



HAL
open science

Sparse regular variation

Nicolas Meyer, Olivier Wintenberger

► **To cite this version:**

| Nicolas Meyer, Olivier Wintenberger. Sparse regular variation. 2020. hal-02168495v2

HAL Id: hal-02168495

<https://hal.science/hal-02168495v2>

Preprint submitted on 6 Jan 2020 (v2), last revised 18 Feb 2021 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sparse regular variation

Nicolas Meyer^{*1} and Olivier Wintenberger^{†1}

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

January 6, 2020

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 simultaneously occur in all directions. Unfortunately, it is defined through weak convergence which does not provide a natural way to capture this sparsity structure. In this paper, we introduce the notion of sparse regular variation which allows to better learn the dependence 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 and prove that under mild assumptions sparse regular variation and regular variation are two equivalent notions. Finally, we provide numerical evidence of our theoretical findings and compare our method with a recent one developed by [Goix et al. \(2017\)](#).

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 standard framework in multivariate Extreme Value Theory (EVT) is based on the concept of regularly varying random vectors. Regular variation has first been defined in terms of vague convergence on the compactified space $[-\infty, \infty]^d$ and several characterizations have subsequently been established, see e.g. [Embrechts et al. \(2013\)](#), [Resnick \(1987\)](#), [Resnick \(2007\)](#), or [Beirlant et al. \(2006\)](#). [Hult and Lindskog \(2006\)](#) extend the notion of regular variation on a general (possibly infinite dimensional) metric space. They introduce the concept of M_0 -convergence of Borel measures which is based on bounded continuous test functions with support bounded away from the origin.

In this article, we use Resnick's setting and define multivariate regular variation through the convergence of the radial and polar coordinates of a random vector (see [Resnick \(1987\)](#), Proposition 5.17 and Corollary 5.18, or [Resnick \(2007\)](#), Theorem 6.1). 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 \mathbb{S}_+^{d-1} of the unit sphere if

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

^{*}nicolas.meyer@upmc.fr (corresponding author)

[†]olivier.wintenberger@upmc.fr

for all $x > 0$ and for all continuity set B of S . Convergence (1.1) can be interpreted as follows: The limit of the radial component $|\mathbf{X}|/t$ has a Pareto distribution with parameter $\alpha > 0$, while 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} . Note that the choice of the norm in (1.1) is arbitrary. Actually, it is even possible to choose two different norms for the radial and angular parts (see Beirlant et al. (2006), Section 8.2.3).

Based on convergence (1.1), several nonparametric estimation techniques have been proposed to estimate S . In the bivariate case, some useful representations of the spectral measure has been introduced 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 marginals by fitted Pareto tails in order to estimate S by an empirical measure. The latter is consistent and asymptotically normal under suitable assumptions. Einmahl and Segers (2009) focus 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 moment constraints. Inference on the spectral measure has also been studied in a Bayesian framework, for instance by Guillotte et al. (2011). In this paper, the authors use censored likelihood methods in the context of infinite dimensional spectral measures. Parametric approaches have also been introduced to tackle the study of extremes in moderate ($d \leq 10$) dimensions, for instance by Coles and Tawn (1991) and Sabourin et al. (2013).

In higher dimensions, mixtures 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.

More recently, the study of the spectral measure's support has become an active topic of research. Indeed, this support gathers information on the dependence structure of extreme values: The subspaces on which the spectral measure puts mass correspond to 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. Unfortunately, the complete support's estimation is often difficult to capture, so that a main goal in the tail dependence's study is rather to identify clusters of components which are likely to be extreme together. This approach has firstly been introduced by Chautru (2015) who uses a clustering technique to exhibit groups of variables with asymptotic dependence. In the same way, Janßen and Wan (2019) use spherical k -means in order to find clusters with the same extremal behavior.

Recently, two approaches based on Principal Component Analysis (PCA) for high-dimensional extremes have been developed. Cooley and Thibaud (2019) define a vector space on the positive orthant \mathbb{R}_+^d in order to conciliate both PCA and regular variation. They summarize the tail dependence through a matrix of pairwise tail dependence metrics and apply some usual decomposition on this matrix. They illustrate their approach with simulations on Swiss rainfall data and financial return data. In a recent work, Sabourin and Drees (2019) assume that the spectral measure concentrates on \mathbb{S}_+^{p-1} with $p < d$, and assume that the parameter p is known. The aim of their paper is to identify this support with an empirical risk minimization's technique.

Some more algorithmic approaches have also been recently introduced. Goix et al. (2017) consider ϵ -thickened rectangles to estimate the directions on which the spectral measure concentrates. This estimation is based on a tolerance parameter $\epsilon > 0$ and brings out a sparse representation of the dependence structure. It leads to an algorithm called DAMEX (for Detecting Anomalies among Multivariate EXtremes) of complexity $O(dn \log n)$, where n corresponds to the number of data points. Subsequently, Chiapino and Sabourin (2016) propose another algorithm (CLEF for CLustering Extremal Features) to group together subsets that are likely to be simultaneously extreme. A $O(dn \log n)$ complexity has also been reached by Simpson et al. (2019) who base their method on hidden regular variation. They introduce a set of parameters $(\tau_C)_{C \subset \{1, \dots, d\}}$ which describe to what extent the feature C gathers ex-

treme values. Most of 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 is firstly mapped to the space $[0, 1]^{d-1}$ before being partitioned in equally sized rectangles. The estimation of the support is based on a standard estimator of the spectral measure, see [Resnick \(2007\)](#), Section 9.2.2. 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 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 $d - 1$, while these types of subsets often concentrate large events. Indeed, in many situations it is very unlikely that a lot of coordinates are simultaneously extreme. In other words, extreme events occur in few directions $i_1, \dots, i_r \in \{1, \dots, d\}$, with $r \ll d$. In this case, the spectral measure puts mass on $\text{Vect}(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_r}) \cap \mathbb{S}_+^{d-1}$, where $\mathbf{e}_1, \dots, \mathbf{e}_d$ denote the vectors of the canonical basis of \mathbb{R}^d . We say then that the spectral measure is *sparse*. Unfortunately, as soon as $r < d$, the weak convergence [\(1.1\)](#) does not hold for subspaces like $\text{Vect}(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_r}) \cap \mathbb{S}_+^{d-1}$, since they are not continuity 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 $\{\mathbf{x} \in \mathbb{R}_+^d, x_1 + \dots + x_d = 1\}$. 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 by [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 sparsity 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 limit measures.

Outline The structure of this paper is as follows. Section [2](#) gathers all theoretical results useful in this paper. Firstly, we introduce 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 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 a regular variation context. We focus on the angular part of the limit after substituting the usual projected vector $\mathbf{X}/|\mathbf{X}|$ in [\(1.1\)](#) by a vector based on the Euclidean projection onto the simplex. We also provide some interpretations of this new angular vector and discuss to what extent this way of projecting allows us to better capture the sparsity structure of the extremes. The concept of sparsely regularly varying random vector is then introduced in Section [4](#). We establish the equivalence, under mild conditions, between this notion and the standard regular variation's concept. Finally, we illustrate in Section [5](#) the performance of our method on simulated data and compare it with the approach of [Goix et al. \(2017\)](#).

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 $\mathbb{R}_+^d = \{\mathbf{x} \in \mathbb{R}^d, x_1 \geq 0, \dots, x_d \geq 0\}$ and $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^d$. For $j = 1, \dots, d$, \mathbf{e}_j denotes the j -th vector of the canonical basis of \mathbb{R}^d , which 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 $(x_{i_1}, \dots, x_{i_r})$ of \mathbb{R}^r . For $p \in [1, \infty]$, we denote by $\|\cdot\|_p$ the ℓ^p -norm in \mathbb{R}^d . We write \xrightarrow{w} for the weak convergence.

For a set E , we denote by $\mathcal{P}(E)$ its power set: $\mathcal{P}(E) = \{A, A \subset E\}$. We also use the notation $\mathcal{P}^*(E) = \mathcal{P}(E) \setminus \{\emptyset\}$. If $E = \{1, \dots, r\}$, we simply write $\mathcal{P}_r = \mathcal{P}(\{1, \dots, r\})$ and $\mathcal{P}_r^* = \mathcal{P}(\{1, \dots, r\}) \setminus \{\emptyset\}$. For a finite set E , we denote by $\#E$ its cardinality. If $\#E = r \geq 1$, then $\#\mathcal{P}(E) = 2^r$. In particular, $\#\mathcal{P}_r = 2^r$ and $\#\mathcal{P}_r^* = 2^r - 1$. Finally, if F is a subset of a set E , we denote by F^c the complement of F (in E).

2 Theoretical background

2.1 Regular variation and spectral measure

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: There exist a random vector Θ on \mathbb{S}_+^{d-1} and a non-degenerate random variable Y such that the following limit holds:

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \frac{\mathbf{X}}{|\mathbf{X}|}\right) \in \cdot \mid |\mathbf{X}| > t\right) \xrightarrow{w} \mathbb{P}((Y, \Theta) \in \cdot), \quad t \rightarrow \infty. \quad (2.1)$$

In this case, there exists $\alpha > 0$ such that Y follows a Pareto distribution with parameter α . Moreover, the radial limit Y is independent of the angular limit Θ . The random vector Θ is called the *spectral vector* and its distribution $S(\cdot) := \mathbb{P}(\Theta \in \cdot)$ is called the *spectral measure*.

Equation (2.1) brings out the two quantities which characterize the regular variation property of \mathbf{X} . On the one hand, the tail index α highlights the intensity of the extremes: The smaller this index is, the larger the extremes are. On the other hand, the spectral vector Θ informs on their localization and their dependence structure: The spectral measure puts mass in a direction of \mathbb{S}_+^{d-1} if and only if extreme events appear in this direction. Hence, estimating the spectral measure is a crucial (but challenging) problem in multivariate EVT.

In arbitrary dimensions, several authors recently focus on the estimation of the spectral measure's support. The main purpose is to detect features that are likely to be extreme together. In other words, we would like to identify some specific subsets of \mathbb{S}_+^{d-1} on which the spectral measure puts mass. To this end, it is convenient to partition the positive unit sphere \mathbb{S}_+^{d-1} in the following way. For $\beta \in \mathcal{P}_d^*$, we define the subsets

$$C_\beta = \{\mathbf{x} \in \mathbb{S}_+^d, x_i > 0 \text{ for } i \in \beta, x_i = 0 \text{ for } i \notin \beta\}. \quad (2.2)$$

This approach can be related to the one developed by [Goix et al. \(2017\)](#) (see Remark 5 and Section 5). Note that, by construction, the subsets C_β are pairwise disjoint and form a partition of \mathbb{S}_+^{d-1} :

$$\mathbb{S}_+^{d-1} = \bigsqcup_{\beta \in \mathcal{P}_d^*} C_\beta,$$

where \bigsqcup denotes a disjoint union. An illustration of these subsets in dimension 3 are given in Figure 1.

