Journal of Machine Learning Research 25 (2024) 1-36 Submitted 11/21; Revised 12/23; Published 4/24

# Spectral learning of multivariate extremes

**Marco Avella Medina** MARCO.AVELLA@COLUMBIA.EDU
*Department of Statistics*
*Columbia University*
*New York, NY, USA*

**Richard A. Davis** RDAVIS@STAT.COLUMBIA.EDU
*Department of Statistics*
*Columbia University*
*New York, NY, USA*

**Gennady Samorodnitsky** GS18@CORNELL.EDU
*School of Operations Research and Information Engineering*
*Cornell University*
*Ithaca, NY, USA*

**Editor:** Aryeh Kontorovich

## Abstract

We propose a spectral clustering algorithm for analyzing the dependence structure of multivariate extremes. More specifically, we focus on the asymptotic dependence of multivariate extremes characterized by the angular or spectral measure in extreme value theory. Our work studies the theoretical performance of spectral clustering based on a random $k$-nearest neighbor graph constructed from an extremal sample, i.e., the angular part of random vectors for which the radius exceeds a large threshold. In particular, we derive the asymptotic distribution of extremes arising from a linear factor model and prove that, under certain conditions, spectral clustering can consistently identify the clusters of extremes arising in this model. Leveraging this result we propose a simple consistent estimation strategy for learning the angular measure. Our theoretical findings are complemented with numerical experiments illustrating the finite sample performance of our methods.

**Keywords:** Angular measure, heavy tails, Laplacian, nearest neighbor graphs, regular variation, spectral clustering

## 1. Introduction

Multivariate extremes arise when one or more of rare extreme events occur simultaneously. They are of paramount importance for understanding environmental risks such as fires or droughts since they are driven by joint extremes of a number of meteorological variables. Similarly, catastrophic financial events are also of a multivariate nature in financial systems driven by core institutions that are connected. In the above examples one is precisely interested in modeling the dependence between rare individual extremes. Multivariate extreme value theory is an active research area that provides tools for modeling such events.

The dependence structure between extreme observations can be complex and typically characterized by different notions of dependence from the ones arising in the non-extreme world. For this reason recent work has sought to rethink various notions of sparsity for

extremes (Goix et al., 2017; Meyer and Wintenberger, 2021; Simpson et al., 2020), concentration inequalities (Goix et al., 2015; Clémençon et al., 2023), conditional independence (Belkin and Niyogi, 2003) and unsupervised learning (Chautru, 2015; Cooley and Thibaud, 2019; Janßen and Wan, 2020; Drees and Sabourin, 2021). See also Engelke and Ivanovs (2021) for a review of recent developments in the literature of multivariate extremes. Much of this line of research tries to connect important ideas from modern statistics and machine learning to the context of multivariate extremes. Our work falls in this category as we propose spectral clustering as a tool for learning the dependence structure of multivariate extremes.

Spectral clustering (Von Luxburg, 2007) and related techniques are very popular and have found success in various applications such as parallel computing (Hendrickson and Leland, 1995; Van Driessche and Roose, 1995), image segmentation (Shi and Malik, 2000) and community detection (Rohe et al., 2011; Lei and Rinaldo, 2015; Zhou and Amini, 2019). The central idea of spectral clustering is to use the eigenvectors of the graph Laplacian matrix constructed from an affinity graph between sample points in order to find clusters in the data. Typically these are obtained by a $K$-means algorithm that take these graph Laplacian eigenvectors as input. We follow this same principle but use as input to our algorithm the angular parts of the observations whose norms exceed a certain large threshold i.e., a standard spectral clustering algorithm is applied to the graph built over the angular parts of these extreme observations.

Because of the nature of the extreme events that we study, we leverage tools from multivariate extreme value theory for analyzing the theoretical properties of our spectral clustering algorithm. In particular, we use multivariate regular variation as a modeling tool since it is closely connected to asymptotic characterizations of multivariate extreme value distributions (Resnick, 2007, 2018). While a precise definition of regular variation is provided in Section 2, the basic idea is that a $d$-dimensional random vector $\mathbf{X}$ is regularly varying if the distribution of the angular part $\mathbf{X}/\|\mathbf{X}\|$ stabilizes (i.e., converges in distribution) as the radial part $\|\mathbf{X}\|$ becomes large and that the radial part has Pareto-like tails. The dependence structure is then governed by the asymptotic distribution of the limiting angular part. In this paper, we consider clustering of the angular parts, which live on a $d$-dimensional unit sphere, of observations with large radii. Learning this measure is challenging because of its multivariate nature and because only a small fraction of the data is considered to be extremes, i.e., those observations whose radii are sufficiently large, are retained for estimation. In contrast, standard modeling approaches built on parametric models are hard to extend to larger dimensions because of their lack of flexibility and computational complexity Davison and Huser (2015).

We will explore the use of spectral clustering for learning the angular measure. The performance of the algorithm critically depends on the properties of the random graph that it takes as input. We will focus on $k$-nearest neighbor graphs and hence a decision has to be made about the size of $k$ for constructing the random graph. In this work we study this question by focusing on a linear factor model. We characterize the asymptotic distribution of the multivariate extremes generated from this model and show that their dependence structure is captured by a discrete angular measure in the limit. We establish a rate of convergence for the angular components of the extremes to their discrete limits. This is a key step in deriving a theoretically valid range of numbers of $k$-nearest neighbors for

constructing a nearest neighbor graph that one should consider in order to guarantee that spectral clustering can be successfully used to learn the asymptotic angular measure.

From a methodological perspective, the work of Janßen and Wan (2020) is perhaps the closest to our approach since they also provide a clustering algorithm for extremes. Their method is however very different as it is based on spherical $k$-means (Dhillon and Modha, 2001), a variant of $k$-means that replaces the usual square loss minimization by an angular dissimilarity measure minimization. The data-generating model we consider is a natural factor model that can be viewed as a generalization of the max-linear model considered in Janßen and Wan (2020). We characterize the limiting distribution of the extremes in this model. We rigorously study the extremal nearest neighbor graphs and show that their connected components can identify the clusters of extremes of our factor model. By construction our algorithm is computationally tractable and model agnostic, so it has a potential of working well beyond the setting covered by our theory.

The rest of the paper is organized as follows. Section 2 provides some background notions from multivariate regular variation necessary for our analysis. Section 3 introduces the proposed spectral clustering algorithm for extremes. In Section 4 we introduce our linear factor model (LFM) and derive the asymptotic distribution of the angular components $\mathbf{X}/\|\mathbf{X}\|$ of observations with high threshold exceedances i.e., observations $\mathbf{X}$ with very large $\|\mathbf{X}\|$. In Section 5 we study the behavior of $k$-nearest neighbor graphs constructed using a sample of extremes. Section 6 contains a number of numerical examples that illustrate our proposed method. We show in Section 6.1 that for a large range of values of $k$ the connected components of the nearest neighbor graph consistently identify the clusters of extremes arising from the linear factor model. This includes an examination of LFM with added noise. The spectral clustering method is still able to estimate the signal reasonably well. The good numerical performance of the method in the LFM plus noise context suggests that it might work well in more general settings. The spectral clustering method is also applied to an environmental data set consisting of daily measurements of five air pollutants over both winter and summer seasons. The analysis suggests that in modeling the extremes, a LFM model with 5 clusters seems appropriate. Moreover, viewed as a time series, the extremal dependence for O3 and NO2 does not extend beyond a second-day time lag. Proofs of the technical results in the body of the paper and their complements are contained in the appendix.

## 2. Background on multivariate regular variation

Regular variation is often the starting point in modeling heavy-tailed data. We will make regular use of this assumption throughout this work. A random vector $\mathbf{X} = (X_1, \ldots, X_d)^\top$ is said to be regularly varying with exponent $\alpha > 0$ if for some norm $\|\cdot\|$ on $\mathbb{R}^d$ and some probability measure $\Gamma$ on the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$, the following limits hold:

$$\lim_{r\to\infty} \mathbb{P}\big(\mathbf{X}/\|\mathbf{X}\| \in \cdot \mid \|\mathbf{X}\| > r\big) \Rightarrow \Gamma(\cdot) \tag{1}$$

and

$$\lim_{r\to\infty} \frac{\mathbb{P}\big(\|\mathbf{X}\| > rx\big)}{\mathbb{P}\big(\|\mathbf{X}\| > r\big)} = x^{-\alpha} \tag{2}$$

for all $x > 0$, where $\Rightarrow$ denotes weak convergence on $\mathbb{S}^{d-1}$. In other words, the law of the angular component $\mathbf{X}/\|\mathbf{X}\|$ stabilizes as the radial component becomes large, and the radial component is regularly varying (equation (2)) with index $\alpha$. The limit probability measure $\Gamma$ is called the *angular measure* (or spectral measure) and describes how likely the extremal observations are to point in different directions. In other words, the angular measure describes the limiting extremal angle for high threshold exceedances that correspond to large $\|\mathbf{X}\|$. The support of this measure is particularly important since it shows which directions of the extremes are feasible and which are not feasible. Throughout the rest of paper we will take $\|\cdot\|$ to be the Euclidean norm.

For example, if $\mathbf{X}$ has a spherically symmetric distribution and the radius $\|\mathbf{X}\|$ has a Pareto distribution with index $\alpha$, then $\mathbf{X}$ is regularly varying with angular measure that is uniform on $\mathbb{S}^{d-1}$. In this case, the random vector is equally likely to have extremes in any direction so we do not expect extremes to be clustered. On the other hand, consider observations generated from a univariate MA(3) process given by $Y_t = Z_t + .5Z_{t-1} - .6Z_{t-2} + 1.5Z_{t-3}$, where $\{Z_t\}$ is an iid sequence of symmetric stable random variables with index $\alpha = 1.8$. The bivariate vector $\mathbf{X}_t = (Y_t, Y_{t-1})^\top$ is regularly varying and the scatter plot of $Y_t$ vs $Y_{t-1}$ is displayed in the left panel of Figure 1. Notice that for large values of $\|\mathbf{X}_t\|$, the points align themselves on rays. In the right panel is a plot of $\mathbf{X}_t/\|\mathbf{X}_t\|$ for those values of $\|\mathbf{X}_t\|$ that exceed the 99.8% empirical quantile of the radii and are grouped in 10 clusters. In this particular case, the spectral distribution consists of 10 point masses (5 pairs of symmetric point masses, indicated by arrows emanating from the origin).

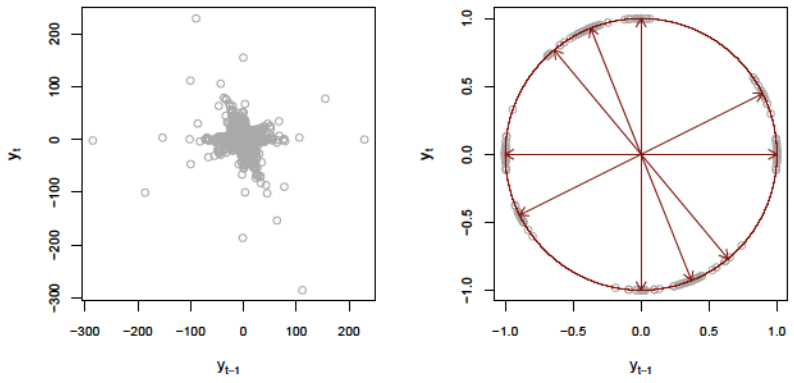



Figure 1: Scatter plot of $(Y_t, Y_{t-1})$ for an MA(3) process (left); spectral measure on $\mathbb{S}^1$ (right)

This simple example illustrates the challenge in finding meaningful low dimensional regions supporting the extremes. In a series of papers (see Meyer and Wintenberger (2021),

Drees and Sabourin (2021), Cooley and Thibaud (2019)), PCA-like analyses have been applied to finding low-dimensional subspaces that contain the bulk of the support of the spectral measure. As seen in this MA(3) example, such strategies might not be well suited for extracting the key features in the extremes which do not necessarily live neatly on a small number of subspaces. The approach taken here will be to use spectral clustering to learn the angular measure. While machine learning ideas provide guidance for thinking about and addressing multivariate extremes, the very nature of rare events will require us to borrow ideas from the theory of multivariate regular variation to analyze the extremal nearest neighbor graphs used by our algorithm.

## 3. Spectral clustering

In this section we describe how to construct random graphs based on a sample of extremes and how to use such graphs to find clusters of extremes via a simple spectral clustering algorithm.

### 3.1 Constructing random graphs

Starting with a sample of $d$-dimensional observations $\mathbf{X}_i$, $i = 1, \ldots, n$, one first needs to identify the extremal part of the sample, on which the extremal estimation will be performed. This is often done by selecting a high threshold $u_n$ and assigning to the extremal part of the sample the observations $\mathbf{X}_i$ satisfying $\|\mathbf{X}_i\| > u_n$.

Assume that $N_n$ observations $\mathbf{X}_i$ (with $i$ in some set $\mathcal{V}_n$ of cardinality $N_n$) are in the extremal part of the sample. Associated with each $i \in \mathcal{V}_n$, is the angular component of the observation $\mathbf{X}_i/\|\mathbf{X}_i\|$ that lives on the unit sphere. This allows us to think of the points in $\mathcal{V}_n$ as points on the unit sphere, forming nodes in a simple graph. We connect nodes $i_1$ and $i_2$ by an edge according to a certain rule. One possible rule chooses $\epsilon > 0$ and connects $i_1, i_2 \in \mathcal{V}_n$ by an edge if

$$\rho\big(\mathbf{X}_{i_1}/\|\mathbf{X}_{i_1}\|, \mathbf{X}_{i_2}/\|\mathbf{X}_{i_2}\|\big) \leq \epsilon\,. \tag{3}$$

One often uses the usual Euclidean distance $\rho$ on the unit sphere $\mathbb{S}^{d-1}$ in (3); but another distance function on the unit sphere could also be used. The random set of edges $\mathcal{E}_n$ created in this fashion define an $\epsilon$-neighborhood graph.

