



HAL
open science

Detection of extremal directions via Euclidean projections

Nicolas Meyer, Olivier Wintenberger

► **To cite this version:**

Nicolas Meyer, Olivier Wintenberger. Detection of extremal directions via Euclidean projections. 2020. hal-02168495v3

HAL Id: hal-02168495

<https://hal.science/hal-02168495v3>

Preprint submitted on 14 May 2020 (v3), 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.

Detection of extremal directions via Euclidean projections

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

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

May 14, 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. However, 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. [Resnick \(1987\)](#), [Beirlant et al. \(2006\)](#), or [Resnick \(2007\)](#), [Embrechts et al. \(2013\)](#). [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 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

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

^{*}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 $|\mathbf{X}|/t$ follows 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 . These approaches tackle nonstandard regular variation for which $\alpha = 1$ and all marginals are tail equivalent (possibly after a standardization). 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 via 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 Guillothe 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. However, the complete support's estimation is often difficult to capture in high dimensions so that dimensionality reduction algorithms have been proposed. There, a main goal in the tail dependence's study is to identify clusters of components which are likely to be extreme together. This approach has firstly been introduced by Chautru (2015) who uses the Principal Nested Spheres 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. 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) provide an incremental-type algorithm (CLustering Extreme Features, CLEF) to group components which may be large together. This algorithm is based on the DAMEX algorithm and also requires a hyperparameter κ_{\min} . Several variants of the CLEF algorithm have then been proposed by Chiapino et al. (2019). These approaches differ in the stopping criteria which are based on asymptotic results of the coefficient of tail dependence. 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 $(\tau_\beta)_{\beta \subset \{1, \dots, d\}}$ which describes to what extent the direction β gathers extreme values. All these approaches address nonstandard regular variation and are thus based on a rank transform which provides empirical normalization of the marginals.

Some alternative approaches have also been recently proposed Lehtomaa and Resnick (2019) analyze 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 authors then build an asymptotically normal test statistic to validate the support estimate. Cooley and Thibaud (2019) define a vector

space on the positive orthant \mathbb{R}_+^d in order to conciliate both Principal Component Analysis 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.

Many of the aforementioned approaches rely on dimension reduction since large events often concentrate in subspaces of dimension smaller than $d - 1$. 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*. Identifying the low-dimensional subspaces on which the spectral measure puts mass is therefore a first major step in the study of multivariate extremes. However, as soon as $r < d$, the weak convergence (1.1) does not hold for subsets 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 identifying the possible sparsity of S is a major challenge in multivariate extreme value analysis. A review on sparsity in multivariate EVT has been established by Engelke and Ivanovs (2020).

Sparsity arises all the more for standard regular variation: There, it is possible that the marginals of \mathbf{X} are not tail equivalent and thus that the support of the spectral measure is included in \mathbb{S}_+^{r-1} for $r \ll d$. This is the approach we use in this article. This framework is for instance used by Sabourin and Drees (2019) who assume that the parameter r is known. The aim of their paper is to identify the sphere \mathbb{S}_+^{r-1} with an empirical risk minimization's technique. On the contrary, nonstandard regular variation entails that all marginals are standardized so that they the support of the spectral measure can not be included in a smaller sphere. Even if both approaches are not comparable, the aim of the spectral measure's support identification is similar: capturing the extremal directions of \mathbf{X} .

In both cases the self-normalized vector $\mathbf{X}/|\mathbf{X}|$ fails to identify the directions on which the spectral measure puts mass. This is why we 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, since 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 should also 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. With this approach we obtain a new angular limit vector \mathbf{Z} whose distribution slightly changes from the spectral measure. We prove that under mild conditions both concepts of regular variation are equivalent and we give the relation between both limit measures. Besides, we study this new angular limit and show that it empirically captures the extremal directions of \mathbf{X} .

Our method proves to be an alternative to the one proposed by Goix et al. (2017). There, the authors work with the spectral measure and propose an algorithmic approach which needs a hyperparameter ϵ . On the other hand we slightly modify the standard case and work with an angular vector \mathbf{Z} to exhibit the sparsity structure of multivariate extremes. A main advantage of our approach is that it does not need any hyperparameter to capture the tail dependence. The numerical results we provide emphasize the efficiency of our method to detect directions $\beta \subset \{1, \dots, d\}$ which may be large together. These results also highlight how the new vector \mathbf{Z} provides an interpretation of the relative importance of a coordinate j in an extremal direction β .

Outline The structure of this paper is as follows. Section 2 gathers all theoretical results useful in this paper. We first introduce the EVT framework and especially multivariate regular variation. 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. We also explain which issues appear in this context. We introduce then the Euclidean projection onto the simplex and list several results which are of constant use for our study. In Section 3 we introduce the notion of sparse regular variation. We study the distribution of the new angular vector and compare it with the spectral measure. Finally we establish the equivalence under mild conditions between sparse regular variation and the standard regular variation's concept. In Section 4 we discuss to what extent the Euclidean projection allows us to better capture the extremal directions of \mathbf{X} . We prove that Θ and \mathbf{Z} similarly behave on so-called maximal subsets. 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).

2 Theoretical background

2.1 Notation

We introduce some standard notation that is used throughout the paper. Symbols in bold such as $\mathbf{x} \in \mathbb{R}^d$ are column vectors with components denoted by $x_j, j \in \{1, \dots, d\}$. Operations and relationships involving such vectors are meant componentwise. We 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 . 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 $\beta = \{\beta_1, \dots, \beta_r\} \subset \{1, \dots, d\}$, then \mathbf{x}_β denotes the vector $(x_{\beta_1}, \dots, x_{\beta_r})$ 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 complementary of F (in E).

2.2 Multivariate regular variation

We consider a nonnegative random vector $\mathbf{X} \in \mathbb{R}_+^d$ and our aim is to assess the tail structure of \mathbf{X} . It is customary in EVT to assume that the random vector \mathbf{X} is regularly varying: There exist a positive sequence $(a_n), a_n \rightarrow \infty$ when $n \rightarrow \infty$, and a nonnegative Radon measure μ on $\mathbb{R}_+^d \setminus \{\mathbf{0}\}$ such that

$$n\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot) \xrightarrow{v} \mu(\cdot), \quad n \rightarrow \infty, \quad (2.1)$$

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

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

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

It is even possible to choose two different norms $|\cdot|$ and $|\cdot|'$ in the definition of T but it will not be useful for our purpose. Classical choices of norms are ℓ^p -norms, $p \in [1, \infty]$. The spectral measure S of the regularly varying random vector \mathbf{X} is then defined as

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

for all Borel subset B of \mathbb{S}_+^{d-1} . The homogeneity property of the tail measure implies that

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

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

Equation (2.2) can be rephrased as $\alpha r^{-(\alpha+1)}drS(d\theta) = \mu \circ T^{-1}(dr, d\theta)$. This gives a decomposition of the tail measure in a radial part and an angular part. The radial component can thus be modeled through a random variable with Pareto(α) distribution while the angular one is characterized by the spectral measure S . The decomposition in Equation (2.2) ensures that the radial and the angular parts are independent. Combining Equations (2.1) and (2.2) entails the following characterization of a regularly varying random vector in \mathbb{R}_+^d :

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

This is the same as

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

see Resnick (1986). We call *spectral vector* a random vector on \mathbb{S}_+^{d-1} whose distribution is S .

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

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \frac{\mathbf{X}}{|\mathbf{X}|}\right) \in \cdot \mid |\mathbf{X}| > t\right) \xrightarrow{w} \mathbb{P}((Y, \Theta) \in \cdot), \quad t \rightarrow \infty. \quad (2.4)$$

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 which has for distribution the spectral measure S . Regarding extreme values, Equation (2.4) 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 especially in high dimensions.

Example 1. We consider two independent regularly varying random variables $X_1, X_2 \in \mathbb{R}_+$ with tail index $\alpha_1 > 0$ and $\alpha_2 > \alpha_1$ respectively. The vector $\mathbf{X} = (X_1, X_2) \in \mathbb{R}_+^2$ is then regularly varying with tail index α_1 and spectral vector $\Theta = (1, 0)$ a.s. This example represents the case of a marginal X_1 with a larger tail than another one X_2 and depicts therefore a situation where extreme events are only due to one component. In this case our aim is to identify this largest component.

Example 2. We keep the same notation as in Example 1. If we define $\mathbf{X} = (X_1, X_1 + X_2)$, then the vector \mathbf{X} is regularly varying with tail index α_1 and spectral measure $\Theta = (1, 1)/|(1, 1)|$ a.s. In this case we have a strong dependence between the two marginals and the spectral measure places mass in the center of the positive unit sphere.

Example 3. We consider a regularly varying random variable $X \in \mathbb{R}_+$ with tail index $\alpha > 0$. Then the vector $\mathbf{X} = (X, X/2)$ satisfies the convergence in distribution

$$\mathbf{X}/t \mid |\mathbf{X}| > t \xrightarrow{d} (Y, Y/2), \quad t \rightarrow \infty.$$

where Y follows a Pareto(α) distribution. This means that \mathbf{X} is regularly varying with tail index α and spectral vector $\Theta = (1, 1/2)/|(1, 1/2)|$ a.s. If we choose the ℓ^1 -norm we obtain $\Theta = (2/3, 1/3)$ a.s. In this case it seems challenging to capture exactly this direction. We are rather willing to detect that the first marginal is larger than the second one while both contribute to the extremal behavior of \mathbf{X} .

