Stochastic ordering and sparse approximation of multivariate extremal dependence

Michela Corradini
School of Mathematics
Cardiff University

A thesis submitted for the degree of
Doctor of Philosophy

July 2024

Abstract

While traditional statistical methodologies focus on the average behaviour of a system and quantifying deviations from it, practical concern often lies on the extremal behaviour of a system. Understanding the statistics of extremes is crucial for answering questions arising from practical applications, such as the severity of floods or the extent of financial losses. Most real-world phenomena involve multiple variables, which may exhibit interaction at extreme levels; this motivates the study of multivariate extremes, the study of distributional tails of multivariate random vectors when two or more variables in the vector can be large together.

This thesis addresses two challenges in multivariate extreme value theory.

First, new (and highly non-trivial) stochastic orderings among multivariate extreme value distributions are revealed. More precisely, we consider the multivariate stochastic orders of upper orthants, lower orthant and positive quadrant dependence (PQD) among simple max-stable distributions and their exponent measures. The main result shows that each of these orders holds for the max-stable distribution if and only if it holds for the corresponding exponent measure. Popular parametric models such as the Dirichlet and Hüsler-Reiß families are shown to be ordered according to the aforementioned multivariate stochastic orderings.

Second, this thesis proposes a new method for estimating a sparse but accurate representation of the spectral measure, which contains the information about the dependence structure of multivariate extremes. In order to obtain such sparse approximations, we introduce techniques from the kernel mean embedding of measures to the context of spectral measure estimation in multivariate extremes. A broad range of numerical experiments shows that this is a promising approach.

Contents

List of Figures

List of Tables

Dissemination of Work

Publications

- Corradini, M. and Strokorb, K. (2024), 'Stochastic ordering in multivariate extremes', Extremes pp. 1–40. (Corradini and Strokorb, 2024)

Talks

- Decompositions of dependence in extremes. Welsh Mathematics Colloquium, Gregynog (Wales), May 2022.
- Kernel embeddings for sparse angular measure approximation. Welsh Mathematics Colloquium, Gregynog (Wales), May 2023.
- Kernel embeddings for sparse angular measure approximation. EVA International Conference, Milan (Italy), June 2023.
- Kernel embeddings for sparse angular measure approximation. STOR-i Extremes Workshop (STEW), Lancaster (UK), September 2023.

To my dad,
Stefano Corradini "Britt",
and to my grandparents,
Aldo and Alma.

Al mio papà,
Stefano Corradini "Britt",
e ai miei nonni,
Aldo e Alma.

Acknowledgements

Firstly, I want to express my deepest gratitude to my supervisor Dr. Kirstin Strokorb for the relentless support and guidance during the last four years. I cannot thank you enough for always having an open door (particularly during very busy times), your availability to help, your empathy and the insightful discussions we've had that enabled me to grow as a researcher and as a person. Likewise, I wish to thank Dr. Bertrand Gauthier for encouraging me to pursue this journey, the valuable input to this project and the availability to help and answer questions throughout. I hope there will be occasions in the future to work with both of you again.

I also wish to acknowledge EPSRC for providing the generous funding without which this research could not have happened. I greatly appreciate the financial support which allowed me to participate and present at international conferences throughout my time as a PhD student.

I would like to say thanks to the School of Mathematics staff for providing a propitious and supporting environment and for several development opportunities, from teaching to outreach. In particular, I'd like to mention Dr. Ana Ros Camacho, Dr. Katerina Kaouri, Prof. Angela Mihai and Prof. Federica Dragoni for your kind words, supporting me in my professional development and most of all, for being inspirational women in mathematics.

A big thank you to all of my fellow PhD colleagues, who have made my experience very enjoyable. Whilst I cannot possibly mention all of you individually, I would like to name a few. To my academic brother Matt, for sharing this journey with me from the beginning till the very end and for always answering all the stupid questions I had during this time. To Lizzie, for being my number one supporter and consuming copious amounts of tea on a daily basis with me. To Matthew, for helping me with all I took on other than my research and acting as my PA throughout (especially when toning down all my angry emails). To Tim and Josh, for allowing me to talk nonsense to you anytime I felt like it (when I should have let you carry on writing your theses instead). Procrastination Squad, it truly has been an honour. Please keep up the good work after I'm gone, since we all need to be distracted a little from time to time.

I would also like to extend my heartfelt gratitude to my family and my friends outside academia.

Grazie mamma, grazie papà. Per avermi sempre supportata in ogni mia scelta e per tutti i vostri sacrifici che mi hanno permesso di essere qui oggi. Spero di avervi resi fieri di me.

To Fran and Jodie, for keeping me sane by reminding me there is an entire world outside academia. To my climbing crew CJ, Joe, Adib, Aric and Panos, for listening to me complain about not being able to solve problems, whether it be a move on a boulder on the wall or the countless difficulties I faced whilst carrying out my research and writing this thesis.

To Russell, for believing in me and encouraging me to follow this path when I was doubting myself.

Lastly, and perhaps most importantly, thank you Mike. Thank you for losing at pool to me that one time and allowing me to mock you about it forever afterwards. On a serious note, thank you for always reading me like an open book, always knowing when things were not going well and, in those moments, for helping me to get out of my head and reminding me of what really matters in life. I can confidently say that, without you, this thesis would never have seen the light of day. From the bottom of my heart, thank you.

Chapter 1

Introduction

Most methodologies that we encounter in statistics focus on the description of average or typical behaviour of a system, to explain changes of typical behaviour or to quantify a deviation from it. On the other hand, especially in view of its impacts, practical concern often lies on the extremal behaviour of a system. How often do we expect to encounter a flood of a certain severity? How large do we expect a certain financial loss is going to be? Which kind of wind speeds should skyscrapers be able to withstand during their design life and beyond? Regulatory frameworks, e.g. in the financial, infrastructure or energy sectors often demand answers to such questions. Due to the rare nature of extreme events, such answers are usually based on limited data and require adequate extrapolation. Understanding and interpreting the statistics of extremes is thus crucial for risk assessment and management, so that they can inform mitigation and adaptation strategies that can reduce extreme events' societal, environmental or industrial impacts and make infrastructure more resilient.

Almost all real-world phenomena involve multiple variables, which may exhibit interaction at extreme levels. Weather phenomena affect an entire landscape and are measured at multiple stations at a certain frequency in time and they have complex interactions; financial assets exhibit co-movements, where sometimes the interaction only becomes apparent at extreme levels, e.g. during financial crises; train delays are linked through complex dependencies, including the underlying rail system. Such phenomena motivate the study of multivariate extremes, i.e. the study of extremal behaviour of more than one variable.

From a theoretical point of view, the information about the extremal dependence between variables can be summarised by a number of different objects, depending on the precise framework at hand. Two such objects are a *simple max-stable distribution* (or, equivalently, extreme value copula) or the *spectral (or angular) measure* of a regularly varying random vector, which are closely linked via *multivariate max-domain-of-attraction conditions*, cf. Chapter 2 and

Section 2.3. Among the multiple challenges associated with the study of multivariate extremes, this thesis contributes to two research directions:

(I) the study of *stochastic ordering* among multivariate extreme value distributions, which is based on Corradini and Strokorb (2024); this part is addressed in Chapters 2 (for background and preliminaries) and 3 (for results),

and

(II) obtaining *sparse representations* of extremal dependence in high-dimensional multivariate extremes; this part is addressed in Chapters 2 and 4 (for background and preliminaries) and 5 (for results).

While we give further background, motivation and context for each of (I) and (II) in Chapters 3 and 5, respectively, let us briefly summarise the main contributions here.

As for (I), we consider the *multivariate stochastic orders* of upper orthants, lower orthants and positive quadrant dependence (PQD) among *simple max-stable distributions* (i.e. extreme value copulas) and their *exponent measures*. Positive quadrant order is a concordance order, where the components of a random vector tend to exhibit stronger positive association, with large and small values more likely to occur jointly, than under independence. Previously, the extremes literature has been focusing on orderings with respect to lower orthants, mostly, as it corresponds to the ordering of stable tail dependence functions, cf. Section 2.1. Considering (non-trivial) stochastic orderings with respect to *upper orthants* among max-stable distributions is new to the best of our knowledge. It corresponds to studying the stochastic ordering of the min-combinations that arise from max-stable random vectors. Generally, relatively little is known about the minima of max-stable random vectors. It is shown for each order that it holds for the max-stable distribution if and only if it holds for the corresponding exponent measure. The finding is in particular non-trivial for upper orthants (and hence PQD order). We also demonstrate that from dimension $d \geq 3$ these three orders are not equivalent and a variety of phenomena can occur, as we illustrate through popular parametric models such as the Dirichlet family and the Hüsler-Reiß family. However, every max-stable distribution PQD-dominates the corresponding independent model and is PQD-dominated by the fully dependent model. For statistical inference, stochastic orderings are typically of concern for distributionally robust inference, as discussed in Chapter 3.

On the other hand, regarding (II), this thesis explores how techniques from the *kernel mean embedding of measures* literature can be utilised to obtain a

sparse approximation of the extremal dependence of a regularly varying random vector. One of the main challenges in multivariate extremes is how to compress dependence information without resorting to the full empirical spectral measure or drawing on parametric models, which can be overly restrictive. We explore the potential for such a kernel-mean-embedding-based approach in approximating the empirical spectral measure with few support points, leading to a sparser approximation than the full empirical angular measure but maintaining a similar accuracy in estimating the original spectral measure. To the best of our knowledge, introducing reproducing kernel Hilbert space techniques to the study of spectral measures has only been done in Avella-Medina et al. (2022) before, and therein in an essentially different way. Our numerical experiments and results show that this is a promising tool to estimate a sparse and accurate representation of the spectral measure in multivariate extremes.

The thesis is accordingly structured as follows.

In Chapter 2 we recall relevant background knowledge from the theory of multivariate extremes. We highlight some important parametric families which will become relevant for the study of stochastic orders in Chapter 3. We cover the connections between multivariate extremes and regular variation, which is the underlying framework for the kernel mean embedding, cf. Chapter 5.

In Chapter 3 we first give some background on multivariate stochastic orderings, before moving on to our theoretical results concerning stochastic orderings in multivariate extremes. Besides exploring the connection between max-stable distributions and their exponent measures, we study these orderings for the parametric families introduced in Chapter 2.

In Chapter 4, we recall definitions and facts about reproducing kernel Hilbert spaces and explain their role in the kernel mean embedding of measures and the definition of the maximum mean discrepancy (MMD), a convenient kernel-based metric to compare the distance of measures. We then cover the sparsity-inducing vertex exchange algorithm, which optimises a regularised version of the MMD, and which gives rise to our new approach of summarising extremal dependence information in Chapter 5.

Finally, in Chapter 5 we apply these ideas to the context of multivariate extremes. We explore the benefits of sparsifying empirical angular measures by means of using the vertex exchange algorithm in order to solve a regularised minimisation problem that is based on the maximum mean discrepancy for measuring distance between measures (as introduced in Chapter 4). In order to study such an approach in an appropriate setting, we first introduce a convenient stochastic

model and show that it defines a rich class of multivariate regularly varying random vectors. This stochastic model is convenient in that it is easy to simulate from and provides us with situations where we know the true spectral measure; moreover, it allows us to test our approach widely since one can retrieve any spectral measure and control the amount of noise around the spectral measure. Subsequently we carry out a broad numerical study based on such models, and document and interpret their outcomes.

Some parts of this thesis are based on the joint publication "Corradini, M. and Strokorb, K. (2024), 'Stochastic ordering in multivariate extremes', Extremes pp. 1–40." (Corradini and Strokorb, 2024). These are:

- Sections 2.1 and 2.2 of Chapter 2,
- Chapter 3,
- Appendices A and B.

The material presented in these appendices (Appendices A and B) has deliberately not been included in the main body of the thesis, as it is largely based on original contributions of lead supervisor K. Strokorb. It is included here for completeness, as Proposition A.0.9 is important to derive Theorem 3.2.1.

For the remaining material, the role of K. Strokorb was supervisory with the research carried out and documented by M. Corradini. In particular, Chapter 3 contains complementary details to Corradini and Strokorb (2024).

Chapter 2

Multivariate extremes and regular variation

This section is preparing the setting for the two main streams of research in this thesis. Our main results on stochastic orderings, presented in Chapter 3, concern stochastic orderings among max-stable distributions, or, as it turns out, equivalently, orderings among their respective exponent measures. Therefore, this chapter reviews some basic well-known results from the theory of multivariate extremes. Secondly, we will take a closer look at three marginally closed parametric families, the Dirichlet family, the Hüsler-Reiß family and the Choquet (Tawn-Molchanov) family of max-stable distributions, each model offering a different insight into phenomena of orderings among multivariate extremes.

Subsequently, we recall the connections between (multivariate) extremes and regular variation, setting the scene for our underlying framework in Chapter 5, when we study how to compress the information contained in an empirical angular measure.

While most of this chapter brings together relevant background knowledge from the literature, the representation of the Dirichlet model with a Gamma generator seems to be new in this generality, cf. Remark 2.2.2.

2.1 Max-stable random vectors and their exponent measures

In this section we recall some definitions and basic facts about representations for max-stable distributions, cf. also Resnick (1987) or Beirlant et al. (2004). Operations and inequalities between vectors are meant componentwise. We abbreviate $\mathbf{0} = (0, 0, \ldots, 0)^\top \in \mathbb{R}^d$.

Definition 2.1.1. A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \in \mathbb{R}^d$ is called *max-stable* if for all $n \geq 1$ there exist suitable norming vectors $\boldsymbol{a}_n > \mathbf{0}$ and $\boldsymbol{b}_n \in \mathbb{R}^d$,

such that the distributional equality

$$\max_{j=1,\dots,n}(\boldsymbol{X}_j) \stackrel{d}{=} \boldsymbol{a}_n \boldsymbol{X} + \boldsymbol{b}_n \tag{2.1}$$

holds, where $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are i.i.d. copies of $\boldsymbol{X}$.

According to the Fisher-Tippett theorem (Fisher and Tippett, 1928), each marginal distribution $G_i(x) = \mathbb{P}(X_i \leq x)$ is a univariate max-stable distribution, that is, either degenerate to a point mass or a generalised extreme value (GEV) distribution of the form

$$G_\xi((x-\mu)/\sigma) \quad \text{with} \quad G_\xi(x) = \begin{cases} \exp(-(1+\xi x)_+^{-1/\xi}) & \text{if } \xi \neq 0, \\ \exp(-e^{-x}) & \text{if } \xi = 0, \end{cases} \tag{2.2}$$

where $\xi \in \mathbb{R}$ is a shape-parameter, while $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters, respectively. We write $\mathrm{GEV}(\mu, \sigma, \xi)$ for short. Please note that the GEV parameters may differ for each marginal distribution and that the marginal distribution already determines the normalizing vectors $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$ in (2.1).

Definition 2.1.2. A max-stable random vector $\boldsymbol{X} = (X_1, \dots, X_d)^\top$ is called *simple max-stable* if it has standard unit Fréchet marginals, that is, $\mathbb{P}(X_i \leq x) = \exp(-1/x)$, $x > 0$, for all $i = 1, \dots, d$.

The terminology *simple max-stable* has been used, for instance, in Molchanov (2008) or Falk et al. (2004) in this way. Therefore, if $\boldsymbol{X}$ is simple max-stable, then

$$\max_{j=1,\dots,n}(\boldsymbol{X}_j) \stackrel{d}{=} n\boldsymbol{X} \tag{2.3}$$

holds, where $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are i.i.d. copies of $\boldsymbol{X}$. Note that the converse is not true, i.e. if (2.3) holds, it is only implied that $\boldsymbol{X}$ is max-stable and has marginal distributions either degenerate to a point mass at zero or a scaled unit Fréchet distribution of the form $\mathbb{P}(X_i \leq x) = \exp(-\sigma_i/x)$, $x > 0$, $\sigma_i > 0$, for all $i = 1, \dots, d$. If all marginal distributions are scaled unit Fréchet distributions (with possibly different σ_i), $\boldsymbol{X}$ has been termed *semi-simple max-stable* in Molchanov (2008).

Any max-stable random vector $\boldsymbol{X}$ with GEV margins $X_i \sim \mathrm{GEV}(\mu_i, \sigma_i, \xi_i)$ can be transformed into a simple max-stable random vector $\boldsymbol{X}^*$ and vice versa via the componentwise order-preserving transformations

$$X_i^* = T(X_i; \mu_i, \sigma_i, \xi_i) \quad \text{and} \quad X_i = T^{-1}(X_i^*; \mu_i, \sigma_i, \xi_i), \quad i = 1, \dots, d, \tag{2.4}$$

where the transformations

$$T(x;\mu,\sigma,\xi) = \Big(1+\xi\frac{x-\mu}{\sigma}\Big)^{1/\xi} \quad \text{and} \quad T^{-1}(x^*;\mu,\sigma,\xi) = \sigma\frac{(x^*)^\xi - 1}{\xi} + \mu,$$

(with the usual interpretation of $(1+\xi x)^{1/\xi}$ as e^x for $\xi = 0$) are order-preserving on the support of $\mathrm{GEV}(\mu,\sigma,\xi)$ and the standard unit Féchet law, respectively. In this sense a simple max-stable random vector can be interpreted as a copula of a general max-stable random vector with non-degenerate margins, which encapsulates its dependence structure. We note that the term copula is often reserved in the literature for the law of a random vector with uniform margins on the interval $[0,1]$. The extremes literature is dominated by a normalization to standard Fréchet margins instead.

There are different ways to describe the distribution of such simple max-stable random vectors. The following will be relevant for us. Note that such vectors take values in the open upper orthant $(0,\infty)^d$ almost surely. Here and hereinafter we shall denote the i-th indicator vector by $\boldsymbol{e}_i$ (all components of $\boldsymbol{e}_i$ are zero except for the i-th component, which takes the value one).

Theorem/Definition 2.1.3 (Representations of simple max-stable distributions, that is, max-stable distributions with standard unit Fréchet margins)**.** *A random vector* $\boldsymbol{X} = (X_1,\dots,X_d)^\top$ *with distribution function* $G(\boldsymbol{x}) = \exp(-V(\boldsymbol{x}))$, $\boldsymbol{x} \in (0,\infty]^d$, *is simple max-stable if and only if the* exponent function V *can be represented in one of the following equivalent ways:*

(i) ***Spectral representation*** *(de Haan, 1984). There exists a finite measure space* $(\Omega,\mathcal{A},\nu)$ *and a measurable function* $f : \Omega \to [0,\infty)^d$ *such that* $\int_\Omega f_i(\omega)\,\nu(\mathrm{d}\omega) = 1$ *for* $i = 1,\dots,d$, *and*

$$V(x_1,\dots,x_d) = \int_\Omega \max_{i=1,\dots,d} \frac{f_i(\omega)}{x_i}\,\nu(\mathrm{d}\omega).$$

(ii) ***Exponent measure*** *(Resnick, 1987). There exists a* (-1)*-homogeneous measure* Λ *on* $[0,\infty)^d \setminus \{\mathbf{0}\}$, *i.e.* $\Lambda(cA) = c^{-1}\Lambda(A)$ *for all* $c > 0$ *and* A *any Borel subset of* $[0,\infty)^d \setminus \{\mathbf{0}\}$, *such that*

$$\Lambda\Big(\big\{y \in [0,\infty)^d \,:\, y_i > 1\big\}\Big) = 1$$

for $i = 1,\dots,d$, *and*

$$V(x_1,x_2,\dots,x_d) = \Lambda\Big(\big\{y \in [0,\infty)^d \,:\, y_i > x_i \text{ for some } i \in \{1,\dots,d\}\big\}\Big).$$

(iii) ***Stable tail dependence function*** *(Ressel, 2013). There exists a* 1*-homogeneous and max-completely alternating function* $\ell : [0,\infty)^d \to [0,\infty)$*, such that* $\ell(\boldsymbol{e}_i) = 1$ *for* $i = 1,\dots,d$*, and*

$$V(x_1,\dots,x_d) = \ell\left(\frac{1}{x_1},\dots,\frac{1}{x_d}\right)$$

(cf. Appendix A for the notion of max-complete alternation).

In fact, the spectral representation can be seen as a polar decomposition of the exponent measure Λ, cf. e.g. Resnick (1987) or Beirlant et al. (2004). Importantly, it is not uniquely determined by the law of $\boldsymbol{X}$. For instance, if $(\Omega, \mathcal{A}, \nu, f)$ is a spectral representation for $\boldsymbol{X}$, and $c > 0$, then $(\Omega, \mathcal{A}, c\nu, c^{-1}f)$ is also a spectral representation for $\boldsymbol{X}$. Uniqueness can however be achieved by imposing further constraints on the tuple $(\Omega, \mathcal{A}, c\nu, c^{-1}f)$. Typical choices for the measure space $(\Omega, \mathcal{A}, \nu)$ are outlined in Theorem 6.1.14 of de Haan and Ferreira (2006) for dimension 2, and include (i) the unit interval with Lebesgue measure or (ii) a sphere $\Omega = \{\boldsymbol{\omega} \in [0,\infty)^d \ : \ \|\boldsymbol{\omega}\| = 1\}$ with respect to some norm $\|\cdot\|$, for instance the ℓ_p-norm

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1,\dots,d} |x_i|^p\right)^{1/p}$$

for some $p \geq 1$. For (i) it is then the spectral map f which highlights the contribution of each direction separately. For (ii) one usually considers the component maps $f_i(\boldsymbol{\omega}) = \omega_i$, so that all the dependence information is then given by the measure ν, then often termed ***angular measure***. In general, the angular measure will not be a probability measure. However, it can be rescaled to a probability measure whilst the scaling constant is multiplied to the component maps. More precisely, for a given spectral representation $(\Omega, \mathcal{A}, \nu, f)$ one may rescale ν to a probability measure and absorb the rescaling constant into the spectral map f. The resulting random vector $\boldsymbol{Z} = (Z_1,\dots,Z_d)^\top$ such that $\mathbb{E}(Z_i) = 1$, $i = 1,\dots,d$, and

$$V(x_1,\dots,x_d) = \mathbb{E}\max_{i=1,\dots,d}\left(\frac{Z_i}{x_i}\right),$$

has been termed ***generator*** of the simple max-stable random vector $\boldsymbol{X}$, cf. Falk (2019). In general, we have

$$\max_{i=1,\dots,d}\left(\frac{1}{x_i}\right) \leq V(x_1,\dots,x_d) \leq \frac{1}{x_1} + \cdots + \frac{1}{x_d},$$

where the boundary cases represent full dependence (lower bound) and full independence (upper bound), respectively.

A useful observation is the following; for a given vector $\boldsymbol{x}$ with values in $\mathbb{R}^d$ and a subset $A \subset \{1,\dots,d\}$, let $\boldsymbol{x}_A$ be the subvector with components in A.

Lemma 2.1.4. *Let $\boldsymbol{Z}$ be a generator for the max-stable law $\boldsymbol{X}$, then $\boldsymbol{Z}_A$ is a generator for $\boldsymbol{X}_A$.*

Proof. Let $\boldsymbol{X}$ be a d-variate simple max-stable random vector with generator $\boldsymbol{Z}$ and stable tail dependence function ℓ. Then

$$\ell(x_1, \dots, x_d) = \mathbb{E} \max_{i=1,\dots,d} \big(x_i Z_i\big), \quad \boldsymbol{x} \in [0, \infty)^d.$$

It follows that for a non-empty subset $A \subset \{1, \dots, d\}$

$$\ell_A(\boldsymbol{x}_A) = \mathbb{E} \max_{i \in A} \big(x_i Z_i\big), \quad \boldsymbol{x}_A \in [0, \infty)^A$$

is the stable tail dependence function of $\boldsymbol{X}_A$, cf. e.g. the introduction of Ressel (2022) or Section 7 (Projection) of Molchanov (2008). Hence $\boldsymbol{Z}_A$ is a generator for $\boldsymbol{X}_A$. □

An important fact about the exponent measure which will be relevant in Chapter 3 is the following: while the support of the exponent measure Λ is contained in $[0, \infty)^d \setminus \{\mathbf{0}\}$, its total mass is infinite.

The stable tail dependence function ℓ goes back to Huang (1992) and has also been called *D-norm* (Falk et al., 2004) of $\boldsymbol{X}$. The boundary cases for the stable tail dependence function ℓ are given by

$$\max_{i=1,\dots,d} x_i \le \ell(x_1, \dots, x_d) \le x_1 + \cdots + x_d, \qquad \text{for } \boldsymbol{x} \in [0, \infty)^d.$$

Since ℓ is 1-homogeneous, it suffices to know its values on the unit simplex $\triangle_d = \{\boldsymbol{x} \in [0, \infty)^d : \|\boldsymbol{x}\|_1 = 1\}$; note that $\triangle_d$ is a special case of the sphere Ω defined above, where the norm chosen is the ℓ_p-norm with $p = 1$; the restriction of ℓ to $\triangle_d$ is called ***Pickands dependence function***

$$A(x_1, \dots, x_d) = \ell(x_1, \dots, x_d), \qquad (x_1, \dots, x_d)^\top \in \triangle_d.$$

There exist further descriptors of the dependence structure, e.g. in terms of Point processes or LePage representation, cf. e.g. Resnick (1987) or, in a very general context, Davydov et al. (2008). Copulas of max-stable random vectors on standard uniform margins are called *extreme value copulas* (Gudendorf and Segers, 2010).

Let us close with a representation that allows for some interesting geometric interpretations and has been instrumental in the derivation of Theorem 2.2.9, which we will discuss below. Molchanov (2008) introduced a convex body $K \subset [0, \infty)^d$, which can be interpreted (up to rescaling) as selection expectation of a random cross polytope associated with the (normalised) spectral measure ν.

Molchanov (2008) shows that the stable tail dependence function is in fact the support function of the convex body K, that is,

$$\ell(\boldsymbol{x}) = \sup\{\langle \boldsymbol{x}, \boldsymbol{k}\rangle \ : \ \boldsymbol{k} \in K\}, \tag{2.5}$$

where sup denotes the supremum. The convex body K is called ***max-zonoid*** (or ***dependency set***) of $\boldsymbol{X}$ and it is uniquely determined by the law of $\boldsymbol{X}$. In fact

$$K = \big\{\boldsymbol{k} \in [0,\infty)^d \ : \ \langle \boldsymbol{k}, \boldsymbol{x}\rangle \leq \ell(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in [0,\infty)^d\big\}. \tag{2.6}$$

In general, it is difficult to translate one representation from Theorem 2.1.3 into another apart from the obvious relations

$$\begin{aligned}\ell(\boldsymbol{x}) &= \mathbb{E} \max_{i=1,\dots,d}(x_i Z_i) \\ &= \int_\Omega \max_{i=1,\dots,d} x_i f_i(\omega)\ \nu(\mathrm{d}\omega) = \Lambda\Big(\big\{y \in [0,\infty)^d \ : \ \max_{i=1,\dots,d}(x_i y_i) > 1\big\}\Big)\end{aligned}$$

for $\boldsymbol{x} \geq \boldsymbol{0}$. For convenience, we have added material in Appendix B how to obtain the boundary of a max-zonoid K from the stable tail dependence function ℓ in the bivariate case, which will help to illustrate some of the results below.