In what follows, we will focus on a different rule, leading to the *k-Nearest Neighbor graphs* ($k$-NN graphs). This rule asserts that a node $i_1 \in \mathcal{V}_n$ is connected to a node $i_2 \in \mathcal{V}_n$ if the point on the unit sphere corresponding to $i_2$ is among the $k$-nearest neighbors of the point corresponding to $i_1$, according to some distance function. This definition leads to a directed graph because the neighborhood relationship is not symmetric. There are two natural ways of making this graph undirected. The first one is to connect $i_1$ and $i_2$ with an undirected edge if either $i_1$ is among the $k$-nearest neighbors of $i_2$ or $i_2$ is among the $k$-nearest neighbors of $i_1$. The second one connects $i_1$ and $i_2$ only if both conditions are met, i.e., when $i_1$ and $i_2$ are mutual nearest neighbors (hence the resulting graph is usually called *mutual* $k$-nearest neighbor graph). Our main results apply to both constructions. We work with weighted graphs, where we assign to the edges a weight equal to the distance between the points on the unit sphere defining the nodes. More specifically, we will take as

input to our algorithm the *weighted adjacency matrix* $\mathbf{W} = [w_{i_1 i_2}]_{i_1, i_2 \in \mathcal{V}_n}$ and

$$w_{i_1 i_2} = \begin{cases} d\big(\mathbf{X}_{i_1}/\|\mathbf{X}_{i_1}\|, \mathbf{X}_{i_2}/\|\mathbf{X}_{i_2}\|\big) & \text{if } i_1 \text{ and } i_2 \text{ are connected,} \\ 0, & \text{if } i_1 \text{ and } i_2 \text{ are not connected.} \end{cases} \tag{4}$$

When defining the weights in (4) $d$ is a certain kernel; a typical example of such a kernel e.g., $d(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|)$, is used in the examples of Section 6. In the following subsections, we describe our algorithm and highlight the theoretical challenges.

### 3.2 The algorithm

The degree of a node $i \in \mathcal{V}_n$ is defined as

$$d_i = \sum_{j \in \mathcal{V}_n} w_{ij}.$$

The degree matrix $D$ is defined as the diagonal matrix with diagonal elements $[d_i]_{i \in \mathcal{V}_n}$ and the normalized symmetric graph Laplacian matrix is defined as

$$L = I - D^{-1/2} W D^{-1/2}, \tag{5}$$

where $I$ is the identity matrix. The spectral clustering algorithm of Ng et al. (2002) proceeds as follows:

1. Compute the first $m$ eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ of $L$ (i.e., the eigenvectors corresponding to the $m$ smallest eigenvalues of $L$) and define an $N_n \times m$ matrix $U$ using these eigenvectors.

2. Form an $N_n \times m$ matrix $V$ by normalizing the rows of $U$ to have unit norm.

3. Treating each of the $N_n$ rows of $V$ as a vector in $\mathbb{R}^m$, cluster them into $m$ clusters $C_1, \ldots, C_m$ using the $K$-means algorithm.

4. Assign the original points $\mathbf{X}_i$ to cluster $C_j$ if and only if row $i$ of the matrix $V$ was assigned to cluster $C_j$.

The motivation for this algorithm is described below.

### 3.3 Connected components, Laplacian and $k$-nearest neighbor graph.

We say that a subset $\mathcal{A} \subset \mathcal{V}_n$ of the vertices of a graph is connected if any two vertices in $\mathcal{A}$ can be joined by a path of edges such that all intermediate vertices also lie in $\mathcal{A}$. If $\mathcal{A}$ is connected and there are no connections between $\mathcal{A}$ and $\mathcal{V}_n \setminus \mathcal{A}$, then $\mathcal{A}$ is called a connected component. It is well known that the number of connected components of a graph $G$ is related to the spectrum of its symmetric graph Laplacian. This is formalized in the following proposition (Von Luxburg, 2007, Proposition 2).

**Proposition 1** *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with non-negative weights. Then the multiplicity $m$ of the eigenvalue 0 of $L$ equals the number of connected components $\mathcal{A}_1, \ldots, \mathcal{A}_m$ in the graph. The eigenspace of eigenvalue 0 is spanned by the indicator vectors $\delta_{\mathcal{A}_1}, \ldots, \delta_{\mathcal{A}_m}$ of those components.*

It follows from this result that if the spectral clustering algorithm is applied to a graph with the number of connected components equal to the parameter $m$ in the algorithm, then the algorithm will identify the connected components. In Sections 4 and 5 we derive the asymptotic angular distribution of multivariate extremes arising from a linear factor model and provide the relevant asymptotic theory for the connected components of the $k_n$-nearest neighbor graph constructed using the angular components of the extremes. Specifically, it will be rigorously established that under certain conditions the spectral clustering of the resulting graph consistently estimates the support of the spectral measure of multivariate extremes arising from this model.

## 4. Linear factor model and convergence of the angular components

We now introduce the generative model that we will be studying in this paper. Let $\mathbf{X}$ be a $d$-dimensional random vector defined by the following linear factor model (LFM)

$$\mathbf{X} = A\mathbf{Z}\,, \tag{6}$$

where $A = [a_{ij}]_{i=1,\ldots d;j=1,\ldots p}$ is a $d \times p$ matrix of nonnegative elements and $\mathbf{Z}$ is a $p$-dimensional random vector of factors consisting of independent and identically distributed random variables, that are either nonnegative or symmetric, and have asymptotically Pareto tails, i.e.,

$$\mathbb{P}(Z_1 > z) \sim cz^{-\alpha},\ \text{as } z \to \infty \tag{7}$$

for some $\alpha > 0$ and $c > 0$. Note that we write $f(x) \sim g(x)$ as $x \to \infty$ to mean that $\lim_{x\to\infty} f(x)/g(x) = 1$. In the examples section, we will add noise to the model in (6) in which case, the model corresponds to a *standard* heavy-tailed linear factor model. One can think of (6) as a linear version of the max-linear model studied in Janßen and Wan (2020), which has the same spectral distribution. We will also relax the assumption that the matrix $A$ is non-negative and will allow the noise to be symmetric. In this case, the spectral distribution of the model is no longer constrained to the positive quadrant of $\mathbb{S}^{d-1}$. Related max-linear models have also been considered in the context of time series models for extremes (Davis and Resnick, 1989; Hall et al., 2002) and more recently in the context of structural equation models (Gissibl and Klüppelberg, 2018; Klüppelberg and Lauritzen, 2019). The asymptotic Pareto assumption in (7) can be weakened to regular variation, at least for the main results in this section. However, this comes at the expense of assuming a more intrinsically complex set of conditions on the choice of thresholding sequences. Additional conditions, such as existence and properties of the density function of the noise, are required for the proofs of the results in Section 5.

It follows immediately from (6) and (7) (see, for example, Basrak et al. (2002), Proposition A.1) that $\mathbf{X}$ is a multivariate regularly varying random vector satisfying (1); namely,

$$\lim_{x\to\infty} \mathbb{P}\left( \frac{\mathbf{X}}{\|\mathbf{X}\|} \in \cdot \mid \|\mathbf{X}\| > x, \right) \Rightarrow \Gamma(\cdot)\,, \tag{8}$$

where $\Rightarrow$ denotes weak convergence on the unit sphere $\mathbb{S}^{d-1}$, $\Gamma$ is a discrete probability measure on $\mathbb{S}^{d-1}$ that, in the nonnegative case, puts mass $\|\mathbf{a}^{(k)}\|^\alpha/w$ at $\mathbf{a}^{(k)}/\|\mathbf{a}^{(k)}\|$ for

$k = 1, \ldots, p$, where $\mathbf{a}^{(k)} = (a_{1k}, a_{2k}, \ldots, a_{dk})^\top$, is the $k^{th}$ column of the matrix $A$ and

$$w = \sum_{k=1}^{p} \|\mathbf{a}^{(k)}\|^\alpha . \tag{9}$$

In other words, $\Gamma$ has the representation

$$\Gamma(\cdot) = w^{-1} \sum_{k=1}^{p} \|\mathbf{a}^{(k)}\|^\alpha \delta_{\frac{\mathbf{a}^{(k)}}{\|\mathbf{a}^{(k)}\|}}(\cdot) , \tag{10}$$

where $\delta_x(\cdot)$ is the Dirac measure that puts unit mass at $x$. On the other hand, in the symmetric case, $\Gamma$ puts mass $\|\mathbf{a}^{(k)}\|^\alpha / 2w$ at $\pm \mathbf{a}^{(k)}/\|\mathbf{a}^{(k)}\|$ for $k = 1, \ldots, p$. That is, the number of point masses in the symmetric case is double of that number in the nonnegative case. [1]

Based on a random sample of iid copies of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of $\mathbf{X}$ as above, we construct an estimate of the location of the point masses that comprise $\Gamma$, i.e.,

$$\mathbf{c}_k = \frac{\mathbf{a}^{(k)}}{\|\mathbf{a}^{(k)}\|} , \ k = 1, \ldots, p \tag{11}$$

in the nonnegative case, and

$$\mathbf{c}_{k,\pm} = \frac{\pm \mathbf{a}^{(k)}}{\|\mathbf{a}^{(k)}\|} , \ k = 1, \ldots, p \tag{12}$$

in the symmetric case. Note that these $\mathbf{c}_k$ are not necessarily distinct. Intuitively, for large $n$, the *angular parts* $\mathbf{X}_i/\|\mathbf{X}_i\|$ of the sample for which $\|\mathbf{X}_i\|$ is large, will cluster around these $\mathbf{c}_k$. In fact, we formalize this intuition and provide a *rate of convergence* for the limiting extremal angles with high threshold exceedances in the next theorem. This will be a key ingredient in our convergence analysis of extremal $k$-NN graphs. For the ease of notation we will prove the following result in the nonnegative case; the symmetric case follows by simply doubling the number of points on the sphere.

**Theorem 2** *If $(u_n)$ is a sequence converging to infinity as $n \to \infty$, then, in the nonnegative case, for any $j = 1, \ldots, p$, the conditional law of*

$$u_n\big(\mathbf{X}/\|\mathbf{X}\| - \mathbf{c}_j\big)$$

*given $\|\mathbf{X}\| > u_n$, $Z_j > u_n / w^{1/\alpha}$, ($w$ defined in (9)) converges weakly to the law of*

$$\frac{1}{\|\mathbf{a}^{(j)}\|^2 W_\alpha} \left(S^*_{1,-j}, \ldots, S^*_{d,-j}\right)^\top ,$$

1. Without much additional effort, one could consider the case that the tails of $Z_1$ are balanced in the sense that $\lim_{x\to\infty} \mathbb{P}(Z_1 > x)/\mathbb{P}(|Z_1| > x) \to p^+ \in [0,1]$. The location of the point masses for $\Gamma$ would be exactly the same as in the symmetric case, but with mass $p^\pm \|\mathbf{a}^{(k)}\|^\alpha / w$ at $\mathbf{c}_{k,\pm}$ defined in (12), where $p^- = 1 - p^+$.

*where $W_\alpha$ has a Pareto distribution (i.e. $\mathbb{P}(W_\alpha > x) = x^{-\alpha}$, $x \geq 1$) that is independent of $Z_1, \ldots, Z_p$,*

$$S^*_{l,-j} = \sum_{i=1}^{d} \left( a_{ij}^2 X_{l,-j} - a_{lj} a_{ij} X_{i,-j} \right) , \tag{13}$$

*and*

$$X_{i,-j} = X_i - a_{ij} Z_j = \sum_{\substack{m=1, \\ m \neq j}}^{p} a_{im} Z_m \, . \tag{14}$$

**Proof** We start by observing that the conditional law of

$$\left( Z_1, \ldots, Z_{j-1}, Z_j / u_n, Z_{j+1}, \ldots, Z_p \right)^\top \tag{15}$$

given $\|\mathbf{X}\|^2 > u_n^2$, $Z_j > u_n / w^{1/\alpha}$, converges in distribution, as $n \to \infty$, to the law of

$$\left( Z_1, \ldots, Z_{j-1}, W_\alpha / w_j, Z_{j+1}, \ldots, Z_p \right)^\top, \tag{16}$$

where $w_j = \|\mathbf{a}^{(j)}\|$ The main ingredients in establishing this result is to note that $Z_j^2$ is regularly varying with index $\alpha/2$ while for $i \neq j$, $Z_i Z_j$ is regularly varying with index $\alpha$ (see Embrechts and Goldie (1980); Theorem 3). Moreover, from the convolution closure property for sums of independent regularly varying random variables, it follows easily that

$$\begin{aligned}
\mathbb{P}(Z_j > u_n x \, , \sum_{k=1}^{d} (\sum_{i=1}^{p} a_{ki} Z_i)^2 > u_n^2, \, ) &\sim \mathbb{P}(Z_j > u_n x, \sum_{i=1}^{p} w_i^2 Z_i^2 > u_n^2) \\
&\sim \sum_{i=1}^{p} \mathbb{P}(Z_j > u_n x, w_i^2 Z_i^2 > u_n^2) \\
&\sim \mathbb{P}(Z_j > u_n x) \, .
\end{aligned}$$

Now to finish the proof of (16), we use these relations and note that for $x \geq 1/w_j$ (and hence $x \geq 1/w^{1/\alpha}$),

$$\begin{aligned}
\mathbb{P}(Z_j > u_n x \, | \, \|\mathbf{X}\|^2 > u_n^2, \, Z_j^2 > u_n^2 / w^{2/\alpha}) &\sim \frac{\mathbb{P}(Z_j > u_n x, \sum_{i=1}^{p} w_i^2 Z_i^2 > u_n^2, Z_j^2 > u_n^2 / w^{2/\alpha})}{\mathbb{P}(\sum_{i=1}^{p} w_i^2 Z_i^2 > u_n^2, Z_j^2 > u_n^2 / w^{2/\alpha})} \\
&\sim \frac{\mathbb{P}(Z_j > u_n x)}{\mathbb{P}(Z_j > u_n / w_j)} \\
&\to w_j^{-\alpha} x^{-\alpha} \\
&= \mathbb{P}(W_\alpha > w_j x) \, .
\end{aligned}$$

We have

$$\begin{aligned}
u_n \big( \mathbf{X} / \|\mathbf{X}\| - \mathbf{c}_j \big) &= u_n \left( \frac{(\sum_{m=1}^{p} a_{1m} Z_m, \ldots, \sum_{m=1}^{p} a_{dm} Z_m)^\top}{\|\mathbf{X}\|} - \frac{\mathbf{a}^{(j)}}{\|\mathbf{a}^{(j)}\|} \right) \\
&= (V_1, \ldots, V_d)^\top,
\end{aligned}$$

say. For $l = 1, \ldots, d$,

$$V_l = u_n \frac{w_j \sum_{m=1}^{p} a_{lm} Z_m - a_{lj} \|\mathbf{X}\|}{w_j \|\mathbf{X}\|}$$
$$= u_n \frac{w_j^2 \left(\sum_{m=1}^{p} a_{lm} Z_m\right)^2 - a_{lj}^2 \|\mathbf{X}\|^2}{w_j \|\mathbf{X}\| \left(w_j \sum_{m=1}^{p} a_{lm} Z_m + a_{lj} \|\mathbf{X}\|\right)} . \quad (17)$$

