Sparse regular variation

Nicolas Meyer, Olivier Wintenberger

Sparse regular variation

Nicolas Meyer*[1] and Olivier Wintenberger†[1]

[1]*Sorbonne Université, LPSM, F-75005, Paris, France*

June 28, 2019

Abstract

Regular variation provides a convenient theoretical framework to study large events. In the multivariate setting, the dependence structure of the positive extremes is characterized by a measure - the spectral measure - defined on the positive orthant of the unit sphere. This measure gathers information on the localization of extreme events and is often sparse since severe events do not occur in all directions. Unfortunately, it is defined through weak convergence which does not provide a natural way to capture its sparse structure. In this paper, we introduce the notion of sparse regular variation, which allows to better learn the sparse structure of extreme events. This concept is based on the euclidean projection onto the simplex for which efficient algorithms are known. We show several results for sparsely regularly varying random vectors. Finally, we prove that under mild assumptions sparse regular variation and regular variation are two equivalent notions.

Keywords: multivariate extremes, projection onto the simplex, regular variation, sparse regular variation, spectral measure

1 Introduction

Estimating the dependence structure of extreme events has proven to be a major issue in many applications. The classical framework in the multivariate Extreme Value Theory (EVT) is based on the concept of regularly varying random vectors. Several characterizations of regular variation has been established (see e.g. Embrechts et al. (1997), Resnick (1987), Resnick (2007), or Beirlant et al. (2006)). A natural way to define multivariate regular variation is through the convergence of the polar coordinates of a random vector. Indeed, a random vector $\mathbf{X} \in \mathbb{R}_+^d$ is said to be regularly varying with tail index $\alpha > 0$ and spectral measure S on the positive orthant of the unit sphere if for all $x > 0$,

$$\mathbb{P}\left(|\mathbf{X}| > tx, \mathbf{X}/|\mathbf{X}| \in \cdot \mid |\mathbf{X}| > t\right) \xrightarrow{d} x^{-\alpha} S(\cdot), \quad t \to \infty\,, \tag{1.1}$$

where $\xrightarrow{d}$ denotes the weak convergence in the space of nonnegative Radon measure on the unit sphere. Convergence (1.1) can be interpreted as follows: the limit of the radial component $|\mathbf{X}|/t$ has a Pareto distribution with parameter $\alpha > 0$ whereas the angular component $\mathbf{X}/|\mathbf{X}|$ has limit measure S. Moreover, both components of the limit are independent. The measure S, called the *spectral measure*, summarizes the tail dependence of the regularly varying random vector $\mathbf{X}$. Estimating this $d-1$ dimensional measure is a challenging problem, especially in high dimensions.

* *nicolas.meyer@upmc.fr* (corresponding author)

† *olivier.wintenberger@upmc.fr*

Based on convergence (1.1), several nonparametric estimation techniques have been proposed to estimate S. Since (1.1) holds for any norms, these estimators mainly differ in terms of the choice of the norm. In particular, some useful representations of the spectral measure has been introduced in the bivariate case by Einmahl et al. (1993), Einmahl et al. (1997), Einmahl et al. (2001) and Einmahl and Segers (2009). In Einmahl et al. (1997), the authors replace the tails of the margins by fitted Pareto tails in order to estimate S by an empirical measure. This latter is consistent and asymptotically normal under suitable assumptions. Einmahl and Segers (2009) focuse on the choice of the ℓ^p norm, for $p \in [1, \infty]$, in order to construct an estimator of the spectral measure which satisfies moments constraints. Inference on the spectral measure has also been studied in a Bayesian framework, for instance in Guillotte et al. (2011). In this paper, the authors use censored likelihood methods in the context of infinite dimensional spectral measures.

In higher dimension, mixture of Dirichlet distributions are often used to model the spectral densities. Boldi and Davison (2007) show that under some conditions these distributions are weakly dense in the set of spectral measures. They propose both frequentist and Bayesian inferences based on EM algorithms and MCMC simulations. Subsequently, Sabourin and Naveau (2014) introduce a re-parametrization of the Bayesian Dirichlet mixture model.

Since extreme events often concentrate on small subspaces of $\mathbb{R}^d_+$, the spectral measure is usually sparse. This means that it does not put mass in some regions of the unit sphere. The subspaces where the spectral measure puts mass are these where extreme events occur. Thus, estimating the spectral measure is a major issue in multivariate EVT but it is a challenging problem, especially in high dimensions. This topic of research is quite recent but some methods have already been proposed. Chautru (2015) proposes a clustering approach to exhibit groups of variables with asymptotic dependence. In the same way, the purpose of the algorithm in Chiapino and Sabourin (2016) is to gather the features that are likely to be extreme simultaneously. Cooley and Thibaud (2016) model the dependence via a matrix of pairwise tail dependence metrics. They define a transformation to consider the positive orthant $\mathbb{R}^d_+$ as a vector space before applying some factorizations on the positive matrix. Goix et al. (2017) consider ϵ-thickened rectangles to estimate the directions on which the spectral measure concentrates. All these approaches are based on the rank transform and try to identify groups of asymptotically dependent extremes.

In a recent work, Lehtomaa and Resnick (2019) analyse extremal dependence with application to risk management. They study the support of the spectral measure by using a grid estimator. The simplex $\mathbb{S}^{d-1}_+$ is firstly mapped to an $d-1$ dimensional space $[0,1]^{d-1}$ before being partitioned in equally sized rectangles. The estimation of the support is based on a classical estimator of the spectral measure, see Resnick (2007), p. 308. The second step is then to build an asymptotically normal test statistic to validate the support estimate.

The main issue in the study of the spectral measure S is that the self-normalized extreme $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$ that appears in (1.1) is inefficient to estimate S in subspaces of dimension smaller than d. Indeed, if S puts mass in such subspaces, then the weak convergence (1.1) does not hold anymore since such subspaces are not continuous sets for S. This is why the difficulty to identify the possible sparsity of S is at the core of the multivariate extremes' study.

Since the self-normalized vector $\mathbf{X}/|\mathbf{X}|$ fails to identify the regions on which the spectral measure puts mass, our aim is to introduce another way of projecting onto the unit sphere. This new projection should take the sparsity of the spectral measure into account by introducing some sparsity in the vector $\mathbf{X}$. In other words, as the limit measure S in (1.1) is likely to be sparse, we need to replace $\mathbf{X}/|\mathbf{X}|$ by a unit vector based on $\mathbf{X}$ which is also likely to be sparse. To this end, we use the euclidean projection of $\mathbf{X}/t$ onto the simplex $\{\mathbf{x} \in \mathbb{R}^d_+, \, x_1 + \ldots + x_d = 1\}$. This projection has been widely studied in learning theory (see e.g. Duchi et al. (2008), Kyrillidis et al. (2013), or Liu and Ye (2009)). Many different efficient algorithms have been proposed, for instance in Duchi et al. (2008) and Condat (2016).

Based on this projection, we define the concept of sparse regular variation for which the self-normalized vector $\mathbf{X}/|\mathbf{X}|$ is replaced by $\pi(\mathbf{X}/t)$ where π denotes the euclidean projection onto the simplex. The limit measure obtained after this substitution is slightly different from the spectral measure S. We study this new angular limit and show that it better captures the possible sparse structure of the extremes. Besides, we prove that under mild conditions both concepts of regular variation are equivalent and we give the relation between both limiting spectral measures.

Outline The structure of this paper is as follows. Section 2 gathers all theoretical results useful in this paper. Firstly, we give the context of regularly varying random vector in the multivariate EVT framework. We detail why the knowledge of the subspaces on which the spectral measure puts mass is a main issue for the study of extreme events, and we explain which difficulties appear in this context. Secondly, we introduce the euclidean projection onto the simplex and list several results which are of constant use for our study. Section 3 is dedicated to the study of this projection in the context of regular variation. We focus on the angular part of the limit after substituing the classical projected vector $\mathbf{X}/|\mathbf{X}|$ in (1.1) by a vector based on the euclidean projection onto the simplex. We also discuss to what extend this way of projecting allows to better capture the sparse structure of the tail dependence. Finally, in Section 4 we introduce the concept of sparsely regularly varying random vector. We establish the equivalence, under mild conditions, between this notion and the classical regular variation's concept.

Notations Denote in bold-face elements $\mathbf{x} = (x_1, \dots, x_d)$ of $\mathbb{R}^d$. We write $\mathbf{x} \leq \mathbf{y}$, $\mathbf{x} < \mathbf{y}$, $\mathbf{x} \geq \mathbf{y}$, etc. where $\leq, <, \geq$ refer to the componentwise partial ordering in $\mathbb{R}^d$. More generally, for $\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$, we write $\mathbf{x} \leq y$ if all components x_i of $\mathbf{x}$ satisfy $x_i \leq y$. In the same way, $\mathbf{x} + y$ is defined as the vector $(x_1 + y, \dots x_d + y)$. We also define $\mathbb{R}^d_+ = \{\mathbf{x} \in \mathbb{R}^d,\, x_1 \geq 0, \dots, x_d \geq 0\}$, $\mathbf{e} = (1, \dots, 1) \in \mathbb{R}^d$, $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^d$. For $j = 1, \dots, d$, $\mathbf{e}_j$ denotes the j-th vector of the canonical basis of $\mathbb{R}^d$, this means that $\mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)$, where the 1 is in position j. For $a \in \mathbb{R}$, a_+ denotes the positive part of a, that is $a_+ = a$ if $a \geq 0$ and $a_+ = 0$ otherwise. If $\mathbf{x} \in \mathbb{R}^d$ and $I = \{i_1, \dots, i_r\} \subset \{1, \dots, d\}$, then $\mathbf{x}_I$ denotes the vector $(x_{i_1}, ..., x_{i_r})$ of $\mathbb{R}^r$. For $p \in [1, \infty]$, we denote by $|\cdot|_p$ the ℓ^p-norm in $\mathbb{R}^d$. We write $\overset{d}{\to}$ for the convergence in distribution and $\overset{v}{\to}$ for the vague convergence. Finally, for a finite set I, $|I|$ denotes the number of elements of I.

2 Regular variation and projection

2.1 Multivariate Extreme Value Theory

We consider a nonnegative random vector $\mathbf{X} = (X_1, \dots, X_d)$ with cumulative distribution function F. Our aim is to assess the tail structure of F. It is customary in EVT to assume that the random vector $\mathbf{X}$ is regularly varying, *i.e.* there exist a positive sequence (a_n), $a_n \to \infty$ when $n \to \infty$, and a nonnegative Radon measure μ on $\mathbb{R}^d_+ \setminus \{\mathbf{0}\}$ such that

$$n\mathbb{P}\left(\frac{\mathbf{X}}{a_n} \in \cdot\right) \overset{v}{\to} \mu(\cdot), \quad n \to \infty, \tag{2.1}$$

where $\overset{v}{\to}$ denotes vague convergence in the space of nonnegative Radon measures on $\mathbb{R}^d_+ \setminus \{\mathbf{0}\}$. The limit measure μ is called the *tail measure* and describes the behavior of the extremes. It satisfies the homogeneity property $\mu(aC) = a^{-\alpha}\mu(C)$ for any set $C \subset \mathbb{R}^d_+ \setminus \{\mathbf{0}\}$ and any $a > 0$. The parameter $\alpha > 0$ is called the *tail index*.

It is often more convenient to represent the extremal behavior of $\mathbf{X}$ through a polar representation (see Beirlant et al. (2006), Section 8.2.2). Choose a norm $|\cdot|$ on $\mathbb{R}^d$ and denote by $\mathbb{S}^{d-1}_+$ the restriction of its unit sphere to the positive orthant: $\mathbb{S}^{d-1}_+ = \{\mathbf{x} \in \mathbb{R}^d_+,\, |\mathbf{x}| = 1\}$. Now define the following transformation

$$\begin{array}{rcccc} T & : & \mathbb{R}^d_+ \setminus \{\mathbf{0}\} & \to & (0, \infty) \times \mathbb{S}^{d-1}_+ \\ & & \mathbf{v} & \mapsto & (r, \boldsymbol{\theta}) = (|\mathbf{v}|, \mathbf{v}/|\mathbf{v}|)\,. \end{array}$$

Note that it is even possible to choose two different norms $|\cdot|$ and $|\cdot|'$ in the definition of T (see Beirlant et al. (2006)) but it will not be useful here. Classical choices of norms are ℓ^p-norms, $p \in [1,\infty]$. Define a measure S on $\mathbb{S}_+^{d-1}$ by setting

$$S(B) = \mu\big(\{\mathbf{v} \in \mathbb{R}_+^d \setminus \{\mathbf{0}\}, |\mathbf{v}| > 1, \mathbf{v}/|\mathbf{v}| \in B\}\big) = \mu\left(T^{-1}\left[(1,\infty) \times B\right]\right) ,$$

for Borel subsets B of $\mathbb{S}_+^{d-1}$. The measure S is called the *spectral measure* of the regularly varying random vector $\mathbf{X}$. It may be seen as the projection of the tail measure μ onto the unit sphere. The homogeneity property of the tail measure implies that

$$r^{-\alpha} S(B) = \mu\big(\{\mathbf{v} \in \mathbb{R}_+^d \setminus \{\mathbf{0}\}, |\mathbf{v}| > r, \mathbf{v}/|\mathbf{v}| \in B\}\big) = \mu\left(T^{-1}\left[(r,\infty) \times B\right]\right) , \tag{2.2}$$

for Borel subsets B of $\mathbb{S}_+^{d-1}$ and $r > 0$.

Equation (2.2) can be rephrased as $\alpha r^{-(\alpha+1)} \mathrm{d}r S(\mathrm{d}\boldsymbol{\theta}) = \mu \circ T^{-1}(\mathrm{d}r, \mathrm{d}\boldsymbol{\theta})$. This gives a decomposition of the tail measure in a radial part and an angular part. The radial component can thus be modeled through a random variable with Pareto(α) distribution, whereas the angular one is characterized by the spectral measure S. The decomposition in Equation (2.2) ensures that the radial and the angular parts are independent.