This partition is helpful to study the extremal structure of \mathbf{X} . Indeed, for $\beta \in \mathcal{P}_d^*$, the inequality $\mathbb{P}(\Theta \in C_\beta) > 0$ means that it is likely to observe simultaneously large values in the directions $i \in \beta$ and small values in the directions $i \in \beta^c$. Then, identifying the subsets C_β which concentrate the mass of the spectral measure allows us to bring out clusters of coordinates which can be simultaneously large. Hence,

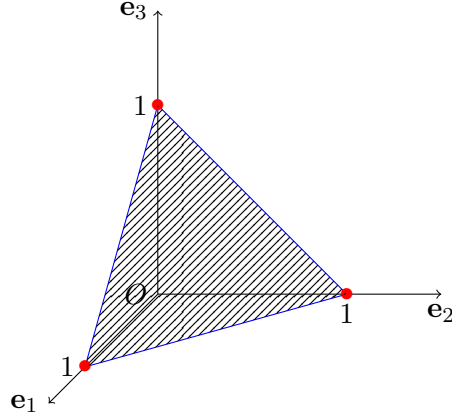


Figure 1: The subsets C_β in dimension 3 for to the ℓ^1 -norm. In red, the subsets $C_{\{1\}}$, $C_{\{2\}}$, and $C_{\{3\}}$. In blue, the subsets $C_{\{1,2\}}$, $C_{\{1,3\}}$, and $C_{\{2,3\}}$. The shaded part corresponds to the interior of the simplex, that is, the subset $C_{\{1,2,3\}}$.

the main first step of the spectral measure's estimation consists in classifying the $2^d - 1$ probabilities $\mathbb{P}(\Theta \in C_\beta)$ depending on their nullity or not. Note that if $\mathbb{P}(\Theta \in C_\beta) > 0$, for $\beta \neq \{1, \dots, d\}$, it means that some coordinates of Θ are equal to zero with positive probability. In this case, we say that the spectral vector (and hence the spectral measure) is sparse.

Remark 1. In EVT, the notion of sparsity can be defined in two different ways. The first one concerns the number of subsets C_β which gather the mass of the spectral measure. "Sparse" means then that this number is much smaller than $2^d - 1$. This is for instance the device of [Goix et al. \(2017\)](#). The second notion deals with the number of 0 in the spectral vector Θ . In this case, "sparse" means that with high probability $|\Theta|_0 \ll d$, where $|\cdot|_0$ denotes the ℓ^0 -norm of Θ , that is, $|\Theta|_0 = \#\{i = 1, \dots, d, \theta_i \neq 0\}$. In all this article, we refer to this second notion. Our aim is to provide a suitable model for extremes which takes this possible sparsity into account.

A standard example of sparsity is the one where the spectral measure only puts mass on the axis: $\mathbb{P}(\Theta \in \sqcup_{1 \leq j \leq d} \{e_j\}) = \mathbb{P}(\Theta \in \sqcup_{1 \leq j \leq d} C_{\{j\}}) = 1$. This means that there is never more than one direction which contributes to the extremal behavior of the data. In this case, we say that the extremes are asymptotically independent. This concept has been studied by many authors, for instance [Ledford and Tawn \(1996\)](#) or [Ramos and Ledford \(2009\)](#).

Even in cases of asymptotically dependence, the mass of the spectral measure often only spreads on low-dimensional subsets C_β , that is, for β such that $\#\beta \ll d$. This is all the more true in high dimensions. Indeed, when d is large, it is very unlikely that all coordinates are extreme together. Regarding the spectral vector, this means that $\mathbb{P}(\Theta \in C_{\{1, \dots, d\}}) = 0$. In such cases, it is then interesting to identify the larger groups of variables $\beta \in \mathcal{P}_d^*$ such that $\mathbb{P}(\Theta \in C_\beta) > 0$. This motivates the notion of maximal subset.

Definition 1 (Maximal subset for Θ). *Let $\beta \in \mathcal{P}_d^*$. We say that a subset C_β is maximal for Θ if*

$$\mathbb{P}(\Theta \in C_\beta) > 0 \quad \text{and} \quad \mathbb{P}(\Theta \in C_{\beta'}) = 0, \quad \text{for all } \beta' \supsetneq \beta. \quad (2.3)$$

In terms of extreme values, the notion of maximality can be rephrased in the following way. Firstly, $\mathbb{P}(\Theta \in C_\beta) > 0$ means that the coordinates of β may be extreme together. Secondly, the condition $\mathbb{P}(\Theta \in C_{\beta'}) = 0$, for all $\beta' \supsetneq \beta$, means that β is not included in a larger group of coordinates β' such that the coordinates of β' may be simultaneously extreme.

Remark 2. A straightforward but useful consequence of Definition 1 is that each subset C_β such that $\mathbb{P}(\Theta \in C_\beta) > 0$ is included in a maximal subset of Θ . Indeed, if there exists no $\beta' \supseteq \beta$, such that $\mathbb{P}(\Theta \in C_{\beta'}) = 0$, then C_β is a maximal subset itself. If not, it means that there exists $\beta' \supseteq \beta$, such that $\mathbb{P}(\Theta \in C_{\beta'}) > 0$. If $C_{\beta'}$ is not maximal, then we repeat this procedure with β' . Since the length of the β 's is finite, the procedure stops and provides $\gamma \in \mathcal{P}_d^*$ such that $\beta \subset \gamma$, $\mathbb{P}(\Theta \in C_\gamma) > 0$ and $\mathbb{P}(\Theta \in C_{\gamma'}) > 0$, for all $\gamma' \supseteq \gamma$.

Why the support's estimation is difficult While the interpretation of the subspaces C_β is rather intuitive, it is quite difficult to estimate the probabilities $\mathbb{P}(\Theta \in C_\beta)$. A natural estimator of the spectral vector Θ is based the second component of convergence (2.1). Indeed, the polar component of \mathbf{X} satisfies

$$\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in \cdot \mid |\mathbf{X}| > t) \xrightarrow{d} \mathbb{P}(\Theta \in \cdot), \quad t \rightarrow \infty. \quad (2.4)$$

This means that the spectral vector Θ can be approximated by the self-normalized extreme $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$, for t large enough. Unfortunately, the supports of Θ and $\mathbf{X}/|\mathbf{X}|$ often drastically differ. Indeed, since \mathbf{X} could model real-world data, the components of \mathbf{X} are almost surely positive. In other words, except for degenerate cases, the random vector $\mathbf{X}/|\mathbf{X}|$ concentrates on the central subspace $C_{\{1, \dots, d\}}$. Equivalently, if $\beta \neq \{1, \dots, d\}$, then $\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in C_\beta) = 0$. This arises while the probability $\mathbb{P}(\Theta \in C_\beta)$ is often positive for some $\beta \neq \{1, \dots, d\}$.

This means that Equation (2.4) is not helpful to study the support of the spectral vector Θ . The self-normalized extreme $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$ does not inform on the behavior of Θ on the C_β 's. This kind of problems arises since the spectral measure may put mass on subspaces included in the boundary of the unit sphere \mathbb{S}_+^{d-1} (in our case the C_β 's, for $\beta \neq \{1, \dots, d\}$), whereas the data generally do 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.

The solution we propose in this article is 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 (2.1).

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

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

In particular, the subsets C_β defined in (2.2) are associated to the ℓ^1 -norm. More generally $\mathbb{S}_+^{d-1}(z) := \{\mathbf{x} \in \mathbb{R}_+^d, x_1 + \dots + x_d = z\}$ for $z > 0$.

2.2 The Euclidean projection onto the simplex

In this 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:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{w} - \mathbf{v}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{w}\|_1 = z. \quad (2.5)$$

Since $\mathbf{v} \geq 0$, the minimization problem (2.5) is equivalent to

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

(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 $\sum_{1 \leq i \leq d} (v_i - \lambda_{\mathbf{v}, z})_+ = z$.

Based on these considerations, we define the application π_z which maps \mathbf{v} to \mathbf{w} :

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

This application is called the *projection onto the positive sphere* $\mathbb{S}_+^{d-1}(z)$. An algorithm which computes $\pi_z(\mathbf{v})$ for $\mathbf{v} \in \mathbb{R}_+^d$ and $z > 0$ is given in [Duchi et al. \(2008\)](#). It is based on a median-search procedure whose expected time complexity is $O(d)$. Unfortunately, this approach is not very intuitive and introduces many variables. Hence, we include it in [Appendix A](#) and detail here a more understandable version of this algorithm with complexity $O(d \log(d))$. [Algorithm 1](#) emphasizes the number of positive coordinates ρ of the projected vector $\pi_z(\mathbf{v})$:

$$\rho = \max \left\{ j \in \{1, \dots, d\}, \mu_j - \frac{1}{j} \left(\sum_{r=1}^j \mu_r - z \right) > 0 \right\}, \quad (2.6)$$

where $\mu_1 \geq \dots, \mu_d$ denote the order coordinates of \mathbf{v} , see [Duchi et al. \(2008\)](#), Lemma 2. The integer ρ corresponds to the ℓ^0 -norm of $\pi_z(\mathbf{v})$ and thus informs on the sparsity of this projected vector. It will therefore be crucial in what follows.

Data: A vector $\mathbf{v} \in \mathbb{R}_+^d$ and a scalar $z > 0$
Result: The projected vector $\mathbf{w} = \pi(\mathbf{v})$
Sort \mathbf{v} in $\boldsymbol{\mu} : \mu_1 \geq \dots \geq \mu_d$;
Find $\rho = \max \left\{ j \in \{1, \dots, n\}, \mu_j - \frac{1}{j} \left(\sum_{r=1}^j \mu_r - z \right) > 0 \right\}$;
Define $\eta = \frac{1}{\rho} \left(\sum_{r=1}^{\rho} \mu_r - z \right)$;
Define \mathbf{w} s.t. $w_i = \max(v_i - \eta, 0)$.

Algorithm 1: Euclidean projection onto the simplex.

Remark 3. The expected linear complexity with respect to the dimension d 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 increases, the study of large events becomes a difficult issue. The recent algorithmic approaches developed by [Simpson et al. \(2019\)](#) or [Goix et al. \(2017\)](#) reach a complexity $O(dn \log(n))$, where n denotes the number of data points. Based on [Algorithm 3](#), we manage to reach a complexity $O(dn)$.

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 $\lambda_{\mathbf{v},1}$:

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

An illustration of π for $d = 2$ is given in [Figure 2](#).

We list below some straightforward results on the projection.

- P1. The projection preserves the order of the coordinates: If $v_{\sigma(1)} \geq \dots \geq v_{\sigma(d)}$ for a permutation σ , then $\pi(\mathbf{v})_{\sigma(1)} \geq \dots \geq \pi(\mathbf{v})_{\sigma(d)}$ for the same permutation.
- P2. If $\pi(\mathbf{v})_j > 0$, then $v_j > 0$. Equivalently, $v_j = 0$ implies $\pi(\mathbf{v})_j = 0$.
- P3. The projection π is continuous, as every projection on a convex, closed set in a Hilbert space.

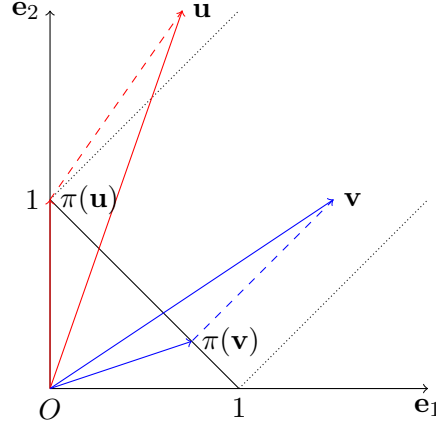


Figure 2: The Euclidean projection onto the simplex \mathbb{S}_+^1 .

The last property will be useful in what follows since π is used to tackle the weak convergence's issue in the spectral measure's definition (2.1). The idea is indeed to substitute the quantity $\mathbf{X}/|\mathbf{X}|$ in (2.1) 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.

We end this section with two important properties satisfied by the projection.

Lemma 1. *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 sparsity structure of extreme events, we are interested in computing probabilities like $\mathbb{P}(\Theta \in C_\beta)$ and $\mathbb{P}(\Theta_{\beta^c} = 0)$, for $\beta \in \mathcal{P}_d^*$. To this end, next lemma will be helpful.