We note that the numerator of (17) reduces to the following expression where the $Z_j^2$ terms cancel out

$$\begin{aligned} Num_l : &= 2a_{lj} Z_j \left( \|\mathbf{a}^{(j)}\|^2 X_{l,-j} - a_{lj} \sum_{i=1}^{d} a_{i,j} X_{i,-j} \right) + \|\mathbf{a}^{(j)}\|^2 X_{l,-j}^2 - a_{lj}^2 \sum_{i=1}^{d} X_{i,-j}^2 \\ &= 2a_{lj} Z_j \left( \sum_{i=1}^{d} \left( a_{ij}^2 X_{l,-j} - a_{lj} a_{ij} X_{i,-j} \right) \right) + \|\mathbf{a}^{(j)}\|^2 X_{l,-j}^2 - a_{lj}^2 \sum_{i=1}^{d} X_{i,-j}^2 , \quad (18) \end{aligned}$$

where $X_{i,-j}$ is as defined in (14). The denominator of (17) is handled in a similar way, but this time the $Z_j^2$ terms do not cancel. Indeed, since

$$\begin{aligned} \|\mathbf{X}\| &= \sqrt{\sum_{k=1}^{d} (a_{kj}^2 Z_j^2 + 2a_{kj} Z_j X_{k,-j} + X_{k,-j}^2)} \\ &= \sqrt{\sum_{k=1}^{d} a_{kj}^2 Z_j^2} \sqrt{1 + \frac{\sum_{k=1}^{d} (X_{k,-j}^2 + 2a_{kj} Z_j X_{k,-j})}{\sum_{k=1}^{d} a_{kj}^2 Z_j^2}} \\ &= w_j |Z_j| \sqrt{1 + \frac{\sum_{k=1}^{d} (X_{k,-j}^2 + 2a_{kj} Z_j X_{k,-j})}{w_j^2 Z_j^2}}, \end{aligned}$$

we can write

$$Den_l = w_j^2 |Z_j| (1 + o_p(1)) \left( w_j a_{lj} Z_j + R + a_{lj} w_j |Z_j| (1 + o_p(1)) \right),$$

where $R$ is a linear function in the variables $Z_1, \ldots, Z_p$ in which $Z_j$ does not appear, and $o_p(1)$ goes to zero in probability given $Z_j > u_n / w^{1/\alpha}$. We view

$$V_l = u_n \frac{Num_l}{Den_l} = \frac{Num_l / u_n}{Den_l / u_n^2}$$

as a ratio of two continuous real-valued functions of the random vector in (15) (plus a vanishing term in $Den_l$), so that the random vector $(V_1, \ldots, V_d)^\top$ becomes a $d$-dimensional vector of such ratios. By the continuous mapping theorem the random vector $(V_1, \ldots, V_d)^\top$ converges weakly to the $d$-dimensional vector of the ratios of the corresponding functions applied to the random vector in (16). These result in

$$2a_{lj} (W_\alpha / w_j) S_{l,-j}^*$$

in the case of $Num_l$ and in

$$W_\alpha^2\, 2a_{lj}w_j$$

in the case of $Den_l$. Putting everything together produces the claim. ■

As explained above, the following corollary is an immediate consequence of Theorem 2.

**Corollary 3** *Let a sequence of levels $(u_n)$ converging to infinity. Then, in the symmetric case, in the notation of Theorem 2, for any $j = 1, \dots, p$, the conditional law of*

$$u_n\big(\mathbf{X}/\|\mathbf{X}\| - \pm\mathbf{c}_j\big)$$

*given $\|\mathbf{X}\| > u_n$, $\pm Z_j > u_n/w^{1/\alpha}$, converges weakly to the law of*

$$\frac{\pm 1}{\|\mathbf{a}^{(j)}\|^2 W_\alpha}\left(S^*_{1,-j}, \dots, S^*_{d,-j}\right)^\top.$$

**Remark 4** *In the nonnegative case, Theorem 2 addresses the conditional convergence of $\mathbf{X}/\|\mathbf{X}\|$, given $\|\mathbf{X}\| > u_n$, $Z_j > u_n/w^{1/\alpha}$, to the location $\mathbf{c}_j$ of the corresponding atom of the spectral measure. It is also possible to address a conditional convergence to the mass $w^{-1}\|\mathbf{a}^{(j)}\|^\alpha$ of this atom. Indeed,*

$$\mathbb{P}\big(Z_j > u_n/w^{1/\alpha}\big|\|\mathbf{X}\| > u_n\big) \to w^{-1}\|\mathbf{a}^{(j)}\|^\alpha \tag{19}$$

*as $n \to \infty$. To see this, write*

$$\mathbb{P}\big(Z_j > u_n/w^{1/\alpha}\big|\|\mathbf{X}\| > u_n\big) = \frac{u_n^\alpha \mathbb{P}\big(Z_j > u_n/w^{1/\alpha}, \|\mathbf{X}\| > u_n\big)}{u_n^\alpha \mathbb{P}\big(\|\mathbf{X}\| > u_n\big)},$$

*and the numerator converges to $c\|\mathbf{a}^{(j)}\|^\alpha$, while the denominator converges to $cw$. If one strengthens the asymptotic Pareto tails assumption* (7) *to include the rate of convergence to the limit, then one would able to establish the rate of convergence in* (19) *as well. The situation is similar in the symmetric case. We do not pursue this in the present paper.*

We now explore the connection between large values of the underlying factors $Z_{i1}, \dots, Z_{ip}$ and large values of $\|\mathbf{X}_i\|$. We will see that under certain conditions, high threshold exceedances of $\|\mathbf{X}_i\|$ are generated by only one underlying factor $Z_{ij}$, $j = 1, \dots, p$. This will be important for our analysis of extremal $k$-NN graphs which will require additional assumptions on the rate of growth of $u_n$. Since $\frac{\alpha+2}{\alpha(\alpha+3)} < \alpha^{-1}$, we can further impose that the sequence $(u_n)$ satisfies the growth conditions

$$n^{-1/\alpha}u_n \to 0 \quad \text{and} \quad n^{-(\alpha+2)/(\alpha(\alpha+3))}u_n \to \infty\,, \tag{20}$$

as $n \to \infty$. Also note that we may choose a further sequence $(h_n)$ such that

$$h_n \to \infty,\ h_n = o(u_n),\ h_n = o\big(u_n^{(\alpha+1)/2}n^{-1/2}\big),\ n^{-1/\alpha}u_n h_n \to \infty \tag{21}$$

as $n \to \infty$. Indeed, the choice

$$h_n = u_n^{(\alpha-1)/4}n^{(2-\alpha)/(4\alpha)}$$

works for this purpose.

For $n = 1, 2, \ldots$, we define the set of indexes corresponding to extreme observations

$$\mathcal{I}_n = \big\{i = 1, \ldots, n : \|\mathbf{X}_i\| > u_n\big\}, \tag{22}$$

and denote its cardinality by $N_n = \text{card}(\mathcal{I}_n)$. From (7) and (20), we see that the mean and variance of $N_n/(nu_n^{-\alpha})$ converge to $cw$ and 0, respectively and hence that

$$N_n/(nu_n^{-\alpha}) \xrightarrow{P} cw, \text{ as } n \to \infty. \tag{23}$$

The following lemma connects exceedances of $u_n$ by $\|\mathbf{X}_i\|$ with exceedances of $h_n$ by $Z_{ij}$, $j = 1, \ldots, p$.

**Lemma 5** *Let $(h_n)$ be a sequence satisfying (21) and consider the event*

$$B_n = \big\{\textit{for any } i \in \mathcal{I}_n \textit{ at most one of } Z_{im},\ m = 1, \ldots, p \textit{ exceeds } h_n\big\}.$$

*Then $\mathbb{P}(B_n) \to 1$ as $n \to \infty$.*

**Proof** Note that (6) implies that the $m^{th}$ component of the $i^{th}$ observation is of the form

$$X_{im} = \sum_{j=1}^{p} a_{mj} Z_{ij},\ m = 1, \ldots, d;\ i = 1, \ldots, n\,, \tag{24}$$

where $Z_{i1}, \ldots, Z_{ip}$ are iid random variables with asymptotic Pareto tails (7). Denote

$$a^* = d^{1/2} \max\{a_{mj},\ m = 1, \ldots, d;\ j = 1, \ldots, p\} \tag{25}$$

Since $u_n > h_n$ for $n$ large, we have

$$\begin{aligned}
\mathbb{P}(B_n^c) &\le \sum_{i=1}^{n} \mathbb{P}\left(\sum_{k=1}^{d} \big(X_{ik}\big)^2 > u_n^2,\ Z_{im} > h_n \text{ for two or more of } \ m = 1, \ldots, p\right) \\
&\le n\mathbb{P}\left(a^* \max_{k=1,\ldots,p} Z_{1k} > u_n,\ Z_{1m} > h_n \text{ for two or more of } \ m = 1, \ldots, p\right) \\
&\le \sum_{k=2}^{p} \binom{p}{k} n\mathbb{P}(Z_1 > h_n)\mathbb{P}\left(Z_1 > u_n/a^*\right) \to 0
\end{aligned}$$

by the last property in (21) and (7). This proves the lemma.

■

Equipped with Lemma 5 we can now proceed to bound the distance between the observed angular parts of the multivariate extremes and their corresponding theoretical asymptotic atoms. Assume, for a moment, that we are in the nonnegative case. We already know that for large $n$, we have that for every $i \in \mathcal{I}_n$ one of the values of $Z_{im}$, $m = 1, \ldots, p$ must exceed $u_n/a^*$ and all other values of these variables cannot exceed $h_n$. We now define the sets of

indexes corresponding to extremes generated by each of the individual factors i.e. we define for $j = 1, \ldots, p$

$$\mathcal{I}_n^{(j)} = \{i = 1, \ldots, n : \|\mathbf{X}_i\| > u_n,\, Z_{ij} > u_n/a^*\}\,. \tag{26}$$

Consequently,

$$\mathcal{I}_n = \bigcup_{j=1}^{p} \mathcal{I}_n^{(j)} \tag{27}$$

and by Lemma 5 for large $n$ this is a disjoint union with probability tending to one. Let $N_n^{(j)}$ be the cardinality of $\mathcal{I}_n^{(j)}$, $j = 1, \ldots, p$. Using the fact that $a^* \geq \|\mathbf{a}^{(j)}\|$ for $j = 1, \ldots, p$, the same argument as in (23) shows that $j = 1, \ldots p$,

$$N_n^{(j)}/(nu_n^{-\alpha}) \xrightarrow{P} c\|\mathbf{a}^{(j)}\|^\alpha, \text{ as } n \to \infty. \tag{28}$$

We enumerate $\mathbf{X}_i/\|\mathbf{X}_i\|$, $i \in \mathcal{I}_n$ as $\mathbf{Y}_i$, $i = 1, \ldots, N_n$, a sample on $\mathbb{S}^{d-1}$ of random size $N_n$. For each $j = 1, \ldots, p$, we enumerate $\mathbf{X}_i/\|\mathbf{X}_i\|$, $i \in \mathcal{I}_n^{(j)}$ as $\mathbf{Y}_i^{(j)}$, $i = 1, \ldots, N_n^{(j)}$, a sample on $\mathbb{S}^{d-1}$ of random size $N_n^{(j)}$. It is straightforward (if a bit tedious) to check the following result.

**Lemma 6** *For large $n$, on the event $B_n$, for $i = 1, \ldots, N_n^{(j)}$,*

$$\left\|\mathbf{Y}_i^{(j)} - \mathbf{c}_j\right\| \leq \frac{8(a^*)^2}{\|\mathbf{a}^{(j)}\|^\alpha} \frac{h_n}{u_n}, \quad j = 1, \ldots, p, \tag{29}$$

*where the $\mathbf{c}_j$ are as defined in* (12).

The situation in the symmetric case is, of course, completely analogous. It follows from the definition of $h_n$ in (21) and (29) that the angular components of the extremes are clustered around the centers $\mathbf{c}_j$. The results in the next section build on Lemma 6 and provide sufficient conditions for our extremal spectral clustering algorithm to be consistent. For this, we provide a careful asymptotic analysis of the extremal $k$-NN graph used by the algorithm.

## 5. Asymptotic analysis of the connected components of the extremal $k$-NN graph

Our analysis consists of two main components. The first one is to show that the extremes generated by different factors will belong to different components of the extremal $k$-NN graph as long as the cluster centers corresponding to the underlying factors are different. The second part will be to argue that all the extremes generated by an underlying factor will also belong to the same component of the extremal $k$-NN graph. This second step turns out to be the more technical one in our analysis and we will only establish this result for $d = 2$. Along the way we derive a few intermediate results that we also highlight in order to better explain the key ingredients of our argument. Going forward, in our proofs, $c > 0$ represents a finite and non-zero constant whose value may change from line-to-line. In the sequel we will assume, without further comments, that the sequence $(u_n)$ satisfies (20). The first step of our program is covered by the following proposition.

**Proposition 7** *Suppose that $k_n = o\big(nu_n^{-\alpha}\big)$ as $n \to \infty$. Then there is a sequence $(B_{n,1})$ of events with $\mathbb{P}(B_{n,1}) \to 1$ as $n \to \infty$ such that, for all $n$ large enough, on the event $B_{n,1}$, any two points $\mathbf{Y}_{i_1}^{(j_1)}$ and $\mathbf{Y}_{i_2}^{(j_2)}$, $i_1 = 1, \ldots, N_n^{(j_1)}$, $i_2 = 1, \ldots, N_n^{(j_2)}$ will belong to two different connected components of the $k_n$-NN graph if $\mathbf{c}_{j_1} \neq \mathbf{c}_{j_2}$.*

**Proof** Define

$$B_{n,1} = B_n \cap \big\{N_n^{(j)} > k_n \ \text{ for } j = 1, \ldots, p\big\}. \tag{30}$$

By Lemma 5, (28) and the assumption on $k_n$ , $\mathbb{P}(B_{n,1}) \to 1$ as $n \to \infty$. By (29) and the triangle inequality, on the events $B_{n,1}$, any point $\mathbf{Y}_{i_1}^{(j_1)}$ has at least $k_n$ neighbours of the type $\mathbf{Y}_{i}^{(j_1)}$, $i = 1, \ldots, N_n^{(j_1)}$, $i \neq i_1$, that are within distance of $c \cdot h_n/u_n$ from it. On the other hand, by (29) and the triangle inequality, its distance from any point $\mathbf{Y}_{i_2}^{(j_2)}$ with $\mathbf{c}_{j_1} \neq \mathbf{c}_{j_2}$ cannot be smaller than

$$\|\mathbf{c}_{j_1} - \mathbf{c}_{j_2}\| - c \cdot h_n/u_n.$$

Therefore, for large $n$, the latter point cannot be among the $k_n$-nearest neighbours of $\mathbf{Y}_{i_1}^{(j_1)}$. ■

We now embark on the second step of our program and establish that, at least in the case $d = 2$, under appropriate conditions, the points $\mathbf{Y}_i^{(j)}$, $i = 1, \ldots, N_n^{(j)}$, belong, with high probability, to the same connected component in the $k_n$-NN graph. We start by investigating the deviations of these points from the center of the cluster, $\mathbf{c}_j$, defined in (12). Since the points $\mathbf{Y}_i^{(j)}$, $i = 1, \ldots, N_n^{(j)}$ are treated as independent, the following result is essentially immediate from Theorem 2.