We combine Equations (2.1) and (2.2) to characterize regularly varying random vector in $\mathbb{R}_+^d$ in the following way:

$$n\mathbb{P}\left(a_n^{-1}|\mathbf{X}| > r, \mathbf{X}/|\mathbf{X}| \in \cdot\right) \xrightarrow{v} r^{-\alpha} S(\cdot), \quad n \to \infty .$$

This is the same as

$$\frac{\mathbb{P}\left(|\mathbf{X}| > tr, \mathbf{X}/|\mathbf{X}| \in \cdot\right)}{\mathbb{P}(|\mathbf{X}| > t)} \xrightarrow{v} r^{-\alpha} S(\cdot), \quad t \to \infty , \tag{2.3}$$

see e.g. Resnick (1986).

We call *spectral vector* a random vector on $\mathbb{S}_+^{d-1}$ whose distribution is S. Equation (2.3) then leads to the following characterization of regular variation. A random vector $\mathbf{X}$ on $\mathbb{R}_+^d$ is regularly varying if there exist a random vector $\boldsymbol{\Theta}$ on $\mathbb{S}_+^{d-1}$ (the spectral vector) and a random variable Y such that the following limit holds:

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \frac{\mathbf{X}}{|\mathbf{X}|}\right) \in \cdot \;\middle|\; |\mathbf{X}| > t\right) \xrightarrow{d} \mathbb{P}((Y, \boldsymbol{\Theta}) \in \cdot), \quad t \to \infty . \tag{2.4}$$

In this case, there exists $\alpha > 0$ such that the distribution of Y is Pareto(α). Moreover, the radial limit Y is independent of the angular limit $\boldsymbol{\Theta}$ which has distribution the spectral measure S.

Equation (2.4) highlights the two quantities which characterize the regular variation propery of $\mathbf{X}$. On the one hand, the tail index α gives the size of the extremes: the smaller this index is, the larger the extremes are. On the other hand, the spectral vector $\boldsymbol{\Theta}$ informs on the localization and the dependence structure of the extremes. Its support is the same as the one of the tail measure μ: the spectral measure puts mass in a direction of $\mathbb{S}_+^{d-1}$ if and only if extreme events appear in this direction.

2.2 Estimating the spectral measure: some issues

If we only focus on the second margin of the couple in (2.4), the convergence becomes

$$\mathbb{P}\left(\mathbf{X}/|\mathbf{X}| \in \cdot \mid |\mathbf{X}| > t\right) \xrightarrow{d} \mathbb{P}(\boldsymbol{\Theta} \in \cdot), \quad t \to \infty . \tag{2.5}$$

This means that the spectral vector $\boldsymbol{\Theta}$ can be approximated by the self-normalized extreme $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$, for t large enough. Therefore, its distribution gathers all information on the localization and the dependence structure of extreme events. Hence, estimating the spectral measure is a crucial (but challenging) problem in multivariate EVT.

A first natural step to study large events is to focus on the subspaces on which this measure puts mass. Indeed, they correspond to the ones where extremes events occur. It is frequent that such events only appear in a few directions of $\mathbb{R}^d_+$. This means that the spectral measure has often a sparse structure. Thus, capturing these directions allows to significantly reduce the dimension. Unfortunately, Equation (2.5) is not helpful to identify this potential sparsity. Indeed, as soon as the spectral measure puts mass on a subspace A of dimension smaller than $d-1$, the weak convergence in Equation (2.5) often fails. On the one hand $\mathbb{P}(\mathbf{\Theta} \in A) > 0$ by assumption, but on the other hand, $\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in A) = 0$ if for instance $\mathbf{X}$ has a density with respect to the Lebesgue measure in $\mathbb{R}^d_+$

Let us develop a relevant example. Recall that $\mathbf{e}_1, \dots, \mathbf{e}_d$ denote the unit vectors of the canonical basis of $\mathbb{R}^d$. If the spectral measure only puts mass on $\sqcup_{1 \le j \le d} \{\mathbf{e}_j\}$, we say that the extremes are asymptotically independent. This means that there is never more than one direction which contributes to the extremal behavior of the data. In this case, the spectral measure concentrates on the axis: $\mathbb{P}(\mathbf{\Theta} \in \sqcup_{1 \le j \le d} \{\mathbf{e}_j\}) = 1$. But except for a degenerate vector $\mathbf{X}$, the probability $\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in \sqcup_{1 \le j \le d} \{\mathbf{e}_j\})$ is equal to 0. In practice, real multivariate data never concentrate on the axis, and the study of the spectral measure can not be done through the self-normalized extreme $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$, even for t very large.

This example shows that Equation (2.5) is not helpful to study the support of the spectral vector $\mathbf{\Theta}$. The self-normalized extreme $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$ does not inform on the possible sparsity of $\mathbf{\Theta}$. This kind of problems arises since the spectral measure may put mass on subspaces of zero Lebesgue measure whereas the data generally does not concentrate on such subspaces. Our goal is thus to circumvent this problem by using another projection. This projection has to capture the dependence structure of extremes by taking into account the potential sparsity of the spectral measure. Basically, it has to be more flexible toward the weak convergence on subspaces of dimension smaller than $d-1$.

The solution we propose in this article in to replace the quantity $\mathbf{X}/|\mathbf{X}|$ by the euclidean projection onto the simplex of $\mathbf{X}/t$. To this end, we have to adapt Equation (2.4). This is the aim of next subsection.

From now on, $|\cdot|$ denotes the ℓ^1-norm and $\mathbb{S}^{d-1}_+$ denotes the simplex in dimension d:

$$\mathbb{S}^{d-1}_+ := \{\mathbf{x} \in \mathbb{R}^d_+,\, x_1 + \dots + x_d = 1\}\,.$$

More generally $\mathbb{S}^{d-1}_+(z) := \{\mathbf{x} \in \mathbb{R}^d_+,\, x_1 + \dots + x_d = z\}$ for $z > 0$.

2.3 The euclidean projection onto the simplex

In the subsection, we introduce the euclidean projection onto the simplex. For more details, see Duchi et al. (2008) and the references therein.

Let $z > 0$ and $\mathbf{v} \in \mathbb{R}^d_+$. We consider the following optimization problem:

$$\underset{\mathbf{w}}{\text{minimize}}\ \frac{1}{2}|\mathbf{w} - \mathbf{v}|_2^2 \quad \text{s.t.} \quad \sum_{i=1}^d w_i = z\,. \tag{2.6}$$

Since $\mathbf{v} \ge 0$, the minimization problem (2.6) is equivalent to

$$\underset{\mathbf{w}}{\text{minimize}}\ \frac{1}{2}|\mathbf{w} - \mathbf{v}|_2^2 \quad \text{s.t.} \quad \sum_{i=1}^d w_i = z, \quad w_i \ge 0\,.$$

(see Duchi et al. (2008), Lemma 3). The Lagrangian of this problem and the complementary slackness KKT condition imply that this problem has a unique solution $\mathbf{w} \in \mathbb{R}^d_+$ which satisfies $w_i = (v_i - \lambda_{\mathbf{v},z})_+$ for $\lambda_{\mathbf{v},z} \in \mathbb{R}$. The constant $\lambda_{\mathbf{v},z}$ is defined by the relation $\sum_{1 \le i \le d} (v_i - \lambda_{\mathbf{v},z})_+ = z$.

Hence, the following application has been defined:

$$\begin{array}{rcl} \pi_z & : & \mathbb{R}^d_+ \rightarrow \mathbb{S}^{d-1}_+(z) \\ & & \mathbf{v} \mapsto \mathbf{w} = (\mathbf{v} - \lambda_{\mathbf{v},z})_+ . \end{array}$$

The application π_z is called the *projection onto the positive sphere* $\mathbb{S}^{d-1}_+(z)$. An linear time algorithm which computes $\pi_z(\mathbf{v})$ for $\mathbf{v} \in \mathbb{R}^d_+$ and $z > 0$ is given in Duchi et al. (2008). We include it for completeness.

Data: A vector $\mathbf{v} \in \mathbb{R}^d_+$ and a scalar $z > 0$
Result: The projected vector $\mathbf{w} = \pi(\mathbf{v})$
Initialize $U = \{1, \dots, d\}$, $s = 0$, $\rho = 0$;
while $U \neq \emptyset$ **do**
  Pick $k \in U$ at random;
  Partition U: $G = \{j \in U, v_j \geq v_k\}$ and $L = \{j \in U, v_j < v_k\}$;
  Calculate $\Delta\rho = |G|$, $\Delta s = \sum_{j \in G} v_j$;
  if $(s + \Delta s) - (\rho + \Delta\rho)v_k < z$ **then**
    $s = s + \Delta s$;
    $\rho = \rho + \Delta\rho$;
    $U \leftarrow L$;
  else
    $U \leftarrow G \setminus \{k\}$;
  end
end
Set $\theta = (s - z)/\rho$.;
Define $\mathbf{w}$ s.t. $w_i = v_i - \theta$.

Algorithm 1: Linear time projection onto the positive sphere $\mathbb{S}^{d-1}_+(z)$.

The linear complexity of this algorithm is essential. Indeed, multivariate extremes have already been studied in low dimensions, especially in two dimensions (for instance in Einmahl et al. (2001) or Einmahl and Segers (2009)). But when the dimension growths, the study of large events becomes a difficult issue. Thus, this linear time algorithm is a non-negligible advantage in order to use this projection to study extreme events.

Remark 1. For a vector $\mathbf{v} \in \mathbb{R}^d_+$, the quantity ρ which appears in Algorithm 1 corresponds to the number of positive coordinates of $\pi(\mathbf{v})$. It can be defined as

$$\rho = \max\left\{ j \in \{1, \dots, d\},\ \mu_j - \frac{1}{j}\left(\sum_{r=1}^{j} \mu_j - z \right) > 0 \right\}, \tag{2.7}$$

where $\mu_1 \geq \ldots \geq \mu_d$ denote the order coordinates of $\mathbf{v} = (v_1, \dots, v_d)$, see Duchi et al. (2008), Lemma 2. The integer ρ will be crucial is many proofs.

Note that the projection satisfies the relation $\pi_z(\mathbf{v}) = z\pi_1(\mathbf{v}/z)$ for all $\mathbf{v} \in \mathbb{R}^d_+$ and $z > 0$. This is why we mainly focus on the projection π_1 onto the simplex $\mathbb{S}^{d-1}_+$. In this case, we shortly denote π for π_1 and $\lambda_{\mathbf{v}}$ for $\lambda_{\mathbf{v},1}$:

$$\begin{array}{rcl} \pi & : & \mathbb{R}^d_+ \rightarrow \mathbb{S}^{d-1}_+ \\ & & \mathbf{v} \mapsto (\mathbf{v} - \lambda_{\mathbf{v}})_+ . \end{array}$$

An illustration of π for $d = 2$ is given in Figure 1.

We now list some results on the projection. A first straightforward result if that the projection preserves the order of the coordinates in the following sense: if $v_{\sigma(1)} \geq \ldots \geq v_{\sigma(d)}$ for a permutation σ, then $\pi(\mathbf{v})_{\sigma(1)} \geq \ldots \geq \pi(\mathbf{v})_{\sigma(d)}$ for the same permutation.

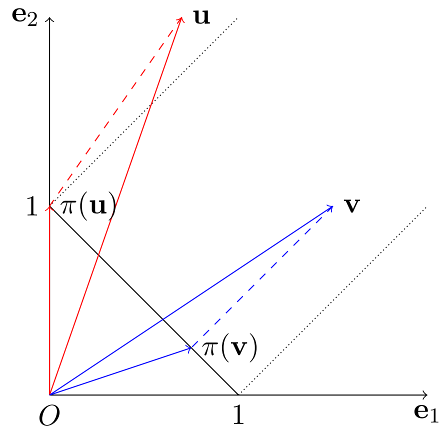

Figure 1: The euclidean projection onto the simplex $\mathbb{S}^1_+$.

Besides, recall that the projection is used to deal with the weak convergence's issue in the spectral measure's definition (2.4). The idea is to substitute the quantity $\mathbf{X}/|\mathbf{X}|$ in (2.4) for $|\cdot| = |\cdot|_1$ by $\pi(\mathbf{X}/t)$ and to manage to get same convergence results. A natural way to do this relies on the continuous mapping theorem. The continuity of the projection π is thus a crucial point.

Lemma 1. *The application π is continuous.*

Another important property satisfied by the projection is the following one.

Lemma 2. *If $0 < z \leq z'$, then $\pi_z \circ \pi_{z'} = \pi_z$.*

This means that projecting onto a sphere and then onto a smaller one is the same as directly projecting onto the smaller sphere. This lemma will be useful to prove some technical results gathering the projection π and regular variation.

Finally, in order to study the sparse structure of extreme events, we are interested in computing probabilities like $\mathbb{P}(\mathbf{\Theta}_I = 0)$ for $I \subset \{1, \dots, d\}$. To this end, next lemma will be helpful.

Lemma 3. *Let $\mathbf{v} \in \mathbb{R}^d_+$ and $I \subset \{1, \dots, d\}$. The following equivalences hold:*

$$\pi(\mathbf{v})_I = 0 \quad \textit{if and only if} \quad 1 \leq \min_{i \in I} \sum_{k=1}^d (v_k - v_i)_+ \,, \tag{2.8}$$

and

$$\pi(\mathbf{v})_I = 0 \textit{ and } \pi(\mathbf{v})_{I^c} > 0 \quad \textit{if and only if} \quad \begin{cases} \max_{i \in I^c} \sum_{j \in I^c} (v_j - v_i) < 1 \,, \\ \min_{i \in I} \sum_{j \in I^c} (v_j - v_i) \geq 1 \,. \end{cases} \tag{2.9}$$

If $\pi(\mathbf{v}) > 0$ (that is, if $I = \emptyset$), then $\pi(\mathbf{v})$ has necessary the following form:

$$\pi(\mathbf{v}) = \mathbf{v} - \frac{1}{d}\Big(\sum_{k=1}^d v_k - 1\Big) = \mathbf{v} - \frac{|\mathbf{v}| - 1}{d} \,.$$

Thus, for $\mathbf{x} \geq 0$, we have the following characterization:

$$\pi(\mathbf{v}) > \mathbf{x} \quad \text{if and only if} \quad \mathbf{v} > \mathbf{x} + \frac{|\mathbf{v}| - 1}{d} \,. \tag{2.10}$$

This equivalence will be of constant use in the sequel.

