HAL open science

Sparse regular variation

Nicolas Meyer, Olivier Wintenberger

To cite this version:

Nicolas Meyer, Olivier Wintenberger. Sparse regular variation. Advances in Applied Probability, 2021, 53 (4), pp.1115 - 1148. 10.1017/apr.2021.14 . hal-02168495v5

HAL Id: hal-02168495

https://hal.science/hal-02168495v5

Submitted on 18 Feb 2021

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

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

Sparse regular variation

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

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

February 18, 2021

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 has often a sparse support 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 prove that under mild assumptions sparse regular variation and regular variation are two equivalent notions and we establish several results for sparsely regularly varying random vectors. Finally, we illustrate on numerical examples how this new concept allows one to detect extremal directions.

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), Resnick (2007), or Embrechts et al. (2013). Alternatively, it can also be defined via 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). Following this approach, 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

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

for all $x > 0$ and for all continuity set B of S. This means that 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 $\mathbf{X}$. It puts mass in

meyer@math.ku.dk (corresponding author)

†*olivier.wintenberger@upmc.fr*

a direction of $\mathbb{S}_+^{d-1}$ if and only if extreme events appear in this direction. Note that the choice of the norm in (1.1) is arbitrary. In this paper, $|\cdot|$ will always denote the ℓ^1-norm, for reasons explained later.

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). Some useful representations of the spectral measure has been introduced in the bivariate case (Einmahl et al. (1993), Einmahl et al. (1997), Einmahl et al. (2001) and Einmahl and Segers (2009)) and in moderate dimensions (Coles and Tawn (1991) and Sabourin et al. (2013)). Inference on the spectral measure has also been studied in a Bayesian framework, for instance by Guillotte et al. (2011). In higher dimensions, mixtures of Dirichlet distributions are often used to model the spectral densities (Boldi and Davison (2007), Sabourin and Naveau (2014)). Some alternative approaches based on grid estimators (Lehtomaa and Resnick (2020)) and Principal Component Analysis (Cooley and Thibaud (2019), Sabourin and Drees (2019)) have also been recently proposed.

More recently, the study of the spectral measure's support has become an active topic of research. In high dimensions, it is likely that this measure only places mass on low-dimensional subspaces. The measure is then said to be *sparse*. Sparsity arises all the more for standard regular variation. There, it is possible that the marginals of $\mathbf{X}$ are not tail equivalent and therefore 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. Then, identifying low-dimensional subspaces on which the spectral measure puts mass allows one to capture clusters of directions which are likely to be extreme together (Chautru (2015), Janßen and Wan (2020)). However, for such subspaces the convergence in (1.1) often fails for topological reasons which makes the identification of these subsets challenging.

Several algorithms have been recently proposed to identify the extremal directions of $\mathbf{X}$. 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. The authors provide an algorithm called DAMEX (for Detecting Anomalies among Multivariate EXtremes) of complexity $O(dn \log n)$, where n corresponds to the number of data points. Subsequently, Chiapino and Sabourin (2016) propose 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 $\kappa_{\min}$. 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. (2020) who base their method on hidden regular variation.

Since the self-normalized vector $\mathbf{X}/|\mathbf{X}|$ fails to capture the support of a sparse spectral measure S, we replace it by another angular component based on the Euclidean projection onto the simplex $\mathbb{S}_+^{d-1} = \{\mathbf{x} \in \mathbb{R}_+^d : x_1 + \ldots + x_d = 1\}$. This projection has been widely studied in learning theory (see e.g. Duchi et al. (2008), Liu and Ye (2009), or Kyrillidis et al. (2013)). 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. With this approach we obtain a new angular limit vector $\mathbf{Z}$ whose distribution slightly differs from the spectral measure and which is more likely to be sparse. We prove that under mild conditions both concepts of regular variation are equivalent and we give the relation between both limit vectors. Besides, we study this new angular limit and show that it allows one to capture the extremal directions of $\mathbf{X}$. The numerical results we provide emphasize the efficiency of our method to detect directions 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 a cluster of extremal directions.

Outline The structure of this paper is as follows. We introduce in Section 2 the concept of sparse regular variation based on the Euclidean projection onto the simplex. This notion gives rise to a new

angular vector, for which we study its distribution and compare it with the spectral measure. In the main theorem of this section we establish the equivalence under mild conditions between sparse regular variation and standard regular variation. In Section 3 we discuss to what extent the Euclidean projection allows us to better capture the extremal directions of $\mathbf{X}$. We use a natural partition of the simplex to address this issue and we prove that $\mathbf{\Theta}$ and $\mathbf{Z}$ similarly behave on so-called maximal directions. Finally we illustrate in Section 4 the performance of our method on simulated data and briefly discuss it with the approach of Goix et al. (2017).

Notation We introduce some standard notation that is used throughout the paper. Symbols in bold such as $\mathbf{x} \in \mathbb{R}^d$ are vectors with components denoted by x_j, $j \in \{1, \dots, d\}$. Operations and relationships involving such vectors are meant componentwise. We define $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^d$, $\mathbb{R}^d_+ = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x} \geq \mathbf{0}\}$, and $B^d_+(0,1) = \{\mathbf{x} \in \mathbb{R}^d_+ : x_1 + \ldots + x_d \leq 1\}$. 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 $\beta = \{i_1, \dots, i_r\} \subset \{1, \dots, d\}$, then $\mathbf{x}_\beta$ denotes the vector $(x_{i_1}, ..., x_{i_r})$ of $\mathbb{R}^r$. For $p \in [1, \infty]$, we denote by $|\cdot|_p$ the ℓ^p-norm in $\mathbb{R}^d$, except for $p = 1$ for which we just write $|\cdot|$. We write $\overset{w}{\to}$ 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. Finally, if F is a subset of a set E, we denote by F^c the complementary set of F in E.

2 Sparse regular variation

2.1 From standard to sparse regular variation

We start from Equation (1.1) which we rephrase as follows:

$$\mathbb{P}\big((|\mathbf{X}|/t, \mathbf{X}/|\mathbf{X}|) \in \cdot \mid |\mathbf{X}| > t\big) \overset{w}{\to} \mathbb{P}((Y, \mathbf{\Theta}) \in \cdot)\,, \quad t \to \infty\,, \tag{2.1}$$

where Y is a Pareto(α)-distributed random variable independent of $\mathbf{\Theta}$. We call the random vector $\mathbf{\Theta}$ the *spectral vector*, its distribution being the spectral measure. If the convergence (2.1) holds we write $\mathbf{X} \in \mathrm{RV}(\alpha, \mathbf{\Theta})$. In many cases the spectral measure is sparse, that is, it places mass on some lower-dimensional subspaces. The self-normalized vector $\mathbf{X}/|\mathbf{X}|$ then fails to estimate the spectral vector $\mathbf{\Theta}$ on such subsets.

Remark 1. The notion of sparsity in EVT can be defined in two different ways. The first one concerns the number of subsets $\{\mathbf{x} \in \mathbb{R}^d_+ : \mathbf{x}_\beta > \mathbf{0} \text{ and } \mathbf{x}_{\beta^c} = \mathbf{0}\}$, $\beta \in \mathcal{P}^*_d$, which gather the mass of the spectral measure (see Section 3 for some insights on these subsets). "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 of null coordinates in the spectral vector $\mathbf{\Theta}$. In this case, "sparse" means that with high probability $|\mathbf{\Theta}|_0 \ll d$, where $|\cdot|_0$ denotes the ℓ^0-norm of $\mathbf{\Theta}$, that is, $|\mathbf{\Theta}|_0 = |\{i = 1, \dots, d : \theta_i \neq 0\}|$. This is denoted by (S2.b) in Engelke and Ivanovs (2020). In all this article we refer to this second notion.

In order to better capture this sparsity, we replace the quantity $\mathbf{X}/|\mathbf{X}|$ by the vector $\pi_1(\mathbf{X}/t)$, where $\pi_z(\mathbf{v})$ denotes the Euclidean projection of the vector $\mathbf{v}$ in $\mathbb{R}^d_+$ onto the positive sphere $\mathbb{S}^{d-1}_+(z) := \{\mathbf{x} \in \mathbb{R}^d_+ : x_1 + \ldots + x_d = z\}$, $z > 0$. The projected vector $\pi_z(\mathbf{v})$ is defined as the unique solution of the following optimization problem (see Duchi et al. (2008) and the references therein):

$$\underset{\mathbf{w}}{\text{minimize}}\ \frac{1}{2}|\mathbf{w} - \mathbf{v}|_2^2 \quad \text{s.t.} \quad |\mathbf{w}|_1 = z\,. \tag{2.2}$$

The *Euclidean projection onto the positive sphere* $\mathbb{S}_+^{d-1}(z)$ is then defined as the application

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

where $\lambda_{\mathbf{v},z}$ is the unique constant satisfying the relation $\sum_{i=1}^d (v_i - \lambda_{\mathbf{v},z})_+ = z$. Several algorithms which compute $\pi_z(\mathbf{v})$ have been introduced (Duchi et al. (2008), Condat (2016)). We present two of them in Appendix A, a first one which gives an intuitive way to compute $\pi_z(\mathbf{v})$ and a second one based on a median-search procedure whose expected time complexity is $O(d)$.

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 write π for π_1. An illustration of π for $d = 2$ is given in Figure 1. We list below some straightforward results satisfied by the projection.

P1. The projection preserves the order of the coordinates: If $v_{\sigma(1)} \geq \ldots \geq v_{\sigma(d)}$ for a permutation σ, then $\pi(\mathbf{v})_{\sigma(1)} \geq \ldots \geq \pi(\mathbf{v})_{\sigma(d)}$ for the same permutation.

P2. If $\pi(\mathbf{v})_j > 0$, then $v_j > 0$. Equivalently, $v_j = 0$ implies $\pi(\mathbf{v})_j = 0$.

P3. The projection π is continuous, as every projection on a convex, closed set in a Hilbert space.

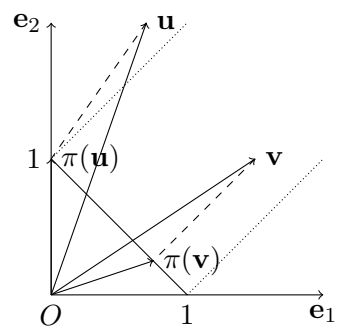

Figure 1: The Euclidean projection onto the simplex $\mathbb{S}_+^1$.

The substitution of the self-normalized vector $\mathbf{X}/|\mathbf{X}|$ by the projected vector $\pi(\mathbf{X}/t)$ motivates the following definition.

Definition 1 (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*

$$\mathbb{P}\big((|\mathbf{X}|/t, \pi(\mathbf{X}/t)) \in \cdot \mid |\mathbf{X}| > t\big) \xrightarrow{w} \mathbb{P}((Y, \mathbf{Z}) \in \cdot)\,, \quad t \to \infty\,. \tag{2.3}$$

In this case we write $\mathbf{X} \in \mathrm{SRV}(\alpha, \mathbf{Z})$.

In this case, standard results on regularly varying random variables state that there exists $\alpha > 0$ such that Y follows a Pareto distribution with parameter α. The continuity of π ensures that regular variation with limit $(Y, \mathbf{\Theta})$ implies sparse regular variation with limit $(Y, \pi(Y\mathbf{\Theta}))$. Contrary to the limit in (2.1), we lose independence between the radial component Y and the angular component $\mathbf{Z}$ of the limit. The dependence relation between both components is given by the following proposition.

Proposition 1. *If $\mathbf{X} \in \mathrm{SRV}(\alpha, \mathbf{Z})$, then for all $r \geq 1$ we have*

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

For $\beta \in \mathcal{P}_d^*$ we denote by $\mathbf{e}(\beta)$ the vector with 1 in position j if $j \in \beta$ and 0 otherwise. Then the vector $\mathbf{e}(\beta)/|\beta|$ belongs to the simplex. We consider the following class of discrete distributions on the simplex:

$$\sum_{\beta \in \mathcal{P}_d^*} p(\beta)\, \delta_{\mathbf{e}(\beta)/|\beta|}\,, \tag{2.4}$$

where $(p(\beta))_\beta$ is a 2^d-1 vector with nonnegative components summing to 1. This is the device developed by Segers (2012). The family of distributions (2.4) is stable under multiplication by a positive random variable and Euclidean projection onto the simplex. Hence, if $\mathbf{\Theta}$ has a distribution of type (2.4), then $\mathbf{Z} = \mathbf{\Theta}$ a.s. The following corollary states that these kind of distributions are the only possible discrete distributions for $\mathbf{Z}$.

Corollary 1. *If the distribution of* $\mathbf{Z}$ *is discrete, then it is of the form* (2.4).

The family of distributions given in (2.4) forms an accurate model for the angular vector $\mathbf{Z}$. Indeed, the distributions of this class place mass on some particular points of the simplex on which extremes values often concentrate in practice. It includes the case of complete dependence which corresponds to the case $p(\{1,\dots,d\}) = 1$ and the case of asymptotic independence which corresponds to the case $p(\{j\}) = 1/d$ for all $j = 1,\dots,d$ (see also Example 2).

2.2 Main result

We establish in this section some relations between both vectors $\mathbf{\Theta}$ and $\mathbf{Z}$. We start this section by introducing some notation which will be useful throughout the article. We consider the sets

$$A_{\mathbf{x}} = \{\mathbf{u} \in \mathbb{S}_+^{d-1} : \mathbf{u} \geq \mathbf{x}\}\,, \quad \mathbf{x} \in B_+^d(0,1)\,,$$

$$\mathcal{X}_\beta = \{\mathbf{x} \in B_+^d(0,1) : \mathbf{x}_\beta > \mathbf{0}_\beta,\, \mathbf{x}_{\beta^c} = \mathbf{0}_{\beta^c}\}\,, \quad \beta \in \mathcal{P}_d^*\,,$$

$$\mathcal{X}_\beta^0 = \{\mathbf{x} \in B_+^d(0,1) : \mathbf{x}_{\beta^c} = \mathbf{0}_{\beta^c}\}\,, \quad \beta \in \mathcal{P}_d^*\,,$$

and we define λ_β as the Lebesgue measure on the set $\mathcal{X}_\beta$. For $\beta, \gamma \in \mathcal{P}_d^*$ such that $\gamma \supset \beta$ we also consider the functions

$$G_\beta(\mathbf{x}) = \mathbb{P}(\mathbf{Z}_\beta > \mathbf{x}_\beta,\, \mathbf{Z}_{\beta^c} \leq \mathbf{x}_{\beta^c})\,, \quad \mathbf{x}_\beta \in B_+^{|\beta|}(0,1)\,, \quad \mathbf{x}_{\beta^c} \in B_+^{|\beta^c|}(0,1)\,,$$

and

$$H_{\beta,\gamma}(\mathbf{u}, v, w) = \mathbb{P}\big(\phi_\gamma(\mathbf{Z})_\beta \geq \mathbf{u},\, \min_{j \in \gamma \setminus \beta} \phi_\gamma(\mathbf{Z})_j > v,\, \max_{j \in \gamma^c} \phi_\gamma(\mathbf{Z})_j \leq w\big)\,, \quad \mathbf{u} \in B_+^{|\beta|}(0,1)\,, \quad v, w \in [0,1]\,,$$

where $\phi_\gamma : \mathbb{S}_+^{d-1} \to \mathbb{R}_+^d$, $\mathbf{u} \mapsto \phi_\gamma(\mathbf{u})$ is defined as

$$\phi_\gamma(\mathbf{u})_j = \begin{cases} u_j + \frac{|\mathbf{u}_{\gamma^c}|}{|\gamma|}\,, & \text{for } j \in \gamma\,, \\ u_j + \frac{|\mathbf{u}_{\gamma^c \setminus \{j\}}|}{|\gamma|+1}\,, & \text{for } j \in \gamma^c\,, \end{cases}$$

see Lemma 3 for more insights on this function.

