measure* σ *has the representation*

$$\sigma(B) = \int_E 1_B\Big(\frac{\vec{f}(z)}{\|\vec{f}(z)\|}\Big)\|\vec{f}(z)\|\nu(dz), \quad B \in \mathcal{B}(S). \tag{2.10}$$

Again, this result is standard but we sketch its proof for the sake of completeness in Appendix A. The classic representation of the simple max-stable cumulative distribution functions is a simple corollary from Proposition 2.6.

Corollary 2.7. *In the situation of Proposition 2.6, by taking* $h(u) := h_x(u) := (\max_{i\in[p]} u_i/x_i)_+$ *for* $x_i \in (0,\infty)^p$ *in* (2.9)*, we obtain* $\mu(\{h > 1\}) = \mu([0,x]^c)$ *and*

$$\mathbb{P}[X \le x] = \exp\{-\mu([0,x]^c)\} = \exp\Big\{-\int_S \Big(\max_{i\in[p]} \frac{u_i}{x_i}\Big)\sigma(du)\Big\}. \tag{2.11}$$

For more details on the characterization of the max-domain of attraction of multivariate max-stable laws in terms of multivariate regular variation, see e.g., Proposition 5.17 in Resnick (1987).
We are now ready to recall the general definitions of the extremal and tail-dependence coefficients of a regularly varying random vector, which have briefly been introduced in Section 1, now with additional notation for the normalizing sequence $\{a_n\}$.

Definition 2.8. Let $X = (X_i)_{1\le i\le p} \in \mathrm{RV}(\{a_n\}, \mu)$. Then, for non-empty sets $K, L \subset [p]$, we let

$$\theta_X(K;\{a_n\}) := \lim_{n\to\infty} n\mathbb{P}\Big[\max_{i\in K} X_i > a_n\Big] \quad \text{and} \quad \lambda_X(L;\{a_n\}) := \lim_{n\to\infty} n\mathbb{P}\Big[\min_{i\in L} X_i > a_n\Big].$$

The $\theta_X(K;\{a_n\})$'s and $\lambda_X(L;\{a_n\})$'s are referred to as the *extremal and tail-dependence coefficients relative to* $\{a_n\}$ of the vector X, respectively.

If it is clear to which random vector we refer to or it does not matter for the argument, we may drop the index X and just write $\theta(K;\{a_n\})$ and $\lambda(K;\{a_n\})$. Sometimes we will view θ and λ as functions of k-tuples and write for example

$$\lambda_X(i_1,\cdots,i_k;\{a_n\}),\ 1 \le i_1,\ldots,i_k \le p,$$

(where some of the arguments $i_1,\ldots,i_k$ may repeat) which corresponds to $\lambda_X(L,\{a_n\})$ where L is the set of all distinct values in $\{i_1,\ldots,i_k\}$.

Remark 2.9. Note that the definitions of $\theta_X(K,\{a_n\})$ and $\lambda_X(L,\{a_n\})$ depend on the choice of the sequence $\{a_n\}$. They are unique, however, up to a multiplicative constant. More precisely, if $\mathrm{index}(X) = \alpha$ and $a_n \sim a'_n, c > 0$, then

$$\theta_X(K;\{ca_n\}) = c^{-\alpha}\theta_X(K;\{a'_n\}) \quad \text{as well as} \quad \lambda_X(L;\{ca_n\}) = c^{-\alpha}\lambda_X(L;\{a'_n\}).$$

Remark 2.10. In the following we will focus on extremal and tail-dependence coefficients of max-stable random vectors, which exist by Definition 2.8 in combination with Proposition 2.6 as long as X is non-degenerate. Observe that if X is non-degenerate simple max-stable, then

$$\lambda(i;\{n\}) = \theta(i;\{n\}) = \lim_{n\to\infty} n\mathbb{P}[X_i > n] = \lim_{n\to\infty} n(1-e^{-\sigma_i/n}) = \sigma_i = \|X_i\|_1,\ 1 \le i \le p.$$

Thus, if all marginals of X are standard 1−Fréchet, i.e., $\|X_i\|_1 = 1$, then setting $a_n = n$ ensures that $\lim_{n\to\infty} n\mathbb{P}[X_i > a_n] = 1$ and one recovers the upper tail-dependence coefficient $\lambda_X(i,j)$ from (1.1), $i,j \in [p]$. More generally, if X is non-degenerate simple max-stable, then we can choose $a_n = n$ as a normalizing sequence and in this case (or if the sequence $\{a_n\}$ does not matter for the argument), we will also write

$$\theta(K) = \theta_X(K) = \theta_X(K;\{a_n\}), \quad \lambda(L) = \lambda_X(L) = \lambda_X(L;\{a_n\}), \quad K, L \subset [p].$$

In the case that $\mathbb{P}[X = (0,\dots,0)] = 1$, we set $\theta_X(K) = \lambda_X(L) = 0$ for all $K, L \subset [p]$.

The following result expresses these functionals in terms of both the tail measure μ and the spectral functions of the vector X. Again, the proof is given in Appendix A.

Corollary 2.11. *Let $X = (X_i)_{1\le i\le p}$ be a simple max-stable vector as in* (2.6)*. Then,*

$$\theta_X(K) = \mu\Big(\bigcup_{i\in K} A_i\Big) \quad \textit{and} \quad \lambda_X(L) = \mu\Big(\bigcap_{i\in L} A_i\Big), \tag{2.12}$$

where $A_i := \{x \in \mathbb{R}^p : x_i > 1\}$ and

$$\theta_X(K) = \int_E \max_{i\in K} f_i(x)\nu(dx) \quad \textit{and} \quad \lambda_X(L) = \int_E \min_{i\in L} f_i(x)\nu(dx). \tag{2.13}$$

2.4 Bivariate tail-dependence measures and spectral distance

In Definition 2.8 we introduced general extremal and tail-dependence coefficients for arbitrary non-empty subsets $K, L \subset [p]$, i.e. for $2^p - 1$ different sets. Often these are too many coefficients for a handy description of the dependence structure. Therefore, one may consider only the pairwise dependence in a simple max-stable vector X which corresponds to the consideration of sets K and L with at most two entries. The set of tail-dependence coefficients with sets containing at most two elements can be written in the so called matrix of bivariate tail-dependence coefficients, which we denote by

$$\Lambda_X = \Lambda = (\lambda_X(i,j))_{1\le i,j\le p} = (\lambda_X(i,j;\{n\}))_{1\le i,j\le p}.$$

For the bivariate tail-dependence we have the alternative representation

$$\begin{aligned}\lambda_X(i,j) &= \lim_{n\to\infty} n\mathbb{P}[X_i > n, X_j > n] = \lim_{n\to\infty} n(\mathbb{P}[X_i > n] + \mathbb{P}[X_j > n] - \mathbb{P}[X_i \vee X_j > n]) \\ &= \|X_i\|_1 + \|X_j\|_1 - \|X_i \vee X_j\|_1. \end{aligned} \tag{2.14}$$

For standardized marginals $\|X_i\| = 1$ this implies $\lambda_X(i,j) = 2 - \|X_i \vee X_j\|_1$. The 1-Fréchet marginals of X imply

$$\mathbb{P}[X_i > n] \sim \frac{\|X_i\|_1}{n} \quad \text{and} \quad \mathbb{P}[X_i \vee X_j > n] \sim \frac{\|X_i \vee X_j\|_1}{n}$$

as $n \to \infty$, where $\|X_i \vee X_j\|_1$ denotes the scale coefficient of the 1-Fréchet distribution of $X_i \vee X_j$. Thus, for standardized marginals $\|X_i\|_1 = 1$, $1 \le i \le p$, the bivariate tail-dependence coefficients also have the following representation for all $1 \le i, j \le p$:

$$\lambda_X(i,j) = \lim_{u\to\infty} \mathbb{P}[X_i > u, X_j > u]/\mathbb{P}[X_i > u] = \lim_{n\to\infty} \mathbb{P}[X_j > n \mid X_i > n]. \tag{2.15}$$

In this form, the bivariate tail-dependence matrix is a popular measure for the extremal dependence in the random vector X. First appearing around the 60's (e.g. de Oliveira (1962)), the bivariate tail-dependence coefficients are frequently considered in the literature, see e.g. Coles et al. (1999), Beirlant et al. (2004), Frahm et al. (2005), Fiebig et al. (2017), Shyamalkumar & Tao (2020) for different considerations (sometimes other names as *coefficient of (upper) tail-dependence* or χ-*measure* are used). In the context of finance and insurance but also in an environmental context this measure is used to describe the extremal risk in the random vector X. Moreover, the characterization of whether X_i and X_j are extremally dependent is usually formulated by these bivariate tail-dependence coefficents: If $\lambda_X(i,j) = 0$, then X_i and X_j are extremally independent, otherwise the two random variables are extremally dependent.
Note that for standardized marginals the relation $\theta_X(i,j) = 2 - \lambda_X(i,j)$ holds. The extremal dependence coefficient in this form has often been used in the literature as a measure for extremal dependence, see e.g. Smith (1990), Schlather & Tawn (2003), Strokorb & Schlather (2015).
In all these references, the tail-dependence coefficient was defined as in (2.15) and standardized (or at least identically distributed) marginal distributions were assumed, as it is common for the analysis of dependence. However, we allow for unequal scales and therefore use the more general form (2.14).

Remark 2.12. The matrix of bivariate tail-dependence coefficients Λ of a simple max-stable vector is necessarily positive semi-definite. Indeed, this follows from the observation that by Corollary 2.11

$$\lambda(i,j) = \int f_i(x) \wedge f_j(x)\nu(dx) = \int \mathrm{Cov}(B(f_i(x)), B(f_j(x)))\nu(dx),$$

where $B = \{B(t),\ t \geq 0\}$ is a standard Brownian motion and since non-negative mixtures of covariance matrices are again covariance matrices. Another way to see this is from the observation that for each n, we have $n\mathbb{P}[X_i > n, X_j > n] = n\mathbb{E}[I(X_i > n)I(X_j > n)]$ is a positive semi-definite function of $i, j \in [p]$, which is related to the fact that $(i,j) \mapsto \lambda(i,j)$ is, up to a multiplicative constant, the covariance function of a certain random exceedance set (see Remark 3.6, below).

The matrix Λ is thus positive semi-definite, has non-negative entries and for standardized marginals of X it holds $\lambda(\{i\}) = 1$, i.e. Λ is a correlation matrix. However, not every correlation matrix with non-negative entries is necessarily a matrix of bivariate tail-dependence coefficients. The realization problem (i.e. the question whether a given matrix is the tail-dependence matrix of some random vector) is a recent topic in the literature (Fiebig et al. 2017, Krause et al. 2018, Shyamalkumar & Tao 2020). We will further discuss this problem in Section 5.
Related to the bivariate dependence coefficients we define an associated function, which will turn out to be a semi-metric on $[p]$.