2.2 Parametric models

Several parametric models for max-stable random vectors have been summarised for instance in Beirlant et al. (2004). In what follows we draw our attention to two of the most popular parametric models, the Dirichlet and Hüsler-Reiß families, as well as the Choquet model (Tawn-Molchanov model), which will reveal some interesting phenomena and (counter-)examples of stochastic ordering relations. For the Dirichlet and Hüsler-Reiß families, we will show they are PQD-ordered according to the natural order within their parameter spaces. The Choquet model will be useful to construct some counterexamples which show that upper orthant and lower orthant ordering among simple max-stable distributions are not equivalent.

2.2.1 Dirichlet model

Coles and Tawn (1991) compute densities of angular measures of simple max-stable random vectors constructed from non-negative functions on the unit simplex $\triangle_d$. In particular, the following *asymmetric Dirichlet model* has been introduced. We summarise some equivalent characterisations, each of which may serve as a definition of the asymmetric Dirichlet model. This model has gained popularity due to its flexibility and simple structure forming the basis of Dirichlet mixture models (Boldi and Davison, 2007; Sabourin and Naveau, 2014).

Theorem/Definition 2.2.1 (Multivariate max-stable Dirichlet distribution)**.** *A random vector* $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ *is simple* max-stable Dirichlet *distributed with parameter vector* $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top \in (0, \infty)^d$*, we write*

$$\boldsymbol{X} = (X_1, \ldots, X_d)^\top \sim \mathrm{MaxDir}(\alpha_1, \ldots, \alpha_d) = \mathrm{MaxDir}(\boldsymbol{\alpha})$$

for short, if and only if one of the following equivalent conditions is satisfied:

(i) ***(Gamma generator)*** *A generator of* $\boldsymbol{X}$ *is the random vector*

$$\boldsymbol{\alpha}^{-1}\boldsymbol{\Gamma} = (\Gamma_1/\alpha_1, \Gamma_2/\alpha_2, \ldots, \Gamma_d/\alpha_d)^\top,$$

where $\boldsymbol{\Gamma} = (\Gamma_1, \ldots, \Gamma_d)^\top$ *consists of independent Gamma distributed variables* $\Gamma_i \sim \Gamma(\alpha_i)$*,* $\alpha_i > 0$*,* $i = 1, \ldots, d$*. Here, the Gamma distribution* $\Gamma(\alpha_i)$ *has the density*

$$\gamma_{\alpha_i}(x) = \frac{x^{\alpha_i - 1}}{\Gamma(\alpha_i)} \exp\big(-x\big).$$

(ii) ***(Dirichlet generator)*** *A generator of* $\boldsymbol{X}$ *is the random vector*

$$(\boldsymbol{\alpha}^{-1}\|\boldsymbol{\alpha}\|_1)\boldsymbol{D} = (\alpha_1 + \cdots + \alpha_d) \cdot (D_1/\alpha_1, D_2/\alpha_2, \ldots, D_d/\alpha_d)^\top,$$

where $\boldsymbol{D}$ *follows a Dirichlet distribution* $\mathrm{Dir}(\alpha_1, \ldots, \alpha_d)$ *on the unit simplex* $\triangle_d$ *with density*

$$d(\omega_1, \ldots, \omega_d) = \Gamma(\|\boldsymbol{\alpha}\|_1) \prod_{i=1}^{d} \frac{\omega_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}, \quad (\omega_1, \ldots, \omega_d)^\top \in \triangle_d.$$

(iii) ***(Angular measure)*** *The density of the angular measure of* $\boldsymbol{X}$ *on* $\triangle_d$ *is given by*

$$h(\omega_1, \ldots, \omega_d) = \frac{\Gamma(\|\boldsymbol{\alpha}\|_1 + 1)}{\|\boldsymbol{\alpha}\boldsymbol{\omega}\|_1} \prod_{i=1}^{d} \frac{\alpha_i^{\alpha_i} \omega_i^{\alpha_i - 1}}{\Gamma(\alpha_i)(\|\boldsymbol{\alpha}\boldsymbol{\omega}\|_1)^{\alpha_i}}, \quad (\omega_1, \ldots, \omega_d)^\top \in \triangle_d. \tag{2.7}$$

Proof. The equivalence of (ii) and (iii) has been proved in Coles and Tawn (1991) (page 382). The equivalence of (i) and (ii) follows similarly to Aulbach et al. (2015) (3) from the fact that $\boldsymbol{D}$ is distributed like $\boldsymbol{\Gamma}/\|\boldsymbol{\Gamma}\|_1$ and the independence of $\boldsymbol{\Gamma}/\|\boldsymbol{\Gamma}\|_1$ and $\|\boldsymbol{\Gamma}\|_1$. More precisely, let ℓ_1 and ℓ_2 be the stable tail dependence functions that arise from the generators (i) and (ii), respectively. Then ℓ_1 and ℓ_2 can be expressed as follows for any $\boldsymbol{x} \in [0, \infty)^d$

$$\ell_1(\boldsymbol{x}) = \mathbb{E} \max_{i=1,\ldots,d} \frac{x_i \Gamma_i}{\alpha_i} = \mathbb{E}\|\boldsymbol{\Gamma}\|_1 \cdot \mathbb{E} \max_{i=1,\ldots,d} \frac{x_i \Gamma_i/\|\boldsymbol{\Gamma}\|_1}{\alpha_i},$$
$$\ell_2(\boldsymbol{x}) = \mathbb{E} \max_{i=1,\ldots,d} \frac{x_i \|\boldsymbol{\alpha}\|_1 D_i}{\alpha_i} = \|\boldsymbol{\alpha}\|_1 \cdot \mathbb{E} \max_{i=1,\ldots,d} \frac{x_i D_i}{\alpha_i}.$$

If suffices to note $\mathbb{E}\|\boldsymbol{\Gamma}\|_1 = \|\boldsymbol{\alpha}\|_1$ in order to conclude $\ell_1 = \ell_2$. □

Remark 2.2.2. To the best of our knowledge the representation through the Gamma generator, albeit inspired by Aulbach et al. (2015) from the fully symmetric case, where $\alpha_1 = \alpha_2 = \cdots = \alpha_d$, is new in this generality, where all parameters (or only some of them) α_i, $i = 1, \ldots, d$, may be mutually distinct. An advantage of the representation with the Gamma generator is that it reveals immediately the closure of the model with respect to taking marginal distributions, cf. Lemma 2.1.4, a result that has been previously obtained in Ballani and Schlather (2011), but with a one-page proof and some intricate density calculations.

Lemma 2.2.3 (Closure of Dirichlet model under taking marginals)**.** *Let* $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \sim \mathrm{MaxDir}(\alpha_1, \ldots, \alpha_d) = \mathrm{MaxDir}(\boldsymbol{\alpha})$ *and* $A \subset \{1, \ldots, d\}$*, then* $\boldsymbol{X}_A \sim \mathrm{MaxDir}(\boldsymbol{\alpha}_A)$.

Proof. Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \sim \mathrm{MaxDir}(\alpha_1, \ldots, \alpha_d)$. Then a generator for $\boldsymbol{X}$ is the random vector

$$\boldsymbol{\alpha}^{-1}\boldsymbol{\Gamma} = (\Gamma_1/\alpha_1, \Gamma_2/\alpha_2, \ldots, \Gamma_d/\alpha_d)^\top,$$

where $\boldsymbol{\Gamma} = (\Gamma_1, \ldots, \Gamma_d)^\top$ consists of independent Gamma distributed variables $\Gamma_i \sim \Gamma(\alpha_i)$, $\alpha_i > 0$, $i = 1, \ldots, d$. By Lemma 2.1.4 the random vector $\boldsymbol{\alpha}_A^{-1}\boldsymbol{\Gamma}_A$ is a generator for $\boldsymbol{X}_A$. Hence, $\boldsymbol{X}_A \sim \mathrm{MaxDir}(\boldsymbol{\alpha}_A)$. □

The angular density representation (2.7) on the other hand is useful to see that different parameter vectors $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$ lead in fact to different multivariate distributions $\mathrm{MaxDir}(\boldsymbol{\alpha}) \neq \mathrm{MaxDir}(\boldsymbol{\beta})$ for $d \geq 2$, so that $(0, \infty)^d$ is indeed the natural parameter space for this model.

2.2.2 Hüsler-Reiß model

The multivariate *Hüsler-Reiß distribution* (Hüsler and Reiß, 1989) forms the basis of the popular Brown-Resnick process (Kabluchko et al., 2009) and has sparked significant interest from the perspectives of spatial modelling (Davison et al., 2019) and more recently in connection with graphical modelling of extremes (Engelke and Hitz, 2020). The natural parameter space for this model is the convex cone of conditionally negative symmetric $d \times d$-matrices, whose diagonal entries are zero

$$\mathcal{G}_d = \left\{ \boldsymbol{\gamma} = (\gamma_{ij})_{i,j \in \{1,\ldots,d\}} \in \mathbb{R}^{d \times d} \; : \; \begin{array}{c} \gamma_{ij} = \gamma_{ji},\ \gamma_{ii} = 0 \text{ for all } i, j \in \{1, \ldots, d\}, \\ v^\top \boldsymbol{\gamma} v \leq 0 \text{ for all } v \in \mathbb{R}^d \\ \text{such that } v_1 + \cdots + v_d = 0 \end{array} \right\}.$$

It is well-known, cf. e.g. Berg et al. (1984, Ch. 3), that for a given $\boldsymbol{\gamma} \in \mathcal{G}_d$, there exists a zero mean Gaussian random vector $\boldsymbol{W} = (W_1, \ldots, W_d)^\top$ with incremental variance

$$\mathbb{E}(W_i - W_j)^2 = \gamma_{ij}, \qquad i, j \in \{1, \ldots, d\}, \tag{2.8}$$

although its distribution is not uniquely specified by this condition. For instance, select $i \in \{1, \ldots, d\}$. Imposing additionally the linear constraint "$W_i = 0$ almost surely" leads to $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ with

$$(\boldsymbol{\Sigma}_i)_{jk} = \frac{1}{2}\big(\gamma_{ij} + \gamma_{ik} - \gamma_{jk}\big), \qquad j, k \in \{1, \ldots, d\},$$

which satisfies (2.8). By writing $\boldsymbol{W} = (W_1, \ldots, W_d)^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ we mean that $\boldsymbol{W}$ is Gaussian with zero mean, i.e. $\mathbb{E}(W_j) = 0$ for all $j = 1, \ldots, d$, and with covariance matrix $\boldsymbol{\Sigma}_i$, i.e. $\mathrm{Cov}(W_j, W_k) = \mathbb{E}(W_j W_k) = (\boldsymbol{\Sigma}_i)_{jk}$ for $j, k \in \{1, \ldots, d\}$.

Theorem/Definition 2.2.4 (Multivariate Hüsler-Reiss model, cf. Kabluchko (2011) Theorem 1)**.** *Let $\boldsymbol{\gamma} \in \mathcal{G}_d$ and let $\boldsymbol{W}$ be a zero mean Gaussian random vector which satisfies (2.8). Consider the simple max-stable random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ defined by the generator $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^\top$ with*

$$Z_i = \exp\left(W_i - \frac{1}{2}\mathrm{Var}(W_i)\right), \qquad i = 1, \ldots, d.$$

Then the distribution of $\boldsymbol{X}$ depends only on $\boldsymbol{\gamma}$ and not on the specific choice of a zero mean Gaussian distribution satisfying (2.8). We call $\boldsymbol{X}$ simple Hüsler-Reiß *distributed with parameter matrix $\boldsymbol{\gamma}$ and write for short*

$$\boldsymbol{X} = (X_1, \ldots, X_d)^\top \sim \mathrm{HR}(\boldsymbol{\gamma}).$$

We also note that for $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{G}_d$, the distributions $\mathrm{HR}(\boldsymbol{\gamma}_1)$ and $\mathrm{HR}(\boldsymbol{\gamma}_2)$ coincide if and only if $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$, so that $\mathcal{G}_d$ is indeed the natural parameter space for these models. This follows directly from the observation that the multivariate Hüsler-Reiß model is also closed under taking marginal distributions and the equivalent statement for bivariate Hüsler-Reiß models, which can be seen for instance from (B.3) below. Indeed, we also state the following lemma for clarity. It follows directly from the generator representation of $\mathrm{HR}(\boldsymbol{\gamma})$ and Lemma 2.1.4.

Lemma 2.2.5 (Closure of Hüsler-Reiß model under taking marginals)**.** *Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \sim \mathrm{HR}(\boldsymbol{\gamma})$ and $A \subset \{1, \ldots, d\}$, then $\boldsymbol{X}_A \sim \mathrm{HR}(\boldsymbol{\gamma}_{A\times A})$, where $\boldsymbol{\gamma}_{A\times A}$ is the restriction of $\boldsymbol{\gamma}$ to the components of A in both rows and columns.*

It is well-known that up to a change of location and scale parameters Hüsler-Reiß distributions are the only possible limit laws of maxima of triangular arrays of multivariate Gaussian distributions, a finding which can be traced back to Hüsler and Reiß (1989) and Brown and Resnick (1977). The following version will be convenient for us.

Theorem 2.2.6 (Triangular array convergence of maxima of Gaussian vectors, cf. Kabluchko (2011) Theorem 2)**.** *Let u_n be a sequence such that $\sqrt{2\pi}u_n e^{u_n^2/2}/n \to 1$ as $n \to \infty$. For each $n \in \mathbb{N}$ let $\boldsymbol{Y}_1^{(n)}, \boldsymbol{Y}_2^{(n)}, \dots, \boldsymbol{Y}_n^{(n)}$ be independent copies of a d-variate zero mean unit-variance Gaussian random vector with correlation matrix $(\rho_{ij}^{(n)})_{i,j\in\{1,\dots,d\}}$. Suppose that for all $i, j \in \{1, \dots, d\}$*

$$4\log(n)(1-\rho_{ij}^{(n)}) \to \gamma_{ij} \in [0,\infty)$$

as $n \to \infty$. Then the matrix $\boldsymbol{\gamma} = (\gamma_{ij})_{i,j\in\{1,\dots,d\}}$ is necessarily and element of $\mathcal{G}_d$. Let $\boldsymbol{M}^{(n)}$ be the componentwise maximum of $\boldsymbol{Y}_1^{(n)}, \boldsymbol{Y}_2^{(n)}, \dots, \boldsymbol{Y}_n^{(n)}$. Then the componentwise rescaled vector $u_n(\boldsymbol{M}^{(n)} - u_n)$ converges in distribution to the Hüsler-Reiß distribution $\mathrm{HR}(\boldsymbol{\gamma})$.

Remark 2.2.7. In the bivariate case we have $\gamma_{12} = \gamma_{21} = \gamma \in [0,\infty)$ and the boundary case $\gamma = 0$ leads to a degenerate random vector with fully dependent components, whereas $\gamma \uparrow \infty$ leads to a random vector with independent components. More generally, one might also admit the value ∞ for γ_{ij} in the multivariate case, as long as the resulting matrix $\boldsymbol{\gamma}$ is *negative definite in the extended sense*, cf. Kabluchko (2011). This extension corresponds to a partition of $\boldsymbol{X}$ into independent subvectors $\boldsymbol{X} = \bigsqcup_A \boldsymbol{X}_A$, where each $\boldsymbol{X}_A$ is a Hüsler-Reiß random vector in the usual sense. Here $\gamma_{ij} = \infty$ precisely when i and j are in different subsets of the partition. Theorem 2.2.6 extends to this situation as well. In fact, is has been formulated in this generality in Kabluchko (2011).

Remark 2.2.8. In Theorem 2.2.6 it is important to consider the componentwise maxima of the rows of a triangular array, where in the n-th row, the dependence structure of the independent Gaussian random vectors $\boldsymbol{Y}_1^{(n)}, \boldsymbol{Y}_2^{(n)}, \dots, \boldsymbol{Y}_n^{(n)}$ is allowed to change with n. Simply taking componentwise maxima of an i.i.d. sequence of Gaussian random vectors with each correlation $\rho_{ij} < 1$, would lead instead to a distributional limit with independent components (Sibuya, 1960), which is why the Gaussian random vector is also often referred to as an asymptotic independent model. Therefore, we need to follow the route of triangular arrays as in Hüsler and Reiß (1989) or Kabluchko (2011), for instance, where the correlation is allowed to change with the row index n.

2.2.3 Choquet model / Tawn-Molchanov model

A popular way to summarise extremal dependence information within a random vector is by considering its *extremal coefficients*, which in the case of a simple max-stable random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$ may be expressed as

$$\theta(A) = \ell(\boldsymbol{e}_A), \qquad \boldsymbol{e}_A = \sum_{i \in A} \boldsymbol{e}_i, \qquad A \subset \{1, \ldots, d\},\ A \neq \emptyset,$$

or, equivalently,

$$\theta(A) = \mathbb{E} \max_{i \in A} Z_i = \int_\Omega \max_{i \in A} f_i(\omega)\ \nu(\mathrm{d}\omega) = \Lambda\Big(\big\{y \in [0,\infty)^d\ :\ \max_{i \in A}(y_i) > 1\big\}\Big), \tag{2.9}$$

where ℓ is the stable tail dependence function, $\boldsymbol{Z}$ a generator, Λ the exponent measure and $(\Omega, \mathcal{A}, \nu, f)$ a spectral representation for $\boldsymbol{X}$. The coefficient $\theta(A)$ takes values in $[1, |A|]$. Loosely speaking, it can be interpreted as the effective number of independent variables among the collection $(X_i)_{i \in A}$. We have $\theta(\{i\}) = 1$ for singletons $\{i\}$ and naturally $\theta(\emptyset) = 0$.

The following result can be traced back to Schlather and Tawn (2002) and Molchanov (2008). Accordingly, the associated max-stable model, which can be parametrised by its extremal coefficients, has been introduced as *Tawn-Molchanov model* in Strokorb and Schlather (2015). It is essentially an application of the the Choquet theorem (see Molchanov (2017) Section 1.2 and Berg et al. (1984) Theorem 6.6.19), which also holds for not necessarily finite capacities (see Schneider and Weil (2008) Theorem 2.3.2). Therefore, it has been relabelled *Choquet model* in Molchanov and Strokorb (2016), cf. Appendix A for background on complete alternation. We write $\mathcal{P}_d$ for the power set of $\{1, \ldots, d\}$ henceforth.

Theorem 2.2.9. *a) [Strokorb and Schlather (2015) Theorem 8 (a)] Let $\theta : \mathcal{P}_d \to \mathbb{R}$. Then θ is the extremal coefficient function of a simple max-stable random vector in $(0,\infty)^d$ if and only if $\theta(\emptyset) = 0$, $\theta(\{i\}) = 1$ for all $i = 1, \ldots, d$ and θ is union-completely alternating.*

b) [Strokorb and Schlather (2015) Theorem 8 (b) and Molchanov and Strokorb (2016) Theorem 3.7 and Prop. 5.3] Let $\theta : \mathcal{P}_d \to \mathbb{R}$ be an extremal coefficient function. Let

$$\ell^*(\boldsymbol{x}) = \int_0^\infty \theta(\{i\ :\ x_i \geq t\})\, dt, \quad \boldsymbol{x} \in [0,\infty)^d \tag{2.10}$$

be the Choquet integral with respect to θ. Then ℓ^* is a valid stable tail dependence function, which retrieves the given extremal coefficients $\ell^*(\boldsymbol{e}_A) = \theta(A)$ for all $A \in \mathcal{P}_d$. Its max-zonoid is given by

$$K^* = \big\{\boldsymbol{k} \in [0,\infty)^d\ :\ \langle \boldsymbol{k}, \boldsymbol{e}_A \rangle \leq \theta(A) \text{ for all } A \in \mathcal{P}_d\big\}.$$

c) *[Strokorb and Schlather (2015) Theorem 32 and Corollary 33, Molchanov and Strokorb (2016) Corollary 5.4] Let ℓ be any stable tail dependence function with extremal coefficient function θ, that is, $\ell(\boldsymbol{e}_A) = \theta(A)$ for all $A \in \mathcal{P}_d$, and let K be its corresponding max-zonoid. Then*

$$\ell(\boldsymbol{x}) \leq \ell^*(\boldsymbol{x}), \quad \boldsymbol{x} \geq 0 \qquad \text{and} \qquad K \subset K^*,$$

where ℓ^* is as in (2.10).

Example 2.2.10 (Choquet model in the bivariate case)**.** Let ℓ be any bivariate stable tail dependence function with bivariate extremal coefficient $\theta = \ell(1,1) \in [1,2]$. Then the associated Choquet model is given by the max-zonoid $K^* = \{(x_1, x_2) \in [0,1]^2 \,:\, x_1 + x_2 \leq \theta\}$ or the stable tail dependence function $\ell^*(x_1, x_2) = \max(x_1 + (\theta-1)x_2, (\theta-1)x_1 + x_2)$. Figure 2.1 displays a situation, where the original ℓ stems from an asymmetric Dirichlet model, see (B.4) and (B.5) in Appendix B for the expression of ℓ for this model. Figure 2.1 visually shows that Theorem 2.2.9, part (c) is indeed satisfied, i.e. that $\ell(\boldsymbol{x}) \leq \ell^*(\boldsymbol{x})$ for $\boldsymbol{x} \geq 0$ and $K \subset K^*$.

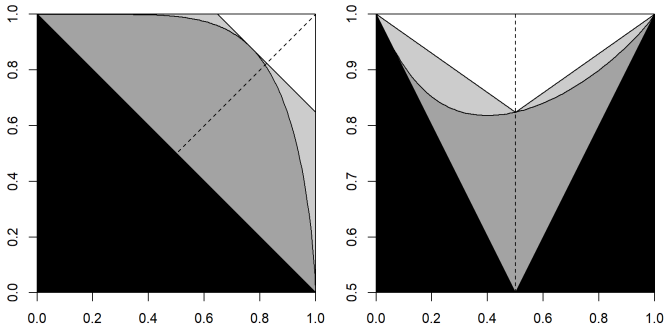

Figure 2.1: Nested max-zonoids and Pickands dependence functions ranging from full dependence (black), an asymmetric Dirichlet model with $\boldsymbol{\alpha} = (30, 0.2)$ (dark grey), its associated Choquet (Tawn-Molchanov) model (light grey) to the fully independent model (white).

In geometric terms, for any given max-zonoid $K \subset [0,1]^d$ the associated Choquet max-zonoid $K^* \subset [0,1]^d$ is bounded by $2^d - 1$ hyperplanes, one for each direction $\boldsymbol{e}_A$, which is the supporting hyperplane of the max-zonoid K in the direction of $\boldsymbol{e}_A$.

The Choquet model is a spectrally discrete max-stable model, whose exponent measure has its support contained in the rays through the vectors $\boldsymbol{e}_A$, $A \subset \{1, \ldots, d\}$, $A \neq \emptyset$. While its natural parameter space is the set of extremal coefficients, we can also describe it via the mass that the model puts on those rays. To this end, let $\tau : \mathcal{P}_d \setminus \{\emptyset\} \to \mathbb{R}$ be given as follows

$$\tau(A) = \sum_{I \subset A} (-1)^{|I|+1} \theta(I \cup (\{1, \ldots, d\} \setminus A)),$$

where we assume $a_1, a_2, \ldots, a_n$ to be the distinct elements from $A \subset \{1, \ldots, d\}$. Then the spectral representation $(\Omega, \mathcal{A}, \nu^*, f)$ with

$$\Omega = \{\boldsymbol{\omega} \in [0,\infty)^2 \,:\, \|\boldsymbol{\omega}\|_\infty = 1\}, \quad f_i(\boldsymbol{\omega}) = \omega_i, \quad \nu^* = \sum_{A \in \mathcal{P}_d \setminus \{\emptyset\}} \tau(A)\, \delta_{\boldsymbol{e}_A} \tag{2.11}$$

corresponds to the stable tail dependence function ℓ^* from Theorem 2.2.9. In terms of an underlying generator for which (2.9) holds true, we may express τ as

$$\tau(A) = \mathbb{E}\Big(\min_{i \in A} Z_i - \max_{i \in \{1,\ldots,d\} \setminus A} Z_i \Big)_+,$$

cf. Papastathopoulos and Strokorb (2016) Lemma 3. Moreover, we recover θ from τ via

$$\theta(A) = \sum_{K \,:\, K \cap A \neq \emptyset} \tau(K),$$

which makes the analogy between extremal coefficient functions θ and capacity functionals of random sets even more explicit.

However, there are two drawbacks with representing the Choquet model by the collection of coefficients $\tau(A)$, $A \subset \{1, \ldots, d\}$, $A \neq \emptyset$. First, this representation is specific to the dimension, in which the model is considered, that is, we cannot simply turn to a subset of these coefficients when considering marginal distributions. Second, one may easily forget that one has in fact not $2^d - 1$ degrees of freedom among these coefficients, but $2^d - 1 - d$, since $\theta(\{i\}) = 1$ for singletons $\{i\}$, which is only encoded through d linear constraints for τ as follows

$$\sum_{K \,:\, i \in K} \tau(K) = 1, \qquad i = 1, \ldots, d. \tag{2.12}$$

A third parametrisation of the Choquet model, which has received little attention so far, but is very relevant for the ordering results in this article (cf. Lemma 3.2.14) and does not have such drawbacks, is the following. Instead of extremal coefficients, let us consider the following *tail dependence coefficients* for $A \subset \{1, \ldots, d\}$, $A \neq \emptyset$:

$$\chi(A) = \mathbb{E} \min_{i \in A} Z_i = \int_\Omega \min_{i \in A} f_i(\omega)\, \nu(\mathrm{d}\omega) = \Lambda\Big(\big\{ y \in [0,\infty)^d \,:\, \min_{i \in A}(y_i) > 1 \big\} \Big).$$

Then it is easily seen that

$$\chi(A) = \sum_{I \subset A,\, I \neq \emptyset} (-1)^{|I|+1} \theta(I) \qquad \text{and} \qquad \theta(A) = \sum_{I \subset A,\, I \neq \emptyset} (-1)^{|I|+1} \chi(I). \tag{2.13}$$

In particular $\chi(\{i\}) = \theta(\{i\}) = 1$ for $i = 1, \ldots, d$, and these operations show explicitly, how θ and χ can be recovered from each other. While θ resembles a capacity functional, χ can be seen as an analog of an inclusion functional, since

$$\chi(A) = \sum_{K \,:\, A \subset K} \tau(K), \tag{2.14}$$

whereas

$$\tau(A) = \sum_{K\,:\,A\subset K} (-1)^{|K\setminus A|}\chi(K).$$

To sum up, we may consider three different parametrizations for the Choquet model:

(i) by the $2^d - 1$ extremal coefficients $\theta(A)$, $A \in \mathcal{P}_d$, $A \neq \emptyset$,

(ii) by the $2^d - 1$ tail dependence coefficients $\chi(A)$, $A \in \mathcal{P}_d$, $A \neq \emptyset$,

(iii) by the $2^d - 1$ mass coefficients $\tau(A)$, $A \in \mathcal{P}_d$, $A \neq \emptyset$.