We consider the following assumption:

(A) For all $\beta, \gamma \in \mathcal{P}_d^*$ such that $\gamma \supset \beta$ and for λ_β-almost every $\mathbf{x} \in \mathcal{X}_\beta$ the function $H_{\beta,\gamma}$ is continuously differentiable in $(\mathbf{x}_\beta, 0, 0)$.

Remark 2 (On assumption (A))**.** Suppose that the distribution of $\mathbf{Z}$ is a mixture of a discrete part $\sum_k a_k \delta_{\mathbf{u}_k}$ and a continuous part $\sum_{\beta \in P} a_\beta f_\beta$ with $P \subset \{\beta \in \mathcal{P}_d^* : |\beta| \geq 2\}$ and with continuous densities f_β. Then, so is the distribution of $\phi_\gamma(\mathbf{Z})$, since $\phi_\gamma(\mathbf{Z})$ is a linear transformation of $\mathbf{Z}$. In this case, assumption (A) is satisfied.

We are now able to state the main result of this paper.

Theorem 1 (Equivalence of regular variation and sparse regular variation)**.** *Under assumption (A) we have the equivalence between regular variation and sparse regular variation. More precisely:*

1. *If* $\mathbf{X} \in \mathrm{RV}(\alpha, \mathbf{\Theta})$*, then* $\mathbf{X} \in \mathrm{SRV}(\alpha, \mathbf{Z})$*, with* $\mathbf{Z} = \pi(Y\mathbf{\Theta})$ *satisfying*

$$G_\beta(\mathbf{x}) = \mathbb{E}\Bigg[\Bigg(1 \wedge \min_{j\in\beta_+}\Big(\frac{|\beta|\Theta_j - |\mathbf{\Theta}_\beta|}{|\beta|x_j - 1}\Big)_+^\alpha \wedge \min_{j\in\beta^c}(|\mathbf{\Theta}_\beta| - |\beta|\Theta_j)_+^\alpha - \max_{j\in\beta_-}\Big(\frac{|\beta|\Theta_j - |\mathbf{\Theta}_\beta|}{|\beta|x_j - 1}\Big)_+^\alpha\Bigg)_+\Bigg], \quad (2.5)$$

for all $\mathbf{x} \in \mathcal{X}_\beta^0$ *such that for all* $j \in \beta$*,* $x_j \neq 1/|\beta|$*, and where* $\beta_+ = \{j \in \beta,\, x_j > 1/|\beta|\}$ *and* $\beta_- = \{j \in \beta,\, x_j < 1/|\beta|\}$*.*

2. *If* $\mathbf{X} \in \mathrm{SRV}(\alpha, \mathbf{Z})$ *with* $\mathbf{Z}$ *satisfying (A), then* $\mathbf{X} \in \mathrm{RV}(\alpha, \mathbf{\Theta})$ *with* $\mathbf{\Theta}$ *satisfying*

$$\mathbb{P}(\mathbf{\Theta} \in A_\mathbf{x}) = \mathbb{P}(\mathbf{Z} \in A_\mathbf{x}) + \alpha^{-1}\sum_{\gamma\supset\beta} \mathrm{d}H_{\beta,\gamma}(\mathbf{x}_\beta, 0, 0)\cdot\Big(\mathbf{x}_\beta - \frac{1}{|\gamma|}, -\frac{1}{|\gamma|}, -\frac{1}{|\gamma|+1}\Big), \quad (2.6)$$

for $\beta \supset \gamma \in \mathcal{P}_d^*$ *and* λ_β*-almost every* $\mathbf{x} \in \mathcal{X}_\beta$*.*

Remark 3 (Discrete distributions)**.** Note that if the angular vector $\mathbf{Z}$ has a discrete distribution, then the assumption (A) is satisfied and we have $\mathrm{d}H(\mathbf{x}_\beta, 0, 0) = 0$ for all $\beta \subset \gamma$ and λ_β-almost every $\mathbf{x} \in \mathcal{X}_\beta$. Hence Theorem 1 ensures that there exists a spectral vector $\mathbf{\Theta}$ such that $\mathbb{P}(\mathbf{\Theta} \in A_\mathbf{x}) = \mathbb{P}(\mathbf{Z} \in A_\mathbf{x})$, i.e. $\mathbf{\Theta} \overset{d}{=} \mathbf{Z}$. Actually, using Corollary 1 and similar arguments than above we obtain $\mathbf{\Theta} = \mathbf{Z}$ a.s.

Equation (2.5) gives a relation between the distribution of $\mathbf{Z}$ and the one of $\mathbf{\Theta}$. 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 with $\beta = \{1, \dots, d\}$ and $\mathbf{x} < 1/d$ which provides the relation

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

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

$$G_\mathbf{Z}(\mathbf{0}) = 1 - \mathbb{E}\Big[\max_{1\le j\le d}(1 - d\Theta_j)^\alpha\Big]. \quad (2.7)$$

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

$$\mathbb{P}(Z_j = 0 \text{ for some } j \le d) = \mathbb{E}\Big[\max_{1\le j\le d}(1 - d\Theta_j)^\alpha\Big]. \quad (2.8)$$

This quantity is null if and only if for all $j = 1, \dots, d$, $\Theta_j = 1/d$ a.s. and is equal to 1 if and only if $\min_{1\le j\le d}\Theta_j = 0$ a.s. This implies that the new angular vector $\mathbf{Z}$ is more likely to be sparse in the sense of Remark 1. In particular, all usual spectral models on $\mathbf{\Theta}$ that are not supported on the axis are not suitable for $\mathbf{Z}$. More insights into the sparsity of the vector $\mathbf{Z}$ is given in Section 3.

Example 1. We consider a spectral vector $\mathbf{\Theta}$ in $\mathbb{S}_+^1$ with a first component Θ_1 uniformly distributed (and then $\Theta_2 = 1 - \Theta_1$ is also uniformly distributed). This fits into the framework of Remark 2. We also assume $\alpha = 1$ for simplicity. The probability that $\mathbf{Z}$ belongs to the first axis is equal to $\mathbb{P}(Z_1 = 1) = \mathbb{P}(Y\Theta_1 - Y\Theta_2 \ge 1) = \mathbb{P}(2\Theta_1 - 1 \ge 1/Y)$ (see Lemma 3). Since the random variable $1/Y$ follows a uniform distribution on $(0, 1)$ and is independent of $\mathbf{\Theta}$ we obtain

$$\mathbb{P}(Z_1 = 1) = \int_0^1 \mathbb{P}(2\Theta_1 - 1 \ge u)\,\mathrm{d}u = \int_0^1 \mathbb{P}\Big(\Theta_1 \ge \frac{u+1}{2}\Big)\,\mathrm{d}u = \int_0^1 \frac{1-u}{2}\,\mathrm{d}u = 1/4. \quad (2.9)$$

Besides, if $x \in (0,1)$, Lemma 3 entails that

$$\mathbb{P}(0 < Z_1 \leq x) = \mathbb{P}(0 < Y\Theta_1 - (Y-1)/2 \leq x) = \int_0^1 \mathbb{P}\big[(1-u)/2 < \Theta_1 \leq (1+(2x-1)u)/2\big]\,\mathrm{d}u = x/2\,.$$

The distribution of Z_1 is thus given by $\delta_0/4 + \delta_1/4 + U(0,1)/2$, where U denotes a uniform distribution on $(0,1)$ and δ_a a Dirac measure at point a. We check in Appendix B that this vector $\mathbf{Z}$ satisfies Equations (2.5) and (2.6).

3 Detection of extremal directions with sparse regular variation

3.1 Sparsity in extremes

This section tackles the issue of detecting extremal directions for a regularly varying random vector $\mathbf{X}$. In such a context it is helpful to partition the underlying space, in our case $\mathbb{S}_+^{d-1}$, with understandable subsets (Chautru (2015), Goix et al. (2017), Simpson et al. (2020)). In this article we consider the subsets

$$C_\beta = \big\{\mathbf{x} \in \mathbb{S}_+^d : x_i > 0 \text{ for } i \in \beta,\, x_i = 0 \text{ for } i \notin \beta\big\}\,, \quad \beta \in \mathcal{P}_d^*\,, \tag{3.1}$$

which form a partition of the simplex. An illustration of these subsets in dimension 3 is given in Figure 2. This partition is helpful to study the tail structure of $\mathbf{X}$. Indeed, for $\beta \in \mathcal{P}_d^*$ the inequality $\mathbb{P}(\mathbf{\Theta} \in C_\beta) > 0$ 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.

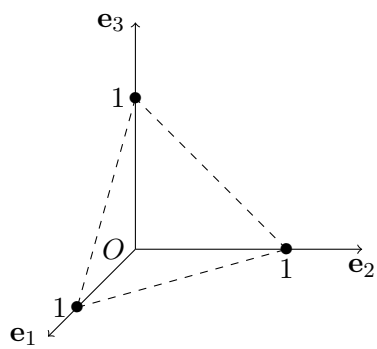

Figure 2: The subsets C_β in dimension 3 for the ℓ^1-norm. The subsets $C_{\{1\}}$, $C_{\{2\}}$, and $C_{\{3\}}$ correspond to the unit vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, respectively. The dashed lines indicate the subsets $C_{\{1,2\}}$, $C_{\{1,3\}}$, and $C_{\{2,3\}}$. Finally, the subset $C_{\{1,2,3\}}$ correspond to the interior of the simplex.

Remark 4. Our approach aims to detect sparse directions that are aligned with the standard coordinate system. For $\mathbf{X} \in \mathbb{R}_+^d$ it allows us to understand how a marginal affects the extremal behavior of $\mathbf{X}$. In applications, if $\mathbf{X}$ represents a phenomenon, the goal is to understand which groups of marginals are the main causes of the extremal behavior of this phenomenon. This is why we do not focus on what happens if the directions do not align with the standard coordinate system.

Example 2. A standard example of sparsity is asymptotic independence for which the spectral measure only puts mass on the axis: $\mathbb{P}(\mathbf{\Theta} \in \sqcup_{1\leq j\leq d}\{\mathbf{e}_j\}) = \mathbb{P}(\mathbf{\Theta} \in \sqcup_{1\leq j\leq d} C_{\{j\}}) = 1$. This means that there is never more than one direction which contributes to the extremal behavior of the vector. This concept has been studied by many authors, see for instance Ledford and Tawn (1996) or Ramos and Ledford (2009).

As for any low-dimensional subspaces, topological issues may arise for the C_β's in the convergence (2.1). Indeed, for $\beta \neq \{1, \dots, d\}$ the subset C_β is included in its boundary (with respect to the topology of the simplex) and the convergence (2.1) fails for such a set. This kind of problems appears since the spectral measure may put mass on low-dimensional subspaces while the data generally do not concentrate on such subspaces. This issue can be circumvent with sparse regular variation. We replace the study of the sets $\{\mathbf{x} \in \mathbb{R}_+^d : |\mathbf{x}| > 1,\, \mathbf{x}/|\mathbf{x}| \in C_\beta\}$ by the one of the sets $\{\mathbf{x} \in \mathbb{R}_+^d : |\mathbf{x}| > 1,\, \pi(\mathbf{x}/t) \in C_\beta\}$ which enjoy better topological properties.

In this context, many other sets have been proposed recently. Goix et al. (2016) define the truncated ϵ-cones as

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

Subsequently, Goix et al. (2017) introduce the notion of ϵ-thickened rectangles:

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

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

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

to focus on groups of variables that may be large together. Similarly to the ideas of Chiapino and Sabourin (2016) we relax the condition on β in the definition of the C_β's which leads to the study of the subsets $\{\mathbf{x} \in \mathbb{S}_+^{d-1} : x_i = 0 \text{ for } i \notin \beta\} = \mathbb{S}_+^{d-1} \cap \mathrm{Vect}(\mathbf{e}_j, j \in \beta)$.

We gather in the following proposition some results regarding the behavior of $\pi(\mathbf{X}/t)$ and $\mathbf{Z}$ on the subsets C_β and $\mathbb{S}_+^{d-1} \cap \mathrm{Vect}(\mathbf{e}_j, j \in \beta)$. In all what follows we write $\Theta_{\beta,j}$ (resp. $\Theta_{\beta,j,+}$) for $\sum_{k\in\beta}(\Theta_k - \Theta_j)$ (resp. $\sum_{k\in\beta}(\Theta_k - \Theta_j)_+$).

Proposition 2. *Let $\mathbf{X} \in \mathrm{RV}(\alpha, \mathbf{\Theta})$ and set $\mathbf{Z} = \pi(Y\mathbf{\Theta})$, where Y is a Pareto(α)-distributed random variable independent of $\mathbf{\Theta}$.*

1. *For any $\beta \in \mathcal{P}_d^*$ we have*

$$\mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta \mid |\mathbf{X}| > t) \to \mathbb{P}(\mathbf{Z} \in C_\beta), \quad t \to \infty, \tag{3.2}$$

and

$$\mathbb{P}(\pi(\mathbf{X}/t)_{\beta^c} = 0 \mid |\mathbf{X}| > t) \to \mathbb{P}(\mathbf{Z}_{\beta^c} = 0), \quad t \to \infty. \tag{3.3}$$

2. *For any $\beta \in \mathcal{P}_d^*$ we have*

$$\mathbb{P}(\mathbf{Z} \in C_\beta) = \mathbb{E}\Big[\Big(\min_{j\in\beta^c} \Theta_{\beta,j,+}^\alpha - \max_{j\in\beta} \Theta_{\beta,j,+}^\alpha\Big)_+\Big], \tag{3.4}$$

and

$$\mathbb{P}(\mathbf{Z}_{\beta^c} = 0) = \mathbb{E}\Big[\min_{j\in\beta^c} \Theta_{\beta,j,+}^\alpha\Big]. \tag{3.5}$$

Regarding the behavior of $\pi(\mathbf{X}/t)$, the convergence (2.3) holds for any pair of Borel sets $A \times B \in (1,\infty) \times \mathbb{S}_+^{d-1}$ such that $\mathbb{P}(Y\mathbf{\Theta} \in \partial\pi^{-1}(B)) = 0$, where $\partial\pi^{-1}(B)$ denotes the boundary of the set $\pi^{-1}(B)$. The first point of Proposition 2 states that the subsets C_β and $\mathbb{S}_+^{d-1} \cap \mathrm{Vect}(\mathbf{e}_j, j \in \beta)$ satisfy this condition. This implies in particular that the sparsity of $\mathbf{Z}$ can be studied through the projected vector $\pi(\mathbf{X}/t)$. This will be illustrated in Section 4.