Definition 2.13. Let $X = (X_i)_{1 \leq i \leq p}$ be a simple max-stable vector. Then, for $i, j \in [p]$, the spectral distance d_X is defined by

$$d_X(i,j) := d(X_i, X_j) := 2\|X_i \vee X_j\|_1 - \|X_i\|_1 - \|X_j\|_1. \tag{2.16}$$

By (2.14)

$$d_X(i,j) = \|X_i\|_1 + \|X_j\|_1 - 2\lambda_X(i,j) = \lambda_X(i) + \lambda_X(j) - 2\lambda_X(i,j). \tag{2.17}$$

If the scales of the marginals of the simple max-stable vector $(X_i)_{1\le i\le p}$ are the same, i.e. $\|X_i\| = c > 0$ for some $c > 0$ and all $1 \le i \le p$, then (2.17) simplifies to

$$d(i,j) = 2(c - \lambda_X(i,j)).$$

For standard 1-Fréchet marginals this further reduces to $d(i,j) = 2(1 - \lambda_X(i,j))$.
The spectral distance for max-stable vectors was already considered in Stoev & Taqqu (2005), equation (2.11). There it was shown that this distance is indeed a semi-metric on $[p]$ (Stoev & Taqqu 2005, Proposition 2.6) and that it metricizes convergence in probability in 1-Fréchet spaces (Stoev & Taqqu 2005, Proposition 2.4). In the form of (2.17), the spectral distance also appears in Fiebig et al. (2017), where it was defined in two steps in (Fiebig et al. 2017, Proposition 34 and 37). There, the use of the spectral distance is based on the fundamental work of (Deza & Laurent 1997, Section 5.2), where it is used in a different context.
In Section 4 we will prove that the spectral distance of a simple max-linear vector X is L^1-embeddable, with representation $d_X(i,j) = \|f_i - f_j\|_{L^1}$, where f_i, f_j are the spectral functions of X. In this form, the spectral distance was already used in Davis & Resnick (1989, 1993), where it was mainly applied for a projection method for prediction of max-stable processes. Davis & Resnick (1993) also gave a connection to the bivariate tail-dependence coefficients $\lambda(i,j)$ as considered in de Oliveira (1962), but only in the case of equally scaled marginals.

3 Tail-dependence via exceedance sets

In this section we develop a unified approach to representing tail-dependence via random exceedence sets, which explains and extends the notion of Bernoulli compatibility discovered in Embrechts et al. (2016) to higher order tail-dependence. Moreover, we introduce a slight extension of the so-called Tawn-Molchanov models and explore their connections to extremal and tail-dependence coefficients.

3.1 Bernoulli compatibility

We will first demonstrate that tail-dependence can be succinctly characterized via a random set obtained as the limit of exceedance sets. Let $X \in \mathrm{RV}_\alpha(\{a_n\}, \mu)$ and consider the exceedance set:

$$\Theta_n := \{i \,:\, X_i > a_n\}.$$

The asymptotic distribution of this random set, conditioned on it being non-empty can be directly characterized in terms of the extremal or tail-dependence coefficients of X. Specifically, these dependence coefficients can be seen as the hitting and inclusion functionals of a limiting random set Θ, respectively. For the precise definitions and related notions from the theory of random sets, we will always refer to the monograph of Molchanov (2017).

Before proceeding with the analysis of Θ we will introduce some appropriate coefficients. Let

$$\beta(J) := \mu(B_J) := \mu\Big(\bigcap_{j\in J} A_j \cap \bigcap_{k\in J^c} A_k^c\Big), \quad \emptyset \ne J \subset [p], \tag{3.1}$$

where again $A_i := \{x \in \mathbb{R}^p \,:\, x_i > 1\}, i \in [p]$. Then, in view of (2.12), since the B_J's are all pairwise disjoint in J,

$$\theta_X(K) = \sum_{J\,:\,J\cap K\ne\emptyset} \beta(J) \quad \text{and} \quad \lambda_X(L) = \sum_{J\,:\,L\subset J} \beta(J). \tag{3.2}$$

This, in view of the so-called Möbius inversion formula, see, e.g., Molchanov (2017), Theorem 1.1.61, yields the inversion formulae:

$$\beta(J) = \sum_{K\,:\,\emptyset\neq K,\ J^c\subset K} (-1)^{|J\cap K|+1}\theta_X(K), \tag{3.3}$$

which is Equation (7) in Schlather & Tawn (2003), Theorem 1. We also have

$$\beta(J) = \sum_{L\,:\,J\subset L\subset[p]} (-1)^{|L\setminus J|}\lambda_X(L). \tag{3.4}$$

Finally, the usual inclusion-exclusion type relationships hold between θ and λ:

$$\theta_X(K) = \sum_{L\,:\,\emptyset\neq L\subset K} (-1)^{|L|-1}\lambda_X(L) \quad \text{and} \quad \lambda_X(L) = \sum_{K\,:\,\emptyset\neq K\subset L} (-1)^{|K|-1}\theta_X(K). \tag{3.5}$$

Although some of the Relations (3.3), (3.4), and (3.5) are available in the literature, we prove them in Appendix A independently with elementary arguments in Lemma A.2.

Observe that the event $\{\Theta_n \cap K \neq \emptyset\}$ is $\{\max_{i\in K} X_i > a_n\}$. This immediately implies that

$$T_n(K) := \mathbb{P}[\Theta_n \cap K \neq \emptyset \mid \Theta_n \neq \emptyset] = \frac{\mathbb{P}[\max_{i\in K} X_i > a_n]}{\mathbb{P}[\max_{i\in[p]} X_i > a_n]} \longrightarrow \frac{\theta_X(K)}{\theta_X([p])}, \quad \text{as } n\to\infty.$$

The functionals $T_n(\cdot)$ are known as the hitting functionals of the conditional distribution of the random set Θ_n. They are completely alternating capacities and their limit yields hitting functionals $T(K) := \theta_X(K)/\theta_X([p])$ of a non-empty random set $\Theta \subset [p]$. This random set Θ may be viewed as the "typical" exceedance set for a regularly varying vector as the threshold a_n approaches infinity. It is immediate from (3.3) and Molchanov (2017), Corollary 1.1.31, that

$$\mathbb{P}[\Theta = J] = \frac{\beta(J)}{\sum_{\emptyset\neq K\subset[p]}\beta(K)}, \quad \emptyset \neq J \subset [p]. \tag{3.6}$$

Observing that $\theta_X([p]) = \sum_{\emptyset\neq K\subset[p]}\beta(K)$, we have thus established the following result.

Proposition 3.1. *Let $X \in \mathrm{RV}(\{a_n\},\mu)$ and define the random exceedance set $\Theta_n := \{i : X_i > a_n\}$. Then, as $n\to\infty$, we have*

$$\mathbb{P}[\Theta_n \in \cdot\,|\{\Theta_n \neq \emptyset\}] \Rightarrow \mathbb{P}[\Theta \in \cdot],$$

where the probability mass function of Θ is as in (3.6) *and the $\beta(J)$'s are as in* (3.1)*. We have moreover that*

$$\mathbb{P}[\Theta \cap K \neq \emptyset] = \frac{\theta_X(K)}{\theta_X([p])} \quad \textit{and} \quad \mathbb{P}[L \subset \Theta] = \frac{\lambda_X(L)}{\theta_X([p])}. \tag{3.7}$$

Remark 3.2. Molchanov & Strokorb (2016) introduced the important class of Choquet random sup-measures whose distribution is characterized by the extremal coefficient functional $\theta(\cdot)$. This is closely related but not identical to our perspective here, which emphasizes threshold-exceedance rather than max-stability.

The above result shows that all tail-dependence coefficients can be succinctly represented (up to a constant) via the random set Θ. This finding allows us to connect the tail-dependence coefficients to so-called Bernoulli-compatible tensors.

Definition 3.3. A k-tensor $T = (T(i_1, \cdots, i_k))_{1 \le i_1, \cdots, i_k \le p}$ is said to be Bernoulli-compatible, if

$$T(i_1, \cdots, i_k) = \mathbb{E}\Big[\xi(i_1) \cdots \xi(i_k)\Big], \tag{3.8}$$

where $\xi(1), \cdots, \xi(p)$ are (possibly dependent) Bernoulli 0 or 1-valued random variables, i.e. $P(\xi(i) = 1) = p_i = 1 - P(\xi(i) = 0)$ for some $p_i \in [0, 1], i \in [p]$. If not all $\xi(i)$'s are identically zero, the tensor T is said to be non-degenerate.

In the case $k = 2$, this definition recovers the notion of Bernoulli compatibility in Embrechts et al. (2016). Proposition 3.1 implies the following result.

Proposition 3.4.

(i) For every Bernoulli-compatible k-tensor $T = (T(i_1, \cdots, i_k))_{[p]^k}$, there exists a simple max-stable random vector X such that

$$T(i_1, \cdots, i_k) = \lambda_X(i_1, \cdots, i_k),$$

for all $i_1, \cdots, i_k \in [p]$.

(ii) Conversely, for every simple max-stable random vector $X = (X_i)_{1 \le i \le p}$, and every $c \ge \theta_X([p])$

$$(T(i_1, \cdots, i_k))_{[p]^k} := \frac{1}{c} \cdot \Big(\lambda_X(i_1, \cdots, i_k)\Big)_{[p]^k} \tag{3.9}$$

is a Bernoulli-compatible k-tensor.

Proof. (i) : Assume (3.8) holds and introduce the random set $\Theta := \{i : \xi(i) = 1\}$. Let $\beta(J) := \mathbb{P}[\Theta = J]$ and define the simple max-stable vector

$$X := \bigvee_{J\,:\,\emptyset \ne J \subset [p]} \beta(J) 1_J Z_J, \tag{3.10}$$

where $1_J = (1_J(i))_{1 \le i \le p}$ contains 1 in the coordinates in J and 0 otherwise and the Z_J's are iid standard 1-Fréchet. In view of Lemma A.1 and since $\lambda_{1_J Z_J}(L) = 1$ for $L \subset J$ and $\lambda_{1_J Z_J}(L) = 0$ for $L \not\subset J$, we have

$$\lambda_X(L) = \sum_{J\,:\,L \subset J} \beta(J) = \mathbb{P}[L \subset \Theta].$$

Since for $L = \{i_1, \cdots, i_k\}$ we have $1_{\{L \subset \Theta\}} = \prod_{j=1}^k \xi(i_j)$, we obtain

$$T(i_1, \cdots, i_k) = \mathbb{E}[\xi(i_1) \cdots \xi(i_k)] = \mathbb{P}[L \subset \Theta] = \lambda_X(L).$$

This completes the proof of (i).

(ii) : If $\mathbb{P}[X = (0, \dots, 0)] = 1$, then $\theta_X([p]) = 0$ and the statement follows by setting all $\xi(i_k)$ identically to 0, so assume $\mathbb{P}[X = (0, \dots, 0)] < 1$ in the following, which implies $\theta_X([p]) > 0$. Let $\Theta \subset [p]$ be a random set such that (3.7) holds, i.e.,

$$\lambda_X(L) = \theta_X([p]) \cdot \mathbb{P}[L \subset \Theta],\ L \subset [p].$$

Define $\xi(i) := B \cdot 1_\Theta(i)$, where B is a Bernoulli random variable, independent of Θ, such that $\mathbb{P}[B = 1] = 1 - \mathbb{P}[B = 0] = q \in (0, 1]$ for all $i \in [p]$. Then, we have that

$$\mathbb{E}[\xi(i_1) \cdots \xi(i_k)] = q\mathbb{E}[1_{\{i_1, \cdots, i_k\} \subset \Theta}] = \frac{q}{\theta_X([p])} \cdot \lambda_X(i_1, \cdots, i_k).$$

This shows that (3.9) holds with potentially any $c \geq \theta_X([p])$. □

Remark 3.5. As it can be seen from the proof the lower bound on the constant c in Proposition 3.4 (ii) cannot be improved. Observe that $\theta([p]) \leq \sum_{i \in [p]} \lambda_X(i)$, where the inequality is strict unless all X_i's are independent. Thus, the above result even in the case $k = 2$ improves upon Theorem 3 in Krause et al. (2018) where the range for the constant c is $c \geq \sum_{i \in [p]} \lambda_X(i)$.