2.3 Sparsity and dimension reduction

The convergence in (2.1) may be seen as standard regular variation in contrast to the nonstandard one: There exists a nonnegative Radon measure $\tilde{\mu}$ on $\mathbb{R}_+^d \setminus \{\mathbf{0}\}$ such that

$$n\mathbb{P}((X_i/a_{n,i})_{1 \leq i \leq d} \in \cdot) \xrightarrow{v} \tilde{\mu}(\cdot), \quad n \rightarrow \infty,$$

where the sequences $(a_{n,i})_n$ are satisfying $n\mathbb{P}(X_i > a_{n,i}) \rightarrow 1$ as $n \rightarrow \infty$ (see Resnick (2007), Section 6.5.6). Actually standard regular variation (2.1) is more general since it allows the tail measure to be sparse which means that it places mass on some lower-dimensional subspaces of \mathbb{R}_+^d . On the other hand, in the nonstandard case, the condition satisfied by the $a_{n,i}$ implies that $\tilde{\mu}(\{\mathbf{x} \in \mathbb{R}_+^d, x_i > 1\}) = 1$ for all $i = 1, \dots, d$. This means that the mass of the measure $\tilde{\mu}$ is concentrated in all directions. The study of extreme values under the nonstandard assumption begins with a modification of the marginals called *rank transform* which provides equally distributed components (see Resnick (1987), Proposition 5.10). This is the approach mostly used in the literature (see Goix et al. (2016), Goix et al. (2017), Chiapino and Sabourin (2016), Simpson et al. (2019), Lehtomaa and Resnick (2019)). In this paper we do not consider any transformation of the marginals and we only focus on standard regular variation defined by (2.1). This means that we assume that the tail measure μ is likely to be sparse.

Remark 1. In the standard case the spectral measure's support can be included in \mathbb{S}_+^{r-1} with $r \ll d$. The first step in the study of multivariate extremes should therefore deal with the identification of this low-dimensional sphere. This case can not arise in the nonstandard case since all marginals have the same tail index which implies that the spectral measure puts mass in all the directions $j = 1, \dots, d$.

Estimating a sparse support boils down to focusing on the low-dimensional subspaces on which this measure puts mass. In high dimensions, several authors recently addressed this issue with the aim of detecting directions that are likely to be extreme together (Chautru (2015), Goix et al. (2017), Chiapino and Sabourin (2016), Chiapino et al. (2019)). This requires dimension reduction techniques in order to move from a high-dimensional sphere \mathbb{S}^{d-1} to these low-dimensional subspaces. Regarding the spectral measure, 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 consider the subsets

$$C_\beta = \{\mathbf{x} \in \mathbb{S}_+^d, x_i > 0 \text{ for } i \in \beta, x_i = 0 \text{ for } i \notin \beta\}, \quad (2.5)$$

for $\beta \in \mathcal{P}_d^*$. This approach can be related to the one developed by Goix et al. (2017) (see Remark 6). 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 is given in Figure 1.

This partition is helpful to study the tail 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, 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.

Remark 2. The notion of sparsity in EVT 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). It corresponds to the assumption (S2.a) in Engelke and Ivanovs (2020). The second notion deals with the number

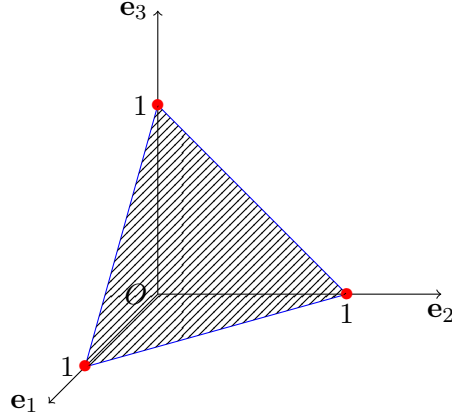


Figure 1: The subsets C_β in dimension 3 for to the ℓ^1 -norm. In red, the subsets $C_{\{1\}}$, $C_{\{2\}}$, and $C_{\{3\}}$. In blue, the subsets $C_{\{1,2\}}$, $C_{\{1,3\}}$, and $C_{\{2,3\}}$. The shaded part corresponds to the interior of the simplex, that is, the subset $C_{\{1,2,3\}}$.

of null coordinates 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, $|\Theta|_0 = \#\{i = 1, \dots, d, \theta_i \neq 0\}$. This is denoted by (S2.b) in [Engelke and Ivanovs \(2020\)](#). 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, see for instance [Ledford and Tawn \(1996\)](#) or [Ramos and Ledford \(2009\)](#). Even in cases of asymptotic 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 relevant to identify the largest 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*

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

In terms of extreme values, the notion of maximality can be rephrased in the following way. First, $\mathbb{P}(\Theta \in C_\beta) > 0$ 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 3. 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 $\beta' \supsetneq \beta$, such that $\mathbb{P}(\Theta \in C_{\beta'}) = 0$, then C_β is a maximal subset itself. If not, we consider $\beta' \supsetneq \beta$ 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 $\gamma' \supsetneq \gamma$.

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 on the second component

of convergence (2.4). Indeed, the angular component of \mathbf{X} satisfies the convergence

$$\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in \cdot \mid |\mathbf{X}| > t) \xrightarrow{w} \mathbb{P}(\Theta \in \cdot), \quad t \rightarrow \infty. \quad (2.6)$$

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. However 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 $\beta \neq \{1, \dots, d\}$. Equation (2.6) is thus 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 $\beta \neq \{1, \dots, d\}$), while the data generally do not concentrate on such subspaces.

In this context many authors approximate the behavior of extremes by working on subsets that are close to the C_β 's and for which the aforementioned issue vanishes. Goix et al. (2016) define the truncated ϵ -cones as

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

and study the behavior of the tail measure on these cones. Subsequently, Goix et al. (2017) introduce the notion of ϵ -thickened rectangles:

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

Chiapino and Sabourin (2016) relax the condition on β^c and define the rectangles

$$\{\mathbf{x} \in \mathbb{R}_+^d, x_i > 1 \text{ for } i \in \beta\}.$$

All these approaches lead to efficient dimensionality reduction algorithms. In this article, we do not address the issue that arises with the weak convergence on the subsets C_β by using alternative subsets. Instead, we rather focus on the vector $\mathbf{X}/|\mathbf{X}|$ in Equation (2.6) and provide a slight modification of this vector to obtain a convergence on the C_β 's. This needs to consider another way of projecting the vector \mathbf{X} on the positive unit sphere. This projection has to circumvent the weak convergence's issue and to highlight the extremal directions of \mathbf{X} .

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 (2.4). 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 (2.5) 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 $z > 0$.

2.4 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:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{w} - \mathbf{v}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{w}\|_1 = z. \quad (2.7)$$

Since $\mathbf{v} \geq 0$, the minimization problem (2.7) 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 $O(d)$. However, 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 $\pi_z(\mathbf{v})$:

$$\rho = \max \left\{ j \in \{1, \dots, d\}, \mu_j - \frac{1}{j} \left(\sum_{r=1}^j \mu_r - z \right) > 0 \right\}, \quad (2.8)$$

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 4. Multivariate extremes have already been studied in low dimensions especially in the bivariate case (see Einmahl et al. (2001) or Einmahl and Segers (2009)). However, the study of large events becomes a challenging issue when the dimension increases. The recent inferential methods developed by Simpson et al. (2019) or Goix et al. (2017) reach both a computational complexity $O(dn \log(n))$ for n representing the number of data points. While the main part of their algorithms is linear, the $n \log(n)$ terms come from the standardization of the marginals which requires to sort them. In our approach we assume that the different marginals do not necessary have the same tail index and therefore no standardization is required (see Section (2.3)). Algorithm 3 ensures then an expected 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 $\lambda_{\mathbf{v},1}$:

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

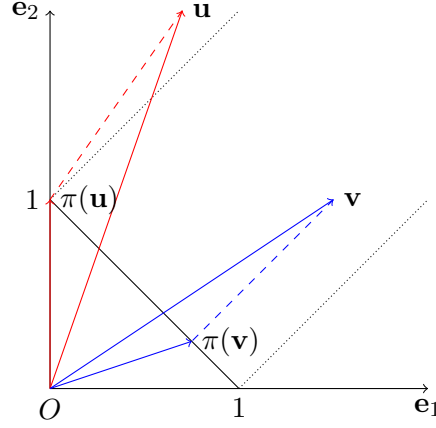


Figure 2: The Euclidean projection onto the simplex \mathbb{S}_+^1 .

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

We list below some straightforward results satisfied by 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.

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 (2.4). The idea is indeed to substitute the quantity $\mathbf{X}/|\mathbf{X}|$ in (2.4) for $|\cdot| = |\cdot|_1$ by $\pi(\mathbf{X}/t)$ and to manage to get same convergence results. A natural way to do this relies on the continuous mapping theorem.

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 tail dependence of \mathbf{X} , 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:*

$$\pi(\mathbf{v})_{\beta^c} = 0 \quad \text{if and only if} \quad 1 \leq \min_{i \in \beta^c} \sum_{k=1}^d (v_k - v_i)_+, \quad (2.9)$$

and

$$\pi(\mathbf{v}) \in C_\beta \quad \text{if and only if} \quad \begin{cases} \max_{i \in \beta} \sum_{j \in \beta} (v_j - v_i) < 1, \\ \min_{i \in \beta^c} \sum_{j \in \beta} (v_j - v_i) \geq 1. \end{cases} \quad (2.10)$$

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

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

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

$$\pi(\mathbf{v}) > \mathbf{x} \quad \text{if and only if} \quad \mathbf{v} > \mathbf{x} + \frac{|\mathbf{v}| - 1}{d}. \quad (2.11)$$

This equivalence will be of constant use in the proofs.