Remark 2. Note that the projection π is not homogeneous. Recall that a function f is said to be homogeneous if there exists $q > 0$ such that for all $t > 0$, $f(t\mathbf{x}) = t^q f(\mathbf{x})$. If f is a continous and homogeneous function and $\mathbf{X}$ is a regularly random vector in $\mathbb{R}^d_+$ with tail index $\alpha > 0$, then the random vector $f(\mathbf{X})$ is regularly varying with tail index α/q (see Jessen and Mikosch (2006)). Such a result cannot be used for the euclidean projection onto the simplex.

The theoretical framework being defined, we now want to use the projection in the context of regularly varying random vectors. This is the purpose of next section.

3 Spectral measure and projection

The purpose of this section is twofold. In the first part, we use the euclidean projection to introduce a new convergence based on (2.4). This new convergence brings out an angular limit vector which differs from the spectral vector. Some results on this limit and its relation with the spectral vector are introduced. The second part is dedicated to sparsity results of this new limit. Finally, we study two classical particular cases of the multivariate EVT: complete dependence and asymptotic independence.

3.1 Regular variation and projection

From now on, and till the end of section 3, we consider a regularly varying random vector $\mathbf{X}$ on $\mathbb{R}^d_+$:

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \frac{\mathbf{X}}{|\mathbf{X}|}\right) \in \cdot \;\middle|\; |\mathbf{X}| > t\right) \xrightarrow{d} \mathbb{P}((Y, \boldsymbol{\Theta}) \in \cdot), \quad t \to \infty\,. \tag{3.1}$$

We know that in this case there exists $\alpha > 0$ such that Y follows a Pareto(α) distribution and also that the limits Y and $\boldsymbol{\Theta}$ are independent. As explained in Section 2, convergence (3.1) is not helpful to capture the possible sparse structure of the spectral vector $\boldsymbol{\Theta}$. Our idea is to substitute the self-normalized extremes $\mathbf{X}/|\mathbf{X}|$ by another vector on the simplex which better highlights the sparsity.

Let us give an intuitive idea to see how the euclidean projection can solve this kind of issue. For $I = \{i_1, \dots, i_r\} \subset \{1, \dots, d\}$, we consider $\mathrm{Vect}(\mathbf{e}_I) = \mathrm{Vect}(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_r})$, the subspace of $\mathbb{R}^d_+$ generated by the r vectors $\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_r}$. As explained in Section 2, if $r < d$, then the quantity $\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in \mathrm{Vect}(\mathbf{e}_I) \mid |\mathbf{X}| > t)$ always equals 0 (except for degenerate cases), whereas $\mathbb{P}(\boldsymbol{\Theta} \in \mathrm{Vect}(\mathbf{e}_I))$ could be positive. The projection π allows to give more weight to such subspaces $\mathrm{Vect}(\mathbf{e}_I)$.

For instance, on Figure 1, the vector $\mathbf{u}$ does not have a null $\mathbf{e}_1$-coordinate but the projected vector $\pi(\mathbf{u})$ does. In this two dimensional case, a positive vector $\mathbf{w}$ satisfies $\pi(\mathbf{w})_1 = 0$ if and only if $w_2 \geq w_1 + 1$. Essentially, estimating the probability $\mathbb{P}(\Theta_1 = 0)$ needs to replace the set $\{\mathbf{w},\, w_1/|\mathbf{w}| = 0\}$ by a larger one which has not zero Lebesgue measure. Our idea here is to use the space $\{\pi(\mathbf{w})_1 = 0\} = \{\mathbf{w}, w_2 \geq w_1 + 1\}$.

Remark 3. This idea of substituing subspaces of zero Lebesgue measure by closer subspaces but of positive Lebesgue measure has already been used. For instance, Goix et al. defined ϵ-thickened rectangles

$$C^\epsilon_\beta = \{\mathbf{v} \in \mathbb{R}^d_+;\, |\mathbf{v}|_\infty \geq 1;\, \forall j \in \beta,\, v_j > \epsilon;\, \forall j \notin \beta,\, v_j \leq \epsilon\}\,,$$

for $\beta \subset \{1, \dots, d\}$ (see Goix et al. (2017)). They worked on the tail measure μ (see (2.1)) and showed that $\mu(C^\epsilon_\beta) \to \mu(C_\beta)$ when $\epsilon \to 0$, where

$$C_\beta = \{\mathbf{v} \in \mathbb{R}^d_+;\, |\mathbf{v}|_\infty \geq 1;\, \forall j \in \beta,\, v_j > 0;\, \forall j \notin \beta,\, v_j = 0\}\,.$$

Unfortunately, these considerations are based on a parameter $\epsilon > 0$ which has to be choosen in practice. One of the advantages of the projection π is that it is does not need any parameter.

With this in mind, we substitute the classical projection $\mathbf{X}/|\mathbf{X}|$ by $\pi(\mathbf{X}/t)$. The first step is to see how this affects the spectral vector. Since the projection π is continuous, the following convergence holds:

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \pi\left(\frac{\mathbf{X}}{t}\right)\right) \in \cdot \,\middle|\, |\mathbf{X}| > t\right) \xrightarrow{d} \mathbb{P}((Y, \pi(Y\mathbf{\Theta})) \in \cdot), \quad t \to \infty\,. \tag{3.2}$$

The limit of the angular component is now $\pi(Y\mathbf{\Theta})$. Of course, we lose independence between the radial component Y and the angular component $\pi(Y\mathbf{\Theta})$ of the limit. The dependence relation between both components will be detailed in Proposition 4.

The aim of this section is to study to what extent the new angular limit $\pi(Y\mathbf{\Theta})$ differs from the spectral vector $\mathbf{\Theta}$. Thus, in the rest of this subsection we establish a relation between these both vectors.

Set $\mathbf{Z} = \pi(Y\mathbf{\Theta}) \in \mathbb{S}^{d-1}_+$. Our aim is to explicit the relation between $\mathbf{\Theta}$ and $\mathbf{Z}$. We define the function $G_{\mathbf{Z}}$ by

$$G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\mathbf{Z} > \mathbf{x}) = \mathbb{P}(Z_1 > x_1, \ldots, Z_d > x_d), \quad \mathbf{x} \in \mathbb{R}^d\,. \tag{3.3}$$

The function $G_{\mathbf{Z}}$ characterizes the distribution of $\mathbf{Z}$. However, note that there is no simple relation between $G_{\mathbf{Z}}$ and the cumulative distribution function of $\mathbf{Z}$ as soon as $d \geq 2$. Since $\mathbf{Z} \in \mathbb{S}^{d-1}_+$, we only focuses on $G_{\mathbf{Z}}(\mathbf{x})$ for $\mathbf{x}$ in $\mathbb{R}^d_+$ such that $\sum_j x_j < 1$, this means for $\mathbf{x} \in \mathcal{B}(0,1) \cap \mathbb{R}^d_+$, where $\mathcal{B}(0,1)$ denotes the (open) unit ball for the ℓ^1-norm. Thus, we consider

$$G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}})\,,$$

where the borelians $A_{\mathbf{x}}$ are defined by

$$A_{\mathbf{x}} = \{\mathbf{u} \in \mathbb{S}^{d-1}_+,\, x_1 < u_1, \ldots, x_d < u_d\}\,. \tag{3.4}$$

Since the family $\mathcal{A} = \{A_{\mathbf{x}},\, \mathbf{x} \in \mathcal{B}(0,1) \cap \mathbb{R}^d_+\}$ generates the borelians of the simplex $\mathbb{S}^{d-1}_+$, the distribution of $\mathbf{Z}$ is completely characterized by $G_{\mathbf{Z}}(\mathbf{x})$ for $\mathbf{x} \in \mathcal{B}(0,1) \cap \mathbb{R}^d_+$.

With Equation (2.10), we can express the condition $\mathbf{Z} > \mathbf{x}$ in terms of $\mathbf{\Theta}$. This is the aim of next Proposition.

Proposition 1. *Let $\mathbf{X}$ be a regularly varying random vector of $\mathbb{R}^d_+$ with tail index $\alpha > 0$ and spectral vector $\mathbf{\Theta}$. For $\mathbf{x} \in \mathcal{B}(0,1) \cap \mathbb{R}^d_+$, such that for all $j = 1, \ldots, d$, $x_j \neq 1/d$, define $J_+ = \{j,\, x_j > 1/d\}$ and $J_- = \{j,\, x_j < 1/d\}$. Then, we have*

$$G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{E}\left[\left(1 \wedge \min_{j \in J_+}\left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)^\alpha_+ - \max_{j \in J_-}\left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)^\alpha_+\right)_+\right]\,, \tag{3.5}$$

with $G_{\mathbf{Z}}$ defined in (3.3).

Proposition 1 gives an interesting relation between the distribution of $\mathbf{Z}$ and the one of $\mathbf{\Theta}$. Unfortunately, its complexity makes it difficult to use. But specific choices for $\mathbf{x}$ will give some useful results.

Remark 4. Note that (3.5) still holds if there exists j_0 such that $\Theta_{j_0} = 0$ a.s. In this case, we know that $Z_{j_0} = \pi(Y\mathbf{\Theta})_{j_0} = 0$. Equation (3.5) gives the same result since, for $x_{j_0} = 0$, we obtain

$$\max_{j \in J_-}\left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)^\alpha_+ \geq \left(\frac{1/d - \Theta_{j_0}}{1/d - x_{j_0}}\right)^\alpha_+ = 1\,,$$

so that $G_{\mathbf{Z}}(\mathbf{x}) = 0$.

A convenient particular case is the one where $\mathbf{x}$ satisfies $\mathbf{x} < 1/d$. There, we obtain

$$G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{E}\left[1 - \max_{1 \le j \le d} \left(\frac{1/d - \Theta_j}{1/d - x_j}\right)^{\alpha}\right].$$

In particular, for $\mathbf{x} = \mathbf{0}$, we get

$$G_{\mathbf{Z}}(\mathbf{0}) = 1 - \mathbb{E}\left[\max_{1 \le j \le d} (1 - d\Theta_j)^{\alpha}\right]. \tag{3.6}$$

Thus, the probability for $\mathbf{Z}$ to have a null component is

$$\mathbb{P}(\exists j = 1, \dots, d\,, Z_j = 0) = \mathbb{E}\left[\max_{1 \le j \le d} (1 - d\Theta_j)^{\alpha}\right]. \tag{3.7}$$

This quantity is null if and only if for all $j = 1, \dots, d$, $\Theta_j = 1/d$ a.s. and is equal to 1 if and only if $\min_{1 \le j \le d} \Theta_j = 0$ a.s. It means that the new angular limit $\mathbf{Z}$ is more likely to be sparse. In particular, all usual spectral models on $\mathbf{\Theta}$ that are not supported on the axis are not suitable for $\mathbf{Z}$. The goal of the next subsection is to study more into details this sparse structure of $\mathbf{Z}$.

3.2 Sparse structure of Z

Since the projection is introduced in order to better capture the sparse structure of the extremes, we give here different results of sparsity for the angular component $\mathbf{Z} = \pi(Y\mathbf{\Theta})$. The general aim is thus to compute probabilities like $\mathbb{P}(\mathbf{Z}_I = 0)$ or $\mathbb{P}(\mathbf{Z}_I = 0, \mathbf{Z}_{I^c} > 0)$, for $I \subset \{1, \dots, d\}$, in order to generalize Equation (3.7). These kinds of results are developed in the next Proposition.

Proposition 2. *Let $\mathbf{X}$ be a regularly varying random vector of $\mathbb{R}^d_+$ with spectral vector $\mathbf{\Theta}$ and tail index $\alpha > 0$. Set $\mathbf{Z} = \pi(Y\mathbf{\Theta})$, where Y follows a Pareto(α) distribution independent of $\mathbf{\Theta}$. For any $I \subsetneq \{1, \dots, d\}$, we have*

$$\mathbb{P}(\mathbf{Z}_I = 0) = \mathbb{E}\left[\min_{j \in I} \left(\sum_{k=1}^{d} (\Theta_k - \Theta_j)_+\right)^{\alpha}\right], \tag{3.8}$$

and

$$\mathbb{P}(\mathbf{Z}_I = 0,\, \mathbf{Z}_{I^c} > 0) = \mathbb{E}\left[\left(\min_{j \in I} \left(\sum_{k \in I^c} (\Theta_k - \Theta_j)_+\right)^{\alpha} - \max_{j \in I^c} \left(\sum_{k \in I^c} (\Theta_k - \Theta_j)_+\right)^{\alpha}\right)_+\right]. \tag{3.9}$$

Remark 5. In particular, Equation (3.8) imply that

$$\begin{aligned}\mathbb{P}(\mathbf{Z}_I = 0) &= \mathbb{E}\left[\min_{j \in I} \left(\sum_{k=1}^{d} (\Theta_k - \Theta_j)_+\right)^{\alpha} \mathbb{1}_{\{\mathbf{\Theta}_I = 0\}}\right] + \mathbb{E}\left[\min_{j \in I} \left(\sum_{k=1}^{d} (\Theta_k - \Theta_j)_+\right)^{\alpha} \mathbb{1}_{\{\exists i \in I,\, \Theta_i > 0\}}\right] \\ &= \mathbb{P}(\mathbf{\Theta}_I = 0) + \mathbb{E}\left[\min_{j \in I} \left(\sum_{k=1}^{d} (\Theta_k - \Theta_j)_+\right)^{\alpha} \mathbb{1}_{\{\exists i \in I,\, \Theta_i > 0\}}\right] \ge \mathbb{P}(\mathbf{\Theta}_I = 0)\,. \end{aligned} \tag{3.10}$$

As expected, the vector $\mathbf{Z}$ is more likely to be sparse than the spectral vector $\mathbf{\Theta}$. Besides, we have the following equivalence:

$$\min_{j \in I} \sum_{k=1}^{d} (\Theta_k - \Theta_j)_+ = 1 \quad \text{a.s.} \quad \text{if and only if} \quad \forall j \in I,\, \Theta_j = 0 \quad \text{a.s.}$$

These results lead to the following corollary.