For (i) and (ii) the constraint for standard unit Fréchet margins is encoded via $\chi(\{i\}) = \theta(\{i\}) = 1$ for $i = 1, \ldots, d$. For (iii) it amounts to the d conditions from (2.12). Only (i) and (ii) do not depend on the dimension, in which the model is considered.

2.3 Max-domain of attraction and regular variation

Here, we recall some of the most fundamental connections between (multivariate) extreme value theory and regular variation. Standard textbook treatments include for instance de Haan and Ferreira (2006) Appendix B or Embrechts et al. (1997) Chapter 3, Beirlant et al. (2004) Chapter 2 for the univariate theory, or Resnick (1987) Section 5, Resnick (2007) Chapter 6 or Kulik and Soulier (2020) Chapter 2 for the multivatiate theory.

Definition 2.3.1. Let X be a random variable with distribution function F, let $X_1, X_2, \ldots$ be independent copies of X, and $b_n \in \mathbb{R}$ and $a_n > 0$ such that the law of

$$\frac{\max_{i=1,\ldots,n}(X_i) - b_n}{a_n} \tag{2.15}$$

converges in distribution to some non-degenerate limiting distribution G, then G is called a *(univariate) extreme value distribution* and X (or its distribution F) is said to be in the *max-domain of attraction* of G.

Remark 2.3.2. The distribution function of the law of (2.15) is given by $x \mapsto F^n(a_n x + b_n)$. That the law of (2.15) converges to a limit G is equivalent to saying that $\lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$ for all continuity points x of G.

Each (univariate) extreme value distribution is necessarily max-stable, and by the Fisher-Tippett theorem (in Jenkinson-van-Mises form), it takes the form of a GEV distribution, cf. (2.2). The extreme value index ξ therein is linked to the tail behaviour of the distribution. The only possible limits are the following three types:

- For $\xi > 0$, we obtain the (heavy-tailed) Fréchet distribution

$$\Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \end{cases}$$

  where $\alpha = 1/\xi > 0$.

- For $\xi = 0$, we obtain the Gumbel distribution $\Lambda(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$.

- For $\xi < 0$, we obtain the (short-tailed) Weibull distribution

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha & x \leq 0 \\ 1 & x > 0, \end{cases}$$

  where $\alpha = -1/\xi > 0$.

Alternatively, we may state the extremal types theorem as follows.

Theorem 2.3.3 (Extremal types theorem (Fisher-Tippett-Gnedenko-de Haan))**.** *Let F be a univariate distribution function and $a_n > 0$ and $b_n \in \mathbb{R}$ be sequences such that $\lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$ for a distribution function G and for all continuity points x of G, then there exists $a > 0$ and $b \in \mathbb{R}$, such that $G(ax + b)$ is either a Fréchet distribution Φ_α for some $\alpha > 0$, a Gumbel distribution Λ, or a Weibull distribution Φ_α for some $\alpha > 0$.*

We shall focus on the heavy-tailed case in what follows, for which the domain-of attraction condition links particularly well with the theory of regular variation.

Definition 2.3.4. A function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is *regularly varying (at infinity)* with *index* $\alpha \in \mathbb{R}$, denoted by $f \in \mathrm{RV}_\alpha$ if for all $x > 0$,

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^\alpha.$$

In case $\alpha = 0$ in this condition, the function f is said to be *slowly varying*.

It is easily seen that any function $f \in \mathrm{RV}_\alpha$ can be written as $f(x) = x^\alpha \ell(x)$, where ℓ is a slowly varying function.

Theorem 2.3.5. *A distribution function F belongs to the maximum domain of attraction of the Fréchet distribution Φ_α if and only if its tail function $\overline{F} = 1 - F$ satisfies $\overline{F} \in \mathrm{RV}_{-\alpha}$.*

If in Theorem 2.3.5, X is distributed according to F, and we set

$$\begin{aligned} a(u) = F^{-1}\left(1 - \frac{1}{u}\right) &= \inf\left\{t > 0 : \mathbb{P}(X \leq t) \geq 1 - \frac{1}{u}\right\} \\ &= \inf\left\{t > 0 : \mathbb{P}(X > t) \leq \frac{1}{u}\right\} \\ &= \inf\left\{t > 0 : \overline{F}(t) \leq \frac{1}{u}\right\}, \end{aligned}$$

the regular variation condition $\overline{F} \in \mathrm{RV}_{-\alpha}$ may also be reformulated as

$$\begin{aligned} \lim_{u\to\infty} u\mathbb{P}(X > a(u)x) &= \lim_{u\to\infty} \frac{\mathbb{P}(X > a(u)x)}{1/u} \\ &= \lim_{t\to\infty} \frac{\mathbb{P}(X > tx)}{\mathbb{P}(X > t)} = \lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha}, \end{aligned}$$

which is our basis for generalizing these univariate notions to the multivariate situation. Note that in the second equality we use that $t = a(u) \to \infty$ as $u \to \infty$, which holds since F in the Fréchet domain of attraction has upper endpoint ∞.

Multivariate regular variation There are multiple equivalent definitions of multivariate regular variation. The following will be most useful for us in Chapter 5 and takes into account already the polar decomposition of the limiting measure in the vague convergence stated below this definition. Although, in principle, we might have chosen an arbitrary norm $\|\cdot\|$ for such a definition, it will be more natural for us to work with the Euclidean norm abbreviated as $\|\cdot\| = \|\cdot\|_2$ here and in Chapter 5.

Definition 2.3.6. A random vector $\boldsymbol{X}$ with values in $\mathbb{R}^d$ is *(multivariate) regularly varying* if there exists a probability measure σ on the sphere $\mathbb{S}_{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$ and a positive function a such that $a(u) \to \infty$ for $u \to \infty$, and for any measureable $B \subset \mathbb{S}_{d-1}$, for which $\sigma(\partial B) = 0$, and $r > 0$,

$$u\,\mathbb{P}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in B,\, \|\boldsymbol{X}\| > ra(u)\right) \longrightarrow r^{-\alpha}\sigma(B) \quad \text{as} \quad u \to \infty. \tag{2.16}$$

In this case, we call $\alpha > 0$ the *index of regular variation* and σ the (Euclidean) *(probability) spectral measure* of the regularly varying random vector $\boldsymbol{X}$, and a an *auxiliary function*.

Given a probability measure σ on $\mathbb{S}_{d-1}$ and $\alpha > 0$, we may define a measure $\Lambda_{\alpha,\sigma}$ on $\mathbb{R}^d \setminus \{\mathbf{0}\}$ via

$$\Lambda_{\alpha,\sigma}\left(\left\{\boldsymbol{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\} \,:\, \|\boldsymbol{x}\| > r, \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \in B\right\}\right) = r^{-\alpha}\sigma(B).$$

The convergence (2.16) is then equivalent to the *vague$^{\#}$-convergence* of

$$t\mathbb{P}(a(t)^{-1}\boldsymbol{X} \in \cdot)$$

to $\Lambda_{\alpha,\sigma}$ in the sense of Kulik and Soulier (2020), Definition 2.1.1.

In fact, if the measure $\Lambda_{\alpha,\sigma}$ is supported only on the punctured upper orthant $[0,\infty)^d \setminus \{\mathbf{0}\}$, or, equivalently the probability spectral measure is concentrated on the upper orthant part of the sphere $\mathbb{S}_+^{(d-1)} = \{\boldsymbol{u} \in [0,\infty)^d \ : \ \|\boldsymbol{u}\|_2 = 1\}$, then the multivaraite regular variation property (2.16) is equivalent to $\boldsymbol{X}$ being in the *max-domain of attraction* of an associated max-stable random vector as follows, cf. Dombry and Ribatet (2015) Theorem 1 (and note that the norm therein may be replaced by any other in the finite-dimensional setting).

Theorem 2.3.7. *For each $n \in \mathbb{N}$ let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be independent copies of a d-dimensional random vector $\boldsymbol{X}$ with positive components. Consider the following two statements:*

(i) $\boldsymbol{X}$ is regularly varying with index α and probability spectral measure σ and auxiliary function a.

(ii) There exists a normalizing sequence $\widetilde{a}_n$, such that, as $n \to \infty$, the normalized componentwise maximum

$$\widetilde{a}_n^{-1} \max\left(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\right)$$

converges to the max-stable random vector

$$\boldsymbol{M}_{\widetilde{\alpha},\widetilde{\sigma}} = \max_{i\geq 1} \Gamma_i^{-1/\widetilde{\alpha}} \boldsymbol{Y}_i,$$

where $\{(\Gamma_i, \boldsymbol{Y}_i)\}_{i=1,2,\ldots}$ is a Poisson process on $(0,\infty) \times \mathbb{S}_+^{(d-1)}$ with intensity measure the product of the Lebesgue measure and $\widetilde{\sigma}$.

Then (i) implies (ii) with $\widetilde{\alpha} = \alpha$, $\widetilde{\sigma} = \sigma$ and $\widetilde{a}_n = a(n)$. Conversely, (ii) implies (i) with $\alpha = \widetilde{\alpha}$, $\sigma = \widetilde{\sigma}$ and the auxiliary function a may be chosen such that $\widetilde{a}_n = a(n)$.

Due to the marginal standardisation to unit Frèchet margins, we have been working only with $\alpha = 1$ in Section 2.1 and spectral measures that are concentrated on different spheres (with respect to other norms). Dombry and Ribatet (2015) have documented translation mechanisms of measure transforms. Up to a marginal standardisation, the random vector $\boldsymbol{Y}$ in Theorem 2.3.7 can be seen as a generator in the sense of Section 2.1.

Chapter 3

Stochastic ordering among multivariate extremes

Research on stochastic orderings and inequalities cover several decades, culminating among a vast literature for instance in the two monographs of Shaked and Shanthikumar (2007) and Müller and Stoyan (2002), the latter with a view towards applications and stochastic models, which appear in queuing theory, survival analysis, statistical physics or portfolio optimisation. Li and Li (2013) summarises developments of stochastic orders in reliability and risk management. While the scientific activities in finance, insurance, welfare economics or management science have been a driving force for many advances in the area, applications of stochastic orders are numerous and not limited to these fields. Importantly, such orderings will often play a role for robust inference, when only partial knowledge about a highly complex stochastic model is available.

Within the Extremes literature, related notions of positive dependence are well-known. It is a long-standing result that multivariate extreme value distributions exhibit positive association (Marshall and Olkin, 1983). More generally, max-infinitely divisible distributions have this property as shown in Resnick (1987), while Beirlant et al. (2004) summarise further implications in terms of positive dependence notions. Recently, an extremal version of the popular multivariate totally positive of order 2 (MTP2) property (Karlin and Rinott (1980); Fallat et al. (2017)) has been studied in the context of multivariate extreme value distributions, especially Hüsler-Reiß distributions, and linked to graphical modelling, sparsity and implicit regularisation in multivariate extreme value models (Röttger et al., 2023). Without any hope of being exhaustive, further fundamental scientific activity of the last decade on comparing stochastic models with a focus on multivariate extremes includes for instance an ordering of multivariate risk models based on extreme portfolio losses (Mainik and Rüschendorf, 2012), inequalities for mixtures on risk aggregation (Chen et al., 2022), a comparison of dependence in multivariate extremes via tail orders (Li, 2013) or stochastic

ordering for conditional extreme value modelling (Papastathopoulos and Tawn, 2015).

Yuen and Stoev (2014) use stochastic dominance results from Strokorb and Schlather (2015) in order to derive bounds on the maximum portfolio loss and extend their work in Yuen et al. (2020) to a distributionally robust inference for extreme Value-at-Risk.

In this chapter we go back to some fundamental questions concerning stochastic orderings among multivariate extreme value distributions. We focus on the order of positive quadrant dependence (PQD order, also termed concordance order), which is defined via orthant orders. Formally, a random vector $\boldsymbol{X}$ is said to be smaller than a random vector $\boldsymbol{Y}$ in the positive quadrant order if $F(\boldsymbol{x}) \leq G(\boldsymbol{x})$ and $\bar{F}(\boldsymbol{x}) \leq \bar{G}(\boldsymbol{x})$ for all $\boldsymbol{x}$, where F and G are the distribution functions and $\bar{F}$ and $\bar{G}$ the survival functions of the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. The relation to orthant orders is given in Definition 3.1.1. Answers are given to the following questions.

- *What is the relation between orders among max-stable distributions and corresponding orders among their exponent measures?* (Theorem 3.2.1 and Corollary 3.2.2)
- *Can we find characterisations in terms of other typical dependency descriptors (stable tail dependence function, generators, max-zonoids)?* (Theorem 3.2.1)
- *What is the role of fully independent and fully dependent model in this framework?* (Corollary 3.2.3)
- *What is the role of Choquet/Tawn-Molchanov models in this framework?* (Corollary 3.2.4 and Lemma 3.2.14)

For lower orthants, the answers are readily deduced from standard knowledge in Extremes. It is dealing with the upper orthants that makes this work interesting. The key element in the proof of our most fundamental characterisation result, Theorem 3.2.1, is based on Proposition A.0.9 below, which may be of independent interest. Stochastic orders are typically considered for probability distributions only. In order to make sense of the first question, we introduce corresponding orders for exponent measures, which turn out natural in this context, cf. Definition 3.1.2.

Second, we draw our attention to two popular parametric families of multivariate extreme value distributions that are closed under taking marginal distributions.

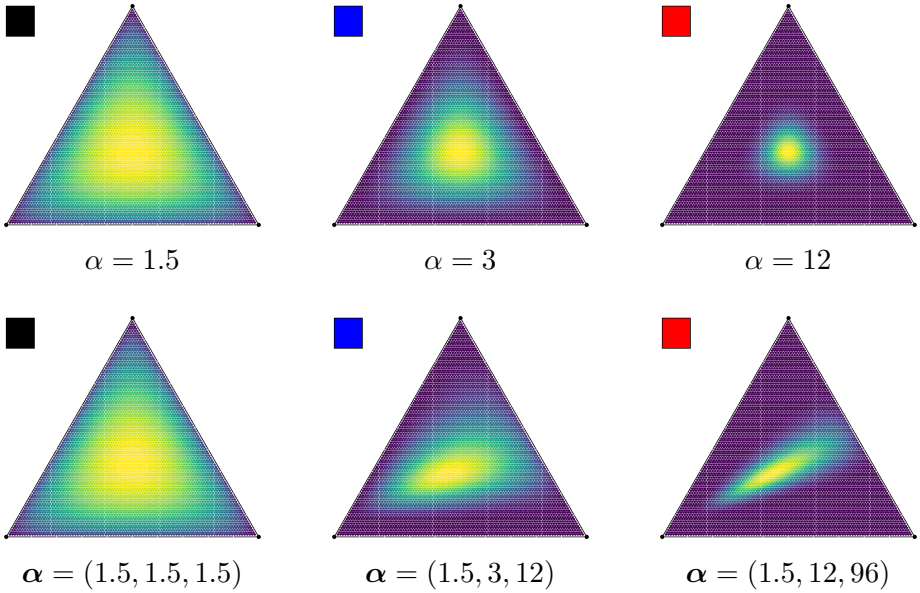

Figure 3.1: Angular densities (heat maps) of the symmetric max-stable Dirichlet model (top) and of the asymmetric max-stable Dirichlet model (bottom), cf. (2.7) for an expression of the density. Larger values are represented by brighter colours. The corresponding max-stable distributions are stochastically ordered in the PQD sense, increasing from left to right (Theorem 3.2.5). The black, blue and red boxes encode the matching with Figures 3.5 and 3.6.

- *Can we find order relations among the Dirichlet and Hüsler-Reiß parametric models?* (Theorem 3.2.5 and Theorem 3.2.9)

The answers are affirmative. For the Hüsler-Reiß model the result may be even strengthened for the supermodular order, which is otherwise beyond the scope of this work. To give an impression of the result for the Dirichlet family, Figure 3.1 depicts six angular densities of the trivariate max-stable Dirichlet model. Aulbach et al. (2015) showed already that the symmetric models associated with the top row densities are decreasing in the lower orthant sense. Our new result covers the asymmetric case depicted in the bottom row; we show that the associated multivariate extreme value distributions are decreasing in the (even stronger) PQD-sense (with a more streamlined proof).

Accordingly, this chapter is structured as follows. In Section 3.1 we review the relevant multivariate stochastic orderings together with important closure properties. This section contains also our (arguably natural) definition for corresponding order notions for exponent measures. All main results are then given in Section 3.2. Auxiliary results are postponed Appendix A. Appendix B contains background material how we obtained the illustrations (max-zonoid envelopes for bivariate Hüsler-Reiß and Dirichlet families) depicted in Figures 2.1, 3.3, 3.4 and 3.7.

3.1 Prerequisites from stochastic orderings

A wealth of stochastic orderings and associated inequalities have been summarised in Müller and Stoyan (2002) and Shaked and Shanthikumar (2007), the most fundamental order being the *usual stochastic order*

$$F \leq_{\text{st}} G$$

between two univariate distributions F and G, which is defined as $F(x) \geq G(x)$ for all $x \in \mathbb{R}$. This means that draws from F are less likely to attain large values than draws from G.

For multivariate distributions definitions of orderings are less straightforward and there are many more notions of stochastic orderings. We will focus on upper orthants, lower orthants and the PQD order here. A subset $U \subset \mathbb{R}^d$ is called an *upper orthant* if it is of the form

$$U = U_{\boldsymbol{a}} = \{\boldsymbol{x} \in \mathbb{R}^d \, : \, x_1 > a_1, \ldots, x_d > a_d\}$$

for some $\boldsymbol{a} \in \mathbb{R}^d$. Similarly, a subset $L \subset \mathbb{R}^d$ is called a *lower orthant* if it is of the form

$$L = L_{\boldsymbol{a}} = \{\boldsymbol{x} \in \mathbb{R}^d \, : \, x_1 \leq a_1, \ldots, x_d \leq a_d\}$$

for some $\boldsymbol{a} \in \mathbb{R}^d$.

Definition 3.1.1 (Multivariate orders LO, UO, PQD, Shaked and Shanthikumar (2007), Sections 6.G and 9.A, Müller and Stoyan (2002), Sections 3.3. and 3.8)**.** Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^d$ be two random vectors.

- $\boldsymbol{X}$ is said to be *smaller than* $\boldsymbol{Y}$ *in the upper orthant order*,

  $$\text{denoted } \boldsymbol{X} \leq_{\text{uo}} \boldsymbol{Y},$$

  if $\mathbb{P}(\boldsymbol{X} \in U) \leq \mathbb{P}(\boldsymbol{Y} \in U)$ for all upper orthants $U \subset \mathbb{R}^d$.
- $\boldsymbol{X}$ is said to be *smaller than* $\boldsymbol{Y}$ *in the lower orthant order*,

  $$\text{denoted } \boldsymbol{X} \leq_{\text{lo}} \boldsymbol{Y},$$

  if $\mathbb{P}(\boldsymbol{X} \in L) \geq \mathbb{P}(\boldsymbol{Y} \in L)$ for all lower orthants $L \subset \mathbb{R}^d$.
- $\boldsymbol{X}$ is said to be *smaller than* $\boldsymbol{Y}$ *in the positive quadrant order* ,

  $$\text{denoted } \boldsymbol{X} \leq_{\text{PQD}} \boldsymbol{Y},$$

  if we have the relations $\boldsymbol{X} \leq_{\text{uo}} \boldsymbol{Y}$ and $\boldsymbol{X} \geq_{\text{lo}} \boldsymbol{Y}$.

Note that the PQD order (also termed concordance order) is a dependence order. If $\boldsymbol{X} \leq_{\mathrm{PQD}} \boldsymbol{Y}$ holds, it implies that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have identical univariate marginals. Several equivalent characterizations of these orders are summarised in the respective sections of Müller and Stoyan (2002) and Shaked and Shanthikumar (2007). In relation to portfolio properties, it is interesting to note that for non-negative random vectors $\boldsymbol{X}, \boldsymbol{Y} \in [0,\infty)^d$

$$\boldsymbol{X} \leq_{\mathrm{uo}} \boldsymbol{Y} \quad \Longleftrightarrow \quad \min_{i=1,\dots,d}(a_i X_i) \leq_{\mathrm{st}} \min_{i=1,\dots,d}(a_i Y_i) \quad \text{for all } \boldsymbol{a} \in (0,\infty)^d; \tag{3.1}$$

$$\boldsymbol{X} \leq_{\mathrm{lo}} \boldsymbol{Y} \quad \Longleftrightarrow \quad \max_{i=1,\dots,d}(a_i X_i) \leq_{\mathrm{st}} \max_{i=1,\dots,d}(a_i Y_i) \quad \text{for all } \boldsymbol{a} \in (0,\infty)^d. \tag{3.2}$$

In addition, if $\boldsymbol{X}, \boldsymbol{Y} \in [0,\infty)^d$ and $\boldsymbol{X} \leq_{\mathrm{lo}} \boldsymbol{Y}$, then

$$\mathbb{E}\, g\Big(\sum_{i=1}^d a_i X_i\Big) \leq \mathbb{E}\, g\Big(\sum_{i=1}^d a_i Y_i\Big),$$

for all $\boldsymbol{a} \in [0,\infty)^d$ and all *Bernstein functions* g, provided that the expectation exists, cf. Shaked and Shanthikumar (2007) 6.G.14 and 5.A.4 for this fact and Appendix A for a definition of Bernstein functions. In particular, such functions are non-negative, monotonously increasing and concave and therefore form a natural class of utility functions, see e.g. Brockett and Golden (1987) and Caballé and Pomansky (1996). Important examples of Bernstein functions include the identity function, $g(x) = \log(1+x)$ or $g(x) = (1+x)^\alpha - 1$ for $\alpha \in (0,1)$.

The multivariate orders from Definition 3.1.1 have several useful **closure properties**. We refer to Müller and Stoyan (2002) Theorem 3.3.19 and Theorem 3.8.7 for a systematic collection, including

- independent or identical concatenation,
- marginalisation,
- distributional convergence,
- applying increasing transformations to the components,
- taking mixtures.

In what follows, we will need a corresponding notion of multivariate orders not only for probability measures on $\mathbb{R}^d$, but also for exponent measures as introduced in Section 2.1. While the support of an exponent measure Λ is contained in $[0,\infty)^d \setminus \{\mathbf{0}\}$, its total mass is infinite. We only know for sure that $\Lambda(B)$ is finite for Borel sets $B \subset \mathbb{R}^d$ *bounded away from the origin* in the sense that there exists $\varepsilon > 0$, such that $B \cap L_{\varepsilon \boldsymbol{e}} = \emptyset$ (recall $L_{\varepsilon \boldsymbol{e}} = \{\boldsymbol{x} \in \mathbb{R}^d : x_1 \leq \varepsilon, \dots, x_d \leq \varepsilon\}$). This means that we need to assume a different view on lower orthants and work with their complements instead, a subtlety, which did not matter previously when

defining such notions for probability measures only. The following notion seems natural in view of Definition 3.1.1 and the results of Section 3.2. Figure 3.2 illustrates the restriction to fewer admissible test sets for these orders for exponent measures.

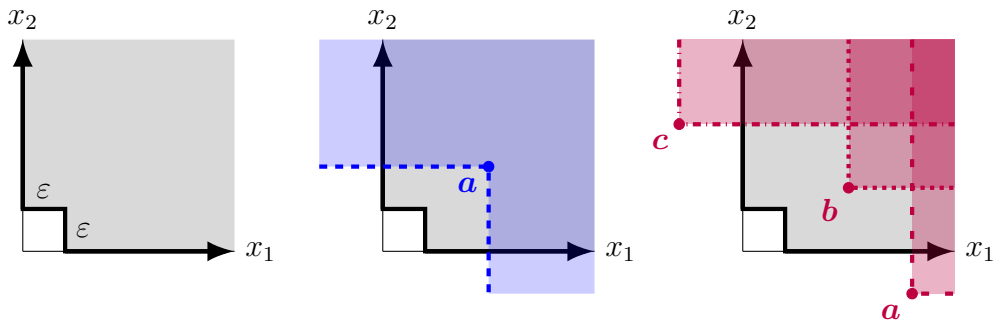

Figure 3.2: Illustration of test sets for multivariate orders for exponent measures in dimension $d = 2$, cf. Definition 3.1.2. Left: Λ is locally finite on the (closed) grey area for all $\varepsilon > 0$, its total (infinite) mass is contained in the union of such sets; middle: admissible complement of a lower orthant $\mathbb{R}^2 \setminus L_{\boldsymbol{a}}$ (blue area) for testing lower orthant order for Λ; right: admissible upper orthants $U_{\boldsymbol{a}}$, $U_{\boldsymbol{b}}$, $U_{\boldsymbol{c}}$ (three red areas) for testing upper orthant order for Λ.

Definition 3.1.2 (Multivariate orders for exponent measures)**.** Let $\Lambda, \widetilde{\Lambda}$ be two infinite measures on $\mathbb{R}^d$ with mass contained in $[0,\infty)^d \setminus \{\mathbf{0}\}$ and taking finite values on Borel sets bounded away from the origin.

- Λ is said to be *smaller than* $\widetilde{\Lambda}$ *in the upper orthant order*, denoted $\Lambda \leq_{\text{uo}} \widetilde{\Lambda}$, if $\Lambda(U) \leq \widetilde{\Lambda}(U)$ for each upper orthant $U \subset \mathbb{R}^d$ that is bounded away from the origin.
- Λ is said to be *smaller than* $\widetilde{\Lambda}$ *in the lower orthant order*, denoted $\Lambda \leq_{\text{lo}} \widetilde{\Lambda}$, if $\Lambda(\mathbb{R}^d \setminus L) \leq \widetilde{\Lambda}(\mathbb{R}^d \setminus L)$ for all lower orthants $L \subset \mathbb{R}^d$ such that $\mathbb{R}^d \setminus L$ is bounded away from the origin.
- Λ is said to be *smaller than* $\widetilde{\Lambda}$ *in the positive quadrant order*, denoted $\Lambda \leq_{\text{PQD}} \widetilde{\Lambda}$,
  if we have the relations $\Lambda \leq_{\text{uo}} \widetilde{\Lambda}$ and $\Lambda \geq_{\text{lo}} \widetilde{\Lambda}$.