Lemma 2. *Let $\mathbf{v} \in \mathbb{R}_+^d$ and $\beta \in \mathcal{P}_d^*$. The following equivalences hold:*

$$\pi(\mathbf{v})_{\beta^c} = 0 \quad \text{if and only if} \quad 1 \leq \min_{i \in \beta^c} \sum_{k=1}^d (v_k - v_i)_+, \quad (2.7)$$

and

$$\pi(\mathbf{v}) \in C_\beta \quad \text{if and only if} \quad \begin{cases} \max_{i \in \beta} \sum_{j \in \beta} (v_j - v_i) < 1, \\ \min_{i \in \beta^c} \sum_{j \in \beta} (v_j - v_i) \geq 1. \end{cases} \quad (2.8)$$

If $\pi(\mathbf{v}) > 0$ (that is, if $\beta = \{1, \dots, d\}$), then $\pi(\mathbf{v})$ has necessary the following form (see Algorithm 1):

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

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

$$\pi(\mathbf{v}) > \mathbf{x} \quad \text{if and only if} \quad \mathbf{v} > \mathbf{x} + \frac{|\mathbf{v}| - 1}{d}. \quad (2.9)$$

This equivalence will be of constant use in the proofs.

Remark 4. 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 continuous 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 π in a regular variation context. This is the purpose of next section.

3 Spectral measure and projection

The aim of this section is twofold. In the first part, we use the Euclidean projection onto the simplex to introduce a new convergence based on (2.1). 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. Secondly, we establish sparsity results for this new vector. Finally, we develop a model with a discrete spectral vector Θ and study how it affects the vector \mathbf{Z} .

3.1 Regular variation and projection

From now on, and until the end of Section 3, we consider a regularly varying random vector \mathbf{X} on \mathbb{R}_+^d :

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \frac{\mathbf{X}}{|\mathbf{X}|}\right) \in \cdot \mid |\mathbf{X}| > t\right) \xrightarrow{w} \mathbb{P}((Y, \Theta) \in \cdot), \quad t \rightarrow \infty. \quad (3.1)$$

In this case, we know that there exists $\alpha > 0$ such that Y follows a $\text{Pareto}(\alpha)$ distribution and also that the limits Y and Θ are independent. We emphasized in Subsection 2.1 that convergence (3.1) is not helpful to capture the possible sparsity structure of the spectral vector Θ . Our idea is to substitute the self-normalized extremes $\mathbf{X}/|\mathbf{X}|$ by another vector on the simplex which better highlights this sparsity.

Here is an intuitive idea to see how the Euclidean projection can solve this kind of issue. As explained in Section 2.1, for $\beta \in \mathcal{P}_d^*$, the quantity $\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in C_\beta \mid |\mathbf{X}| > t)$ is always equal to 0 (except for degenerate cases), whereas $\mathbb{P}(\Theta \in C_\beta)$ could be positive. This arises since for $t > 0$, the sets $\{\mathbf{x} \in \mathbb{R}_+^d, |\mathbf{x}| > 1, \mathbf{x}/|\mathbf{x}| \in C_\beta\}$ have zero Lebesgue measure for $\beta \neq \{1, \dots, d\}$, and real-world data do not concentrate on such subspaces. Our idea is to replace these subsets by closer ones, but with positive Lebesgue measure. Based on the projection π , we use the subsets $\{\mathbf{x} \in \mathbb{R}_+^d, |\mathbf{x}| > 1, \pi(\mathbf{x}/t) \in C_\beta\}$.

Example 1. Let us take the example of the two-dimensional case illustrated in Figure 2. Here, estimating for instance the probability $\mathbb{P}(\Theta \in C_{\{2\}}) = \mathbb{P}(\Theta_1 = 0)$ with the set of zero Lebesgue measure $\{\mathbf{x}, \mathbf{x}/|\mathbf{x}| \in C_{\{2\}}\}$ seems unachievable. Our idea here is to rather use the set $\{\mathbf{x}, \pi(\mathbf{x}) \in C_{\{2\}}\} = \{\mathbf{x}, x_2 \geq x_1 + 1\}$ which has positive Lebesgue measure. In a sense, the projection allows us to give more weight to the subsets C_β , for $\beta \neq \{1, \dots, d\}$.

Remark 5. The idea of substituting the subspaces $\{\mathbf{x} \in \mathbb{R}_+^d, |\mathbf{x}| > 1, \mathbf{x}/|\mathbf{x}| \in C_\beta\}$ which have zero Lebesgue measure by closer subspaces with positive Lebesgue measure has already been used in the literature. For instance, [Goix et al. \(2017\)](#) define ϵ -thickened rectangles R_β^ϵ , defined by

$$R_\beta^\epsilon = \{\mathbf{x} \in \mathbb{R}_+^d, |\mathbf{x}|_\infty > 1, x_i > \epsilon \text{ for } i \in \beta, x_i \leq \epsilon \text{ for } i \notin \beta\},$$

for $\beta \in \mathcal{P}_d^*$ and $\epsilon > 0$. Unfortunately, these considerations are based on a hyperparameter $\epsilon > 0$ which has to be tuned in practice. One of the advantages of the projection π is that it does not need any hyperparameter. A more detailed comparison of these two methods will be discussed in Section 5.

With this in mind, we substitute the usual projection $\mathbf{X}/|\mathbf{X}|$ by $\pi(\mathbf{X}/t)$. The first step is to see how this affects the spectral vector. The continuity of the projection π implies that

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \pi\left(\frac{\mathbf{X}}{t}\right)\right) \in \cdot \mid |\mathbf{X}| > t\right) \xrightarrow{w} \mathbb{P}((Y, \pi(Y\Theta)) \in \cdot), \quad t \rightarrow \infty. \quad (3.2)$$

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

Following Equation (3.2), we set $\mathbf{Z} = \pi(Y\Theta) \in \mathbb{S}_+^{d-1}$. The aim of this section is to study to what extent the new angular limit \mathbf{Z} differs from the spectral vector Θ and how it helps to study the extremal dependence of \mathbf{X} . A first crucial point is that convergence (3.2) holds for Borel sets $A \in \mathbb{S}_+^{d-1}$ which satisfy $\mathbb{P}(Y\Theta \in \partial\pi^{-1}(A)) = 0$. Next proposition states that the subsets C_β and $\text{Vect}(\mathbf{e}_j, j \in \beta)$, $\beta \in \mathcal{P}_d^*$, satisfy this condition.

Proposition 1. *Let \mathbf{X} be a regularly varying random vector in \mathbb{R}_+^d with spectral vector Θ and tail index $\alpha > 0$. Set $\mathbf{Z} = \pi(Y\Theta)$, where Y is a Pareto(α)-distributed random variable independent of Θ . For any $\beta \in \mathcal{P}_d^*$, the following convergences hold:*

$$\mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta \mid |\mathbf{X}| > t) \rightarrow \mathbb{P}(\mathbf{Z} \in C_\beta), \quad t \rightarrow \infty, \quad (3.3)$$

$$\mathbb{P}(\pi(\mathbf{X}/t)_{\beta^c} = 0 \mid |\mathbf{X}| > t) \rightarrow \mathbb{P}(\mathbf{Z}_{\beta^c} = 0), \quad t \rightarrow \infty. \quad (3.4)$$

Both convergences imply that the sparsity structure of \mathbf{Z} can be studied through the projected vector $\pi(\mathbf{X}/t)$. We insist on the fact that for $\beta \neq \{1, \dots, d\}$, the convergences (3.3) and (3.4) do not hold if we replace \mathbf{Z} by Θ and $\pi(\mathbf{X}/t)$ by $\mathbf{X}/|\mathbf{X}|$. From a statistical point of view, Proposition 1 is helpful since it allows us to estimate the sparse behavior of \mathbf{Z} based on the one of \mathbf{X} . This will be developed in Section 5.

A first interpretation of \mathbf{Z} At a first glance, using the vector \mathbf{Z} instead of Θ in order to capture the tail dependence of \mathbf{X} seems less interpretable. Nevertheless, some properties of the projection make this new vector more understandable, in particular its interpretation regarding \mathbf{X} .

The first property deals with the extremality of a component with respect to the others. For $j = 1, \dots, d$, we apply Equation (2.8) of Lemma 2 with $\beta = \{j\}$ and obtain the following equivalences:

$$\pi(\mathbf{X}/t) \in C_{\{j\}} \iff \min_{i \neq j} (X_j/t - X_i/t) \geq 1 \iff X_j \geq \max_{i \neq j} X_i + t, \quad t > 0.$$

Then, applying Proposition 1 to the subset $C_{\{j\}}$ leads to the convergence

$$\mathbb{P}(X_j \geq \max_{i \neq j} X_i + t \mid |\mathbf{X}| > t) = \mathbb{P}(\pi(\mathbf{X}/t) \in C_{\{j\}} \mid |\mathbf{X}| > t) \rightarrow \mathbb{P}(\mathbf{Z} \in C_{\{j\}}) = \mathbb{P}(Z_j = 1), \quad t \rightarrow \infty.$$

At a non-asymptotic level, this means that for t "high enough", we have the approximation

$$\mathbb{P}(X_j \geq \max_{i \neq j} X_i + t \mid |\mathbf{X}| > t) \approx \mathbb{P}(Z_j = 1).$$

This means that \mathbf{Z} concentrates on the j -th axis if the j -th coordinate of \mathbf{X} is much larger than the others, that is, if extreme values appear in this direction.

More generally, if we fix $\beta \in \mathcal{P}_d^*$ with length $r = \#\beta$, then Equation (2.7) of Lemma 2 leads to the following equivalences:

$$\begin{aligned} \pi(\mathbf{X}/t)_{\beta^c} = 0 &\iff 1 \leq \min_{i \in \beta^c} \sum_{k=1}^d (X_k/t - X_i/t)_+ \iff t \leq \sum_{k=1}^d (X_k - \max_{i \in \beta^c} X_i)_+ \\ &\iff t \leq \sum_{k \in \beta} (X_k - \max_{i \in \beta^c} X_i)_+ \iff t \leq \sum_{k \in \beta} X_k - r \max_{i \in \beta^c} X_i, \end{aligned}$$

where the positive part can be withdrawn since the projection keeps the order of the coordinates (see the property P1, Subsection 2.2). All in all, we obtain the equivalence

$$\pi(\mathbf{X}/t)_{\beta^c} = 0 \iff r^{-1} \sum_{k \in \beta} X_k \geq \max_{i \in \beta^c} X_i + t.$$

Then, following Proposition 1, we obtain

$$\mathbb{P}\left(r^{-1} \sum_{k \in \beta} X_k \geq \max_{i \in \beta^c} X_i + t \mid |\mathbf{X}| > t\right) = \mathbb{P}(\pi(\mathbf{X}/t)_{\beta^c} = 0 \mid |\mathbf{X}| > t) \rightarrow \mathbb{P}(\mathbf{Z}_{\beta^c} = 0), \quad t \rightarrow \infty,$$

and it leads to the following approximation:

$$\mathbb{P}\left(r^{-1} \sum_{k \in \beta} X_k \geq \max_{i \in \beta^c} X_i + t \mid |\mathbf{X}| > t\right) \approx \mathbb{P}(\mathbf{Z}_{\beta^c} = 0),$$

for t "high enough". This can be interpreted in the following way: The vector \mathbf{Z} does not concentrate on the directions $j \in \beta^c$ if the average value of X_j for $j \in \beta$ is much larger than all the components of \mathbf{X} on β^c . In other words, there is an important gap between the average value of the marginals on β and the value of the marginals on β^c .