Remark 5. 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 can therefore not be used for the Euclidean projection onto the simplex.

3 Sparse regular variation

3.1 Regular variation and projection

In all this section we consider a random vector \mathbf{X} in \mathbb{R}_+^d . Recall that if \mathbf{X} is regularly varying, then there exists $\alpha > 0$ and a random variable Y following a $\text{Pareto}(\alpha)$ distribution independent of Θ such that

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

We emphasized in Subsection 2.3 that convergence (3.1) is not helpful to capture the sparsity of the spectral vector Θ . Thus, we substitute the self-normalized extreme $\mathbf{X}/|\mathbf{X}|$ 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 2.3, 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 $\beta \neq \{1, \dots, d\}$ while real-world data do not concentrate on such subspaces. The main issue is that a vector $\mathbf{x} \in \mathbb{R}_+^d$ has to satisfy the condition $\mathbf{x}_{\beta^c} = 0$ so that the self-normalized vector $\mathbf{x}/|\mathbf{x}|$ belongs to C_β . Our idea is to use the Euclidean projection to relax this condition. We replace the set $\{\mathbf{x} \in \mathbb{R}_+^d, |\mathbf{x}| > 1, \mathbf{x}/|\mathbf{x}| \in C_\beta\}$ by $\{\mathbf{x} \in \mathbb{R}_+^d, |\mathbf{x}| > 1, \pi(\mathbf{x}/t) \in C_\beta\}$ which have positive Lebesgue measure. There, a vector \mathbf{x} satisfies $\pi(\mathbf{x}) \in C_\beta$ even if $\mathbf{x}_{\beta^c} > 0$. Figure 2 provides an illustration of the subsets $\{\mathbf{x} \in \mathbb{R}_+^d, |\mathbf{x}| > 1, \pi(\mathbf{x}/t) \in C_\beta\}$ for $\beta = \{1\}$ or $\beta = \{2\}$: They correspond to the areas between the axis and the dotted lines.

Example 4. 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.

Remark 6. 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. [Goix et al. \(2017\)](#) define for instance ϵ -thickened rectangles

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

for $\beta \in \mathcal{P}_d^*$ and $\epsilon > 0$. These subsets rely 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 established in Section 5.

With this in mind, we substitute the usual projection $\mathbf{X}/|\mathbf{X}|$ by $\pi(\mathbf{X}/t)$. The continuity of the projection π implies that

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

where $\mathbf{Z} = \pi(Y\Theta)$. Contrary to the limit in (3.1), we lose independence between the radial component Y and the angular component \mathbf{Z} of the limit. The dependence relation between both components will be detailed in Proposition 2.

3.2 The distribution of \mathbf{Z}

The purpose of this section is to explicit a relation between the new angular vector \mathbf{Z} and the spectral vector Θ . To this end, we define the function $G_{\mathbf{Z}}$ by

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

The function $G_{\mathbf{Z}}$ characterizes the distribution of \mathbf{Z} . However, 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

$$A_{\mathbf{x}} = \{\mathbf{u} \in \mathbb{S}_+^{d-1}, x_1 < u_1, \dots, x_d < u_d\}, \quad (3.4)$$

$\mathbf{x} \in \mathcal{B}(0, 1) \cap \mathbb{R}_+^d$. 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 (2.11), we can express the condition $\mathbf{Z} > \mathbf{x}$ in terms of Θ .

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

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

Proposition 1 gives a relation between the distribution of \mathbf{Z} and the one of Θ . While its complexity makes it difficult to use in all generality, specific choices for \mathbf{x} lead to 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

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

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

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

This quantity is null if and only if for all $j = 1, \dots, d$, $\Theta_j = 1/d$ a.s. and is equal to 1 if and only if $\min_{1 \leq j \leq d} \Theta_j = 0$ a.s. This implies that 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} . More insight into the sparsity of the vector \mathbf{Z} is given in Section 4.

3.3 Sparse regular variation

In Section (3.1) we proved that the angular part $\pi(\mathbf{X}/t)$ of a regularly varying random vector \mathbf{X} converges to $\mathbf{Z} = \pi(Y\Theta)$ when $t \rightarrow \infty$, see Equation (3.2). This encourages to introduce the following definition.

Definition 2 (Sparse regular variation). *A random vector $\mathbf{X} \in \mathbb{R}_+^d$ is sparsely regularly varying if there exist a random vector \mathbf{Z} defined on the simplex \mathbb{S}_+^{d-1} and a nondegenerate random variable Y such that*

$$\mathbb{P}\left(\left(\frac{|\mathbf{X}|}{t}, \pi\left(\frac{\mathbf{X}}{t}\right)\right) \in \cdot \mid |\mathbf{X}| > t\right) \xrightarrow{w} \mathbb{P}((Y, \mathbf{Z}) \in \cdot), \quad t \rightarrow \infty. \quad (3.8)$$

In this case, the general theory of regular variation states that there exists $\alpha > 0$ such that Y is Pareto(α)-distributed. In Section 3.1 we established that regular variation with limit (Y, Θ) implies sparse regular variation with limit $(Y, \pi(Y\Theta))$. Our aim is now to prove that under mild assumption the converse implication also holds. We consider a sparsely regularly varying random vector \mathbf{X} and we define the function $G_{\mathbf{Z}}$ by $G_{\mathbf{Z}}(\mathbf{x}) = \mathbb{P}(\mathbf{Z} > \mathbf{x})$ for $\mathbf{x} \in \mathcal{B}(0, 1) \cap \mathbb{R}_+^d$.

Our first result concerns the dependence between the radial limit Y and the angular limit \mathbf{Z} in (3.8). This is established in the following proposition.

Proposition 2. *Let \mathbf{X} be a sparsely regularly varying random vector on \mathbb{R}_+^d . Then, for all $r \geq 1$,*

$$\mathbf{Z} \mid Y > r \stackrel{d}{=} \pi(r\mathbf{Z}). \quad (3.9)$$

As already mentioned in Section 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 (3.9) and will be helpful in the proof of Theorem 1.

Our aim is now to prove that under some assumptions on $G_{\mathbf{Z}}$ sparse regular variation implies regular variation. Note that if convergence (3.8) holds, then the radial component $|\mathbf{X}|$ 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 $t \rightarrow \infty$. The 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

$$\lim_{\epsilon \rightarrow 0} \liminf_{t \rightarrow \infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1 + \epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \mid |\mathbf{X}| > t\right) = l(A), \quad (3.10)$$

and

$$\lim_{\epsilon \rightarrow 0} \limsup_{t \rightarrow \infty} \epsilon^{-1} \mathbb{P}\left(\frac{|\mathbf{X}|}{t} \in (1, 1 + \epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \mid |\mathbf{X}| > t\right) = l(A), \quad (3.11)$$

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 consider the family $\mathcal{A} = \{A_{\mathbf{x}}, \mathbf{x} \in \mathcal{B}(0, 1) \cap \mathbb{R}_+^d\}$, where the $A_{\mathbf{x}}$ are defined in (3.4).

Remark 8. In Lemma 3, $|\cdot|$ denotes any norm of \mathbb{R}^d but in what follows we apply this lemma with the ℓ^1 -norm.

Our aim is now to prove that a random vector \mathbf{X} which satisfies (3.8) is regularly varying. This requires some assumptions on the function $G_{\mathbf{Z}}$. We denote by λ the Lebesgue measure on the positive unit sphere $\mathcal{B}(0, 1) \cap \mathbb{R}_+^d$ and we consider the following assumptions on $G_{\mathbf{Z}}$:

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

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

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 simply write \mathcal{Z} for $\mathcal{Z}(G_{\mathbf{Z}})$ and $\mathcal{A}_{\mathcal{Z}}$ for $\mathcal{A}_{\mathcal{Z}(G_{\mathbf{Z}})}$.

Theorem 1. *Let \mathbf{X} be a sparsely regularly varying random vector on \mathbb{R}_+^d with angular limit \mathbf{Z} . 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*

$$\mathbb{P}(\Theta \in A_{\mathbf{x}}) = \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + \alpha^{-1} dG_{\mathbf{Z}}(\mathbf{x})(\mathbf{x} - 1/d), \quad (3.12)$$

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

Theorem 1 states under mild assumptions the equivalence between regular variation and sparse regular variation. Moreover the distribution of \mathbf{Z} completely characterizes the one of Θ . Equation (3.12) completes the result (3.5) obtained in Proposition 1.

Let us summarize the results we obtained. Proposition 1 characterizes the distribution of $\mathbf{Z} = \pi(Y\Theta)$ when \mathbf{X} is regularly varying with spectral vector Θ . Conversely if \mathbf{X} is a sparsely regularly varying random vector, then Theorem 1 states that \mathbf{X} is regularly varying with a spectral vector Θ which satisfies Equation (3.12). 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.

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 $\mathbb{P}(\exists 1 \leq i \leq d, \Theta_i = 1) = 1$.