Corollary 1. *With the same notations as in Proposition 2, we have*

$$\mathbb{P}(\mathbf{Z}_I = 0) \geq \mathbb{P}(\mathbf{\Theta}_I = 0)\,.$$

Moreover, $\mathbb{P}(\mathbf{Z}_I = 0) = 1$ *if and only if* $\mathbb{P}(\mathbf{\Theta}_I = 0) = 1$.

This corollary states that the euclidean projection tends to introduce more sparsity. But eventually both angular vectors $\mathbf{\Theta}$ and $\mathbf{Z}$ put mass on the same subspaces. This means that we do not lose any information on the support of the spectral measure by studying $\mathbf{Z}$ instead of $\mathbf{\Theta}$.

Another main advantage of using the euclidean projection is gathered in the next proposition. Recall that with the classical projection $\mathbf{X}/|\mathbf{X}|$, if $I \subset \{1, \dots, d\}$ and $B_I = \{\mathbf{x} \in \mathbb{R}^d_+,\, \mathbf{x}_I = 0\}$, then the convergence $\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in B_I \mid |\mathbf{X}| > t) \to \mathbb{P}(\mathbf{\Theta} \in B_I)$ often does not hold (see Section 2). On the contrary, based on the second component of convergence (3.2), we know that $\mathbb{P}(\pi(\mathbf{X}/t) \in A \mid |\mathbf{X}| > t) \to \mathbb{P}(\pi(Y\mathbf{\Theta}) \in A)$, as soon as $\mathbb{P}(Y\mathbf{\Theta} \in \partial\pi^{-1}(A)) = 0$. Next proposition states that the sets B_I satisfy this condition. Hence, the projection π allows to circumvent the issue discussed in Section 2.

Proposition 3. *Let* $\mathbf{X}$ *be a regularly varying random vector in* $\mathbb{R}^d_+$ *with spectral vector* $\mathbf{\Theta}$ *and tail index* $\alpha > 0$. *Set* $\mathbf{Z} = \pi(Y\mathbf{\Theta})$, *where* Y *follows a Pareto(α) distribution independent of* $\mathbf{\Theta}$. *For any* $I \subset \{1, \dots, d\}$, *we have*

$$\mathbb{P}\left(\pi\left(\frac{\mathbf{X}}{t}\right)_I = 0 \,\middle|\, |\mathbf{X}| > t\right) \to \mathbb{P}(\mathbf{Z}_I = 0)\,. \tag{3.11}$$

This means that the sparse structure of $\mathbf{Z}$ can be inferred through the euclidean projection $\pi(\mathbf{X}/t)$. We insist that this convergence does not hold if we replace $\mathbf{Z}$ by $\mathbf{\Theta}$ and $\pi(\mathbf{X}/t)$ by $\mathbf{X}/|\mathbf{X}|$.

All in all this shows that using the projected vector $\pi(\mathbf{X}/t)$ instead of the self-normalized vector $\mathbf{X}/|\mathbf{X}|$ has several advantages. Firstly, it better captures the sparse structure of the extremes (see Corollary 1). Secondly, since the directions on which $\mathbf{Z}$ puts mass are the same as those where $\mathbf{\Theta}$ puts mass, we do not lose any information (Corollary 1). Finally, Proposition 3 states that the projection π is more efficient with the weak convergence since convergence (3.2) holds for some zero Lebesgue measure subspaces whereas convergence (3.1) does not.

We end this subsection by applying Proposition 2 with some particular choices of $I \subset \{1, \dots, d\}$. Firstly, if we consider the case where $I = \{1, \dots, d\}$, then we obtain the probability that all coordinates are positive. This has already been computed in (3.6), it is equal to $G_{\mathbf{Z}}(\mathbf{0}) = 1 - \mathbb{E}\left[\max_{1 \leq j \leq d}(1 - d\Theta_j)^\alpha\right]$.

Another particular case of Proposition 2 is the one where I^c corresponds to a single coordinate j_0. In this case, since $\mathbf{Z}$ belongs to the simplex, both probabilities $\mathbb{P}(\mathbf{Z}_I = 0)$ and $\mathbb{P}(\mathbf{Z}_I = 0, Z_{j_0} > 0)$ are equal. Their common value corresponds to the probability that $\mathbf{Z}$ is concentrated on the j_0-th axis. It is equal to

$$\mathbb{P}(Z_{j_0} = 1) = \mathbb{E}\left[\min_{j \neq j_0}(\Theta_{j_0} - \Theta_j)^\alpha_+\right]. \tag{3.12}$$

Note that this quantity can also be computed with Equation (2.9) of Lemma 3. Moreover we have the following equalities:

$$\begin{aligned}\mathbb{P}(Z_{j_0} = 1) &= \mathbb{E}\left[\min_{j \neq j_0}(\Theta_{j_0} - \Theta_j)^\alpha_+ \mathbb{1}_{\{\Theta_{j_0} = 1\}}\right] + \mathbb{E}\left[\min_{j \neq j_0}(\Theta_{j_0} - \Theta_j)^\alpha_+ \mathbb{1}_{\{\Theta_{j_0} < 1\}}\right] \\ &= \mathbb{P}(\Theta_{j_0} = 1) + \mathbb{E}\left[\min_{j \neq j_0}(\Theta_{j_0} - \Theta_j)^\alpha_+ \mathbb{1}_{\Theta_{j_0} < 1}\right] \geq \mathbb{P}(\Theta_{j_0} = 1)\,.\end{aligned}$$

This shows again that the vector $\mathbf{Z}$ is more likely to be sparse than the spectral vector $\mathbf{\Theta}$.

3.3 Special cases: asymptotic independence and complete dependence

We study here two particular cases in multivariate EVT. The first one is the complete dependence's case, which is defined by the relation $\mathbb{P}(\boldsymbol{\Theta} = 1/d) = 1$. It means that the spectral measure is a Dirac point at $(1/d, \ldots, 1/d)$. In terms of extremes, it means that all coordinates simultaneously contribute to large events. Since

$$\mathbb{P}(\mathbf{Z} = 1/d) = \mathbb{P}\left(Y\boldsymbol{\Theta} - (Y-1)/d = 1/d\right) = \mathbb{P}(\boldsymbol{\Theta} = 1/d)\,,$$

we have the equivalence

$$\mathbf{Z} = 1/d \quad \text{a.s.} \quad \text{if and only if} \quad \boldsymbol{\Theta} = 1/d \quad \text{a.s.} \tag{3.13}$$

The other case is the asymptotic independence's one, which appears when $\boldsymbol{\Theta}$ only concentrates on the axis. It means that $\mathbb{P}(\boldsymbol{\Theta} \in \sqcup_{1\leq k\leq d}\, \mathbf{e}_k) = 1$. Note that this case has already been partially discussed in Section 2 to illustrate the difficulty to estimate the spectral vector $\boldsymbol{\Theta}$. As for the complete dependence's case, we want to express the asymptotic independence in terms of $\mathbf{Z}$. To this end, we write

$$\begin{aligned}
\mathbb{P}(\exists 1 \leq i \leq d,\, Z_i = 1) &= \mathbb{P}(\exists 1 \leq i \leq d,\, \forall j \neq i,\, Z_j = 0) \\
&= \mathbb{P}\left(\exists 1 \leq i \leq d,\, \forall j \neq i,\, 1 \leq Y(\Theta_i - \Theta_j)_+\right) \\
&= \mathbb{P}\Big(\exists 1 \leq i \leq d,\, Y^{-\alpha} \leq \min_{j\neq i}(\Theta_i - \Theta_j)_+^\alpha\Big) \\
&= \mathbb{P}\Big(Y^{-\alpha} \leq \max_{1\leq i\leq d}\min_{j\neq i}(\Theta_i - \Theta_j)_+^\alpha\Big) \\
&= \int_0^1 \mathbb{P}\Big(u \leq \max_{1\leq i\leq d}\min_{j\neq i}(\Theta_i - \Theta_j)_+^\alpha\Big)\,\mathrm{d}u \\
&= \mathbb{E}\Big[\max_{1\leq i\leq d}\min_{j\neq i}(\Theta_i - \Theta_j)_+^\alpha\Big]\,.
\end{aligned}$$

Thus, since $\max_{1\leq i\leq d}\min_{j\neq i}(\Theta_i - \Theta_j)_+^\alpha \leq 1$, we have the equivalence

$$\mathbb{P}(\exists 1 \leq i \leq d,\, Z_i = 1) = 1 \quad \text{if and only if} \quad \mathbb{P}\Big(\max_{1\leq i\leq d}\min_{j\neq i}(\Theta_i - \Theta_j)_+^\alpha = 1\Big) = 1\,.$$

This last probability can be rewritten as follows:

$$\mathbb{P}\Big(\max_{1\leq i\leq d}\min_{j\neq i}(\Theta_i - \Theta_j)_+^\alpha = 1\Big) = \mathbb{P}\Big(\exists 1 \leq i \leq d,\, \min_{j\neq i}(\Theta_i - \Theta_j)_+ = 1\Big) = \mathbb{P}\Big(\exists 1 \leq i \leq d,\, \Theta_i = 1\Big)\,.$$

This proves the equivalence between $\mathbb{P}(\exists 1 \leq i \leq d,\, Z_i = 1) = 1$ and $\mathbb{P}\left(\exists 1 \leq i \leq d,\, \Theta_i = 1\right) = 1$. Based on this result and Proposition 3, it is thus possible to test asymptotic independence by studying $\pi(\mathbf{X}/t)$. This justifies afterwards the choice of the projection π to study the extremal dependence structure.

All in all, this means that these two classical cases of multivariate EVT can be studied through the distribution of $\mathbf{Z}$. We do not lose any information by studying $\mathbf{Z}$ instead of $\boldsymbol{\Theta}$ in the asymptotically independent and completely dependent settings.

4 Sparse regular variation

We consider in this section a random vector $\mathbf{X}$ in $\mathbb{R}_+^d$. In Section 3, we assumed that $\mathbf{X}$ was regularly varying. In this case convergence (3.2) holds and allows to study the properties of $\mathbf{Z} = \pi(Y\boldsymbol{\Theta})$. Our aim is now to establish a converse result. Thus, we do not assume anymore that $\mathbf{X}$ is regularly varying. We only start from convergence (3.2) which encourages to introduce the following definition.

Definition 1 (Sparse regular variation). *A random vector* $\mathbf{X}$ *on* $\mathbb{R}^d_+$ *is sparsely regularly varying if there exist a random vector* $\mathbf{Z}$ *defined on the simplex* $\mathbb{S}^{d-1}_+$ *and a random variable* Y *such that*

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \pi\left(\frac{\mathbf{X}}{t}\right)\right) \in \cdot \;\middle|\; |\mathbf{X}| > t\right) \xrightarrow{d} \mathbb{P}((Y, \mathbf{Z}) \in \cdot), \quad t \to \infty. \tag{4.1}$$

In this case, the general theory of regular variation states that there exists $\alpha > 0$ such that Y is Pareto(α) distributed. Note that we lose independence between the radial component of the limit Y and the angular one $\mathbf{Z}$. The dependence structure between Y and $\mathbf{Z}$ will be detailed in Proposition 4.

A direct consequence of this definition is that if the random vector $\mathbf{X}$ is regularly varying with spectral vector $\boldsymbol{\Theta}$ and radial limit Y, then, by continuity of π, $\mathbf{X}$ is sparsely regularly varying with angular limit $\mathbf{Z} = \pi(Y\boldsymbol{\Theta})$.

In all this section, we consider a sparsely regularly varying random vector $\mathbf{X}$. Recall that the function $G_{\mathbf{Z}}$ is defined by $G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\mathbf{Z} > \mathbf{x})$. However, note that for the moment we can not write $G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\pi(Y\boldsymbol{\Theta}) > \mathbf{x})$ since there is no guarantee of the existence of $\boldsymbol{\Theta}$. Our aim is twofold. The first goal is to study the dependence between the radial limit Y and the angular limit $\mathbf{Z}$ in (4.1). Secondly, we prove that under some assumptions on $G_{\mathbf{Z}}$ the vector $\mathbf{X}$ is regularly varying.

The following proposition gives the explicit dependence structure between $\mathbf{Z}$ and Y.

Proposition 4. *Let* $\mathbf{X}$ *be a sparsely regularly varying random vector on* $\mathbb{R}^d_+$*. Then, for all* $r \geq 1$,

$$\mathbf{Z} \mid Y > r \stackrel{d}{=} \pi(r\mathbf{Z}). \tag{4.2}$$

As mentioned before, we do not have independence between the angular component $\mathbf{Z}$ and the radial one Y. However, the dependence between $\mathbf{Z}$ and Y is completely determined by Equation (4.2), and will be helpful in the proof of next theorem.

Our aim is now to prove that, under some conditions, if $\mathbf{X}$ is a sparsely regularly varying vector, then $\mathbf{X}$ is regularly varying. Note that if convergence (4.1) holds, then $|\mathbf{X}|$ is regularly varying. So the only step is to prove the convergence of the angular component, that is, of the self-normalized extreme $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$ when $t \to \infty$. We will base our proof on the following lemma.

Lemma 4. *Let* $\mathbf{X}$ *be a random vector on* $\mathbb{R}^d_+$ *and* $\alpha > 0$*. The following assumptions are equivalent.*

1. $\mathbf{X}$ *is regularly varying with tail index* α.
2. $|\mathbf{X}|$ *is regularly varying with tail index* α *and there exists a finite measure* l *on* $\mathbb{S}^{d-1}_+$ *such that*

$$\lim_{\epsilon \to 0} \liminf_{t \to \infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; |\mathbf{X}| > t\right) = l(A), \tag{4.3}$$

and

$$\lim_{\epsilon \to 0} \limsup_{t \to \infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; |\mathbf{X}| > t\right) = l(A), \tag{4.4}$$

for all $A \in \mathcal{B}(\mathbb{S}^{d-1}_+)$ *such that* $l(\partial A) = 0$.