Remark 3.1.3. Exponent measures Λ and $\widetilde{\Lambda}$ are Radon measures on $[0,\infty]^d \setminus \{\mathbf{0}\}$ (the one-point uncompactification of $[0,\infty]^d$). Any Borel set $B \subset [0,\infty]^d \setminus \{\mathbf{0}\}$ bounded away from the origin is relatively compact in this space, hence $\Lambda(B)$ and $\widetilde{\Lambda}(B)$, including $\Lambda(U)$, $\widetilde{\Lambda}(U)$, $\Lambda(\mathbb{R}^d \setminus L)$ and $\widetilde{\Lambda}(\mathbb{R}^d \setminus L)$ as above, are all finite.

3.2 Main results

First we present some fundamental characterisations of LO, UO and PQD order among simple max-stable distributions and their exponent measures, then we study these orders among the introduced parametric families. While we focus on simple max-stable distributions in what follows, we would like to stress that applying componentwise identical isotonic transformations to random vectors preserves orthant and concordance orders; in this sense the following properties can be seen as statements about the respective copulas. In particular, among max-stable random vectors, it suffices to establish these orders among simple max-stable random vectors and they translate immediately to all counterparts with different marginal distributions, cf. (2.4).

3.2.1 Fundamental results

We start by assembling the most fundamental relations for multivariate orders among simple max-stable random vectors. While the statements about lower orthant orders are almost immediate from existing theory and definitions, the relations for upper orthants are a bit more intricate and non-standard in the area. In particular, showing that "$\Lambda \leq_{\text{uo}} \widetilde{\Lambda}$ implies $G \leq_{\text{uo}} \widetilde{G}$" turns out to be non-trivial. The key ingredient in the proof of the following theorem is Proposition A.0.9 for part b).

Theorem 3.2.1 (Orthant orders characterisations)**.** *Let G and $\widetilde{G}$ be d-variate simple max-stable distributions with exponent measures Λ and $\widetilde{\Lambda}$, generators $\boldsymbol{Z}$ and $\widetilde{\boldsymbol{Z}}$, stable tail dependence functions ℓ and $\widetilde{\ell}$ and max-zonoids K and $\widetilde{K}$, respectively.*

a) The following statements are equivalent.

(i) $G \leq_{\text{lo}} \widetilde{G}$;
(ii) $\Lambda \leq_{\text{lo}} \widetilde{\Lambda}$;
(iii) $\mathbb{E}(\max_{i=1,\dots,d}(a_i Z_i)) \leq \mathbb{E}(\max_{i=1,\dots,d}(a_i \widetilde{Z}_i))$ *for all* $\boldsymbol{a} \in (0,\infty)^d$;
(iv) $\ell \leq \widetilde{\ell}$;
(v) $K \subset \widetilde{K}$.

b) The following statements are equivalent.

(i) $G \leq_{\text{uo}} \widetilde{G}$;
(ii) $\Lambda \leq_{\text{uo}} \widetilde{\Lambda}$;
(iii) $\mathbb{E}(\min_{i\in A}(a_i Z_i)) \leq \mathbb{E}(\min_{i\in A}(a_i \widetilde{Z}_i))$ *for all* $\boldsymbol{a} \in (0,\infty)^d$ *and* $A \subset \{1,\dots,d\}$, $A \neq \emptyset$.

c) If $d = 2$, the following statements are equivalent.

(i) $G \leq_{\mathrm{PQD}} \widetilde{G}$;
(ii) $G \leq_{\mathrm{uo}} \widetilde{G}$;
(iii) $G \geq_{\mathrm{lo}} \widetilde{G}$.

Proof. In what follows, let $\boldsymbol{X} \sim G$ and $\widetilde{\boldsymbol{X}} \sim \widetilde{G}$.

a) Because of (3.2), it suffices to compare $G(\boldsymbol{x})$ and $\widetilde{G}(\boldsymbol{x})$ for $\boldsymbol{x} \in (0,\infty)^d$ only. The same is true for ℓ and $\widetilde{\ell}$ as they are continuous on $[0,\infty)^d$. At the same time the test sets for the relation $\Lambda \leq_{\mathrm{lo}} \widetilde{\Lambda}$ in Definition 3.1.2 are precisely of the form $\mathbb{R}^d \setminus L_{\boldsymbol{x}}$, where $\boldsymbol{x} \in (0,\infty)^d$. So the equivalence of (i), (ii), (iii) and (iv) follows directly from the relations

$$G(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \in L_{\boldsymbol{x}}) = \exp(-\Lambda([0,\infty]^d \setminus L_{\boldsymbol{x}})) = \exp(-\Lambda(\mathbb{R}^d \setminus L_{\boldsymbol{x}}))$$

with

$$\Lambda(\mathbb{R}^d \setminus L_{\boldsymbol{x}}) = \ell(1/x_1, \ldots, 1/x_d) = \mathbb{E}(\max(Z_1/x_1, \ldots, Z_d/x_d)),$$

and the respective tilde-counterparts. Likewise, the equivalence of (iv) and (v) is immediate from (2.6) and (2.5).

b) We start by showing the equivalence between (ii) and (iii). The test sets for the relation $\Lambda \leq_{\mathrm{uo}} \widetilde{\Lambda}$ in Definition 3.1.2 are precisely the upper orthants $U_{\boldsymbol{x}}$, where at least one component of $\boldsymbol{x}$ is larger than zero. Let $\boldsymbol{a} \in (0,\infty)^d$ and $A \subset \{1,\ldots,d\}$, $A \neq \emptyset$. Define $\boldsymbol{x} \in \mathbb{R}^d$ by setting $x_i = 1/a_i$ if $i \in A$ and $x_i = -1$ else. Then $U_{\boldsymbol{x}}$ is an admissible test set and

$$\Lambda(U_{\boldsymbol{x}}) = \Lambda\left(\left\{\boldsymbol{y} \in [0,\infty)^d \setminus \{\boldsymbol{0}\} \,:\, \min_{i\in A}(a_i y_i) > 1\right\}\right) = \mathbb{E}\Big(\min_{i\in A}(a_i Z_i)\Big). \tag{3.3}$$

Likewise, $\widetilde{\Lambda}(U_{\boldsymbol{x}}) = \mathbb{E}(\min_{i\in A}(a_i \widetilde{Z}_i))$ and we may deduce the implication (ii)⇒(iii). Conversely, assume (iii) and note that the same argument implies $\Lambda(U_{\boldsymbol{x}}) \leq \widetilde{\Lambda}(U_{\boldsymbol{x}})$ for any $\boldsymbol{x}$, which has at least one positive component, whilst all other components of $\boldsymbol{x}$ are negative. What remains to be seen is the same relation for upper orthants $U_{\boldsymbol{x}}$, for which at least one component of $\boldsymbol{x}$ is positive, but where among the non-positive components, there may be zeroes. Let $\boldsymbol{x} \in \mathbb{R}^d$ be such a vector. For $n \in \mathbb{N}$ let $\boldsymbol{x}_n \in \mathbb{R}^d$ be an identical vector, but with zero entries replaced by $1/n$. Then $\Lambda(U_{\boldsymbol{x}_n}) \leq \widetilde{\Lambda}(U_{\boldsymbol{x}_n})$ for all $n \in \mathbb{N}$ by the previous argument, whilst $U_{\boldsymbol{x}_n} \uparrow U_{\boldsymbol{x}}$, such that $\Lambda(U_{\boldsymbol{x}_n}) \to \Lambda(U_{\boldsymbol{x}})$ and $\widetilde{\Lambda}(U_{\boldsymbol{x}_n}) \to \widetilde{\Lambda}(U_{\boldsymbol{x}})$ as $n \to \infty$. This shows (iii)⇒(ii).

Next, we establish (i)⇒(iii). Assume (i). Since the order UO is closed under marginalisation, it suffices to consider $A = \{1,\ldots,d\}$ in (iii), see also Lemma 2.1.4. Set $x_i = 1/a_i$, $i = 1,\ldots,d$, such that (3.3) holds (as well as the

tilde-version) and note that the closure of $U_{\boldsymbol{x}}$ in $[0,\infty]^d\setminus\{\mathbf{0}\}$ is a continuity set for each of the (-1)-homogeneous measures Λ and $\widetilde{\Lambda}$. Hence, since each max-stable vector satisfies its own Domain-of-attraction conditions (cf. e.g. Resnick (1987) Section 5.4.2), we have

$$\Lambda(U_{\boldsymbol{x}}) = \lim_{n\to\infty} n\mathbb{P}(\boldsymbol{X}\in nU_{\boldsymbol{x}}) = \lim_{n\to\infty} n\mathbb{P}(\boldsymbol{X}\in U_{n\boldsymbol{x}})$$

and the analog for $\widetilde{\Lambda}$ and $\widetilde{\boldsymbol{X}}$. The implication (i)⇒(iii) follows.

Lastly, let us establish (iii)⇒(i). Suppose (iii) holds. We abbreviate $\chi^{(a)}(A) = \mathbb{E}(\min_{i\in A}(a_iZ_i))$ and analogously $\widetilde{\chi}^{(a)}(A) = \mathbb{E}(\min_{i\in A}(a_i\widetilde{Z}_i))$, such that (iii) translates into

$$\chi^{(a)}(A) \le \widetilde{\chi}^{(a)}(A)$$

for all $\boldsymbol{a}\in(0,\infty)^d$ and $A\subset\{1,\dots,d\}$, $A\neq\emptyset$. With $\theta^{(a)}(A) = \mathbb{E}(\max_{i\in A}(a_iZ_i))$ we find that

$$\chi^{(a)}(A) = \sum_{I\subset A,\, I\neq\emptyset} (-1)^{|I|+1}\theta^{(a)}(I) \qquad \text{and} \qquad \theta^{(a)}(A) = \sum_{I\subset A,\, I\neq\emptyset} (-1)^{|I|+1}\chi^{(a)}(I)$$

(similarly to (2.13) and analogously for the tilde-version), where $\theta^{(a)}$ can be interpreted as directional extremal coefficient function. It is easily seen that $\theta^{(a)}$ with $\theta^{(a)}(\emptyset)=0$ is union-completely alternating, cf. Lemma A.0.3.

Because of (3.1), in order to arrive at (i), it suffices to establish

$$\mathbb{P}\Big(\min_{i=1,\dots,d}(a_iX_i) > 1\Big) \le \mathbb{P}\Big(\min_{i=1,\dots,d}(a_i\widetilde{X}_i) > 1\Big)$$

for all $\boldsymbol{a}\in(0,\infty)^d$. The left-hand side can be rewritten as

$$\begin{aligned}
\mathbb{P}\Big(\min_{i=1,\dots,d}(a_iX_i) > 1\Big) &= 1 - \sum_{I\subset\{1,\dots,d\},\, I\neq\emptyset} (-1)^{|I|+1}\mathbb{P}\Big(\max_{i\in I}(a_iX_i)\le 1\Big)\\
&= 1 - \sum_{I\subset\{1,\dots,d\},\, I\neq\emptyset} (-1)^{|I|+1}\exp\big(-\ell(\boldsymbol{a}_I)\big)\\
&= -\underbrace{\sum_{I=\emptyset}(-1)^{|I|+1}\exp\big(-\ell(\boldsymbol{a}_I)\big)}_{=-1} - \sum_{I\subset\{1,\dots,d\},\, I\neq\emptyset} (-1)^{|I|+1}\exp\big(-\ell(\boldsymbol{a}_I)\big).
\end{aligned}$$

Hence, in the notation of Lemma A.0.3 we obtain

$$\begin{aligned}
\mathbb{P}\Big(\min_{i=1,\dots,d}(a_iX_i) > 1\Big) &= -\sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1}\exp\big(-\ell(\boldsymbol{a}_I)\big)\\
&= -\sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1}\exp\big(-\theta^{(a)}(I)\big)\\
&= \underbrace{\sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1}}_{=0} - \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1}\exp\big(-\theta^{(a)}(I)\big).
\end{aligned}$$

With $g(x) = 1 - \exp(-x)$, this gives

$$\begin{aligned}\mathbb{P}\Big(\min_{i=1,\dots,d}(a_i X_i) > 1\Big) &= \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1}\cdot 1 - \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1}\exp\big(-\theta^{(a)}(I)\big)\\ &= \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1}\big[1-\exp\big(-\theta^{(a)}(I)\big)\big]\\ &= \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1} g\big(\theta^{(a)}(I)\big)\end{aligned}$$

Analogously,

$$\mathbb{P}\Big(\min_{i=1,\dots,d}(a_i \widetilde{X}_i) > 1\Big) = \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|+1} g\big(\widetilde{\theta}^{(a)}(I)\big)$$

The assertion follows then directly from Proposition A.0.9, since g is a Bernstein function.

c) The statement follows from the relation

$$\begin{aligned}&\mathbb{E}(\min(a_1 Z_1, a_2 Z_2)) + \mathbb{E}(\max(a_1 Z_1, a_2 Z_2))\\ &= \mathbb{E}(a_1 Z_1 + a_2 Z_2) = a_1\mathbb{E}(Z_1) + a_2\mathbb{E}(Z_2) = a_1 + a_2,\end{aligned}$$

and likewise

$$\mathbb{E}(\min(a_1 \widetilde{Z}_1, a_2 \widetilde{Z}_2)) + \mathbb{E}(\max(a_1 \widetilde{Z}_1, a_2 \widetilde{Z}_2)) = a_1 + a_2,$$

hence both sides are equal to $a_1 + a_2$. So we have

$$\mathbb{E}(\min(a_1 Z_1, a_2 Z_2)) \le \mathbb{E}(\min(a_1 \widetilde{Z}_1, a_2 \widetilde{Z}_2))$$

if and only if

$$\mathbb{E}(\max(a_1 Z_1, a_2 Z_2)) \ge \mathbb{E}(\max(a_1 \widetilde{Z}_1, a_2 \widetilde{Z}_2)).$$

The result then follows from part a) and b). □

The assumption $d = 2$ is important in part c); these equivalences are no longer true in higher dimensions, cf. Example 3.2.15 below. Theorem 3.2.1 implies further that the orthant ordering of two generators $\boldsymbol{Z}$ and $\widetilde{\boldsymbol{Z}}$ implies the respective ordering of the corresponding distributions G and $\widetilde{G}$ and exponent measures Λ and $\widetilde{\Lambda}$. However, the converse is false and most generators will not satisfy orthant orderings, even when the corresponding distributions do. An interesting case for this phenomenon is the Hüsler-Reiß family, cf. Example 3.2.13 below. The following corollary is another immediate consequence of Theorem 3.2.1.

Corollary 3.2.2 (PQD/concordance order characterisation)**.** *Let G and $\widetilde{G}$ be d-variate simple max-stable distributions with exponent measures Λ and $\widetilde{\Lambda}$, then*

$$G \le_{\mathrm{PQD}} \widetilde{G} \quad \Longleftrightarrow \quad \Lambda \le_{\mathrm{PQD}} \widetilde{\Lambda}.$$

It is well-known that for any stable tail dependence function ℓ of a simple max-stable random vector

$$\ell_{\mathrm{dep}}(\boldsymbol{x}) = \|\boldsymbol{x}\|_\infty \leq \ell(\boldsymbol{x}) \leq \|\boldsymbol{x}\|_1 = \ell_{\mathrm{indep}}(\boldsymbol{x}), \qquad \boldsymbol{x} \geq \boldsymbol{0}, \tag{3.4}$$

where ℓ_{dep} represents the degenerate max-stable random vector, whose components are fully dependent, and ℓ_{indep} corresponds to the max-stable random vector with completely independent components. From the perspective of stochastic orderings this means that every max-stable random vector is dominated by the fully independent model, while it dominates the fully dependent model with respect to the lower orthant order. It seems less well-known that the converse ordering holds true for upper orthants, so that we arrive at the following corollary.

Corollary 3.2.3 (PQD/concordance for independent and fully dependent model)**.** *Let G_{indep}, G_{dep} and G be d-dimensional simple max-stable distributions, where G_{indep} represents the model with fully independent components, and G_{dep} represents the model with fully dependent components. Then*

$$G_{\mathrm{indep}} \leq_{\mathrm{PQD}} G \leq_{\mathrm{PQD}} G_{\mathrm{dep}}.$$

Proof. In view of (3.4) and Theorem 3.2.1 b), if suffices to investigate the upper and lower bounds of $\mathbb{E}(\min_{i\in A}(a_i Z_i))$ for $\boldsymbol{a} \in (0,\infty)^d$ and $A \subset \{1,\dots,d\}$, $A \neq \emptyset$, where $\boldsymbol{Z}$ is a generator for G. We have

$$\mathbb{E}\big(\min_{i\in A}(a_i Z_i)\big) \leq \min_{i\in A}(\mathbb{E}(a_i Z_i)) = \min_{i\in A}(a_i)$$

and

$$\mathbb{E}\big(\min_{i\in A}(a_i Z_i)\big) \geq \begin{cases} a_j & \text{if } A = \{j\}, \\ 0 & \text{else,} \end{cases}$$

and the upper and lower bounds are attained by generators of the fully dependent model ($\boldsymbol{Z}$ being almost surely $\boldsymbol{e} = (1,1,\dots,1)^\top$) and the independent model ($\boldsymbol{Z}$ being uniformly distributed among the set $\{d\boldsymbol{e}_1, d\boldsymbol{e}_2, \dots, d\boldsymbol{e}_d\}$), respectively, which implies the assertion. □

Similarly Theorem 2.2.9 can be strengthened as follows. Whilst previously only the lower orthant order was known, we have in fact PQD/concordance ordering.

Corollary 3.2.4 (PQD/concordance for the associated Choquet model)**.** *Let $\boldsymbol{X}$ be a simple max-stable random vector with extremal coefficients $(\theta(A))$, $A \subset \{1,\dots,d\}$, $A \neq \emptyset$ and $\boldsymbol{X}^*$ the Choquet (Tawn-Molchanov) random vector with identical extremal coefficients. Then*

$$\boldsymbol{X}^* \leq_{\mathrm{PQD}} \boldsymbol{X}.$$

Proof. The lower orthant order $\boldsymbol{X}^* \geq_{\text{lo}} \boldsymbol{X}$ is known from Theorem 2.2.9. Let $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ be generators of the respective models. Since they share identical extremal coefficients, they also share identical tail dependence coefficients $\chi(A) = \mathbb{E}(\min_{i\in A} Z_i) = \mathbb{E}(\min_{i\in A} Z_i^*)$, $A \subset \{1,\dots,d\}$, $A \neq \emptyset$, which can be retrieved from θ via (2.13). In general, we have for $A \subset \{1,\dots,d\}$, $A \neq \emptyset$, $\boldsymbol{a} \in (0,\infty)^d$

$$\mathbb{E}\big(\min_{i\in A}(a_i Z_i)\big) \geq \min_{i\in A}(a_i)\mathbb{E}\big(\min_{i\in A}(Z_i)\big) = \min_{i\in A}(a_i)\cdot\chi(A).$$

The Choquet model attains the lower bound, since with (2.11) and (2.14)

$$\begin{aligned}\mathbb{E}\big(\min_{i\in A}(a_i Z_i^*)\big) &= \sum_{L\subset\{1,\dots,d\},L\neq\{\emptyset\}} \tau(L)\min_{i\in A}(a_i(\boldsymbol{e}_L)_i)\\ &= \sum_{L\subset\{1,\dots,d\},A\subset L} \tau(L)\min_{i\in A}(a_i) = \min_{i\in A}(a_i)\cdot\chi(A).\end{aligned}$$

So by Theorem 3.2.1 we also have $\boldsymbol{X}^* \leq_{\text{uo}} \boldsymbol{X}$, hence the assertion. □

3.2.2 Parametric models

In general, parametric families of multivariate distributions do not necessarily exhibit stochastic orderings. One of the few more interesting known examples among multivariate max-stable distributions is the Dirichlet family, for which it has been shown that it is ordered in the symmetric case (Aulbach et al., 2015, Proposition 4.4), that is, for $\alpha \leq \beta$ we have

$$\text{MaxDir}((\alpha,\alpha,\dots,\alpha)) \geq_{\text{lo}} \text{MaxDir}((\beta,\beta,\dots,\beta)). \tag{3.5}$$

Figure 3.3 illustrates (3.5) in the bivariate situation and shows a bivariate example that these distributions are otherwise not necessarily ordered in the asymmetric case.

Here, we extend (3.5) in several ways: (i) going beyond the symmetric situation considering the fully asymmetric model, (ii) considering PQD/concordance order, (iii) shortening the proof by exploiting a connection to the theory of majorisation. Figure 3.4 provides an illustration of the stochastic ordering for the asymmetric Dirichlet family in the bivariate case. In Figure 3.1 we see how the mass of the angular measure of the symmetric and asymmetric Dirichlet model is more concentrated from left plot to right plot. This also corresponds to their stochastic ordering, with the right one being the most dominant model in terms of PQD order.

Theorem 3.2.5 (PQD/concordance order of Dirichlet family)**.** *Consider the max-stable Dirichlet family from Theorem/Definition 2.2.1. If $\alpha_i \leq \beta_i$, $i = 1,\dots,d$, then*

$$\text{MaxDir}(\alpha_1,\alpha_2,\dots,\alpha_d) \leq_{\text{PQD}} \text{MaxDir}(\beta_1,\beta_2,\dots,\beta_d).$$

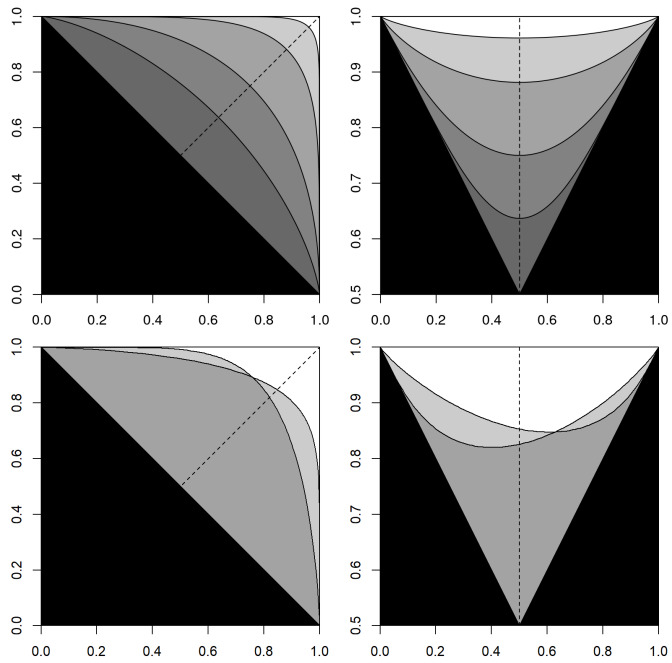

Figure 3.3: Top: Nested max-zonoids (left) and ordered (hypographs of) Pickands dependence functions (right) from the fully symmetric Dirichlet family for $\alpha \in \{0.0625, 0.25, 1, 4\}$. Smaller values of α correspond to larger sets and larger Pickands dependence functions and are closer to the independence model represented by the box $[0,1]^2$ or the constant function, which is identically 1. The fully dependent model is represented in black. Bottom: Non-nested max-zonoids and non-ordered Pickands dependence function from the asymmetric Dirichlet family for $(\alpha_1, \alpha_2) \in \{(0.15, 12), (4, 0.2)\}$.

In order to prove Theorem 3.2.5 we will use a simple inequality that follows from the theory of majorisation (Marshall et al., 2011).

Proposition 3.2.6 (Marshall and Proschan (1965) Corollary 3, Marshall et al. (2011) Proposition B.2.b.)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be continuous and convex and let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, then*

$$\mathbb{E}\, g\left(\sum_{i=1}^{n} \frac{X_i}{n}\right)$$

is nonincreasing in $n = 1, 2, \ldots,$ i.e. for natural numbers $1 \leq k < n$,

$$\mathbb{E}\, g\left(\sum_{i=1}^{k} \frac{X_i}{k}\right) \geq \mathbb{E}\, g\left(\sum_{i=1}^{n} \frac{X_i}{n}\right).$$

Corollary 3.2.7. *Let $g : \mathbb{R} \to \mathbb{R}$ be continuous and convex and, let $Z^{(\alpha)} \sim \Gamma(\alpha)$ follow a univariate Gamma distribution with shape parameter $\alpha > 0$ and unit scale, then*

$$\mathbb{E}\, g\left(\frac{Z^{(\alpha)}}{\alpha}\right)$$

is nonincreasing in $\alpha \in (0, \infty)$.

Proof. We consider first the case that $\alpha = (k/n) \cdot \beta$ for some natural numbers $1 \leq k < n$. Then consider independent and identically distributed random variables $\Gamma_1, \Gamma_2, \ldots$ following a $\Gamma(\beta/n)$ distribution. Then also $X_i = (n/\beta)\Gamma_i$, $i = 1, 2, \ldots$, are independent and identically distributed. Then Proposition 3.2.6 gives

$$\begin{aligned}\mathbb{E}\, g\left(\frac{\sum_{i=1}^k \Gamma_i}{\alpha}\right) &= \mathbb{E}\, g\left(\frac{n}{\beta} \cdot \sum_{i=1}^k \frac{\Gamma_i}{k}\right) = \mathbb{E}\, g\left(\sum_{i=1}^k \frac{X_i}{k}\right) \\ &\geq \mathbb{E}\, g\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \mathbb{E}\, g\left(\frac{n}{\beta} \cdot \sum_{i=1}^n \frac{\Gamma_i}{n}\right) = \mathbb{E}\, g\left(\frac{\sum_{i=1}^n \Gamma_i}{\beta}\right).\end{aligned}$$

By the convolution stability of the Gamma distribution

$$\sum_{i=1}^k \Gamma_i \sim \Gamma(\alpha) \qquad \text{and} \qquad \sum_{i=1}^n \Gamma_i \sim \Gamma(\beta).$$

Note that the argument above holds for fixed k and n, and no limiting considerations were needed so far. But we may conclude that the assertion holds for α and β that differ by a rational multiplier k/n, where k and n may be arbitrary natural numbers subject to $1 \leq k < n$. If we only know $\alpha < \beta$, consider a decreasing sequence $\beta_n \downarrow \beta$, such that α and β_n differ by a rational multiplier. This gives

$$\mathbb{E}g(Z^{(\alpha)}/\alpha) \geq \limsup_{n\to\infty} \mathbb{E}g(Z^{(\beta_n)}/\beta_n)$$

by the above argument. On the other hand, Fatou's lemma gives

$$\mathbb{E}g(Z^{(\beta)}/\beta) \leq \liminf_{n\to\infty} \mathbb{E}g(Z^{(\beta_n)}/\beta_n).$$

This finishes the proof. □

Proof of Theorem 3.2.5. If $\boldsymbol{\alpha} = \boldsymbol{\beta}$, the statement is clear. Else, because the parameter space of the Dirichlet model is $(0, \infty)^d$, we can find a chain of parameter vectors $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(0)} \leq \boldsymbol{\alpha}^{(1)} \leq \cdots \leq \boldsymbol{\alpha}^{(m)} = \boldsymbol{\beta}$, such that for each $i = 0, \ldots, m-1$, the vectors $\boldsymbol{\alpha}^{(i)}$ and $\boldsymbol{\alpha}^{(i+1)}$ differ only by one component. Hence it suffices to consider the case, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ differ only in one component. Without loss of generality, let this be the first component.