3.2 The distribution of \mathbf{Z}

The new angular vector \mathbf{Z} being defined, the aim is now to explicit some links between Θ and \mathbf{Z} . To this end, we define the function $G_{\mathbf{Z}}$ by

$$G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\mathbf{Z} > \mathbf{x}) = \mathbb{P}(Z_1 > x_1, \dots, Z_d > x_d), \quad \mathbf{x} \in \mathbb{R}^d. \quad (3.5)$$

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 focus 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 write

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

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

$$A_{\mathbf{x}} = \{\mathbf{u} \in \mathbb{S}_+^{d-1}, x_1 < u_1, \dots, x_d < u_d\}. \quad (3.6)$$

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

Following Equation (2.9), we can express the condition $\mathbf{Z} > \mathbf{x}$ in terms of Θ .

Proposition 2. *Let \mathbf{X} be a regularly varying random vector of \mathbb{R}_+^d with tail index $\alpha > 0$ and spectral vector Θ . For $\mathbf{x} \in \mathcal{B}(0, 1) \cap \mathbb{R}_+^d$, such that for all $j = 1, \dots, d$, $x_j \neq 1/d$, define $J_+ = \{j, x_j > 1/d\}$ and $J_- = \{j, x_j < 1/d\}$. Then, we have*

$$G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{E} \left[\left(1 \wedge \min_{j \in J_+} \left(\frac{\Theta_j - 1/d}{x_j - 1/d} \right)_+^\alpha - \max_{j \in J_-} \left(\frac{\Theta_j - 1/d}{x_j - 1/d} \right)_+^\alpha \right) \right], \quad (3.7)$$

with $G_{\mathbf{Z}}$ defined in (3.5).

Proposition 2 gives an interesting relation between the distribution of \mathbf{Z} and the one of Θ . Unfortunately, its complexity makes it difficult to use. But specific choices for \mathbf{x} will give some useful results.

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

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

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

$$G_{\mathbf{Z}}(\mathbf{0}) = 1 - \mathbb{E} \left[\max_{1 \leq j \leq d} (1 - d\Theta_j)^\alpha \right]. \quad (3.8)$$

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

$$\mathbb{P}(\exists j = 1, \dots, d, Z_j = 0) = \mathbb{E} \left[\max_{1 \leq j \leq d} (1 - d\Theta_j)^\alpha \right]. \quad (3.9)$$

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 $\min_{1 \leq j \leq d} \Theta_j = 0$ a.s. As expected, the new angular vector \mathbf{Z} is more likely to be sparse. In particular, all usual spectral models on Θ 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 the sparsity structure of \mathbf{Z} .

3.3 Sparsity structure of \mathbf{Z}

Since the projection is introduced in order to better capture the sparsity structure of the extremes, we give here different results of sparsity for the angular component $\mathbf{Z} = \pi(Y\Theta)$. The general aim is thus to compute probabilities like $\mathbb{P}(\mathbf{Z} \in C_\beta)$ or $\mathbb{P}(\mathbf{Z}_{\beta^c} = 0)$, for $\beta \in \mathcal{P}_d^*$, in order to generalize Equation (3.9).

Proposition 3. *Let \mathbf{X} be a regularly varying random vector of \mathbb{R}_+^d with spectral vector Θ and tail index $\alpha > 0$. Set $\mathbf{Z} = \pi(Y\Theta)$, where Y is a Pareto(α)-distributed random variable independent of Θ . For any $\beta \in \mathcal{P}_d^*$, we have*

$$\mathbb{P}(\mathbf{Z}_{\beta^c} = 0) = \mathbb{E} \left[\min_{j \in \beta^c} \left(\sum_{k=1}^d (\Theta_k - \Theta_j)_+ \right)^\alpha \right], \quad (3.10)$$

and

$$\mathbb{P}(\mathbf{Z} \in C_\beta) = \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \right]. \quad (3.11)$$

If we consider the case where $\beta = \{1, \dots, d\}$, then we obtain the probability that all coordinates are positive. This has already been computed in (3.8). It is equal to $G_{\mathbf{Z}}(\mathbf{0}) = 1 - \mathbb{E} [\max_{1 \leq j \leq d} (1 - d\Theta_j)^\alpha]$.

Another particular case of Proposition 3 is the one where β corresponds to a single coordinate j_0 . In this case, since \mathbf{Z} belongs to the simplex, both probabilities $\mathbb{P}(\mathbf{Z}_{\beta^c} = 0)$ and $\mathbb{P}(\mathbf{Z} \in C_\beta)$ are equal. Their common value corresponds to the probability that \mathbf{Z} concentrates on the j_0 -th axis, which is equal to

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

Then, Equation (3.12) can be developed in the following way:

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

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

Remark 6. Following Equation (3.10), we write

$$\mathbb{P}(\mathbf{Z}_{\beta^c} = 0) \geq \mathbb{E} \left[\min_{j \in \beta^c} \left(\sum_{k=1}^d (\Theta_k - \Theta_j)_+ \right)^\alpha \mathbf{1}_{\{\Theta_{\beta^c} = 0\}} \right] = \mathbb{E} \left[\left(\sum_{k=1}^d \Theta_k \right)^\alpha \mathbf{1}_{\{\Theta_{\beta^c} = 0\}} \right] = \mathbb{P}(\Theta_{\beta^c} = 0). \quad (3.13)$$

This can also be seen as a direct consequence of Property P2, see Subsection 2.2. This property also gives

$$\mathbb{P}(\mathbf{Z}_\beta > 0) \leq \mathbb{P}(\Theta_\beta > 0). \quad (3.14)$$

This inequality will be useful in some proofs.

Our goal is now to compare the probabilities $\mathbb{P}(\Theta \in C_\beta)$ and $\mathbb{P}(\mathbf{Z} \in C_\beta)$, for $\beta \in \mathcal{P}_d^*$. Based on Proposition 3, we state a first inequality between these two quantities.

Corollary 1. *We use the same notations as in Proposition 3. For $\beta \in \mathcal{P}_d^*$, if $\mathbb{P}(\Theta \in C_\beta) > 0$, then $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$.*

Corollary 1 implies that we do not lose any information on the support of the spectral measure by studying \mathbf{Z} instead of Θ . But it is possible that the distribution of \mathbf{Z} puts some mass on a subset C_β while the one of Θ does not. Nevertheless, if the overestimation is not too large, \mathbf{Z} gives a reduce numbers of directions (regarding the total number $2^d - 1$) in which extreme events could appear. So the use of \mathbf{Z} provides some trends in the dependence structure of \mathbf{X} .

Example 2. We detail here an example which shows that the converse implication of Corollary 1 does not hold. We consider a spectral vector Θ in \mathbb{S}_+^1 with a first component Θ_1 uniformly distributed (and hence $\Theta_2 = 1 - \Theta_1$ is also uniformly distributed). On the one hand, the probability that Θ belongs to an axis is equal to

$$\mathbb{P}(\Theta \in C_{\{1\}} = 0) = \mathbb{P}(\Theta_2 = 0) = 0.$$

On the other hand, following Lemma 2, the probability that \mathbf{Z} belongs to an axis is equal to

$$\mathbb{P}(\mathbf{Z} \in C_{\{1\}}) = \mathbb{P}(Y\Theta_1 - Y\Theta_2 \geq 1) = \mathbb{P}(2\Theta_1 - 1 \geq 1/Y).$$

If we assume that $\alpha = 1$ in order to simplify the calculations, then $1/Y$ is uniformly distributed, and thus, by independence of Θ and Y , we obtain

$$\mathbb{P}(\mathbf{Z} \in C_{\{1\}}) = \int_0^1 \mathbb{P}(2\Theta_1 - 1 \geq u) du = \int_0^1 \mathbb{P}\left(\Theta_1 \geq \frac{u+1}{2}\right) du = \int_0^1 \frac{1-u}{2} du = \frac{1}{4}.$$

Example 2 shows that it is possible to find some β such that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$ and $\mathbb{P}(\Theta \in C_\beta) = 0$. In order to have a partial converse result, and similarly to Definition 2, we introduce the notion of maximal subset for \mathbf{Z} .

Definition 2 (Maximal subset for \mathbf{Z}). *Let $\beta \in \mathcal{P}_d^*$. We say that a subset C_β is maximal for \mathbf{Z} if*

$$\mathbb{P}(\mathbf{Z} \in C_\beta) > 0 \quad \text{and} \quad \mathbb{P}(\mathbf{Z} \in C_{\beta'}) = 0, \quad \text{for all } \beta' \supsetneq \beta. \quad (3.15)$$

Next Theorem states that maximal subsets for Θ and \mathbf{Z} are equivalent notions.

Theorem 1. *We use the same notations as in Proposition 3 and fix $\beta \in \mathcal{P}_d^*$. Then, C_β is a maximal subset for Θ if and only if C_β is a maximal subset for \mathbf{Z} .*

Example 2 shows it may exists $\beta \in \mathcal{P}_d^*$ such that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$ and $\mathbb{P}(\Theta \in C_\beta) = 0$. In this case, Theorem 1 states that the subset C_β is not maximal for \mathbf{Z} since it is not maximal for Θ . Following Remark 2, we consider a maximal subset γ for \mathbf{Z} such that $\beta \subset \gamma$. Then, Theorem 1 states that $\mathbb{P}(\Theta \in C_\gamma) > 0$. This means that even β does not gather itself coordinates on which extreme values simultaneously occur, there exists a superset of β which actually contains extremes. Thus, β still gives information on the study of large events.

3.4 A discrete model for the spectral measure

We introduce here a known discrete model on Θ and compute the corresponding distribution of \mathbf{Z} .

Asymptotic independence and complete dependence We first study two particular cases in multivariate EVT. The first one is the complete dependence's case, which is defined by the relation $\mathbb{P}(\forall i = 1, \dots, d, \Theta_i = 1/d) = 1$. Equivalently, the spectral measure is a Dirac mass at $(1/d, \dots, 1/d)$. In terms of extremes, it means that all coordinates simultaneously contribute to large events. Note that if $\mathbf{u} = r(1/d, \dots, 1/d) \in \mathbb{R}_+^d$, $r \geq 1$, the projected vector $\pi(\mathbf{u})$ corresponds to the self-normalization: $\pi(\mathbf{u}) = (1/d, \dots, 1/d) = \mathbf{u}/|\mathbf{u}|$. This implies that in case of complete dependence, $\mathbf{Z} = \Theta = (1/d, \dots, 1/d)$ a.s.

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

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

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

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

This last probability can be rewritten as follows:

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

This proves the equivalence between $\mathbb{P}(\exists 1 \leq i \leq d, Z_i = 1) = 1$ and $\mathbb{P}(\exists 1 \leq i \leq d, \Theta_i = 1) = 1$. Based on this result and Proposition 1, 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, these two standard 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 Θ in the asymptotically independent and completely dependent settings.

A discrete model We now extend the previous examples to a general discrete model. If $\beta \in \mathcal{P}_d^*$, we denote by $\mathbf{e}(\beta)$ the vector with 1 in position i if $i \in \beta$ and 0 otherwise. Note that for all $\beta \in \mathcal{P}_d^*$, the vector $\mathbf{e}(\beta)/\#\beta$ belongs to the simplex \mathbb{S}_+^{d-1} .

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

$$\sum_{\beta \in \mathcal{P}_d^*} p(\beta) \delta_{\mathbf{e}(\beta)/\#\beta}, \quad (3.16)$$

where $(p(\beta))_\beta$ is a $2^d - 1$ vector with nonnegative components summing to 1. This is the device developed in Segers (2012). Note that this class of distributions includes the previous cases, with respectively $p(\{1, \dots, d\}) = 1$ for complete dependence, and $p(\{j\}) = 1/d$, for all $j = 1, \dots, d$, for asymptotic independence.