Remark 3.6. In the case of two-point sets, we have that the bivariate tail-dependence coefficient

$$\lambda(i, j) = \theta([p]) \times \mathbb{P}[i, j \in \Theta] = \theta([p])\mathbb{E}[1_\Theta(i)1_\Theta(j)], \quad i, j \in [p], \tag{3.11}$$

is proportional to the so-called covariance function $(i, j) \mapsto \mathbb{P}[i, j \in \Theta] = \mathbb{E}[1_\Theta(i)1_\Theta(j)]$ of the random set Θ. This shows again that the bivariate tail-dependence function $(i, j) \mapsto \lambda(i, j)$ is positive semidefinite.

Remark 3.7. Relation (3.11) recovers a simple proof of the Bernoulli compatibility of TD matrices established in Theorem 3.3 of Embrechts et al. (2016). Namely, their result states that $\Lambda = (\lambda_{i,j})_{p \times p}$ is a matrix of bivariate tail-dependence coefficients, if and only if $\Lambda = c\mathbb{E}[\xi\xi^\top]$ for some $c > 0$ and a random vector $\xi = (\xi_i)_{1 \leq i \leq p}$ with Bernoulli entries taking values in $\{0, 1\}$. Clearly, there is a one-to-one correspondence between a random set $\Theta \subset [p]$ and a Bernoulli random vector: $\Theta := \{i : \xi_i = 1\}$ and $\xi = (1_\Theta(i))_{1 \leq i \leq p}$. The characterization result then follows from (3.11).

3.2 Generalized Tawn-Molchanov models

In the previous section we defined in (3.1) coefficients $\beta(J)$ to characterize the distribution of the limiting exceedance set Θ. These coefficients were then used in (3.10) to construct a max-stable random vector in order to prove Proposition 3.4. This special random vector is in fact nothing else than a generalized version of the so-called Tawn-Molchanov model which we will introduce formally in this section.

The following result is a slight extension and re-formulation of existing results in the literature, which have first appeared in Schlather & Tawn (2002, 2003) (see also Strokorb & Schlather (2015), Molchanov & Strokorb (2016) for extensions) in the context of finding necessary and sufficient conditions for a set of $2^p - 1$ numbers $\{\theta(K) \mid \emptyset \neq K \subset [p]\}$ to be the extremal coefficients of a max-stable vector X. The novelty here is that we consider max-stable vectors with possibly non-identical marginals and treat simultaneously the cases of extremal as well as tail-dependence coefficients.

Theorem 3.8. *The function* $\{\theta(K),\ K \subset [p]\}$ *(*$\{\lambda(L),\ L \subset [p]\}$*, respectively) yields the extremal (tail-dependence, respectively) coefficients of a simple max-stable vector* $X = (X_i)_{1\le i\le p}$ *if and only if the* $\beta(J)$*'s in* (3.3) *(*(3.4)*, respectively) are non-negative for all* $\emptyset \neq J \subset [p]$*. In this case, let* $Z_J, J \subset [p]$*, be iid standard* 1*-Fréchet random variables and define*

$$X^* = (X_i^*)_{1\le i\le p} := \bigvee_{\emptyset\neq J\subset[p]} \beta(J) 1_J Z_J, \tag{3.12}$$

where $1_J = (1_J(i))_{1\le i\le p}$ *contains* 1 *in the coordinates in* J *and* 0 *otherwise. Then,* X^* *is a max-stable random vector whose extremal (tail-dependence) coefficients are precisely the* $\theta(K)$*'s (*$\lambda(L)$*'s, respectively).*

The proof is given in Appendix A. The vector X^* defined in (3.12) is referred to as the Tawn-Molchanov or simply TM-model associated with the extremal (tail-dependence) coefficients $\{\theta(K)\}$ ($\{\lambda(L)\}$, respectively).

Remark 3.9. The distribution of the random set Θ introduced in Section 3.1 can be understood in terms of the Tawn-Molchanov model (3.12) using the single large jump heuristic. Given that $\Theta_n = \{i : X_i^* > n\} \neq \emptyset$, for large n, only one of the Z_J's is extreme enough to contribute to the exceedance set. Thus, with high probability, Θ_n equals the corresponding J in (3.12). The probability of the set J to occur is asymptotically proportional to the weight $\beta(J)$, which explains the formula (3.6).

We have seen in Section 2.4 that extremal dependence can also be measured in terms of spectral distance. In the following section we will explore further the connections between spectral distance and the just introduced Tawn-Molchanov models and see how the latter naturally lead to a decomposition of the former which is equivalent to ℓ_1-embeddability.

4 Embeddability and rigidity of the spectral distance

So far, we have mainly considered the overall tail-dependence of X or the tail-dependence function $\lambda(L)$ for arbitrary $L \subset [p]$. In this section we will focus on the bivariate dependence as in Section 2.4. Specifically, we look at the spectral distance and prove that it is both L^1- and, equivalently, ℓ_1-embeddable. For special spectral distances, namely those corresponding to line metrics, we prove that they are rigid and completely determine the tail-dependence of a TM-model.

4.1 L^1-embeddability of the spectral distance

Recall that a function $d : T \times T \to [0, \infty)$ on a non-empty set T is called a semi-metric on T if (i) $d(u, u) = 0$, $u \in T$ (ii) $d(u, v) = d(v, u)$, $u, v \in T$ and (iii) $d(u, w) \le d(u, v) + d(v, w)$, $u, v, w \in T$. The semi-metric is a metric if $d(u, v) = 0$ only if $u = v$.

Definition 4.1. A semi-metric d on a set T is said to be $L^1(E, \nu)$-embeddable (or short L^1-embeddable, when the measure space is understood) if there exists a collection of functions $f_t \in L^1(E, \nu)$, $t \in T$, such that

$$d(s, t) = \|f_s - f_t\|_{L^1} = \int_E |f_s(u) - f_t(u)|\nu(du),\ s, t \in T.$$

The concept of L^1-embeddability is extensively discussed in Deza & Laurent (1997). An overview can also be found in Matoušek (2013). Our first theorem in this section shows that the spectral distance matrix d_X of a max-stable vector X as defined in (2.16) is L^1-embeddable.

Theorem 4.2.

(i) For a simple max-stable vector X with bivariate tail-dependence coefficients $\lambda_{i,j} = \lambda_X(i,j)$, the spectral distance

$$d(i,j) := \lambda_{i,i} + \lambda_{j,j} - 2\lambda_{i,j} \tag{4.1}$$

(see Definition 2.13 and (2.17)) is an L^1-embeddable semi-metric.

(ii) Conversely, for every L^1-embeddable semi-metric d on $[p]$, there exists a simple max-stable vector X such that (4.1) holds with $\lambda_{i,j} := \lambda_X(i,j),\ 1 \le i,j \le p$. Moreover, there exists a $c \ge 0$ such that X may be chosen to have equal marginal distributions with $\|X_i\|_1 = c, i \in [p]$.

(iii) The semi-metric d in parts (i) and (ii) is a metric if and only if $\mathbb{P}[X_i \neq X_j] = 1$ for all $i \neq j$.

Proof. Part (i): Suppose that $X = (X_i)_{1\le i\le p}$ is simple max-stable and let $f_i \in L^1_+([0,1])$ be as in (2.6), where for simplicity and without loss of generality we choose $\nu =$Leb. In view of Relation (2.13), we obtain

$$\lambda_X(i,j) = \int_{[0,1]} f_i(x) \wedge f_j(x)dx,\ i,j \in [p].$$

Now the identity $|a-b| = a+b-2(a\wedge b)$ implies

$$\begin{aligned} d(i,j) &:= \int_{[0,1]} |f_i(x) - f_j(x)|dx = \int_{[0,1]} f_i(x)dx + \int_{[0,1]} f_j(x)dx - 2\int_{[0,1]} f_i(x)\wedge f_j(x)dx \\ &= \lambda_X(i,i) + \lambda_X(j,j) - 2\lambda_X(i,j). \end{aligned} \tag{4.2}$$

This shows that the semi-metric in (4.1) is L^1-embeddable. Note that d is a metric if and only if $f_i(\cdot) \neq f_j(\cdot)$, almost everywhere, or equivalently $X_i \neq X_j$ a.s., for all $i \neq j$.

Part (ii): Suppose now that $d(i,j) = \|g_i - g_j\|_{L^1}$ for some $g_i \in L^1(E,\nu),\ i \in [p]$. For simplicity and without loss of generality, we can assume that $(E,\mathcal{E},\nu) = ([0,1],\mathcal{B}[0,1],\text{Leb})$. Define the function $g^*(x) := \max_{i\in[p]} |g_i(x)|$ and let

$$f_i(x) = \begin{cases} g^*(2x) - g_i(2x) & ,\ x \in [0,1/2] \\ g^*(2x-1) + g_i(2x-1) & ,\ x \in (1/2,1]. \end{cases}$$

This way, we clearly have that the f_i's are non-negative elements of $L^1([0,1])$ and

$$\|f_i - f_j\|_{L^1} = \|g_i - g_j\|_{L^1} = d(i,j), \quad i,j \in [p].$$

Letting $X_i := I(f_i)$ be the extremal integrals defined in (2.7), we obtain as in (4.2) that

$$d(i,j) = \|f_i - f_j\|_{L^1} = \lambda_X(i,i) + \lambda_X(j,j) - 2\lambda_X(i,j), \quad i,j \in [p].$$

This proves the first claim in part (ii). It remains to argue that (with this particular choice of f_i's) the scales of the X_i's are all equal. Note that $\|X_i\|_1 = \|f_i\|_{L^1}$ and since

$$\int_0^{1/2} g^*(2x) - g_i(2x)dx = \frac{1}{2}\int_0^1 g^*(u) - g_i(u)du$$

$$\int_{1/2}^{1} g^*(2x-1) + g_i(2x-1)dx = \frac{1}{2}\int_0^1 g^*(u) + g_i(u)du,$$

we obtain $\|X_i\|_1 = \|f_i\|_{L^1} = \int_0^1 g^*(u)du$, for all $i \in [p]$, which completes the proof of part (ii).

Part (iii): The claim follows from the observation that $X_i := I(f_i) = I(f_j) =: X_j$ almost surely if and only if $f_i = f_j$ a.e., or equivalently, $\|f_i - f_j\|_{L^1} = 0$. □

Remark 4.3. The construction in the proof of part (ii) of Theorem 4.2 still works for f_i replaced by $\tilde{f}_i = f_i + \tilde{c}$ for any $\tilde{c} > 0$. Thus, the constant c can be chosen equal to or larger than $\int_0^1 g^*(u)du$, where $g^*(x) := \max_{i\in[p]} |g_i(x)|$ and $g_i \in L^1(E,\nu)$, $i \in [p]$ such that $d(i,j) = \|g_i - g_j\|_{L^1}$. In particular, for $\int_0^1 g^*(u)du \le 1$, one may choose X with standardized marginals, i.e. $\|X_i\|_1 = 1, i \in [p]$.

4.2 ℓ_1-embeddability of the spectral distance

In Theorem 4.2 we have shown the equivalence between L^1-embeddable metrics and spectral distances of simple max-stable vectors. In this section, we will additionally state an explicit formula for the ℓ_1-embedding of the spectral distance. Thereby we show that L^1- and ℓ_1-embeddability are equivalent and, in passing, we recover and provide novel probabilistic interpretations of the so-called cut-decomposition of ℓ_1-embeddable metrics (Deza & Laurent 1997).