Let $\boldsymbol{Z}$ be a Gamma generator for MaxDir($\boldsymbol{\alpha}$) and $\widetilde{\boldsymbol{Z}}$ be a Gamma generator for for MaxDir($\boldsymbol{\beta}$) in the sense of Theorem/Definition 2.2.1. Then we may assume that $Z_i = \widetilde{Z}_i$ for $i = 2, \ldots, d$, whereas $\alpha_1 Z_1 \sim \Gamma(\alpha_1)$ and $\beta_1 \widetilde{Z}_1 \sim \Gamma(\beta_1)$ are independent from $(Z_2, \ldots, Z_d)^\top$, and $\alpha_1 < \beta_1$ by assumption. We will need to show (cf. Theorem 3.2.1) that for fixed $\boldsymbol{a} \in (0, \infty)^d$ and $A \subset \{1, \ldots, d\}$, $A \neq \emptyset$

$$\mathbb{E} \min_{i\in A} \left(a_i Z_i\right) \leq \mathbb{E} \min_{i\in A} \left(a_i \widetilde{Z}_i\right) \qquad \text{and} \qquad \mathbb{E} \max_{i=1,\ldots,d} \left(a_i Z_i\right) \geq \mathbb{E} \max_{i=1,\ldots,d} \left(a_i \widetilde{Z}_i\right).$$

Due to the setting above, it suffices to consider only subsets A with $1 \in A$, and due to the marginal standardisation $\mathbb{E}(Z_1) = 1$, it suffices to restrict our attention to $A \setminus \{1\} \neq \emptyset$. Setting $V_A = \min_{i \in A \setminus \{1\}} (a_i Z_i)$ and $W = \max_{i=2,\dots,d} (a_i Z_i)$ this means the assertion will follow from

$$\mathbb{E} \min \left(a_1 Z_1, V_A\right) \leq \mathbb{E} \min \left(a_1 \widetilde{Z}_1, V_A\right) \text{ and } \mathbb{E} \max \left(a_1 Z_1, W\right) \geq \mathbb{E} \max \left(a_1 \widetilde{Z}_1, W\right).$$

Indeed, this is implied by Corollary 3.2.7, when considering the continuous convex functions $g_c(x) = -\min(a_1 x, c)$ or $g_c(x) = \max(a_1 x, c)$ for a constant $c \in \mathbb{R}$. □

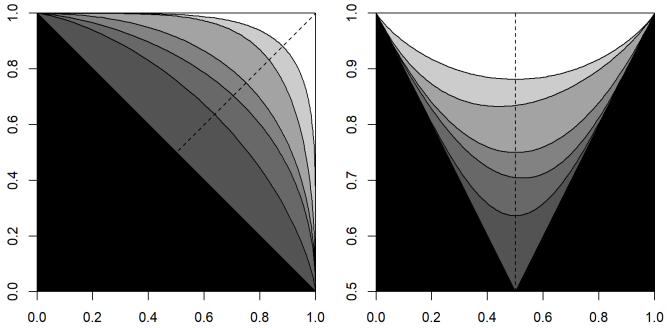

Figure 3.4: Nested max-zonoids and ordered Pickands dependence functions of the asymmetric max-stable Dirichlet family for $(\alpha_1, \alpha_2) \in \{(0.25, 0.25), (1, 0.25), (1, 1), (1, 4), (4, 4)\}$. Componentwise smaller values of (α_1, α_2) correspond to larger sets and larger Pickands dependence functions and are closer to the independence model.

Example 3.2.8. In order to draw attention to some further consequences of Theorem 3.2.5, let $\boldsymbol{X} \sim \mathrm{MaxDir}(\boldsymbol{\alpha})$ and $\boldsymbol{Y} \sim \mathrm{MaxDir}(\boldsymbol{\beta})$ where $\alpha_i \leq \beta_i$, $i = 1, \dots, d$, so that $\boldsymbol{X} \leq_{\mathrm{PQD}} \boldsymbol{Y}$, hence $\boldsymbol{X} \leq_{\mathrm{uo}} \boldsymbol{Y}$ and $\boldsymbol{X} \geq_{\mathrm{lo}} \boldsymbol{Y}$, which implies

$$\begin{aligned} \min_{i=1,\dots,d} (a_i X_i) &\leq_{\mathrm{st}} \min_{i=1,\dots,d} (a_i Y_i) && \text{for all } \boldsymbol{a} \in (0, \infty]^d. \\ \max_{i=1,\dots,d} (a_i X_i) &\geq_{\mathrm{st}} \max_{i=1,\dots,d} (a_i Y_i) && \text{for all } \boldsymbol{a} \in [0, \infty)^d, \end{aligned}$$

cf. (3.1), (3.2) and Lemma 2.2.3. Exemplarily, we consider a range of trivariate symmetric and asymmetric max-stable Dirichlet distributions $\mathrm{MaxDir}(\alpha_1, \alpha_2, \alpha_3)$ with parameters $(\alpha_1, \alpha_2, \alpha_3)$ given in Figure 3.1. The colouring is chosen such that red models PQD-dominate blue models, which PQD-dominate black models.

In addition, we consider the portfolio with equal weights $(1, 1, 1)$ and the resulting min-projections $\min(X_1, X_2, X_3)$ and max-projections $\max(X_1, X_2, X_3)$, where $(X_1, X_2, X_3) \sim \mathrm{MaxDir}(\alpha_1, \alpha_2, \alpha_3)$. Figures 3.5 and 3.6 display their distribution functions on the Gumbel scale. As commonly of interest for extreme value distributions, instead of the quantile function Q, we show the equivalent return level plot, which displays the return levels $Q(1-p)$ for the return period of

$1/p$ observations. The plots of these functions are based on empirical estimates from one million simulated observations from the respective models, and their orderings are as expected from the theory, i.e. quantile functions increase as the dominance of the model grows, while distribution functions decrease.

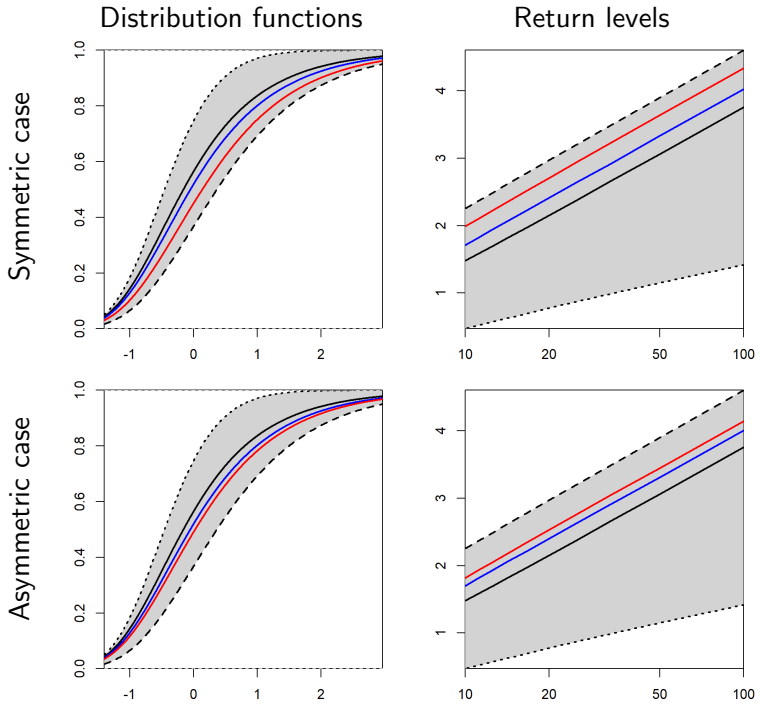

Figure 3.5: Distribution functions (left) and return levels (right) for return periods 10 to 100 (on logarithmic scale) of $\min(X_1, X_2, X_3)$, where $(X_1, X_2, X_3) \sim \text{MaxDir}(\alpha_1, \alpha_2, \alpha_3)$ on standard Gumbel scale with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ as chosen in Figure 3.1. Top: symmetric case; bottom: asymmetric case. Black, blue and red colouring encodes the matching with Figure 3.1. The grey areas represent the range between the fully dependent (dashed line) and fully independent (dotted line) cases.

Another prominent family of multivariate max-stable distributions that turns out to be stochastically ordered in the PQD/concordance order is the Hüsler-Reiß family. It can be shown using the limit result from Theorem 2.2.6 together with Slepian's normal comparison lemma and some closure properties of the PQD/concordance order. Figure 3.7 provides an illustration in terms of nested max-zonoids and ordered Pickands dependence functions in the bivariate case. However, while these models are ordered, we would like to point out that none of the typically chosen families of log-Gaussian generators satisfy any of the orthant orders, cf. Example 3.2.13.

Theorem 3.2.9 (PQD/concordance order of Hüsler-Reiß family)**.** *Consider the max-stable Hüsler-Reiß family from Theorem/Definition 2.2.4. If $\gamma_{i,j} \leq \widetilde{\gamma}_{i,j}$ for all $i, j \in \{1, \ldots, d\}$, then*

$$\text{HR}(\boldsymbol{\gamma}) \geq_{\text{PQD}} \text{HR}(\widetilde{\boldsymbol{\gamma}}).$$

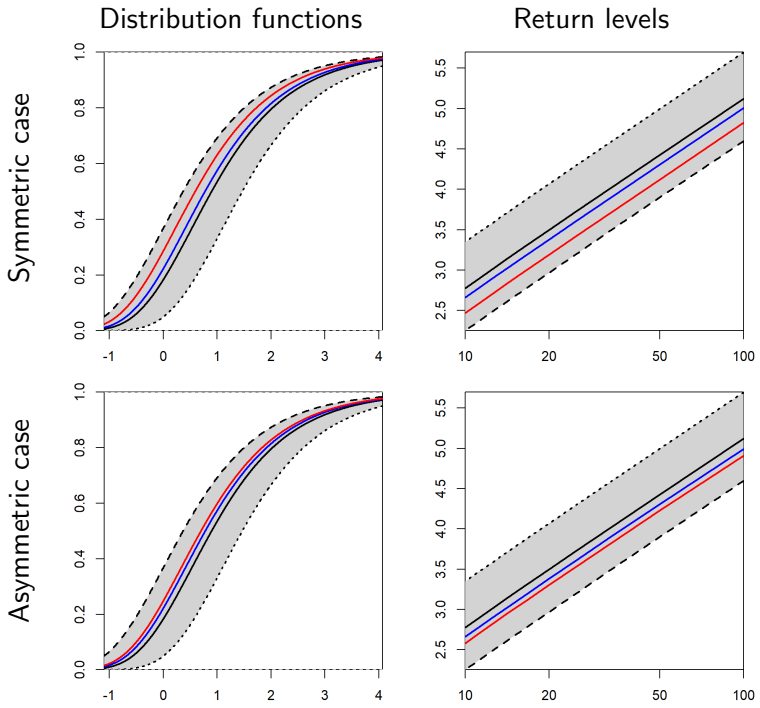

Figure 3.6: Distribution functions (left) and return levels (right) for return periods 10 to 100 (on logarithmic scale) of $\max(X_1, X_2, X_3)$, where $(X_1, X_2, X_3) \sim \mathrm{MaxDir}(\alpha_1, \alpha_2, \alpha_3)$ on standard Gumbel scale with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ as chosen in Figure 3.1. Top: symmetric case; bottom: asymmetric case. Black, blue and red colouring encodes the matching with Figure 3.1. The grey areas represent the range between the fully dependent (dashed line) and fully independent (dotted line) cases.

Remark 3.2.10. With almost identical proof, we may even deduce

$$\mathrm{HR}(\boldsymbol{\gamma}) \geq_{\mathrm{sm}} \mathrm{HR}(\widetilde{\boldsymbol{\gamma}}),$$

where $\geq_{\mathrm{sm}}$ denotes the *supermodular* order, cf. Müller and Stoyan (2002) Section 3.9 or Shaked and Shanthikumar (2007) Section 9.A.4. We have therefore included the respective arguments in the proof, too, although considering the supermodular order is otherwise beyond the scope of this work.

Proof. Set

$$\rho_{ij}^{(n)} = \exp(-\gamma_{ij}/(4\log(n)) \qquad \text{and} \qquad \widetilde{\rho}_{ij}^{(n)} = \exp(-\widetilde{\gamma}_{ij}/(4\log(n))$$

for $i, j \in \{1, \ldots, d\}$, $n \in \mathbb{N}$, so that $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \mathcal{G}_d$ ensures that the resulting matrices are correlation matrices, cf. e.g. Berg et al. (1984) Theorem 3.2.2. By construction, $\rho_{ij}^{(n)} \geq \widetilde{\rho}_{ij}^{(n)}$ for all i, j, n. And so the normal comparison lemma (Slepian, 1962), cf. also Tong (1980) Section 2.1. or Müller and Stoyan (2002) Example 3.8.6, implies that $\boldsymbol{Y} \geq_{\mathrm{PQD}} \widetilde{\boldsymbol{Y}}$ if $\boldsymbol{Y}$ and $\widetilde{\boldsymbol{Y}}$ are zero mean unit-variance Gaussian random vectors with correlations $\boldsymbol{\rho}$ and $\widetilde{\boldsymbol{\rho}}$, respectively. In fact, even

$\boldsymbol{Y} \geq_{\text{sm}} \widetilde{\boldsymbol{Y}}$ holds for the supermodular order (Müller and Stoyan, 2002, Theorem 3.13.5).

Consider the triangular arrays with independent $\boldsymbol{Y}_i^{(n)} \sim \boldsymbol{Y}$, $i = 1, \ldots, n$ and $\widetilde{\boldsymbol{Y}}_i^{(n)} \sim \widetilde{\boldsymbol{Y}}$, $i = 1, \ldots, n$. Since $t(1 - \exp(-a/t)) \to a$ as $t \to \infty$, Theorem 2.2.6 gives that

$$u_n(\boldsymbol{M}^{(n)} - u_n) = u_n \cdot \left(\max_{i=1,\ldots,n} (\boldsymbol{Y}_i^{(n)})_1 - u_n, \ldots, \max_{i=1,\ldots,n} (\boldsymbol{Y}_i^{(n)})_d - u_n \right)^\top$$

converges in distribution to $\mathrm{HR}(\boldsymbol{\gamma})$ and the corresponding tilde-version, while the closure under independent conjunction (Shaked and Shanthikumar (2007) Theorem 9.A.5) together with Shaked and Shanthikumar (2007) Theorem 9.A.4 implies $u_n(\boldsymbol{M}^{(n)} - u_n) \geq_{\text{PQD}} u_n(\widetilde{\boldsymbol{M}}^{(n)} - u_n)$ for all $n \in \mathbb{N}$. In fact, even $u_n(\boldsymbol{M}^{(n)} - u_n) \geq_{\text{sm}} u_n(\widetilde{\boldsymbol{M}}^{(n)} - u_n)$ for all $n \in \mathbb{N}$ as the supermodular order is also closed under independent conjuction (Müller and Stoyan, 2002, Theorem 3.9.14) and note Shaked and Shanthikumar (2007) Theorem 9.A.12. The assertion follows now from the closure of the PQD-order under distributional limits (Shaked and Shanthikumar (2007) Theorem 9.A.5). We even have $\mathrm{HR}(\boldsymbol{\gamma}) \geq_{\text{sm}} \mathrm{HR}(\widetilde{\boldsymbol{\gamma}})$, as the supermodular order satisfies the same closure property with respect to distributional limits (Müller and Stoyan, 2002, Theorem 3.9.12). □

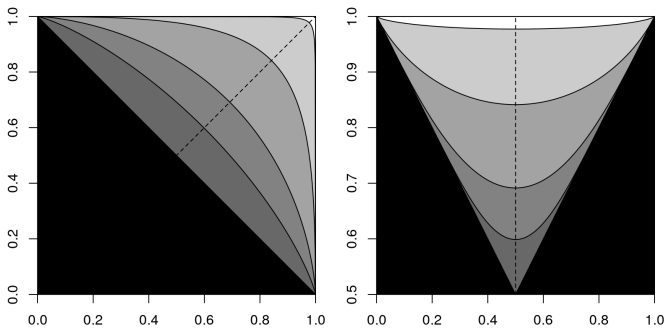

Figure 3.7: Nested max-zonoids and ordered Pickands dependence functions from the bivariate Hüsler-Reiß family for $\sqrt{\gamma} \in \{0.5, 1, 2, 4\}$. Larger values of γ correspond to larger sets and larger Pickands dependence functions and are closer to the independence model.

Remark 3.2.11. Theorem 3.2.9 includes the assumption that both parameter matrices $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ constitute a valid set of parameters for the Hüsler-Reiß model, i.e. they need to be elements of $\mathcal{G}_d$. In dimensions $d \geq 3$ it is possible that increasing (or decreasing) any of the parameters of a given valid $\boldsymbol{\gamma}$ will result in a set of parameters that is not valid for the Hüsler-Reiß model.

Remark 3.2.12. Since the orthant orders are closed under independent conjunction, Theorem 3.2.9 extends to the generalised Hüsler-Reiß model, where we can

allow some parameter values of $\boldsymbol{\gamma}$ to assume the value ∞, as long as $\boldsymbol{\gamma}$ remains negative definite in the extended sense (see Remark 2.2.7).

Example 3.2.13 (Ordering of $G/\widetilde{G}$ does not imply generator ordering for $\boldsymbol{Z}/\widetilde{\boldsymbol{Z}}$ – the case of Hüsler-Reiß log-Gaussian generators)**.** Consider the non-degenarate bivariate Hüsler-Reiß model with $\gamma_{12} = \gamma_{21} = \gamma \in (0, \infty)$ and let additionally $a \in [0, 1]$. Then the zero mean bivariate Gaussian model $(W_1, W_2)^\top$ with $\mathbb{E}(W_1) = \gamma a^2$, $\mathbb{E}(W_2) = \gamma(1-a)^2$, $\mathrm{Cov}(W_1, W_2) = 0.5\gamma \cdot (a^2 + (1-a)^2 - 1)$ satisfies $\mathbb{E}(W_1 - W_2)^2 = \gamma$, hence leads to a generator for the bivariate Hüsler-Reiß distribution in the sense of Theorem/Definition 2.2.4. WLOG $a \in (0, 1]$ (otherwise consider $1-a$ instead of a). Then $\log(Z_1)$ follows a non-degenerate univariate Gaussian distribution with mean $-0.5\gamma a^2$ and variance γa^2. The family of such distributions is not ordered in $\gamma > 0$ (cf. e.g. Shaked and Shanthikumar (2007) Example 1.A.26 or Müller and Stoyan (2002) Theorem 3.3.13). Hence, the bivariate family $(\log(Z_1), \log(Z_2))^\top$ can also not be ordered according to orthant order, nor can any multivariate family, for which this constitutes a marginal family. Accordingly, the corresponding log-Gaussian generators $\boldsymbol{Z}$ of the Hüsler-Reiß model will not be ordered, even when the resulting max-stable model and exponent measures are ordered as seen in Theorem 3.2.9.

While Dirichlet and Hüsler-Reiß families are ordered in the PQD/concordance sense according to the natural ordering of their parameter spaces, we would like to provide some examples that show that UO and LO ordering among simple max-stable distributions are in fact not equivalent.

To this end, we revisit the Choquet max-stable model from Section 2.2.3. We write $\mathrm{Choquet}_{\mathrm{EC}}(\theta)$ for the simple max-stable Choquet distribution if it is parameterised by its extremal coefficients $\theta(A)$, $A \subset \{1, \ldots, d\}$, $A \neq \emptyset$ and $\mathrm{Choquet}_{\mathrm{TD}}(\chi)$ if it is parameterised by its tail dependence coefficients $\chi(A)$, $A \subset \{1, \ldots, d\}$, $A \neq \emptyset$.

Lemma 3.2.14 (LO and UO order of Choquet family/Tawn-Molchanov model)**.** *Consider the family of max-stable Choquet models from Section 2.2.3. Then the LO order is characterised by the ordering of extremal coefficients, we have*

$$\theta \leq \widetilde{\theta} \qquad \Longleftrightarrow \qquad \mathrm{Choquet}_{\mathrm{EC}}(\theta) \leq_{\mathrm{lo}} \mathrm{Choquet}_{\mathrm{EC}}(\widetilde{\theta});$$

and the UO order is characterised by the ordering of tail dependence coefficients, that is

$$\chi \leq \widetilde{\chi} \qquad \Longleftrightarrow \qquad \mathrm{Choquet}_{\mathrm{TD}}(\chi) \leq_{\mathrm{uo}} \mathrm{Choquet}_{\mathrm{TD}}(\widetilde{\chi}).$$

Proof. The LO part is immediate from $\theta \leq \widetilde{\theta}$ implying the inclusion of associated max-zonoids $K^* \subset \widetilde{K}^*$ or Choquet integrals $\ell^* \leq \widetilde{\ell}^*$ (cf. Theorem 2.2.9) and then follows directly from Theorem 3.2.1 part a). For the UO part, note from the Proof of Corollary 3.2.4 that for $A \subset \{1, \dots, d\}$, $A \neq \emptyset$, $\boldsymbol{a} \in (0, \infty)^d$

$$\mathbb{E}\big(\min_{i \in A}(a_i Z_i^*)\big) = \min_{i \in A}(a_i) \cdot \chi(A) \qquad \text{and} \qquad \mathbb{E}\big(\min_{i \in A}(a_i \widetilde{Z}_i^*)\big) = \min_{i \in A}(a_i) \cdot \widetilde{\chi}(A)$$

if $\boldsymbol{Z}$ and $\widetilde{\boldsymbol{Z}}$ are generators of the respective models, hence the assertion with Theorem 3.2.1 part b). □

As we know already from Theorem 3.2.1 part c), in dimension $d = 2$, it is easily seen that $\chi \leq \widetilde{\chi}$ is equivalent to $\theta \geq \widetilde{\theta}$, alternatively recall $\theta_{12} + \chi_{12} = 2$. Starting from dimension $d = 3$, this is no longer the case and one can easily construct examples, where only LO or UO ordering holds.

Example 3.2.15. Table 3.1 lists valid parameter sets for four different trivariate Choquet/Tawn-Molchanov models. Among these, we can easily read off that

- $\boldsymbol{B} \leq_{\text{uo}} \boldsymbol{D}$, but there is no order between $\boldsymbol{B}$ and $\boldsymbol{D}$ according to lower orthants;
- $\boldsymbol{C} \leq_{\text{lo}} \boldsymbol{B}$, but there is no order between $\boldsymbol{B}$ and $\boldsymbol{C}$ according to upper orthants.

Of course, it is still possible that Choquet models are ordered according to PQD order, e.g.

- $\boldsymbol{A} \leq_{\text{PQD}} \boldsymbol{B}$.

It is also possible to have both LO and UO order in the same direction, e.g.

- $\boldsymbol{A} \leq_{\text{uo}} \boldsymbol{C}$ and $\boldsymbol{A} \leq_{\text{lo}} \boldsymbol{C}$.

However, note that such an order can only arise if the bivariate marginal distributions all agree.

Table 3.1: Valid parameter sets of four trivariate Choquet/Tawn-Molchanov models A,B,C,D, cf. Section 2.2.3. The models are exchangeable so that parameters τ_A, χ_A, θ_A depend on sets A only through their cardinality. Since $\chi_1 = \theta_1 = 1$ these parameters need not be listed. We have the relations $\tau_1 + 2\tau_{12} + \tau_{123} = 1$, $\chi_{123} = \tau_{123}$, $\chi_{12} = \chi_{123} + \tau_{12}$, $\theta_{12} = 1 + \tau_1 + \tau_{12}$ and $\theta_{123} = \theta_{12} + \tau_1$.

	τ_1	τ_{12}	τ_{123}	χ_{12}	χ_{123}	θ_{12}	θ_{123}
A	0.3	0.2	0.3	0.5	0.3	1.5	1.8
B	0.1	0.3	0.3	0.6	0.3	1.4	1.5
C	0.4	0.1	0.4	0.5	0.4	1.5	1.9
D	0.3	0.0	0.7	0.7	0.7	1.3	1.6

Chapter 4

Kernel embedding of measures

We start by recalling basic definitions and facts on reproducing kernel Hilbert spaces in Section 4.1, before explaining their role in the kernel mean embedding of measures in Section 4.2. This will allow us to define the maximum mean discrepancy (MMD), a computationally convenient surrogate for the Wasserstein distance to compare measures (on the sphere) in what follows. In particular, as we seek a sparse representation of the empirical spectral measure in the framework of regular variation later on (Chapter 5), we shall consider a regularised minimisation problem in terms of the MMD (4.3), for which fast solvers such as the vertex exchange algorithm (Section 4.3) exist.

4.1 Background on reproducing kernel Hilbert spaces

Let us recall some fundamental definitions and facts related to (reproducing Kernel) Hilbert spaces, mainly to fix notation. Many ideas go back as far as Aronszajn (1950); over the last decades a vast amount of textbook references have emerged, including Berlinet and Thomas-Agnan (2004), Paulsen and Raghupathi (2016) and Saitoh and Sawano (2016).

Definition 4.1.1. Let V be a vector space over $\mathbb{R}$. A map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called an *inner product* if it satisfies the following conditions for all vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ and all scalars $a, b \in \mathbb{R}$:

(a) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$;

(b) $\langle a\boldsymbol{x} + b\boldsymbol{y}, \boldsymbol{z} \rangle = a\langle \boldsymbol{x}, \boldsymbol{z} \rangle + b\langle \boldsymbol{y}, \boldsymbol{z} \rangle$;

(c) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, with equality if and only if $\boldsymbol{x} = \boldsymbol{0}$.

A vector space V over $\mathbb{R}$ together with an inner product $\langle \cdot, \cdot \rangle$ is called an *inner product space*.

If $(V, \langle \cdot, \cdot \rangle)$ is an inner product space, and we set

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}, \quad \boldsymbol{x} \in V,$$

then $\|\cdot\|$ is a *norm*. An inner product space is called a *Hilbert space* if it is *complete* with respect to its induced norm.