In this case, $l(A) = \alpha \mathbb{P}(\boldsymbol{\Theta} \in A)$*, where* $\boldsymbol{\Theta}$ *is the spectral vector of* $\mathbf{X}$.

Remark 6. The assertion 2. of Lemma 4 can be weakened by taking A in a family of borelians that generates $\mathcal{B}(\mathbb{S}^{d-1}_+)$. In what follows, we will consider the family $\mathcal{A} = \{A_{\mathbf{x}}, \mathbf{x} \in \mathcal{B}(0,1) \cap \mathbb{R}^d_+\}$, where the $A_{\mathbf{x}}$ are defined in (3.4).

Remark 7. In Lemma 4, $|\cdot|$ denotes any norm of $\mathbb{R}^d$, but in what follows we will use this lemma for the ℓ^1-norm.

We now prove that under mild assumptions on $G_{\mathbf{Z}}$, a random vector $\mathbf{X}$ which satisfies (4.1) is regularly varying. The assumptions on $G_{\mathbf{Z}}$ are the following ones:

(A1) The function $G_{\mathbf{Z}}$ is differentiable for almost every $\mathbf{x} \in \mathcal{B}(0,1) \cap \mathbb{R}^d_+$.

(A2) $\mathbb{P}(\mathbf{Z} \in \partial A_{\mathbf{x}}) = 0$ for almost every $\mathbf{x} \in \mathcal{B}(0,1) \cap \mathbb{R}^d_+$.

Let us denote by $\mathcal{Z}(G_{\mathbf{Z}})$ the set of vectors $\mathbf{x}$ in $\mathcal{B}(0,1) \cap \mathbb{R}^d_+$ which satisfy (A1) and (A2). Then, the family $\mathcal{A}_{\mathcal{Z}(G_{\mathbf{Z}})} := \{A_{\mathbf{x}}, \mathbf{x} \in \mathcal{Z}(G_{\mathbf{Z}})\}$ generates the borelians of $\mathbb{S}^{d-1}_+$. If there is no confusion, we will simply write $\mathcal{Z}$ for $\mathcal{Z}(G_{\mathbf{Z}})$ and $\mathcal{A}_{\mathcal{Z}}$ for $\mathcal{A}_{\mathcal{Z}(G_{\mathbf{Z}})}$.

Theorem 1. *Let $\mathbf{X}$ be a vector on $\mathbb{R}^d_+$ that is sparsely regularly varying. Assume that $G_{\mathbf{Z}}(\cdot) = \mathbb{P}(\mathbf{Z} > \cdot)$ satisfies (A1) and (A2). Then $\mathbf{X}$ is regularly varying with spectral vector $\mathbf{\Theta}$ which satisfies*

$$\mathbb{P}(\mathbf{\Theta} \in A_{\mathbf{x}}) = \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + \alpha^{-1} \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d)\,, \tag{4.5}$$

for all $\mathbf{x} \in \mathcal{Z}$.

This shows under mild assumptions the equivalence of regular variation and sparse regular variation. Moreover, the distribution of $\mathbf{Z}$ completely characterizes the one of $\mathbf{\Theta}$. Equation (4.5) completes the result (3.5) obtained in Proposition 1.

Let us summarize the results we obtained. Proposition 1 characterizes the distribution of $\mathbf{Z} = \pi(Y\mathbf{\Theta})$ when $\mathbf{X}$ is regularly varying with spectral vector $\mathbf{\Theta}$. Conversely, suppose that $\mathbf{X}$ is a sparsely regularly varying random vector. Then Theorem 1 states that $\mathbf{X}$ is regularly varying with spectral vector $\mathbf{\Theta}$ satisfying Equation (4.5). This ensures that $\mathbf{Z} = \pi(Y\mathbf{\Theta})$, with Y a Pareto(α) random variable independent of $\mathbf{\Theta}$. In other words, we have an almost complete equivalence between the usual regular variation and sparse regular variation.

5 Discussion

We discuss here a possible model on $\mathbf{Z}$ and its relation with $\mathbf{\Theta}$. For $d \geq 1$, there are $2^d - 1$ non-empty subsets of $\{1, \ldots, d\}$. If c is one of theses subsets, we denote by $\mathbf{e}_c$ the vector with 1 in position i if $i \in c$ and 0 otherwise. For instance, $\mathbf{e}_{\{1,2,3\}} = (1, 1, 1, 0, \ldots, 0)$, $\mathbf{e}_{\{1,d\}} = (1, 0, \ldots, 0, 1)$, and so on. Note that for all non-empty subsets c of $\{1, \ldots, d\}$, $\mathbf{e}_c/|c|$ belongs to $\mathbb{S}^{d-1}_+$.

We consider the following class of discrete distributions on the simplex:

$$\sum_{\emptyset \neq c \subset \{1,\ldots,d\}} p(c)\, \delta_{\mathbf{e}_c/|c|}\,, \tag{5.1}$$

where $(p(c))_c$ is a $2^d - 1$ vector with nonnegative components summing to 1. This is the device developed in Segers (2012). In terms of stable tail dependence function $l(\mathbf{x}) = \mathbb{E}[\max_j(x_j\Theta_j)]$, this means that

$$l(\mathbf{x}) = \sum_{\emptyset \neq c \subset \{1,\ldots,d\}} \frac{p(c)}{|c|} \max_{j \in c}(x_j)\,,$$

see Segers (2012), Examples 3.4 and 3.5 for more details. Note that this class of distributions includes the complete dependence's and the asymptotic independence's cases developed in Section 3. Moreover, a distribution of this class satisfies both assumptions (A1) and (A2).

The family of distributions (5.1) is stable after multiplying by a positive random variable and projecting onto the simplex with π. Hence, if $\mathbf{\Theta}$ has a distribution of type (5.1), then $\mathbf{Z} = \mathbf{\Theta}$ a.s. Moreover, Equations (3.5) and (4.5) show that there is a bijection between the distribution of $\mathbf{\Theta}$ and the one of $\mathbf{Z}$. Hence, if $\mathbf{Z}$ has a distribution of type (5.1), then $\mathbf{\Theta}$ has the same distribution.

This shows that the family of distributions (5.1) forms an accurate model for the angular vector $\mathbf{Z}$. Indeed, it is stable for the transformation $\mathbf{\Theta} \mapsto \mathbf{Z}$. Besides, the distributions of this class have sparse supports. Finally, they put weights on some particular points of the simplex on which extremes values often concentrate in practice.

Conclusion In this paper, we introduce the notion of sparsely regularly varying random vectors in order to tackle the issues that arise with the classical notion of regular variation in a high dimensional setting. The idea to replace the self-normalized vector $\mathbf{X}/|\mathbf{X}|$ by the projected one $\pi(\mathbf{X}/t)$ allows to better capture the sparsity structure of the extremal dependence. Our main result is the equivalence between sparse regular variation and regular variation under some mild assumptions.

The benefits of this new way of projecting are multiple. The first one is the sparser structure of the new angular vector $\mathbf{Z}$ compared to the one of $\mathbf{\Theta}$. Besides, contrary to the classical regular variation's framework, the sparsity of $\mathbf{Z}$ can be directly captured by studying $\pi(\mathbf{X}/t)$, as stated in Proposition 3. This means that the projection π manages to circumvent to issue of the weak convergence in the definition of regularly random vectors. Finally, the results of Propostion 1 and Theorem 1 state that under some assumptions, there is a bijection between the spectral vector $\mathbf{\Theta}$ and the new angular vector $\mathbf{Z}$.

Practically speaking, the advantages of using the projection π are twofold. Firstly, the euclidean projection onto the simplex does not introduce any parameter to estimate. In Goix et al. (2017) for instance, the introduction of ϵ-thickened rectangles leads to the choice of the optimal parameter ϵ. The same issue arises in Chiapino and Sabourin (2016) with the choice of $\kappa_{\min}$. The only choice that has to be made in our case is the parameter t. At the best of our knowledge, choosing the best threshold t in EVT is an issue that has not been completely overcome yet. Secondly, the algorithm which computes the projection π takes linear time. Hence, the study of extreme events with π can be done in high dimensions.

Finally, we know by Corollary 1 that the directions on which $\mathbf{\Theta}$ and $\mathbf{Z}$ put mass are the same. Hence, in order to study $\mathbf{\Theta}$, it is possible to study $\mathbf{Z}$, and this can be done through the study of $\pi(\mathbf{X}/t)$ (see Proposition 3). Analyzing the projected vector $\pi(\mathbf{X}/t)$ often highlights a sparse structure and then leads to dimension reduction. In particular, asymptotic independence can be infered with $\mathbf{Z}$, and thus with $\pi(\mathbf{X}/t)$. The future work is thus to build a statistical method based on the projection π to study the angular vector $\mathbf{Z}$. Starting from a sample $\mathbf{X}_1, \dots, \mathbf{X}_n$, this method should identify the directions on which the spectral measure puts mass, that is, the directions where extreme events occur.

6 Proofs

Proof of Lemma 1. Let $\epsilon > 0$, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^d$ such that $|\mathbf{x} - \mathbf{y}| \leq \frac{\epsilon}{d+1}$. We have the following inequalities:

$$|\pi(\mathbf{x}) - \pi(\mathbf{y})| = |(\mathbf{x} - \boldsymbol{\lambda}_\mathbf{x})_+ - (\mathbf{y} - \boldsymbol{\lambda}_\mathbf{y})_+| \leq |(\mathbf{x} - \boldsymbol{\lambda}_\mathbf{x}) - (\mathbf{y} - \boldsymbol{\lambda}_\mathbf{y})| \leq |\mathbf{x} - \mathbf{y}| + d|\lambda_\mathbf{y} - \lambda_\mathbf{x}|, \tag{6.1}$$

where the second inequality is a consequence of the inequality $|\mathbf{a}_+ - \mathbf{b}_+| \leq |\mathbf{a} - \mathbf{b}|$ for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$.

The goal is now to control the quantity $|\lambda_\mathbf{y} - \lambda_\mathbf{x}|$. Without loss of generality, we may assume that $\lambda_\mathbf{x} \leq \lambda_\mathbf{y}$. Then, by definition of $\lambda_\mathbf{x}$ and $\lambda_\mathbf{y}$, we have

$$
\begin{aligned}
|(\mathbf{y}-\boldsymbol{\lambda}_{\mathbf{y}})_+| = 1 = |(\mathbf{x}-\boldsymbol{\lambda}_{\mathbf{x}})_+| &= \sum_{i=1}^{d}(x_i-\lambda_{\mathbf{x}})\mathbb{1}_{\{x_i>\lambda_{\mathbf{x}}\}} \\
&= \sum_{i=1}^{d}(x_i-\lambda_{\mathbf{y}}+\lambda_{\mathbf{y}}-\lambda_{\mathbf{x}})\mathbb{1}_{\{x_i>\lambda_{\mathbf{x}}\}} \\
&= \sum_{i=1}^{d}(x_i-\lambda_{\mathbf{y}})\mathbb{1}_{\{x_i>\lambda_{\mathbf{x}}\}} + \sum_{i=1}^{d}(\lambda_{\mathbf{y}}-\lambda_{\mathbf{x}})\mathbb{1}_{\{x_i>\lambda_{\mathbf{x}}\}}\,,
\end{aligned}
\tag{6.2}
$$

As $|(\mathbf{x}-\boldsymbol{\lambda}_{\mathbf{x}})_+| = 1$, at least one i satisfies $x_i \geq \lambda_{\mathbf{x}}$, so

$$
\sum_{i=1}^{d}(\lambda_{\mathbf{y}}-\lambda_{\mathbf{x}})\mathbb{1}_{\{x_i>\lambda_{\mathbf{x}}\}} \geq \lambda_{\mathbf{y}}-\lambda_{\mathbf{x}}\,.
$$

Besides, $\mathbb{1}_{\{x_i>\lambda_{\mathbf{x}}\}} \geq \mathbb{1}_{\{x_i>\lambda_{\mathbf{y}}\}}$, since we assumed $\lambda_{\mathbf{x}} \leq \lambda_{\mathbf{y}}$. Using Equation (6.2), we then deduce that

$$
|(\mathbf{y}-\boldsymbol{\lambda}_{\mathbf{y}})_+| \geq \sum_{i=1}^{d}(x_i-\lambda_{\mathbf{y}})\mathbb{1}_{\{x_i>\lambda_{\mathbf{y}}\}} + (\lambda_{\mathbf{y}}-\lambda_{\mathbf{x}}) = |(\mathbf{x}-\boldsymbol{\lambda}_{\mathbf{y}})_+| + (\lambda_{\mathbf{y}}-\lambda_{\mathbf{x}})\,,
$$

which means that

$$
(\lambda_{\mathbf{y}}-\lambda_{\mathbf{x}}) \leq |(\mathbf{y}-\lambda_{\mathbf{y}})_+| - |(\mathbf{x}-\lambda_{\mathbf{y}})_+|\,.
$$

We use again the inequality $|\mathbf{a}_+ - \mathbf{b}_+| \leq |\mathbf{a}-\mathbf{b}|$ for $\mathbf{a},\mathbf{b} \in \mathbb{R}^d$, and obtain

$$
(\lambda_{\mathbf{y}}-\lambda_{\mathbf{x}}) \leq |(\mathbf{y}-\lambda_{\mathbf{y}})_+| - |(\mathbf{x}-\lambda_{\mathbf{y}})_+| \leq |(\mathbf{y}-\lambda_{\mathbf{y}})_+ - (\mathbf{x}-\lambda_{\mathbf{y}})_+| \leq |(\mathbf{y}-\lambda_{\mathbf{y}}) - (\mathbf{x}-\lambda_{\mathbf{y}})| = |\mathbf{y}-\mathbf{x}|\,.
$$

Finaly, inequality (6.1) gives

$$
|\pi(\mathbf{x})-\pi(\mathbf{y})| \leq |\mathbf{x}-\mathbf{y}| + d|\lambda_{\mathbf{x}}-\lambda_{\mathbf{y}}| \leq |\mathbf{x}-\mathbf{y}| + d|\mathbf{x}-\mathbf{y}| \leq (d+1)\frac{\epsilon}{d+1} = \epsilon\,,
$$

which proves the continuity of π. □

Proof of Lemma 2. We use the relation $\pi_z(\mathbf{v}) = z\pi(\mathbf{v}/z)$ to simplify the problem:

$$
\begin{aligned}
&\forall\, 0 < z \leq z',\ \forall\, \mathbf{v} \in \mathbb{R}^d_+,\ \pi_z(\pi_{z'}(\mathbf{v})) = \pi_z(\mathbf{v}) \\
\iff &\forall\, 0 < z \leq z',\ \forall\, \mathbf{v} \in \mathbb{R}^d_+,\ z\pi(z^{-1}\pi_{z'}(\mathbf{v})) = z\pi(\mathbf{v}/z) \\
\iff &\forall\, 0 < z \leq z',\ \forall\, \mathbf{v} \in \mathbb{R}^d_+,\ \pi(z'z^{-1}\pi(\mathbf{v}/z')) = \pi(\mathbf{v}/z) \\
\iff &\forall\, a \geq 1,\ \forall\, \mathbf{u} \in \mathbb{R}^d_+,\ \pi(a\pi(\mathbf{u})) = \pi(a\mathbf{u})\,.
\end{aligned}
$$

So we need to prove this last equality. Let $a \geq 1$ and $\mathbf{u} \in \mathbb{R}^d_+$. We divide the proof into three steps. Recall that an expression of ρ is given in (2.7).