Definition 4.4. A semi-metric d on T is said to be ℓ_1-embeddable in $(\mathbb{R}^m, \|\cdot\|_{\ell_1})$ (or short ℓ_1-embeddable) for some integer $m \ge 1$ if there exist $x_t = (x_t(k))_{1\le k\le m} \in \mathbb{R}^m$, $t \in T$, such that

$$d(i,j) = \|x_i - x_j\|_{\ell_1} = \sum_{k=1}^{m} |x_i(k) - x_j(k)| \quad \text{for all} \quad i,j \in T.$$

Proposition 4.5. *A semi-metric d on the finite set $[p]$ is embeddable in $L^1(E,\mathcal{E},\nu)$ if and only if*

$$d(i,j) = \sum_{J:\emptyset\neq J\subset[p]} \beta(J)|1_J(i) - 1_J(j)|, \quad i,j \in [p], \tag{4.3}$$

for some non-negative $\beta(J)$'s. This means that d is L^1-embeddable if and only if it is ℓ_1-embeddable in $\mathbb{R}^m$, where $m = |\mathcal{J}|$ and $\mathcal{J} = \{\emptyset \neq J \subset [p] : \beta(J) > 0\}$. Indeed, (4.3) is equivalent to $d(i,j) = \|x_i - x_j\|_{\ell_1}$, with $x_i = (x_i(J))_{J\in\mathcal{J}} := (\beta(J)1_J(i))_{J\in\mathcal{J}} \in \mathbb{R}^m_+$, $i,j \in [p]$.

Proof. By Theorem 4.2, d is L^1-embeddable if and only if (4.1) holds, where $\lambda_{i,j} = \lambda_X(\{i,j\})$ for some simple max-stable random vector. In view of (3.7) for the special case of $J = \{i\}$, using that $\mathbb{P}[J \subset \Theta] = \mathbb{E}[1_{\{J\subset\Theta\}}]$, we have

$$\frac{1}{\theta[p]} \cdot d(i,j) = \mathbb{P}[i \in \Theta] + \mathbb{P}[j \in \Theta] - 2\mathbb{P}[\{i,j\} \subset \Theta] = \mathbb{E}[|1_{\{i\in\Theta\}} - 1_{\{j\in\Theta\}}|]. \tag{4.4}$$

Taking X^* to be the (generalized) TM-model with matching extremal coefficients to those of X, by Relations (3.6) and (4.4) we obtain (4.3). □

Remark 4.6. Equation (4.4) shows that the spectral distance d is proportional to the probability that the limiting exceedance set Θ covers *one and only one* of the points i and j.

Remark 4.7. Proposition 4.5 recovers the well-known result that L^1- and ℓ_1-embeddability are equivalent (see Theorem 4.2.6 in Deza & Laurent 1997).

Proposition 4.5 also provides a *probabilistic* interpretation of the so-called cut-decomposition of ℓ_1-embeddable metrics. To connect to the rich literature on the subject, we will introduce some terminology following Chapter 4 of the monograph of Deza & Laurent (1997).
Let $J \subset [p]$ be a non-empty set and define the so-called cut semi-metric:

$$\delta(J)(i,j) = \begin{cases} 1 & , \text{ if } i \neq j \text{ and } |J \cap \{i,j\}| = 1 \\ 0 & , \text{ otherwise.} \end{cases} \tag{4.5}$$

The positive cone $\mathrm{CUT}_p := \{\sum_{J\subset[p]} c_J \delta(J),\ c_J \geq 0\}$ is referred to as the cut cone of non-negative functions defined on $[p]$. Notice that CUT_p consists of semi-metrics. Therefore, Proposition 4.5 entails that the cut cone CUT_p comprises all ℓ_1-embeddable metrics on p points (Proposition 4.2.2 in Deza & Laurent 1997). Relation (4.3), moreover, provides a decomposition of any such metric as a positive linear combination of cut semi-metrics. The coefficients of this decomposition are precisely the coefficients of *some* Tawn-Molchanov model. Finally, in view of (4.4), the random exceedance set Θ of this TM-model is such that

$$d(i,j) = \theta([p]) \cdot \mathbb{E}[|1_\Theta(i) - 1_\Theta(j)|].$$

Remark 4.8. For a given spectral distance d, Proposition 4.5 provides a decomposition and thereby shows the ℓ_1-embeddability of d in $\mathbb{R}^m$, where $m = |\mathcal{J}|$ and $\mathcal{J} = \{\emptyset \neq J \subset [p] : \beta(J) > 0\}$. Without further knowledge about the number of J such that $\beta(J) > 0$ we can always choose $m = 2^p - 2$, since we may set $\beta([p]) = 0$ as it does not affect d. However, by Caratheodory's theorem each ℓ_1-embeddable metric on $[p]$ is in fact known to be ℓ_1-embeddable in $\mathbb{R}^{\binom{p}{2}}$, see (Matoušek 2013, Proposition 1.4.2). We would like to mention that finding the corresponding "minimal" TM-model (i.e. the one with minimal $|\mathcal{J}|$) and analyzing the properties of such representations could be an interesting topic for further research.

Observe that

$$\delta(J)(i,j) = |1_J(i) - 1_J(j)| = |1_{J^c}(i) - 1_{J^c}(j)| = \delta(J^c)(i,j), \quad i,j \in [p],$$

where $J^c = [p] \setminus J$, which implies that, in general, the decomposition of d in Proposition 4.5 is not unique. Furthermore, $\beta([p]) \geq 0$ does not affect d in (4.3), since $|1_{[p]}(i) - 1_{[p]}(j)| = 0$. The next definition guarantees that, apart from those unavoidable ambiguities, the representation in (4.3) is essentially unique.

Definition 4.9. An ℓ_1-embeddable metric d is said to be rigid if for any two representations

$$d(i,j) = \sum_{J\,:\,\emptyset\neq J\subset[p]} \beta(J)|1_J(i) - 1_J(j)|, \quad i,j \in [p],$$

and

$$d(i,j) = \sum_{J\,:\,\emptyset\neq J\subset[p]} \tilde{\beta}(J)|1_J(i) - 1_J(j)|, \quad i,j \in [p],$$

with non-negative $\beta(J), \tilde{\beta}(J), \emptyset \neq J \subset [p]$, the equality

$$\beta(J) + \beta(J^c) = \tilde{\beta}(J) + \tilde{\beta}(J^c)$$

holds for all $\emptyset \neq J \subsetneq [p]$.

Observe that each semimetric d on p points can be identified with a vector $d = (d(i,j),\ 1 \le i < j \le p)$ in $\mathbb{R}^N$, where $N := \binom{p}{2}$. Thus, sets of such semimetrics can be treated as subsets of the Euclidean space $\mathbb{R}^N$. By Corollary 4.3.3 in Deza & Laurent (1997), the metric d is rigid, if and only if it lies on a simplex face of the cut-cone CUT_p. That is, if and only if the set $\{J_1, \cdots, J_m\} = \{\emptyset \neq J \subset [p] : \beta(J) > 0\}$ is such that the cut semimetrics $\delta(J_i),\ i = 1, \cdots, m$ (defined in (4.5)) lie on an affinely independent face of CUT_p. Recall that the points $\delta_i \in \mathbb{R}^N,\ i = 1, \cdots, m$ are affinely independent if and only if $\{\delta_i - \delta_1,\ i = 2, \cdots, m\}$ are linearly independent. In general, the description of the faces of the cut-cone is challenging, but the next section deals with a special class of metrics which are always rigid.

4.3 Rigidity of line metrics

In this section we show that so-called line metrics are rigid (cf. Definition 4.9) and that for spectral distances corresponding to line metrics the bivariate tail-dependence coefficients, in combination with the marginal distribution, fully determine the higher order tail-dependence coefficients of the underlying random vector and thus the coefficients of the corresponding Tawn-Molchanov model.

Definition 4.10. A metric d on $[p]$ is said to be a line metric if there exist a permutation $\pi = (\pi_i)_{1 \le i \le p}$ of $[p]$ and some weights $w_k \ge 0$, $1 \le k \le p-1$, such that

$$d(\pi_i, \pi_j) = \sum_{k=i}^{j-1} w_k.$$

In other words, d is a line metric if all points of $[p]$ can be ordered with different distances on some line and the distance between any two points equals the distance along that line.

Theorem 4.11. *Let d be a line metric, where without loss of generality the indices are ordered in such a way that for all $1 \le i < j \le p$ and some $w_k \ge 0$*

$$d(i,j) = \sum_{k=i}^{j-1} w_k. \tag{4.6}$$

(i) The line metric d is ℓ_1-embeddable and rigid.

Assume in addition that X follows a (generalized) TM-model as in (3.12) with given univariate $\lambda(i) = \lambda_{i,i}$ and bivariate tail-dependence coefficients $\lambda(i,j) = \lambda_{i,j}$ satisfying (4.1) with d as in (4.6). Then:

(ii) For every non-empty set $J \subset [p]$, we have

$$\lambda(J) = \lambda(i,j), \quad \text{where } i = \min(J) \text{ and } j = \max(J).$$

(iii) For the coefficients $\beta(J)$ of the (generalized) TM-model, we have that for all $1 \le k \le p-1$,

$$\beta([1:k]) = \lambda(k) - \lambda(k, k+1), \quad \beta([k+1:p]) = \lambda(k+1) - \lambda(k, k+1), \tag{4.7}$$

where $[i:j] := \{i, i+1, \ldots, j-1, j\}, i < j \in [p]$,

$$\beta([p]) = \lambda(1,p), \tag{4.8}$$

and $\beta(J) = 0$ for all other $J \subset [p]$.

Proof. Part (i): To see that d is ℓ_1-embeddable, set $\beta([1:k]) = w_k, k \in [p-1]$, and $\beta(J) = 0$ for all other sets $\emptyset \neq J \subset [p]$, which gives

$$d(i,j) = \sum_{k=i}^{j-1} w_k = \sum_{k=i}^{j-1} \beta([1:k]) = \sum_{J\,:\,\emptyset \neq J \subset [p]} \beta(J)|1_J(i) - 1_J(j)| = \sum_{J\,:\,\emptyset \neq J \subset [p]} \beta(J)\delta(J).$$

Thus, d is ℓ_1-embeddable by Proposition 4.5.
Let now $\beta(J), \emptyset \neq J \subset [p]$ be the coefficients of a representation (4.3) of d. We will show that

$$\beta(J) > 0 \;\; \Rightarrow \;\; J = [1:k] \text{ or } J = [k:p] \text{ for some } k \in [p]. \tag{4.9}$$

To this end, note that (4.6) implies, for any $i \leq j \in [p]$, that $d(i,j) = \sum_{k=i}^{j-1} d(k,k+1)$ and thus

$$\sum_{J\,:\,\emptyset \neq J \subset [p]} \beta(J)|1_J(i) - 1_J(j)| = \sum_{k=i}^{j-1} \sum_{J\,:\,\emptyset \neq J \subset [p]} \beta(J)|1_J(k) - 1_J(k+1)|,$$

or, equivalently,

$$\sum_{J\,:\,\emptyset \neq J \subset [p]} \beta(J) \left(|1_J(i) - 1_J(j)| - \sum_{k=i}^{j-1} |1_J(k) - 1_J(k+1)| \right) = 0. \tag{4.10}$$

Since

$$\sum_{k=i}^{j-1} |1_J(k) - 1_J(k+1)| \geq |1_J(i) - 1_J(j)|$$

and all $\beta(J)$ are non-negative, (4.10) implies that

$$|1_J(i) - 1_J(j)| = \sum_{k=i}^{j-1} |1_J(k) - 1_J(k+1)|$$

for those J with $\beta(J) > 0$ and all $i \leq j \in [p]$. This leads to the following four possible cases:

(i) If $1, p \in J$, then $J = [p]$.

(ii) If $1, p \in J^c$, then $J = \emptyset$.

(iii) If $1 \in J, p \in J^c$, then there exists one $k \in [p]$ such that $J = [1:k]$.