The second point of Proposition 2 provides some interesting relations between the sparsity of $\mathbf{Z}$ and $\mathbf{\Theta}$. If we consider $\beta = \{1, \dots, d\}$, then we obtain the probability that all coordinates are positive which has already been computed in (2.7). It is equal to $G_{\mathbf{Z}}(\mathbf{0}) = 1 - \mathbb{E}\left[\max_{1\leq j\leq d}(1 - d\Theta_j)^\alpha\right]$. Another particular case of Equations (3.4) and (3.5) 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

$$\mathbb{P}(Z_{j_0} = 1) = \mathbb{E}\Big[\min_{j \neq j_0} (\Theta_{j_0} - \Theta_j)_+^\alpha\Big]\,. \tag{3.6}$$

Then, Equation (3.6) can be developed in the following way:

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

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

Remark 5. Following Equation (3.5), we write

$$\mathbb{P}(\mathbf{Z}_{\beta^c} = 0) \geq \mathbb{E}\Big[\min_{j \in \beta^c} \Theta^\alpha_{\{1,\dots,d\},\,j,\,+} \mathbb{1}_{\{\mathbf{\Theta}_{\beta^c}=0\}}\Big] = \mathbb{E}\Big[\big(\sum_{k \leq d} \Theta_k\big)^\alpha \mathbb{1}_{\{\mathbf{\Theta}_{\beta^c}=0\}}\Big] = \mathbb{P}(\mathbf{\Theta}_{\beta^c} = 0)\,. \tag{3.7}$$

This can also be seen as a direct consequence of Property P2, see Section 2.1. This property also gives

$$\mathbb{P}(\mathbf{Z}_\beta > 0) \leq \mathbb{P}(\mathbf{\Theta}_\beta > 0)\,, \tag{3.8}$$

an inequality which will be useful in some proofs.

3.2 Maximal directions

We focus in this section on the subsets C_β. A positive value for $\mathbb{P}(\mathbf{\Theta} \in C_\beta)$ entails that the marginals X_j for $j \in \beta$ take simultaneously large values while the ones in β^c do not (Chautru (2015), Simpson et al. (2020), Goix et al. (2017)). Our aim is to use Proposition 2 to compare the nullity or not of the probabilities $\mathbb{P}(\mathbf{\Theta} \in C_\beta)$ and $\mathbb{P}(\mathbf{Z} \in C_\beta)$. To this end, it is relevant to focus on the largest group of variables $\beta \in \mathcal{P}_d^*$ such that $\mathbb{P}(\mathbf{\Theta} \in C_\beta) > 0$. This motivates the notion of maximal direction.

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

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

We similarly define maximal directions for $\mathbf{Z}$.

Remark 6. A straightforward but useful consequence of Definition 2 is that each direction β such that $\mathbb{P}(\mathbf{\Theta} \in C_\beta) > 0$ is included in a maximal direction of $\mathbf{\Theta}$. Indeed, if there exists no $\beta' \supsetneq \beta$ such that $\mathbb{P}(\mathbf{\Theta} \in C_{\beta'}) = 0$, then β is maximal itself. If not, we consider $\beta' \supsetneq \beta$ such that $\mathbb{P}(\mathbf{\Theta} \in C_{\beta'}) > 0$. If β' 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 $\beta \subset \gamma$, $\mathbb{P}(\mathbf{\Theta} \in C_\gamma) > 0$ and $\mathbb{P}(\mathbf{\Theta} \in C_{\gamma'}) = 0$, for all $\gamma' \supsetneq \gamma$.

The notion of maximal directions is justified by the following theorem.

Theorem 2. *Let $\beta \in \mathcal{P}_d^*$.*

1. *If $\mathbb{P}(\mathbf{\Theta} \in C_\beta) > 0$, then $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$.*
2. *The direction β is maximal for $\mathbf{\Theta}$ if and only if it is maximal for $\mathbf{Z}$.*

Theorem 2 implies that we do not lose any information on the extremal directions of β by studying $\mathbf{Z}$ instead of $\mathbf{\Theta}$. But it is possible that the distribution of $\mathbf{Z}$ puts some mass on a subset C_β while the one of $\mathbf{\Theta}$ does not. In such a case, the associated direction β is necessarily non-maximal.

Example 3. In Example 1 we proved that if Θ_1 follows a uniform distribution on $(0,1)$, then $\mathbb{P}(\mathbf{Z} \in C_{\{1\}}) = 1/4$ while $\mathbb{P}(\Theta_1 = 1) = 0$. This proves that the direction $\beta = \{1\}$ is non-maximal for $\mathbf{Z}$.

Example 3 shows there may exist $\beta \in \mathcal{P}_d^*$ such that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$ and $\mathbb{P}(\mathbf{\Theta} \in C_\beta) = 0$. In this case, Theorem 2 states that the direction β is not maximal for $\mathbf{Z}$ since it is not maximal for $\mathbf{\Theta}$. Following Remark 6, we consider a maximal direction γ of $\mathbf{Z}$ such that $\beta \subset \gamma$. Then Theorem 2 states that $\mathbb{P}(\mathbf{\Theta} \in C_\gamma) > 0$. 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 directions 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 direction regarding the spectral vector $\mathbf{\Theta}$.

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

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

Following Equation (C.20) we obtain that

$$0 < \mathbb{P}(\pi(Y\mathbf{\Theta})_{\beta^c} = 0) = \mathbb{P}\big(|\beta|^{-1}|\mathbf{\Theta}_\beta| \geq \max_{i\in\beta^c} \Theta_i + Y^{-1}\big)\,. \tag{3.9}$$

This implies that with positive probability $|\beta|^{-1}\sum_{k\in\beta}\Theta_k \geq \Theta_i$ 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 (3.9) yields to

$$0 < \mathbb{P}\big(|\beta|^{-1}|\mathbf{\Theta}_\beta| \geq \max_{i\in\beta^c} \Theta_i + Y^{-1}\big) = \lim_{t\to\infty} \mathbb{P}\big(|\beta|^{-1}|\mathbf{X}_\beta| \geq \max_{i\in\beta^c} X_i + t \mid |\mathbf{X}| > t\big)\,.$$

This means that the extreme values of $\mathbf{X}_\beta$ are likely to be larger than the extreme ones of $\mathbf{X}_{\beta^c}$. This does not contradict the fact that $\mathbb{P}(\mathbf{\Theta} \in C_\beta) = 0$ 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 direction 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 directions $\beta \subsetneq \gamma$ such that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$. 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 deeper interpretation of non-maximal directions is deferred to future work.

Example 4. 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 $a_1 > a_2 > \ldots > a_d > 0$ and we also assume for simplicity that $\mathbf{a} \in \mathbb{S}_+^{d-1}$. We define the vector $\mathbf{X}$ by setting $\mathbf{X} = X\mathbf{a} = (a_1X, \ldots, a_dX) \in \mathbb{R}_+^d$. Then, the vector $\mathbf{X}$ is regularly varying with tail index α and a spectral vector given by $\mathbf{\Theta} = \mathbf{a}$ a.s. This means that the direction $\{1, \ldots, d\}$ is the only one on which the spectral measure places mass, and it is a maximal one. Hence, the angular vector $\mathbf{Z}$ satisfies $\mathbb{P}(\mathbf{Z} \in C_{\{1,\ldots,d\}}) > 0$. 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 $\{1\}, \{1,2\}, \{1,2,3\}, \ldots, \{1, \ldots, d-1\}$.

We first consider the direction $\{1\}$ and compute the probability that $\mathbf{Z}$ belongs to the subset $C_{\{1\}}$. Following Equation (C.3), we obtain that

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

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

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

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 (C.4) implies that $\mathbf{Z}$ belongs to the subset $C_{\{1,\ldots,r\}}$ 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 $C_{\{1,\ldots,r\}}$ is equal to

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

If we take $\alpha = 1$ for the sake of simplicity, then we obtain $\mathbb{P}(\mathbf{Z} \in C_{\{1,\ldots,r\}}) = r(a_r - a_{r+1})$ and thus the probability that $\mathbf{Z}$ belongs to the subset $C_{\{1,\ldots,r\}}$ 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 directions. It highlights the relative importance of a coordinate regarding the extreme values of a group of directions this coordinate belongs to.

4 Numerical results

This section is devoted to a statistical illustration of sparse regular variation. We highlight how our approach manages to detect sparsity in the tail dependence. We provide a method in order to approximate the probabilities $\mathbb{P}(\mathbf{Z} \in C_\beta)$ and apply it to several numerical results.

4.1 The framework

We consider an iid sequence of regularly varying random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with generic distribution $\mathbf{X} \in \mathrm{RV}(\alpha, \mathbf{\Theta})$. We set $\mathbf{Z} = \pi(Y\mathbf{\Theta})$ where Y follows a Pareto(α) distribution independent of $\mathbf{\Theta}$. Our aim is to capture the directions $\beta \in \mathcal{P}_d^*$ such that $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$. Thanks to Proposition 2 the latter probability is defined through the limit

$$\mathbb{P}(\mathbf{Z} \in C_\beta) = \lim_{t\to\infty} \mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta \mid |\mathbf{X}| > t) = \lim_{t\to\infty} \frac{\mathbb{P}(\pi(\mathbf{X}/t) \in C_\beta, |\mathbf{X}| > t)}{\mathbb{P}(|\mathbf{X}| > t)}\,. \tag{4.1}$$

The goal is then to approximate this probability with the sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$. We define the quantity

$$T_\beta(t) = \frac{\sum_{j=1}^{n} \mathbb{1}\{\pi(\mathbf{X}_j/t) \in C_\beta, |\mathbf{X}_j| > t\}}{\sum_{j=1}^{n} \mathbb{1}\{|\mathbf{X}_j| > t\}}\,, \quad t > 0\,, \quad \beta \in \mathcal{P}_d^*\,, \tag{4.2}$$

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 then entails 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 allows one to classify the directions 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 (4.2) and only focus on $\sum_{j=1}^n \mathbb{1}\{\pi(\mathbf{X}_j/t) \in C_\beta, |\mathbf{X}_j| > t\}$. The selection of the largest vectors $\mathbf{X}_j$ whose norm is above t then boils down to keeping only a proportion, say k, of vectors. It is customary in EVT to choose a level k which satisfies $k \to \infty$ and $k/n \to 0$ when $n \to \infty$.

Remark 7 (The approach proposed by Goix et al. (2017))**.** In order to detect anomalies among multivariate extremes, Goix et al. (2017) propose a similar approach with the ℓ^∞-norm based on the ϵ-thickened rectangles

$$R_\beta^\epsilon = \{\mathbf{x} \in \mathbb{R}_+^d : |\mathbf{x}|_\infty > 1,\, x_j > \epsilon \text{ for all } j \in \beta,\, x_j \le \epsilon \text{ for all } j \in \beta^c\}\,, \quad \beta \in \mathcal{P}_d^*\,.$$

Starting from the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ with generic random vector $\mathbf{X} = (X^1, \dots, X^d)$ with marginal distribution $F_1, \dots, F_d$, 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 $\hat{F}_j : x \mapsto \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i^j < x}$ is the empirical counterpart of F_j. This rank transformation provides standardized marginals to the vectors $\mathbf{V}_i$. Denoting by $\tilde{\mathbf{\Theta}}_\infty$ the nonstandard spectral vector with respect to the ℓ^∞-norm and by $C_{\beta,\infty} = \big\{\mathbf{x} \in \mathbb{S}_{+,\infty}^d,\, x_i > 0 \text{ for } i \in \beta,\, x_i = 0 \text{ for } i \notin \beta\big\}$ the associated subsets, Goix et al. (2017) use the approximation

$$T_\beta(k, \epsilon) = \frac{1}{k}\sum_{i=1}^n \mathbb{1}_{\mathbf{V}_i \in (n/k) R_\beta^\epsilon} \approx c\,\mathbb{P}(\tilde{\mathbf{\Theta}}_\infty \in C_{\beta,\infty})\,, \quad c > 0\,,$$

for k large, and ϵ close to zero (the ratio n/k plays the role of the large threshold t, see de Haan and Ferreira (2006)). The authors propose an algorithm called DAMEX whose goal is to identify the subsets $C_{\beta,\infty}$ such that $\mathbb{P}(\tilde{\mathbf{\Theta}}_\infty \in C_{\beta,\infty}) > 0$.

Remark 8 (On the choice of the norm)**.** After some calculations we observe that if the spectral vectors $\tilde{\mathbf{\Theta}}$ and $\tilde{\mathbf{\Theta}}_\infty$ correspond to nonstandard regular variation (with $\alpha = 1$), then they satisfy the relation

$$\mathbb{P}(\tilde{\mathbf{\Theta}} \in B) = \frac{\mathbb{E}\big[|\tilde{\mathbf{\Theta}}_\infty| \mathbb{1}_{\{\tilde{\mathbf{\Theta}}_\infty/|\tilde{\mathbf{\Theta}}_\infty| \in B\}}\big]}{\mathbb{E}[|\tilde{\mathbf{\Theta}}_\infty|]}\,, \tag{4.3}$$

for all $B \in \mathbb{S}_+^{d-1}$. Since $\tilde{\mathbf{\Theta}}_\infty/|\tilde{\mathbf{\Theta}}_\infty| \in C_\beta$ if and only if $\tilde{\mathbf{\Theta}}_\infty \in C_{\beta,\infty}$ we obtain the equivalence

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

Hence, the directions in which extremes gather are the same regardless the choice of the norm. This means that after a standardization of the marginals we can compare the performance of our method with the one of Goix et al. (2017). This is what we do in the first numerical example below.

Note that if $\mathbf{X}$ is regularly varying with tail index 1, then $\mathbf{X}^q$ is regularly varying with tail index $1/q$ and Equation (4.3) implies that the corresponding spectral measures concentrate on the same subsets C_β (see also Remark 11).

Remark 9. 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, we obtain a bias which provides a difference between some directions β for which $T_\beta(t)$ takes small values while the theoretical quantities $\mathbb{P}(\mathbf{Z} \in C_\beta)$ are null. We follow the idea of Goix et al. (2017), Remark 4, to deal with this issue. We define a threshold value under which the empirical quantities $T_\beta(t)$ 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 directions β such that $T_\beta(t) > 0$. In this case the number of selected 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.

Algorithm 1: Extremal dependence structure of sparsely regularly varying random vectors

Data: $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}_+^d$, $t > 0$, $p \geq 0$
Result: A list $\mathcal{C}$ of directions β
Compute $\pi(\mathbf{X}_j/t)$, $j = 1, \ldots, n$;
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 9;
Define $\mathcal{C} = \{\beta,\, T_\beta(t) > 0\}$.

4.2 Experimental results

In this section we consider two different cases of numerical data for which we apply Algorithm 1. 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 $\mathcal{C} = \{\beta,\, T_\beta(t) > 0\}$ 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. 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. 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 to study the procedure given in Algorithm 1 and to see how it manages to detect the sparsity in the extremes. We also analyze the influence of the tail index α by choosing different values for this parameter. This is done by considering the random vector $\mathbf{X}^\alpha = (X_1^\alpha, \ldots, X_d^\alpha)$ whose tail index is $1/\alpha$ for $\mathbf{X} \in \mathrm{RV}(1, \mathbf{\Theta})$.