The family of distributions (3.16) is stable after multiplying by a positive random variable and projecting onto the simplex with π . Hence, if Θ has a distribution of type (3.16), then $\mathbf{Z} = \Theta$ a.s. This shows that (3.16) forms an accurate model for the angular vector \mathbf{Z} . Indeed, it is stable for the transformation $\Theta \mapsto \mathbf{Z}$. Besides, the distributions of this class have sparse supports. Finally, they put mass on some particular points of the simplex on which extremes values often concentrate in practice.

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 us to study the properties of $\mathbf{Z} = \pi(Y\Theta)$. 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 3 (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 non-degenerate random variable Y such that*

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \pi\left(\frac{\mathbf{X}}{t}\right)\right) \in \cdot \mid |\mathbf{X}| > t\right) \xrightarrow{w} \mathbb{P}((Y, \mathbf{Z}) \in \cdot), \quad t \rightarrow \infty. \quad (4.1)$$

In this case, the general theory of regular variation states that there exists $\alpha > 0$ such that Y is Pareto(α)-distributed. With this definition in mind, we can rephrase the ideas of the beginning of Section 3, and particularly Equation (3.2), in the following way: Regular variation with limit (Y, Θ) implies sparse regular variation with limit $(Y, \pi(Y\Theta))$.

From now on, we consider a sparsely regularly varying random vector \mathbf{X} . Recall that we defined the function $G_{\mathbf{Z}}$ by $G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\mathbf{Z} > \mathbf{x})$ for $\mathbf{x} \in \mathcal{B}(0, 1) \cap \mathbb{R}_+^d$. However, note that for the moment we can not write $G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\pi(Y\Theta) > \mathbf{x})$ since the existence of Θ is not guaranteed. Our aim then 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.

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

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

As already mentioned in Subsection 3.1, 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 Theorem 2.

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 we need to focus on the convergence of the angular component, that is, of the self-normalized extreme $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$ when $t \rightarrow \infty$. This idea is thus to provide a result which characterizes regular variation for a vector \mathbf{X} when $|\mathbf{X}|$ is already regularly varying. This is the purpose of next lemma.

Lemma 3. *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

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

and

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

for all continuity set $A \in \mathcal{B}(\mathbb{S}_+^{d-1})$ of l .

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

Remark 7. The assertion 2 of Lemma 3 can be weakened by taking A in a family of Borel sets 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 (3.6).

Remark 8. In Lemma 3, $|\cdot|$ 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. We denote by λ the Lebesgue measure on the positive unit sphere $\mathcal{B}(0, 1) \cap \mathbb{R}_+^d$. The assumptions on $G_{\mathbf{Z}}$ are the following ones:

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

For instance, the distributions of the model (3.16) satisfy both assumptions.

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 Borel sets 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 2. Let \mathbf{X} be a sparsely regularly varying random vector on \mathbb{R}_+^d . Assume that $G_{\mathbf{Z}}(\cdot) = \mathbb{P}(\mathbf{Z} > \cdot)$ satisfies (A1) and (A2). Then \mathbf{X} is regularly varying with spectral vector Θ which satisfies

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

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 Θ . Equation (4.5) completes the result (3.7) obtained in Proposition 2.

Let us summarize the results we obtained. Proposition 2 characterizes the distribution of $\mathbf{Z} = \pi(Y\Theta)$ when \mathbf{X} is regularly varying with spectral vector Θ . Conversely, suppose that \mathbf{X} is a sparsely regularly varying random vector. Then Theorem 2 states that \mathbf{X} is regularly varying with a spectral vector Θ which satisfies Equation (4.5). This ensures that $\mathbf{Z} = \pi(Y\Theta)$, where Y is a Pareto(α)-distributed random variable independent of Θ . In other words, we have an almost complete equivalence between the usual regular variation's concept and sparse regular variation.

5 Numerical results

This section is devoted to a statistical illustration of sparse regular variation. The idea is to highlight how our approach manages to detect the tail dependence's sparsity. In the first subsection, we provide a method in order to approximate the probabilities $\mathbb{P}(\mathbf{Z} \in C_{\beta})$, $\beta \in \mathcal{P}_d^*$, and we introduce the approach developed by Goix et al. (2017). The second subsection is dedicated to numerical results.

5.1 The framework

We consider an iid sequence of regularly varying random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ with generic distribution \mathbf{X} , and with tail index α and spectral vector $\Theta \in \mathbb{S}_+^{d-1}$. We set $\mathbf{Z} = \pi(Y\Theta)$, where Y follows a Pareto(α) distribution independent of Θ . Our aim is to capture the features $\beta \in \mathcal{P}_d^*$ in which the extreme values of \mathbf{X} occur. Recall that these directions are characterized by the fact that $\mathbb{P}(\Theta \in C_\beta) > 0$, and therefore, by Corollary 1, $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$. Thanks to Proposition 1, the latter probability is defined through the limit

$$\mathbb{P}(\mathbf{Z} \in C_\beta) = \lim_{t \rightarrow \infty} \mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta \mid |\mathbf{X}| > t) = \lim_{t \rightarrow \infty} \frac{\mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta, |\mathbf{X}| > t)}{\mathbb{P}(|\mathbf{X}| > t)}. \quad (5.1)$$

The goal is then to approximate this probability with the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$. For $t > 0$, and $\beta \in \mathcal{P}_d^*$, we define the quantity

$$\hat{T}_n(t, \beta) = \frac{\sum_{j=1}^n \mathbb{1}\{\pi(\mathbf{X}_j/t) \in C_\beta, |\mathbf{X}_j| > t\}}{\sum_{j=1}^n \mathbb{1}\{|\mathbf{X}_j| > t\}}, \quad (5.2)$$

which corresponds to the proportion of data \mathbf{X}_j whose projected vector $\pi(\mathbf{X}_j/t)$ belongs to C_β among the data whose ℓ^1 -norm is above t . Intuitively, the larger the variable $\hat{T}_n(t, \beta)$, the more likely the feature β concentrates extreme values.

The Law of Large Numbers ensures that

$$\lim_{n \rightarrow \infty} \hat{T}_n(t, \beta) = \frac{\mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta, |\mathbf{X}| > t)}{\mathbb{P}(|\mathbf{X}| > t)}, \quad \text{a.s.} \quad (5.3)$$

Hence, Equations (5.3) and (5.1) lead to the following approximation:

$$\hat{T}_n(t, \beta) \approx \frac{\mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta, |\mathbf{X}| > t)}{\mathbb{P}(|\mathbf{X}| > t)} \approx \mathbb{P}(\mathbf{Z} \in C_\beta), \quad (5.4)$$

where the first approximation holds for n large, and the second one for t large. With this approximation, we can classify the subsets C_β depending on the nullity or not of the associated quantity $\hat{T}_n(t, \beta)$.

The approach proposed by Goix et al. (2017) In order to detect anomalies among multivariate extremes, Goix et al. (2017) propose a similar approach by using the ℓ^∞ -norm. They define the ϵ -thickened rectangles by

$$R_\beta^\epsilon := \{\mathbf{x} \in \mathbb{R}_+^d, |\mathbf{x}|_\infty > 1, x_j > \epsilon \text{ for all } j \in \beta, x_j \leq \epsilon \text{ for all } j \in \beta^c\},$$

for $\beta \in \mathcal{P}_d^*$ (see Remark 5). In order to go back to the ℓ^∞ positive unit sphere $\mathbb{S}_{+, \infty}^{d-1}$, we define

$$C_{\beta, \infty}^\epsilon := \{\mathbf{x} / |\mathbf{x}|_\infty, \mathbf{x} \in R_\beta^\epsilon\}, \quad \text{and} \quad C_{\beta, \infty} := \{\mathbf{x} \in \mathbb{S}_{+, \infty}^{d-1}, x_i > 0 \text{ for } i \in \beta, x_i = 0 \text{ for } i \notin \beta\}.$$

Denoting by Θ_∞ the spectral vector with respect to the ℓ^∞ -norm, the convergence

$$\mathbb{P}(\Theta_\infty \in C_{\beta, \infty}^\epsilon) \rightarrow \mathbb{P}(\Theta_\infty \in C_{\beta, \infty}), \quad \epsilon \rightarrow 0,$$

proved in Goix et al. (2017) ensures that $\mathbb{P}(\Theta_\infty \in C_{\beta, \infty}^\epsilon)$ approximates the quantity $\mathbb{P}(\Theta_\infty \in C_{\beta, \infty})$.

Similarly to Equation (5.2), we define the quantity

$$\hat{T}_n^\epsilon(t, \beta) = \frac{\sum_{j=1}^n \mathbb{1}\{\mathbf{X}_j / |\mathbf{X}_j|_\infty \in C_{\beta, \infty}^\epsilon, |\mathbf{X}_j|_\infty > t\}}{\sum_{j=1}^n \mathbb{1}\{|\mathbf{X}_j|_\infty > t\}}, \quad (5.5)$$

and the Law of Large Numbers ensures that, for $\epsilon > 0$ and $t > 0$ fixed,

$$\lim_{n \rightarrow \infty} \hat{T}_n^\epsilon(t, \beta) = \frac{\mathbb{P}(\mathbf{X} / |\mathbf{X}|_\infty \in C_{\beta, \infty}^\epsilon, |\mathbf{X}|_\infty > t)}{\mathbb{P}(|\mathbf{X}|_\infty > t)} = \mathbb{P}(\mathbf{X} / |\mathbf{X}|_\infty \in C_{\beta, \infty}^\epsilon \mid |\mathbf{X}|_\infty > t), \quad \text{a.s.} \quad (5.6)$$

Hence, the estimation of $\mathbb{P}(\Theta_\infty \in C_\beta)$ is based on the following sequence of approximations:

$$\hat{T}_n^\epsilon(t, \beta) \approx \frac{\mathbb{P}(\mathbf{X}/|\mathbf{X}|_\infty \in C_{\beta, \infty}^\epsilon, |\mathbf{X}|_\infty > t)}{\mathbb{P}(|\mathbf{X}|_\infty > t)} \approx \mathbb{P}(\Theta_\infty \in C_{\beta, \infty}^\epsilon) \approx \mathbb{P}(\Theta_\infty \in C_{\beta, \infty}), \quad (5.7)$$

where the first approximation holds for n , the second one for t large, and the last one for ϵ close to zero. All these considerations lead to an algorithm, called DAMEX, introduced in [Goix et al. \(2017\)](#), Section 4.2.

Remark 9 (On the choice of the norm). We already mentioned that the spectral vector can be defined for any norm in \mathbb{R}^d . The choice of the ℓ^1 -norm in this article is deeply linked to the use of the projection π . On the other hand, [Goix et al. \(2017\)](#) choose the ℓ^∞ -norm. After some calculations, we observe that both spectral vectors Θ and Θ_∞ satisfy the relation

$$\mathbb{P}(\Theta \in B) = \frac{\mathbb{E}[|\Theta_\infty|^\alpha \mathbf{1}_{\{\Theta_\infty/|\Theta_\infty| \in B\}}]}{\mathbb{E}[|\Theta_\infty|^\alpha]},$$

for all $B \in \mathbb{S}_+^{d-1}$. In particular,

$$\mathbb{P}(\Theta \in C_\beta) = \frac{\mathbb{E}[|\Theta_\infty|^\alpha \mathbf{1}_{\{\Theta_\infty/|\Theta_\infty| \in C_\beta\}}]}{\mathbb{E}[|\Theta_\infty|^\alpha]}.$$

Since $\Theta_\infty/|\Theta_\infty| \in C_\beta$ if and only if $\Theta_\infty \in C_{\beta, \infty}$, we obtain the equivalence

$$\mathbb{P}(\Theta \in C_\beta) > 0 \quad \text{if and only if} \quad \mathbb{P}(\Theta_\infty \in C_{\beta, \infty}) > 0.$$

In other words, the choice of the norm has no impact on the directions in which extremes gather.