(iii) If $1 \in J^c, p \in J$, then there exists one $k \in [p]$ such that $J = [k:p]$.

We have thus shown (4.9) and in order to show that d is rigid, we only need to consider sets of the form $J = [1:k], J^c = [k+1:p], k \in [p-1]$. For those sets we get

$$\beta([1:k]) + \beta([k+1:p]) = \sum_{J:\emptyset \neq J \subset [p]} \beta(J)|1_J(k) - 1_J(k+1)| = d(k,k+1) = w_k, \tag{4.11}$$

and thus the sum $\beta(J) + \beta(J^c) = w_k$ is invariant for all representations (4.3) of d and d is rigid.

Part (ii): Let $\emptyset \neq J \subset [p]$ and set $i = \min(J), j = \max(J)$. Then, from part *(i)* and (3.2),

$$\begin{aligned}\lambda(J) &= \sum_{K:J\subset K} \beta(K) = \sum_{k\in[p]:J\subset[1:k]} \beta([1:k]) + \sum_{k\in[2:p]:J\subset[k:p]} \beta([k:p]) \\ &= \sum_{k=j}^{p} \beta([1:k]) + \sum_{k=1}^{i} \beta([k:p]) \\ &= \sum_{k\in[p]:i,j\in[1:k]} \beta([1:k]) + \sum_{k\in[2:p]:i,j\in[k:p]} \beta([k:p]) = \lambda(\{i,j\}) = \lambda(i,j),\end{aligned}$$

where we used the fact that $\beta(J) = 0$, for all $J \subset [2 : p-1]$ established in the proof of part *(i)*. This completes the proof of *(ii)*.

Part (iii): We have from (4.11) that

$$\beta([1:k]) + \beta([k+1:p]) = d(k,k+1) = \lambda(k) + \lambda(k+1) - 2\lambda(k,k+1),$$

and it follows for $k \in [1 : p-1]$ by *(i)* and (3.2) that

$$\begin{aligned}\lambda(k) - \lambda(k+1) &= \sum_{k\in J} \beta(J) - \sum_{k+1\in J} \beta(J) \\ &= \sum_{j=k}^{p} \beta([1:j]) + \sum_{j=1}^{k} \beta([j:p]) - \sum_{j=k+1}^{p} \beta([1:j]) - \sum_{j=1}^{k+1} \beta([j:p]) \\ &= \beta([1:k]) - \beta([k+1:p]).\end{aligned}$$

Together, this gives (4.7). Furthermore, (4.8) follows from

$$\lambda(1,p) = \sum_{J:1,p\in J} \beta(J) = \beta([1:p]).$$

That $\beta(J) = 0$ if J is not of the form $[1:k]$ or $[k:p], k \in p$, has already been shown in (i).

□

Remark 4.12. Consider a max-stable vector X with standard 1-Fréchet marginals, i.e., $\|X_i\|_1 = \lambda_X(i) = 1,\ i \in [p]$. Theorem 4.11 shows that if the spectral distance $d_X(i,j) = 2(1-\lambda_X(i,j)),\ i,j \in [p]$ is a line metric on $[p]$, then

$$\beta([1:k]) = \beta([k+1:p]) = 1 - \lambda_X(k,k+1),\ 1 \le k \le p-1,\ \ \beta([1:p]) = \lambda_X(1,p),$$

and for all other $\emptyset \neq J \subset [p], \beta(J) = 0$. In particular, all higher order extremal coefficients of X are then completely determined by the bivariate tail-dependence coefficients and given from (3.2) by

$$\begin{aligned}\theta_X(K) &= \sum_{J:J\cap K\neq\emptyset} \beta(J) = \sum_{j=\min K}^{p} \beta([1:j]) + \sum_{j=1}^{\max K} \beta([j:p]) - \beta([1:p]) \\ &= \sum_{j=\min K}^{p} (1 - \lambda_X(j,j+1)) + \sum_{j=1}^{\max K} (1 - \lambda_X(j,j+1)) - \lambda_X(1,p).\end{aligned}$$

Remark 4.13. The random set Θ corresponding to such line-metric tail-dependence is a random segment with one of its endpoints anchored at 1 or p. This is a direct consequence of the characterisation of $\beta(J)$ in from Theorem 4.11 *(iii)* and (3.6).

Remark 4.14. In practical applications, the non-parametric inference on higher-order tail-dependence coefficients can be very challenging or virtually impossible. Only, say, the bivariate tail-dependence coefficients $\Lambda = (\lambda_X(i,j))_{p\times p}$ of the vector X may be estimated well. Given such constraints, one may be interested in providing upper and lower bounds on $\lambda_X(\{1,\cdots,p\})$, which provide the worst- and best-case scenarios for the probability of simultaneous extremes.
If the spectral distance turns out to be a line metric and the marginal distributions are known, then Theorem 4.11 provides a way to precisely calculate $\lambda_X(\{1,\cdots,p\})$. However, in general this problem falls in the framework of computational risk management (see e.g. Embrechts & Puccetti 2010) as well as the distributionally robust inference perspective (see, e.g. Yuen et al. 2020, and the references therein). The problem can be stated as a linear optimization problem in dimension $2^p - 1$, similar to the approach in Yuen et al. (2020). Unfortunately, the exponential growth of complexity of the problem makes it computationally intractable for $p \geq 15$. In fact, the exact solution to such types of optimization problems may be NP-hard. This underscores the importance of the line of research initiated by Shyamalkumar & Tao (2020) where new approximate solutions or model-regularized approaches to distributionally robust inference in high-dimensional extremes are of great interest.

5 Computational complexity of decision problems

In this section we will use known results about the algorithmic complexity of ℓ_1-embeddings to derive that the so-called tail dependence realization problem is NP-complete, thereby confirming a conjecture from Shyamalkumar & Tao (2020). While a formal introduction to the theory of algorithmic complexity is beyond the scope of this paper, we shall informally recall the basic notions needed in our context following the treatment in (Deza & Laurent 1997, Section 2.3).
Consider a class of computational problems D, where each instance $\mathcal{I}$ of D can be encoded with a finite number of bits $|\mathcal{I}|$. D is said to be a decision problem, if for any input instance $\mathcal{I}$ there is a correct answer, which is either "yes" or "no". The goal is to determine this answer based on any input $\mathcal{I}$ by using a computer (i.e., a deterministic Turing machine).
The decision problem D is said to belong to:

- The class **P** (for polynomial complexity), if there is an algorithm (i.e., a deterministic Turing machine), that can produce the correct answer in polynomial time, i.e. its running time is of the order $\mathcal{O}(|\mathcal{I}|^k)$ for some $k \in \mathbb{N}$.

- The class **NP** (nondeterministic polynomial time) if the problem admits a polynomially-verifiable positive certificate. More precisely, this means that for each instance $\mathcal{I}$ of D with positive ("yes") answer, there exists a finite-bit certificate $\mathcal{C}$ of size $|\mathcal{C}|$ that can be verified by an algorithm / deterministic Turing machine with running time $\mathcal{O}(|\mathcal{C}|^l)$ for some $l \in \mathbb{N}$. (The certificate needs not be constructed in polynomial time.)

- The class **NP-hard** if *any* problem in NP reduces to D in polynomial time. This means that for every problem D' in NP, the correct answer to this decision problem for any instance $\mathcal{I}'$ of D' can be found by first applying an algorithm that runs in polynomial time of $|\mathcal{I}'|$ to

transform $\mathcal{I}'$ into an instance $\mathcal{I}$ of D and then solve the decision problem D for this instance $\mathcal{I}$. Note that this definition does not require that D itself is in NP.

- The class **NP-complete** if D is both in NP and is NP-hard.

A decision problem which has received some attention recently, see Fiebig et al. (2017), Embrechts et al. (2016), Krause et al. (2018), and Shyamalkumar & Tao (2020), is the realization problem of a TD matrix with standardized entries on the diagonal, namely finding an algorithm with the following input and output:

Tail dependence realization (TDR) problem

- Input: A non-negative, symmetric $p \times p$-matrix $L = (L_{i,j})$ with $L_{i,i} = 1, i \in [p]$.
- Output: An answer to the question: Does there exist a simple max-stable vector $X = (X_1, \ldots, X_p)$ such that
$$\lambda_X(i,j) = L_{i,j}, \quad i,j \in [p],$$
i.e. L is the matrix of bivariate tail-dependence coefficients of a max-stable vector with standard 1-Fréchet-margins?

This problem may at a first glance look similar to deciding whether a given matrix is a valid covariance matrix. Indeed, as a strengthening of Remark 3.7, it can be shown that there exists a bijection between TD matrices as in the above problem and a subset of the so-called Bernoulli-compatible random matrices, i.e. expected outer products $E(YY^t)$ of random (column) vectors Y with Bernoulli margins, see Embrechts et al. (2016) and Fiebig et al. (2017). But while it is a simple task to check if a matrix is the covariance matrix of some random vector, for example by finding the eigenvalues of this matrix, it can become more difficult to check whether a matrix is the covariance matrix or outer product of a restricted space of random variables. Practical and numerical aspects of deciding whether a given matrix is a TD matrix have been studied in Krause et al. (2018) and Shyamalkumar & Tao (2020), including a discussion on the computational complexity of the problem. Indeed, they point out that due to results by Pitowsky (1991), checking whether a matrix is Bernoulli-compatible is an NP-complete problem. However, some subtlety arises as in order to check whether a $p \times p$-matrix L is a so-called tail coefficient matrix, i.e. a TD matrix with 1's on the diagonal, it needs to be checked that $p^{-1}L$ is Bernoulli-compatible, see Shyamalkumar & Tao (2020). Thus, the problem narrows down to checking Bernoulli compatibility of the subclass of matrices with $1/p$ on their diagonal and this may have a different complexity than the general membership problem. Due to the similarity in the above mentioned problems, Shyamalkumar & Tao (2020) conjecture that the TDR problem is NP-complete as well.
We add to the discussion by using results about computational complexity of problems related to cut metrics and metric embeddings, see Section 4.4 in Deza & Laurent (1997) for a brief overview over some relevant results. To this end, let us first introduce a problem which is related to the TDR problem but easier to handle for the subsequent complexity analysis.

Spectral distance realization (SDR) problem with unconstrained, identical margins

- Input: A non-negative, symmetric $p \times p$-matrix $d = (d(i,j))_{p\times p}$.
- Output: An answer to the question: Does there exist a simple max-stable vector $X = (X_1, \dots, X_p)$ and some $c > 0$ such that $\|X_i\|_1 = c, i \in [p]$, and

$$d(i,j) = 2(c - \lambda_X(i,j)), \quad i,j \in [p], \tag{5.1}$$

  i.e. d is the spectral distance of X?

With the help of our previous results and the known computational complexity of ℓ_1-embeddings it is simple to establish the computational complexity of the above problem.

Theorem 5.1. *The SDR problem with unconstrained, identical margins is NP-complete.*

Proof. Due to Theorem 4.2 (i)-(ii), the spectral distance $d(i,j) = 2(c - \lambda_X(i,j))$ of a simple max-stable random vector with $\|X_i\|_1 = c, i \in [p]$, is L^1-embeddable and for each L^1-embeddable semi-metric d there exists a simple max-stable vector X with $\|X_i\|_1 = c, i \in [p]$, for some $c > 0$ such that d is the spectral distance of X. Thus, the question is equivalent to checking that d is L^1-embeddable and this is equivalent to checking that d is ℓ_1-embeddable, see Remark 4.7. The latter problem is NP-complete by Avis & Deza (1991), see also (P5) in Deza & Laurent (1997). □

Remark 5.2. In the SDR problem one could add more assumptions about d in the first place under "Input", for example that the entries on the diagonal of d are equal to 0 or that d is a distance matrix. Alternatively, one could also just assume under "Input" that d is a $p \times p$-matrix. Since a positive answer to the question would always ensure that d is a distance matrix and all mentioned properties (non-negativity, symmetry, triangle inequality) could be checked in a number of steps which is a polynomial in p these additional assumptions do not change the NP-completeness of the problem.