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:

$$\sum_{\beta \in \mathcal{P}_d^*} p(\beta) \delta_{\mathbf{e}(\beta)/\#\beta}, \tag{3.13}$$

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 (3.13) is stable after multiplying by a positive random variable and projecting onto the simplex with π . Hence, if Θ has a distribution of type (3.13), then $\mathbf{Z} = \Theta$ a.s. This shows that (3.13) 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 Sparsity in multivariate extremes

In Section 3 we provided an alternative notion of regular variation which coincide under mild assumptions to the standard one. In this context our aim is to study the tail dependence of large events. We

explained in Section 2.3 that the largest directions of a random vector \mathbf{X} can not be easily identified via an estimation of the spectral measure's support. The aim of this section is to provide several results which prove that the vector \mathbf{Z} is more likely to be sparse and allows to capture empirically the extremal directions of \mathbf{X} .

4.1 Sparsity structure of \mathbf{Z}

We start by focusing on the behavior of $\pi(\mathbf{X}/t)$ on the subsets C_β defined in (2.5). These subspaces are used in order to identify groups of directions which may take large values simultaneously while the other directions are small (Goix et al. (2016), Goix et al. (2017)). In order to obtain dimension reduction, we can also relax in the definition of the C_β 's the condition that all directions in β are positive. Then we introduce the subsets $\text{Vect}(\mathbf{e}_j, j \in \beta)$ which are used to identify groups of directions which are small simultaneously. Identifying such a subset allows to reduce the study of extremes on its complementary. Another idea is to keep the largest directions and to relax the condition that the coordinates in β^c are equal to zero (see Chiapino and Sabourin (2016) and Chiapino et al. (2019)).

Regarding the behavior of $\pi(\mathbf{X}/t)$, convergence (3.2) holds for any pair of Borel sets $A \times B \in (1, \infty) \times \mathbb{S}_+^{d-1}$ such that $\mathbb{P}(Y\Theta \in \partial\pi^{-1}(B)) = 0$. Next proposition states that the subsets C_β and $\text{Vect}(\mathbf{e}_j, j \in \beta)^c$ satisfy this condition.

Proposition 3. *Let \mathbf{X} be a sparsely regularly varying random vector in \mathbb{R}_+^d with radial limit Y and angular limit \mathbf{Z} . Then for any $\beta \in \mathcal{P}_d^*$ the following convergences hold:*

$$\mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta \mid |\mathbf{X}| > t) \rightarrow \mathbb{P}(\mathbf{Z} \in C_\beta), \quad t \rightarrow \infty, \quad (4.1)$$

$$\mathbb{P}(\pi(\mathbf{X}/t)_{\beta^c} = 0 \mid |\mathbf{X}| > t) \rightarrow \mathbb{P}(\mathbf{Z}_{\beta^c} = 0), \quad t \rightarrow \infty. \quad (4.2)$$

Both convergences imply that the sparsity of \mathbf{Z} can be studied through the projected vector $\pi(\mathbf{X}/t)$. Recall that for $\beta \neq \{1, \dots, d\}$ the convergences (4.1) and (4.2) do not hold if we replace \mathbf{Z} by Θ and $\pi(\mathbf{X}/t)$ by $\mathbf{X}/|\mathbf{X}|$. Proposition 3 is helpful in a statistical setting since it allows us to estimate the sparse behavior of \mathbf{Z} based on the one of \mathbf{X} . This will be illustrated in Section 5.

An interpretation of \mathbf{Z} Proposition 3 provides a link between the angular vector \mathbf{Z} and the projected vector $\pi(\mathbf{X}/t)$ on the two types of subsets C_β and $\text{Vect}(\mathbf{e}_j, j \in \beta)^c$. Based on this proposition and on some properties of the Euclidean projection we are able to better highlight the role of \mathbf{Z} regarding the extreme values of \mathbf{X} .

A first property deals with the extremality of a component with respect to the others. For $j = 1, \dots, d$, we apply Equation (2.10) 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 3 to the subset $C_{\{j\}}$ leads to the convergence

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

This means that for t "high enough" we have the approximation

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

In other words the vector \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 (2.9) 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 in Section 2.4). All in all, we obtain the equivalence

$$\pi(\mathbf{X}/t)_{\beta^c} = 0 \iff r^{-1} \sum_{k \in \beta} X_k \geq \max_{i \in \beta^c} X_i + t. \quad (4.3)$$

Then, following Proposition 3, 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 means that the vector \mathbf{Z} does not concentrate on the directions $j \in \beta^c$ if the average value of the marginals X_j for $j \in \beta$ is much larger than all the components of \mathbf{X} on β^c . Thus, regarding the extreme values of \mathbf{X} , the vector \mathbf{Z} highlights the relative importance of the marginals in β with respect to the ones in β^c .

Our aim is now to compare the limit probabilities in Proposition 3 which involve the angular limit \mathbf{Z} with the associated ones involving the spectral vector Θ .

Proposition 4. *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*

$$\mathbb{P}(\mathbf{Z}_{\beta^c} = 0) = \mathbb{E}\left[\min_{j \in \beta^c} \left(\sum_{k=1}^d (\Theta_k - \Theta_j)_+\right)^\alpha\right], \quad (4.4)$$

and

$$\mathbb{P}(\mathbf{Z} \in C_\beta) = \mathbb{E}\left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+\right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+\right)^\alpha\right)_+\right]. \quad (4.5)$$

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 (3.6). 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 4 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

$$\mathbb{P}(Z_{j_0} = 1) = \mathbb{E}\left[\min_{j \neq j_0} (\Theta_{j_0} - \Theta_j)_+^\alpha\right]. \quad (4.6)$$

Then, Equation (4.6) can be developed in the following way:

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

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

Remark 9. Following Equation (4.4), we write

$$\mathbb{P}(\mathbf{Z}_{\beta^c} = 0) \geq \mathbb{E}\left[\min_{j \in \beta^c} \left(\sum_{k=1}^d (\Theta_k - \Theta_j)_+\right)^\alpha \mathbf{1}_{\{\Theta_{\beta^c}=0\}}\right] = \mathbb{E}\left[\left(\sum_{k=1}^d \Theta_k\right)^\alpha \mathbf{1}_{\{\Theta_{\beta^c}=0\}}\right] = \mathbb{P}(\Theta_{\beta^c} = 0). \quad (4.7)$$

This can also be seen as a direct consequence of Property P2, see Subsection 2.4. This property also gives

$$\mathbb{P}(\mathbf{Z}_\beta > 0) \leq \mathbb{P}(\Theta_\beta > 0), \quad (4.8)$$

an inequality which will be useful in some proofs.

Our aim is now to compare the probabilities $\mathbb{P}(\Theta \in C_\beta)$ and $\mathbb{P}(\mathbf{Z} \in C_\beta)$ for $\beta \in \mathcal{P}_d^*$. The following result is a consequence of Proposition 4.

Proposition 5. *For $\beta \in \mathcal{P}_d^*$, if $\mathbb{P}(\Theta \in C_\beta) > 0$, then $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$.*

Proposition 5 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. However, if the overestimation is not too large, the angular vector \mathbf{Z} gives a reduced numbers of directions (regarding the total number $2^d - 1$) in which extreme events could appear. Hence, it provides a first selection among the subsets C_β 's.

Example 5. We detail here an example which shows that the converse implication of Proposition 5 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 the first axis is equal to

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

On the other hand following Lemma 2 the probability that \mathbf{Z} belongs to the first 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

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

4.2 Maximal subsets

We focus in this section on the subsets C_β . A positive value for $\mathbb{P}(\Theta \in C_\beta)$ entails that the directions in β take simultaneously large values while the ones in β^c do not (Chautru (2015), Simpson et al. (2019), Goix et al. (2017)). In this case Proposition 5 ensures that the associated probability $\mathbb{P}(\mathbf{Z} \in C_\beta)$ is also positive. Example 5 shows that the converse implication is not always true. In order to have a partial converse result, and similarly to Definition 3, we introduce the notion of maximal subset for \mathbf{Z} .

Definition 3 (Maximal subset for \mathbf{Z}). Let $\beta \in \mathcal{P}_d^*$. We say that a subset C_β is maximal for \mathbf{Z} if

$$\mathbb{P}(\mathbf{Z} \in C_\beta) > 0 \quad \text{and} \quad \mathbb{P}(\mathbf{Z} \in C_{\beta'}) = 0, \quad \text{for all } \beta' \supsetneq \beta. \quad (4.9)$$

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

Theorem 2. For $\beta \in \mathcal{P}_d^*$, C_β is a maximal subset for Θ if and only if C_β is a maximal subset for \mathbf{Z} .

Example 5 shows it may exist $\beta \in \mathcal{P}_d^*$ such that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$ and $\mathbb{P}(\Theta \in C_\beta) = 0$. In this case, Theorem 2 states that the subset C_β is not maximal for \mathbf{Z} since it is not maximal for Θ . Following Remark 3, we consider a maximal subset γ for \mathbf{Z} such that $\beta \subset \gamma$. Then Theorem 2 states that $\mathbb{P}(\Theta \in C_\gamma) > 0$. This means that even if the direction 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.

A natural procedure to capture the extremal direction of \mathbf{X} is then the following one. Based on the Euclidean projection π we identify the subsets C_β on which the distribution of \mathbf{Z} places mass. Hopefully, the selected subsets are low-dimensional. Among these subsets we select the maximal ones which also correspond to the maximal subsets regarding the spectral vector Θ .