**Lemma 8** *In the nonnegative case, for any $j = 1, \ldots, p$, the conditional law of*

$$u_n\big(\mathbf{Y}_1^{(j)} - \mathbf{c}_j\big)$$

*given $N_n^{(j)} \geq 1$, converges weakly to the law of*

$$\frac{1}{\|\mathbf{a}^{(j)}\|^2 W_\alpha} \left(S^*_{1,-j}, \ldots, S^*_{d,-j}\right)^T$$

*that is specified in the statement of Theorem 2. An analogous statement holds in the symmetric case.*

**Remark 9** *It is a straightforward calculation to check that, if $j = 1, \ldots, d$,*

$$\sum_{l=1}^{d} a_{lj} S^*_{l,-j} = \sum_{l=1}^{d} \sum_{i=1}^{d} \left(a_{lj} a_{ij}^2 X_{l,-j} - a_{lj}^2 a_{ij} X_{i,-j}\right) = 0\,. \tag{31}$$

*Therefore, the normalized deviations of the points $\mathbf{Y}_i^{(j)}$, $i = 1, \ldots, N_n^{(j)}$ from the center of the $j$th cluster are, in the limit, supported by a $(d-1)$-dimensional subspace.*

Using the information in Lemma 8 we now proceed to prove that under appropriate conditions, the points $\mathbf{Y}_i^{(j)}$, $i = 1, \ldots, N_n^{(j)}$, belong, with high probability, to the same connected component of the $k_n$-NN graph. We will need some additional notation in order to state the result. For a fixed $j = 1, \ldots, d$ we write for $m = 1, \ldots, d$,

$$Y_{im}^{(j)} = \sum_{l=1}^{p} a_{ml} Z_{il}^{(*,j)},\ i = 1, \ldots, N_n^{(j)}; \tag{32}$$

the notation should be compared with (24). That is, $\{(Z_{i1}^{(*,j)}, \ldots, Z_{ip}^{(*,j)})^\top, i = 1, \ldots, N_n^{(j)}\}$ are iid random vectors distributed according to the conditional distribution of $(Z_1, \ldots, Z_p)^\top$ given $\|\mathbf{X}\| > u_n, Z_j > u_n/w^{1/\alpha}$. Since the connectivity of any nearest neighbor graph is not affected by shifting and scaling, it is sufficient to consider the $k_n$-NN graph constructed on the deviations of the points $\mathbf{Y}_i^{(j)}$, $i = 1, \ldots, N_n^{(j)}$ from the cluster center.

Continuing with the notation used in the proof of Theorem 2 we isolate the main term in the deviations from the cluster center by writing

$$\begin{aligned} u_n\big(\mathbf{Y}_i^{(j)} - \mathbf{c}_j\big) &= \Big(S_{1,-j}^{(*,j)}, \ldots, S_{d,-j}^{(*,j)}\Big)^\top / \big(w_j^2 (Z_{ij}^{(*,j)}/u_n)\big) \\ &\quad + \left[u_n\big(\mathbf{Y}_i^{(j)} - \mathbf{c}_j\big) - \Big(S_{1,-j}^{(*,j)}, \ldots, S_{d,-j}^{(*,j)}\Big)^\top / \big(w_j^2 (Z_{ij}^{(*,j)}/u_n)\big)\right] \\ &= \mathbf{M}^{(i)} + \mathbf{D}^{(i)}, \quad i = 1, \ldots, N_n^{(j)}, \end{aligned} \tag{33}$$

where $S_{l,-j}^{(*,j)} = Y_{il}^{(j)} - a_{lj} Z_{ij}^{(*,j)} = \sum_{\substack{m=1,\\ m\neq j}}^{p} a_{lm} Z_{im}^{(*,j)}$ is analogous to (14). In the case $d = 2$, it follows from (31) that for some nonzero deterministic vector $\mathbf{b}$ in $\mathbb{R}^2$,

$$\mathbf{M}^{(i)} = \frac{1}{w_j^2} \frac{S_{2,-j}^{(*,j)}}{Z_{ij}^{(*,j)}/u_n} \mathbf{b},\ i = 1, \ldots, N_n^{(j)}.$$

For notational simplicity we continue the discussion with $j = 1$, and in this case these are essentially univariate iid random variables with the distribution of

$$T_n = \frac{a_{22} Z_2 + \cdots + a_{2p} Z_p}{w_1^2 Z_1/u_n} \tag{34}$$

given

$$(a_{11} Z_1 + \cdots + a_{1,p} Z_p)^2 + (a_{21} Z_1 + \cdots + a_{2p} Z_p)^2 > u_n^2, \quad Z_1 > u_n/w^{1/\alpha}\,. \tag{35}$$

Finally, we let $F_{T_n}$ denote the conditional law of $T_n$ in (34) given the conditions in (35). For technical reasons we require further conditions on the latent factors in our subsequent results. We assume that the generic noise variable $Z$ in (6) and (7) is positive or symmetric, and has a probability density function $f_Z$ such that

$$f_Z(z) \text{ is bounded away from 0 on compact intervals and bounded from above,} \tag{36}$$

and

$$B^{-1} z^{-(\alpha+1)} \leq f_Z(z) \leq B z^{-(\alpha+1)}, \tag{37}$$

$\alpha > 1$, for all $z \geq z_0$, some $B \geq 1$.

The following lemma shows that the conditional density function $f_{T_n}(t) = \partial F_{T_n}(t)/\partial t$ enjoys some useful regularity properties.

**Lemma 10** *Assume* (36) *and* (37). *For $\alpha > 1$ the conditional density function $f_{T_n}$ is such that:*

*(i) There exists an $G \in (0, \infty)$ such that for all $n$ large enough, $f_{T_n}(t) \leq G$ for all $t$.*

*(ii) $f_{T_n}$ is uniformly bounded from below on compact intervals, uniformly in $n$.*

*(iii) There is a constant $D \geq 1$ and a number $t_0 \geq 0$ such that $D^{-1}t^{-(\alpha+1)} \leq f_{T_n}(t) \leq Dt^{-(\alpha+1)}$ uniformly for all $n$ large enough and all $t \geq t_0$.*

It is clear that an analogous result holds for the appropriate conditional densities in the symmetric case.

The following intermediate result is the key ingredient for completing our analysis of the connected component of the extremal $k_n$-NN graph, at least in the case $d = 2$, assuming certain regularity conditions on the noise variables.

**Lemma 11** *Assume* (36), (37) *and let $d = 2$, $\tau > 1$ and consider the random variable $m_n$ defined by*

$$m_n = N_n^{(1)}/\lceil \tau \log N_n^{(1)} \rceil, \text{ so that by (23) } m_n \sim \frac{cw_1 n u_n^{-\alpha}}{\tau \log(n u_n^{-\alpha})}, \quad n \to \infty. \tag{38}$$

*Define the intervals*

$$I_{i,n} = \left(F_{T_n}^{-1}\big((i-1)/m_n\big), F_{T_n}^{-1}\big(i/m_n\big)\right), \ i = 1, \ldots, m_n \tag{39}$$

*as well as intervals along vector $\mathbf{b}$ by*

$$J_{i,n} = I_{i,n}\mathbf{b}, \ i = 1, \ldots, m_n \, .$$

*Then, on an event with probability tending to one, there is a finite positive integer $K_0$ such that for all $n$ large enough and all $i = 2, \ldots, m_n - K_0$, every point in $J_{i,n}$ is closer to every point in the intervals $J_{i-1,n}$ and $J_{i+1,n}$ than to any point in an interval $J_{i_1,n}$ with $|i - i_1| > K_0$.*

The proofs of these two lemmas are contained in the Appendix.

We are now ready to state the main result showing that the extremes generated from the same underlying factor will also belong to the same connected component of the extremal $k_n$-NN graph with probability tending to one, under appropriate regularity conditions. This time, for simplicity, we only consider the symmetric case.

**Theorem 12** *Assume* (36), (37) *and let $d = 2$. Then, if $k_n > G \log n$ with large enough $G > 0$, there is a sequence $(B_{n,2})$ of events with $\mathbb{P}(B_{n,2}) \to 1$ as $n \to \infty$ such that, for all $n$ large enough, on the event $B_{n,2}$, any two points $\mathbf{Y}_{i_1}^{(j)}$ and $\mathbf{Y}_{i_2}^{(j)}$, $i_1 = 1, \ldots, N_n^{(j)}$, $i_2 = 1, \ldots, N_n^{(j)}$ will belong to the same connected component of the $k_n$-NN graph for any $j = 1, \ldots, p$.*

The proof of the theorem has been relegated to the appendix.

It follows from Proposition 7 and Theorem 12 that, with probability tending to one as $n \to \infty$, the extremal $k_n$-NN graph obtained from a sample drawn from (6) will have exactly $m \leq p$ connected components corresponding to the $m$ distinct asymptotic point masses (12) of the model. In other words, the extremal $k_n$-NN graph consistently identifies the underlying clusters through its connected components. This in turn implies that spectral clustering will be consistent by Proposition 1. We have therefore shown the following main practical result.

**Corollary 13** *Assume* (36), (37), $d = 2$, $k_n = o(nu_n^{-\alpha})$ *and* $k_n > G \log n$. *Then, spectral clustering will consistently identify the clusters of extremes arising from the linear factor model.*

**Remark 14** *In practice consistent clustering can be achieved by taking* $k_n > G_0 \log N_n$ *for some* $G_0 > 0$ *and* $k_n = o(N_n)$ *since* $N_n/(nu_n^{-\alpha}) \xrightarrow{P} cw$, *as* $n \to \infty$. *In our experiments we chose* $k_n = \lceil \frac{N_n}{\tau \log N_n} \rceil + 1$ *for some* $\tau > 1$.

Corollary 13 suggests a simple strategy for estimating the asymptotic angular measure of the extremes generated from the linear factor model (6). Assume we run spectral clustering on the extremal $k_n$-NN graph. Then we can denote by $\hat{\mathcal{I}}_n^{(j)}$ the set of indices corresponding to the $j$th cluster found by the algorithm for $j = 1, \ldots, m$. With these sets we can define $\hat{N}_n^{(j)} = \text{card}(\mathcal{I}_n^{(j)})$ and estimate the centers of the spectral measure and their respective masses as

$$\hat{\mathbf{c}}_j = \frac{1}{\hat{N}_n^{(j)}} \sum_{i \in \hat{\mathcal{I}}_n^{(j)}} \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|} \quad \text{and} \quad \hat{\pi}_j = \frac{\hat{N}_n^{(j)}}{N_n}, \quad j = 1, \ldots, p. \tag{40}$$

The following result is an inmediate consequence of the main results of this section.

**Corollary 15** *Suppose* $m = p$ *and that the conditions of Proposition 7 and Theorem 12 hold. Then,* $\hat{\mathbf{c}}_j \xrightarrow{P} \mathbf{c}_j$ *and* $\hat{\pi}_j \xrightarrow{P} w^{-1} \|\mathbf{a}^{(j)}\|^\alpha$ *for all* $j = 1, \ldots, p$.

Note that in practice one can also normalize the estimates $\hat{\mathbf{c}}_j$ to ensure that they lie in the unit sphere for all $n$. Clearly the resulting estimators remain consistent under the conditions of Corollary 15.

Even though the theoretical results of this section use the assumption $\alpha > 1$ in (37), we believe they should also hold when $\alpha \in (0, 1]$. The numerical experiments shown in the next section supports this assertion.

## 6. Numerical illustrations

In all the examples considered below we compute weighted adjacency matrices using the exponential kernel $d(\mathbf{x}, \mathbf{y}) = \exp(-s\|\mathbf{x} - \mathbf{y}\|)$ with $s = 1$ and select the number of clusters as suggested by the screeplots of the fully connected weighted adjacency matrices $W$. It matched nicely the correct number of clusters, when known. We consider sample sizes $n = \{1000, 5000, 25000, 125000\}$ and take a sample of extremes corresponding to observations whose Euclidean norm is larger or equal to the following vector of corresponding sample

quantiles: $\beta = \{0.9, 0.96, 0.984, 0.9968\}$, respectively. These quantiles were chosen to lead to samples of extremes of sizes $N_n = \{100, 200, 400, 800\}$, correspondingly. For these extremes we define $k_n$-nearest neighbors graphs with $k_n = \lceil \frac{N_n}{C \log N_n} \rceil + 1$, the corresponding values of the constants are in the vector $C = \{3, 5, 7, 9\}$ .