Remark 10. At the end of our procedure, we obtain groups of directions β such that $\hat{T}_n(t, \beta) > 0$. Since we deal with non-asymptotic data, many $\hat{T}_n(t, \beta)$ have small values while the theoretical quantities $\mathbb{P}(\mathbf{Z} \in C_\beta)$ are null. We follow the idea of [Goix et al. \(2017\)](#), Remark 4, to deal with this issue. We define a threshold value under which the empirical quantities $\hat{T}_n(t, \beta)$ are set to 0. We use a threshold of the form $p/\#B$, where $B = \{\beta, \hat{T}_n(t, \beta) > 0\}$ and where the hyperparameter $p \geq 0$ is fixed by the user. Of course, it is possible to set p to 0 and then we select all subsets C_β such that $\hat{T}_n(t, \beta) > 0$. In this case, the number of selected C_β is still much smaller than the total number $2^d - 1$. We do not detail more the choice of p and defer this issue to future work.

Taking this hyperparameter p into account, we are now able to introduce the algorithm used to study the dependence structure of sparsely regularly varying random vectors.

Data: $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}_+^d$, $t > 0$, $p \geq 0$

Result: A list \mathcal{C} of subsets C_β

Compute $\pi(\mathbf{X}_j/t)$, $j = 1, \dots, n$;

Assign to each $\pi(\mathbf{X}_j/t)$ the subsets C_β it belongs to;

Compute $\hat{T}_n(t, \beta)$;

Set to 0 the $\hat{T}_n(t, \beta)$ below the threshold discussed in [Remark 10](#);

Define $\mathcal{C} = \{C_\beta, \hat{T}_n(t, \beta) > 0\}$.

Algorithm 2: Extremal dependence structure of sparsely regularly varying random vectors

5.2 Experimental results

We consider two different cases for the numerical results. For each case, we generate datasets of size $n \in \{10^4, 5 \cdot 10^4, 10^5\}$ and we compute the quantities $\hat{T}_n(t, \beta)$ and $\hat{T}_n^\epsilon(t, \beta)$, for $\beta \in \mathcal{P}_d^*$. We repeat this

procedure over $N = 100$ simulations. Note that there are two different types of errors which could arise. The first one corresponds to the occurrence of a feature β while it should not appear theoretically. The second one corresponds to the absence of a feature β while it should appear theoretically. The results correspond then to the average number of errors among the $N = 100$ simulations. The code can be found at https://github.com/meyernicolas/projection_extremes.

The purpose of these experiments is twofold. The first idea is to study Algorithm 2 and to see if it manages to detect the sparsity structure of the extremal data. The second goal of these simulations is to highlight some evidence in favor of our method compared to the DAMEX algorithm, which is based on a hyperparameter ϵ . The results will show that there exists no natural choice for this hyperparameter. In other words, it may happen that for a fixed simulation study, there exists a specific hyperparameter ϵ_0 for which the DAMEX algorithm leads to better results than our approach. But as soon as we use different simulated data, this specific ϵ_0 is no longer appropriate.

Remark 11 (Choice of the parameters). It is common in EVT to define a level of exceedances $k = n\mathbb{P}(|\mathbf{X}| > t)$ and to rather work with k instead of t . For our simulations, we choose $k = \sqrt{n}$, following Goix et al. (2017), who also suggest choosing ϵ of order $k^{-1/4}$, that is, of order $n^{-1/8}$. This choice of ϵ is based on theoretical results (Goix et al. (2017), Theorem 1), but the authors then advise to rather choose $\epsilon = 0.01$, which gives better results on their simulations. In order to have a very large scale of comparison, we use different $\epsilon \in \{0.01, 0.1, 0.5, 1, 5, 10\}$. Finally, we consider $p = 0.3$ which is larger than the value chosen in Goix et al. (2017) but leads to better results for both methods.

Asymptotic independence We consider iid vectors $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^{40}$ with all marginals independent and Pareto(1)-distributed. This leads to asymptotic independence, which has already been discussed in Subsection 3.4. Equivalently, $\mathbb{P}(\Theta \in C_\beta) = 1/d$ for β such that $\#\beta = 1$ (and therefore $\mathbb{P}(\Theta \in C_\beta) = 0$ elsewhere). In other words, the spectral measure concentrates on the axis. Our aim is thus to recover these 40 directions among the $2^{40} - 1 \approx 10^{12}$ subsets C_β .

Table 1 shows the average number of errors among the 100 experiments. For the Euclidean projection, the number of errors is quite low in compared to the total number of subsets C_β , especially when n increases. As expected, the angular vector \mathbf{Z} is helpful to detect asymptotic independence since it is likely to concentrate on the axis.

For the DAMEX algorithm, a large ϵ leads theoretically to more mass assigned on the axis. The asymptotic independent case should therefore gives better results for large ϵ . It is the case for our numerical results which become better when ϵ increases. This algorithm also gives better results than the one we propose for $\epsilon \geq 5$. However, some results seems difficult to interpret. Firstly, for $n = 10^4$, a choice of $\epsilon = 0.01$ leads to better results than $\epsilon = 0.1$ or $\epsilon = 0.5$. Secondly, for $\epsilon < 0.5$, the number of errors increases with n .

	Euclidean projection	DAMEX $\epsilon = 0.01$	DAMEX $\epsilon = 0.1$	DAMEX $\epsilon = 0.5$	DAMEX $\epsilon = 1$	DAMEX $\epsilon = 5$	DAMEX $\epsilon = 10$
$n_1 = 10^4$	15.04	41	136.43	64.97	37.39	10.41	6.51
$n_2 = 5 \cdot 10^4$	1.36	264	196.54	64.33	33.27	1.05	0.98
$n_3 = 10^5$	0.47	356	221.30	64.81	0.99	0.39	0.53

Table 1: Average number of errors in the asymptotically independent case ($d = 40$).

A dependent case We now consider a dependent case where extremes do not appear on the axis. In order to include dependence we start from a regularly varying random variable $V \in \mathbb{R}_+$ with tail index $\alpha > 0$. Then, for $r \geq 2$, we consider $r - 1$ independent variables $P_1, \dots, P_{r-1} \in \mathbb{R}_+$, independent of X ,

such that P_j is regularly varying with tail index $\alpha' > \alpha$. Finally, we consider the vector $\mathbf{V} \in \mathbb{R}_+^r$ whose components are defined as follows:

$$V_1 = V \quad \text{and} \quad V_{j+1} = a_j V + P_j, \text{ for } j = 1, \dots, r-1, \quad (5.8)$$

where a_1, \dots, a_{r-1} are positive constants. In this case, the random vector \mathbf{X} is regularly varying with tail index α and a spectral vector Θ which concentrates on the interior of \mathbb{S}_+^{r-1} , that is, on the subset $C_{\{1, \dots, r\}}$.

For our simulations, we consider two vectors $\mathbf{V}_1, \mathbf{V}_2 \in \mathbb{R}_+^{10}$ defined as in (5.8), with $\alpha = 1$, $\alpha' = 2$, and $a_j \in \{0.1, 10\}$. This choice of a_j implies that the vector \mathbf{V}_i , $i = 1, 2$, does not concentrate too much in the center of the subset $C_{\{1, \dots, 10\}}$ but rather near the axis. Finally, we define the vector $\mathbf{X} \in \mathbb{R}_+^{20}$ as the concatenation of \mathbf{V}_1 and \mathbf{V}_2 . In this dependent case, the mass of the spectral measure associated to \mathbf{X} concentrates on both subsets $C_{\{1, \dots, 10\}}$ and $C_{\{11, \dots, 20\}}$. Our aim is then to recover these two subsets among the $2^{20} \approx 10^6$ subsets C_β , based on iid vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ with the same distribution as \mathbf{X} .

Table 2 shows the average number of errors among the 100 experiments. As for the previous case, the Euclidean projection leads to a quite low number of errors compared to the total number of subsets C_β . This means that the vector \mathbf{Z} mainly concentrates on the desired subsets $C_{\{1, \dots, 10\}}$ and $C_{\{11, \dots, 20\}}$. Besides, the number of errors slightly decreases when n increases.

Regarding the DAMEX algorithm, it is more difficult to interpret the results. For $n = 10^4$, the choice of $\epsilon = 0.01$ seems very efficient but the number of errors then drastically increases with n . Contrary to the previous example, large values of ϵ do not appear suitable here, even if they provide more stable results than in the asymptotic independent case. It seems that the best compromise for the different values of n is $\epsilon = 1$. Besides, on this type of data, the Euclidean projection provides better results than the DAMEX algorithm for all different choice of ϵ we make.

	Euclidean projection	DAMEX $\epsilon = 0.01$	DAMEX $\epsilon = 0.1$	DAMEX $\epsilon = 0.5$	DAMEX $\epsilon = 1$	DAMEX $\epsilon = 5$	DAMEX $\epsilon = 10$
$n_1 = 10^4$	5.02	3.00	22.26	5.27	4.20	8.75	7.52
$n_2 = 5 \cdot 10^4$	4.48	68.05	9.95	8.32	6.02	8.39	7.10
$n_3 = 10^5$	4.18	47.42	8.21	7.11	5.81	7.82	6.88

Table 2: Average number of errors in a dependent case ($d = 20$).

These simulations show that there is no easy way to find an optimal value for ϵ . Large ϵ provide good results in the asymptotic independent case, even slightly better than the ones obtained with the Euclidean projection. On the contrary, in the dependent case we propose, the DAMEX seems less efficient, and it is not obvious which value of ϵ should be chosen.

6 Conclusion

In this paper, we introduce the notion of sparsely regularly varying random vectors in order to tackle the issues that arise with the standard 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 us 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 Θ , which implies that the new vector \mathbf{Z} seems more suitable to study extremes in high dimensions. Besides, contrary to the standard regular variation's framework, the sparsity of \mathbf{Z} can be directly captured by studying $\pi(\mathbf{X}/t)$, as stated in Proposition

1. This means that the projection π manages to circumvent to issue of the weak convergence in the definition of regularly varying random vectors. Finally, the results of Proposition 2 and Theorem 2 state that under some assumptions, there is a bijection between the spectral vector Θ and the new angular vector \mathbf{Z} .

Practically speaking, the advantages of the projection π are twofold. Firstly, the Euclidean projection onto the simplex does not introduce any extra parameter. The introduction of ϵ -thickened rectangles in Goix et al. (2017) leads to the choice of a suitable ϵ . The numerical results introduced in Section 5 provide empirical evidence that there is no optimal ϵ . Secondly, the algorithm which computes the projection π takes linear time. Hence, the study of extreme events with π can be done in reasonable time in high dimensions. More generally, the numerical results provide some good results for our approach and encourage to further develop the statistical study of sparsely regularly varying random vectors. In particular, the future work should address the question of the threshold t , or equivalently the level k , and of the threshold introduced in Remark 10.

7 Proofs

Proof of Lemma 1. 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 (2.6).

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

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

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

and 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 proves that $\rho_{a\mathbf{u}} \leq \rho_{\mathbf{u}}$.

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 $\rho_{\mathbf{u}} < k \leq d$.