Remark 10 (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.

Remark 11. Assume that $\pi(\mathbf{X}/t) \in C_\beta$. This implies that for all $j \in \beta$ we have $X_j > \max_{i \in \beta^c} X_i$ and $|\mathbf{X}_\beta| - \max_{i \in \beta^c} X_i > t$, see (C.4). Hence for $k \in \beta$ and $q \geq 1$ we obtain

$$|\mathbf{X}_\beta^q| - \max_{i \in \beta^c} X_i^q = X_k^q - (\max_{i \in \beta^c} X_i)^q + |\mathbf{X}_{\beta \setminus \{k\}}^q| = (X_k - (\max_{i \in \beta^c} X_i)) \sum_{l=0}^{q} X_k^l (\max_{i \in \beta^c} X_i)^{q-l-1} + |\mathbf{X}_{\beta \setminus \{k\}}^q|$$

$$\geq X_k - (\max_{i\in\beta^c} X_i) + |\mathbf{X}_{\beta\setminus\{k\}}| = |\mathbf{X}_\beta| - (\max_{i\in\beta^c} X_i) \geq t\,,$$

as soon as we assume that all marginals satisfy $X_j > 1$. The relation $|\mathbf{X}^q_\beta| - \max_{i\in\beta^c} X^q_i$ implies that $\pi(\mathbf{X}^q/t)_{\beta^c} = 0$, see (C.3). In other words, it means that $\pi(\mathbf{X}^q/t)$ belongs to C_γ for $\gamma \subset \beta$.

In particular, if the spectral measure of $\mathbf{X}$ only concentrates on the axis, these directions can be easily detected with the study of $\pi(\mathbf{X}^q/t)$, $q \geq 1$, than with the one of $\pi(\mathbf{X}/t)$.

Asymptotic independence We consider an iid sequence of random vectors $\mathbf{N}_1, \ldots, \mathbf{N}_n$ 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 $\mathbf{X}_1, \ldots, \mathbf{X}_n$ such that the marginals X^j_i of $\mathbf{X}_i = (X^1_i, \ldots, X^d_i)$ are defined as

$$X^j_i = \frac{1}{1 - \hat{F}_j(N^j_i)}\,, \quad 1 \leq j \leq d\,,$$

where $\hat{F}_j$ is the empirical version of the cumulative distribution function of $N_j \sim \mathcal{N}(0,1)$. This provides a sample of regularly varying random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and the assumption on the correlation leads to asymptotic independence, i.e. $\mathbb{P}(\mathbf{\Theta} \in C_\beta) = \mathbb{P}(\mathbf{Z} \in C_\beta) = 1/d$ for all β such that $|\beta| = 1$ (Sibuya (1960)). The aim of our procedure is to recover these 40 directions among the $2^{40} - 1 \approx 10^{12}$ ones.

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

$$\mathbf{\Sigma} = \mathrm{Diag}({\sigma'}^{-1/2}_{1,1}, \ldots, {\sigma'}^{-1/2}_{d,d}) \cdot \mathbf{\Sigma}'^T \cdot \mathbf{\Sigma}' \cdot \mathrm{Diag}({\sigma'}^{-1/2}_{1,1}, \ldots, {\sigma'}^{-1/2}_{d,d})\,,$$

where $\mathrm{Diag}(\sigma^{-1/2}_{1,1}, \ldots, \sigma^{-1/2}_{d,d})$ denotes the diagonal matrix of $\mathcal{M}_d(\mathbb{R})$ whose diagonal is given by the vector $({\sigma'}^{-1/2}_{1,1}, \ldots, {\sigma'}^{-1/2}_{d,d})$. This provides a covariance matrix with diagonal entries equal to 1 and off-diagonal entries less than 1. A given matrix $\mathbf{\Sigma}$ provides then a dependence structure for $\mathbf{N}$ and thus for $\mathbf{X}$. We generate $N_{\text{model}} = 20$ different matrices $\mathbf{\Sigma}$ and for each of these dependence structures we generate $N = 100$ sample $\mathbf{N}_1, \ldots, \mathbf{N}_n$. We summarize in Table 1 and in Table 2 the two types of errors averaged among the $N \cdot N_{\text{model}} = 2000$ simulations.

In this case, the standard spectral vector $\mathbf{\Theta}$ and the non-standard one $\tilde{\mathbf{\Theta}}$ coincide. Hence it is possible to compare $\mathbf{\Theta}$ and $\tilde{\mathbf{\Theta}}_\infty$ (see Remark 8) which is done by computing the quantities $T_\beta(k, \epsilon)$ as well as the two types of errors for the DAMEX algorithm. We only study the effect of α for our approach since the DAMEX algorithm is not sensitive to marginals.

Errors of Type 1	Euclidean proj. $\alpha = 1$	Euclidean proj. $\alpha = 1/2$	Euclidean proj. $\alpha = 2$	DAMEX $\epsilon = 0.05$	DAMEX $\epsilon = 0.1$	DAMEX $\epsilon = 0.5$
$n = 10^4$	22.62	21.76	2.50	3034.70	2899.05	987.63
$n = 5 \cdot 10^4$	19.43	6.49	69.9	6972.52	4646.43	271.87
$n = 10^5$	1.83	0.65	99.79	8401.21	4813.46	235.80

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

For the Euclidean projection with $\alpha = 1$ we observe that our algorithm manages to capture almost all $d = 40$ directions regardless the value of n, and the error of type 2 decreases when n increases. On the other hand, our algorithm still captures some extra-directions, especially for $n = 10^4$ and $n = 5 \cdot 10^4$.

Errors of Type 2	Euclidean proj. $\alpha = 1$	Euclidean proj. $\alpha = 1/2$	Euclidean proj. $\alpha = 2$	DAMEX $\epsilon = 0.05$	DAMEX $\epsilon = 0.1$	DAMEX $\epsilon = 0.5$
$n = 10^4$	0.07	0.02	40.00	39.43	13.76	0.00
$n = 5 \cdot 10^4$	0.00	0.00	4.89	3.69	0.01	0.00
$n = 10^5$	0.00	0.00	0.41	0.07	0.00	0.00

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

This may be a consequence of the choice of p in Remark 9 which is probably too high and for which a deeper study should be conducted. The error of type 1 is then much lower for $n = 10^5$. We observe that for $\alpha = 1/2$ we obtain better results, while the number of errors is higher for $\alpha = 2$. Since the extremal directions are in this case only one-dimensional, this confirms numerically the observations of Remark 11.

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 n. Hence it seems that our procedure leads to the best compromise between both types of errors.

A dependent case We now consider a dependent case where extremes occur on lower-dimensional directions. In order to include dependence we recall that a vector $\mathbf{P}(k) = (P_1, P_1+P_2, \ldots, P_1+P_k) \in \mathbb{R}_+^k$ with P_j following Pareto(α_j), $\alpha_1 < \alpha_j$ for all $2 \leq j \leq k$, is regularly varying with tail index α_1 and spectral vector $\mathbf{\Theta} = (1/k, \ldots, 1/k)$. For our simulations we consider $s_1 = 10$ independent copies $\mathbf{P}_1, \ldots, \mathbf{P}_{s_1}$ of $\mathbf{P}(2) \in \mathbb{R}_+^2$ with $\alpha_1 = 1$ and $\alpha_2 = 2$ and $s_2 = 10$ independent copies $\mathbf{R}_1, \ldots, \mathbf{R}_{s_2}$ of $\mathbf{P}(3) \in \mathbb{R}_+^3$ with $\alpha_1 = 1$ and $\alpha_2 = \alpha_3 = 2$. 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 $j = 1, 3, \ldots, 17, 19$ and on the points $(\mathbf{e}_j + \mathbf{e}_{j+1} + \mathbf{e}_{j+2})/3$ for $j = 21, 24, \ldots, 45, 48$. Besides, as discussed after Proposition 1, in this case the angular vector $\mathbf{Z}$ and the spectral vector $\mathbf{\Theta}$ are equal almost surely. Our aim is to recover the $s_1 = 10$ two-dimensional directions $(\mathbf{e}_j + \mathbf{e}_{j+1})/2$ for $j = 1, 3, \ldots, 17, 19$ and also the $s_2 = 10$ three-dimensional directions $(\mathbf{e}_j + \mathbf{e}_{j+1} + \mathbf{e}_{j+2})/3$ for $j = 21, 24, \ldots, 45, 48$ based on a sample of iid random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with the same distribution as $\mathbf{X}$. Hence we would like to recover $s = s_1 + s_2 = 20$ directions among the $2^{50} - 1 \approx 10^{15}$ directions.

	Errors of type 1			Errors of type 2		
	Eucl. proj $\alpha = 1$	Eucl. proj $\alpha = 1/2$	Eucl. proj $\alpha = 2$	Eucl. proj $\alpha = 1$	Eucl. proj $\alpha = 1/2$	Eucl. proj $\alpha = 2$
$n = 10^4$	8.22	0.01	26.23	0.76	0.74	7.78
$n = 5 \cdot 10^4$	0.32	0.00	59.31	0.04	0.08	1.11
$n = 10^5$	0.04	0.00	78.97	0.03	0.01	0.35

Table 3: Average number of errors of type 1 and 2 in a dependent case ($d = 50$).

As for asymptotic independence, we remark that the number of errors decreases when n increases for almost all cases. For $\alpha = 1$ and $\alpha = 1/2$ our algorithm is not only able to detect all the $s = 20$ directions on which the distribution of $\mathbf{Z}$ puts mass, but it also does not identify some extra directions. These results are all the more accurate since the identification of the $s = 20$ directions is done among a very large number of directions, in this case $2^{50} - 1 \approx 10^{15}$. Besides, we obtain very low errors of both

types for $\alpha = 1$ and $\alpha = 1/2$. The better results we obtain for the error of type 1 in the case $\alpha = 1/2$ compared to the ones in the case $\alpha = 1$ can however not be explained by Remark 11. We defer this question to future work.

4.3 Sparse regular variation and non-maximal directions

In this section we illustrate some interpretation of the vector $\mathbf{Z}$ regarding extremal directions in non-maximal directions (see the discussion in Section 3.2). We consider a vector $\mathbf{a} \in \mathbb{S}_+^{r-1}$ as in Example 4 and a Pareto(α) distributed random variables P and define $P\mathbf{a} = (a_1 P, \ldots, a_r P) \in \mathrm{RV}(\alpha, \mathbf{a})$. Then, combining this device with the one of the dependent case we consider $\mathbf{P} = (a_1 P, a_2 P + P_2, \ldots, a_r P + P_r) \in \mathrm{RV}(\alpha, \mathbf{a})$ where $P_2, \ldots, P_r$ are iid random variables following a Pareto distribution with parameter $\alpha' > \alpha$. Hence the degenerate spectral vector $\mathbf{\Theta} = \mathbf{a}$ only places mass in the direction $\{1, \ldots, r\}$, which is thus maximal, while the vector $\mathbf{Z}$ places mass in all non-maximal directions $\{1\}, \{1, 2\}, \{1, 2, 3\}, \ldots, \{1, \ldots, d-1\}$, see Example 4. In our simulations, we choose $\alpha = 1$, $\alpha' = 2$, and $r = 3$, and we consider a vector $\mathbf{a} = (7, 6, 4)/|(7, 6, 4)|$. We then aggregate $s = 20$ iid copies $\mathbf{P}_k$ of the vector $\mathbf{P}$ and obtain $\mathbf{X} = (\mathbf{P}_1, \ldots, \mathbf{P}_s) \in \mathrm{RV}(1, \mathbf{\Theta})$ with $\mathbf{\Theta}$ placing mass on the three-dimensional maximal directions $\{j, j+1, j+2\}$ for $j \in J = \{1, 4, 7, \ldots, 58\}$ while the angular vector $\mathbf{Z}$ places mass on the aforementioned maximal directions but also on $s = 20$ two-dimensional and $s = 20$ one-dimensional directions.

The aim of the simulations is to see to what extent our procedure manages to recover the 40 non-maximal directions aforementioned. The columns of Table 4 gives the averaged number of the directions that have been recovered by our algorithm, depending on there size. Recall that for each type of directions the theoretical number of these directions that should appear is $s = 20$. Finally, the last column deals with the number of directions that should not appear theoretically. All the results are averaged among the $N = 100$ simulations.

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

Table 4: Average number of directions recovered by Algorithm 1 ($d = 60$).

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

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 deeper interpretation of $\mathbf{Z}$ in terms of extremes.

5 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.

Regarding extremes values, the vector $\mathbf{Z}$ enjoys many useful properties. This vector is sparser than the spectral vector $\mathbf{\Theta}$ 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 subsets C_β which highlight the tail dependence of $\mathbf{X}$. Proposition 2 provides a natural way to capture the behavior of $\mathbf{Z}$ on these subsets and proves that the Euclidean projection manages to circumvent the weak convergence's issue which arises in the standard definition of regularly varying random vectors.

Practically speaking, Section 4 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. 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 we obtain highlight the efficiency of our method to detect extremal directions. The future work should address the question of the threshold t, or equivalently the level k, and the bias issue introduced in Remark 9. Moreover, a comparison between $\mathbf{Z}$ and $\mathbf{\Theta}$ on non-maximal is also a crucial point to tackle. To this end, a deeper study of the behavior of $\mathbf{Z}$ on these kind of subsets should be conducted.

A Appendix A: Algorithms

We introduce here two algorithms which compute the Euclidean projection $\pi_z(\mathbf{v})$ given $\mathbf{v} \in \mathbb{R}_+^d$ and $z > 0$.

Algorithm 2: Euclidean projection onto the simplex.

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

Algorithm 2 emphasizes the number of positive coordinates $\rho_{\mathbf{v},z}$ of the projected vector $\pi_z(\mathbf{v})$:

$$\rho_{\mathbf{v},z} = \max\left\{j = 1, \ldots, d : \mu_j - \frac{1}{j}\Big(\sum_{r \leq j} \mu_r - z\Big) > 0\right\}, \tag{A.1}$$

where $\mu_1 \geq \ldots \geq \mu_d$ denote the order coordinates of $\mathbf{v}$, see Duchi et al. (2008), Lemma 2. In other words a coordinate j satisfies $\pi_z(\mathbf{v})_j > 0$ if and only if

$$v_j - \frac{1}{\sum_{k \leq d} \mathbb{1}_{v_k \geq v_j}} \Big(\sum_{k=1}^{d} v_k \mathbb{1}_{v_k \geq v_j} - z\Big) > 0\,. \tag{A.2}$$

The integer $\rho_{\mathbf{v},z}$ corresponds to the ℓ^0-norm of $\pi_z(\mathbf{v})$ and thus informs on the sparsity of this projected vector. For $z = 1$ we simply write $\rho_{\mathbf{v}}$.

A major remark is that Algorithm 2 allows one to compute $\pi_z(\mathbf{v})$ as soon as we know the set β of positive coordinates of this vector. Indeed if $\beta = \{j \leq d : \pi_z(\mathbf{v})_j > 0\}$, then $\pi_z(\mathbf{v})_j = v_j - (|\mathbf{v}_\beta| - z)/|\beta|$ for $j \in \beta$.