### 6.1 Linear factor model with and without noise

As a first example, consider $d-$dimensional vectors that follow the $p-$dimensional linear factor model

$$\mathbf{X} = A\mathbf{Z} + \sigma\boldsymbol{\varepsilon}\,, \tag{41}$$

where $A \in \mathbb{R}^{d\times p}$ is a matrix of factor loadings, $\mathbf{Z} = (Z_1, \ldots, Z_p)^\top$ is a $p$-dimensional vector consisting of iid standard Fréchet distributed components ($\alpha = 1$), $\sigma \geq 0$ regulates the signal to noise ratio and $\boldsymbol{\varepsilon}$ is a noise vector obtained by multiplying a univariate independent standard Fréchet with an independent $p$-dimensional random vector of iid standard normals, i.e.,

$$\boldsymbol{\varepsilon} = \mathbf{N}\eta\,, \tag{42}$$

where $\eta$ is standard Fréchet, $\mathbf{N} = (N_1, \ldots, N_p)^\top$ is a $p-$random vector consisting of iid standard normals, and $\mathbf{Z}, \eta$, and $\mathbf{N}$ are independent. Now using computations similar to those given in Section 4, it can be shown that

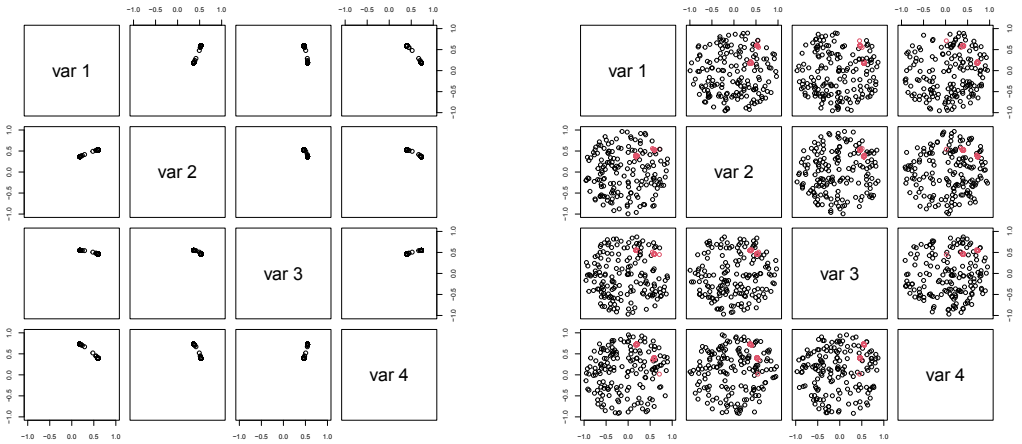
$$\begin{aligned}
\frac{\mathbb{P}(\|\mathbf{X}\| > x)}{\mathbb{P}(Z_1 > x)} &\sim \frac{\mathbb{P}(\sum_{i=1}^p \|\mathbf{a}^{(i)}\|^2 Z_i^2 + \sigma^2\|\mathbf{N}\|^2\eta^2 > x^2)}{\mathbb{P}(Z_1 > x)} \\
&\sim \frac{\sum_{i=1}^p \mathbb{P}(\|\mathbf{a}^{(i)}\|^2 Z_i^2 > x^2) + \mathbb{P}(\sigma^2\|\mathbf{N}\|^2\eta^2 > x^2)}{\mathbb{P}(Z > x)} \\
&\to \sum_{i=1}^p \|\mathbf{a}^{(i)}\| + \sigma\mathbb{E}\|\mathbf{N}\|\,, \quad \text{as } x \to \infty,
\end{aligned} \tag{43}$$

where the last line follows from an application of Breiman's lemma, see Breiman (1965). Taking this calculation one step further, we find that the angular measure associated with the model (41) is (see (10)),

$$\Gamma(\cdot) = w^{-1}\left(\sum_{i=1}^p \|\mathbf{a}^{(i)}\|\delta_{\frac{\mathbf{a}^{(i)}}{\|\mathbf{a}^{(i)}\|}}(\cdot) + \sigma\mathbb{E}\|\mathbf{N}\|\delta_{\frac{\mathbf{N}}{\|\mathbf{N}\|}}(\cdot)\right)\,, \tag{44}$$

where $w = \sum_{i=1}^p \|\mathbf{a}^{(i)}\| + \sigma\mathbb{E}\|\mathbf{N}\|$. In other words, $\Gamma$ has discrete mass points at $\frac{\mathbf{a}^{(i)}}{\|\mathbf{a}^{(i)}\|}$ with probability $\|\mathbf{a}^{(i)}\|/w$, $i = 1, \ldots, p$ and a uniform distribution $\mathbf{N}/\|\mathbf{N}\|$ on $\mathbb{S}^{d-1}$ with probability $\sigma\mathbb{E}\|\mathbf{N}\|/w$. This latter piece corresponds to the noise component $\sigma\varepsilon$. So the goal here is to identify the discrete components of $\Gamma$ using our method when the model does not strictly follow the LFM. Figure 2 shows pairwise scatter plots of the angular components of extremes generated from a pure signal and a noisy LFM with $\sigma > 0$.

We note that if $\sigma = 0$, then model (41) is approximately equal to the max-linear model $X = (\vee_{j=1}^k a_{1j}Z_j, \ldots, \vee_{j=1}^k a_{pj}Z_j)^\top$ and will in fact have the same asymptotic spectral measure. Intuitively, this model generates $p$ clusters of extremes since the noise term is only adding uniform noise to the angular measure.




(a) Pure signal LFM

(b) Noisy LFM

Figure 2: Pairwise scatterplots of the angular part of the extremes generated from (41) with factor loading matrix (45), $n = 125000$, $N_n = 400$ and $\sigma = \{0, 1\}$. In both cases there are two clear clusters corresponding to the signal. The red points in subfigure (b) denote extremes attributed to the signal $A^\top \mathbf{Z}_i$.

As part of a simulation study, we consider $\sigma = \{0, 1, 3, 5\}$ and choose

$$A = \begin{pmatrix} 0.1 & 0.9 \\ 0.2 & 0.8 \\ 0.3 & 0.7 \\ 0.4 & 0.6 \end{pmatrix}. \tag{45}$$

This model is similar to one of the max-linear models considered in the simulations of Janßen and Wan (2020) where our factor loading matrix $A$ can be viewed as a deterministic version of their random factor loadings. In the simulations we took two clusters for the pure signal model where $\sigma = 0$ and three clusters for the noisy model when $\sigma > 0$ as these values are suggested by the typical screeplots we observed; see Figure 3.

We compute the normalized columns of the $A$ matrix which correspond to the location of the point masses of the spectral distribution (these are the $\mathbf{c}_k$, $k = 1, 2$ in (12)). After applying our method to a single realization of size $n = 125000$ with $N_n = 400$, $k_n = \lceil \frac{400}{5 \log 400} \rceil + 1 = 15$, visualized in the pairwise scatter plots of Figure 2, we obtained the estimates of the $\mathbf{c}_k$ represented in Figure 4. These masses on the sphere are estimated by taking the mean of all members in each of the identified clusters, seen in Figure 5, and then normalizing it to lie on the unit sphere. The two panels in Figure 4 correspond to the cases of 2 clusters and no noise and two clusters with uniform noise. In the first plot, the heat map does a good job in recreating the relatives size of the mass locations. In the second panel, the first two columns of the matrix, also reproduce the relative sizes of the columns (increasing in the first and decreasing in the second) of the $A$ matrix. The third column corresponds to the cluster of points that have not been assigned to either of the first two clusters. As such they are essentially scattered uniformly around the unit sphere but away from the locations of the point masses corresponding to the first two columns. This is reflected in the third cluster having more negative values as indicated by the softer (red colors) in the heat map.

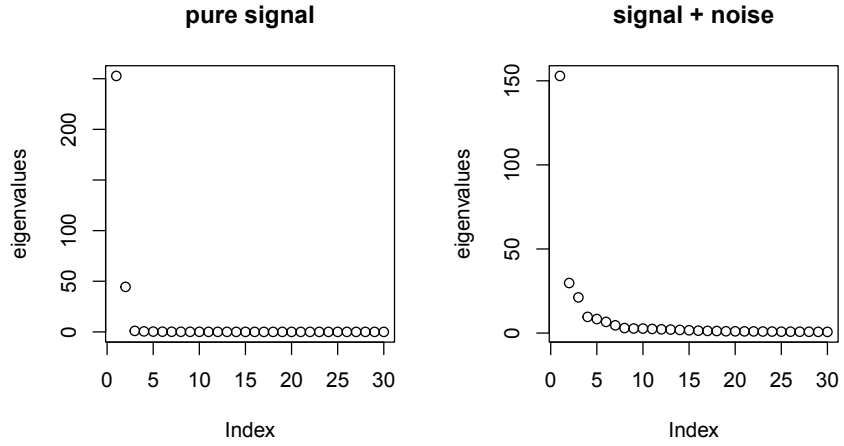
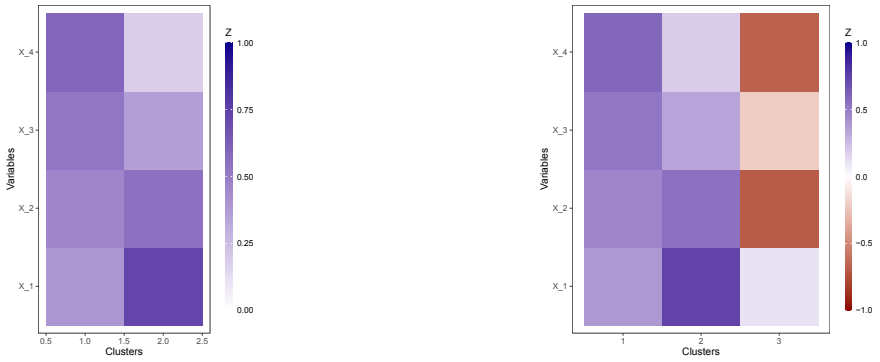


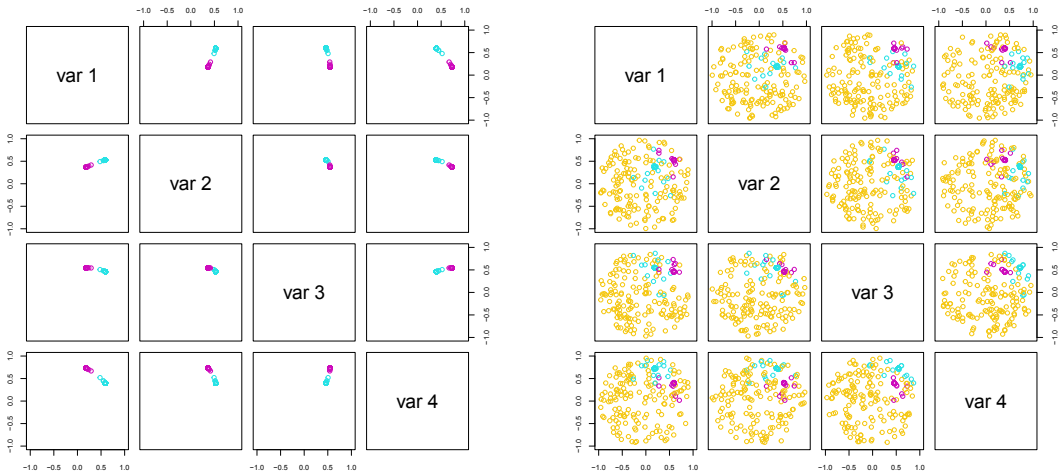

Figure 3: Screeplots of fully connected kernel matrix of pure signal and noisy linear factor models noise models.




(a) pure signal

(b) uniform noise on the sphere

Figure 4: The heat maps show the estimated cluster centers based on the cluster assignments displayed in Figure 5. The extremal sample corresponds to four dimensional extremes generated from LFM given by (41) with loading matrix (45) and $\sigma = 0$ and $\sigma = 1$ respectively. In both cases we took $n = 125000$, $N_n = 400$ and $k_n = 15$.




(a) Pure signal LFM

(b) Noisy LFM

Figure 5: Cluster assignments output of spectral clustering applied to data generated from (41) with $n = 125000$, $N_n = 400$ and $\sigma = \{0, 1\}$. In both cases spectral clustering used an extremal 15-NN graph.

A small simulation study was conducted for this LFM model with and without noise. Based on the screeplots, we used 2 clusters in the noiseless case and 3 clusters in the case with noise. The two normalized columns of the $A$ were estimated and the boxplot of the

estimation error measured in Frobenius norm are displayed in Figure 6 for $\sigma = 0$ and in Figure 7 for the case $\sigma > 0$. The succession of boxplots in each row correspond to an increasing $N_n$, with the centers and width becoming smaller. Note that the scales on the plots change across the row. The boxplots in blue correspond to our method with difference choices of nearest neighbors as a function of $C$, and the yellow boxplot is based on the spherical $k$-means approach considered in Janßen and Wan (2020). In the $\sigma = 0$ case, our method performs about the same or slightly better than the spherical $k$-means method. However, as one adds noise to the model, our method generally outperforms spherical $k$-means. In models with larger noise, it can be more difficult to estimate the LFM signal. So to compare performance across difference sample sizes and level of noise, we can calibrate by calculating a notion of *signal to noise ratio*. In this context we consider the part of the mass in the angular measure associated to the *signal* in (41), which as a function of $\sigma$ is given by

$$\mathrm{SNR}(\sigma) := \frac{\sum_{i=1}^{p} \|\mathbf{a}^{(i)}\|}{\sum_{i=1}^{p} \|\mathbf{a}^{(i)}\| + \sigma \mathbb{E}\|\mathbf{N}\|} .$$

In the absence of any noise, i.e., $\sigma = 0$, then SNR is 1 while as $\sigma \to \infty$, SNR converges to 0. For the simulation example above for which $d = 4, p = 2$, we have $\mathbb{E}\|\mathbf{N}\| = \sqrt{2}\Gamma(5/2)/\Gamma(2) = 1.880$. Hence $\mathrm{SNR}(\sigma) = 2.065/(2.065 + \sigma 1.880)$. In looking at the various plots in Figure 7, it is instructive to compute the *effective sample size* given by ESS$=$ SNR$\times N_n$. This number essentially gives the expected sample size of the number of observations, from the total $N_n$, that *come from the signal*. With this index in mind, plots that have the same ESS values (reported in the caption of Figure 7) generally show similar results since the procedures are applied to the roughly the same number of extreme observations attributed to the signal component in the model. We finally note that in this simulation $\alpha = 1$ which is not currently covered by our LFM theory but is the setting proposed in the simulations of Janßen and Wan (2020). We carried out simulations with $\alpha = 0.5$ and $\alpha = 2$ and obtained qualitatively the same type of results as the ones reported here. The only noticeable difference was that spherical $k$-means seems to work better with larger $\alpha$ in the noisy model, but is much worse for small $\alpha$. In both cases spectral clustering outperformed spherical $k$-means.

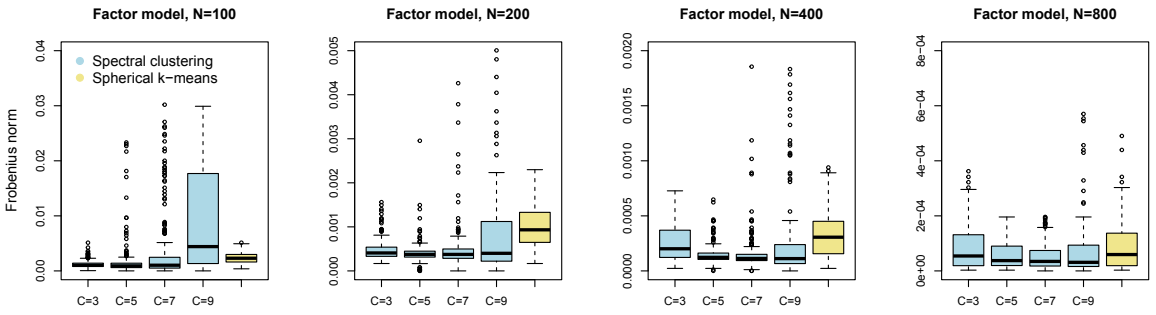



Figure 6: Estimation error measured in Frobenius norm when $\sigma = 0$.