What happens on non-maximal subsets? While the study of maximal subsets is the same for \mathbf{Z} and Θ , we develop here some ideas which highlight the use of \mathbf{Z} regarding non-maximal subsets. We consider a direction $\beta \in \mathcal{P}_d^*$ and the associated subset C_β , and we assume that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$ and $\mathbb{P}(\Theta \in C_\beta) = 0$. Then, the subsets C_β is necessary non-maximal for \mathbf{Z} . It satisfies the following inequalities:

$$0 < \mathbb{P}(\mathbf{Z} \in C_\beta) = \mathbb{P}(\mathbf{Z}_\beta > 0, \mathbf{Z}_{\beta^c} = 0) \leq \mathbb{P}(\mathbf{Z}_{\beta^c} = 0) = \mathbb{P}(\pi(Y\Theta)_{\beta^c} = 0).$$

Following Equation (4.3), we set $r = \#\beta$ and obtain that

$$0 < \mathbb{P}(\pi(Y\Theta)_{\beta^c} = 0) = \mathbb{P}\left(r^{-1} \sum_{k \in \beta} \Theta_k \geq \max_{i \in \beta^c} \Theta_i + Y^{-1}\right). \quad (4.10)$$

This implies that with positive probability $r^{-1} \sum_{k \in \beta} \Theta_k \geq \Theta_i$ for all $i \in \beta^c$. If we consider for instance $\beta = \{j\}$, then we obtain that with positive probability $\Theta_j \geq \Theta_i$ for all $i \neq j$. More generally, regarding the vector \mathbf{X} , Equation (4.10) yields to

$$0 < \mathbb{P}\left(r^{-1} \sum_{k \in \beta} \Theta_k \geq \max_{i \in \beta^c} \Theta_i + Y^{-1}\right) = \lim_{t \rightarrow \infty} \mathbb{P}\left(r^{-1} \sum_{k \in \beta} X_k \geq \max_{i \in \beta^c} X_i + t \mid |\mathbf{X}| > t\right).$$

This means that the extreme values of \mathbf{X}_β are likely to be larger than the extreme ones of \mathbf{X}_{β^c} . This does not contradict the fact that $\mathbb{P}(\Theta \in C_\beta) = 0$ which only implies that it is unlikely to observe simultaneously large values in the direction β and small values in the direction β^c .

Hence, if we detect a maximal subset C_γ , we first infer that it is likely that the directions in γ are large together while the ones in γ^c take small values. The marginals in γ form a cluster of extremal directions for which the relative importance of each direction can be studied via the identification of non-maximal subsets $C_\beta \subsetneq C_\gamma$. Indeed, if we detect such a subset, it means that in the cluster γ the directions in β are likely to be larger than the ones in $\gamma \setminus \beta$. A more deeper interpretation of non-maximal subsets is deferred to future work.

Example 6. We develop the device of Example 3. We consider a regularly random variable X in \mathbb{R}_+ with tail index $\alpha > 0$ and a vector $\mathbf{a} \in \mathbb{R}_+^d$. We assume that the coordinates of \mathbf{a} satisfy the inequality $a_1 > a_2 > \dots > a_d > 0$ and we also assume for simplicity that $\mathbf{a} \in \mathbb{S}_+^{d-1}$. We define the

vector \mathbf{X} by setting $\mathbf{X} = X\mathbf{a} = (a_1X, \dots, a_dX) \in \mathbb{R}_+^d$. Then, the vector \mathbf{X} is regularly varying with tail index α and a spectral vector given by $\Theta = \mathbf{a}$ a.s. This means that the subsets $C_{\{1, \dots, d\}}$ is the only one on which the spectral measure places mass, and it is a maximal one. Hence, the angular vector \mathbf{Z} satisfies $\mathbb{P}(\mathbf{Z} \in C_{\{1, \dots, d\}}) > 0$. However, it is possible that the distribution of \mathbf{Z} also puts mass on lower-dimensional subsets. Since the Euclidean projection keeps the order of the marginals, the only possible groups of directions are $\{1\}, \{1, 2\}, \{1, 2, 3\}, \dots, \{1, \dots, d-1\}$.

We first consider the direction $\{1\}$ and compute the probability that \mathbf{Z} belongs to the subset $C_{\{1\}}$. Following Equation (2.9), we obtain that

$$\mathbb{P}(\mathbf{Z} \in C_{\{1\}}) = \mathbb{P}\left(\min_{j \leq 2} (Y\Theta_1 - Y\Theta_j) \geq 1\right) = \mathbb{P}\left(Y\Theta_1 \geq \max_{j \leq 2} Y\Theta_j + 1\right).$$

Then, we use the relation $\Theta = \mathbf{a}$ a.s. which entails that

$$\mathbb{P}(\mathbf{Z} \in C_{\{1\}}) = \mathbb{P}(Ya_1 \geq Ya_2 + 1) = \mathbb{P}(Y \geq (a_1 - a_2)^{-1}) = (a_1 - a_2)^\alpha.$$

Hence, the probability that \mathbf{Z} belongs to the first axis depends on the difference between the first and the second coordinate of \mathbf{a} .

More generally, for $1 \leq r \leq d-1$, Equation (2.10) implies that \mathbf{Z} belongs to the subset $C_{\{1, \dots, r\}}$ if and only if

$$\max_{1 \leq i \leq r} \sum_{j=1}^r (Y\Theta_j - Y\Theta_i) < 1, \quad \text{and} \quad \min_{r+1 \leq i \leq d} \sum_{j=1}^r (Y\Theta_j - Y\Theta_i) \geq 1.$$

Thus, the probability that \mathbf{Z} belongs to the subset $C_{\{1, \dots, r\}}$ is equal to

$$\begin{aligned} \mathbb{P}(\mathbf{Z} \in C_{\{1, \dots, r\}}) &= \mathbb{P}\left(\sum_{j=1}^r (Ya_j - Ya_r) < 1, \sum_{j=1}^r (Y\Theta_j - Y\Theta_{r+1}) \geq 1\right) \\ &= \mathbb{P}\left((\tilde{a}_r - ra_{r+1})^{-1} \leq Y < (\tilde{a}_r - a_r)^{-1}\right) \quad \text{where } \tilde{a}_r = a_1 + \dots + a_r \\ &= (\tilde{a}_r - ra_{r+1})^\alpha - (\tilde{a}_r - a_r)^\alpha. \end{aligned}$$

If we take $\alpha = 1$ for the sake of simplicity, we obtain that $\mathbb{P}(\mathbf{Z} \in C_{\{1, \dots, r\}}) = r(a_r - a_{r+1})$ and thus the probability that \mathbf{Z} belongs to the subset $C_{\{1, \dots, r\}}$ depends only on the distance between a_r and a_{r+1} .

This example emphasizes the use of the vector \mathbf{Z} on non-maximal subsets. It highlights the relative importance of a coordinate regarding the extreme values of a group of directions this coordinate belongs to.

5 Numerical results

This section is devoted to a statistical illustration of sparse regular variation. We highlight how our approach manages to detect the tail dependence's sparsity. 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. \(2017\)](#). We develop then several numerical results and compare both methods.

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 directions $\beta \in \mathcal{P}_d^*$ in which the extreme values of \mathbf{X} occur. These directions are characterized by the fact that $\mathbb{P}(\Theta \in C_\beta) > 0$ and therefore, by

Proposition 5, $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$. Thanks to Proposition 3 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

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

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 $T_\beta(t)$, the more likely the direction β gathers extreme values.

The Law of Large Numbers ensures that

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

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

$$T_\beta(t) \approx \frac{\mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta, |\mathbf{X}| > t)}{\mathbb{P}(|\mathbf{X}| > t)} \approx \mathbb{P}(\mathbf{Z} \in C_\beta),$$

where the first approximation holds for n large and the second one for t large. This approximation enables to classify the subsets C_β depending on the nullity or not of the associated quantity $T_\beta(t)$. Actually once t is fixed we can get rid of the denominator in (5.2) and only focus on $\sum_{j=1}^n \mathbb{1}\{\pi(\mathbf{X}_j/t) \in C_\beta, |\mathbf{X}_j| > t\}$ for $\beta \in \mathcal{P}_d^*$. There, the selection of the largest vectors \mathbf{X}_j whose norm is above t boils down to keep only a proportion, say k , of vectors. It is customary in EVT to choose a level k which satisfies $k \rightarrow \infty$ and $k/n \rightarrow 0$ when $n \rightarrow \infty$.

The approach proposed by Goix et al. (2017) In order to detect anomalies among multivariate extremes, Goix et al. (2017) propose a similar approach based on 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 6). Starting from the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ with generic random vector $\mathbf{X} = (X^1, \dots, X^d)$ with joint (resp. marginal) distribution \mathbf{F} (resp. F_1, \dots, F_d), the authors define the vectors $\mathbf{V}_i = (1/(1 - \hat{F}_j(X_i^j)))_{j=1, \dots, d}$ for $i = 1, \dots, n$, where

$$\hat{F}_j : x \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i^j < x} \quad (5.4)$$

is the empirical counterpart of F_j . This rank transformation provides standardized marginals to the vectors \mathbf{V}_i . It corresponds to the empirical counterpart of the transformation $x \mapsto 1/(1 - F_j(x))$ which transforms the marginals into Pareto(1) random variables. This means that they focus on the nonstandard case (see Section 2.2).