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

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

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,

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

which gives

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

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

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

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

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

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

and finally

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

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 $j \in \{1, \dots, d\}$, 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 $j \leq \rho$,

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

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

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

which concludes the proof. \square

Proof of Lemma 2. Let $\mathbf{v} \in \mathbb{R}_+^d$. We sort \mathbf{v} in $\boldsymbol{\mu}$ such that $\mu_1 \geq \dots \geq \mu_d$. 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 (2.7) For $i \in \beta^c$, let $j \in \{1, \dots, d\}$ such that $\mu_j = v_i$, and let $\gamma^c \subset \{1, \dots, d\}$ be the subset of such j . By definition of $\rho_{\mathbf{v}}$, the projected vector $\pi(\mathbf{v})$ satisfies $\pi(\mathbf{v})_{\beta^c} = 0$ if and only if for all $j \in \gamma^c$, $j > \rho_{\mathbf{v}}$, which means

$$\mu_j - \frac{1}{j} \left(\sum_{k=1}^j \mu_k - 1 \right) \leq 0. \quad (7.1)$$

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 (7.1) can be rephrased as

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

This inequality is equivalent to

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

which proves (2.7).

For (2.8), set $r = |\beta| \geq 1$ (note that $\beta = \emptyset$ is not possible). Then, the condition $\pi(\mathbf{v}) \in C_\beta$ imply that $\rho_{\mathbf{v}} = r$. Thus, we obtain

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

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

$$\max_{i \in \beta} \sum_{j \in \beta} (v_j - v_i) < 1.$$

On the other hand, the second equality is equivalent to

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

□

Proof of Proposition 1. We only prove (3.3) (the proof of (3.4) is similar).

Let $\beta \in \mathcal{P}_d^*$. Following Lemma 2, we have the equivalence

$$\pi(Y\Theta) \in C_\beta \quad \text{if and only if} \quad \begin{cases} \max_{i \in \beta} \sum_{j \in \beta} (\Theta_j - \Theta_i) < 1/Y, \\ \min_{i \in \beta^c} \sum_{j \in \beta} (\Theta_j - \Theta_i) \geq 1/Y. \end{cases}$$

Hence, (3.3) is equivalent to

$$\mathbb{P}((|\mathbf{X}|/t, \mathbf{X}/|\mathbf{X}|) \in D_\beta \mid |\mathbf{X}| > t) \rightarrow \mathbb{P}((Y, \Theta) \in D_\beta), \quad (7.2)$$

with

$$D_\beta = \left\{ (r, \theta) \in (1, \infty) \times \mathbb{S}_+^{d-1}, \forall i \in \beta, \sum_{j \in \beta} (\theta_j - \theta_i) < 1/r, \text{ and } \forall i \in \beta^c, \sum_{j \in \beta} (\theta_j - \theta_i) \geq 1/r \right\}.$$

This convergence holds if $\mathbb{P}((Y, \Theta) \in \partial D_\beta) = 0$.

The boundary ∂D_β of D_β is included in the union of the subsets

$$\bigcup_{i \in \beta} \partial \left\{ (r, \theta) \in (1, \infty) \times \mathbb{S}_+^{d-1}, \sum_{j \in \beta} (\theta_j - \theta_i) < 1/r \right\},$$

and

$$\bigcup_{i \in \beta^c} \partial \left\{ (r, \theta) \in (1, \infty) \times \mathbb{S}_+^{d-1}, \sum_{j \in \beta} (\theta_j - \theta_i) \geq 1/r \right\},$$

and for all $i = 1, \dots, d$, we have the equality

$$\partial \left\{ (r, \theta) \in (1, \infty) \times \mathbb{S}_+^{d-1}, \sum_{j \in \beta} (\theta_j - \theta_i) < 1/r \right\} = \left\{ (r, \theta) \in (1, \infty) \times \mathbb{S}_+^{d-1}, \sum_{j \in \beta} (\theta_j - \theta_i) = 1/r \right\},$$

and similarly

$$\partial \left\{ (r, \theta) \in (1, \infty) \times \mathbb{S}_+^{d-1}, \sum_{j \in \beta} (\theta_j - \theta_i) \geq 1/r \right\} = \left\{ (r, \theta) \in (1, \infty) \times \mathbb{S}_+^{d-1}, \sum_{j \in \beta} (\theta_j - \theta_i) = 1/r \right\}.$$

This implies that

$$\mathbb{P}((Y, \Theta) \in \partial D_\beta) \leq \sum_{i \in \beta} \mathbb{P} \left(\sum_{j \in \beta} (\Theta_j - \Theta_i) = Y^{-1} \right) + \sum_{i \in \beta^c} \mathbb{P} \left(\sum_{j \in \beta} (\Theta_j - \Theta_i) = Y^{-1} \right),$$

and all these probabilities are null since Y is a continuous random variable independent of Θ . Hence, we proved that $\mathbb{P}((Y, \Theta) \in \partial D_\beta) = 0$ which implies that convergence (7.2) holds and then convergence (3.3) holds as well. \square

Proof of Proposition 2. Fix $\mathbf{x} \in \mathcal{B}(0, 1) \cap \mathbb{R}_+^d$, with $x_j \neq 1/d$ for all $j = 1, \dots, d$. We use (2.9) to write

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

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

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

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

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

\square

Proof of Proposition 3. We fix $\beta \in \mathcal{P}_d^*$ and use Lemma 2. The probability that \mathbf{Z}_{β^c} is null is equal to

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

which proves (3.10).

For Equation (3.11), we use Lemma 2, so that the probability that \mathbf{Z} is concentrated on C_β is equal to

$$\begin{aligned} \mathbb{P}(\mathbf{Z} \in C_\beta) &= \mathbb{P}\left(\max_{j \in \beta} \sum_{k \in \beta} (Y\Theta_k - Y\Theta_j) < 1, \min_{j \in \beta^c} \sum_{k \in \beta} (Y\Theta_k - Y\Theta_j) \geq 1\right) \\ &= \mathbb{P}\left(\left(\max_{j \in \beta} \sum_{k \in \beta} (\Theta_k - \Theta_j)_+\right)^\alpha < Y^{-\alpha}, \min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+\right)^\alpha \geq Y^{-\alpha}\right) \\ &= \int_0^1 \mathbb{P}\left(\max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+\right)^\alpha < u \leq \min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+\right)^\alpha\right) du \\ &= \mathbb{E}\left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+\right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+\right)^\alpha\right)_+\right]. \end{aligned}$$

This concludes the proof of the proposition. \square

Proof of Corollary 1. The proof of Corollary 1 is based on the following lemma, whose result will also be used in other proofs.

Lemma 4. *Let $\beta \in \mathcal{P}_d^*$. Then we have the inequality*

$$\mathbb{P}(\boldsymbol{\Theta} \in C_\beta) \leq \mathbb{P}\left(\max_{j \in \beta} \sum_{k \in \beta} (\Theta_k - \Theta_j)_+ < 1\right) \quad (7.3)$$

Proof of Lemma 4. While Lemma 4 is stated and used in this way, we rather prove the following inequality:

$$\mathbb{P}\left(\max_{j \in \beta} \sum_{k \in \beta} (\Theta_k - \Theta_j)_+ = 1\right) \leq \mathbb{P}(\boldsymbol{\Theta} \notin C_\beta).$$

The first probability can be rephrased as follows:

$$\mathbb{P}\left(\max_{j \in \beta} \sum_{k \in \beta} (\Theta_k - \Theta_j)_+ = 1\right) = \mathbb{P}\left(\sum_{k \in \beta} (\Theta_k - \min_{j \in \beta} \Theta_j) = 1\right) = \mathbb{P}\left(\sum_{k \in \beta} \Theta_k = 1 + \#\beta \min_{j \in \beta} \Theta_j\right).$$

But since $\boldsymbol{\Theta} \in \mathbb{S}_+^{d-1}$, the equality $\sum_{k \in \beta} \Theta_k = 1 + \#\beta \min_{j \in \beta} \Theta_j$ holds only if there exists $k \in \beta$ such that $\Theta_k = 0$. Thus, we obtain the inequality

$$\mathbb{P}\left(\max_{j \in \beta} \sum_{k \in \beta} (\Theta_k - \Theta_j)_+ = 1\right) \leq \mathbb{P}(\exists k \in \beta, \Theta_k = 0) \leq \mathbb{P}(\boldsymbol{\Theta} \notin C_\beta),$$

which concludes the proof. \square

We now prove Corollary 1. We fix $\beta \in \mathcal{P}_d^*$ and assume that $\mathbb{P}(\Theta \in C_\beta) > 0$. Then, starting from Equation (3.11), we write

$$\begin{aligned}
\mathbb{P}(\mathbf{Z} \in C_\beta) &\geq \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbf{1}_{\Theta \in C_\beta} \right] \\
&= \mathbb{E} \left[\left(\left(\sum_{k \in \beta} \Theta_k \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbf{1}_{\Theta \in C_\beta} \right] \\
&= \mathbb{E} \left[\left(1 - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right) \mathbf{1}_{\Theta \in C_\beta} \right] \\
&= \mathbb{E} \left[\left(1 - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right) \mid \Theta \in C_\beta \right] \mathbb{P}(\Theta \in C_\beta).
\end{aligned}$$

The expectation is positive by Lemma 4 and the probability $\mathbb{P}(\Theta \in C_\beta)$ is positive by assumption. This shows that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$. \square

Proof of Theorem 1. We separately prove both implications.

We first consider $\beta \in \mathcal{P}_d^*$ such that C_β is a maximal subset for Θ :

$$\mathbb{P}(\Theta \in C_\beta) > 0 \quad \text{and} \quad \mathbb{P}(\Theta \in C_{\beta'}) = 0, \quad \text{for } \beta' \supsetneq \beta.$$

By Corollary 1, we already know that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$. Besides, if $\beta' \supsetneq \beta$, then Equation (3.14) gives

$$\mathbb{P}(\mathbf{Z} \in C_{\beta'}) \leq \mathbb{P}(\mathbf{Z}_{\beta'} > 0) \leq \mathbb{P}(\Theta_{\beta'} > 0).$$

and this last probability equals zero since C_β is a maximal subset for Θ . This proves that C_β is a maximal subset for \mathbf{Z} .

We now consider $\beta \in \mathcal{P}_d^*$ such that C_β is a maximal subset for \mathbf{Z} :

$$\mathbb{P}(\mathbf{Z} \in C_\beta) > 0 \quad \text{and} \quad \mathbb{P}(\mathbf{Z} \in C_{\beta'}) = 0, \quad \text{for } \beta' \supsetneq \beta.$$

First note that, for $\beta' \supsetneq \beta$, $\mathbb{P}(\Theta \in C_{\beta'}) = 0$. If not, Corollary 1 implies that $\mathbb{P}(\mathbf{Z} \in C_{\beta'}) > 0$, which contradicts the maximality of C_β for \mathbf{Z} .