In what follows, we will consider Hilbert spaces $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$, where $\mathcal{H}$ is a set of functions on a compact set $\mathcal{X} \subset \mathbb{R}^d$ for some $d \in \mathbb{N}$ and simply abbreviate $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ by $\mathcal{H}$ for notational convenience. If $\mathcal{H}$ is such a Hilbert space, we shall denote the *evaluation functional* for evaluation at $\boldsymbol{x} \in \mathcal{X}$ by

$$F_{\boldsymbol{x}} : \mathcal{H} \to \mathbb{R}, \quad F_{\boldsymbol{x}}[f] = f(\boldsymbol{x}).$$

Definition 4.1.2. A Hilbert space of functions $\mathcal{H}$ is called a *reproducing kernel Hilbert space* (RKHS) if for every $\boldsymbol{x} \in \mathcal{X}$, the evaluation functional $F_{\boldsymbol{x}}$ is bounded, i.e. for every $\boldsymbol{x} \in \mathcal{X}$ there exists some $C > 0$ such that

$$|F_{\boldsymbol{x}}[f]| = |f(\boldsymbol{x})| \leq C\|f\|_{\mathcal{H}} \quad \text{for all } f \in \mathcal{H}.$$

The evaluation functional $F(\boldsymbol{x})$ is in the dual space of $\mathcal{H}$, hence by the Riesz representation theorem there exists a function $k_{\boldsymbol{x}} \in \mathcal{H}$ such that

$$F_{\boldsymbol{x}}[f] = f(\boldsymbol{x}) = \langle f, k_{\boldsymbol{x}} \rangle_{\mathcal{H}};$$

this is called the *reproducing property*.

Definition 4.1.3. The function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined by $k(\boldsymbol{x}, \boldsymbol{y}) = \langle k_{\boldsymbol{y}}, k_{\boldsymbol{x}} \rangle_{\mathcal{H}}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ is called the *reproducing kernel* of the RKHS $\mathcal{H}$.

An important property of the reproducing kernel k associated to a RKHS $\mathcal{H}$ is that it is a positive semi-definite function, as we will see in Lemma 4.1.5.

Definition 4.1.4. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is *symmetric positive semi-definite* if

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0, \tag{4.1}$$

for any $n \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathbb{R}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$. The function k is called *symmetric positive definite* if equality in (4.1) implies $a_1 = a_2 = \cdots = a_n = 0$ whenever the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are distinct.

Lemma 4.1.5. *The reproducing kernel k of a RKHS $\mathcal{H}$ is a positive semi-definite function.*

Proof. Using the definition of the reproducing kernel k of a RKHS $\mathcal{H}$, we have

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \langle k_{\boldsymbol{x}_j}, k_{\boldsymbol{x}_i}\rangle_{\mathcal{H}} \\ &= \left\langle \sum_{j=1}^{n} a_j k_{\boldsymbol{x}_j}, \sum_{j=1}^{n} a_j k_{\boldsymbol{x}_j}, \right\rangle_{\mathcal{H}} \\ &= \left\| \sum_{j=1}^{n} a_j k_{\boldsymbol{x}_j} \right\|_{\mathcal{H}}^2 \geq 0.\end{aligned}$$

If the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are distinct, equality holds only if $a_1 = a_2 = \cdots = a_n = 0$. Hence, the reproducing kernel k is positive semi-definite. □

It is well-known that each positive semi-definite function gives rise to a reproducing kernel Hilbert space, which is unique up to isomorphisms between Hilbert spaces.

Theorem 4.1.6. *(Moore-Aronszajn Theorem) Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric positive semi-definite function. There exists a unique reproducing kernel Hilbert space $\mathcal{H}$ with reproducing kernel k.*

Examples of positive definite kernels A plethora of kernel functions on $\mathbb{R}^d$ are widely known and intensively studied, cf. e.g. Buhmann (2000), Genton (2002) or Wendland (2005), for instance. One of the most common kernels used in machine learning, when the domain $\mathcal{X} = \mathbb{R}^d$, is the Gaussian kernel

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{2\sigma^2} \right), \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d,$$

with bandwidth parameter $\sigma > 0$, For this thesis, it is kernel functions on the unit sphere $\mathbb{S}_{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$ in $\mathbb{R}^d$ that are most relevant, cf. Chapter 5. Table 4.1 gives the analytic expression and parameter ranges for several such kernels. All of the kernel functions in Table 4.1 are *isotropic* on the sphere $\mathbb{S}_{d-1}$, that is, they can be expressed as functions of the *great circle distance*

$$\theta(\boldsymbol{x}, \boldsymbol{y}) = \arccos(\langle \boldsymbol{x}, \boldsymbol{y} \rangle), \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}_{d-1} \subset \mathbb{R}^d,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathbb{R}^d$. In principle, it is possible that such a function is positive definite only up to a certain dimension $d \in \mathbb{N}$, cf. Gneiting (2013) for several examples. The ones listed in Table 4.1 are valid positive definite kernels on the sphere $\mathbb{S}_{d-1}$ in any dimension $d \in \mathbb{N}$, cf. Gneiting (2013).

Table 4.1: Kernel functions on $\mathbb{S}_{d-1}$ in terms of the great circle distance $\theta \in [0, \pi]$. Here, b is a scale parameter, τ is a shape parameter and α and ν are smoothness parameters. The parameter ranges ensure in each case that they are valid kernel functions in any dimension $d \in \mathbb{N}$.

Family	Expression	Parameter ranges
Powered exponential	$\exp(-(\frac{\theta}{b})^{\alpha})$	$b > 0$, $\alpha \in (0, 1]$
Matérn	$\frac{2^{\nu-1}}{\Gamma(\nu)}(\frac{\theta}{b})^{\nu} K_{\nu}(\frac{\theta}{b})$	$b > 0$, $\nu \in (0, 1/2]$
Generalised Cauchy	$(1 + (\frac{\theta}{b})^{\alpha})^{-\tau/\alpha}$	$b > 0$, $\tau > 0$, $\alpha \in (0, 1]$
Dagum	$1 - ((\frac{\theta}{b})^{\tau}/(1 + (\frac{\theta}{b})^{\tau})^{\alpha/\tau}$	$b > 0$, $\tau \in (0, 1]$, $\alpha \in (0, \tau)$

4.2 Kernel mean embedding of measures

Consider a reproducing kernel Hilbert space $\mathcal{H}$ with kernel $k : \mathcal{X} \times \mathcal{X} \to \mathcal{H}$; by definition $k(\boldsymbol{x}, \boldsymbol{y}) = \langle k_{\boldsymbol{y}}, k_{\boldsymbol{x}} \rangle_{\mathcal{H}}$. We can then view the evaluation of the kernel as an inner product in $\mathcal{H}$ induced by the map $k_{\boldsymbol{y}} : \mathcal{X} \to \mathcal{H}$, $\boldsymbol{y} \mapsto k_{\boldsymbol{y}}$. The map $k_{\boldsymbol{y}}$ is called the *canonical feature map*. Classical kernel methods use this property to formulate algorithms that require calculating an inner product in $\mathcal{H}$ by evaluating the kernel function instead. This is called the *kernel trick* and is often computationally cheaper than calculating the mapping of the points into $\mathcal{H}$ explicitly. The embedding of $\mathcal{X}$ into the RKHS $\mathcal{H}$ can be seen more generally as the embedding of the Dirac measure

$$\delta_{\boldsymbol{y}} = \begin{cases} 0, & \boldsymbol{y} \notin A \\ 1, & \boldsymbol{y} \in A \end{cases}$$

where A is a measurable subset of $\mathcal{X}$. This can then be further generalised using the kernel mean embedding, where a signed finite measure is mapped to an element of a reproducing kernel Hilbert space (RKHS). For a general review, see Muandet et al. (2017); other related references are Sriperumbudur et al. (2011), Simon-Gabriel and Schölkopf (2018) and Simon-Gabriel et al. (2023).

Definition 4.2.1. Let $\mathcal{M}$ be the space of signed finite measures on a measurable space $\mathcal{X}$ and let $|\mu|$ be the total variation measure of μ. The *kernel mean embedding* of a measure μ into a RKHS $\mathcal{H}$ with reproducing kernel k is defined as the map $K : \mathcal{M}_k \to \mathcal{H}$,

$$K(\mu) = \int_{\mathcal{X}} k_{\boldsymbol{x}} \, \mathrm{d}\mu(\boldsymbol{x}), \tag{4.2}$$

where $\mathcal{M}_k = \{\mu \in \mathcal{M} : \int_{\mathcal{X}} \sqrt{k(\boldsymbol{x}, \boldsymbol{x})} \, \mathrm{d}|\mu|(\boldsymbol{x}) < \infty\}$ and the integral should be interpreted as a Bochner integral.

Definition 4.2.2. A kernel function k is called *characteristic* if the map $K : \mathcal{M}_k \to \mathcal{H}$, $\mu \mapsto K(\mu)$ is injective.

The kernel mean embedding can be used to define a metric, called the *maximum mean discrepancy*, as follows.

Definition 4.2.3. The *maximum mean discrepancy* of two measures μ and ν from $\mathcal{M}_k$ is given by

$$\mathrm{MMD}[\mathcal{H}, \mu, \nu] = \sup_{\|f\|_{\mathcal{H}} \leq 1} \left\{ \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) - \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}\nu(\boldsymbol{x}) \right\}, \tag{4.3}$$

where $\mathcal{H}$ is a reproducing kernel Hilbert space.

The maximum mean discrepancy is simply a particular case of an *integral probability metric*

$$\sup_{f \in \mathcal{F}} \left\{ \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) - \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}\nu(\boldsymbol{x}) \right\},$$

where we choose $\mathcal{F}$ to be the set of functions that belong to the unit ball in a RKHS $\mathcal{H}$. Other choices of $\mathcal{F}$ give rise to different distance measures. For example, setting $\mathcal{F} = \{f : \|f\|_\infty \leq 1\}$, where $\|f\|_\infty = \sup_{\boldsymbol{x} \in \mathcal{X}} |f(\boldsymbol{x})|$, gives rise to the total variation distance. Another example is the choice $\mathcal{F} = \{f : \|f\|_L \leq 1\}$, where $\|f\|_L = \sup_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}} \{|f(\boldsymbol{x}) - f(\boldsymbol{y})| / \rho(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{x} \neq \boldsymbol{y}\}$ is the Lipschitz seminorm of a real-valued function f for some metric ρ on $\mathcal{X}$. This choice yields the Wasserstein distance. Further examples can be found in Müller (1997).

The following proposition expresses the maximum mean discrepancy between two measures in terms of their respective kernel mean embeddings.

Proposition 4.2.4. *The maximum mean discrepancy between two measures μ and ν from $\mathcal{M}_k$ can be expressed as the distance between the corresponding mean embeddings in the reproducing kernel Hilbert space $\mathcal{H}$, i.e.*

$$\mathrm{MMD}[\mathcal{H}, \mu, \nu] = \|K(\mu) - K(\nu)\|_{\mathcal{H}}. \tag{4.4}$$

Proof. Let $f \in \mathcal{H}$. Using the reproducing property of $\mathcal{H}$, we have

$$\begin{aligned} \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) &= \int_{\mathcal{X}} \langle f, k_{\boldsymbol{x}} \rangle_{\mathcal{H}} \, \mathrm{d}\mu(\boldsymbol{x}) \\ &= \left\langle f, \int_{\mathcal{X}} k_{\boldsymbol{x}} \, \mathrm{d}\mu(\boldsymbol{x}) \right\rangle_{\mathcal{H}} \\ &= \langle f, K(\mu) \rangle_{\mathcal{H}}. \end{aligned}$$

Then for two measures μ and ν, we have

$$\begin{aligned} &\left| \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) - \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}\nu(\boldsymbol{x}) \right| \\ &= |\langle f, K(\mu) - K(\nu) \rangle_{\mathcal{H}}| \leq \|f\|_{\mathcal{H}} \|K(\mu) - K(\nu)\|_{\mathcal{H}}, \end{aligned}$$

where the last step follows from the Cauchy-Schwarz inequality. Hence,

$$\sup_{\|f\|_{\mathcal{H}}\leq 1}\left\{\int_{\mathcal{X}} f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})-\int_{\mathcal{X}} f(\boldsymbol{x})\,\mathrm{d}\nu(\boldsymbol{x})\right\}\leq \|K(\mu)-K(\nu)\|_{\mathcal{H}}.$$

On the other hand the upper bound $\|K(\mu)-K(\nu)\|_{\mathcal{H}}$ is attained as a supremum by

$$f=\frac{K(\mu)-K(\nu)}{\|K(\mu)-K(\nu)\|_{\mathcal{H}}},$$

and hence we can rewrite $\mathrm{MMD}[\mathcal{H},\mu,\nu]=\|K(\mu)-K(\nu)\|_{\mathcal{H}}$. □

Using a characteristic kernel k ensures that $\|K(\mu)-K(\nu)\|_{\mathcal{H}}=0$ if and only if $\mu=\nu$, i.e. no information about the measures μ and ν is lost in their mappings in the RKHS $\mathcal{H}$. Steinwart and Ziegel (2021) show that all the kernels in Table 4.1 are characteristic.

Given a kernel k and a set of sample locations $\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$, the matrix $\boldsymbol{K}$ with entries $\boldsymbol{K}_{ij}=k(\boldsymbol{x}_i,\boldsymbol{x}_j)$ is called the *kernel matrix*. We can rewrite the maximum mean discrepancy in terms of the kernel matrix as follows. First, note that

$$\mathrm{MMD}^2[\mathcal{H},\mu,\nu]=\|K(\mu)\|_{\mathcal{H}}^2+\|K(\nu)\|_{\mathcal{H}}^2-2\langle K(\mu),K(\nu)\rangle_{\mathcal{H}}. \tag{4.5}$$

Now consider two discrete measures $\mu=\sum_{i=1}^N \omega_i\delta_{\boldsymbol{x}_i}$ and $\nu=\sum_{j=1}^N \upsilon_j\delta_{\boldsymbol{x}_j}$. Then we can rewrite

$$\langle K(\mu),K(\nu)\rangle_{\mathcal{H}}=\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)\omega_i\upsilon_j=\boldsymbol{\omega}^T\boldsymbol{K}\boldsymbol{\upsilon};$$

this gives

$$\mathrm{MMD}^2[\mathcal{H},\mu,\nu]=(\boldsymbol{\omega}-\boldsymbol{\upsilon})^T\boldsymbol{K}(\boldsymbol{\omega}-\boldsymbol{\upsilon}).$$

In what follows, we will look at minimising (a regularised version of) the squared maximum mean discrepancy between two discrete measures μ and ν. A detailed study of the following minimisation problems and methods to solve them is given in Gauthier and Suykens (2018).

Squared maximum mean discrepancy minimisation problem Let $\mu=\sum_{i=1}^N \omega_i\delta_{\boldsymbol{x}_i}$ be fixed, i.e. $\boldsymbol{\omega}=(\omega_i)_{i=1,\ldots,N}$ with weights $\omega_i>0$ is fixed, then we define

$$D(\boldsymbol{\upsilon})=\frac{1}{2}(\boldsymbol{\omega}-\boldsymbol{\upsilon})^T\boldsymbol{K}(\boldsymbol{\omega}-\boldsymbol{\upsilon}), \tag{4.6}$$

for $\boldsymbol{\upsilon}\in\mathbb{R}^N$ and where the scalar $1/2$ is added for simplification purposes. Obviously, when minimising $D(\boldsymbol{\upsilon})$ and $\boldsymbol{K}$ stems from a characteristic kernel and the

support points $\boldsymbol{x}_i$, $i = 1, \ldots, N$, are mutually distinct, then only $\boldsymbol{v} = \boldsymbol{\omega}$ minimises (4.6). Since the main purpose of our application is to find a *sparse* approximation of the measure μ represented by the vector $\boldsymbol{\omega}$, we add a regularisation term which will control the sparsity of the solution.

Regularised minimisation problem For a given penalisation direction $\boldsymbol{d} \in \mathbb{R}^n$ with components $d_i > 0$, for $\alpha > 0$ and $\boldsymbol{v} \in \mathbb{R}^N$, we introduce the regularised minimisation problem

$$\underset{\boldsymbol{v}}{\text{minimise}}\ D_\alpha(\boldsymbol{v}) = \frac{1}{2}(\boldsymbol{\omega} - \boldsymbol{v})^T \boldsymbol{K}(\boldsymbol{\omega} - \boldsymbol{v}) + \alpha \boldsymbol{d}^T \boldsymbol{v} \ \text{ subject to } \ \boldsymbol{v} \geq 0. \tag{4.7}$$

Here, the parameter $\alpha \geq 0$ controls the sparsity of the solution to the minimisation problem (4.7). The smaller α is, the less sparse the solution will be. It is possible to compute values $\alpha_{\text{crit}} = \alpha_0 > \alpha_1 > \cdots > \alpha_{\text{final}} = 0$, where the support of the solution of (4.7) changes; this is called the regularisation path method, see e.g. Gauthier and Suykens (2018). However, this method is computationally expensive and hence for large values of N becomes computationally prohibitive. Therefore, we introduce yet another minimisation problem, which is equivalent to (4.7).

Constrained minimisation problem For $\kappa > 0$ and for $\boldsymbol{v} \in \mathbb{R}^N$, we introduce the constrained minimisation problem

$$\underset{\boldsymbol{v}}{\text{minimise}}\ D(\boldsymbol{v}) = \frac{1}{2}(\boldsymbol{\omega} - \boldsymbol{v})^T \boldsymbol{K}(\boldsymbol{\omega} - \boldsymbol{v}) \ \text{ subject to } \ \boldsymbol{v} \geq 0 \ \text{ and } \ \boldsymbol{d}^T \boldsymbol{v} = \kappa. \tag{4.8}$$

In this case, the parameter $\kappa > 0$ controls the sparsity of the solution. The smaller κ is, the sparser the solution will be. In Section 4.3, we describe an algorithm to solve this constrained minimisation problem.

As mentioned, the two minimisation problems (4.7) and (4.8) are equivalent, and one can easily recover α from κ (and viceversa) using the following proposition from Gauthier and Suykens (2018).

Proposition 4.2.5. *Let $\boldsymbol{v}_\alpha$ be a solution to problem (4.7) for a fixed $\alpha > 0$; then $\boldsymbol{v}_\alpha$ is also a solution to the constrained problem with $\kappa = \boldsymbol{d}^T \boldsymbol{v}_\alpha$. Conversely, let $\boldsymbol{v}_\kappa$ be a solution to problem (4.8) for a fixed $\kappa > 0$; then $\boldsymbol{v}_\kappa$ is also a solution to the regularised problem with $\alpha = \boldsymbol{v}_\kappa^T \boldsymbol{K}(\boldsymbol{\omega} - \boldsymbol{v}_\kappa)/\kappa$.*

This also implies that if κ in (4.8) exceeds or is equal to $\boldsymbol{d}^T \boldsymbol{\omega}$, the solution to (4.8) is going to be equal to $\boldsymbol{\omega}$ itself. Therefore, we shall assume here and hereinafter that κ ranges between 0 and $\boldsymbol{d}^T \boldsymbol{\omega}$. If we work with probability measures only, and $d_i = 1$ for all $i = 1, \ldots, N$, this amounts to the upper bound $\boldsymbol{d}^T \boldsymbol{\omega} = 1$ for κ.

4.3 Vertex exchange algorithm

In this section, we cover an algorithm, called the *vertex exchange algorithm*, which can be used to solve the constrained minimisation problem (4.8). For more details on the algorithm, see Böhning (1986), Pronzato and Pázman (2013) or Gauthier and Suykens (2018), for instance. For convenience, we abbreviate the index set $[N] = \{1, \dots, N\}$. We use $\odot$ to denote the *Hadamard product* between matrices (i.e. their pointwise multiplication).

To begin with, let us clarify why we use the term vertex to denote observations. Recall that we denote the simplex of probability vectors in $\mathbb{R}^N$ by

$$\triangle_N = \{\boldsymbol{p} \in [0,1]^N \, : \, p_1 + \cdots + p_N = 1\}.$$

For each *observation* $\boldsymbol{x}$ in the data matrix $\boldsymbol{X} \in \mathbb{R}^{N \times d}$, the point measure $\delta_{\boldsymbol{x}}$ is a *vertex* in the convex set of probability measures

$$\left\{ \sum_{i=1,\dots,N} \upsilon_i \delta_{\boldsymbol{x}_i} : \boldsymbol{\upsilon} \in \triangle_N \right\}$$

spanned by the point measures $\{\delta_{\boldsymbol{x}_i}\}_{i=1,\dots,N}$. This is illustrated in Figure 4.1.

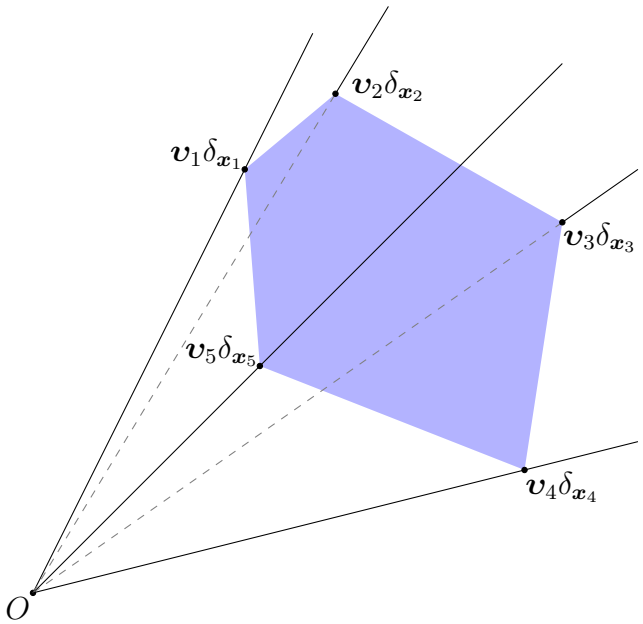

Figure 4.1: The cone of measures spanned by the vertices $\upsilon_i \delta_{\boldsymbol{x}_i}$, $i = 1, \dots, N$, for $N = 5$. The blue shaded area represents the constraint $\boldsymbol{d}^T \boldsymbol{\upsilon} = \kappa$.

In what follows, we give an explanation of how the vertex exchange algorithm works, and provide some pseudocode in Algorithm 1. The key idea is to iteratively move mass along a direction of pairs of support points (vertices) until the target

function does not change anymore. To be more precise, recall that the algorithm is used to solve the constrained minimisation problem

$$\underset{\boldsymbol{\upsilon}}{\text{minimise}}\ D(\boldsymbol{\upsilon}) = \frac{1}{2}(\boldsymbol{\omega}-\boldsymbol{\upsilon})^T\boldsymbol{K}(\boldsymbol{\omega}-\boldsymbol{\upsilon}) \ \text{ subject to } \ \boldsymbol{\upsilon} \geq 0 \ \text{ and } \ \boldsymbol{d}^T\boldsymbol{\upsilon} = \kappa,$$

where $0 < \kappa \leq \boldsymbol{d}^T\boldsymbol{\omega}$. Using standard matrix calculus, it can be shown that the gradient of the quantity $D(\boldsymbol{\upsilon})$ is given by

$$\nabla D(\boldsymbol{\upsilon}) = \boldsymbol{K}\boldsymbol{\upsilon} - \boldsymbol{K}\boldsymbol{\omega}.$$

We initialise the algorithm by finding the solution to (4.8) with the added constraint $\upsilon_i = \kappa/d_i$ and $\boldsymbol{\upsilon}_j = 0$ for all $j \neq i$. The single best vertex (index) b which minimises this problem can easily be calculated using

$$b = \underset{j=1,\ldots,N}{\arg\min}[\text{diag}(\boldsymbol{K}) \odot \boldsymbol{\xi}^2 - 2\boldsymbol{K}\boldsymbol{\omega} \odot \boldsymbol{\xi}]_j, \tag{4.9}$$

where $\boldsymbol{\xi} = \kappa\boldsymbol{d}^{-1}$. The algorithm iteratively selects two vertices $u \in [N]$, "up", and $d \in [N]$, "down", as follows:

$$u = \underset{i=1,\ldots,N}{\arg\min}[\nabla D(\boldsymbol{\upsilon}) \odot \boldsymbol{\xi}]_i, \quad d = \underset{i\in S_{\boldsymbol{\upsilon}}}{\arg\max}[\nabla D(\boldsymbol{\upsilon}) \odot \boldsymbol{\xi}]_i, \tag{4.10}$$

where $S_{\boldsymbol{\upsilon}}$ is the set of support positions of $\boldsymbol{\upsilon}$.

Note that in the first iteration, the vertex d must be equal to b, since $S_{\boldsymbol{\upsilon}} = \{b\}$. Weight is then transferred from the d-th to the u-th component of $\boldsymbol{\upsilon}$ as follows. Let $\boldsymbol{e}_i \in \mathbb{R}^N$ be the i-th element of the standard basis of $\mathbb{R}^N$, i.e. $[\boldsymbol{e}_i]_i = 1$ and all other components equal to zero, and let $\boldsymbol{\delta} = \xi_u\boldsymbol{e}_u - \xi_d\boldsymbol{e}_d$. The optimal step size is given by

$$r = \min\left\{\left[\frac{\boldsymbol{\upsilon}}{\boldsymbol{\xi}}\right]_d, -\frac{\boldsymbol{\delta}^T\nabla(\boldsymbol{\upsilon})}{\boldsymbol{\delta}^T\boldsymbol{K}\boldsymbol{\delta}}\right\}. \tag{4.11}$$

We then update our solution $\boldsymbol{\upsilon}$ and the gradient $\nabla D(\boldsymbol{\upsilon})$ using the optimal stepsize r and repeat the steps until the maximum number of iterations is reached. Whilst each iteration is computationally inexpensive, with a complexity of $\mathcal{O}(N)$, we also implement the following stopping criterion, which allows us to perform fewer iterations and still obtain a good approximation of the solution to (4.8) for a given κ; at each iteration we calculate the value of $D(\boldsymbol{\upsilon})$ and check the difference to its previous value. If this is smaller than a pre-set tolerance level, we stop the algorithm. At this stage, the algorithm is close to converging and at each subsequent iteration, $D(\boldsymbol{\upsilon})$ endures minimal change, i.e. $\boldsymbol{\upsilon}$ approximates the true solution to a satisfying degree.