In order to go back to the ℓ^∞ positive unit sphere $\mathbb{S}_{+, \infty}^{d-1}$, we define

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

and we denote by Θ_∞ the spectral vector with respect to the ℓ^∞ -norm. Then, Goix et al. (2017) (Lemma 1) establish the convergence

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

with $c = \mu([\mathbf{0}, \mathbf{1}]^c) > 0$. Besides, the authors provide an empirical version of μ by defining

$$\mu_n(A) = \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{\mathbf{v}_i \in (n/k)A},$$

where $k = k(n) > 0$ is such that $k \rightarrow \infty$ and $k/n \rightarrow 0$ when $n \rightarrow \infty$ (the ratio n/k playing the role of the large threshold t). All in all, the estimation of the quantity $\mathbb{P}(\Theta_\infty \in C_{\beta,\infty})$ relies on the following sequence of approximation:

$$T_\beta(k, \epsilon) := \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{\mathbf{v}_i \in (n/k)R_\beta^\epsilon} \approx \mu(R_\beta^\epsilon) \approx c \mathbb{P}(\Theta_\infty \in C_{\beta,\infty}),$$

where the first approximation holds for k large, and the second one for ϵ close to zero. All these considerations lead to an algorithm, called DAMEX, introduced in [Goix et al. \(2017\)](#), Section 4.2, whose goal is to identify the subsets C_β such that $\mathbb{P}(\Theta_\infty \in C_{\beta,\infty}) > 0$.

Remark 10 (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 related 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 if the spectral vectors Θ and Θ_∞ correspond to the nonstandard case (with $\alpha = 1$), then they satisfy the relation

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

for all $B \in \mathbb{S}_+^{d-1}$. This entails in particular that

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

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

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

Hence, the directions in which extremes gather are the same regardless the choice of the norm.

Remark 11. At the end of the procedure we obtain a group of directions β such that $T_\beta(t) > 0$. Since we deal with non-asymptotic data many $T_\beta(t)$, we obtain a bias which provides a difference between some directions β for which $T_\beta(t)$ takes small values and 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 $T_\beta(t)$ are set to 0. We use a threshold of the form $p/\#\mathcal{C}$, where $\mathcal{C} = \{\beta, T_\beta(t) > 0\}$ and where the hyperparameter $p \geq 0$ is fixed by the user. It is of course possible to set p to 0 which boils down to selecting all the subsets C_β such that $T_\beta(t) > 0$. 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 $T_\beta(t);$

Set to 0 the $T_\beta(t)$ below the threshold discussed in Remark 11;

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

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

5.2 Experimental results and comparison with DAMEX

In this section we consider two different cases of numerical data on which we apply Algorithm 2. For each case we generate data sets of size $n \in \{10^4, 5 \cdot 10^4, 10^5\}$, we compute the quantities $T_\beta(t)$ and we repeat this procedure over $N = 100$ simulations. Regarding the outcome $\mathcal{C} = \{C_\beta, T_\beta(t) > 0\}$ of our procedure, two different types of errors could arise. The first one corresponds to the occurrence of a direction β while it should not appear theoretically. This error will be called error of type 1 or simply T1. The second type of error corresponds to the absence of a direction β while it should appear theoretically. This error will be called error of type 2 or simply T2. The results correspond to the average number of each error among the N simulations. The code can be found at https://github.com/meyernicolas/projection_extremes.

The purpose of the experiments is twofold. The first idea is to study the procedure given in Algorithm 2 and to see how it manages to detect the sparsity of the extremal data. The second aim of these simulations is to compare our method to the DAMEX algorithm. Therefore, we also compute the quantities $T_\beta(k, \epsilon)$ for $\beta \in \mathcal{P}_d^*$ as well as the two types of errors for this algorithm. The DAMEX is based on a hyperparameter ϵ for which there exists no natural choice, the results show. Indeed it may happen that for a fixed simulation study there exists a specific hyperparameter ϵ_0 for which the DAMEX 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 12 (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 large scale of comparison we consider $\epsilon \in \{0.05, 0.1, 0.5\}$. 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 an iid sequence of random vectors $\mathbf{N}_1, \dots, \mathbf{N}_n$ in \mathbb{R}^{40} with generic random vector \mathbf{N} whose distribution is a multivariate Gaussian distribution with all univariate marginals equal to $\mathcal{N}(0, 1)$ and the correlations less than 1: $\mathbb{E}[N^i N^j] < 1$ for all $1 \leq i \neq j \leq d$. We transform the marginals with a rank transform which consists in considering the vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ such that the marginals X_i^j of $\mathbf{X}_i = (X_i^1, \dots, X_i^d)$ are defined as

$$X_i^j = \frac{1}{1 - \hat{F}_j(N_i^j)}, \quad 1 \leq j \leq d,$$

where \hat{F}_j is the empirical version of the cumulative distribution function of $N_j \sim \mathcal{N}(0, 1)$, see (5.4). This provides a sample of regularly varying random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ and the assumption on the correlation leads to asymptotic independence (see Sibuya (1960)). This case has been discussed in Subsection 3.4. Equivalently, it means that $\mathbb{P}(\Theta \in C_\beta) = \mathbb{P}(\mathbf{Z} \in C_\beta) = 1/d$ for all β such that $\#\beta = 1$ (and therefore $\mathbb{P}(\Theta \in C_\beta) = \mathbb{P}(\mathbf{Z} \in C_\beta) = 0$ elsewhere). The aim of our procedure is then to recover these 40 directions among the $2^{40} - 1 \approx 10^{12}$ subsets C_β .

Regarding the multivariate Gaussian random vectors $\mathbf{N}_1, \dots, \mathbf{N}_n$, the simulation of these vectors depends only on their covariance matrix. We proceed as follows. We generate a matrix Σ' with entries $\sigma'_{i,j}$ following independent uniform distributions on $(-1, 1)$. Then, we define the matrix Σ as

$$\Sigma := \text{Diag}(\sigma'_{1,1}^{-1/2}, \dots, \sigma'_{d,d}^{-1/2}) \cdot \Sigma' \cdot \text{Diag}(\sigma'_{1,1}^{-1/2}, \dots, \sigma'_{d,d}^{-1/2}),$$

where $\text{Diag}(\sigma'_{1,1}^{-1/2}, \dots, \sigma'_{d,d}^{-1/2})$ denotes the diagonal matrix of $\mathcal{M}_d(\mathbb{R})$ whose diagonal is given by the vector $(\sigma'_{1,1}^{-1/2}, \dots, \sigma'_{d,d}^{-1/2})$. This provides a covariance matrix with diagonal entries equal to 1 and off-diagonal entries less than 1. A given matrix Σ provides then a dependence structure for $\mathbf{N}_1, \dots, \mathbf{N}_n$ and

thus for $\mathbf{X}_1, \dots, \mathbf{X}_n$. We generate $N_{\text{model}} = 20$ different matrices Σ , and for each of these dependence structures we generate $N = 100$ sample $\mathbf{N}_1, \dots, \mathbf{N}_n$. We summarize in Table 1 and in Table 2 the two types of errors averaged among the $N \cdot N_{\text{model}} = 2000$ simulations.

	Euclidean projection T1	DAMEX T1 $\epsilon = 0.05$	DAMEX T1 $\epsilon = 0.1$	DAMEX T1 $\epsilon = 0.5$
$n_1 = 10^4$	22.62	3034.70	2899.05	987.63
$n_2 = 5 \cdot 10^4$	19.43	6972.52	4646.43	271.87
$n_3 = 10^5$	1.83	8401.21	4813.46	235.80

Table 1: Average number of errors of type 1 in an asymptotic independence case ($d = 40$).

	Euclidean projection T2	DAMEX T2 $\epsilon = 0.05$	DAMEX T2 $\epsilon = 0.1$	DAMEX T2 $\epsilon = 0.5$
$n_1 = 10^4$	0.07	39.43	13.76	0.00
$n_2 = 5 \cdot 10^4$	0.00	3.69	0.01	0.00
$n_3 = 10^5$	0.00	0.07	0.00	0.00

Table 2: Average number of errors of type 2 in an asymptotic independence case ($d = 40$).

For the Euclidean projection, we observe that our algorithm manages to capture almost all $d = 40$ directions regardless the value of n . The first column of Table 2 gives indeed an error which is very close (or even equal to) to 0. On the other hand, it seems that for $n = n_1 = 10^4$ and $n = n_2 = 5 \cdot 10^4$, the procedure also captures extra-subsets that should not be identified. This is maybe a consequence of the choice of p in Remark 11 which may be too high and for which a deeper study should be conducted. However, the number of errors of type 1 becomes very low for $n = n_3 = 10^5$. For this data size, we therefore obtain very accurate results since our algorithm manages to capture almost all the relevant directions and no extra ones. As expected, the angular vector \mathbf{Z} is helpful to detect asymptotic independence.

Regarding the DAMEX algorithm, a large ϵ leads theoretically to more mass assigned on the axis. This explains why in our simulations choosing a large ϵ reduces the error of type 2. With $\epsilon = 0.5$ the algorithm manages to capture all the $d = 40$ axes. However, the error of type 1 is quite large regardless the choice of ϵ or of n : The DAMEX captures a too large number of subsets. Hence, it seems that our procedure leads to the best compromise between both types of errors.