Algorithm 3 is a expected linear-time algorithm based on a median-search procedure.

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

Data: A vector $\mathbf{v} \in \mathbb{R}_+^d$ and a scalar $z > 0$
Result: The projected vector $\mathbf{w} = \pi(\mathbf{v})$
Initialize $U = \{1, \ldots, d\}$, $s = 0$, $\rho = 0$;
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 $\Delta\rho = |G|, \quad \Delta s = \sum_{j \in G} v_j$;
  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 $\eta = (s - z)/\rho.$;
Output: $\mathbf{w}$ s.t. $w_i = v_i - \eta$.

B Appendix B: Continuation of Example 1

Recall that Θ_1 is uniformly distributed on $(0, 1)$ and that the distribution of Z_1 is given by $\delta_0/4 + \delta_1/4 + U(0, 1)/2$.

We first check that Equation (2.5) holds for $\beta = \{1\}$ and $\beta = \{1, 2\}$. For $\beta = \{1\}$ we have $G_\beta(x, 0) = \mathbb{P}(Z_1 = 1) = 1/4$ while $\mathbb{E}[(\Theta_1 - \Theta_2)_+] = \int_0^{1/2}(1 - 2u)\,\mathrm{d}u = 1/4$. For $\beta = \{1, 2\}$, consider x_1, x_2 in $(0, 1)$ such that $x_1 + x_2 < 1$. On the one hand the quantity $G_\beta(x_1, x_2)$ corresponds to $\mathbb{P}(Z_1 > x_1, Z_2 > x_2) = \mathbb{P}(x_1 < Z_1 < 1 - x_2) = 1/2 - (x_1 + x_2)/2$. On the other hand, if we assume that $x_1, x_2 < 1/2$, then a similar calculation as in Example 1 leads to

$$
\begin{aligned}
\mathbb{E}\Big[\Big(1 - \Big(\frac{2\Theta_1 - 1}{2x_1 - 1}\Big)_+ \vee \Big(\frac{2\Theta_2 - 1}{2x_2 - 1}\Big)_+\Big)_+\Big] \\
&= \mathbb{E}\Big[\Big(1 - \Big(\frac{1 - 2\Theta_1}{1 - 2x_1}\Big)_+ \vee \Big(\frac{2\Theta_1 - 1}{1 - 2x_2}\Big)_+\Big)_+\Big] \\
&= \int_0^{1/2} \Big(1 - \frac{1 - 2u}{1 - 2x_1}\Big)_+ \mathrm{d}u + \int_{1/2}^1 \Big(1 - \frac{2u - 1}{1 - 2x_1}\Big)_+ \mathrm{d}u \\
&= \int_{x_1}^{1/2} \frac{u - x_1}{1/2 - x_1}\,\mathrm{d}u + \int_{1/2}^{1-x_2} \frac{1 - x_2 - u}{1/2 - x_2}\,\mathrm{d}u \\
&= \frac{[1/4 - x_1 + x_1^2]/2}{1/2 - x_1} + \frac{[(1 - x_2)^2 - (1 - x_2) + 1/4]/2}{1/2 - x_2}
\end{aligned}
$$

$$= \frac{1/2 - x_1}{2} + \frac{1/2 - x_2}{2}$$
$$= 1/2 - (x_1 + x_2)/2\,.$$

This proves that Equation (2.5) holds true.

Moving on to Equation (2.6), we first consider $\beta = \{1, 2\}$ and $\mathbf{x} = (x_1, x_2)$ with $x_1, x_2 > 0$. Since Θ is uniformly distributed on $(0, 1)$ we obtain that $\mathbb{P}(\mathbf{\Theta} \in A_{\mathbf{x}}) = \mathbb{P}(x_1 < \Theta < 1 - x_2) = 1 - (x_1 + x_2)$. On the other hand we already proved that $\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) = 1/2 - (x_1 + x_2)/2$. We now have to compute the differential of $H_{\beta,\beta}(\mathbf{x}) = G_\beta(\mathbf{x})$. For $\epsilon > 0$ we obtain

$$\begin{aligned} G_\beta(x_1 + \epsilon, x_2) - G_\beta(x_1, x_2) &= -\mathbb{P}(x_1 < Z_1 \leq x_1 + \epsilon, Z_2 > x_2) \\ &= -\mathbb{P}(x_1 < Z_1 \leq x_1 + \epsilon, Z_1 < 1 - x_2) \\ &= -\mathbb{P}(x_1 < Z_1 \leq x_1 + \epsilon) \quad \text{for } \epsilon \text{ small enough} \\ &= -\epsilon/2\,. \end{aligned}$$

Thus the differential of G_β satisfies the relation

$$dG_\beta(x_1, x_2) \cdot (x_1 - 1/2, x_2 - 1/2) = (1/2 - x_1 + 1/2 - x_2)/2 = 1/2 - (x_1 + x_2)/2\,.$$

Hence the relation (2.6) is satisfied. Now for $\beta = \{1\}$ and for $x \in (0, 1)$ we have $\mathbb{P}(\Theta_1 > x) = 1 - x$ and $\mathbb{P}(Z_1 > x) = \mathbb{P}(Z_1 > x, Z_2 = 0) + \mathbb{P}(Z_1 > x, Z_2 > 0) = 1/4 + (1 - x)/2$. Hence we have to prove that the sum in (2.6) adds up to $-1/4 + (1 - x)/2$. For $\gamma = \beta$ we obtain that $H_{\beta,\beta}(x, u) = \mathbb{P}(Z_1 + Z_2 > x, Z_2 \leq u) = \mathbb{P}(Z_2 \leq u)$ which is constant with respect to x and satisfies

$$H_{\beta,\beta}(x, \epsilon) - H_{\beta,\beta}(x, 0) = \mathbb{P}(0 < Z_2 \leq \epsilon) = \epsilon/2\,.$$

This implies that $\mathrm{d}H_{\beta,\beta}(x, 0)(x - 1, -1/2) = -1/4$. For $\gamma = \{1, 2\}$ we have the relation $H_{\beta,\gamma}(x, u) = \mathbb{P}(Z_1 > x, Z_2 > u) = G_\gamma(x, u)$ which has already been studied above with the case $\beta = \{1, 2\}$. This found that $\mathrm{d}G_\gamma(x, u) = (-1/2, -1/2)$, and thus $\mathrm{d}G_\gamma(x, 0)(x - 1/2, -1/2) = (1 - x)/2$. Hence we proved that

$$\mathrm{d}H_{\beta,\beta}(x, 0) \cdot (x - 1, -1/2) + \mathrm{d}H_{\beta,\{1,2\}}(x, 0) \cdot (x - 1/2, -1/2) = -1/4 + (1 - x)/2\,.$$

C Appendix C: Proofs

C.1 Some results on the projection

We start with this section with three Lemmas which gather some useful properties satisfied by the projection π.

Lemma 1 (Iteration of the projection)**.** *If $0 < z \leq z'$, then $\pi_z \circ \pi_{z'} = \pi_z$.*

Proof of Lemma 1. The proof of this result relies on the relation $\pi_z(\mathbf{v}) = z\pi(\mathbf{v}/z)$ and on the characterization (A.2).

First we simplify the problem via the equivalences

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

So we fix $a \geq 1$ and $\mathbf{u} \in \mathbb{R}_+^d$ and we prove this last equality by proving first that $\rho_{a\pi(\mathbf{u})} = \rho_{a\mathbf{u}}$ and second that the positive coordinates of both vectors $\pi(a\pi(\mathbf{u}))$ and $\pi(a\mathbf{u})$ coincide.

STEP 1: We prove that $\rho_{a\pi(\mathbf{u})} = \rho_{a\mathbf{u}}$.

The characterization (A.2) entails that a coordinate j satisfies $\pi(a\pi(\mathbf{u}))_j > 0$ if and only if

$$a\pi(\mathbf{u})_j > \frac{1}{\sum_{k\le d} \mathbb{1}_{a\pi(\mathbf{u})_k \ge a\pi(\mathbf{u})_j}} \Big(\sum_{k=1}^{d} a\pi(\mathbf{u})_k \mathbb{1}_{a\pi(\mathbf{u})_k \ge a\pi(\mathbf{u})_j} - 1 \Big) . \tag{C.1}$$

Since $a \ge 1$ this assumption holds only if $\pi(\mathbf{u})_j = u - \lambda_{\mathbf{u}} > 0$. Hence, since π preserves the order of the coordinates, we obtain that (C.1) is equivalent to

$$a(u_j - \lambda_{\mathbf{u}}) > \frac{1}{\sum_{k\le d} \mathbb{1}_{au_k \ge au_j}} \Big(\sum_{k=1}^{d} a(u_k - \lambda_{\mathbf{u}}) \mathbb{1}_{au_k \ge au_j} - 1 \Big) . \tag{C.2}$$

The terms with $\lambda_{\mathbf{u}}$ vanish and (C.2) is then equivalent to $\pi(a\mathbf{u})_j > 0$. Hence the vectors $\pi(a\pi(\mathbf{u}))$ and $\pi(a\pi(\mathbf{u}))$ have the same positive coordinates, i.e. $\rho_{a\pi(\mathbf{u})} = \rho_{a\mathbf{u}}$.

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

Let us denote by γ the set of coordinates j such that $\pi(a\pi(\mathbf{u}))_j > 0$. STEP 1 ensures that it corresponds to the set of coordinates j such that $\pi(a\mathbf{u})_j > 0$. We prove that these both components coincide. Recall that for $k \in \gamma$ we have $\pi(\mathbf{u})_k = u_k - \lambda_{\mathbf{u}} > 0$. Then the result follows from equalities

$$\pi(a\pi(\mathbf{u}))_j = a\pi(\mathbf{u})_j - \frac{|a\pi(\mathbf{u})_\gamma| - 1}{|\gamma|} = a(u_j - \lambda_{\mathbf{u}}) - \frac{|a\mathbf{u}_\gamma| - a\lambda_{\mathbf{u}}|\gamma| - 1}{|\gamma|} = au_j - \frac{|a\mathbf{u}_\gamma| - 1}{|\gamma|} = \pi(a\mathbf{u})_j ,$$

for $j \in \gamma$. □

In the following lemma we compare the behavior of the vectors $\mathbf{u}/|\mathbf{u}|$ and $\pi(\mathbf{u}/t)$ on the sets $A_{\mathbf{x}}$.

Lemma 2 (Euclidean projection and self-normalization)**.** *Let $t > 0$, $\epsilon > 0$ small enough, $\gamma \in \mathcal{P}_d^*$, and $\mathbf{x} \in B_+^d(0,1)$. We consider $\mathbf{u} \in \mathbb{R}_+^d$ such that $|\mathbf{u}|/t \in (1, 1+\epsilon]$ and $\pi(\mathbf{u}/t) \in C_\gamma$.*

1. *If $\mathbf{u}/|\mathbf{u}| \in A_{\mathbf{x}}$, then $\pi(\mathbf{u}/t) \in A_{\mathbf{x}-\epsilon/|\gamma|}$. In particular, this holds only for $\{j : x_j > 0\} \subset \gamma$.*
2. *If $\pi(\mathbf{u}/t) \in A_{\mathbf{x}(1+\epsilon)}$, then $\mathbf{u}/|\mathbf{u}| \in A_{\mathbf{x}}$.*

Proof of Lemma 2.

1. The assumption $\pi(\mathbf{u}/t) \in C_\gamma$ implies that $\lambda_{\mathbf{u}/t} = (|\mathbf{u}_\gamma|/t - 1)/|\gamma|$ so that for any $j \le d$ we have

$$\pi(\mathbf{u}/t)_j = \max(u_j/t - \lambda_{\mathbf{u}/t}, 0) \ge \frac{u_j}{t} - \frac{|\mathbf{u}_\gamma|/t - 1}{|\gamma|} > \frac{u_j}{|\mathbf{u}|} - \frac{|\mathbf{u}|/t - 1}{|\gamma|} \ge \frac{u_j}{|\mathbf{u}|} - \frac{\epsilon}{|\gamma|} ,$$

where we used that $1 < |\mathbf{u}|/t \le 1 + \epsilon$. Then the assumption $\mathbf{u}/|\mathbf{u}| \in A_{\mathbf{x}}$ concludes the proof.

2. If $\pi(\mathbf{u}/t) \in A_{\mathbf{x}(1+\epsilon)}$, then we have the inequality

$$\frac{u_j}{|\mathbf{u}|} = \frac{t}{|\mathbf{u}|}\frac{u_j}{t} \ge \frac{1}{1+\epsilon}\pi(\mathbf{u}/t)_j \ge x_j , \quad 1 \le j \le d .$$

□

For $\gamma \in \mathcal{P}_d^*$ recall that the function $\phi_\gamma : \mathbb{S}_+^{d-1} \to \mathbb{R}_+^d$ is defined by

$$\phi_\gamma(\mathbf{u})_j = \begin{cases} u_j + \frac{|\mathbf{u}_{\gamma^c}|}{|\gamma|} , & j \in \gamma , \\ u_j + \frac{|\mathbf{u}_{\gamma^c \setminus \{j\}}|}{|\gamma|+1} , & j \in \gamma^c . \end{cases}$$

Besides, we have defined the quantities $v_{\beta,i} = \sum_{j\in\beta}(v_j - v_i)$ and $v_{\beta,i,+} = \sum_{j\in\beta}(v_j - v_i)_+$ for $\mathbf{v} \in \mathbb{S}_+^{d-1}$, $\beta \in \mathcal{P}_d^*$, and $1 \le i \le d$.

Lemma 3 (Euclidean projection and sparsity)**.** *Let $\beta \in \mathcal{P}_d^*$.*

1. *For $\mathbf{v} \in \mathbb{R}_+^d$ we have the equivalences*

$$\pi(\mathbf{v})_{\beta^c} = 0 \quad \textit{if and only if} \quad 1 \leq \min_{i \in \beta^c} v_{\beta,i,+}\,, \tag{C.3}$$

and

$$\pi(\mathbf{v}) \in C_\beta \quad \textit{if and only if} \quad \begin{cases} \max_{i\in\beta} v_{\beta,i} < 1\,, \\ \min_{i\in\beta^c} v_{\beta,i} \geq 1\,. \end{cases} \tag{C.4}$$

2. *For $\mathbf{x} \in \mathcal{X}_\beta$, $\gamma \supset \beta$, $\mathbf{u} \in \mathbb{S}_+^{d-1}$, and $a \geq 1$ we have the equivalence*

$$\pi(a\mathbf{u}) \in A_{\mathbf{x}} \cap C_\gamma \quad \textit{if and only if} \quad \begin{cases} \phi_\gamma(\mathbf{u})_j \geq \frac{1}{a}\big(x_j + \frac{a-1}{|\gamma|}\big)\,, \quad j \in \beta\,, \\ \min_{j\in\gamma\setminus\beta} \phi_\gamma(\mathbf{u})_j > \frac{a-1}{a|\gamma|}\,, \\ \max_{j\in\gamma^c} \phi_\gamma(\mathbf{u})_j \leq \frac{a-1}{a(|\gamma|+1)}\,, \end{cases} \tag{C.5}$$

Proof of Lemma 3. 1. The characterization (A.2) ensures that $\pi(\mathbf{v})_i = 0$ if and only if

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

which can be rephrased as

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

This proves (C.3).