Algorithm 1 Vertex exchange

Input:

data matrix $\boldsymbol{X}$ of N observations $\{\boldsymbol{x}_i\}_{i=1,\dots,N}$ in Ω; positive definite kernel $k : \Omega \times \Omega \to \mathbb{R}$; vector of probability weights $\boldsymbol{\omega} \in \mathbb{R}^N$; penalisation direction $\boldsymbol{d} \in \mathbb{R}^N$; sparsity tuning parameter $\kappa \in (0, \boldsymbol{d}^T\boldsymbol{\omega}]$; tolerance level for stopping criterion $\epsilon > 0$; maximal number of iterations $I \in \mathbb{N}$;

Preliminary:

compute kernel matrix $\boldsymbol{K} = \{k(\boldsymbol{x}_i, \boldsymbol{x}_j)\}_{i,j=1,\dots,N}$ and vectors $\boldsymbol{\xi} = \kappa\boldsymbol{d}^{-1}$ and $\boldsymbol{g} = \boldsymbol{K}\boldsymbol{\omega}$;

Initialisation:

find index $b \in [N]$ of single best vertex using (4.9); initialise non-negative weight vector $\boldsymbol{\upsilon}^{(1)}$ with one support position at b and weight $\upsilon_b^{(1)} = \xi_b$; set $S_{\boldsymbol{\upsilon}^{(1)}} = \{b\}$; compute initial gradient $\nabla D(\boldsymbol{\upsilon}^{(1)}) = \boldsymbol{K}\boldsymbol{\upsilon}^{(1)} - \boldsymbol{g}$; initialise vector $\boldsymbol{\rho}$ of length I and compute $\boldsymbol{\rho}_1 = (\boldsymbol{\omega} - \boldsymbol{\upsilon}^{(1)})^T\boldsymbol{K}(\boldsymbol{\omega} - \boldsymbol{\upsilon}^{(1)})$;

for $i = 2$ to I **do**
  compute $u \in [N]$ and $d \in [N]$ using (4.10);
  determine optimal step size r from (4.11) using $\boldsymbol{\upsilon} = \boldsymbol{\upsilon}^{(i)}$;
  set $\boldsymbol{\upsilon}^{(i+1)} = \boldsymbol{\upsilon}^{(i)} + r\boldsymbol{\delta}$, and $S_{\boldsymbol{\upsilon}^{(i+1)}} = S_{\boldsymbol{\upsilon}^{(i)}} \cup \{u\}$;
  compute new gradient $\nabla D(\boldsymbol{\upsilon}^{(i+1)}) = \nabla D(\boldsymbol{\upsilon}^{(i)}) + r\boldsymbol{K}\boldsymbol{\delta}$;
  compute $\boldsymbol{\rho}_{(i)} = (\boldsymbol{\omega} - \boldsymbol{\upsilon}^{(i)})\boldsymbol{K}(\boldsymbol{\omega} - \boldsymbol{\upsilon}^{(i)})$
  if $(\boldsymbol{\rho}_{(i-1)} - \boldsymbol{\rho}_{(i)})/\boldsymbol{\omega}^T\boldsymbol{g} < \epsilon$ **then**
    break
  end if
end for

Output:

vector of non-negative weights $\boldsymbol{\upsilon} \in \mathbb{R}^N$; support positions $S_{\boldsymbol{\upsilon}} \subseteq [N]$ of $\boldsymbol{\upsilon}$

Outlook In our practical implementation (Section 5.2 below), we have typically started with equal probability weights $\omega_i = 1/N$ from the empirical spectral measure, penalisation direction $\boldsymbol{d} = \boldsymbol{1}$ ($d_i = 1$, $i = 1, \dots, N$), leading to the bounds $0 < \kappa \leq 1$. While the vertex exchange algorithm's search takes place in the cone spanned by the Dirac-measures at the support points (cf. Figure 4.1), we shall interpret its output measure ν subsequently as a probability measure after renormalisation of its weight vector $\boldsymbol{\upsilon}$.

Chapter 5

Sparse representations of extremal dependence in high dimensions for large data sets

The lack of meaningful sparsity notions and sparsity enforcement in multivariate extreme value analysis has triggered a range of novel approaches to deal with the implied shortcomings when analysing dependence among multivariate extremes, in particular in high-dimensional settings and when dealing with very large data sets. A first review of recent developments has been given in Engelke and Ivanovs (2021). These include identifying clusters in regular variation (Chautru, 2015; Janßen and Wan, 2020; Fomichov and Ivanovs, 2023; Avella-Medina et al., 2024), identifying groups of variables, which may take large values jointly (Goix et al., 2017; Meyer and Wintenberger, 2023) or incorporating notions of conditional independence and defining directed or undirected graphs in multivariate extremes (Gissibl and Klüppelberg, 2018; Engelke and Hitz, 2020). Cooley and Thibaud (2019) introduce a PCA-type approach to decompose an extreme signal based on an eigen-decomposition of a tail dependence matrix, used by Rohrbeck and Cooley (2023) to simulate hazard event sets for river flows, whereas Drees and Sabourin (2021) prove consistency for a PCA-methodology to identify a linear subspace, in which regular variation takes place. Quite recently, Avella-Medina et al. (2022) propose kernel PCA in order to analyse the extremal dependence structures, especially among linear factor models, where subsequently preimages of kernel PCA are used for clustering extremes. To the best of our knowledge, this is the first time RKHS-techniques have been introduced in multivariate extremes in order to analyse dependence structures.

Here, we also incorporate RKHS techniques into the analysis of dependence structures in a regular variation framework. However, our notions of sparsity and accuracy for a spectral measure are fundamentally different from Avella-Medina et al. (2022): given data from a regularly varying random vector, we aim

to estimate a sparse and as accurate as possible representation of its spectral measure, where we measure sparsity through the number of its support points and assess accuracy through the Wasserstein distance.

More precisely, we consider the following regular variation framework, cf. also Section 2.3: let $\boldsymbol{X}$ be a random vector in $\mathbb{R}^d$, which is multivariate regularly varying, that is, (2.16) is satisfied, i.e.

$$u\,\mathbb{P}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in B,\, \|\boldsymbol{X}\| > ra(u)\right) \longrightarrow r^{-\alpha}\sigma(B) \quad \text{as} \quad u \to \infty, \tag{5.1}$$

where $\|\cdot\| = \|\cdot\|_2$ is the Euclidean norm, a is an auxiliary function tending to ∞, and (5.1) holds for any $r > 0$ and Borel set $B \subset \mathbb{S}_{d-1}$. Here, α is the index of regular variation, and the measure σ is the spectral (or angular) measure of $\boldsymbol{X}$, which is the object that contains all the extremal dependence information about $\boldsymbol{X}$.

If we have data in the form of independent copies $\boldsymbol{X}_i$, $i = 1, \ldots, n$, of $\boldsymbol{X}$, a simple canonical estimator for σ is the *empirical spectral/angular measure*:

$$\widehat{\sigma}_{n,k} = \sum_{i=1}^{k} \frac{1}{k} \delta_{\boldsymbol{X}_{(i)}/\|\boldsymbol{X}_{(i)}\|}, \tag{5.2}$$

where $\boldsymbol{X}_{(i)}$, $i = 1, \ldots, k$, are the k largest vectors among $\boldsymbol{X}_i$, $i = 1, \ldots, n$, where we mean "largest" according to the Euclidean norm. The empirical angular measure is known to be a consistent estimator for σ for an *intermediate* sequence $k = k_n$, i.e. $k_n \to \infty$, but $k_n/n \to 0$ (Einmahl and Segers, 2009). On the other hand, for very large data sets $\widehat{\sigma}_{n,k}$ will not be a sparse representation of σ. Since it is always based on k support points, a lot of information needs to be saved in order to represent the dependence information σ, and exploring such information, especially in high dimensions, will not be very insightful.

This is where our new approach for estimating σ comes in. We seek to obtain a representation of σ from $\widehat{\sigma}_{n,k}$ using much fewer support points that contain essentially (almost) the same information and where the proximity to the actual spectral measure is not too heavily compromised (or even better). In other words, we seek to compress the information of $\widehat{\sigma}_{n,k}$ relatively fast, where we measure the loss of information in terms of Wasserstein distance to the true spectral measure.

In order to do so, we borrow ideas from the theory of kernel mean embedding and maximum mean discrepancy minimisation and apply them to this context, cf. Chapter 4. More precisely, we apply the vertex exchange algorithm (Algorithm 1) to the empirical angular measure to obtain a much sparser representation of σ.

In order to demonstrate the feasibility and value of such an approach, we first need to refine the regular variation framework, which is still very general, to have

a large class of models, where we can test such an approach in a meaningful way. This is done in Section 5.1. We introduce a stochastic model for our numerical experiments and show that it falls under the multivariate regular variation framework. Subsequently, we apply the vertex exchange algorithm to a broad range of settings arising from the model of Section 5.1. Such experiments and numerical outcomes are documented in Section 5.2.

5.1 Stochastic model for numerical experiments

In order to explore the capability of the vertex exchange algorithm to compress effectively information from the empirical angular measure, we need a reasonable framework to test it in situations, where (a) we know the true spectral measure, but (b) have also a sufficient amount of noise and variation in the way the model is exhibiting regular variation. For instance, we refrain from testing our model in a spectrally discrete max-stable model as in Janßen and Wan (2020), as such a model produces artificially many ties among its components, which are very unlikely to be seen in real data.

Instead, we introduce a refined model, from which it is easy to simulate; in this section we demonstrate that it is indeed regularly varying.

Theoretical model for numerical experiments Let $\mathbb{S}_{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$ be the unit sphere in $\mathbb{R}^d$ and σ_0 and σ_1 probability measures on $\mathbb{S}_{d-1}$. Then we consider the d-dimensional random vector

$$\boldsymbol{X} = R_0\boldsymbol{S}_0 + R_1\boldsymbol{S}_1, \tag{5.3}$$

where $\boldsymbol{S}_0$ and $\boldsymbol{S}_1$ are drawn from σ_0 and σ_1, respectively. A conceptual figure of model (5.3) is given below in Figure 5.1.

The first direction $\boldsymbol{S}_0$ is multiplied with a positive scalar R_0 with distribution function F_{R_0} such that $\overline{F}_{R_0} = 1 - F_{R_0}$ is regularly varying with index $\alpha > 0$. This implies that, with

$$a(u) = F_{R_0}^{-1}\left(1 - \frac{1}{u}\right) = \inf\left\{t \geq 0 : \overline{F}_{R_0}(t) \leq \frac{1}{u}\right\}, \tag{5.4}$$

the following holds:

$$u\,\mathbb{P}(R_0 > ra(u)) \longrightarrow r^{-\alpha} \quad \text{as} \quad u \to \infty$$

for all $r > 0$, cf. Section 2.3.

The second direction $\boldsymbol{S}_1$ is multiplied with $R_1 = h(R_0)$, where h is a continuous function such $0 \leq h(x) \leq x$, for every x, $h(x) = o(x)$, and $g(x) = x + h(x)$ is

monotonously increasing to ∞. As we shall see below, these conditions on h ensure that the second summand $R_1\boldsymbol{S}_1$ assumes the role of a perturbation term. In fact, we have the following result, which shows that $\boldsymbol{X}$ is a regularly varying random vector, and the features of the regular variation are fully determined by the regularly varying random vector $R_0\boldsymbol{S}_0$. We write $f \sim g$ if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

Theorem 5.1.1. *Let $\boldsymbol{X}$ be the random vector from (5.3). Then*

(a) $\mathbb{P}(\|X\| > x) \sim \mathbb{P}(R_0 > x)$ *as* $x \to \infty$*;*
in particular, $\|\boldsymbol{X}\|$ is regularly varying with the same tail index as R_0 and

$$u\,\mathbb{P}(\|\boldsymbol{X}\| > ra(u)) \longrightarrow r^{-\alpha} \quad \text{as} \quad u \to \infty,$$

for $a(u)$ as in (5.4);

(b) the weak convergence

$$\mathcal{L}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \,\middle|\, \|\boldsymbol{X}\| > t\right) \longrightarrow \mathcal{L}(\boldsymbol{S}_0) = \sigma_0 \quad \text{as} \quad t \to \infty$$

holds true;

(c) $\boldsymbol{X}$ is multivariate regularly varying in the sense that for any measureable $B \subset \mathbb{S}_{d-1}$, for which $\sigma_0(\partial B) = 0$, and any $r > 0$

$$u\,\mathbb{P}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in B,\, \|\boldsymbol{X}\| > ra(u)\right) \longrightarrow r^{-\alpha}\sigma_0(B) \quad \text{as} \quad u \to \infty.$$

Note that (c) is immediate from (a) and (b), since

$$\begin{aligned}
&u\,\mathbb{P}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in B,\, \|\boldsymbol{X}\| > ra(u)\right) \\
&= u\,\mathbb{P}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in B \,\middle|\, \|\boldsymbol{X}\| > ra(u)\right) \mathbb{P}(\|\boldsymbol{X}\| > ra(u)) \\
&= u\,\mathbb{P}(\|\boldsymbol{X}\| > ra(u))\,\mathbb{P}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in B \,\middle|\, \|\boldsymbol{X}\| > ra(u)\right).
\end{aligned}$$

The proof of parts (a) and (b) of Theorem 5.1.1 is given in the next subsection.

5.1.1 Proof of Theorem 5.1.1

To begin with, we will need a few technical lemmas.

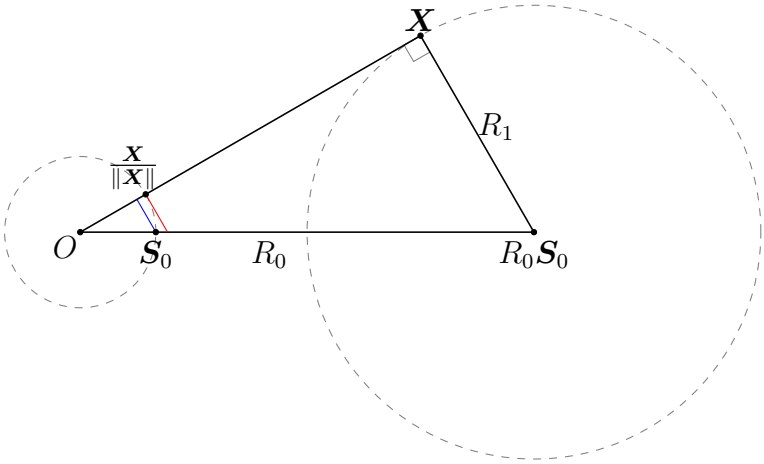

Figure 5.1: Illustration of the theoretical model (5.3), that is, $\boldsymbol{X} = R_0\boldsymbol{S}_0 + R_1\boldsymbol{S}_1$. Note that $\boldsymbol{S}_1 = (\boldsymbol{X} - R_0\boldsymbol{S}_0)/R_1$ is only implicitly depicted via the direction that $\boldsymbol{X}$ takes starting from $R_0\boldsymbol{S}_0$. The blue and red lines have been added to illustrate part of the proof of Proposition 5.1.10 below.

Definition 5.1.2 (Beirlant et al. (2004), Proposition 2.5)**.** Let ℓ be a slowly varying function, then any slowly varying function ℓ^* satisfying

$$\ell(x)\ell^*(x\ell(x)) \to 1, \quad x \to \infty$$

is called *de Bruijn conjugate* of ℓ.

Proposition 5.1.3 (Beirlant et al. (2004), Proposition 2.5)**.** *For any slowly varying function ℓ, a de Bruijn conjugate ℓ^* exists. It is asymptotically unique in the following sense: let $\bar{\ell}$ be another de Bruijn conjugate for ℓ, then $\ell^* \sim \bar{\ell}$. Moreover, $(\ell^*)^* \sim \ell$.*

Lemma 5.1.4. *Let $g : [a, \infty) \to [b, \infty)$ be an unbounded continuous and strictly increasing function from the interval $[a, \infty)$ to $[b, \infty)$ for some $a, b > 0$, given by*

$$g(x) = x \pm h(x),$$

where $h(x) = o(x)$.

(a) Then the inverse function $g^{-1} : [b, \infty) \to [a, \infty)$ can be expressed as

$$g^{-1}(x) = x \mp h^*(x)$$

for $h^*(x) = o(x)$.

(b) If, in addition, $h \geq 0$, then $h^* \geq 0$.

(c) If, in addition, $h > 0$, then $h^* > 0$.

Proof. (a) We rewrite g as

$$g(x) = x\left(1 \pm \frac{h(x)}{x}\right) = x\ell(x),$$

and observe that the thereby defined function ℓ is slowly varying with $\lim_{x\to\infty} \ell(x) = 1$. Let ℓ^* be the de Bruijn conjugate of ℓ. Then

$$\ell(x)\ell^*(x\ell(x)) \to 1 \quad \text{as} \quad x \to \infty. \tag{5.5}$$

Hence $\ell^*(y) \to 1$ as $y \to \infty$. Moreover,

$$\lim_{y\to\infty} \frac{y}{g^{-1}(y)}\ell^*(y) = \lim_{x\to\infty} \frac{g(x)}{x}\ell^*(g(x)) = 1,$$

where the last equality follows from rewriting (5.5) in terms of g. Therefore,

$$1 \mp \lim_{y\to\infty} \frac{h^*(y)}{y} = \lim_{y\to\infty} \frac{y \mp h^*(y)}{y} = \lim_{y\to\infty} \frac{g^{-1}(y)}{y} = \lim_{y\to\infty} \left(\frac{y\ell^*(y)}{g^{-1}(y)}\right)^{-1} \ell^*(y) = 1.$$

Hence, $h^*(y) = o(y)$.

(b) Let

$$\text{graph}(g) = \{(x, g(x)) \,:\, x \in [a, \infty)\}$$

be the graph of the function $g : [a, \infty) \to [b, \infty)$ and

$$\text{graph}(g^{-1}) = \{(x, g^{-1}(x)) \mid x \in [b, \infty)\}$$

be the graph of its inverse $g^{-1} : [b, \infty) \to [a, \infty)$.

Case 1: $g(x) = x + h(x)$ and $g^{-1}(x) = x - h^*(x)$. Then $h \geq 0$ is equivalent to $g(x) \geq x$, which is equivalent to

$$\text{graph}(g) \subset \{(x, y) \mid x \leq y\}. \tag{5.6}$$

Similarly, $h^* \geq 0$ is equivalent to $g^{-1}(x) \leq x$, which is equivalent to

$$\text{graph}(g^{-1}) \subset \{(x, y) \mid x \geq y\}. \tag{5.7}$$

It suffices to note that (5.6) and (5.7) are equivalent. So $h \geq 0$ implies $h^* \geq 0$.

Case 2: $g(x) = x - h(x)$ and $g^{-1}(x) = x + h^*(x)$. This case follows with identical proof by swapping the roles of g and g^{-1}.

(c) The implication $h > 0$ implies $h^* > 0$ follows as in (b) with minor modifications (strict inequalities instead of non-strict ones).

□

Definition 5.1.5 (Foss et al. (2013), Section 2.8)**.** Let F be a univariate distribution function and $\overline{F} = 1-F$ its survival function. Then F is called $o(x)$-*insensitive* if

$$\overline{F}(x \pm y) \sim \overline{F}(x)$$

uniformly in $|y| \leq f(x)$ for all functions f such that $f(x) = o(x)$.

Lemma 5.1.6 (Foss et al. (2013), Section 2.8)**.** *The class of regularly varying distributions is $o(x)$-insensitive.*

Lemma 5.1.7. *Let R_0 and h be as in (5.3). Let Y be any random variable such that $|Y - R_0| \leq h(R_0)$ almost surely. Then Y and R_0 are tail-equivalent in the sense that*

$$\mathbb{P}(Y > r) \sim \mathbb{P}(R_0 > r).$$

In addition, if $a(u)$ is a normalising function for R_0 as in (5.4), then it will also be a normalising function for Y, that is, for any $r > 0$

$$u\,\mathbb{P}(Y > ra(u)) \longrightarrow r^{-\alpha} \quad \text{as} \quad u \to \infty.$$

Proof. By assumption the probability $\mathbb{P}(Y > r)$ is bounded by

$$\mathbb{P}(R_0 - h(R_0) > r) \leq \mathbb{P}(Y > r) \leq \mathbb{P}(R_0 + h(R_0) > r).$$

By Lemma 5.1.4 this implies

$$\mathbb{P}(R_0 > r + h^*(r)) \leq \mathbb{P}(Y > r) \leq \mathbb{P}(R_0 > r - h^{**}(r))$$

for functions $h^*(x) = o(x)$ and $h^{**}(x) = o(x)$. Since R_0 is regularly varying, Lemma 5.1.6 implies that left-hand side and right-hand side are asymptotically equivalent to $\mathbb{P}(R_0 > r)$. Hence, the same holds true for $\mathbb{P}(Y > r)$.

The second part follows then from

$$\lim_{u\to\infty} u\mathbb{P}(Y > ra(u)) = \lim_{u\to\infty} u\mathbb{P}(R_0 > ra(u)) \cdot \lim_{u\to\infty} \frac{\mathbb{P}(Y > ra(u))}{\mathbb{P}(R_0 > ra(u))} = r^{-\alpha} \cdot 1. \quad \square$$

In our setup, Y can be for instance $R_0 - R_1$, $R_0 + R_1$ or $\|\boldsymbol{X}\|$, which sits inbetween the first two expressions. In particular, this shows that $\mathbb{P}(\|\boldsymbol{X}\| > t)$ is regularly varying with the same tail index as R_0 and $u\,\mathbb{P}(\|\boldsymbol{X}\| > ra(u)) \longrightarrow r^{-\alpha}$ as $u \to \infty$, for $a(u)$ as in (5.4). This shows part (a) of Theorem 5.1.1. Note that part (a) of Theorem 5.1.1 is a particular case of a primary result in de Haan et al. (2015).

Lemma 5.1.8. *Consider the model $\boldsymbol{X} = R_0\boldsymbol{S}_0 + R_1\boldsymbol{S}_1$ as in (5.3). For any $\varepsilon > 0$,*

$$\mathbb{P}\left(\frac{R_1}{R_0} > \varepsilon \,\middle|\, \|\boldsymbol{X}\| > r\right) \longrightarrow 0 \quad \text{as} \quad r \to \infty.$$

Proof. Note that

$$\begin{aligned}
&\mathbb{P}\left(\frac{R_1}{R_0} > \varepsilon \;\middle|\; \|\boldsymbol{X}\| > r\right) \\
&= \frac{\mathbb{P}\left(\frac{R_1}{R_0} > \varepsilon, \|\boldsymbol{X}\| > r\right)}{\mathbb{P}(\|\boldsymbol{X}\| > r)} \cdot \frac{\mathbb{P}(R_0 + R_1 > r)}{\mathbb{P}(R_0 + R_1 > r)} \cdot \frac{\mathbb{P}(R_0 > r)}{\mathbb{P}(R_0 > r)} \\
&\leq \frac{\mathbb{P}\left(\frac{R_1}{R_0} > \varepsilon,\ R_0 + R_1 > r\right)}{\mathbb{P}(R_0 + R_1 > r)} \cdot \frac{\mathbb{P}(R_0 + R_1 > r)}{\mathbb{P}(R_0 > r)} \cdot \frac{\mathbb{P}(R_0 > r)}{\mathbb{P}(\|\boldsymbol{X}\| > r)}.
\end{aligned}$$

By Lemma 5.1.7, the second and third terms in the product above converge to 1 as $r \to \infty$. For the first term, we have

$$\begin{aligned}
\frac{\mathbb{P}\left(\frac{R_1}{R_0} > \varepsilon,\ R_0 + R_1 > r\right)}{\mathbb{P}(R_0 + R_1 > r)} &= \mathbb{P}\left(\frac{R_1}{R_0} > \varepsilon \;\middle|\; R_0 + R_1 > r\right) \\
&= \mathbb{P}\left(\frac{h(R_0)}{R_0} > \varepsilon \;\middle|\; R_0 + h(R_0) > r\right) \\
&= \mathbb{P}\left(\frac{h(R_0)}{R_0} > \varepsilon \;\middle|\; R_0 > r - h^{**}(r)\right),
\end{aligned}$$

with $h^{**}(r) = o(r)$ due to Lemma 5.1.4. Finally, we have

$$\lim_{r\to\infty} \mathbb{P}\left(\frac{h(R_0)}{R_0} > \varepsilon \;\middle|\; R_0 > r - h^{**}(r)\right) = \lim_{r\to\infty} \mathbb{P}\left(\frac{h(R_0)}{R_0} > \varepsilon \;\middle|\; R_0 > r\right) = 0,$$

since $h(r) = o(r)$. □

Corollary 5.1.9. *For any $\varepsilon > 0$,*

$$\mathbb{P}\left(\frac{R_1/R_0}{\sqrt{1-(R_1/R_0)^2}} > \varepsilon \;\middle|\; \|\boldsymbol{X}\| > r\right) \longrightarrow 0 \quad \text{as} \quad r \to \infty.$$

Proof. The result follows from Lemma 5.1.8, together with the observation that

$$\frac{x}{\sqrt{1-x^2}} \sim x \quad \text{as} \quad x \to 0.$$

□

Proposition 5.1.10. *Let* dist_2 *be the Euclidean distance in* $\mathbb{R}^d$*. Then*

$$\mathrm{dist}_2\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}, \boldsymbol{S}_0\right) \leq \frac{R_1/R_0}{\sqrt{1-(R_1/R_0)^2}}$$

Proof. Let $\boldsymbol{O}$ be the the origin in $\mathbb{R}^d$. For a given $\boldsymbol{S}_0$ and R_0, the radius $R_1 = h(R_0)$ is fully determined by R_0 and the distance $\mathrm{dist}_2(\boldsymbol{X}/\|\boldsymbol{X}\|, \boldsymbol{S}_0)$ becomes largest in the situation where $\boldsymbol{S}_1$ lies on the unit circle in such a way that the angle $\sphericalangle \boldsymbol{S}_0\boldsymbol{O}\boldsymbol{X}$ becomes largest. This is the situation depicted in Figure 5.1 when the line $\boldsymbol{O}\boldsymbol{X}$ becomes a tangent line to the circle around $R_0\boldsymbol{S}_0$ with radius R_1, i.e. when the angle $\sphericalangle O\boldsymbol{X}(R_0\boldsymbol{S}_0)$ is a right angle. Now choose $\boldsymbol{A}$ and $\boldsymbol{B}$ in such a

way (Figure 5.2) that the three triangles $\boldsymbol{O}(R_0\boldsymbol{S}_0)\boldsymbol{X}$, $\boldsymbol{OA}(\boldsymbol{X}/\|\boldsymbol{X}\|)$ and $\boldsymbol{OS}_0\boldsymbol{B}$ are similar. This gives

$$\frac{\mathrm{dist}_2(\boldsymbol{S}_0, \boldsymbol{B})}{1} = \frac{R_1}{R_0}$$

and

$$\frac{\mathrm{dist}_2(\boldsymbol{X}/\|\boldsymbol{X}\|, \boldsymbol{A})}{1} = \frac{\mathrm{dist}_2(\boldsymbol{S}_0, \boldsymbol{B})}{\mathrm{dist}_2(\boldsymbol{O}, \boldsymbol{B})} = \frac{\mathrm{dist}_2(\boldsymbol{S}_0, \boldsymbol{B})}{\sqrt{1 - \mathrm{dist}_2(\boldsymbol{S}_0, \boldsymbol{B})^2}},$$

where the last identity follows from Pythagoras' theorem. Hence,

$$\mathrm{dist}_2\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}, \boldsymbol{A}\right) = \frac{R_1/R_0}{\sqrt{1 - (R_1/R_0)^2}},$$

and it remains to be seen that $\mathrm{dist}_2(\boldsymbol{X}/\|\boldsymbol{X}\|, \boldsymbol{S}_0) \leq \mathrm{dist}_2(\boldsymbol{X}/\|\boldsymbol{X}\|, \boldsymbol{A})$. However, this follows from the fact that the angle $\sphericalangle \boldsymbol{AS}_0(\boldsymbol{X}/\|\boldsymbol{X}\|)$ is larger than 90°. In order to see this, note that (i) the perpendicular projection of $\boldsymbol{X}/\|\boldsymbol{X}\|$ onto the line $\boldsymbol{OS}_0$, say $\boldsymbol{P}$, must lie between $\boldsymbol{O}$ and $\boldsymbol{S}_0$, and (ii), since $(\boldsymbol{X}/\|\boldsymbol{X}\|)\boldsymbol{A}$ is a tangent line to the unit circle, $\boldsymbol{A}$ must lie outside the unit circle. Hence $90° = \sphericalangle \boldsymbol{AP}(\boldsymbol{X}/\|\boldsymbol{X}\|) \leq \sphericalangle \boldsymbol{AS}_0(\boldsymbol{X}/\|\boldsymbol{X}\|)$. This finishes the proof. □

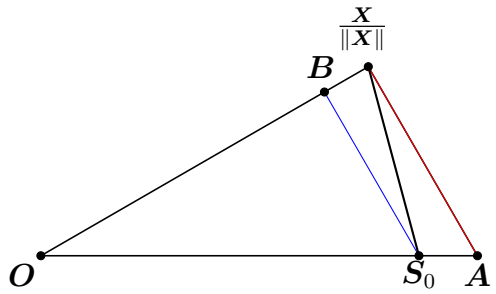

Figure 5.2: Refinement of part of Figure 5.1 to illustrate the geometric argument used in the proof of Proposition 5.1.10.