Unfortunately, the constant c in (5.1) is not part of the input in the algorithm and thus cannot be fixed a priori. If we could for example set $c = 1$ and thus ask if for a given d a simple max-stable vector X with standard 1-Fréchet-margins exists such that $d(i,j) = 2(1 - \lambda_X(i,j))$, then this is equivalent to checking that $\lambda_{i,j} := 1 - d(i,j)/2$ is a TD matrix. But while such an arbitrary fixation of c may change the nature of the problem, the following statement points out an a posteriori feasible range for c.

Lemma 5.3. *If the outcome of the SDR problem with unconstrained, identical margins is a positive answer to the question, then* (5.1) *holds for a suitable chosen max-stable vector X and every $c \geq (2^p - 2) \max_{i,j\in[p]} d(i,j)$.*

The proof is given in Appendix A. From the previous lemma we see that the SDR problem with unconstrained, identical margins is equivalent to

Spectral distance realization (SDR) problem with d-constrained margins

- Input: A non-negative, symmetric $p \times p$-matrix $d = (d(i,j))_{p\times p}$.
- Output: An answer to the question: Does there exist a simple max-stable vector $X = (X_1, \ldots, X_p)$ such that $\|X_i\|_1 = (2^p - 2)\max_{i,j\in[p]} d(i,j), i \in [p]$, and

$$d(i,j) = 2((2^p - 2)\max_{i,j\in[p]} d(i,j) - \lambda_X(i,j)), \quad i,j \in [p],$$

  i.e. d is the spectral distance of X?

Finally, by changing from X to $\tilde{X} := X/((2^p - 2)\max_{i,j\in[p]} d(i,j))$ the spectral distance $d_{\tilde{X}}$ of $\tilde{X}$ and bivariate tail-dependence coefficients $\lambda_{\tilde{X}}(i,j)$ scale accordingly by Lemma A.1 and we see that the latter problem is actually equivalent to

Spectral distance realization (SDR) problem with constrained, standard margins

- Input: A non-negative, symmetric $p \times p$-matrix $d = (d(i,j))_{p\times p}$.
- Output: An answer to the question: Does there exist a simple max-stable vector $X = (X_1, \ldots, X_p)$ such that $\|X_i\|_1 = 1, i \in [p]$, and

$$\frac{d(i,j)}{(2^p - 2)\max_{i,j\in[p]} d(i,j)} = 2(1 - \lambda_X(i,j)), \quad i,j \in [p],$$

  i.e. $\lambda(i,j) := 1 - d(i,j)/(2(2^p - 2)\max_{i,j\in[p]} d(i,j))$ is the matrix of bivariate tail-dependence coefficients of a max-stable vector with standard 1-Fréchet-margins?

From the last line in the above problem we can see that our SDR problem with constrained, standard margins can be solved if we have an algorithm to check that λ of the given form is a TD matrix. But since we know by the stated equivalence of all three SDR problems in combination with Theorem 5.1 that all of them are NP-complete, we know that this algorithm has to be NP-complete as well. This leads to the following result.

Theorem 5.4. *The TDR problem is NP-complete.*

Proof. We need to show that the TDR problem is both in NP and NP-hard. That the TDR problem is in NP has been shown in (Shyamalkumar & Tao 2020, p. 255), with the help of Caratheodory's theorem. We start with the first statement and follow the typical way to prove this by reducing a known NP-complete problem to TDR. Indeed, any input matrix $d(i,j)$ to any of the three equivalent, and by Theorem 5.1 NP-complete, SDR problems can be transformed in polynomial time to the matrix $\lambda(i,j) := 1 - d(i,j)/(2(2^p - 2)\max_{i,j\in[p]} d(i,j))$. By the statement of the third SDR problem, the question with input d can be answered by using λ as an input to the TDR problem. Thus, an NP-complete problem reduces in polynomial time to the TDR problem and the TDR problem is NP-hard, thus NP-complete.

□

References

Avis, D. & Deza, M. (1991), 'The cut cone, L^1 embeddability, complexity, and multicommodity flows', Networks **21**(6), 595–617.
URL: *https://doi.org/10.1002/net.3230210602*

Basrak, B. & Planinić, H. (2019), 'A note on vague convergence of measures', Statist. Probab. Lett. **153**, 180–186.
URL: *https://doi.org/10.1016/j.spl.2019.06.004*

Beirlant, J., Goegebeur, Y., Teugels, J. & Segers, J. (2004), Statistics of extremes, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester.
URL: *https://doi.org/10.1002/0470012382*

Castillo, E. (1988), Extreme value theory in engineering, Statistical Modeling and Decision Science, Academic Press, Inc., Boston, MA.
URL: *https://doi.org/10.1016/c2009-0-22169-6*

Coles, S. (2001), An introduction to statistical modeling of extreme values, Springer Series in Statistics, Springer-Verlag London, London.
URL: *http://dx.doi.org/10.1007/978-1-4471-3675-0*

Coles, S., Heffernan, J. & Tawn, J. (1999), 'Dependence measures for extreme value analyses', Extremes **2**, 339–365.
URL: *https://doi.org/10.1023/A:1009963131610*

Davis, R. A. & Resnick, S. I. (1989), 'Basic properties and prediction of max-ARMA processes', Adv. in Appl. Probab. **21**(4), 781–803.
URL: *https://doi.org/10.2307/1427767*

Davis, R. A. & Resnick, S. I. (1993), 'Prediction of stationary max-stable processes', Ann. Appl. Probab. **3**(2), 497–525.
URL: *https://doi.org/10.1214/aoap/1177005435*

de Haan, L. (1984), 'A spectral representation for max–stable processes', Annals of Probability **12**(4), 1194–1204.
URL: *https://doi.org/10.1214/aop/1176993148*

de Haan, L. & Ferreira, A. (2007), Extreme value theory: an introduction, Springer Science & Business Media.
URL: *https://doi.org/10.1007/0-387-34471-3*

de Oliveira, J. T. (1962), 'Structure theory of bivariate extremes extensions', Estudos de Math. Estat. Econom. **7**, 165–195.

Deza, M. M. & Laurent, M. (1997), Geometry of cuts and metrics, Vol. 15 of Algorithms and Combinatorics, Springer-Verlag, Berlin.
URL: *https://doi.org/10.1007/978-3-642-04295-9*

Dyszewski, P. & Mikosch, T. (2020), 'Homogeneous mappings of regularly varying vectors', Ann. Appl. Probab. **30**, 2999–3026.
URL: *https://doi.org/10.1214/20-AAP1579*

Embrechts, P., Hofert, M. & Wang, R. (2016), 'Bernoulli and tail-dependence compatibility', Ann. Appl. Probab. **26**(3), 1636–1658.
URL: *https://doi.org/10.1214/15-AAP1128*

Embrechts, P. & Puccetti, G. (2010), 'Bounds for the sum of dependent risks having overlapping marginals', J. Multivariate Anal. **101**(1), 177–190.
URL: *https://doi.org/10.1016/j.jmva.2009.07.004*

Fiebig, U.-R., Strokorb, K. & Schlather, M. (2017), 'The realization problem for tail correlation functions', Extremes **20**(1), 121–168.
URL: *https://doi.org/10.1007/s10687-016-0250-8*

Finkenstädt, B. & Rootzén, H. (2003), Extreme Values in Finance, Telecommunications, and the Environment, Monographs on Statistics and Applied Probability, Chapman & Hall CRC Press, New York.
URL: *https://doi.org/10.1201/9780203483350*

Frahm, G., Junker, M. & Schmidt, R. (2005), 'Estimating the tail-dependence coefficient: properties and pitfalls', Insurance Math. Econom. **37**(1), 80–100.
URL: *https://doi.org/10.1016/j.insmatheco.2005.05.008*

Hult, H. & Lindskog, F. (2006), 'Regular variation for measures on metric spaces', Publ. Inst. Math. (Beograd) (N.S.) **80(94)**, 121–140.
URL: *http://dx.doi.org/10.2298/PIM0694121H*

Krause, D., Scherer, M., Schwinn, J. & Werner, R. (2018), 'Membership testing for bernoulli and tail-dependence matrices', Journal of Multivariate Analysis **168**, 240–260.
URL: *https://doi.org/10.1016/j.jmva.2018.07.014*

Kulik, R. & Soulier, P. (2020), Heavy-tailed time series, Springer Series in Operations Research and Financial Engineering, Springer, New York.
URL: *https://doi.org/10.1007/978-1-0716-0737-4*

Matoušek, J. (2013), Lecture notes on metric embeddings, Technical report, Institute of Theoretical Computer Science, ETH Zürich.
URL: *https://kam.mff.cuni.cz/~matousek/ba-a4.pdf*

Molchanov, I. (2017), Theory of random sets, Vol. 87 of Probability Theory and Stochastic Modelling, Springer-Verlag, London. Second edition.
URL: *https://doi.org/10.1007/978-1-4471-7349-6*

Molchanov, I. & Strokorb, K. (2016), 'Max-stable random sup-measures with comonotonic tail dependence', Stochastic Process. Appl. **126**(9), 2835–2859.
URL: *https://doi.org/10.1016/j.spa.2016.03.004*

Pitowsky, I. (1991), 'Correlation polytopes: their geometry and complexity', Mathematical Programming **50**(1), 395–414.
URL: *https://doi.org/10.1007/BF01594946*

Rachev, S. T. (2003), Handbook of heavy tailed distributions in finance: Handbooks in finance, Book 1, Elsevier.
URL: *https://doi.org/10.1016/B978-0-444-50896-6.X5000-6*

Resnick, S. I. (1987), Extreme Values, Regular Variation and Point Processes, Springer-Verlag, New York.
URL: *https://doi.org/10.1007/978-0-387-75953-1*

Resnick, S. I. (2007), Heavy-tail phenomena, Springer Series in Operations Research and Financial Engineering, Springer, New York. Probabilistic and statistical modeling.
URL: *https://doi.org/10.1007/978-0-387-45024-7*

Schlather, M. & Tawn, J. (2002), 'Inequalities for the extremal coefficients of multivariate extreme value distributions', Extremes **5**(1), 87–102.
URL: *https://doi.org/10.1023/A:1020938210765*

Schlather, M. & Tawn, J. A. (2003), 'A dependence measure for multivariate and spatial extreme values: properties and inference', Biometrika **90**(1), 139–156.
URL: *https://doi.org/10.1093/biomet/90.1.139*

Shyamalkumar, N. D. & Tao, S. (2020), 'On tail dependence matrices', Extremes **23**, 245–285.
URL: *https://doi.org/10.1007/s10687-019-00366-y*

Smith, R. L. (1990), 'Max-stable processes and spatial extremes', Unpublished manuscript **205**, 1–32.