STEP 1: We prove that $\rho_{a\mathbf{u}} \leq \rho_{\mathbf{u}}$.

Fix $j \in \{1,\ldots,d\}$ such that $\pi(a\mathbf{u})_j > 0$. This means that

$$
au_j - \frac{1}{j}\left(\sum_{r=1}^{j} au_{(r)} - 1\right) > 0\,.
$$

Thus,

$$
u_j - \frac{1}{j}\sum_{r=1}^{j} u_{(r)} + \frac{1}{ja} > 0\,.
$$

Since $a \geq 1$, we obtain

$$u_j - \frac{1}{j}\sum_{r=1}^{j} u_{(r)} + \frac{1}{j} > 0\,,$$

which means that $\pi(\mathbf{u})_j > 0$. This gives $\rho_{a\mathbf{u}} \leq \rho_{\mathbf{u}}$.

STEP 2: We prove that $\rho_{a\pi(\mathbf{u})} = \rho_{a\mathbf{u}}$.

We recall that the definition of $\pi(\mathbf{u})$ is given by $\pi(\mathbf{u})_k = (u_k - \lambda_{\mathbf{u}})$ for $1 \leq k \leq \rho_{\mathbf{u}}$ and $\pi(\mathbf{u})_k = 0$ for $k > \rho_{\mathbf{u}}$.

- We first prove that $\rho_{a\mathbf{u}} \leq \rho_{a\pi(\mathbf{u})}$. Fix $j \in \{1,\ldots,d\}$ such that $\pi(a\mathbf{u})_j > 0$. Then

$$au_j - \frac{1}{j}\Big(\sum_{r=1}^{j} au_{(r)} - 1\Big) > 0\,.$$

Since $\pi(a\mathbf{u})_j > 0$, we have $j \leq \rho_{a\mathbf{u}}$, and with STEP 1 we obtain $j \leq \rho_{a\mathbf{u}} \leq \rho_{\mathbf{u}}$. So for all $r \leq j \leq \rho_{\mathbf{u}}$, $\pi(\mathbf{u})_r = (u_r - \lambda_{\mathbf{u}})$. Thus,

$$a(\pi(u)_j - \lambda_{\mathbf{u}}) - \frac{1}{j}\Big(\sum_{r=1}^{j} a(\pi(u)_{(r)} - \lambda_{\mathbf{u}}) - 1\Big) > 0\,,$$

which gives

$$a\pi(u)_j - \frac{1}{j}\Big(\sum_{r=1}^{j} a\pi(u)_{(r)} - 1\Big) > 0\,.$$

This means that $\pi(a\pi(\mathbf{u}))_j > 0$. Hence, $\rho_{a\mathbf{u}} \leq \rho_{a\pi(\mathbf{u})}$.

- We now prove that $\rho_{a\pi(\mathbf{u})} \leq \rho_{a\mathbf{u}}$. Fix $j \in \{1,\ldots,d\}$ such that $\pi(a\mathbf{u})_j = 0$. Then

$$au_j - \frac{1}{j}\Big(\sum_{r=1}^{j} au_{(r)} - 1\Big) \leq 0\,.$$

If $j \leq \rho_{\mathbf{u}}$, then for all $r \leq j$, $u_r = \pi(\mathbf{u})_r + \lambda_{\mathbf{u}}$, so that

$$a(\pi(\mathbf{u})_j + \lambda_{\mathbf{u}}) - \frac{1}{j}\Big(\sum_{r=1}^{j} a(\pi(\mathbf{u})_{(r)} + \lambda_{\mathbf{u}}) - 1\Big) \leq 0\,,$$

and finally

$$a\pi(\mathbf{u})_j - \frac{1}{j}\Big(\sum_{r=1}^{j} a\pi(\mathbf{u})_{(r)} - 1\Big) \leq 0\,,$$

which means that $\pi(a\pi(\mathbf{u}))_j = 0$.

If $j > \rho_{\mathbf{u}}$, then $\pi(\mathbf{u})_j = 0$, so $a\pi(\mathbf{u})_j = 0$, and finally $\pi(a\pi(\mathbf{u}))_j = 0$. Hence, $\rho_{a\pi(\mathbf{u})} \leq \rho_{a\mathbf{u}}$.

All in all, we proved that if $j \in \{1,\ldots,d\}$, then $\pi(a\mathbf{u})_j > 0$ if and only if $\pi(a\pi(\mathbf{u}))_j > 0$, which concludes STEP 2.

STEP 3: We prove that $\pi(a\mathbf{u}) = \pi(a\pi(\mathbf{u}))$.

With STEP 2, we know that $\rho := \rho_{a\pi(\mathbf{u})} = \rho_{a\mathbf{u}}$. This proves that for $j > \rho$, $\pi(a\mathbf{u})_j$ and $\pi(a\pi(\mathbf{u}))_j$ are both null. Moreover, by definition of the projection π, if $j \leq \rho$,

$$\pi(a\mathbf{u})_j = au_j - \frac{1}{\rho}\Big(\sum_{r=1}^{\rho} au_{(r)} - 1\Big)\,.$$

Since $\rho \leq \rho_{\mathbf{u}}$ (with STEP 1), we use that for all $r \leq \rho$, $\pi(\mathbf{u})_{(r)} = u_{(r)} - \lambda_{\mathbf{u}}$. Thus, we obtain

$$\pi(a\mathbf{u}) = a(\pi(\mathbf{u})_j - \lambda_{\mathbf{u}}) - \frac{1}{\rho}\Big(\sum_{r=1}^{\rho} a(\pi(\mathbf{u})_{(r)} - \lambda_{\mathbf{u}}) - 1\Big) = au_j - \frac{1}{\rho}\Big(\sum_{r=1}^{\rho} au_{(r)} - 1\Big) = \pi(a\pi(\mathbf{u}))_j\,,$$

which concludes the proof. □

Proof of Lemma 3. Let $\mathbf{v} \in \mathbb{R}_+^d$. We sort $\mathbf{v}$ in $\boldsymbol{\mu}$ such that $\mu_1 \geq \ldots \geq \mu_d$. Firstly, note that if two coordinates of $\mathbf{v}$ are equal, then the corresponding coordinates of $\pi(\mathbf{v})$ are equal too. Thus, they are both null or both positive. So the way these two coordinates are ordered in $\boldsymbol{\mu}$ does not matter.

Let us prove the equivalence (2.8). For $i \in I$, let $j \in \{1, \ldots, d\}$ such that $\mu_j = v_i$, and let $J \subset \{1, \ldots, d\}$ be the subset of such j. By definition of $\rho_{\mathbf{v}}$ (see Remark 1), the projected vector $\pi(\mathbf{v})$ satisfies $\pi(\mathbf{v})_I = 0$ if and only if for all $j \in J$, $j > \rho_{\mathbf{v}}$, which means that for all $j \in J$,

$$\mu_j - \frac{1}{j}\Big(\sum_{k=1}^{j} \mu_k - 1\Big) \leq 0\,. \tag{6.3}$$

Note that $j = \sum_{k=1}^d \mathbb{1}_{v_k \geq v_i}$ and $\sum_{k=1}^j \mu_k = \sum_{k=1}^d v_k \mathbb{1}_{v_k \geq v_i}$, so that condition (6.3) can be rephrased as

$$v_i - \frac{1}{\sum_{k=1}^d \mathbb{1}_{v_k \geq v_i}}\Big(\sum_{k=1}^{d} v_k \mathbb{1}_{v_k \geq v_i} - 1\Big) \leq 0\,.$$

This inequality is equivalent to

$$1 \leq \sum_{k=1}^{d} (v_k - v_i)_+\,,$$

which proves (2.8).

For (2.9), set $r = |I^c| \geq 1$ (note that $I^c = \emptyset$ is not possible). Then, the conditions $\pi(\mathbf{v})_I = 0$ and $\pi(\mathbf{v})_{I^c} > 0$ imply that $\rho_{\mathbf{v}} = r$. Thus, we obtain

$$\forall i \in I^c,\ v_i = \pi(\mathbf{v})_i + \frac{1}{r}\Big(\sum_{j \in I^c} v_j - 1\Big) \quad \text{and} \quad \forall i \in I,\ v_i \leq \frac{1}{r}\Big(\sum_{j \in I^c} v_j - 1\Big)\,.$$

On the one hand, since $\pi(\mathbf{v})_i > 0$ for $i \in I^c$, the first equality is equivalent to

$$\max_{i \in I^c} \sum_{j \in I^c} (v_j - v_i) < 1\,.$$

On the other hand, the second equality is equivalent to

$$\min_{i \in I} \sum_{j \in I^c} (v_j - v_i) \geq 1\,.$$

□

Proof of Proposition 1. Fix $\mathbf{x} \in \mathcal{B}(0,1) \cap \mathbb{R}_+^d$, with $x_j \neq 1/d$ for all $j = 1, \ldots, d$. We use (2.10) to write

$$G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\mathbf{Z} > x) = \mathbb{P}(\pi(Y\boldsymbol{\Theta}) > x) = \int_1^\infty \mathbb{P}(y\boldsymbol{\Theta} - (y-1)/d > \mathbf{x})\, \mathrm{d}(-y^{-\alpha})\,.$$

Set $J_+ = \{j,\, x_j > 1/d\}$ and $J_- = \{j,\, x_j < 1/d\}$. Then, for $j \in J_+$, the condition $y\Theta_j - (y-1)/d > x_j$ becomes $[(\Theta_j - 1/d)/(x_j - 1/d)]_+ > 1/y$. Similarly, for $j \in J_-$, the condition $y\Theta_j - (y-1)/d > x_j$ becomes $[(\Theta_j - 1/d)/(x_j - 1/d)]_+ < 1/y$. So we can rewrite the previous integral as

$$G_{\mathbf{Z}}(\mathbf{x}) = \int_1^\infty \mathbb{P}\left(\left\{\forall j \in J_+,\, y^{-\alpha} < \left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)_+^\alpha\right\} \cap \left\{\forall j \in J_-,\, y^{-\alpha} > \left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)_+^\alpha\right\}\right) \mathrm{d}(-y^{-\alpha})\,.$$

Thus, by the change of variable $u = y^{-\alpha}$, we obtain

$$\begin{aligned} G_{\mathbf{Z}}(\mathbf{x}) &= \int_0^1 \mathbb{P}\left(\left\{\forall j \in J_+,\, u < \left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)_+^\alpha\right\} \cap \left\{\forall j \in J_-,\, u > \left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)_+^\alpha\right\}\right) \mathrm{d}u \\ &= \int_0^1 \mathbb{P}\left(\max_{j\in J_-}\left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)_+^\alpha < u < \min_{j \in J_+}\left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)_+^\alpha\right) \mathrm{d}u \\ &= \mathbb{E}\left[\left(1 \wedge \min_{j\in J_+}\left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)_+^\alpha - \max_{j\in J_-}\left(\frac{\Theta_j - 1/d}{x_j - 1/d}\right)_+^\alpha\right)_+\right]. \end{aligned}$$

□

Proof of Proposition 2. We fix $I \subset \{1, \dots, d\}$ and use Lemma 3. The probability that $\mathbf{Z}_I$ is null is equal to

$$\begin{aligned} \mathbb{P}(\mathbf{Z}_I = 0) &= \mathbb{P}\Big(1 \le \min_{j\in I} \sum_{k=1}^d (Y\Theta_k - Y\Theta_j)_+\Big) \\ &= \mathbb{P}\Big(Y^{-\alpha} \le \min_{j\in I}\Big(\sum_{k=1}^d (\Theta_k - \Theta_j)_+\Big)^\alpha\Big) \\ &= \int_0^1 \mathbb{P}\Big(u \le \min_{j\in I}\Big(\sum_{k=1}^d (\Theta_k - \Theta_j)_+\Big)^\alpha\Big)\,\mathrm{d}u \\ &= \mathbb{E}\left[\min_{j\in I}\Big(\sum_{k=1}^d (\Theta_k - \Theta_j)_+\Big)^\alpha\right], \end{aligned}$$

which proves (3.8).