For (C.4) the assumption $\pi(\mathbf{v}) \in C_\beta$ can be rephrased as follows:

$$\forall i \in \beta,\, v_i = \pi(\mathbf{v})_i + \big(|\mathbf{v}_\beta| - 1\big)/|\beta| \quad \text{and} \quad \forall i \in \beta^c,\, v_i \leq \big(|\mathbf{v}_\beta| - 1\big)/|\beta|\,.$$

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 $\min_{i\in\beta^c} \sum_{j\in\beta}(v_j - v_i) \geq 1$. This proves (C.4).
2. By definition of the projection π the equivalence holds:

$$\begin{cases} \pi(a\mathbf{u})_j \geq x_j\,, \quad j \in \beta\,, \\ \pi(a\mathbf{u})_j > 0\,, \quad j \in \gamma \setminus \beta\,, \\ \pi(a\mathbf{u})_j = 0\,, \quad j \in \gamma^c\,, \end{cases} \quad \Longleftrightarrow \quad \begin{cases} au_j - \frac{|a\mathbf{u}_\gamma| - 1}{|\gamma|} \geq x_j\,, \quad j \in \beta\,, \\ au_j - \frac{|a\mathbf{u}_\gamma| - 1}{|\gamma|} > 0\,, \quad j \in \gamma \setminus \beta\,, \\ au_j \leq \frac{|a\mathbf{u}_\gamma| - 1}{|\gamma|}\,, \quad j \in \gamma^c\,. \end{cases}$$

Then, we write $a|\mathbf{u}_\beta| = a - a|\mathbf{u}_{\beta^c}|$ and the former conditions are equivalent to

$$\begin{cases} au_j + \frac{a|\mathbf{u}_{\gamma^c}|}{|\gamma|} \geq x_j + \frac{a-1}{|\gamma|}\,, \quad j \in \beta\,, \\ au_j + \frac{a|\mathbf{u}_{\gamma^c}|}{|\gamma|} > \frac{a-1}{|\gamma|}\,, \quad j \in \gamma \setminus \beta\,, \\ au_j + \frac{a|\mathbf{u}_{\gamma^c}|}{|\gamma|} \leq \frac{a-1}{|\gamma|}\,, \quad j \in \gamma^c\,. \end{cases}$$

We conclude the proof by writing $\mathbf{u}_{\beta^c} = |\mathbf{u}_{\beta^c \setminus \{j\}}| + u_j$ for $j \in \gamma^c$. □

C.2 Proof of Proposition 1

This proof is a consequence of Lemma 1. Indeed for $r \geq 1$, $t > 0$, and $A \subset \mathbb{S}_+^{d-1}$ we have

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

where the last equality results from Lemma 1. Then, when $t \to \infty$ the continuity of $\pi_{1/r}$ and π entails that

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

and applying again Lemma 1 concludes the proof.

C.3 Proof of Corollary 1

Let $\mathbf{a} \in \mathbb{S}_+^{d-1}$ such that $\mathbb{P}(\mathbf{Z} = \mathbf{a}) > 0$ and define $\beta = \{j : a_j > 0\}$. The goal is to prove that $\mathbf{a} = \mathbf{e}(\beta)/|\beta|$.

Since $\mathbb{P}(\mathbf{Z} = \mathbf{a} \mid Y > r) \to \mathbb{P}(\mathbf{Z} = \mathbf{a})$ when $r \to 1$, there exists $r_0 > 1$ such that $\mathbb{P}(\mathbf{Z} = \mathbf{a} \mid Y > r) > 0$ for all $r \in (1, r_0)$. Proposition 1 and Lemma 3 then imply that for all $r < r_0$ we have

$$0 < \mathbb{P}(\pi(r\mathbf{Z}) = \mathbf{a}) = \mathbb{P}\Big(Z_j - \frac{|\mathbf{Z}_\beta|}{|\beta|} = \frac{a_j - 1/|\beta|}{r} \text{ for } j \in \beta,\, Z_j - \frac{|\mathbf{Z}_\beta|}{|\beta|} \leq \frac{-1}{|\beta| r} \text{ for } j \in \beta^c\Big)$$

If there exists j such that $a_j \neq 1/|\beta|$, then the quantities $\frac{a_j - 1/|\beta|}{r}$ are all distinct when r varies in $(1, r_0)$. This contradicts the fact that $\mathbf{Z}$ has a discrete distribution. Hence we have $a_j = 1/|\beta|$ for all $j \in \beta$.

C.4 Proof of Theorem 1

C.4.1 The distribution of Z in terms of Θ

We consider $\beta \in \mathcal{P}_d^*$ and $\mathbf{x} \in \mathcal{X}_\beta^0$ such that $x_j \neq 1/|\beta|$ for all $j \in \beta$. We define the quantities $B_j = [(|\beta|\Theta_j - |\mathbf{\Theta}_\beta|)/(|\beta|x_j - 1)]_+$ for $j \in \beta$. Then we obtain that

$$\begin{aligned}G_\beta(\mathbf{x}) &= \mathbb{P}(\pi(Y\mathbf{\Theta})_\beta > \mathbf{x}, \pi(Y\mathbf{\Theta})_{\beta^c} = \mathbf{0}_{\beta^c}) \\ &= \mathbb{P}(Y\Theta_j - (Y|\mathbf{\Theta}_\beta| - 1)/|\beta| > x_j \text{ for } j \in \beta,\, Y\Theta_j \leq (Y|\mathbf{\Theta}_\beta| - 1)/|\beta| \text{ for } j \in \beta^c).\end{aligned}$$

For $j \in \beta_+$ the condition $Y\Theta_j - (Y|\mathbf{\Theta}_\beta| - 1)/|\beta| > x_j$ is equivalent to $B_j > 1/Y$, whereas for $j \in \beta_-$ it is equivalent to $B_j < 1/Y$. Moreover, for $j \in \beta^c$ the condition $Y\Theta_j \leq (Y|\mathbf{\Theta}_\beta| - 1)/|\beta|$ is equivalent to $|\mathbf{\Theta}_\beta| - |\beta|\Theta_j \geq 1/Y$. All in all we obtain that $G_\beta(\mathbf{x})$ is equal to

$$\mathbb{P}\big(Y^{-\alpha} < B_j^\alpha \text{ for } j \in \beta_+,\ Y^{-\alpha} > B_j^\alpha \text{ for } j \in \beta_-,\ Y^{-\alpha} \leq (|\mathbf{\Theta}_\beta| - |\beta|\Theta_j)_+^\alpha \text{ for } j \in \beta^c\big)\,.$$

Since the random variable $Y^{-\alpha}$ follows a uniform distribution on $(0, 1)$ and is independent of $\mathbf{\Theta}$ we obtain the desired result:

$$\begin{aligned}G_\beta(\mathbf{x}) &= \int_0^1 \mathbb{P}\Big(u < \min_{j \in \beta_+} B_j^\alpha,\, u > \max_{j \in \beta_-} B_j^\alpha,\, u \leq \min_{j \in \beta^c}(|\mathbf{\Theta}_\beta| - |\beta|\Theta_j)_+^\alpha\Big)\,\mathrm{d}u \\ &= \int_0^1 \mathbb{P}\Big(\max_{j \in \beta_-} B_j^\alpha < u < \min_{j \in \beta_+} B_j^\alpha \wedge \min_{j \in \beta^c}(|\mathbf{\Theta}_\beta| - |\beta|\Theta_j)_+^\alpha\Big)\,\mathrm{d}u \\ &= \mathbb{E}\Big[\Big(1 \wedge \min_{j \in \beta_+} B_j^\alpha \wedge \min_{j \in \beta^c}(|\mathbf{\Theta}_\beta| - |\beta|\Theta_j)_+^\alpha - \max_{j \in \beta_-} B_j^\alpha\Big)_+\Big]\,.\end{aligned}$$

C.4.2 Equivalence of regular variation and sparse regular variation

The proof of this result is divided into two steps. The first one consists in characterizing regular variation via the convergence of $\mathbb{P}(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t)$ when $t \to \infty$ and $\epsilon \to 0$. This result is stated in the following lemma.

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

(1) $\mathbf{X}$ is regularly varying with tail index α.

(2) a. $|\mathbf{X}|$ is regularly varying with tail index α.

b. For all $\beta \in \mathcal{P}_d^*$ and λ_β-almost every $\mathbf{x} \in \mathcal{X}_\beta$ the quantities

$$(\alpha\epsilon)^{-1} \limsup_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big) \tag{C.6}$$

and

$$(\alpha\epsilon)^{-1} \liminf_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big) \tag{C.7}$$

have a common limit $l(A_{\mathbf{x}})$ when $\epsilon > 0$ converges to 0, and the function $\mathbf{x} \mapsto l(A_{\mathbf{x}})$ is continuous at $\mathbf{x}$.

In this case, l extends to a unique probability measure on $\mathcal{B}(\mathbb{S}_+^{d-1})$ which coincides with the spectral measure of $\mathbf{X}$.

The second step then consists in proving that under assumption (A) of Theorem 1 the second assumption of Lemma 4 holds true.

Proof of Lemma 4. We first prove that (1) implies (2). If $\mathbf{X} \in \mathrm{RV}(\alpha, \mathbf{\Theta})$, then $|\mathbf{X}|$ is regularly varying with index α and Portmanteau's Theorem ensures that

$$\lim_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big) = \mathbb{P}(Y \leq 1+\epsilon)\mathbb{P}(\mathbf{\Theta} \in A_{\mathbf{x}}) = (1-(1+\epsilon)^{-\alpha})\mathbb{P}(\mathbf{\Theta} \in A_{\mathbf{x}})\,, \tag{C.8}$$

for $\epsilon > 0$, $\beta \in \mathcal{P}_d^*$, and $\mathbf{x} \in \mathcal{X}_\beta$ such that $\mathbb{P}(\mathbf{\Theta} \in \partial A_{\mathbf{x}}) = 0$. Since $\mathbb{P}(\mathbf{\Theta} \in \partial A_{\mathbf{x}}) \leq \sum_{i\in\beta} \mathbb{P}(\Theta_i = x_i)$, the set of all $\mathbf{x} \in \mathcal{X}_\beta$ for which the convergence (C.8) does not hold is at most countable and thus is λ_β-negligible. After dividing both sides of (C.8) by $\alpha\epsilon$ and taking the limit when ϵ converges to 0 we obtain the convergence to $\mathbb{P}(\mathbf{\Theta} \in A_{\mathbf{x}})$.

We now prove that (2) implies (1). We consider $\beta \in \mathcal{P}_d^*$ and denote by $\bar{\mathcal{X}}_\beta$ the set of all $\mathbf{x} \in \mathcal{X}_\beta$ such that the common limit $l(A_{\mathbf{x}})$ of (C.6) and (C.7) exists and is continuous. We also define $\bar{\mathcal{X}} = \{\mathbf{x} \in B_+^d(0,1) : (\mathbf{x}_\beta, \mathbf{0}_{\beta^c}) \in \mathcal{X}_\beta \text{ for all } \beta\}$.

For $\epsilon > 0$ and $u > 1$ we decompose the interval (u, ∞) as follows:

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

where $\bigsqcup$ denotes a disjoint union. Then we obtain for $\mathbf{x} \in \bar{\mathcal{X}}$ and $t > 0$,

$$\begin{aligned}
&\mathbb{P}\big(|\mathbf{X}|/t > u,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big) \qquad\qquad (\text{C.9})\\
&\quad= \sum_{k=0}^{\infty} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \in (1, 1+\epsilon],\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > t\right)\\
&\quad= \sum_{k=0}^{\infty} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \,\middle|\, |\mathbf{X}| > tu(1+\epsilon)^k\right) \frac{\mathbb{P}(|\mathbf{X}| > tu(1+\epsilon)^k)}{\mathbb{P}(|\mathbf{X}| > t)}\,.
\end{aligned}$$

Fatou's Lemma together with the fact that $|\mathbf{X}|$ is regularly varying with tail index α imply the two following inequalities:

$$\begin{aligned}
&\liminf_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t > u, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big) \\
&\qquad \geq \sum_{k=0}^{\infty} \big(u(1+\epsilon)^k\big)^{-\alpha} \liminf_{t\to\infty} \mathbb{P}\left(\frac{|\mathbf{X}|}{tu(1+\epsilon)^k} \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \,\middle|\, \frac{|\mathbf{X}|}{u(1+\epsilon)^k} > t\right) \\
&\qquad = \sum_{k=0}^{\infty} \big(u(1+\epsilon)^k\big)^{-\alpha} \liminf_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big) \\
&\qquad = \frac{u^{-\alpha}}{1-(1+\epsilon)^{-\alpha}} \liminf_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big)\,,
\end{aligned}$$

and

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

Finally, we use Equations (C.6) and (C.7) and the relation $1-(1+\epsilon)^{-\alpha} \sim \alpha\epsilon$ which entail that

$$\begin{aligned}
u^{-\alpha} l(A_{\mathbf{x}}) \leq \liminf_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t > u,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big) \\
\leq \limsup_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t > u,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big) \leq u^{-\alpha} l(A_{\mathbf{x}})\,.
\end{aligned}$$

This proves that

$$\lim_{t\to\infty} \mathbb{P}\big(|\mathbf{X}|/t > u,\, \mathbf{X}/|\mathbf{X}| \in A \mid |\mathbf{X}| > t\big) = u^{-\alpha} l(A)\,, \tag{C.10}$$

for all $A = A_{\mathbf{x}}$ such that $\mathbf{x} \in \bar{\mathcal{X}}$. The convergence also holds true for $A = \mathbb{S}_+^{d-1}$ (resp. $A = \emptyset$) with $l(\mathbb{S}_+^{d-1}) = 1$ (resp. $l(\emptyset) = 0$). Then, by inclusion exclusion we obtain the convergence of $\mathbb{P}\big(\mathbf{X}/|\mathbf{X}| \leq \mathbf{x} \mid |\mathbf{X}| > t\big)$ for all $\mathbf{x} \in \bar{\mathcal{X}}$ and we denote by $F(\mathbf{x})$ this limit which is continuous at any point $\mathbf{x} \in \bar{\mathcal{X}}$. In particular this implies the convergence $\mathbb{P}(X_j/|\mathbf{X}| \leq x \mid |\mathbf{X}| > t) \to 1 - l(A_{x\mathbf{e}_j}) =: L_j(x)$ for almost every $x \in (0,1]$. The functions L_j are non-decreasing and continuous at almost every $x \in (0,1]$, thus we extend it to a right continuous function on $[0,1]$. Then F is continuous from above (see the definition page 177 in Billingsley (1995)). For $\mathbf{a} < \mathbf{b} \in \bar{\mathcal{X}}$ we have the inequality $\mathbb{P}\big(\mathbf{a} < \mathbf{X}/|\mathbf{X}| \leq \mathbf{b} \mid |\mathbf{X}| > t\big) \geq 0$ which implies that

$$\Delta_{(\mathbf{a},\mathbf{b}]} F := \sum_{\mathbf{u} \in \mathcal{V}(\mathbf{a},\mathbf{b})} \operatorname{sign}(\mathbf{u}) F(\mathbf{u}) \geq 0\,, \tag{C.11}$$

where $\mathcal{V}(\mathbf{a},\mathbf{b})$ denotes all the vertex of $(\mathbf{a},\mathbf{b}]$, and $\operatorname{sign}(\mathbf{v}) = 1$ if $u_k = a_k$ for an even number of k's and $\operatorname{sign}(\mathbf{v}) = -1$ if $u_k = a_k$ for an odd number of k's. Since F is continuous from above, the inequality (C.11) holds true for all $\mathbf{a} < \mathbf{b} \in B_+^d(0,1)$. Then, Billingsley (1995) Theorem 12.5 ensures that there exists a measure S on $\mathbb{R}^d$ such that $S((\mathbf{a},\mathbf{b}]) = \Delta_{(\mathbf{a},\mathbf{b}]} F$. It is straightforward to see that S is a probability measure with support in $\mathbb{S}_+^{d-1}$ and that its restriction to $\mathbb{S}_+^{d-1}$ coincides with l. Thus Equation (C.10) can be rewritten as

$$\mathbb{P}\big(|\mathbf{X}|/t > u,\, \mathbf{X}/|\mathbf{X}| \in \cdot \mid |\mathbf{X}| > t\big) \xrightarrow{d} u^{-\alpha} l(\cdot)\,, \quad t \to \infty\,, \quad u > 1\,.$$

and therefore proves that $\mathbf{X}$ is regularly varying with tail index α and spectral measure $l(\cdot)$. $\square$

Our aim is now to prove that the condition (2) holds true for a sparsely regularly varying random vector which satisfies assumption (A).