Remark 13. In this example we standardized the marginals of \mathbf{N} with a rank transform in order to obtain a regularly varying random vector \mathbf{X} . Following Remark 4, we obtain here a similar computational time for both procedures. However, the modification of the marginals is not a necessary step in our algorithm as soon as the random vector \mathbf{X} is already regularly varying. On the other hand, this step is required for the DAMEX algorithm. This difference provides then cases where the computational time of the latter increases due to this extra step. Besides, this time could also be longer since the DAMEX algorithm is based on a hyperparameter and tuning such a parameter requires to compute the algorithm for different values of ϵ .

A dependent case We now consider a dependent case where extremes occur on lower-dimensional subsets. In order to include dependence we proceed as in Example 2. We consider two independent random variables P and P' following respectively a Pareto(1) and a Pareto(2) distribution. Then, the two-dimensional vector \mathbf{P} whose marginals are $P_1 = P$ and $P_2 = P + P'$ is regularly varying with tail

index 1 and with spectral vector $\Theta_{\mathbf{P}} = (1/2, 1/2)$ a.s. Similarly, we consider three independent random variables R , R' , and R'' following respectively a Pareto(1), a Pareto(2), and a Pareto(2) distribution. Then, if we consider the three-dimensional vector \mathbf{R} whose marginals are $R_1 = R$, $R_2 = R + R'$, and $R_3 = R + R''$, then the vector \mathbf{R} is regularly varying with tail index 1 and with spectral vector $\Theta_{\mathbf{R}} = (1/3, 1/3, 1/3)$ a.s.

For our simulations we consider $s_1 = 10$ independent copies $\mathbf{P}_1, \dots, \mathbf{P}_{s_1}$ of $\mathbf{P} \in \mathbb{R}_+^2$ and $s_2 = 10$ independent copies $\mathbf{R}_1, \dots, \mathbf{R}_{s_2}$ of $\mathbf{R} \in \mathbb{R}_+^3$. We aggregate these vectors and form a vector \mathbf{X} in \mathbb{R}_+^{50} which is then regularly varying with a discrete spectral measure placing mass on the points $(\mathbf{e}_j + \mathbf{e}_{j+1})/2$ for $j = 1, 3, \dots, 17, 19$ and on the points $(\mathbf{e}_j + \mathbf{e}_{j+1} + \mathbf{e}_{j+2})/3$ for $j = 1, 4, \dots, 25, 28$. Besides, as discussed in Section 3.4, in this case the angular vector \mathbf{Z} and the spectral vector Θ are equal almost surely.

For our simulations we consider a sample of iid random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ with the same distribution as \mathbf{X} . Our aim is then to recover the $s_1 = 10$ two-dimensional subsets C_β which contain the directions $(\mathbf{e}_j + \mathbf{e}_{j+1})/2$ for $j = 1, 3, \dots, 17, 19$ and also the $s_2 = 10$ three-dimensional subsets C_β which contain the directions $(\mathbf{e}_j + \mathbf{e}_{j+1} + \mathbf{e}_{j+2})/3$ for $j = 1, 4, \dots, 25, 28$. All in all, we would like to recover these $s = s_1 + s_2 = 20$ subsets among the $2^{50} - 1 \approx 10^{15}$ subsets.

	Euclidean projection T1	DAMEX T1 $\epsilon = 0.05$	DAMEX T1 $\epsilon = 0.1$	DAMEX T1 $\epsilon = 0.5$
$n_1 = 10^4$	8.22	1868.11	1448.21	117.18
$n_2 = 5 \cdot 10^4$	0.32	2776.68	613.66	17.78
$n_3 = 10^5$	0.04	2560.76	295.86	1.15

Table 3: Average number of errors of type 1 in a dependent case ($d = 50$).

	Euclidean projection T2	DAMEX T2 $\epsilon = 0.05$	DAMEX T2 $\epsilon = 0.1$	DAMEX T2 $\epsilon = 0.5$
$n_1 = 10^4$	0.76	12.02	0.01	0.00
$n_2 = 5 \cdot 10^4$	0.04	0.00	0.00	0.00
$n_3 = 10^5$	0.03	0.00	0.00	0.00

Table 4: Average number of errors of type 2 in a dependent case ($d = 50$).

Table 3 (resp. Table 4) show the average number of errors of type 1 (resp. of type 2) among the $N = 100$ experiments. This number decreases with n regardless the type of error or the algorithm. Except for the number of error of type 1 with $n = n_1 = 10^4$, our algorithm provides small values of errors. This means that not only it is able to detect all the $s = 20$ subsets on which the distribution of \mathbf{Z} puts mass, but it also does not identify subsets which does not gather the mass of \mathbf{Z} . These results are all the more accurate since the identification of the $s = 20$ subsets is done among a very large number of subsets, indeed $2^{50} - 1 \approx 10^{15}$.

Compared to the DAMEX algorithm, we obtain lower values of errors of Type 1, which means that our procedure less overestimates the number of identified subsets. On the other hand, it is possible that our procedure miss some β 's that should appear theoretically. However, this case happens only rarely since the averaged error of type 2 is smaller than 1. Regarding this error, the DAMEX obtains perfect results for $n = n_2 = 5 \cdot 10^4$, and $n = n_3 = 10^5$. If we combine both types of errors, our algorithm provides quite accurate results and manages to capture the theoretical directions.

5.3 Sparse regular variation and non-maximal subsets

In this section we illustrate some interpretation of the vector \mathbf{Z} regarding extremal directions in non-maximal subsets. Following Example 6, we consider a vector $\mathbf{a} \in \mathbb{S}_+^{r-1}$, $r \geq 2$, such that $a_1 > \dots > a_r > 0$. If X is a random variable following a Pareto distribution with parameter $\alpha > 0$, then the vector $\mathbf{a}X = (a_1X, \dots, a_rX) \in \mathbb{R}_+^r$ is regularly varying with tail index α and spectral vector $\Theta_r = \mathbf{a}$ a.s. We combine this device with the one of Example 2 and consider independent random variables $\tilde{X}_2, \dots, \tilde{X}_r$ following a Pareto distribution with parameter $\alpha' > \alpha$. Then, the random vector

$$\tilde{\mathbf{X}} = \mathbf{a}X + (0, \tilde{X}_2, \dots, \tilde{X}_r) = (a_1X, a_2\tilde{X} + X_2, \dots, a_rX + \tilde{X}_r)$$

is regularly varying with tail index α and spectral vector $\Theta_r = \mathbf{a}$ a.s.

The spectral vector Θ_r places mass only on the subset $C_{\{1, \dots, r\}}$ which is maximal. Regarding the vector \mathbf{Z} , it is likely that it places mass on some non-maximal subsets $\{1\}, \{1, 2\}, \{1, 2, 3\}, \dots, \{1, \dots, d-1\}$, see Example 6. In our simulations, we fix $\alpha = 1$, $\alpha' = 2$, and $r = 3$, and we consider a vector $\mathbf{a} = (7, 6, 4)/|(7, 6, 4)|$. This leads to a regularly varying random vector $\tilde{\mathbf{X}}$ in \mathbb{R}_+^3 . Finally, we consider $s = 20$ iid copies $\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_s$ of the vector $\tilde{\mathbf{X}}$ and aggregate them to build the vector $\mathbf{X} \in \mathbb{R}^{60}$. The vector \mathbf{X} is then regularly varying with tail index 1 and with spectral measure placing mass on the three-dimensional maximal subsets $C_{\{j, j+1, j+2\}}$, for $j \in J := \{1, 4, 7, \dots, 58\}$. Regarding the vector \mathbf{Z} , it is likely that its distribution puts mass on the two-dimensional subsets $C_{\{j, j+1\}}$ for $j \in J$, and on the one-dimensional subsets $C_{\{j\}}$ for $j \in J$.

The aim of our simulations is twofold. The first step is to see if our procedure manages to capture the $s = 20$ maximal subsets $C_{\{j, j+1, j+2\}}$ for $j \in J$. The second purpose is to recover the 40 non-maximal subsets aforementioned. The first column of Table 5 gives the averaged number of the three-dimensional maximal subsets that have been recovered by our algorithm. Similarly, the second (resp. third) column gives the number of the two-dimensional (resp. one-dimensional) non-maximal subsets that have been identified by our algorithm. Recall that for each type of subsets, the theoretical number of these subsets is $s = 20$. Finally, the last column deals with the number of subsets that should not appear theoretically. All the results are averaged among the $N = 100$ simulations. In this case, we do not compare the results with the DAMEX. Indeed, this latter algorithm requires to standardize the marginals so that there is no partial order between them.

	Three-dimensional subsets	Two-dimensional subsets	One-dimensional subsets	Other subsets
$n_1 = 10^4$	13.16	12.28	17.92	14.96
$n_2 = 5 \cdot 10^4$	18.40	18.04	19.91	17.35
$n_3 = 10^5$	17.95	17.39	19.92	0.68

Table 5: Average number of subsets recovered by Algorithm 2 ($d = 60$).

For $n = n_1 = 10^4$ we observe that the procedure manages to identify most of the one-dimensional subsets, while the average number of two-dimensional subsets is quite smaller than the theoretical one. The same arise for the maximal subsets for which we only manage to recover two third of the theoretical ones. We also obtain a non-negligible number of extra-subsets which should not be identified. For $n = n_2 = 5 \cdot 10^4$, the three types of subsets are quite well recovered by our algorithm, with once again very good result for the one-dimensional ones. The number of extra-subsets is still relatively high. For $n = n_3 = 10^5$, we keep a high level of accuracy regarding the identification of the three types of subsets while the number of extra-subsets drastically decreases. This last case provides therefore accurate results.