Secondly, Equation (3.11) of Proposition 3 gives

$$\begin{aligned}
\mathbb{P}(\mathbf{Z} \in C_\beta) &= \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \right] \\
&= \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbf{1}_{\Theta \in C_\beta} \right] \\
&\quad + \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbf{1}_{\Theta \notin C_\beta} \right] \\
&= (A) + (B).
\end{aligned} \tag{7.4}$$

The first term (A) has already been calculated in the proof of Corollary 1. It is equal to

$$(A) = \mathbb{E} \left[1 - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \mid \Theta \in C_\beta \right] \mathbb{P}(\Theta \in C_\beta).$$

For the second term (B), note that the assumption $\Theta \notin C_\beta$ implies that there exists $l \in \beta$ such that $\Theta_l = 0$, or that there exists $r \in \beta^c$ such that $\Theta_r > 0$. We then decompose (B) into two terms:

$$\begin{aligned} & \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbb{1}_{\Theta \notin C_\beta} \right] \\ & \leq \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbb{1}_{\exists l \in \beta, \Theta_l = 0} \right] \\ & + \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbb{1}_{\exists \beta' \supseteq \beta, \Theta \in C_{\beta'}} \right]. \end{aligned}$$

The first expectation is equal to

$$\mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \left(\sum_{k \in \beta} (\Theta_k)_+ \right)^\alpha \right)_+ \mathbb{1}_{\exists l \in \beta, \Theta_l = 0} \right],$$

and is thus zero. The second expectation is smaller than $\mathbb{P}(\exists \beta' \supseteq \beta, \Theta \in C_{\beta'})$ which is zero. Indeed, if $\mathbb{P}(\exists \beta' \supseteq \beta, \Theta \in C_{\beta'}) > 0$, then by Corollary 1, we also have $\mathbb{P}(\exists \beta' \supseteq \beta, \mathbf{Z} \in C_{\beta'}) > 0$, which contradicts the maximality of C_β for \mathbf{Z} . All in all, this proves that (B) = 0.

Going back to Equation (7.4), we have proved that

$$\mathbb{P}(\mathbf{Z} \in C_\beta) = (A) = \mathbb{E} \left[1 - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \mid \Theta \in C_\beta \right] \mathbb{P}(\Theta \in C_\beta).$$

By Lemma 4, we know that the expectation is positive. Hence, the assumption $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$ implies that $\mathbb{P}(\Theta \in C_\beta) > 0$, which proves that C_β is a maximal subset of Θ . \square

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

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

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

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

Finally, we conclude the proof with Lemma 1:

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

□

Proof of Lemma 3. 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 Θ the spectral vector of \mathbf{X} and consider a random variable Y which follows a Pareto(α) distribution and is independent of Θ . For $A \in \mathcal{B}(\mathbb{S}_+^{d-1})$ such that $\mathbb{P}(\Theta \in \partial A) = 0$, and $\epsilon > 0$, we have

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

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

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 $t \rightarrow \infty$, and by $l_\epsilon^-(A)$ the liminf in (4.4) when $t \rightarrow \infty$. For $u \geq 1$, we decompose the interval (u, ∞) as follows:

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

Then for $t > 0$,

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

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

$$\frac{\mathbb{P} \left(\frac{|\mathbf{X}|}{u(1 + \epsilon)^k} > t \right)}{\mathbb{P}(|\mathbf{X}| > t)} = \mathbb{P} \left(|\mathbf{X}| > tu(1 + \epsilon)^k \mid |\mathbf{X}| > t \right) \rightarrow (u(1 + \epsilon)^k)^{-\alpha}, \quad t \rightarrow \infty. \quad (7.5)$$

Besides, we know by (4.3) that

$$\liminf_{t \rightarrow \infty} \epsilon^{-1} \mathbb{P} \left(\frac{|\mathbf{X}|}{tu(1 + \epsilon)^k} \in (1, 1 + \epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \mid \frac{|\mathbf{X}|}{u(1 + \epsilon)^k} > t \right) = l_\epsilon^-(A). \quad (7.6)$$

We now gather (7.5) and (7.6) and use Fatou's lemma to conclude:

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

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

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

$$\limsup_{t \rightarrow \infty} \epsilon^{-1} \mathbb{P} \left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \mid \frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t \right) = l_{\epsilon}^{+}(A). \quad (7.7)$$

Thus, Equations (7.5) and (7.7) and Fatou's lemma allow us to write

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

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

This proves that

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

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 Θ defined by $\mathbb{P}(\Theta \in \cdot) = \alpha^{-1} l(\cdot)$. \square

Proof of Theorem 2. The proof is based on Lemma 3. 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 (3.6). 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:

$$\pi((1+\epsilon)\mathbf{Z}) > \mathbf{x} \iff \mathbf{Z} > \frac{\mathbf{x} + \epsilon/d}{1+\epsilon}. \quad (7.8)$$

This is a consequence of Equation (2.9) 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

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

converges when $t \rightarrow \infty$, $\epsilon \rightarrow 0$. Following Equation (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 $t \rightarrow \infty$. Then, Proposition 4 gives

$$\begin{aligned} \mathbb{P}(Y \in (1, 1 + \epsilon], \mathbf{Z} \in A_{\mathbf{x}}) &= \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}} \mid Y > 1 + \epsilon) \mathbb{P}(Y > 1 + \epsilon) \\ &= \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\pi((1 + \epsilon)\mathbf{Z}) \in A_{\mathbf{x}}) (1 + \epsilon)^{-\alpha} \\ &= [1 - (1 + \epsilon)^{-\alpha}] \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + [\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\pi((1 + \epsilon)\mathbf{Z}) \in A_{\mathbf{x}})] (1 + \epsilon)^{-\alpha}. \end{aligned} \tag{7.9}$$

The first term divided by ϵ converges to $\alpha \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}})$ when $\epsilon \rightarrow 0$. We use (7.8) 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

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

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

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

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

For the second step, we define

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

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

We first deal with the limsup. Assume that $|\mathbf{X}|/t \in (1, 1 + \epsilon]$. Then $(|\mathbf{X}|/t - 1 - \epsilon)/d \leq 0$. 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

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

and thus

$$\limsup_{t \rightarrow \infty} (\star) \leq \epsilon^{-1} [\mathbb{P}(Y \in (1, 1 + \epsilon], \mathbf{Z} \in A_{\mathbf{x} - \epsilon/d}) - \mathbb{P}(Y \in (1, 1 + \epsilon], \mathbf{Z} \in A_{\mathbf{x}})] =: \epsilon^{-1} [P_1(\epsilon) - P_2(\epsilon)].$$

We use Proposition 4 and Equation (7.8) 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 $\epsilon \rightarrow 0$, whereas the second one is equal to

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

This proves that $\epsilon^{-1}P_1(\epsilon)$ converges to $\alpha G_{\mathbf{Z}}(\mathbf{x}) + dG_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d)$ when $\epsilon \rightarrow 0$. For $P_2(\epsilon)$, we refer to (7.9) in which we proved that $\epsilon^{-1}P_2(\epsilon)$ converges to $\alpha G_{\mathbf{Z}}(\mathbf{x}) + dG_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d)$ when $\epsilon \rightarrow 0$. All in all we proved that $\epsilon^{-1}[P_1(\epsilon) - P_2(\epsilon)] \rightarrow 0$, when $\epsilon \rightarrow 0$.

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

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

and thus

$$\liminf_{t \rightarrow \infty} (\star) \geq \epsilon^{-1} [\mathbb{P}(Y \in (1, 1 + \epsilon), \mathbf{Z} \in A_{(1 + \epsilon)\mathbf{x}}) - \mathbb{P}(Y \in (1, 1 + \epsilon), \mathbf{Z} \in A_{\mathbf{x}})] =: \epsilon^{-1}[P_3(\epsilon) - P_4(\epsilon)].$$

We use again Proposition 4 and Equation (7.8) to compute $P_3(\epsilon)$:

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

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

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

This proves that $P_3(\epsilon)$ converges to $\alpha G_{\mathbf{Z}}(\mathbf{x}) + dG_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d)$ when $\epsilon \rightarrow 0$. Note that $P_4(\epsilon) = P_2(\epsilon)$, so that $P_4(\epsilon)$ converges to $\alpha G_{\mathbf{Z}}(\mathbf{x}) + dG_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d)$ when $\epsilon \rightarrow 0$. All in all we proved that $\epsilon^{-1}[P_3(\epsilon) - P_4(\epsilon)] \rightarrow 0$, when $\epsilon \rightarrow 0$.

Gathering all these results together, we can write

$$\epsilon^{-1}[P_3(\epsilon) - P_4(\epsilon)] \leq \liminf_{t \rightarrow \infty} (\star) \leq \limsup_{t \rightarrow \infty} (\star) \leq \epsilon^{-1}[P_1(\epsilon) - P_2(\epsilon)].$$

Since $\epsilon^{-1}[P_1(\epsilon) - P_2(\epsilon)]$ and $\epsilon^{-1}[P_3(\epsilon) - P_4(\epsilon)]$ converge to 0 as $\epsilon \rightarrow 0$, we proved that $\lim_{\epsilon \rightarrow 0} \liminf_{t \rightarrow \infty} (\star) = \lim_{\epsilon \rightarrow 0} \limsup_{t \rightarrow \infty} (\star) = 0$.

To conclude the proof, we write

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

and both steps lead to

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

and

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

Since $|\mathbf{X}|$ is regularly varying with tail index α , we apply Lemma 3 to conclude that \mathbf{X} is regularly varying with tail index α and with spectral vector Θ satisfying $\mathbb{P}(\Theta \in A_{\mathbf{x}}) = \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + \alpha^{-1}dG_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d)$. \square

Acknowledgement We are grateful to two referees and an associate editor for numerous remarks that significantly improved the paper.

A Appendix

We introduce here the linear-time algorithm given in Duchi et al. (2008). It is based on a random selection of the coordinates.

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 $\eta = (s - z)/\rho$;
Define \mathbf{w} s.t. $w_i = v_i - \eta$.

Algorithm 3: Linear time projection onto the positive sphere $\mathbb{S}_+^{d-1}(z)$.

References

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D., and Ferro, C. (2006). *Statistics of Extremes: Theory and Applications*. 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. *Electronic Journal of Statistics*, 9(1):383–418.
- Chiapino, M. and Sabourin, A. (2016). Feature clustering for extreme events analysis, with application to extreme stream-flow data. In *International Workshop on New Frontiers in Mining Complex Patterns*, pages 132–147. Springer.
- Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392.
- Condat, L. (2016). Fast projection onto the simplex and the ℓ_1 ball. *Mathematical Programming*, 158(1-2):575–585.
- Cooley, D. and Thibaud, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106(3):587–604.
- Duchi, J., Shalev-Shwartz, S., Singer, Y., and Chandra, T. (2008). Efficient projections onto the ℓ_1 -ball for learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, pages 272–279. ACM.
- 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. *The Annals of Statistics*, 29(5):1401–1423.
- 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. *The Annals of Statistics*, 37(5B):2953–2989.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
- Goix, N., Sabourin, A., and Cléménçon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*, 161:12–31.
- Guillette, 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.
- Hult, H. and Lindskog, F. (2006). Regular variation for measures on metric spaces. *Publications de l’Institut Mathématique*, 80(94):121–140.
- Janßen, A. and Wan, P. (2019). k -means clustering for extremes. *arXiv:1904.02970*.
- Jessen, A. H. and Mikosch, T. (2006). Regularly varying functions. *Publications de L’institut Mathématique*, 80(94):171–192.
- Kyrillidis, A., Becker, S., Cevher, V., and Koch, C. (2013). Sparse projections onto the simplex. In *International Conference on Machine Learning*, volume 28, pages 235–243.
- Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.
- 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. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 657–664, New-York.
- Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):219–241.
- Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer, New-York.
- Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New-York.
- Sabourin, A. and Drees, H. (2019). Principal component analysis for multivariate extremes. *arXiv:1906.11043*.
- Sabourin, A. and Naveau, P. (2014). Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization. *Computational Statistics & Data Analysis*, 71:542–567.
- Sabourin, A., Naveau, P., and Fougères, A.-L. (2013). Bayesian model averaging for multivariate extremes. *Extremes*, 16(3):325–350.
- Segers, J. (2012). Max-stable models for multivariate extremes. *Revstat Statistical Journal*, 10:61–82.
- Simpson, E., Wadsworth, J., and Tawn, J. (2019). Determining the dependence structure of multivariate extremes. *arXiv:1809.01606*.