Proof of Theorem 1 We consider a random vector $\mathbf{X} \in \mathrm{SRV}(\alpha, \mathbf{Z})$ which satisfies assumption (A). For $\beta \in \mathcal{P}_d^*$ and $\mathbf{x} \in \mathcal{X}_\beta$ we decompose the probability $\mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big)$ as follows:

$$\sum_{\gamma \supset \beta} \mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}},\, \pi(\mathbf{X}/t) \in C_\gamma \mid |\mathbf{X}| > t\big)\,,$$

where the restriction of the sum to the directions $\gamma \supset \beta$ is a consequence of the first point of Lemma 2. Moreover, for $\gamma \supset \beta$ Lemma 2 then ensures that the probability $\mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}},\, \pi(\mathbf{X}/t) \in C_\gamma \mid |\mathbf{X}| > t\big)$ is upper-bounded by

$$\mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \pi(\mathbf{X}/t) \in A_{\mathbf{x}-\epsilon/|\gamma|} \cap C_\gamma \mid |\mathbf{X}| > t\big)\,, \tag{C.12}$$

assuming that $\epsilon > 0$ is small enough so that the vector $\mathbf{x} - \epsilon/|\gamma|$ has positive coordinates, and lower-bounded by

$$\mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \pi(\mathbf{X}/t) \in A_{\mathbf{x}(1+\epsilon)} \cap C_\gamma \mid |\mathbf{X}| > t\big)\,. \tag{C.13}$$

In order to deal simultaneously with both probabilities in (C.12) and (C.13) we write

$$\mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon,\, \pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma \mid |\mathbf{X}| > t\big)\,, \tag{C.14}$$

where $\psi_\epsilon(\mathbf{x})$ is either equal to $\mathbf{x} - \epsilon/|\gamma|$ or to $\mathbf{x}(1+\epsilon)$.

The proof is then divided into two steps. The first step consists in proving that these two probabilities (C.12) and (C.13) converge when $t \to \infty$ for λ_β-almost every $\mathbf{x} \in \mathcal{X}_\beta$. Then, the goal is to prove that after a division by ϵ these two limits converge to the same quantity when $\epsilon \to 0$. We extract the first step as a separate lemma.

Lemma 5. *For all $\beta \subset \gamma \in \mathcal{P}_d^*$, for λ_β-almost every $\mathbf{x} \in \mathcal{X}_\beta$, for almost every $\epsilon > 0$ we have the following convergence:*

$$\lim_{t\to\infty} \mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t > 1+\epsilon \mid |\mathbf{X}| > t\big) = \mathbb{P}(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, Y > 1+\epsilon)\,. \tag{C.15}$$

Proof. We use Proposition 1 which entails that the right-hand side in (C.15) is equal to

$$\begin{aligned}\mathbb{P}(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, Y > 1+\epsilon) &= \mathbb{P}(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma \mid Y > 1+\epsilon)\mathbb{P}(Y > 1+\epsilon)\\ &= \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma)(1+\epsilon)^{-\alpha}\,.\end{aligned}$$

Regarding the left-hand side in (C.15), it can be rewritten as follows:

$$\begin{aligned}&\mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t > 1+\epsilon \mid |\mathbf{X}| > t\big)\\ &\qquad = \mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma \mid |\mathbf{X}| > t(1+\epsilon)\big)\frac{\mathbb{P}(|\mathbf{X}| > t(1+\epsilon))}{\mathbb{P}(|\mathbf{X}| > t)}\,.\end{aligned}$$

The ratio $\mathbb{P}(|\mathbf{X}| > t(1+\epsilon))/\mathbb{P}(|\mathbf{X}| > t)$ converges to $(1+\epsilon)^{-\alpha}$ when $t \to \infty$ since $|\mathbf{X}|$ is regularly varying with tail index α. Besides the probability $\mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma \mid |\mathbf{X}| > t(1+\epsilon)\big)$ converges when $t \to \infty$ if and only if $\mathbb{P}\big(\pi((1+\epsilon)\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma \mid |\mathbf{X}| > t\big)$ converges when $t \to \infty$, and then both probabilities have the same limit. We use Lemma 1 and the relation $\pi_z(\mathbf{v}) = z\pi(\mathbf{v}/z)$ for $\mathbf{v} \in \mathbb{R}_+^d$ and $z > 0$ which entail that

$$\pi((1+\epsilon)\mathbf{X}/t) = (1+\epsilon)\pi_{1/(1+\epsilon)}(\mathbf{X}/t) = (1+\epsilon)\pi_{1/(1+\epsilon)}(\pi(\mathbf{X}/t)) = \pi((1+\epsilon)\pi(\mathbf{X}/t))\,.$$

Hence the convergence in Equation (C.15) holds if and only if

$$\mathbb{P}\big(\pi((1+\epsilon)\pi(\mathbf{X}/t)) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma \mid |\mathbf{X}| > t\big) \to \mathbb{P}(\pi((1+\epsilon)\mathbf{Z}) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma)\,, \quad t \to \infty\,. \tag{C.16}$$

The equivalence (C.5) then entails that the former convergence holds if and only if the probability

$$\mathbb{P}\bigg(\phi_\gamma(\pi(\mathbf{X}/t))_\beta \geq \frac{\psi_\epsilon(\mathbf{x})}{1+\epsilon} + \frac{\epsilon}{(1+\epsilon)|\gamma|},\, \min_{i\in\gamma\setminus\beta} \phi_\gamma(\pi(\mathbf{X}/t))_i > \frac{\epsilon}{(1+\epsilon)|\gamma|}, \\ \max_{j\in\gamma^c} \phi_\gamma(\pi(\mathbf{X}/t))_j \leq \frac{\epsilon}{(1+\epsilon)(|\gamma|+1)} \,\bigg|\, |\mathbf{X}| > t\bigg)$$

converges to

$$\mathbb{P}\bigg(\phi_\gamma(\mathbf{Z})_\beta \geq \frac{\psi_\epsilon(\mathbf{x})}{1+\epsilon} + \frac{\epsilon}{(1+\epsilon)|\gamma|},\, \min_{i\in\gamma\setminus\beta} \phi_\gamma(\mathbf{Z})_i > \frac{\epsilon}{(1+\epsilon)|\gamma|},\, \max_{j\in\gamma^c} \phi_\gamma(\mathbf{Z})_j \leq \frac{\epsilon}{(1+\epsilon)(|\gamma|+1)}\bigg)\,,$$

when $t \to \infty$.

For $\gamma \supset \beta$ the set of all $\epsilon > 0$ for which there exists $\mathbb{P}(\min_{i\in\gamma\setminus\beta} \phi_\gamma(\mathbf{Z})_i = \epsilon/[(1+\epsilon)|\gamma|]) > 0$ or for which $\mathbb{P}(\max_{j\in\gamma^c} \phi_\gamma(\mathbf{Z})_j = \epsilon/[(1+\epsilon)(|\gamma|+1)]) > 0$ is at most countable so that for almost every $\epsilon > 0$ we have

$$\mathbb{P}\Big(\min_{j\in\gamma\setminus\beta} \phi_\gamma(\mathbf{Z})_j = \frac{\epsilon}{(1+\epsilon)|\gamma|}\Big) = \mathbb{P}\Big(\max_{i\in\gamma^c} \phi_\gamma(\mathbf{Z})_i = \frac{\epsilon}{(1+\epsilon)(|\gamma|+1)}\Big) = 0\,, \text{ for all } \gamma \supset \beta\,.$$

Let us denote by $\mathcal{E}_\beta$ the set of all these $\epsilon > 0$. The same argument implies that for λ_β-almost every $\mathbf{x} \in \mathcal{X}_\beta$ we have

$$\mathbb{P}\Big(\phi_\gamma(\mathbf{Z})_j = \frac{\psi_\epsilon(\mathbf{x})_j}{1+\epsilon} + \frac{\epsilon}{(1+\epsilon)|\gamma|}\Big) = 0, \quad \text{for all } j \in \beta\,, \quad \gamma \supset \beta, \text{ and } \epsilon \in \mathcal{E}_\beta\,.$$

Let us denote by $\bar{\mathcal{X}}_\beta$ the set of all these $\mathbf{x}$. We have proved that for all $\epsilon \in \mathcal{E}_\beta$ and $\mathbf{x} \in \bar{\mathcal{X}}_\beta$ we have the convergence

$$\lim_{t\to\infty} \mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t > 1+\epsilon \mid |\mathbf{X}| > t\big) = \mathbb{P}(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, Y > 1+\epsilon)\,.$$

This concludes the proof of the Lemma 5. □

End of the proof of Theorem 1 We use the same notation $\bar{\mathcal{X}}_\beta$ and $\mathcal{E}_\beta$ as in the proof of Lemma 5. We start with the following decomposition:

$$\mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t \leq 1+\epsilon \mid |\mathbf{X}| > t\big) = \mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t \leq 1+\epsilon^2 \mid |\mathbf{X}| > t\big) \\ + \mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t \in (1+\epsilon^2, 1+\epsilon] \mid |\mathbf{X}| > t\big)\,.$$

The introduction of the term with ϵ^2 on the right-hand side follows from the fact that the probability $\mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma \mid |\mathbf{X}| > t\big)$ may not converge. The latter probability is bounded by $\mathbb{P}\big(|\mathbf{X}|/t \leq 1+\epsilon^2 \mid |\mathbf{X}| > t\big) \to 1-(1+\epsilon^2)^{-\alpha} = o(\epsilon)$. Moreover, Lemma 5 ensures that

$$\mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t \in (1+\epsilon^2, 1+\epsilon] \mid |\mathbf{X}| > t\big) \to \mathbb{P}\big(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, Y \in (1+\epsilon^2, 1+\epsilon]\big)\,,$$

when $t \to \infty$ for $\epsilon, \epsilon^2 \in \mathcal{E}_\beta$ and $\mathbf{x} \in \bar{\mathcal{X}}_\beta$. Hence we obtain

$$\lim_{t\to\infty} \mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t \leq 1+\epsilon \mid |\mathbf{X}| > t\big) \\ = \lim_{t\to\infty} \mathbb{P}\big(\pi(\mathbf{X}/t) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, |\mathbf{X}|/t \in (1+\epsilon^2, 1+\epsilon] \mid |\mathbf{X}| > t\big) + o(\epsilon)$$

$$= \mathbb{P}\big(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, Y \in (1+\epsilon^2, 1+\epsilon]\big) + o(\epsilon)$$
$$= \mathbb{P}\big(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, Y \le 1+\epsilon\big) + o(\epsilon)\,.$$

Then, Proposition 1 implies that

$$\begin{aligned}
&\mathbb{P}\big(\mathbf{Z}_\beta \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma,\, Y \le 1+\epsilon\big) \\
&\qquad = \mathbb{P}\big(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma\big) - \mathbb{P}\big(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma \mid Y > 1+\epsilon\big)(1+\epsilon)^{-\alpha} \\
&\qquad = \mathbb{P}\big(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma\big) - \mathbb{P}\big(\pi((1+\epsilon)\mathbf{Z}) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma\big)(1+\epsilon)^{-\alpha} \\
&\qquad = \big[1-(1+\epsilon)^{-\alpha}\big]\mathbb{P}\big(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma\big) \qquad \text{(C.17)} \\
&\qquad\qquad + (1+\epsilon)^{-\alpha}\Big[\mathbb{P}\big(\mathbf{Z} \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma\big) - \mathbb{P}\big(\pi((1+\epsilon)\mathbf{Z}) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma\big)\Big]\,. \qquad \text{(C.18)}
\end{aligned}$$

The sets $A_{\psi_\epsilon(\mathbf{x})}$ decrease when $\epsilon \to 0$ and satisfy $\cap_{\epsilon>0} A_{\psi_\epsilon(\mathbf{x})} = A_{\mathbf{x}}$. Hence the term in (C.17) divided by ϵ converges to $\alpha\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}} \cap C_\gamma)$. For the term in (C.18) we use (C.5) which states that the event $\{\pi((1+\epsilon)\mathbf{Z}) \in A_{\psi_\epsilon(\mathbf{x})} \cap C_\gamma\}$ corresponds to

$$\begin{cases}
\phi_\gamma(\mathbf{Z})_j \ge \frac{\psi_\epsilon(\mathbf{x})_j}{(1+\epsilon)} + \frac{\epsilon}{|\gamma|(1+\epsilon)}\,, \quad j \in \beta\,, \\
\min_{j\in\gamma\setminus\beta} \phi_\gamma(\mathbf{Z})_j > \frac{\epsilon}{|\gamma|(1+\epsilon)} \\
\max_{j\in\gamma^c} \phi_\gamma(\mathbf{Z})_j \le \frac{\epsilon}{(|\gamma|+1)(1+\epsilon)}\,.
\end{cases}$$

Hence the difference of probabilities in (C.18) corresponds to the difference

$$H_{\beta,\gamma}\big(\psi_\epsilon(\mathbf{x}), 0, 0\big) - H_{\beta,\gamma}\Big(\frac{\psi_\epsilon(\mathbf{x})}{(1+\epsilon)} + \frac{\epsilon}{|\gamma|(1+\epsilon)}, \frac{\epsilon}{|\gamma|(1+\epsilon)}, \frac{\epsilon}{(|\gamma|+1)(1+\epsilon)}\Big)\,.$$

After a division by ϵ this difference converges to

$$\mathrm{d}H_{\beta,\gamma}(\mathbf{x}_\beta, 0, 0) \cdot (\mathbf{x}_\beta - 1/|\gamma|, -1/|\gamma|, -1/(|\gamma|+1))\,, \quad \epsilon \to 0\,,$$

for $\psi_\epsilon(\mathbf{x}) = \mathbf{x}/(1+\epsilon)$ and $\psi_\epsilon(\mathbf{x}) = \mathbf{x} - \epsilon/|\gamma|$.