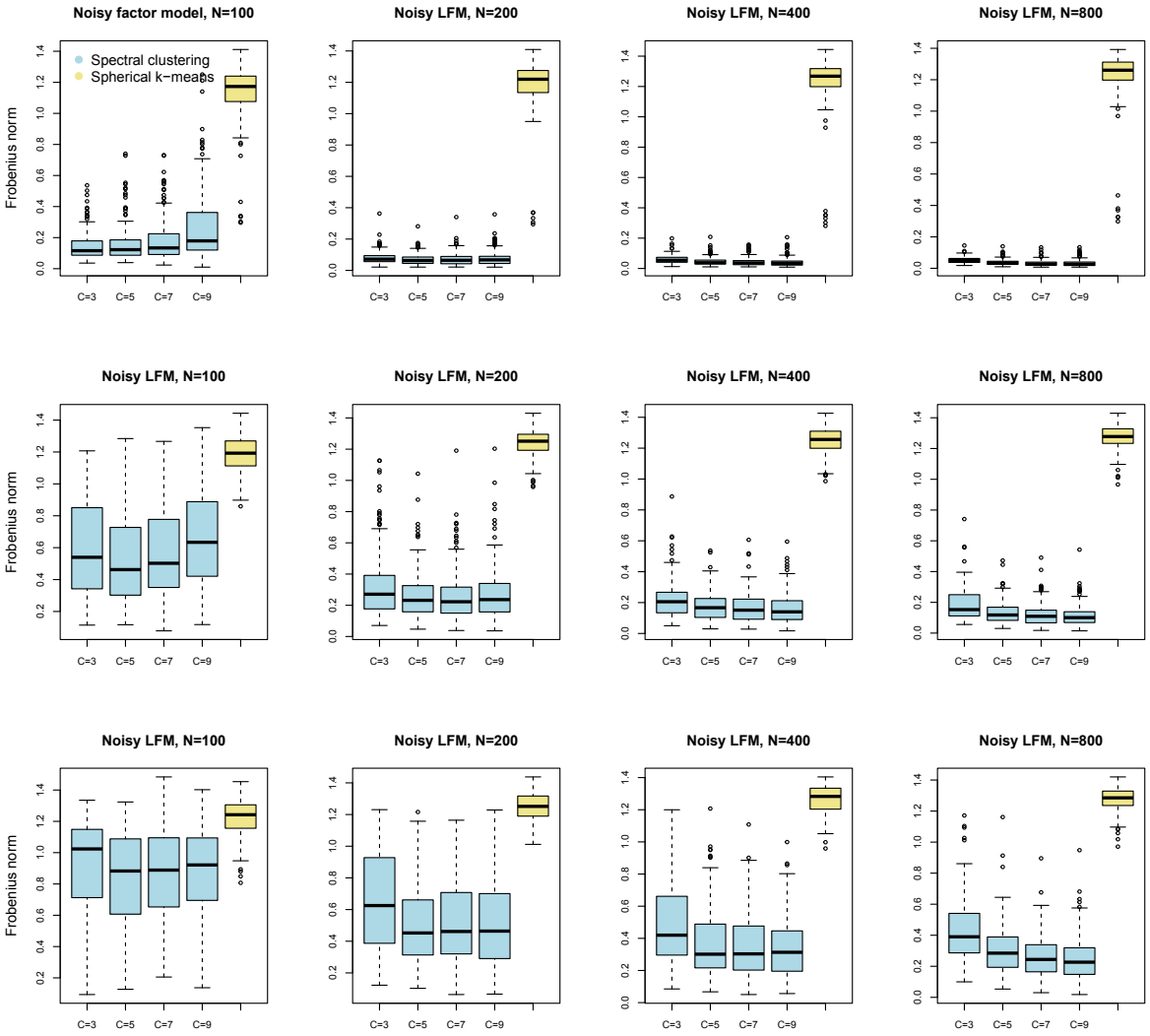



Figure 7: Estimation error of cluster centers measured in Frobenius norm. data was generated from the noisy LFM (41). The sample sizes increases from left to right as $n = \{1000, 5000, 25000, 125000\}$ and $N_n = \{100, 200, 400, 800\}$, and from noise level increases from the top down as $\sigma = \{1, 3, 5\}$. Across rows the ESS are: $\{52, 105, 209, 418\}$, $\{27, 54, 107, 214\}$, $\{18, 36, 72, 144\}$

### 6.2 Bivariate extremes from MA(3)

We consider the model discussed in the introduction and represented in Figure 1. More specifically, the model is $Y_t = Z_t + .5Z_{t-1} - .6Z_{t-2} + 1.5Z_{t-3}$, where $\{Z_t\}$ is an iid symmetric stable random variables with index $\alpha = 1.8$. We analyze the extremal dependence structure

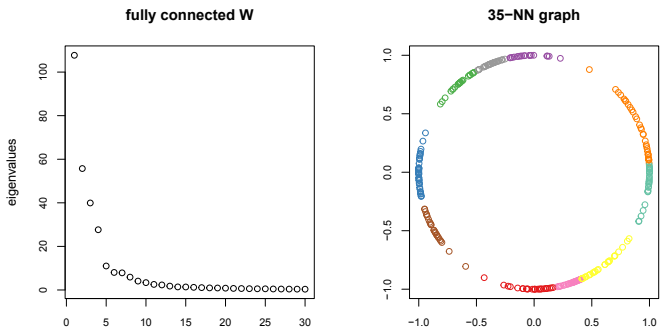



Figure 8: Screeplot of kernel matrix and clustering performance of 2 dimensional MA(3) extremes when $n = 25000$ and $N_n = 400$.

of the bivariate vector $\mathbf{X}_t = (Y_t, Y_{t-1})^\top$ by looking for clusters in the extremes of $\mathbf{X}_t$. This model can be written in the form (6) since we can define $\mathbf{Z}_t = (Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}, Z_{t-4})$ and hence

$$\mathbf{X}_t = \begin{pmatrix} 1 & 0.5 & -0.6 & 1.5 & 0 \\ 0 & 1 & 0.5 & -0.6 & 1.5 \end{pmatrix} \mathbf{Z}_t.$$

Note that even though in this case the sample $\{\mathbf{X}_t\}$ is not independent, the asymptotic distribution obtained in Theorem 2 still holds. In particular, the angular distribution is supported in the points (12) i.e.

$$\mathbf{c}_{1,\pm} = \pm(1,0), \quad \mathbf{c}_{2,\pm} = \pm(.5,1)/\sqrt{1.25}, \quad \mathbf{c}_{3,\pm} = \pm(-0.6,0.5)/\sqrt{0.61},$$
$$\mathbf{c}_{4,\pm} = \pm(1.5,-0.6)/\sqrt{2.61} \quad \text{and} \quad \mathbf{c}_{5,\pm} = \pm(0,1).$$

Figure 8 illustrates the behavior of spectral clustering for this model when $N_n = 400$ and $k_n = \lceil \frac{400}{2 \log 400} \rceil + 1 = 35$. The screeplot suggests 5 clusters corresponding to the 5 columns in the factor loading matrix. However, due to the symmetry of the actual factors in $\mathbf{Z}_t$, each column and its negative value constitute a cluster. So the 10 clusters actually reflect the 5 factors since each is paired with its negative counterpart. It is worth noting that in Figure 1 we had a larger sample size of $100,000$ and stricter quantile threshold of 0.998 resulting in smaller number of observations considered as extremes, but with an empirical distribution visibly closer to the prescribed asymptotic discrete distribution. Therefore the simulation scenario considered here is more difficult. Figure 9 illustrates the convergence of the method. While the spectral $k$-means method of Janßen and Wan (2020) performs slightly better than our spectral clustering for $N_n \leq 200$, our proposed method appears better with much smaller variability for a larger number of extremes. The choice $k_n$ of nearest neighbors did not appreciably impact the performance of spectral clustering across the different sample sizes.

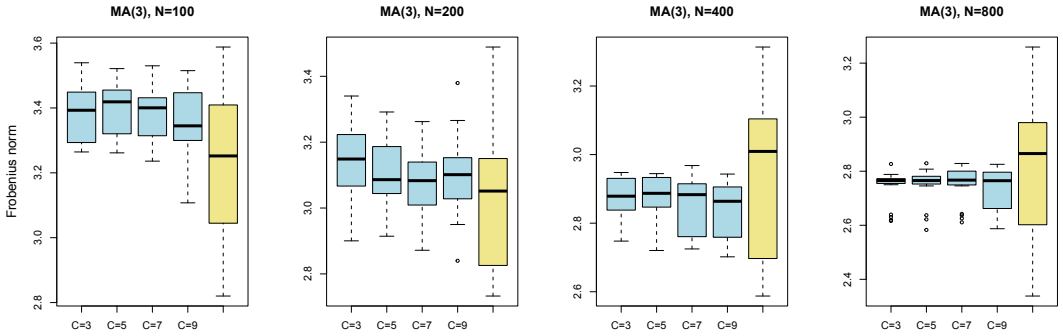



Figure 9: Estimation error of the matrix of atoms of the spectral measure of the symmetric MA(3) model. The sample sample sizes were $n = \{1000, 5000, 25000, 125000\}$ giving $N_n = \{100, 200, 400, 800\}$. We used $k_n = \lceil \frac{N_n}{C \log N_n} \rceil + 1$ nearest neighbor graphs.

### 6.3 Air pollution data

We revisit the data analyzed by Heffernan and Tawn (2004) and Janßen and Wan (2020). It is available in the `R` package `texmex` and consists of daily measurements of five air pollutants in the city of Leeds, UK. It was collected between 1994 and 1998, and split into summer and winter months yielding a total of 578 and 532 observations respectively. Following standard practice in multivariate extremes data analysis we standardize the marginal distribution of the data to focus on the extremal dependence. More specifically, we transform the marginals of the original observations $\mathbf{X}_i$ as in Janßen and Wan (2020) i.e., we let

$$Y_{ij} := 1/\{1 - F_{nj}(X_{ij})\},$$

where $F_{nj}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$ denotes the $j$th marginal empirical cumulative distribution function, $x \in \mathbb{R}$ and $j = 1, \ldots, d$. We then proceed to define the extremal observations as the 10% of the transformed observations $\{\mathbf{Y}_i\}$ with largest Euclidean norm and analyze their angular components with our algorithm. We analyze this data using spectral clustering with the exponential kernel and $s = 1$ as in the simulated data. The screeplots in Figure 10 suggest that one should consider 5 clusters for this data.

Figure 11 shows the estimated cluster centers $\mathbf{c}_j$ for $j = 1, \ldots, 5$. We note that the "elbow plot" considered by Janßen and Wan (2020) suggested the authors to use 4 or 5 clusters in their article. Our results for 5 clusters is consistent with their analysis. Specifically, the normalized cluster centers in the heat map of Figure 11 show that the extremes of the five air pollutants act mostly independent. Looking a bit more closely, both NO and NO2 share common strength in clusters 2 and 3, which is much stronger in winter than in summer. PM10 also shares a common source (cluster 2) with NO and NO2, which is more pronounced in winter than summer. For the O3 and NO2 pollutants, we examined time lagged dependence by applying the spectral clustering algorithm to the vector $\mathbf{X}_t = (X_t, X_{t-1}, X_{t-2}, X_{t-3})^T$, where $X_t$ represents either the measured value of O3 or NO2

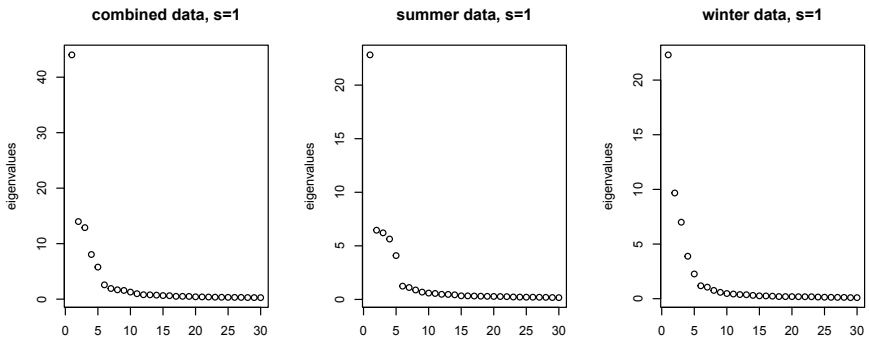
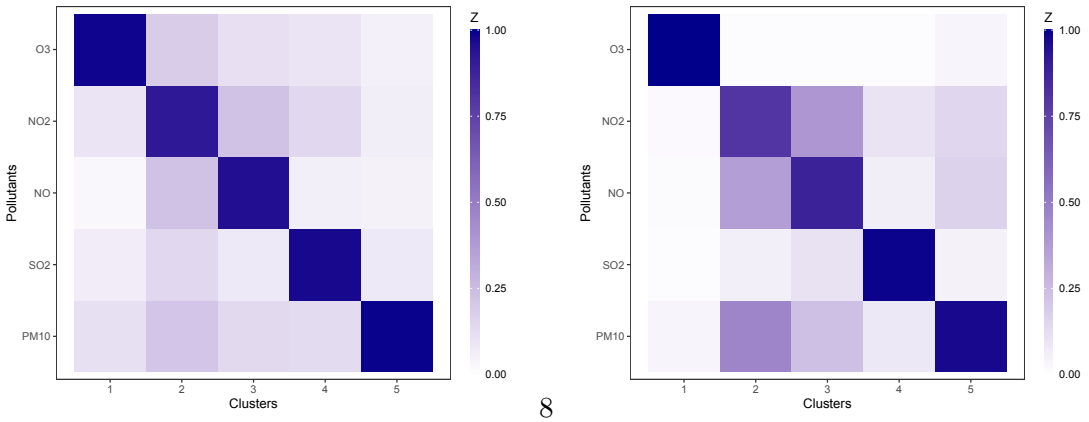


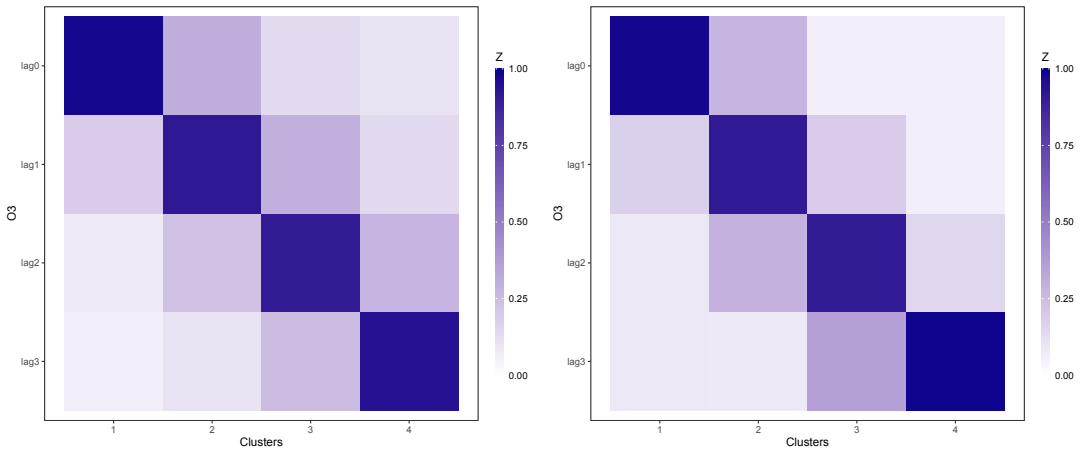

Figure 10: Screeplots of the kernel matrix of the air pollution data extremes obtained with the exponential kernel and bandwidth parameter $s = 1$ .

on day $t$. The resulting heat plots for the cluster centers (4) are displayed in Figures 12 (O3) and 13 (NO2). The super and sub diagonals reflect some extremal dependence at time lag 1 for O3 in both summer and winter. This dependence mostly dissipates after one day. The situation for NO2 is a bit more complex. One still discerns some extremal dependence at a one day lag as indicated by the high-temperature in the heat maps along the diagonal and subdiagonal. However, some clusters have similar shading for its center of mass, e.g., clusters 1 and 4 for winter, which suggests poor delineation between the clusters. In addition, there is a stronger day effect in the summer than winter for NO2 and the dependence does not necessarily die out after one day lag as in the O3 case.