Stoev, S. & Taqqu, M. S. (2005), 'Extremal stochastic integrals: a parallel between max–stable processes and $\alpha-$stable processes', Extremes **8**, 237–266.
URL: *https://doi.org/10.1007/s10687-006-0004-0*

Strokorb, K., Ballani, F. & Schlather, M. (2015), 'Tail correlation functions of max-stable processes: construction principles, recovery and diversity of some mixing max-stable processes with identical TCF', Extremes **18**(2), 241–271.
URL: *https://doi.org/10.1007/s10687-014-0212-y*

Strokorb, K. & Schlather, M. (2015), 'An exceptional max-stable process fully parameterized by its extremal coefficients', Bernoulli **21**(1), 276–302.
URL: *https://doi.org/10.3150/13-BEJ567*

Yuen, R., Stoev, S. & Cooley, D. (2020), 'Distributionally robust inference for extreme Value-at-Risk', Insurance Math. Econom. **92**, 70–89.
URL: *https://doi.org/10.1016/j.insmatheco.2020.03.003*

A Proofs and auxiliary results

A.1 Proofs for Section 2

Proof of Proposition 2.5. Here, for brevity, we shall write $\{h \in A\}$ for the pre-image set $h^{-1}(A) = \{x \in \mathbb{R}^p : h(x) \in A\}$. By the continuity of h, it follows that for all $a \geq 0$, the set $\{h \geq a\}$ is closed and $\{h > a\}$ is open. Hence $\{h = a\} = \{h \geq a\} \setminus \{h > a\} \supset \partial\{h > a\}$, where $\partial A = A^{\mathrm{cl}} \setminus A^{\mathrm{int}}$ denotes the boundary of the set A. Since $h(0) = 0$ (by continuity and homogeneity), we have that for all $a > 0$, the closed set $\{h \geq a\}$ does not contain 0 and hence it is bounded away from 0. Thus, $\mu(\{h \geq a\}) < \infty$. Since $\{h = t\} = t \cdot \{h = 1\},\ t > 0$, the scaling property (2.2) of μ implies that $\mu(\{h = t\}) = t^{-\alpha}\mu(\{h = 1\})$ and if $\mu(\{h = t\}) > 0$ for some (any) $t > 0$, then $\mu(\{h = t\}) > 0$, for all $t > 0$. On the other hand, we have that $\{h \geq a\} = \cup_{t \geq a}\{h = t\}$, where the latter union involves an uncountable collection of disjoint sets. Thus, $\mu(\{h = t\})$ must vanish for all $t > 0$. This means that $\mu(\partial\{\mu > a\}) = 0$, or that $\{h > a\}$ are μ-continuity sets for all $a > 0$. This allows us to apply the definition of regular variation (2.1) and obtain

$$n\mathbb{P}[h(X) > a_n] = n\mathbb{P}[X \in a_n \cdot \{h > 1\}] \to \mu(\{h > 1\}), \quad \text{as } n \to \infty.$$

Now, (2.3) entails

$$\begin{aligned}
\mu(\{h > 1\}) &= \int_S \int_0^\infty 1_{\{h(ru) > 1\}} \alpha r^{-\alpha-1} dr\sigma(du) \\
&= \int_S \int_0^\infty 1_{\{h(u) > 1/r\}} \alpha r^{-\alpha-1} dr\sigma(du) \\
&= \int_S \int_0^\infty 1_{\{h^\alpha(u) > x\}} dx\sigma(du) = \int_S h^\alpha(u)\sigma(du),
\end{aligned}$$

where in the last two displays we used the homogeneity of h and the change of variables $x := r^{-\alpha}$. This completes the proof of the first relation in (2.8). The second relation therein follows from the observation that $\sigma(\cdot)/\sigma(S)$ is a probability distribution. □

Proof of Proposition 2.6. Since X has non-negative components, to establish its regular variation, it is enough consider measures supported only on $\mathbb{R}^p_+$ and show that

$$\mu_n(\cdot) := n\mathbb{P}[X \in n\cdot] \overset{\mathrm{M}_0}{\Longrightarrow} \mu, \quad \text{as } n \to \infty,$$

where $\mu[0,x]^c = -\log(\mathbb{P}[X \leq x])$ with $[0,x]^c := \mathbb{R}^p_+ \setminus [0,x]$.
Fix an $x \in [0,\infty)^p \setminus \{0\}$. The definition of simple max-stability (2.5) entails $\mathbb{P}[X \leq nx]^n = \mathbb{P}[X \leq x]$. Thus, for all $n \in \mathbb{N}$,

$$\mathbb{P}[X \leq nx]^n = \Big(1 - \frac{n\mathbb{P}[X \in n \cdot [0,x]^c]}{n}\Big)^n = \Big(1 - \frac{\mu_n[0,x]^c}{n}\Big)^n = \mathbb{P}[X \leq x].$$

Observe that $\mathbb{P}[X \leq x]$ is *positive*. Indeed, it is easy to see that since X is non-degenerate $\xi := \max_{1 \leq i \leq p} x_i^{-1} X_i$ is 1-Fréchet and thus $\mathbb{P}[\xi \leq 1] = \mathbb{P}[X \leq x] > 0$, for all $x \in \mathbb{R}^p_+$. This means that for the boundedly finite measures $\mu_n(\cdot) = n\mathbb{P}[X \in n\cdot]$, we have

$$\lim_{n\to\infty} \mu_n([0,x]^c) = -\log \mathbb{P}[X \leq x]. \tag{A.1}$$

The latter relation shows that the sequence of measures $\{\mu_n\}$ is relatively compact in $M_0(\mathbb{R}^p)$, equipped with the M_0-convergence topology. Indeed, by (Hult & Lindskog 2006, Theorem 2.7), it suffices to show that for all $\varepsilon > 0$ and $\eta > 0$, there exists an $M = M(\varepsilon, \eta) > 0$, such that

$$\sup_n \mu_n(B(0,\varepsilon)^c) < \infty \quad \text{and} \quad \sup_n \mu_n\big(\mathbb{R}_+^p \setminus [0,M]^p\big) < \eta.$$

The first condition follows from (A.1) and since $\mathbb{P}[X \le x] > 0$. The second condition follows from the fact that $-\log(\mathbb{P}[X \le x]) \downarrow 0$, as $x \uparrow \infty$, which is true since X has a valid probability distribution.
The relative compactness of the measures $\{\mu_n\}$ entails that $\mu_{n'} \overset{\mathrm{M_0}}{\Longrightarrow} \mu$ for some $\mu \in M_0$ and a sub-sequence $n' \to \infty$. However, by (A.1) and Proposition 2.4 we have

$$\begin{aligned} \mu[0,x]^c &= -\log \mathbb{P}[X \le x] \\ &= \int_E \max_{1\le i\le p} \frac{f_i(u)}{x_i} \nu(du) \\ &= \int_E \int_0^\infty 1_{[0,x]^c}(r\vec{f}(u)) r^{-2} dr \nu(du) \end{aligned} \tag{A.2}$$

for all $x \in [0,\infty)^p \setminus \{0\}$, and the limit measure is uniquely determined by its values on all the complements of rectangles containing the origin. Furthermore, we see from (A.2) that for non-degenerate X, the limit measure μ is non-degenerate as well. This proves that $X \in RV(\{n\}, \mu)$ where $\mathbb{P}[X \le x] = \exp\{-\mu[0,x]^c\}$.
Having established regular variation, the first equality in Relation (2.9) follows from the μ-continuity of the set $\{h > 1\}$ as argued in the proof of Proposition 2.5. The rest of Relation (2.9) follows from (A.2).
Finally, the representation in (2.10) follows from the fact that σ is determined by $\int_S g(u)\sigma(du)$, for all continuous functions $g : S \to \mathbb{R}_+$. Indeed, for every such g, the function $h(x) := g(x/\|x\|)\|x\|1_{\{x\neq 0\}}$ is continuous, non-negative and 1-homogeneous and hence by (2.9)

$$\int_S g(u)\sigma(du) = \int_S g\Big(\frac{\vec{f}(z)}{\|\vec{f}(z)\|}\Big)\|\vec{f}(z)\|\nu(dz).$$

This, since g is arbitrary, proves (2.10). □

Proof of Corollary 2.11. Relation (2.12) follows by applying (2.9) to h replaced by the continuous and homogeneous functions $h_{\max,L}(x) := (\max_{i\in L} x_i)_+$ and $h_{\min,L}(x) := (\min_{i\in L} x_i)_+$, respectively. Indeed, observe that

$$n\mathbb{P}\Big[\min_{i\in L} X_i > n\Big] = n\mathbb{P}\Big[h_{\min,L}(X) > n\Big] \to \mu(\{h_{\min,L} > 1\}) = \lambda(L), \quad \text{as } n \to \infty.$$

On the other hand, $\{x : h_{\min,L}(x) > 1\} = \bigcap_{i\in L} A_i$.
The formula for $\lambda(L)$ in Relation (2.13) follows from the above and Equation (2.9) since $h_{\min,L}(\vec{f}) = \min_{i\in L} f_i(x)$. The derivations of the formulae for $\theta(K)$ are similar. □

We conclude this section with the auxiliary result, that the spectral distance and the tail-dependence coefficients are linear under max-linear combinations, in the sense of the following lemma.

Lemma A.1. *Let $X^{(t)} = (X_i^{(t)})_{1\le i\le p}$, $1 \le t \le n$, be independent simple max-stable vectors with tail measures $\mu^{(t)}$, $1 \le t \le n$, and let $\gamma_t \ge 0$, $1 \le t \le n$, be some non-negative weights. Define $\bar{X} = \bigvee_{t=1}^n \gamma_t X^{(t)}$. Then,*

$$d_{\bar{X}}(i,j) = \sum_{t=1}^n \gamma_t d_{X^{(t)}}(i,j) \quad \text{for all} \quad i,j \in [p]$$

and

$$\lambda_{\bar{X}}(L) = \sum_{t=1}^n \gamma_t \lambda_{X^{(t)}}(L) \quad \text{for all} \quad L \subset [p].$$

Proof of Lemma A.1. By the independence of $X^{(t)}$, $1 \le t \le n$, and Proposition 2.6 it applies

$$\mathbb{P}\big[\bar{X} \le x\big] = \prod_{t=1}^n \mathbb{P}\Big[X^{(t)} \le \frac{1}{\gamma_t} x\Big] = \prod_{t=1}^n \exp\Big\{ -\mu^{(t)}\big[0, \frac{1}{\gamma_t}x\big]^c\Big\} = \exp\Big\{ -\sum_{t=1}^n \gamma_t \mu^{(t)}[0,x]^c\Big\}$$

for all $x \in \mathbb{R}_+^p \setminus \{0\}$, where in the last step the homogeneity of $\mu^{(t)}$ was applied. Thus, $\bar{X}$ has the tail measure $\mu_{\bar{X}} = \sum_{t=1}^n \gamma_t \mu^{(t)}$, i.e. the tail measure of the max-linear combination $\bar{X}$ is the corresponding linear combination of the tail measures of the components. In particular, $\bar{X}$ has 1-Fréchet marginals with scale coefficient $\| \bigvee_{t=1}^n \gamma_t X_i^{(t)}\|_1 = \sum_{t=1}^n \gamma_t \|X_i^{(t)}\|_1$.
Hence, by the definition of the spectral distance $d_{\bar{X}}$ in Definition 2.13 we obtain

$$\begin{aligned} d_{\bar{X}}(i,j) &= 2\Big\| \bigvee_{t=1}^n \gamma_t X_i^{(t)} \vee \bigvee_{t=1}^n \gamma_t X_j^{(t)} \Big\|_1 - \Big\| \bigvee_{t=1}^n \gamma_t X_i^{(t)} \Big\|_1 - \Big\| \bigvee_{t=1}^n \gamma_t X_j^{(t)} \Big\|_1 \\ &= 2\sum_{t=1}^n \gamma_t \|X_i^{(t)} \vee X_j^{(t)}\|_1 - \sum_{t=1}^n \gamma_t \|X_i^{(t)}\|_1 - \sum_{t=1}^n \gamma_t \|X_j^{(t)}\|_1 = \sum_{t=1}^n \gamma_t d_{X^{(t)}}(i,j). \end{aligned}$$