For Equation (3.9), set $I^c = \{i_1, \dots, i_r\} \subset \{1, \dots, d\}$. We use Lemma 3, so that the probability that $\mathbf{Z}$ is concentrated on the r-dimensional subspace $\mathrm{Vect}(\mathbf{e}_{I^c}) = \mathrm{Vect}(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_r})$ is equal to

$$\begin{aligned} \mathbb{P}(\mathbf{Z}_I = 0, \mathbf{Z}_{I^c} > 0) &= \mathbb{P}\Big(Y^{-1} > \max_{j\in I^c} \sum_{k\in I^c}(\Theta_k - \Theta_j)\,;\, Y^{-1} \le \min_{j\in I}\sum_{k\in I^c}(\Theta_k - \Theta_j)\Big) \\ &= \mathbb{P}\Big(Y^{-\alpha} > \Big(\max_{j\in I^c}\sum_{k\in I^c}(\Theta_k - \Theta_j)_+\Big)^\alpha;\, Y^{-\alpha} \le \min_{j\in I}\Big(\sum_{k\in I^c}(\Theta_k - \Theta_j)_+\Big)^\alpha\Big) \\ &= \int_0^1 \mathbb{P}\Big(\max_{j\in I^c}\Big(\sum_{k\in I^c}(\Theta_k - \Theta_j)_+\Big)^\alpha < u \le \min_{j\in I}\Big(\sum_{k\in I}(\Theta_k - \Theta_j)_+\Big)^\alpha\Big)\,\mathrm{d}u \\ &= \mathbb{E}\left[\Big(\min_{j\in I}\Big(\sum_{k\in I}(\Theta_k - \Theta_j)_+\Big)^\alpha - \max_{j\in I^c}\Big(\sum_{k\in I}(\Theta_k - \Theta_j)_+\Big)^\alpha\Big)_+\right]. \end{aligned}$$

This concludes the proof of the proposition. □

Proof of Proposition 3. For $I \subset \{1,\dots,d\}$, set $B_I = \{\mathbf{x} \in \mathbb{R}^d_+,\, \mathbf{x}_I = 0\}$. We want to show that $\mathbb{P}(\pi(\mathbf{X}/t) \in B_I \mid |\mathbf{X}| > t) \to \mathbb{P}(\mathbf{Z} \in B_I)$.

We know that

$$\mathbb{P}(\mathbf{Z} \in B_I) = \mathbb{P}(\pi(Y\mathbf{\Theta})_I = 0) = \mathbb{P}\Big(\forall i \in I,\, Y^{-1} \le \sum_{k=1}^d (\Theta_k - \Theta_i)_+\Big) = \mathbb{P}((Y,\mathbf{\Theta}) \in D_I)\,,$$

where $D_I = \{(r,\theta) \in (1,\infty) \times \mathbb{S}^{d-1}_+,\, \forall i \in I,\, r^{-1} \le \sum_{k=1}^d (\theta_k - \theta_i)_+\}$. This means that we need to show that $\mathbb{P}((|\mathbf{X}|/t, \mathbf{X}/|\mathbf{X}|) \in D_I \mid |\mathbf{X}| > t) \to \mathbb{P}((Y,\mathbf{\Theta}) \in D_I)$. This convergence holds if $\mathbb{P}((Y,\mathbf{\Theta}) \in \partial D_I) = 0$. Since

$$\partial D_I = \partial\Big\{\bigcap_{i\in I} D_i\Big\} \subset \bigcup_{i \in I} \partial D_i = \bigcup_{i\in I}\Big\{r^{-1} = \sum_{k=1}^d (\theta_k - \theta_i)_+\Big\}\,,$$

we have $\mathbb{P}((Y,\mathbf{\Theta}) \in \partial D_I) \le \sum_{i \in I} \mathbb{P}(Y^{-1} = \sum_{k=1}^d (\Theta_k - \Theta_i)_+) = 0$, since Y is a continuous random variable independent of $\mathbf{\Theta}$. This gives the desired result. □

Proof of Proposition 4. Fix $r \ge 1$ and $A \in \mathcal{B}(\mathbb{S}^{d-1}_+)$. For $t > 0$, the following sequence of equalities holds:

$$\begin{aligned}
\mathbb{P}\left(\pi\left(\frac{\mathbf{X}}{t}\right) \in A,\, \frac{|\mathbf{X}|}{t} > r \,\middle|\, |\mathbf{X}| > t\right) &= \mathbb{P}\left(\pi\left(\frac{\mathbf{X}}{t}\right) \in A,\, \frac{|\mathbf{X}|}{t} > r\right) \frac{1}{\mathbb{P}(|\mathbf{X}| > t)}\\
&= \mathbb{P}\left(\pi\left(\frac{\mathbf{X}}{t}\right) \in A \,\middle|\, \frac{|\mathbf{X}|}{t} > r\right) \frac{\mathbb{P}(|\mathbf{X}| > tr)}{\mathbb{P}(|\mathbf{X}| > t)}\\
&= \mathbb{P}\left(\pi\left(r\frac{\mathbf{X}}{tr}\right) \in A \,\middle|\, \frac{|\mathbf{X}|}{t} > r\right) \mathbb{P}(|\mathbf{X}| > tr \mid |\mathbf{X}| > t)\\
&= \mathbb{P}\left(r\pi_{1/r}\left(\frac{\mathbf{X}}{tr}\right) \in A \,\middle|\, |\mathbf{X}| > tr\right) \mathbb{P}(|\mathbf{X}| > tr \mid |\mathbf{X}| > t)\\
&= \mathbb{P}\left(r\pi_{1/r}\left(\pi\left(\frac{\mathbf{X}}{tr}\right)\right) \in A \,\middle|\, |\mathbf{X}| > tr\right) \mathbb{P}(|\mathbf{X}| > tr \mid |\mathbf{X}| > t)\,,
\end{aligned}$$

where last equality results from Lemma 2. Now, when $t \to \infty$, assumption (4.1) and the continuity of $\pi_{1/r}$ and π give

$$\mathbb{P}(\mathbf{Z} \in A,\, Y > r) = \mathbb{P}\left(r\pi_{1/r}(\mathbf{Z}) \in A\right) \mathbb{P}(Y > r)\,.$$

Finally, we conclude the proof with Lemma 2:

$$\mathbb{P}(\mathbf{Z} \in A \mid Y > r) = \mathbb{P}(\pi(r\mathbf{Z}) \in A)\,.$$

□

Proof of Lemma 4. We first prove that 1 implies 2: assume that $\mathbf{X}$ is regularly varying with index α. Then $|\mathbf{X}|$ is regularly varying with the same index. Denote by $\mathbf{\Theta}$ the spectral vector of $\mathbf{X}$ and consider a random variable Y which follows a Pareto(α) distribution and is independent of $\mathbf{\Theta}$. For $A \in \mathcal{B}(\mathbb{S}^{d-1}_+)$ such that $\mathbb{P}(\mathbf{\Theta} \in \partial A) = 0$, and $\epsilon > 0$, we have

$$\begin{aligned}
\epsilon^{-1} \lim_{t\to\infty} \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon],\, \frac{\mathbf{X}}{|\mathbf{X}|} \in A \,\middle|\, |\mathbf{X}| > t\right) &= \epsilon^{-1}\mathbb{P}(Y \in (1,1+\epsilon], \mathbf{\Theta} \in A)\\
&= \epsilon^{-1}\mathbb{P}(Y \le 1+\epsilon)\mathbb{P}(\mathbf{\Theta} \in A)\\
&= \epsilon^{-1}(1 - (1+\epsilon)^{-\alpha})\mathbb{P}(\mathbf{\Theta} \in A)\,.
\end{aligned}$$

This last quantity converges to $\alpha\mathbb{P}(\mathbf{\Theta} \in A)$ when $\epsilon \to 0$, which proves that $\mathbf{X}$ satisfies (4.3) and (4.4) with $l(\cdot) = \alpha\mathbb{P}(\mathbf{\Theta} \in \cdot)$.

We now prove that 2 implies 1. Fix $\epsilon > 0$, $u > 1$, and $A \in \mathcal{B}(\mathbb{S}_+^{d-1})$ such that $l(\partial A) = 0$. Denote by $l_\epsilon^+(A)$ the limsup in (4.3) when $t \to \infty$, and by $l_\epsilon^-(A)$ the liminf in (4.4) when $t \to \infty$. For $u \geq 1$, we decompose the interval (u, ∞) as follows:

$$(u, \infty) = \bigsqcup_{k=0}^{\infty} \big(u(1+\epsilon)^k, u(1+\epsilon)^{k+1}\big]\,.$$

Then for $t > 0$,

$$\begin{aligned}
\mathbb{P}\left(\frac{|\mathbf{X}|}{t} > u, \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; |\mathbf{X}| > t\right) &= \sum_{k=0}^{\infty} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; |\mathbf{X}| > t\right) \\
&= \sum_{k=0}^{\infty} \frac{\mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A\right)}{\mathbb{P}(|\mathbf{X}| > t)} \\
&= \epsilon \sum_{k=0}^{\infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; \frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right) \frac{\mathbb{P}\left(\frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right)}{\mathbb{P}(|\mathbf{X}| > t)}\,.
\end{aligned}$$

Since $|\mathbf{X}|$ is regularly varying with tail index α, the limit of the right part of the sum can be computed as follows:

$$\frac{\mathbb{P}\left(\frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right)}{\mathbb{P}(|\mathbf{X}| > t)} = \mathbb{P}\Big(|\mathbf{X}| > tu(1+\epsilon)^k \mid |\mathbf{X}| > t\Big) \to \big(u(1+\epsilon)^k\big)^{-\alpha}, \quad t \to \infty\,. \tag{6.4}$$

Besides, we know by (4.3) that

$$\liminf_{t \to \infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; \frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right) = l_\epsilon^-(A)\,. \tag{6.5}$$

We now gather (6.4) and (6.5) and use Fatou's lemma to conclude:

$$\begin{aligned}
&\liminf_{t \to \infty} \mathbb{P}\left(\frac{|\mathbf{X}|}{t} > u, \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; |\mathbf{X}| > t\right) \\
&\quad \geq \epsilon \sum_{k=0}^{\infty} \liminf_{t \to \infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; \frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right) \frac{\mathbb{P}\left(\frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right)}{\mathbb{P}(|\mathbf{X}| > t)} \\
&\quad = \epsilon \sum_{k=0}^{\infty} l_\epsilon^-(A) \big(u(1+\epsilon)^k\big)^{-\alpha} \\
&\quad = u^{-\alpha} l_\epsilon^-(A) \frac{\epsilon}{1 - (1+\epsilon)^{-\alpha}}\,,
\end{aligned}$$

and this last quantity converges to $u^{-\alpha} l(A) \alpha^{-1}$ when $\epsilon \to 0$.

In the same way, we know by (4.4) that

$$\limsup_{t \to \infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \;\middle|\; \frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right) = l_\epsilon^+(A)\,. \tag{6.6}$$

Thus, Equations (6.4) and (6.6) and Fatou's lemma allow to write

$$\begin{aligned}
&\limsup_{t\to\infty} \mathbb{P}\left(\frac{|\mathbf{X}|}{t} > u, \frac{\mathbf{X}}{|\mathbf{X}|} \in A \,\middle|\, |\mathbf{X}| > t\right) \\
&\qquad \le \epsilon \sum_{k=0}^{\infty} \limsup_{t\to\infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \,\middle|\, \frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right) \frac{\mathbb{P}\left(\frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right)}{\mathbb{P}(|\mathbf{X}| > t)} \\
&\qquad = \epsilon \sum_{k=0}^{\infty} l_\epsilon^+(A) \big(u(1+\epsilon)^k\big)^{-\alpha} \\
&\qquad = u^{-\alpha} l_\epsilon^+(A) \frac{\epsilon}{1-(1+\epsilon)^{-\alpha}},
\end{aligned}$$

and this last quantity converges to $u^{-\alpha} l(A) \alpha^{-1}$ when $\epsilon \to 0$.

This proves that

$$\mathbb{P}\left(\frac{|\mathbf{X}|}{t} > u, \frac{\mathbf{X}}{|\mathbf{X}|} \in A \,\middle|\, |\mathbf{X}| > t\right) \to u^{-\alpha} l(A) \alpha^{-1}, \quad t \to \infty,$$

for all $u > 1$ and all $A \in \mathcal{B}(\mathbb{S}_+^{d-1})$ such that $l(\partial A) = 0$. Thus, the random vector $\mathbf{X}$ is regularly varying with tail index α and spectral vector $\mathbf{\Theta}$ defined by $\mathbb{P}(\mathbf{\Theta} \in \cdot) = \alpha^{-1} l(\cdot)$. □

Proof of Theorem 1. The proof is based on Lemma 4. Firstly, note that if (4.1) holds, then $|\mathbf{X}|$ is regularly varying with tail index α. Hence, the main part of the proof is to show that convergences (4.3) and (4.4) hold for all $A = A_{\mathbf{x}}$, $\mathbf{x} \in \mathcal{Z}$, where the $A_{\mathbf{x}}$ are defined in (3.4). We divide our proof into two steps.

Before dealing with these two steps, we make a brief remark which will be of constant use. For $\epsilon > 0$ and $\mathbf{x} > 0$, we have the following equivalence:

$$\pi((1+\epsilon)\mathbf{Z}) > \mathbf{x} \iff \mathbf{Z} > \frac{\mathbf{x} + \epsilon/d}{1+\epsilon}. \tag{6.7}$$

This is a consequence of Equation (2.10) and the fact that $\mathbf{Z}$ belongs to the simplex.

Let us move to the proof. We fix $\mathbf{x} \in \mathcal{Z}$ and $\epsilon > 0$. The first step consists in proving that

$$\epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon], \pi\left(\frac{\mathbf{X}}{t}\right) \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t\right)$$

converges when $t \to \infty$, $\epsilon \to 0$. By (4.1) and assumption (A2), we know that this quantity converges to $\epsilon^{-1}\mathbb{P}(Y \in (1, 1+\epsilon], \mathbf{Z} \in A_{\mathbf{x}})$ when $t \to \infty$. Then, Proposition 4 gives

$$\begin{aligned}
\mathbb{P}(Y \in (1, 1+\epsilon], \mathbf{Z} \in A_{\mathbf{x}}) &= \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}} \mid Y > 1+\epsilon)\mathbb{P}(Y > 1+\epsilon) \\
&= \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) \in A_{\mathbf{x}})(1+\epsilon)^{-\alpha} \\
&= \left[1 - (1+\epsilon)^{-\alpha}\right] \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + \left[\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) \in A_{\mathbf{x}})\right](1+\epsilon)^{-\alpha}.
\end{aligned} \tag{6.8}$$

The first term converges to $\alpha \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}})$ when $\epsilon \to 0$. We use (6.7) to compute the second term:

$$\begin{aligned}
\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) \in A_{\mathbf{x}}) &= \mathbb{P}(\mathbf{Z} > \mathbf{x}) - \mathbb{P}\left(\mathbf{Z} > \frac{\mathbf{x} + \epsilon/d}{1+\epsilon}\right) \\
&= G_{\mathbf{Z}}(\mathbf{x}) - G_{\mathbf{Z}}\left(\mathbf{x} + \frac{\epsilon}{1+\epsilon}(1/d - \mathbf{x})\right).
\end{aligned}$$

Since $\mathbf{x}$ is a differentiability point of $G_{\mathbf{Z}}$, we obtain

$$\epsilon^{-1}\mathbb{P}(Y \in (1, 1+\epsilon],\, \mathbf{Z} \in A_{\mathbf{x}}) = \alpha \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + \frac{1}{1+\epsilon} \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d) + o(1)\,,$$

when $\epsilon \to 0$. This means that

$$\epsilon^{-1}\mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon], \pi\left(\frac{\mathbf{X}}{t}\right) \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t \right)$$

converges to $\alpha\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d)$ when $t \to \infty$, $\epsilon \to 0$.