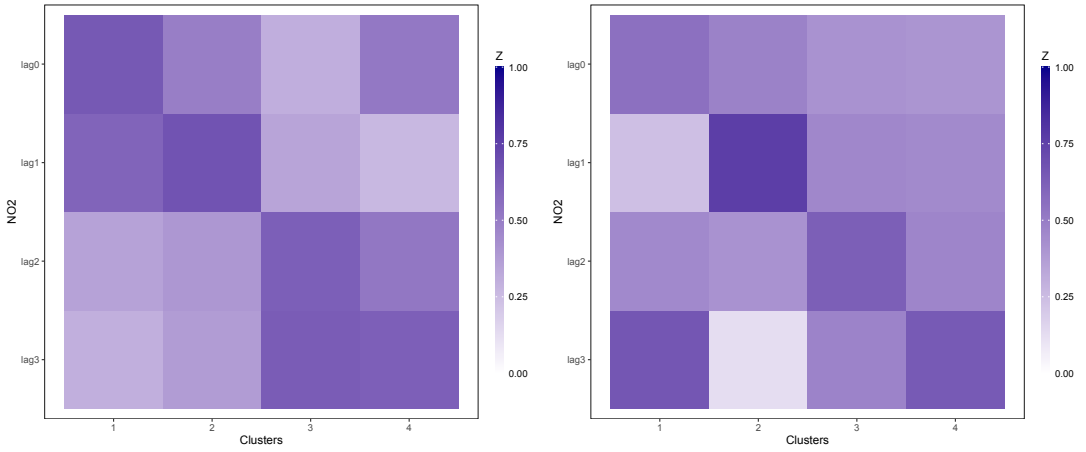

(a) summer (b) winter

Figure 11: Five dimensional extremes from air pollution summer and winter data. The heat maps show the estimated cluster centers using spectral clustering with 5 clusters and 9-nearest neighbors.




(a) summer (b) winter

Figure 12: Four dimensional time series data constructed with lags 0-3 of O3 for summer and winter data respectively. The heat maps show the estimated cluster centers using spectral clustering with 5 clusters and 9-nearest neighbors.




(a) summer (b) winter

Figure 13: Four dimensional time series data constructed with lags 0-3 of NO2 for summer and winter data respectively. The heat maps show the estimated cluster centers using spectral clustering with 5 clusters and 9-nearest neighbors.

## 7. Discussion

In this work we introduced a spectral clustering approach for learning the angular measure of multivariate extremes. We proved that this approach leads to consistent clustering for a natural linear factor model and showed the good finite sample performance of our methods in numerical experiments. The encouraging results suggest the method might be applied in more general contexts. We are particularly interested in exploring two type of extensions. First, high dimensional scenarios where the dimension of the extremes $d$ might be larger than the number of observed extremes $N_n$. This would require introducing appropriate notions of sparsity and regularization. Second, it seems natural to investigate generative models that lead to continuous angular measures in the limit. This scenario implies one would need to carefully introduce more general definitions of extremal clusters and different analysis of the convergence of $k$-nearest neighbor graphs.

## Acknowledgments

This research was partially supported by NSF grants DMS-2015379 (Avella Medina and Davis) at Columbia and DMS-2015242 (Samorodnitsky) at Cornell.

## Appendix

Before proving Lemmas 10 and 11 we will give a result regarding random partitions of uniform random variables that we will leverage as the continuity of $F_{T_n}$ implies that $F_{T_n}(T_n) \sim \text{Unif}(0,1)$. We remind the reader that in our proofs $c > 0$ represents a finite and non-zero constant whose value may change from line-to-line.

**Lemma 16** *Let $U_1, \dots, U_N \overset{iid}{\sim} Unif(0,1)$ and consider the random partition of the unit interval $I_{j,N} = [\frac{j-1}{m_N}, \frac{j}{m_N})$, where $m_N = \frac{N}{\tau \log(N)}$, $\tau > 1$ and $j = 1, \dots, m_N$. Then, with probability at least $1 - \frac{N^{1-\tau}}{\tau \log(N)}(1 + N^{-0.2\tau})$*

*(i) Every $I_{j,N}$ contains at least one of the variables $U_1, \dots, U_N$.*

*(ii) No $I_{j,N}$ contains more than $3\tau \log(N)$ of the variables $U_1, \dots, U_N$.*

**Proof** Consider the event $E_{N,1} = \{\text{Every } I_{j,N} \text{ contains at least one of the variables } U_1, \dots, U_N\}$ and note that a union bound gives

$$\begin{aligned}
\mathbb{P}(E_{N,1}) &\geq 1 - \sum_{j=1}^{m_N} \mathbb{P}(U_k \notin I_{j,N},\, \forall k = 1, \dots, N) \\
&= 1 - m_N \left(1 - \frac{1}{m_N}\right)^N \\
&\geq 1 - m_N e^{-N/m_N} \\
&= 1 - \frac{N^{1-\tau}}{\tau \log(N)}\,. \qquad (46)
\end{aligned}$$

Now consider the event $E_{N,2} = \{\text{No } I_{j,N} \text{ contains more than } 3\tau\log(N) \text{ of the variables } U_1,\dots,U_N\}$. It follows again from a union bound that yields

$$\begin{aligned}\mathbb{P}(E_{N,2}) &\geq 1 - \sum_{j=1}^{m_N} \mathbb{P}(I_{j,N} \text{ has more than } 3\tau\log(N) \text{ of the } U_1,\dots,U_N)\\ &= 1 - m_N\mathbb{P}(S_N > 3\tau\log N),\end{aligned}$$

where $S_N \sim \mathrm{Bin}(N, \frac{1}{m_N})$. Invoking Bernstein's inequality we see that

$$\mathbb{P}(E_{N,2}) \geq 1 - m_N e^{-\frac{1}{2}\frac{(3\tau\log N - \tau\log N)^2}{\tau\log N + 2\tau\log N/3}} = 1 - \frac{N^{-(1.2\tau-1)}}{\tau\log N}. \tag{47}$$

Combining (46) and (47) shows that $(i)$ and $(ii)$ hold with the desired probability. ■

**Proof of Lemma 10**

We prove the lemma for positive $Z$. The same type of arguments work in the symmetric case and are therefore omitted. Note that we can write for $t > 0$,

$$\begin{aligned}&f_{T_n}(t) = \frac{w_1^2}{c_p u_n}\cdot\\ &\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Big[z_1 f_Z(z_1)\, f_Z(z_2)\cdots f_Z(z_{p-1}) f_Z\big[\big(tz_1w_1^2/u_n - (c_2z_2+\cdots+c_{p-1}z_{p-1})\big)/c_p\big]\\ &\cdot\mathbf{1}\Big(z_1 > u_n/w_1^{1/\alpha},\, (a_{11}z_1 + a_{12}z_2 + \cdots + a_{1p}\tilde{z}_p)^2 + (a_{21}z_1 + a_{22}z_2\cdots + a_{2p}\tilde{z}_p)^2 > u_n^2\Big)\Big]\\ &\qquad\qquad dz_1\cdots dz_{p-1}\end{aligned} \tag{48}$$

$$\begin{aligned}&\div\Big[\mathbb{P}\Big((a_{11}Z_1+\cdots+a_{1p}Z_p)^2 + (a_{21}Z_1+\cdots+a_{2p}Z_p)^2 > u_n^2,\quad Z_1 > u_n/w_1^{1/\alpha}\Big)\Big]\\ &:= M_n(t)/D_n\,,\end{aligned}$$

where $\tilde{z}_p = \big(tz_1w_1^2/u_n - (c_2z_2+\cdots+c_{p-1}z_{p-1})\big)/c_p$, $c_i = a_{2i}$, $i = 1,\dots,p$. We already know that

$$D_n \sim cu_n^{-\alpha},\ n\to\infty. \tag{49}$$

Next, from (36), $\sup_z f_Z(z) = M < \infty$, we conclude by (37) that

$$M_n(t) \leq \frac{Mw_1^2}{c_p u_n}\int_{u_n/w_1^{1/\alpha}}^{\infty} z_1 f_Z(z_1)\,dz_1 \sim cu_n^{-\alpha}, \quad \text{as } n\to\infty\,.$$

Hence there exists an $G \in (0,\infty)$ such that for all $n$ large enough,

$$f_{T_n}(t) \leq G \ \text{ for all } t. \tag{50}$$

This shows $(i)$. Let us now turn to claim $(ii)$ for concreteness consider $0 < t \leq 1$. Note that, for large $n$, the indicator in (48) is bounded from below by the indicator of the set

$E = \{C^{-1}u_n < z_1 < Cu_n, |z_i| \le 1, i = 2, \dots, p-1\}$ for some large $C$. Then, on $E$, the argument of the last function $f_Z$ in (48) is within a compact interval, so we obtain

$$M_n(t) \ge c\frac{Mw_1^2}{c_p u_n} \int_{C^{-1}u_n}^{Cu_n} z_1 f_Z(z_1)\, dz_1 \sim c u_n^{-\alpha}\,,$$

where the last relation follows from a direct application of (37). Along with (49) this establishes $(ii)$.

Finally, note that

$$\begin{aligned} M_n(t) \le & \frac{w_1^2}{c_p u_n} \int_{u_n/w_1^{1/\alpha}}^{\infty} z_1 f_Z(z_1) \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_Z(z_2)\cdots f_Z(z_{p-1}) \Big[ \\ & f_Z\big[\big(tz_1w_1^2/u_n - (c_2z_2 + \cdots + c_{p-1}z_{p-1})\big)/c_p\big]\Big] dz_1 \cdots dz_{p-1} \\ = & \frac{w_1^2 u_n}{c_p} \int_{u_n/w_1^{1/\alpha}}^{\infty} z_1 f_Z(u_n z_1) \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_Z(z_2)\cdots f_Z(z_{p-1}) \Big[ \\ & f_Z\big[\big(tz_1w_1^2 - (c_2z_2 + \cdots + c_{p-1}z_{p-1})\big)/c_p\big]\Big] dz_1 \cdots dz_{p-1}\,. \end{aligned}$$

Using the upper bound in (37) it is easy to see that for some $c > 0$ and sufficiently large $t$,

$$\begin{aligned} & \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_Z(z_2)\cdots f_Z(z_{p-1}) \Big[ f_Z\big[\big(t - (c_2z_2 + \cdots + c_{p-1}z_{p-1})\big)/c_p\big]\Big] dz_2 \cdots dz_{p-1} \qquad (51) \\ & \le c t^{-(\alpha+1)}\,. \end{aligned}$$

Indeed, the integral is, up to a constant, equal to the density of a linear combination of $Z_1, \dots, Z_{p-1}$. Therefore, for all $y$ large enough, uniformly in $n$,

$$M_n(t) \le c u_n \int_{1/w_1^{1/\alpha}}^{\infty} z_1 f_Z(u_n z_1)(tz_1)^{-(\alpha+1)}\, dz_1 \le c u_n^{-\alpha} t^{-(\alpha+1)}\,,$$

where once again we have used the upper bound in (37). Together with (49) this shows the upper bound in $(iii)$. The lower bound in $(iii)$ can be established in an identical way using the lower bound in (37). ■

**Proof of Lemma 11**

It follows from Lemma 16 and Lemma 10 $(ii)$ that, outside of an event $\Omega_n^{(1)}$ with $\mathbb{P}\big(\Omega_n^{(1)}\big) \to 0$, each one of the intervals $I_{i,n}$ contains at least one of the points

$$T_{ni} = \frac{a_{21}Z_{2,i}^{(*,1)} + \cdots + a_{p1}Z_{p,i}^{(*,1)}}{w_1^2 Z_{1,i}^{(*,1)}/u_n}, \quad i = 1, \dots, N_n^{(1)},$$

and none of the intervals contains more than $3\tau \log N_n^{(1)}$ of these points. Note that (50) implies that

$$\frac{\partial}{\partial t} F_{T_n}^{-1}(t) = \frac{1}{f_{T_n}(F_{T_n}^{-1}(t))} \ge \frac{1}{G}, \quad \forall t \in [0,1]$$

and hence by the fundamental theorem of calculus

$$F_{T_n}^{-1}\left(\frac{i}{m_n}\right) - F_{T_n}^{-1}\left(\frac{i-1}{m_n}\right) \geq \frac{1}{Gm_n}$$

This shows that the length of the intervals $I_{i,n}$ satisfies

$$|I_{i,n}| \geq l_n/G, \quad i = 1, \ldots, m_n, \tag{52}$$

where $l_n = \frac{1}{m_n}$. Since the conditional law of $(Z_1/u_n, Z_2, \ldots, Z_p)$ given (35) converges weakly, as $n \to \infty$, to the law of

$$\big(W_\alpha/w_1, Z_2, \ldots, Z_p\big)$$

as defined in Theorem 2, we see that

$$F_{T_n} \Rightarrow G := \text{the law of } \frac{c_2 Z_2 + \cdots + c_p Z_p}{W_\alpha}.$$