This example highlights the relevance of our approach to identify clusters of directions that are simultaneously large but also to study the relative importance of the coordinates in a given cluster. This second aspect provides a more deeper interpretation of \mathbf{Z} in terms of extremes.

6 Conclusion

The notion of sparse regular variation that is introduced in this paper is a way to tackle the issues that arise in the study of tail dependence with the standard concept of regular variation. Replacing the self-normalized vector $\mathbf{X}/|\mathbf{X}|$ by the projected one $\pi(\mathbf{X}/t)$ allows us to capture the extremal directions of \mathbf{X} . Our main result is the equivalence between both concepts of regular variation under mild assumptions. The results of Proposition 1 and Theorem 1 establish a bijection between the spectral vector Θ and the angular vector \mathbf{Z} .

Regarding extremes values, the vector \mathbf{Z} enjoys many useful properties. This vector is sparser than the spectral vector Θ which entails that it seems more suitable to identify extremal directions, especially in high dimensions. Indeed, large events often appear due to a simultaneous extreme behavior of a small number of coordinates. This similarity between extreme values and the vector \mathbf{Z} appears even more with the study of the subsets C_β which highlights the tail dependence of \mathbf{X} . Proposition 3 provides a natural way to capture the behavior of \mathbf{Z} on these subsets. In other words the Euclidean projection π manages to circumvent the weak convergence's issue which arises in the standard definition of regularly varying random vectors.

Practically speaking, Section 5 illustrates the advantages of our approach for the study of large events. First, using the Euclidean projection allows to study tail dependence without introducing any hyperparameter. On the contrary, the introduction of ϵ -thickened rectangles in Goix et al. (2017) requires to identify a suitable ϵ . Hence, our procedure reduces the algorithmic complexity by avoiding running the given code for different ϵ and to compare the outcomes. Since the projection can be computed in expected linear time, the study of extreme events can then be done in reasonable time in high dimensions. More generally, the numerical results provide 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 bias issue introduced in Remark 11. Moreover, a comparison between \mathbf{Z} and Θ on non-maximal is also a crucial point to tackle. The results we obtained in Section 5.3 highlight the behavior of \mathbf{Z} on these kind of subsets and a deeper study should be conducted.

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 (2.8).

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,

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

which gives

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

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

- We now prove that $\rho_{a\pi(\mathbf{u})} \leq \rho_{a\mathbf{u}}$. Fix $j \in \{1, \dots, d\}$ 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 (2.9) 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 (2.9).

For (2.10), 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

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

\square

Proof of Proposition 1. Fix $\mathbf{x} \in \mathcal{B}(0, 1) \cap \mathbb{R}_+^d$, with $x_j \neq 1/d$ for all $j = 1, \dots, d$. We use (2.11) 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

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

□

Proof of Proposition 2. 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 (3.8) 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 (3.10) and (3.11) 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 (3.10) when $t \rightarrow \infty$, and by $l_\epsilon^-(A)$ the liminf in (3.11) 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:

$$\frac{\mathbb{P} \left(\frac{|\mathbf{X}|}{u(1 + \epsilon)^k} > t \right)}{\mathbb{P}(|\mathbf{X}| > t)} = \mathbb{P} \left(|\mathbf{X}| > tu(1 + \epsilon)^k \mid |\mathbf{X}| > t \right) \rightarrow (u(1 + \epsilon)^k)^{-\alpha}, \quad t \rightarrow \infty. \quad (7.2)$$

Besides, we know by (3.10) that

$$\liminf_{t \rightarrow \infty} \epsilon^{-1} \mathbb{P} \left(\frac{|\mathbf{X}|}{tu(1 + \epsilon)^k} \in (1, 1 + \epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \mid \frac{|\mathbf{X}|}{u(1 + \epsilon)^k} > t \right) = l_\epsilon^-(A). \quad (7.3)$$

We now gather (7.2) and (7.3) 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 (3.11) that

$$\limsup_{t \rightarrow \infty} \epsilon^{-1} \mathbb{P} \left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon], \frac{\mathbf{X}}{|\mathbf{X}|} \in A \mid \frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t \right) = l_\epsilon^+(A). \quad (7.4)$$

Thus, Equations (7.2) and (7.4) 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 1. The proof is based on Lemma 3. Firstly, note that if (3.8) holds, then $|\mathbf{X}|$ is regularly varying with tail index α . Hence, the main part of the proof is to show that convergences (3.10) and (3.11) hold for all $A = A_{\mathbf{x}}$, $\mathbf{x} \in \mathcal{Z}$, where the $A_{\mathbf{x}}$ are defined in (3.4). We divide our proof into two steps.

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

$$\pi((1+\epsilon)\mathbf{Z}) > \mathbf{x} \iff \mathbf{Z} > \frac{\mathbf{x} + \epsilon/d}{1+\epsilon}. \quad (7.5)$$

This is a consequence of Equation (2.11) 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 (3.8) 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 2 gives

$$\begin{aligned} \mathbb{P}(Y \in (1, 1+\epsilon], \mathbf{Z} \in A_{\mathbf{x}}) &= \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}} \mid Y > 1+\epsilon) \mathbb{P}(Y > 1+\epsilon) \\ &= \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) \in A_{\mathbf{x}}) (1+\epsilon)^{-\alpha} \\ &= [1 - (1+\epsilon)^{-\alpha}] \mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + [\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) - \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) \in A_{\mathbf{x}})] (1+\epsilon)^{-\alpha}. \end{aligned} \quad (7.6)$$

The first term divided by ϵ converges to $\alpha\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}})$ when $\epsilon \rightarrow 0$. We use (7.5) 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 2 and Equation (7.5) to compute $P_1(\epsilon)$ and $P_2(\epsilon)$. For $P_1(\epsilon)$, we have the following relations

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

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 (7.6) 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 $\pi(\mathbf{X}/t)_j > x_j(1+\epsilon)$, then $X_j/t - (|\mathbf{X}|/t - 1)/d > x_j(1+\epsilon)$, and therefore $X_j/t > x_j(1+\epsilon)$. Finally we obtain that $X_j/|\mathbf{X}| > x_j$. So we proved that if $\pi(\mathbf{X}/t)_j > x_j(1+\epsilon)$, then $X_j/|\mathbf{X}| > x_j$. These considerations give the following inequality:

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

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 2 and Equation (7.5) 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

Proof of Proposition 3. We only prove (4.1) (the proof of (4.2) 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, (4.1) is equivalent to

$$\mathbb{P}((|\mathbf{X}|/t, \mathbf{X}/|\mathbf{X}|) \in D_\beta \mid |\mathbf{X}| > t) \rightarrow \mathbb{P}((Y, \Theta) \in D_\beta), \quad (7.7)$$

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 (7.7) holds and then convergence (4.1) holds as well. \square

Proof of Proposition 4. 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 (4.4).

For Equation (4.5), 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 Proposition 5. The proof of Proposition 5 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*

$$\mathbb{P}(\Theta \in C_\beta) \leq \mathbb{P}\left(\max_{j \in \beta} \sum_{k \in \beta} (\Theta_k - \Theta_j)_+ < 1\right). \quad (7.8)$$

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

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

The first probability can be rephrased as follows:

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

Since $\Theta \in \mathbb{S}_+^{d-1}$, 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

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

which concludes the proof. \square

We now prove Proposition 5. We fix $\beta \in \mathcal{P}_d^*$ and assume that $\mathbb{P}(\Theta \in C_\beta) > 0$. Then, starting from Equation (4.5), 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 2. We separately prove both implications.

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

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

Following Proposition 5, we already know that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$. Besides, if $\beta' \supsetneq \beta$, then Equation (4.8) 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, Proposition 5 implies that $\mathbb{P}(\mathbf{Z} \in C_{\beta'}) > 0$, which contradicts the maximality of C_β for \mathbf{Z} .

Secondly, Equation (4.5) of Proposition 4 gives

$$\begin{aligned}
\mathbb{P}(\mathbf{Z} \in C_\beta) &= \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \right] \\
&= \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbf{1}_{\Theta \in C_\beta} \right] \\
&\quad + \mathbb{E} \left[\left(\min_{j \in \beta^c} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha - \max_{j \in \beta} \left(\sum_{k \in \beta} (\Theta_k - \Theta_j)_+ \right)^\alpha \right)_+ \mathbf{1}_{\Theta \notin C_\beta} \right] \\
&= (A) + (B).
\end{aligned} \tag{7.9}$$

The first term (A) has already been calculated in the proof of Proposition 5. 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:

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

The first expectation is equal to

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

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 Proposition 5, 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 (7.9), 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

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., and Teugels, J. L. (2006). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd., Chichester.
- 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.
- Chiapino, M., Sabourin, A., and Segers, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes*, 22(2):193–222.
- 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.
- Engelke, S. and Ivanovs, J. (2020). Sparse structures for multivariate extremes. *arXiv:2004.12182*.
- Goix, N., Sabourin, A., and Cléménçon, S. (2016). Sparse representation of multivariate extremes with applications to anomaly ranking. In *AISTATS*, pages 75–83.
- 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. (1986). Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18(1):66–138.
- 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.
- Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.
- Simpson, E., Wadsworth, J., and Tawn, J. (2019). Determining the dependence structure of multivariate extremes. *arXiv:1809.01606*.