For the second step, we define

$$(\star) := \epsilon^{-1}\left[\mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t \right) - \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon], \pi\left(\frac{\mathbf{X}}{t}\right) \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t \right) \right],$$

and the goal is to prove that $\lim_{\epsilon\to 0} \limsup_{t\to\infty} (\star) = \lim_{\epsilon \to 0} \liminf_{t\to\infty}(\star) = 0$.

We first deal with the lim sup. Assume that $|\mathbf{X}|/t \in (1, 1+\epsilon]$. Then $(|\mathbf{X}|/t - 1 - \epsilon)/d \le 0$. Thus, if $x_j < X_j/|\mathbf{X}|$, then $x_j + (|\mathbf{X}|/t - 1 - \epsilon)/d < X_j/|\mathbf{X}| < X_j/t$. This implies that $x_j - \epsilon/d < X_j/|\mathbf{X}| - (|\mathbf{X}|/t - 1)/d$. The left member is positive for $\epsilon > 0$ small enough, so we proved that if $x_j < X_j/|\mathbf{X}|$, then $x_j - \epsilon/d < \pi(\mathbf{X}/t)$.

These considerations imply that

$$(\star) \le \epsilon^{-1}\left[\mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon], \pi\left(\frac{\mathbf{X}}{t}\right) \in A_{\mathbf{x}-\epsilon/d} \,\middle|\, |\mathbf{X}| > t \right) - \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon], \pi\left(\frac{\mathbf{X}}{t}\right) \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t \right) \right],$$

and thus

$$\limsup_{t\to\infty} (\star) \le \epsilon^{-1} [\underbrace{\mathbb{P}\left(Y \in (1, 1+\epsilon], \mathbf{Z} \in A_{\mathbf{x}-\epsilon/d}\right)}_{(1)} - \underbrace{\mathbb{P}\left(Y \in (1, 1+\epsilon], \mathbf{Z} \in A_{\mathbf{x}}\right)}_{(2)}] =: f_+(\epsilon)\,.$$

We use Proposition 4 and Equation (6.7) to compute (1) and (2). For (1), we have the following equalities:

$$\begin{aligned}
\mathbb{P}(Y \in (1, 1+\epsilon], \mathbf{Z} \in A_{\mathbf{x}-\epsilon/d}) &= \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}-\epsilon/d}) - \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}-\epsilon/d} \mid Y > 1+\epsilon)\mathbb{P}(Y > 1+\epsilon) \\
&= \mathbb{P}(\mathbf{Z} > \mathbf{x} - \epsilon/d) - \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) > \mathbf{x} - \epsilon/d)(1+\epsilon)^{-\alpha} \\
&= \mathbb{P}(\mathbf{Z} > \mathbf{x} - \epsilon/d) - \mathbb{P}(\mathbf{Z} > \mathbf{x}/(1+\epsilon))(1+\epsilon)^{-\alpha} \\
&= G_{\mathbf{Z}}(\mathbf{x} - \epsilon/d)[1 - (1+\epsilon)^{-\alpha}] + [G_{\mathbf{Z}}(\mathbf{x} - \epsilon/d) - G_{\mathbf{Z}}(\mathbf{x} - \epsilon\mathbf{x}/(1+\epsilon))](1+\epsilon)^{-\alpha}\,.
\end{aligned}$$

The first term is equal to $G(\mathbf{x})\alpha\epsilon + o(\epsilon)$ when $\epsilon \to 0$, whereas the second one is equal to

$$G_{\mathbf{Z}}(\mathbf{x}-\epsilon/d) - G_{\mathbf{Z}}(\mathbf{x}) + G_{\mathbf{Z}}(\mathbf{x}) - G_{\mathbf{Z}}(\mathbf{x} - \epsilon\mathbf{x}/(1+\epsilon)) = \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(-\epsilon/d) - \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(-\epsilon\mathbf{x}/(d(1+\epsilon))) + o(\epsilon), \quad \epsilon \to 0\,.$$

This proves that $\epsilon^{-1}(1)$ converges to $\alpha G(\mathbf{x}) + \mathrm{d}G(\mathbf{x})(\mathbf{x} - 1/d)$ when $\epsilon \to 0$. For (2), we refer to (6.8) in which we proved that $\epsilon^{-1}(2)$ converges to $\alpha G(\mathbf{x}) + \mathrm{d}G(\mathbf{x})(\mathbf{x} - 1/d)$ when $\epsilon \to 0$. All in all we proved that $f_+(\epsilon) \to 0$, when $\epsilon \to 0$.

We similarly proceed for the lim inf. Assume that $|\mathbf{X}|/t \in (1, 1+\epsilon]$. Thus, if $\pi(\mathbf{X}/t)_j > x^j(1+\epsilon)$, then $X_j/t - (|\mathbf{X}|/t - 1)/d > x_j(1+\epsilon)$, and therefore $X_j/t > x_j(1+\epsilon)$. Finally we obtain that $X_j/|\mathbf{X}| > x_j$. So we proved that if $\pi(\mathbf{X}/t)_j > x^j(1+\epsilon)$, then $X_j/|\mathbf{X}| > x_j$. These considerations give the following inequality:

$$(\star) \ge \epsilon^{-1}\left[\mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon), \pi\left(\frac{\mathbf{X}}{t}\right) \in A_{(1+\epsilon)\mathbf{x}} \,\middle|\, |\mathbf{X}| > t \right) - \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1+\epsilon), \pi\left(\frac{\mathbf{X}}{t}\right) \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t \right) \right],$$

and thus

$$\liminf_{t\to\infty}(\star) \geq \epsilon^{-1}[\underbrace{\mathbb{P}\left(Y \in (1,1+\epsilon), \mathbf{Z} \in A_{(1+\epsilon)\mathbf{x}}\right)}_{(1')} - \underbrace{\mathbb{P}\left(Y \in (1,1+\epsilon), \mathbf{Z} \in A_{\mathbf{x}}\right)}_{(2')}] = f_-(\epsilon)\,.$$

We use again Proposition 4 and Equation (6.7) to compute $(1')$:

$$\begin{aligned}\mathbb{P}(Y \in (1,1+\epsilon], \mathbf{Z} \in A_{\mathbf{x}-\epsilon/d}) &= \mathbb{P}(\mathbf{Z} \in A_{(1+\epsilon)\mathbf{x}}) - \mathbb{P}(\mathbf{Z} \in A_{(1+\epsilon)\mathbf{x}} \mid Y > 1+\epsilon)\mathbb{P}(Y > 1+\epsilon)\\ &= \mathbb{P}(\mathbf{Z} > (1+\epsilon)\mathbf{x}) - \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) > (1+\epsilon)\mathbf{x})(1+\epsilon)^{-\alpha}\\ &= \mathbb{P}(\mathbf{Z} > (1+\epsilon)\mathbf{x}) - \mathbb{P}(\mathbf{Z} > \mathbf{x} + \epsilon/((1+\epsilon)d))(1+\epsilon)^{-\alpha}\\ &= G_{\mathbf{Z}}((1+\epsilon)\mathbf{x})[1-(1+\epsilon)^{-\alpha}] + [G_{\mathbf{Z}}((1+\epsilon)\mathbf{x}) - G_{\mathbf{Z}}(\mathbf{x}+\epsilon/(d(1+\epsilon)))](1+\epsilon)^{-\alpha}\\ &= G_{\mathbf{Z}}((1+\epsilon)\mathbf{x})\alpha\epsilon + [\mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\epsilon(\mathbf{x}-1/d)/(1+\epsilon))] + o(\epsilon), \quad \epsilon \to 0\,.\end{aligned}$$

The first term is equal to $G_{\mathbf{Z}}(\mathbf{x})\alpha\epsilon + o(\epsilon)$, when $\epsilon \to 0$, whereas the second one is equal to

$$G_{\mathbf{Z}}((1+\epsilon)\mathbf{x}) - G_{\mathbf{Z}}(\mathbf{x}) + G_{\mathbf{Z}}(\mathbf{x}) - G_{\mathbf{Z}}(\mathbf{x}+\epsilon/(d(1+\epsilon))) = \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\epsilon(\mathbf{x}-1/d)) + o(\epsilon), \quad \epsilon \to 0\,.$$

This proves that $\epsilon^{-1}(1')$ converges to $\alpha G_{\mathbf{Z}}(\mathbf{x}) + \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\mathbf{x}-1/d)$ when $\epsilon \to 0$. Note that $(2') = (2)$, so that $\epsilon^{-1}(2')$ converges to $\alpha G_{\mathbf{Z}}(\mathbf{x}) + \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\mathbf{x}-1/d)$ when $\epsilon \to 0$. All in all we proved that $f_-(\epsilon) \to 0$, when $\epsilon \to 0$.

Gathering all these results together, we can write

$$f_-(\epsilon) \leq \liminf_{t\to\infty}(\star) \leq \limsup_{t\to\infty}(\star) \leq f_+(\epsilon)\,.$$

Since $f_-(\epsilon)$ and $f_+(\epsilon)$ converge to 0 as $\epsilon \to 0$, we proved that $\lim_{\epsilon\to 0}\liminf_{t\to\infty}(\star) = \lim_{\epsilon\to 0}\limsup_{t\to\infty}(\star) = 0$.

To conclude the proof, we write

$$\epsilon^{-1}\mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1,1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t\right) = (\star) + \epsilon^{-1}\mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1,1+\epsilon], \pi\left(\frac{\mathbf{X}}{t}\right) \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t\right),$$

and both steps lead to

$$\lim_{\epsilon\to 0}\liminf_{t\to\infty}\epsilon^{-1}\mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1,1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t\right) = \alpha G_{\mathbf{Z}}(\mathbf{x}) + \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\mathbf{x}-1/d)\,,$$

and

$$\lim_{\epsilon\to 0}\limsup_{t\to\infty}\epsilon^{-1}\mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1,1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t\right) = \alpha G_{\mathbf{Z}}(\mathbf{x}) + \mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\mathbf{x}-1/d)\,.$$

Since $|\mathbf{X}|$ is regularly varying with tail index α, we apply Lemma 4 to conclude that $\mathbf{X}$ is regularly varying with tail index α and with spectral vector $\mathbf{\Theta}$ satisfying $\mathbb{P}(\mathbf{\Theta} \in A_{\mathbf{x}}) = \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + \alpha^{-1}\mathrm{d}G_{\mathbf{Z}}(\mathbf{x})(\mathbf{x}-1/d)$. □

References

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D., and Ferro, C. (2006). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Wiley.

Boldi, M.-O. and Davison, A. (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 69(2):217–229.

Chautru, E. (2015). Dimension reduction in multivariate extreme value analysis. *Electron. J. Statist.*, 9(1):383–418.

Chiapino, M. and Sabourin, A. (2016). Feature clustering for extreme events analysis, with application to extreme stream-flow data. In *New Frontiers in Mining Complex Patterns - 5th International Workshop, NFMCP 2016*, pages 132–147.

Condat, L. (2016). Fast projection onto the simplex and the l1 ball. *Mathematical Programming*, 158(1):575–585.

Cooley, D. and Thibaud, E. (2016). Decompositions of dependence for high-dimensional extremes. *Biometrika*.

Duchi, J. et al. (2008). Efficient projections onto the ℓ^1-ball for learning in high dimensions. *Proceedings of the 25th International Conference on Machine Learning*, pages 272–279.

Einmahl, J., de Haan, L., and Huang, X. (1993). Estimating a multidimensional extreme-value distribution. *Journal of Multivariate Analysis*, 47(1):35 – 47.

Einmahl, J., de Haan, L., and Piterbarg, V. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Tilburg University, Open Access publications from Tilburg University*, 29.

Einmahl, J., de Haan, L., and Sinha, A. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Processes and their Applications*, 70(2):143 – 171.

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

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.

Goix, N., Sabourin, A., and Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*.

Guillotte, S., Perron, F., and Segers, J. (2011). Non-parametric bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 73(3):377–406.

Jessen, A. H. and Mikosch, T. (2006). Regularly varying functions. *Publications de l'Institut Mathématique*, 80(94)(100):171–192.

Kyrillidis, A., Becker, S., Cevher, V., and Koch, C. (2013). Sparse projections onto the simplex. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages II–235–II–243. JMLR.org.

Lehtomaa, J. and Resnick, S. (2019). Asymptotic independence and support detection techniques for heavy-tailed multivariate data. arXiv: 1904.00917.

Liu, J. and Ye, J. (2009). Efficient euclidean projections in linear time. *Proceedings of the 26th Annual International Conference on Machine Learning*.

Resnick, S. (1986). Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18(1):66–138.

Resnick, S. (1987). *Extreme values, regular variation, and point processes*. Springer.

Resnick, S. (2007). *Heavy-Tail Phenomena*. Series in Operations Research and Financial Engineering. Springer, New York. Probabilistic and statistical modeling.

Sabourin, A. and Naveau, P. (2014). Bayesian dirichlet mixture model for multivariate extremes: A reparametrization. *Computational Statistics and Data Analysis*, 71:542 – 567.

Segers, J. (2012). Max-stable models for multivariate extremes. *Revstat Statistical Journal*, 10.