It follows that the values $F_{T_n}(t_0)$ converge, as $n \to \infty$, to a finite limit. Therefore, there is $0 < \delta < 1$ such that $F_{T_n}^{-1}\big((i-1)/m_n\big) \geq t_0$ for all $n$ large enough and all $i \geq (1-\delta)m_n$. We conclude by Lemma 10 $(iii)$ that for such $n$ and $i$,

$$\begin{aligned} l_n =& F_{T_n}\Big(F_{T_n}^{-1}\big(i/m_n\big)\Big) - F_{T_n}\Big(F_{T_n}^{-1}\big((i-1)/m_n\big)\Big) \\ \in& \big(D^{-1}, D\big) \int_{F_{T_n}^{-1}\big((i-1)/m_n\big)}^{F_{T_n}^{-1}\big(i/m_n\big)} t^{-(\alpha+1)}\, dt. \end{aligned} \tag{53}$$

Furthermore,

$$\int_{F_{T_n}^{-1}\big((i-1)/m_n\big)}^{F_{T_n}^{-1}\big(i/m_n\big)} t^{-(\alpha+1)}\, dt \geq \Big(F_{T_n}^{-1}\big(i/m_n\big)\Big)^{-(\alpha+1)} \Big(F_{T_n}^{-1}\big(i/m_n\big) - F_{T_n}^{-1}\big((i-1)/m_n\big)\Big), \tag{54}$$

while leveraging again Lemma 10 $(iii)$ we see that

$$\frac{m_n - i}{m_n} = \int_{F_{T_n}^{-1}\big(i/m_n\big)}^{\infty} f_{T_n}(t)\, dt \leq D \int_{F_{T_n}^{-1}\big(i/m_n\big)}^{\infty} t^{-(\alpha+1)}\, dt = \frac{D}{\alpha}\Big(F_{T_n}^{-1}\big(i/m_n\big)\Big)^{-\alpha}. \tag{55}$$

Combining (54) and (55), we conclude that

$$\int_{F_{T_n}^{-1}\big((i-1)/m_n\big)}^{F_{T_n}^{-1}\big(i/m_n\big)} t^{-(\alpha+1)}\, dt \geq c\left(\frac{m_n - i}{m_n}\right)^{(\alpha+1)/\alpha} \Big(F_{T_n}^{-1}\big(i/m_n\big) - F_{T_n}^{-1}\big((i-1)/m_n\big)\Big),$$

and so by (53),

$$l_n \geq c\left(\frac{m_n - i}{m_n}\right)^{(\alpha+1)/\alpha} |I_{i,n}|.$$

Since an upper bound can be obtained in the same way, we conclude that for some $D_1 \geq 1$, for all $n$ large enough and all $i \geq (1-\delta)m_n$,

$$D_1^{-1} l_n \left(\frac{m_n - i}{m_n}\right)^{-(\alpha+1)/\alpha} \leq |I_{i,n}| \leq D_1 l_n \left(\frac{m_n - i}{m_n}\right)^{-(\alpha+1)/\alpha}. \tag{56}$$

It follows from (56) that for $K_0$ fixed , $|i-j| \leq K$ and all $(1-\delta)m_n \leq i, j \leq m_n - K_0$,

$$\frac{|I_{i,n}|}{|I_{j,n}|} \geq D_1^{-2} \left(\frac{m_n - i}{m_n - j}\right)^{-\frac{(\alpha+1)}{\alpha}} \geq D_1^{-2} \left(1 + \frac{K}{K_0}\right)^{-\frac{(\alpha+1)}{\alpha}} \geq C_1\,, \tag{57}$$

where the second to last inequality holds for sufficiently large $m_n$. A similar argument can be used to get the upper bound

$$\frac{|I_{i,n}|}{|I_{j,n}|} \leq D_1^{2} \left(1 - \frac{K}{K_0}\right)^{-\frac{(\alpha+1)}{\alpha}} \leq C_2\,. \tag{58}$$

It follows that for $K_0$ fixed and large enough $n$,

$$C_1 \leq \frac{|I_{i,n}|}{|I_{j,n}|} \leq C_2\,. \tag{59}$$

Therefore, using (59), $|i-j| \leq K$ and choosing $K \geq 2C_2/C_1$,

$$|I_{i,n}| + |I_{i+1,n}| \leq 2C_2|I_{i-1,n}| \leq K|I_{i-k,n}|, \tag{60}$$

where the last inequality holds for all $k = 1, \ldots, K$. Therefore dividing (60) by $K$ and summing over $k$ yields

$$|I_{i,n}| + |I_{i+1,n}| < \sum_{k=1}^{K} |I_{i-k,n}| \tag{61}$$

Note that (61) is sufficient to guarantee that any point in $I_{i,n}$ is closer to any point in $I_{i+1,n}$ that to any point in an interval $I_{j,n}$ with $j < i - K$. A similar argument shows that any point in $I_{i,n}$ is closer to any point in $I_{i-1,n}$ that to any point in an interval $I_{j,n}$ with $j > i + K$. We conclude that, outside of the the event $\Omega_n^{(1)}$, in a $k_n$-NN graph with

$$k_n > 3(K+1)\tau \log N_n^{(1)}, \tag{62}$$

then all points $\left(T_{nj},\, j = 1, \ldots, N_n^{(1)}\right)$ within $I_{i,n}$ for some $i$ in the range $(1-\delta)m_n \leq i \leq m_n - K_0$ will be connected both to each other and to such a point in each $I_{i-1,n}$ and $I_{i+1,n}$. The next observation to make is that, as long as $\delta$ is small enough, the sequence $\left(F_{T_n}^{-1}(1-\delta)\right)$ is bounded from above. Therefore, by Lemma 10 $(ii)$, uniformly in large enough $n$, the density $f_{T_n}$ is bounded from below by, say, $a > 0$ on the interval $\left(0, F_{T_n}^{-1}(1-\delta)\right)$. Therefore, for all large enough $n$,

$$|I_{i,n}| \,\leq\, l_n/a, \quad 1 \leq i \leq (1-\delta)m_n. \tag{63}$$

To see this, note that

$$\frac{\partial}{\partial t} F_{T_n}^{-1}(t) = \frac{1}{f_{T_n}(F_{T_n}^{-1}(t))} \leq \frac{1}{a}, \quad \forall t \in [0,1]$$

and hence by the fundamental theorem of calculus

$$|I_{i,n}| = F_{T_n}^{-1}\left(\frac{i}{m_n}\right) - F_{T_n}^{-1}\left(\frac{i-1}{m_n}\right) \leq \frac{1}{am_n} = \frac{l_n}{a}.$$

It follows from (52) and (63) that if $K > \frac{2G}{a}$ then any point in $I_{i,n}$ is closer to any point in $I_{i-1,n}$ and in $I_{i+1,n}$ than to any point in an interval $I_{j,n}$ with $j < i - K$ or with $j > i + K$. Therefore, on the event $\Omega_n^{(1)}$, in a $k_n$-NN graph satisfying (62), all points $\big(T_{nj},\, j = 1, \ldots, N_n^{(1)}\big)$ within $I_{i,n}$ in the range $1 \le i \le (1-\delta)m_n$ will be connected both to each other and to such a point in each $I_{i-1,n}$ and $I_{i+1,n}$. Indeed, to show this it suffices to show again that (61) holds true in the range $1 \le i \le (1-\delta)m_n$. It is easy to see that (52), (63) and $K > \frac{2G}{a}$ entail

$$|I_{i,n}| + |I_{i+1,n}| \le \frac{2l_n}{a} < \frac{Kl_n}{G} \le \sum_{k=1}^{K} |I_{i-k,n}|.$$

Finally, it is obvious that if $K > K_0$, then on the same event $\Omega_n^{(1)}$, in a $k_n$-NN graph satisfying (62), all points $\big(T_{nj},\, j = 1, \ldots, N_n^{(1)}\big)$ within $I_{i,n}$ in the range $m_n - K_0 < i \le m_n$ will be connected both to each other and to a such a point in each $I_{i-1,n}$ and $I_{i+1,n}$.

Summarizing the above discussion we conclude that on the event $\Omega_n^{(1)}$, in a $k_n$-NN graph satisfying (62) with $K$ large enough, all points $\big(T_{nj},\, j = 1, \ldots, N_n^{(1)}\big)$ within $I_{i,n}$ in the entire range $1 \le i \le m_n$ will be connected both to each other and to a such a point in each $I_{i-1,n}$ and $I_{i+1,n}$. In particular, the $k_n$-NN graph will be connected.

We now translate this discussion to the random vectors $\mathbf{M}^{(i)}$, $i = 1, \ldots, N_n^{(1)}$. We define intervals along vector $\mathbf{b}$ by

$$J_{i,n} = I_{i,n}\mathbf{b},\ = 1, \ldots, N_n^{(1)}\,.$$

Then, outside of the event $\Omega_n^{(1)}$, each one of these intervals contains at least one of the points $\big(\mathbf{M}^{(i)},\, i = 1, \ldots, N_n^{(1)}\big)$ and none of the intervals contains more than $3\tau \log N_n^{(1)}$ of these points. By (52) the lengths of these intervals satisfy for some $G_1 > 0$,

$$|J_{i,n}| \ge l_n/G_1, \quad i = 1, \ldots, m_n\,.$$

We finally note that by (23), with probability tending to one $N_n \sim Cnu_n^{-\alpha}$, and therefore $G > 0$ and $n$ large enough ensure that (62) holds provided $k_n > G \log n$. This concludes the proof. ∎

**Proof of Theorem 12**

Lemma 11 gives us the connectivity of the extremal $k_n$-NN graph for $k_n$ satisfying (62) with $K$ large enough. The next step is to understand by how much the points $\big(\mathbf{M}^{(i)},\, i = 1, \ldots, N_n^{(1)}\big)$ are shifted by adding to them $\big(\mathbf{D}^{(i)},\, i = 1, \ldots, N_n^{(1)}\big)$ in (33). Denote $\Omega_n^{(2)} = B_n^c$ as defined in Lemma 5. Then $\mathbb{P}\big(\Omega_n^{(2)}\big) \to 0$ as $n \to \infty$ and it is elementary to check that outside of $\Omega_n^{(2)}$ we have $\|\mathbf{D}^{(i)}\| \le ch_n^2/u_n$ for all $i = 1, \ldots, N_n^{(1)}$. Recall that by the choice of $h_n$ we have

$$h_n^2/u_n = o(l_n) \ \text{ as } n \to \infty.$$

If we define new sets by

$$\tilde{J}_{i,n} = \big\{\mathbf{M}^{(j)} + \mathbf{D}^{(j)} : \mathbf{M}^{(j)} \in J_{i,n}\big\},\ i = 1, \ldots, m_n,$$

then it follows immediately that for large $n$, outside of the event $\Omega_n^{(2)}$, the new sets have the property described by Lemma 11, perhaps with a larger $K_0$. We already know that this means that for large $n$, outside of $\Omega_n^{(1)} \cup \Omega_n^{(2)}$, the extremal $k_n$-NN graph with $k_n$ satisfying (62) with $K$ large enough, is connected. □

## References


Bojan Basrak, Richard A Davis, and Thomas Mikosch. A characterization of multivariate regular variation. *Annals of Applied Probability*, 12(3):908–920, 2002.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

Leonard Breiman. On some limit theorems similar to the arc-sin law. *Theory of Probability & Its Applications*, 10(2):323–331, 1965.

Emilie Chautru. Dimension reduction in multivariate extreme value analysis. *Electronic Journal of Statistics*, 9(1):383–418, 2015.

Stéphan Clémençon, Hamid Jalalzai, Stéphane Lhaut, Anne Sabourin, and Johan Segers. Concentration bounds for the empirical angular measure with statistical learning applications. *Bernoulli*, 29(4):2797–2827, 2023.

Daniel Cooley and Emeric Thibaud. Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106(3):587–604, 2019.

Richard A Davis and Sidney I Resnick. Basic properties and prediction of max-arma processes. *Advances in Applied Probability*, 21(4):781–803, 1989.

Anthony C Davison and Raphaël Huser. Statistics of extremes. *Annual Review of Statistics and its Application*, 2:203–235, 2015.

Inderjit S Dhillon and Dharmendra S Modha. Concept decompositions for large sparse text data using clustering. *Machine Learning*, 42(1):143–175, 2001.

Holger Drees and Anne Sabourin. Principal component analysis for multivariate extremes. *Electronic Journal of Statistics*, 15(1):908–943, 2021.

Paul Embrechts and Charles M Goldie. On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society*, 29 (2):243–256, 1980.

Sebastian Engelke and Jevgenijs Ivanovs. Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8:241–270, 2021.

Nadine Gissibl and Claudia Klüppelberg. Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A):2693–2720, 2018.

Nicolas Goix, Anne Sabourin, and Stéphan Clémen. Learning the dependence structure of rare events: a non-asymptotic study. In *Conference on Learning Theory*, pages 843–860. PMLR, 2015.

Nicolas Goix, Anne Sabourin, and Stephan Clémençon. Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*, 161:12–31, 2017.

Peter Hall, Liang Peng, and Qiwei Yao. Moving-maximum models for extrema of time series. *Journal of Statistical Planning and Inference*, 103(1-2):51–63, 2002.

Janet E Heffernan and Jonathan A Tawn. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B*, 66(3):497–546, 2004.

Bruce Hendrickson and Robert Leland. An improved spectral graph partitioning algorithm for mapping parallel computations. *SIAM Journal on Scientific Computing*, 16(2):452–469, 1995.

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

Claudia Klüppelberg and Steffen Lauritzen. Bayesian networks for max-linear models. In *Network Science*, pages 79–97. Springer, 2019.

Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, 43(1):215–237, 2015.

Nicolas Meyer and Olivier Wintenberger. Sparse regular variation. *Advances in Applied Probability*, 53(4):1115–1148, 2021.

Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856, 2002.

Sidney I Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media, 2007.

Sidney I Resnick. *Extreme values, regular variation and point processes*. Springer, New York, 2018.

Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, 39(4):1878–1915, 2011.

Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

Emma S Simpson, Jennifer L Wadsworth, and Jonathan A Tawn. Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532, 2020.

Rafael Van Driessche and Dirk Roose. An improved spectral bisection algorithm and its application to dynamic load balancing. *Parallel Computing*, 21(1):29–48, 1995.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4): 395–416, 2007.

Zhixin Zhou and Arash A Amini. Analysis of spectral clustering algorithms for community detection: the general bipartite setting. *Journal of Machine Learning Research*, 20(1): 1774–1820, 2019.