All in all, we proved that for λ_β-almost every $\mathbf{x} \in \mathcal{X}_\beta$ both limits

$$\lim_{\epsilon\to 0} \liminf_{t\to\infty} \epsilon^{-1}\mathbb{P}\big(|\mathbf{X}|/t \le 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big)$$

and

$$\lim_{\epsilon\to 0} \limsup_{t\to\infty} \epsilon^{-1}\mathbb{P}\big(|\mathbf{X}|/t \le 1+\epsilon,\, \mathbf{X}/|\mathbf{X}| \in A_{\mathbf{x}} \mid |\mathbf{X}| > t\big)$$

exist and are equal to

$$\begin{aligned}
l(A_{\mathbf{x}}) &= \sum_{\gamma\supset\beta} \alpha\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}} \cap C_\gamma) + \sum_{\gamma\supset\beta} \mathrm{d}H_{\beta,\gamma}(\mathbf{x}_\beta, 0, 0) \cdot \Big(\mathbf{x}_\beta - \frac{1}{|\gamma|}, -\frac{1}{|\gamma|}, -\frac{1}{|\gamma|+1}\Big) \\
&= \alpha\mathbb{P}(\mathbf{Z} \in A_{\mathbf{x}}) + \sum_{\gamma\supset\beta} \mathrm{d}H_{\beta,\gamma}(\mathbf{x}_\beta, 0, 0) \cdot \Big(\mathbf{x}_\beta - \frac{1}{|\gamma|}, -\frac{1}{|\gamma|}, -\frac{1}{|\gamma|+1}\Big)\,,
\end{aligned}$$

Then assumption (A) ensures that for λ_β-almost every $\mathbf{x} \in \mathcal{X}_\beta$ the function $\mathbf{x} \mapsto l(A_{\mathbf{x}})$ is continuous at $\mathbf{x}$ and Lemma 4 allows us to conclude.

C.5 Proof of Proposition 2

Recall that we have defined the quantities $\theta_{\beta,i} = \sum_{j\in\beta}(\theta_j - \theta_i)$ and $\theta_{\beta,i,+} = \sum_{j\in\beta}(\theta_j - \theta_i)_+$ for $\boldsymbol{\theta} \in \mathbb{S}_+^{d-1}$, $\beta \in \mathcal{P}_d^*$, and $1 \le i \le d$.

1. We only prove the convergence (3.2) (the proof of (3.3) is similar). For $\beta \in \mathcal{P}_d^*$, Lemma 3 ensures that $\pi(Y\boldsymbol{\Theta}) \in C_\beta$ if and only if $\max_{i\in\beta}\Theta_{\beta,i} < 1/Y$ and $\min_{i\in\beta^c}\Theta_{\beta,i} \ge 1/Y$. Hence (3.2) is equivalent to

$$\mathbb{P}\big((|\mathbf{X}|/t, \mathbf{X}/|\mathbf{X}|) \in D_\beta \mid |\mathbf{X}| > t\big) \to \mathbb{P}((Y,\boldsymbol{\Theta}) \in D_\beta)\,, \tag{C.19}$$

with

$$D_\beta = \big\{(r,\theta) \in (1,\infty)\times\mathbb{S}_+^{d-1} : \theta_{\beta,i} < 1/r \text{ for } i\in\beta, \text{ and } \theta_{\beta,i} \ge 1/r \text{ for } i \in \beta^c\big\}\,,$$

and this convergence holds if $\mathbb{P}((Y,\boldsymbol{\Theta}) \in \partial D_\beta) = 0$. The boundary ∂D_β satisfies the inequality

$$\mathbb{P}((Y,\boldsymbol{\Theta}) \in \partial D_\beta) \le \sum_{i=1}^d \mathbb{P}(\Theta_{\beta,i} = Y^{-1})\,,$$

and all the terms of the sum are null since Y is a continuous random variable independent of $\boldsymbol{\Theta}$. Thus $\mathbb{P}((Y,\boldsymbol{\Theta}) \in \partial D_\beta) = 0$, which implies that convergence (C.19) holds and then that convergence (3.2) holds as well.

2. Following Lemma 3, the probability that $\mathbf{Z}$ belongs to C_β is equal to

$$\begin{aligned}
\mathbb{P}(\mathbf{Z} \in C_\beta) &= \mathbb{P}\Big(\max_{j\in\beta}\sum_{k\in\beta}(Y\Theta_k - Y\Theta_j) < 1, \min_{j\in\beta^c}\sum_{k\in\beta}(Y\Theta_k - Y\Theta_j) \ge 1\Big)\\
&= \mathbb{P}\Big(\max_{j\in\beta}\Theta_{\beta,j,+}^\alpha < Y^{-\alpha}, \min_{j\in\beta^c}\Theta_{\beta,j,+}^\alpha \ge Y^{-\alpha}\Big)\\
&= \int_0^1 \mathbb{P}\Big(\max_{j\in\beta}\Theta_{\beta,j,+}^\alpha < u \le \min_{j\in\beta^c}\Theta_{\beta,j,+}^\alpha\Big)\,\mathrm{d}u\\
&= \mathbb{E}\Big[\Big(\min_{j\in\beta^c}\Theta_{\beta,j,+}^\alpha - \max_{j\in\beta}\Theta_{\beta,j,+}^\alpha\Big)_+\Big]\,.
\end{aligned}$$

Similarly, Equation (3.5) is also a consequence of Lemma 3 since we have the relations

$$\begin{aligned}
\mathbb{P}(\mathbf{Z}_{\beta^c} = 0) &= \mathbb{P}\Big(1 \le \min_{j\in\beta^c}\sum_{k=1}^d (Y\Theta_k - Y\Theta_j)_+\Big) = \mathbb{P}\big(Y^{-\alpha} \le \min_{j\in\beta^c}\Theta_{\beta,j,+}^\alpha\big)\\
&= \int_0^1 \mathbb{P}\big(u \le \min_{j\in\beta^c}\Theta_{\beta,j,+}^\alpha\big)\,\mathrm{d}u = \mathbb{E}\big[\min_{j\in\beta^c}\Theta_{\beta,j,+}^\alpha\big]\,.
\end{aligned} \tag{C.20}$$

This proves (3.5) and concludes the proof of the proposition.

C.6 Proof of Theorem 2

We first state an inequality which will be used to prove both results of Theorem 2.

Lemma 6. *For $\beta \in \mathcal{P}_d^*$ we have the inequality*

$$\mathbb{P}(\boldsymbol{\Theta} \in C_\beta) \le \mathbb{P}\big(\max_{j\in\beta}\Theta_{\beta,j,+} < 1\big)\,. \tag{C.21}$$

Proof of Lemma 6. The relation in (C.21) is equivalent to

$$\mathbb{P}\big(\max_{j\in\beta}\Theta_{\beta,\,j,\,+}=1\big)\leq\mathbb{P}(\boldsymbol{\Theta}\notin C_\beta)\,.$$

The probability on the left-hand side can be rewritten as

$$\mathbb{P}\big(\max_{j\in\beta}\Theta_{\beta,\,j,\,+}=1\big)=\mathbb{P}\Big(\sum_{k\in\beta}(\Theta_k-\min_{j\in\beta}\Theta_j)=1\Big)=\mathbb{P}\Big(\sum_{k\in\beta}\Theta_k=1+|\beta|\min_{j\in\beta}\Theta_j\Big)\,.$$

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

$$\mathbb{P}\Big(\max_{j\in\beta}\Theta_{\beta,\,j,\,+}=1\Big)\leq\mathbb{P}(\Theta_k=0\text{ for some }k\in\beta)\leq\mathbb{P}(\boldsymbol{\Theta}\notin C_\beta)\,,$$

which concludes the proof. □

We now move on to the proof of Theorem 2. The proof of these results relies on Equation (3.4) and Lemma 6 above.

1. If $\beta\in\mathcal{P}_d^*$ is such that $\mathbb{P}(\boldsymbol{\Theta}\in C_\beta)>0$, then Equation (3.4) entails that

$$\begin{aligned}\mathbb{P}(\mathbf{Z}\in C_\beta)&\geq\mathbb{E}\Big[\Big(\min_{j\in\beta^c}\Theta^\alpha_{\beta,\,j,\,+}-\max_{j\in\beta}\Theta^\alpha_{\beta,\,j,\,+}\Big)_+\mathbb{1}_{\boldsymbol{\Theta}\in C_\beta}\Big]\\&=\mathbb{E}\Big[\Big(1-\max_{j\in\beta}\Theta^\alpha_{\beta,\,j,\,+}\Big)\mathbb{1}_{\boldsymbol{\Theta}\in C_\beta}\Big]\\&=\mathbb{E}\Big[\Big(1-\max_{j\in\beta}\Theta^\alpha_{\beta,\,j,\,+}\Big)\mid\boldsymbol{\Theta}\in C_\beta\Big]\mathbb{P}(\boldsymbol{\Theta}\in C_\beta)\,.\end{aligned}\tag{C.22}$$

The expectation is positive by Lemma 6 and the probability $\mathbb{P}(\boldsymbol{\Theta}\in C_\beta)$ is positive by assumption. This shows that $\mathbb{P}(\mathbf{Z}\in C_\beta)>0$.

2. We separately prove both implications.

- We first consider a maximal direction β of $\boldsymbol{\Theta}$. The first point of the theorem ensures that $\mathbb{P}(\mathbf{Z}\in C_\beta)>0$. Besides, if $\beta'\supsetneq\beta$, then Equation (3.8) gives

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

and this last probability equals zero since β is a maximal direction for $\boldsymbol{\Theta}$. This proves that β is a maximal direction for $\mathbf{Z}$.

- We now consider a maximal direction β of $\mathbf{Z}$. We claim that for $\beta'\supsetneq\beta$, $\mathbb{P}(\boldsymbol{\Theta}\in C_{\beta'})=0$. If not, the first point of the theorem entails that $\mathbb{P}(\mathbf{Z}\in C_{\beta'})>0$ which contradicts the maximality of β for $\mathbf{Z}$.

Secondly, Equation (3.4) gives that

$$\begin{aligned}\mathbb{P}(\mathbf{Z}\in C_\beta)&=\mathbb{E}\Big[\Big(\min_{j\in\beta^c}\Theta^\alpha_{\beta,\,j,\,+}-\max_{j\in\beta}\Theta^\alpha_{\beta,\,j,\,+}\Big)_+\Big]\\&=\mathbb{E}\Big[\Big(\min_{j\in\beta^c}\Theta^\alpha_{\beta,\,j,\,+}-\max_{j\in\beta}\Theta^\alpha_{\beta,\,j,\,+}\Big)_+\mathbb{1}_{\boldsymbol{\Theta}\in C_\beta}\Big]+\mathbb{E}\Big[\Big(\min_{j\in\beta^c}\Theta^\alpha_{\beta,\,j,\,+}-\max_{j\in\beta}\Theta^\alpha_{\beta,\,j,\,+}\Big)_+\mathbb{1}_{\boldsymbol{\Theta}\notin C_\beta}\Big]\\&=E_1+E_2\,.\end{aligned}\tag{C.23}$$

The first term E_1 has already been calculated in (C.22). It is equal to

$$E_1=\mathbb{E}\Big[1-\max_{j\in\beta}\Theta^\alpha_{\beta,\,j,\,+}\mid\boldsymbol{\Theta}\in C_\beta\Big]\mathbb{P}(\boldsymbol{\Theta}\in C_\beta)\,.$$

For the second term E_2, the assumption $\mathbf{\Theta} \notin C_\beta$ 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 E_2 into two terms:

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\min_{j\in\beta^c} \Theta^\alpha_{\beta,\,j,\,+} - \max_{j\in\beta} \Theta^\alpha_{\beta,\,j,\,+}\Big)_+ \mathbb{1}_{\mathbf{\Theta}\notin C_\beta}\Big] \\
&\leq \mathbb{E}\Big[\Big(\min_{j\in\beta^c} \Theta^\alpha_{\beta,\,j,\,+} - \max_{j\in\beta} \Theta^\alpha_{\beta,\,j,\,+}\Big)_+ \mathbb{1}_{\exists l\in\beta,\,\Theta_l=0}\Big] \\
&+ \mathbb{E}\Big[\Big(\min_{j\in\beta^c} \Theta^\alpha_{\beta,\,j,\,+} - \max_{j\in\beta} \Theta^\alpha_{\beta,\,j,\,+}\Big)_+ \mathbb{1}_{\exists \beta'\supsetneq\beta,\,\mathbf{\Theta}\in C_{\beta'}}\Big]\,.
\end{aligned}$$

The first expectation is then equal to

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

and thus vanishes. The second expectation is smaller than $\mathbb{P}(\mathbf{\Theta} \in C_{\beta'}$ for some $\beta' \supsetneq \beta)$ which is equal to zero. Indeed, if $\mathbb{P}(\mathbf{\Theta} \in C_{\beta'}$ for some $\beta' \supsetneq \beta) > 0$, then by Equation (3.4), we also have $\mathbb{P}(\mathbf{Z} \in C_{\beta'}$ for some $\beta' \supsetneq \beta) > 0$ which contradicts the maximality of β for $\mathbf{Z}$. All in all this proves that $E_2 = 0$.

Going back to Equation (C.23), we have proved that

$$\mathbb{P}(\mathbf{Z} \in C_\beta) = E_1 = \mathbb{E}\Big[1 - \max_{j\in\beta} \Theta^\alpha_{\beta,\,j,\,+} \mid \mathbf{\Theta} \in C_\beta\Big]\mathbb{P}(\mathbf{\Theta} \in C_\beta)\,.$$

By Lemma 6 we know that the expectation is positive. Hence, the assumption $\mathbb{P}(\mathbf{Z} \in C_\beta) > 0$ implies that $\mathbb{P}(\mathbf{\Theta} \in C_\beta) > 0$ which proves that β is a maximal direction of $\mathbf{\Theta}$.

Acknowledgments We are grateful to two referees for careful reading of the paper and for useful suggestions. We also would like to thank Johan Segers for insightful comments which led to complete the proof of Theorem 1.

References

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd., Chichester.

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, 3rd edition.

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.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New-York.

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 Clémençon, S. (2016). Sparse representation of multivariate extremes with applications to anomaly ranking. In *AISTATS*, pages 75–83.

Goix, N., Sabourin, A., and Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*, 161:12–31.

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

Janßen, A. and Wan, P. (2020). k-means clustering of extremes. *Electronic Journal of Statistics*, 14(1):1211–1233.

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. I. (2020). Asymptotic independence and support detection techniques for heavy-tailed multivariate data. *Insurance: Mathematics and Economics*, 93:262–277.

Liu, J. and Ye, J. (2009). Efficient Euclidean projections in linear time. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 657–664, New-York.

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

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

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New-York.

Sabourin, A. and Drees, H. (2019). Principal component analysis for multivariate extremes. *arXiv:1906.11043*.

Sabourin, A. and Naveau, P. (2014). Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization. *Computational Statistics & Data Analysis*, 71:542–567.

Sabourin, A., Naveau, P., and Fougeres, 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. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532.