By the linear representation of $\mu_{\bar{X}}$ and (2.12) it follows for the tail-dependence coefficients

$$\lambda_{\bar{X}}(L) = \mu_{\bar{X}}\Big(\bigcap_{i\in L} A_i\Big) = \sum_{t=1}^n \gamma_t \mu^{(t)}\Big(\bigcap_{i\in L} A_i\Big) = \sum_{t=1}^n \gamma_t \lambda_{X^{(t)}}(L)$$

□

A.2 Proofs for Section 3

Lemma A.2. *For λ and θ in* (3.2) *and β in* (3.1)*, we have the inversion formulae* (3.3), (3.4) *as well as* (3.5)*. Namely, the following formulae hold:*

$$\beta(J) = \sum_{K\,:\,\emptyset\ne K,\ J^c\subset K} (-1)^{|J\cap K|+1}\theta(K), \qquad \beta(J) = \sum_{L\,:\,J\subset L\subset[p]} (-1)^{|L\setminus J|}\lambda(L)$$

$$\theta(K) = \sum_{L\,:\,\emptyset\ne L\subset K} (-1)^{|L|-1}\lambda(L), \qquad \lambda(L) = \sum_{K\,:\,\emptyset\ne K\subset L} (-1)^{|K|-1}\theta(K).$$

Proof. For simplicity, introduce the indicator functions $I_i := 1_{A_i}$, where $A_i = \{x \in \mathbb{R}^p : x_i > 1\}$. In view of (3.1) and (2.12), we have

$$\beta(J) = \int \Big(\prod_{i \in J} I_i \times \prod_{j \in J^c} (1 - I_j) \Big) d\mu$$

as well as

$$\lambda(L) = \int \Big(\prod_{i \in L} I_i \Big) d\mu. \tag{A.3}$$

This immediately entails

$$\beta(J) = \int \Big(\sum_{J \subset L \subset [p]} (-1)^{|L \setminus J|} \prod_{i \in L} I_i \Big) d\mu = \sum_{L \,:\, J \subset L \subset [p]} (-1)^{|L \setminus J|} \lambda(L),$$

which proves (3.4).
The inclusion-exclusion formula for θ in terms of the λ's is immediate from (A.3) and the observation that

$$\theta(K) = \int \Big(1 - \prod_{i \in K} (1 - I_i) \Big) d\mu = \sum_{L \,:\, \emptyset \neq L \subset K} (-1)^{|L|-1} \int \Big(\prod_{i \in L} I_i \Big) d\mu = \sum_{L \,:\, \emptyset \neq L \subset K} (-1)^{|L|-1} \lambda(L).$$

Now, using (A.3) we obtain

$$\begin{aligned} \lambda(L) &= \int \Big(\prod_{i \in L} (1 - (1 - I_i)) \Big) d\mu \\ &= \int \Big(1 + \sum_{K \,:\, \emptyset \neq K \subset L} (-1)^{|K|} \prod_{i \in K} (1 - I_i) \Big) d\mu. \end{aligned} \tag{A.4}$$

Observe that by Newton's binomial formula:

$$0 = (1 + (-1))^{|L|} = \sum_{K \subset L} (-1)^{|K|} = 1 + \sum_{K \,:\, \emptyset \neq K \subset L} (-1)^{|K|},$$

and hence

$$1 = \sum_{K \,:\, \emptyset \neq K \subset L} (-1)^{|K|-1}.$$

Using the latter expression for the constant 1 in the right-hand side of (A.4), we obtain

$$\lambda(L) = \sum_{K \,:\, \emptyset \neq K \subset L} (-1)^{|K|-1} \int \Big(1 - \prod_{i \in K} (1 - I_i) \Big) d\mu = \sum_{K \,:\, \emptyset \neq K \subset L} (-1)^{|K|-1} \theta(K),$$

completing the proof of the inclusion-exclusion formula for the λ's via the θ's in (3.5).
To complete the proof we need to establish the expression of $\beta(J)$'s via the $\theta(K)$'s in (3.3). We do so by passing through the $\lambda(L)$'s first. Namely, by the established (3.4) and (3.5), we have

$$\beta(J) = \sum_{L \,:\, J \subset L \subset [p]} (-1)^{|L \setminus J|} \lambda(L)$$

$$
\begin{aligned}
&= \sum_{L\,:\,J\subset L\subset[p]} (-1)^{|L\setminus J|}\Big(\sum_{K\,:\,\emptyset\neq K\subset L} (-1)^{|K|-1}\theta(K)\Big)\\
&= \sum_{K\,:\,\emptyset\neq K\subset[p]} \theta(K)(-1)^{|J|+|K|-1-|J\cup K|}\times\Big(\sum_{L\,:\,(K\cup J)\subset L} (-1)^{|L|-|J\cup K|}\Big)\\
&=: \sum_{K\,:\,\emptyset\neq K\subset[p]} \theta(K)(-1)^{|J\cap K|+1}\times C(K,J).
\end{aligned}
$$

Observe that $C(K,J) := \sum_{L\,:\,(J\cup K)\subset L}(-1)^{|L|-|J\cup K|} = 0$ if $(J\cup K)^c \neq \emptyset$. Indeed, the latter sum is simply $(1+(-1))^{|[p]\setminus(K\cup L)|} = 0$. On the other hand, if $J\cup K = [p]$, we trivially have $C(K,L) = 1$. This, since $J\cup K = [p]$ is equivalent to $J^c \subset K$, immediately implies

$$
\beta(J) = \sum_{K\,:\,\emptyset\neq K,\ J^c\subset K} \theta(K)(-1)^{|J\cap K|+1}.
$$

This proves (3.3). □

Proof of Theorem 3.8. For this proof it is convenient to let $\|u\| := \max_{i\in[p]}|u_i|$ be the sup-norm. ('if') Suppose that all $\beta(J)$'s in (3.3) (or in (3.4)) are non-negative and define X^* as in (3.12). Clearly, X^* is max-stable and we shall determine its spectral measure σ^* in the sup-norm. Observe that for all $x\in(0,\infty)^p$ we have $\beta(J)Z_J 1_J \leq x$, if and only if $Z_J \leq \min_{i\in J} x_i/\beta(J)$ and since the Z_J's are iid standard 1-Fréchet:

$$
\mathbb{P}[X^*\leq x] = \exp\Big\{-\sum_{\emptyset\neq J\subset[p]}\frac{\beta(J)}{\min_{i\in J}x_i}\Big\}.
$$

Letting $\sigma^*(du) := \sum_{\emptyset\neq J\subset[p]}\beta(J)\delta_{1_J}(du)$, we obtain that

$$
\int_S\Big(\max_{i\in[p]}\frac{u_i}{x_i}\Big)\sigma^*(du) = \sum_{\emptyset\neq J\subset[p]}\frac{\beta(J)}{\min_{i\in J}x_i}. \tag{A.5}
$$

This shows that

$$
\mathbb{P}[X^*\leq x] = \exp\{-\mu^*[0,x]^c\} = \exp\Big\{-\int_S\Big(\max_{i\in[p]}\frac{u_i}{x_i}\Big)\sigma^*(du)\Big\},
$$

which in view of (2.11) shows that σ^* is the spectral measure of X^*, where μ^* denotes the tail measure of X^*.
Let now $\emptyset\neq J\subset[p]$ and define the set

$$
C_J := \{u\in S\,:\,u_i = 1 \text{ for all } i\in J \text{ and } u_j = 0, \text{ for all } j\in J^c\ \}.
$$

We will argue that, with B_J, A_j as in (3.1),

$$
\sigma^*(C_J) = \mu^*(B_J) := \mu^*\Big(\bigcap_{i\in J}A_i\cap\bigcap_{j\in J^c}A_j^c\Big).
$$

Indeed, by (2.4), we have

$$
\sigma^*(C_J) = \mu^*(\widetilde{C}_J) := \mu^*\{x\in\mathbb{R}_+^d\,:\,\|x\|>1,\ x/\|x\|\in C_J\}.
$$

Note that if $x \in \widetilde{C}_J$, then $x_i = \|x\| > 1$ for all $i \in J$, and $x_j = 0 < 1$, for all $j \in J^c$, so that $x \in B_J$. This means that $\widetilde{C}_J \subset B_J$, and hence $\sigma^*(C_J) \leq \mu^*(B_J)$.
By the construction of σ^*, on the other hand, we have $\sigma^*(C_J) = \sigma^*(\{1_J\}) = \beta(J)$ and since the B_J's partition the set $\{\|x\| > 1\} \cap \mathbb{R}^p_+$, we get

$$\mu^*(\{\|x\| > 1\}) = \mu^*(\{\|x\| > 1\} \cap \mathbb{R}^p_+) = \sum_J \mu^*(B_J) \geq \sum_J \sigma^*(C_J) = \sum_J \beta(J) = \sigma^*(S),$$

where the last relation follows from (A.5) by setting $x_i = 1, i \in [p]$. Since $\sigma^*(S) = \mu^*(\{\|x\| > 1\})$ we obtain from the above inequality that $\mu^*(B_J) = \sigma^*(C_J) = \beta(J)$, for all $J \subset [p]$.
We have thus shown that the functionals $\beta(J)$ that we started with are indeed the ones which determine the extremal (tail dependence) coefficients of X^* via (3.2). This completes the proof of the 'if' part.

('only if') Conversely, if $\{\theta(K),\ K \subset [p]\}$ (or $\{\lambda(L),\ L \subset [p]\}$) are the extremal coefficients (tail-dependence coefficients, respectively) of a max-stable vector X with tail measure μ. Then, as already argued above (2.12) holds, and hence the $\beta(J)$'s defined as in (3.1) are non-negative and satisfy Relations (3.3) (or (3.4)). This completes the proof. □

A.3 Proofs for Section 5

Proof of Lemma 5.3. Assume that for an input matrix d for the SDR problem with unconstrained, identical margins the answer is positive, which is equivalent to d being L^1-embeddable, see the proof of Theorem 5.1. According to Proposition 4.5 and Lemma A.1 we can then choose the realizing max-stable vector X as a (generalized) TM-model with coefficients $\beta(J), \emptyset \neq J \subset [p]$, such that

$$c = \sum_{J \ni i} \beta(J), \quad i \in [p],$$

and

$$d(i,j) = \sum_{J\,:\,\emptyset \neq J \subset [p]} \beta(J) |1_J(i) - 1_J(j)|, \quad i,j \in [p]. \tag{A.6}$$

We note that the particular choice of $\beta([p]) \geq 0$ does only affect the value of c and is not determined by (A.6) as $|1_{[p]}(i) - 1_{[p]}(j)| = 0$. Thus, for each realizing max-stable vector X with $\|X_i\| = c$ we can for each $\tilde{c} > c$ find a realizing max-stable vector X with $\|X_i\| = \tilde{c}$ by increasing the value of $\beta([p])$ in the generalized TM-model.
Since all $\beta(J)'s$ are non-negative, (A.6) implies furthermore that for all $\emptyset \neq J \subsetneq [p]$ the inequality

$$\beta(J) \leq \max_{i,j \in [p]} d(i,j)$$

holds. Thus we know that for all $i \in [p]$

$$c = \sum_{J \ni i} \beta(J) \leq \beta([p]) + (2^p - 2) \max_{i,j \in [p]} d(i,j)$$

As any choice of $\beta([p]) \geq 0$ leads to a realizing (generalized) TM-model for given d we see that any value $c \geq (2^p - 2) \max_{i,j \in [p]} d(i,j)$ is possible for the marginal scale of this model. □