Corollary 5.1.11. *(a) For any* $\varepsilon > 0$*,*

$$\mathbb{P}\left(\mathrm{dist}_2\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}, \boldsymbol{S}_0\right) > \varepsilon \,\middle|\, \|\boldsymbol{X}\| > r\right) \longrightarrow 0 \quad \textit{as} \quad r \to \infty.$$

(b) the weak convergence

$$\mathcal{L}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \,\middle|\, \|\boldsymbol{X}\| > r\right) \longrightarrow \mathcal{L}(\boldsymbol{S}_0) = \sigma_0 \quad \textit{as} \quad r \to \infty$$

holds true.

Proof. (a) This is an immediate consequence of Corollary 5.1.9 and Proposition 5.1.10.

(b) This follows directly from part (a), since convergence in probability implies weak convergence (see e.g. Billingsley (1999), Theorem 3.1).

□

5.2 Numerical experiments

In order to explore the effectiveness of the vertex exchange algorithm, let us first make the framework we use to generate synthetic data more precise.

5.2.1 Setup

We have seen in Section 5.1 that we may interpret the stochastic model (5.3) as follows:

$$\underbrace{\boldsymbol{X}}_{\text{regularly varying}} = \underbrace{R_0}_{\substack{\text{determines index} \\ \text{of regular variation of } \boldsymbol{X}}} \underbrace{\boldsymbol{S}_0}_{\substack{\text{determines} \\ \text{spectral measure of } \boldsymbol{X}}} + \underbrace{R_1 \boldsymbol{S}_1}_{\text{distortion}},$$

In particular, we have control over the spectral measure σ of $\boldsymbol{X}$ via the choice of model for $\boldsymbol{S}_0 \sim \sigma_0$, since $\sigma = \sigma_0$, cf. Theorem 5.1.1. In subsequent numerical experiments, we will choose σ_0 to be a weighted measure on a finite set of cluster centers $\mathcal{C} \subset \mathbb{S}_{d-1}$ and σ_1 the uniform distribution on $\mathbb{S}_{d-1}$. In order to explore a broad range of configurations of cluster centers $\mathcal{C} \subset \mathbb{S}_{d-1}$, we first specify their number $c = |\mathcal{C}|$ and draw the points in $\mathcal{C}$ uniformly and independently from $\mathbb{S}_{d-1}$ and then explore the distribution of outcomes across these experiments. Since we focus on recovering the dependence structure, we simply draw R_0 from a standard Pareto distribution, $\mathbb{P}(R_0 > r) = r^{-1}$ for $r > 1$. For the perturbation function h in the distortion term we will typically consider $h(x) = \log(x)$. Alternatively, one might consider, for instance, $h(x) = \sqrt{x}$. Figure 5.3 shows an example in dimension $d = 2$.

We simulate data across different dimensions d and choose as the number of cluster centres in our setup $c = \lceil d \log(d) \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function. In all experiments, we assign equal weights $1/c$ to each cluster. Hence the true $\sigma = \sigma_0$ is the uniform distribution on those centers.

In extreme value analysis, the choice of the threshold (or equivalently k) to select the largest observations is typically a critical issue. As we have already a wide range of parameters for exploration, we opt for a pragmatic way of dealing with this issue here (similar to Janßen and Wan (2020)), and simply choose k as 10% of the number of observations.

On the other hand, when setting up meaningful experiments, we need to think carefully that we do not set ourselves an impossible task. Generating too few data points could lead to a situation, where some clusters are not evenly (or not often enough) drawn, when we simulate $\boldsymbol{S}_0 \sim \sigma_0 = \sigma$. Then, essentially no estimation procedure would be able to pick up points in those centers without data (not even the empirical spectral measure), and wrong conclusions would be drawn due a

Table 5.1: Stochastic model and vertex exchange algorithm parameters used for synthetic experiments. For the stochastic model parameters, d denotes the dimension; c is the number of clusters; w is the weight assigned to each cluster; h is the perturbation function; n is the number of observations generated; k is the number of selected observations. For the vertex exchange algorithm parameters, κ is the sparsity parameter; K is the kernel function; b is the kernel scale parameter; I is the potential maximal number of iterations.

Stochastic model Parameters							**Vertex Exchange Algorithm Parameters**				**Rep.**
d	c	w	h	n	k	p	κ	K	b	I	
3	4	0.25	log	1000	100	10	0.3	Powered exp. ($\alpha = 1$)	1	200	100
3	4	0.25	log	1000	100	10	0.6	Powered exp. ($\alpha = 1$)	1	200	100
3	4	0.25	log	1000	100	10	0.9	Powered exp. ($\alpha = 1$)	1	200	100
5	9	1/9	log	2500	250	15	0.3	Powered exp. ($\alpha = 1$)	1	500	100
5	9	1/9	log	2500	250	15	0.6	Powered exp. ($\alpha = 1$)	1	500	100
5	9	1/9	log	2500	250	15	0.9	Powered exp. ($\alpha = 1$)	1	500	100
7	14	1/14	log	5000	500	20	0.3	Powered exp. ($\alpha = 1$)	1	1000	100
7	14	1/14	log	5000	500	20	0.6	Powered exp. ($\alpha = 1$)	1	1000	100
7	14	1/14	log	5000	500	20	0.9	Powered exp. ($\alpha = 1$)	1	1000	100
20	60	1/60	log	32000	3200	30	0.3	Powered exp. ($\alpha = 1$)	1	6400	100
20	60	1/60	log	32000	3200	30	0.6	Powered exp. ($\alpha = 1$)	1	6400	100
20	60	1/60	log	32000	3200	30	0.9	Powered exp. ($\alpha = 1$)	1	6400	100
25	80	1/80	log	48000	4800	35	0.3	Powered exp. ($\alpha = 1$)	1	9600	100
25	80	1/80	log	48000	4800	35	0.6	Powered exp. ($\alpha = 1$)	1	9600	100
25	80	1/80	log	48000	4800	35	0.9	Powered exp. ($\alpha = 1$)	1	9600	100
30	103	1/103	log	70000	7000	40	0.3	Powered exp. ($\alpha = 1$)	1	14000	100
30	103	1/103	log	70000	7000	40	0.6	Powered exp. ($\alpha = 1$)	1	14000	100
30	103	1/103	log	70000	7000	40	0.9	Powered exp. ($\alpha = 1$)	1	14000	100
7	14	1/14	log	5000	500	20	0.3	Matérn ($\nu = 0.25$)	1	1000	100
7	14	1/14	log	5000	500	20	0.6	Matérn ($\nu = 0.25$)	1	1000	100
7	14	1/14	log	5000	500	20	0.9	Matérn ($\nu = 0.25$)	1	1000	100
7	14	1/14	log	5000	500	20	0.3	Gen. Cauchy ($\alpha = \tau = 1$)	1	1000	100
7	14	1/14	log	5000	500	20	0.6	Gen. Cauchy ($\alpha = \tau = 1$)	1	1000	100
7	14	1/14	log	5000	500	20	0.9	Gen. Cauchy ($\alpha = \tau = 1$)	1	1000	100
7	14	1/14	log	5000	500	20	0.3	Dagum ($\alpha = 0.5$, $\tau = 1$)	1	1000	100
7	14	1/14	log	5000	500	20	0.6	Dagum ($\alpha = 0.5$, $\tau = 1$)	1	1000	100
7	14	1/14	log	5000	500	20	0.9	Dagum ($\alpha = 0.5$, $\tau = 1$)	1	1000	100
7	14	1/14	log	5000	500	20	0.3	Powered exp. ($\alpha = 1$)	0.5	1000	100
7	14	1/14	log	5000	500	20	0.6	Powered exp. ($\alpha = 1$)	0.5	1000	100
7	14	1/14	log	5000	500	20	0.9	Powered exp. ($\alpha = 1$)	0.5	1000	100
7	14	1/14	log	5000	500	20	0.3	Powered exp. ($\alpha = 1$)	1.5	1000	100
7	14	1/14	log	5000	500	20	0.6	Powered exp. ($\alpha = 1$)	1.5	1000	100
7	14	1/14	log	5000	500	20	0.9	Powered exp. ($\alpha = 1$)	1.5	1000	100
7	14	1/14	log	5000	500	20	0.3	Powered exp. ($\alpha = 1$)	3	1000	100
7	14	1/14	log	5000	500	20	0.6	Powered exp. ($\alpha = 1$)	3	1000	100
7	14	1/14	log	5000	500	20	0.9	Powered exp. ($\alpha = 1$)	3	1000	100

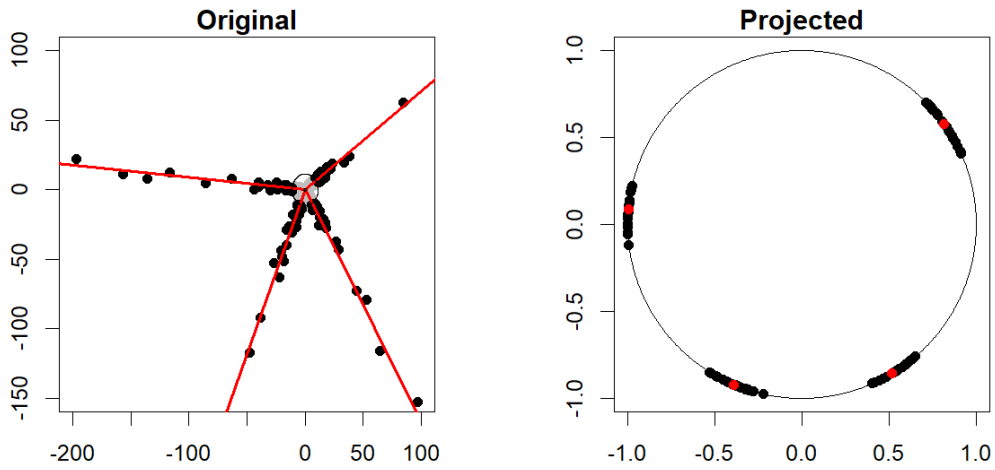

Figure 5.3: Illustration of 100 largest points (with respect to Euclidean norm) of a sample of length 1000 from model (5.3) in dimension $d = 2$. Here, $\boldsymbol{S}_0$ is a discrete distribution on the unit circle taking 4 specific values for potential directions (depicted in red), $\boldsymbol{S}_1$ is uniformly distributed, R_0 is drawn from a standard Pareto distribution $\mathbb{P}(R_0 > r) = r^{-1}$ for $r > 1$, and $R_1 = \log(R_0)$. Left: Original sample (clipping). Right: Projections onto the unit circle.

too complex model that is not sufficiently well represented by the data at hand. Therefore, in order to formulate a reasonable setup, we determine first, how large k needs to be so that

$$\mathbb{P}(\text{There exist at least } p \text{ points in each cluster.}) \gtrapprox 0.95,$$

where p ranges between 10 and 40, cf. Table 5.1. In order to determine such k, we need to understand quantiles of the multinomial distribution, which due to the combinatorial explosion for even moderately large number of clusters, we approximate by a multivariate normal distribution.

The vertex exchange algorithm is then applied to the empirical angular (probability) measure $\widehat{\sigma}_{n,k}$ from (5.2) with penalisation direction $\boldsymbol{d} = \mathbf{1}$. This means the k largest (in norm) $\boldsymbol{X}_{(1)}, \dots \boldsymbol{X}_{(k)}$ act as support points in the vertex exchange algorithm (Algorithm 1) to return a final estimate $\widehat{\sigma}_{n,k}^{(\kappa)}$ for the angular measure σ based on the choice of the tuning parameter κ. In particular, the threshold-determining k here corresponds to N in the vertex exchange algorithm. We experiment with different values of the tuning parameter κ to illustrate how it controls the sparsity of the approximation. Recall that lower κ means a sparser solution, while $\kappa = 1$ recovers the empirical angular measure. As kernel function, we primarily use the powered exponential kernel from Table 4.1 with smoothness parameter $\alpha = 1$ and scale parameter $b = 1$, but also explore the impact of changing kernels with different scale parameters.

Table 5.1 summarises the stochastic model parameters and the vertex exchange parameters used in all numerical experiments. The column I (maximal number of iterations) is merely a numerical safety measure, which is irrelevant subsequently, as our stopping rules never hit I. However, note that each experiment is repeated 100 times, and within each experiment a new set of cluster centers is drawn. So we are evaluating performance on a population of models in a certain dimension, rather than fixing cluster centers arbitrarily.

The core pieces of our implementation are documented in Appendix C. All experiments have been run in R Core Team (2023). In order to evaluate the Wasserstein distance of empirical spectral measures $\widehat{\sigma}_{n,k}$ and resulting sparse approximations $\widehat{\sigma}_{n,k}^{(\kappa)}$ to the true spectral measure σ, we use the package `transport` (Schuhmacher et al., 2024).

5.2.2 Results

Recall that the aim was to obtain a sparse, but as accurate as possible representation of the angular measure σ of the random vector $\boldsymbol{X}$ from independent samples of $\boldsymbol{X}$, and that we refrain from a too wide exploration of the choice of the threshold k in our data experiments (by default the top 10% sample with respect to the norm). Instead, we focus on the role of the tuning parameter κ, and see to what extent it can indeed enforce sparsity in the approximations $\widehat{\sigma}_{n,k}^{(\kappa)}$ obtained by the vertex exchange algorithm (Algorithm 1) without compromising accuracy too heavily (or even not at all).

We measure *sparsity* here in terms of the number of data points needed to represent σ. As expected, the sparsity of the approximations $\widehat{\sigma}_{n,k}^{(\kappa)}$ obtained through the vertex exchange algorithm decreases as the value of κ increases across all of our experiments, as can be seen in Figures 5.4, 5.5, 5.7, 5.8. Exemplarily, let us focus on Figure 5.4 and dimension $d = 7$. Here, the full empirical spectral measure $\widehat{\sigma}_{n,k}$ is supported by 500 points on the sphere, whereas all vertex exchange approximations $\widehat{\sigma}_{n,k}^{(\kappa)}$ live on less than 70 points. The significant reduction in terms of sparsity is well visible in the plot on the right-hand side of Figure 5.4. In the higher-dimensional experiments (Figure 5.5) the compression is even more impressive reducing, for instance, an empirical representation from 7000 points to a value around 100 (when $d = 30$), a compression to less than 1.5%.

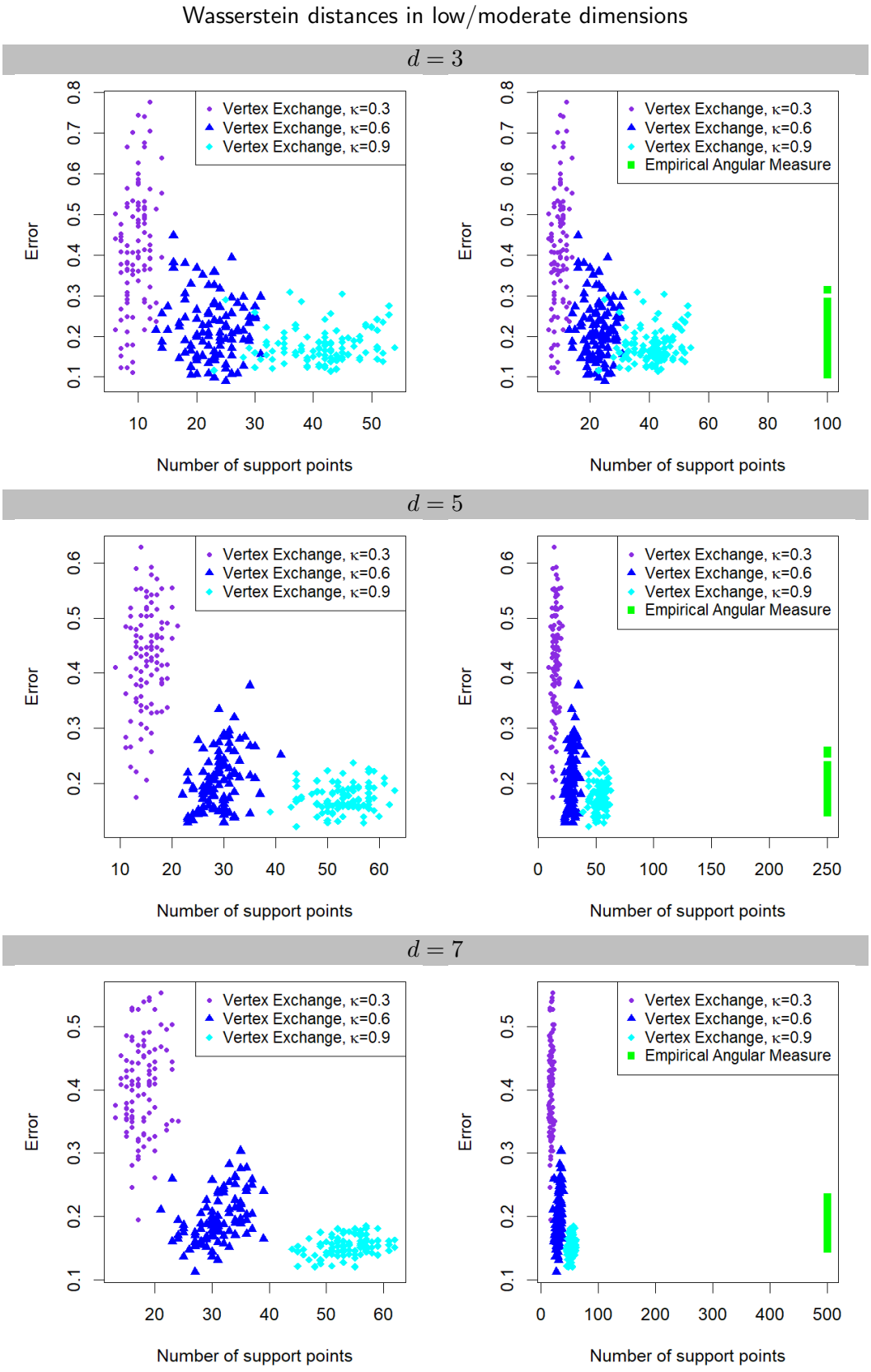

Figure 5.4: Left: Number of support points (sparsity) plotted against the error (Wasserstein distance) of the vertex exchange approximations with $\kappa = \{0.3, 0.6, 0.9\}$. Right: same, but with the empirical angular measure included.

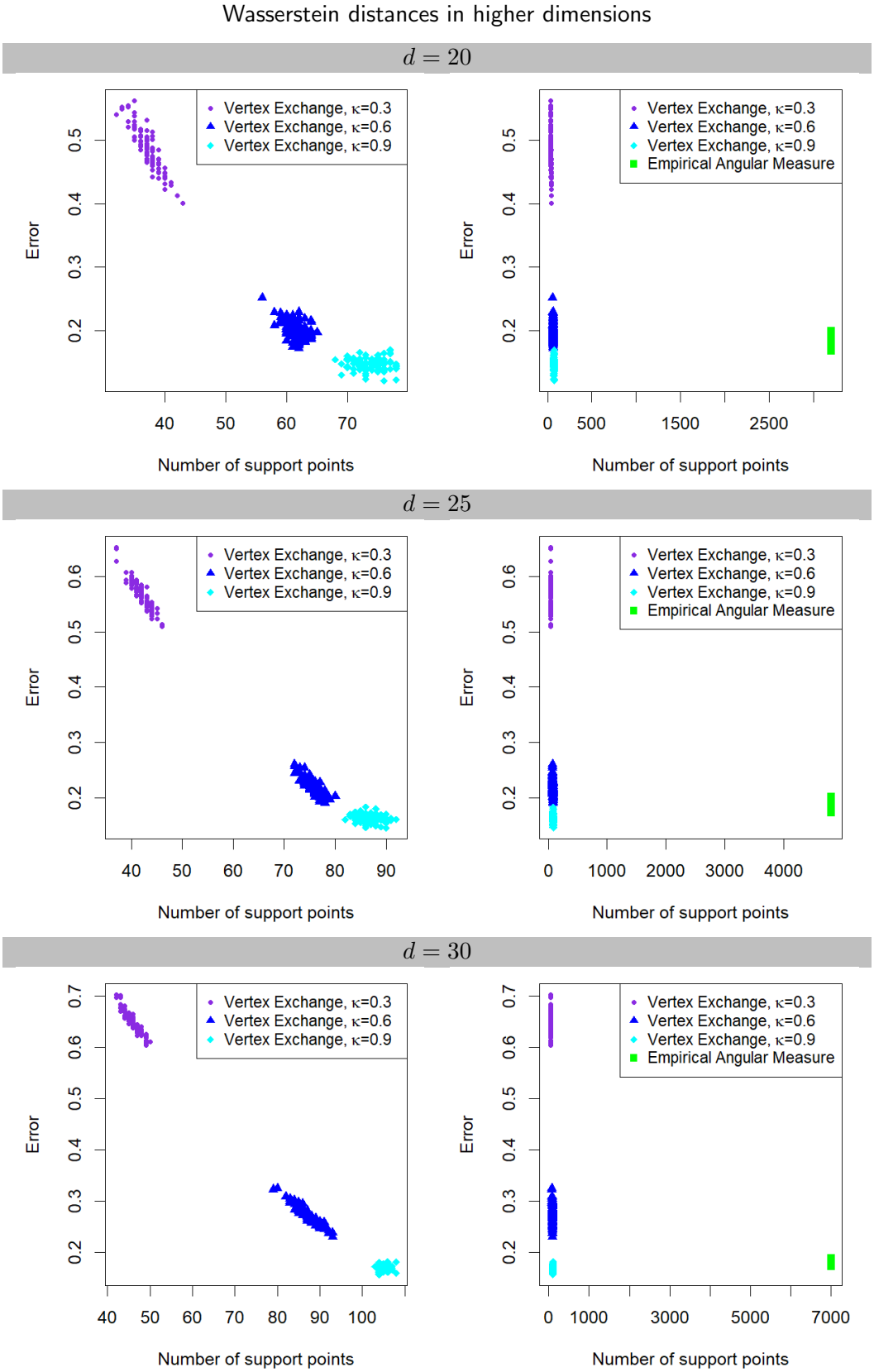

Figure 5.5: Left: number of support points (sparsity) plotted against the error (Wasserstein distance) of the vertex exchange approximations with $\kappa = \{0.3, 0.6, 0.9\}$. Right: same, but with the empirical angular measure included.

In order to compare the *accuracy* of the approximations obtained through the vertex exchange algorithm and the empirical spectral measure, we calculate the *Wasserstein distance*

$$W_1(\sigma, \widehat{\sigma}) = \inf_{\mu \in \Gamma(\sigma, \widehat{\sigma})} \int_{\mathbb{S}^+_{d-1} \times \mathbb{S}^+_{d-1}} \|\boldsymbol{x} - \boldsymbol{y}\|_2 \, \mu(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{y})$$

of our approximation $\widehat{\sigma}$ to the true spectral measure σ (which we know to be the uniform distribution of the cluster centers, cf. Section 5.2.1). Here $\Gamma(\sigma, \widehat{\sigma})$ is the space of all probability measures on $\mathbb{S}^+_{d-1} \times \mathbb{S}^+_{d-1}$ having first marginal σ and second marginal $\widehat{\sigma}$. For discrete measures, it can be computed using the package `transport` (Schuhmacher et al., 2024).

A priori, our methodology involving the vertex exchange algorithm only ensures proximity of the output $\widehat{\sigma}^{(\kappa)}_{n,k}$ to the empirical spectral measure $\widehat{\sigma}_{n,k}$ (a consistent estimator for σ) in terms of MMD, cf. (4.8), and only up to regularisation. We may view the MMD distance using a characteristic kernel as a surrogate for the Wasserstein distance; however it is not clear what we can expect in terms of accuracy for our outputs, since in theory we only know that the MMD is dominated by the Wasserstein distance (up to a multiplicative normalising constant), cf. Sriperumbudur et al. (2010, Thm. 21) (and not vice versa).

However, we would normally expect the choice of κ to represent a trade-off between sparsity and accuracy, i.e. higher κ enforcing less sparsity and giving typically the potential for a more accurate representation of σ. That this is indeed the case can again be seen in all of our data examples, cf. Figures 5.4, 5.5, 5.7, 5.8. Again, let us focus on dimension $d = 7$; Figure 5.6 compares the Wasserstein distances between the approximations $\widehat{\sigma}^{(\kappa)}_{n,k}$ and the empirical angular measure $\widehat{\sigma}_{n,k}$ to the true spectral measure σ. As κ increases from 0.3 to 0.9, we observe a significant decrease in the Wasserstein distance, and, naturally, higher values of κ lead to smaller variance of the error. In particular, note that when $\kappa = 0.9$, we typically even obtain a smaller Wasserstein distance than the empirical spectral measure $\widehat{\sigma}_{n,k}$; but already $\kappa = 0.6$ leads to only a minor information loss. That being said, all of these value for κ lead to a significantly reduced amount of data points to be saved, as discussed above. They are very sparse representation of σ compared to the empirical angular measure. Hence, within the trade-off between sparsity and accuracy, the vertex exchange approximations $\widehat{\sigma}^{(\kappa)}_{n,k}$ show a very promising behaviour. Overall, the benefits become most visible once we consider higher dimensions and larger data sets, cf. the comparison of left-hand and right-hand sides of the boxplots in Figure 5.6 as well as the corresponding sparsity-accuracy scatterplots in Figures 5.4 and 5.5.

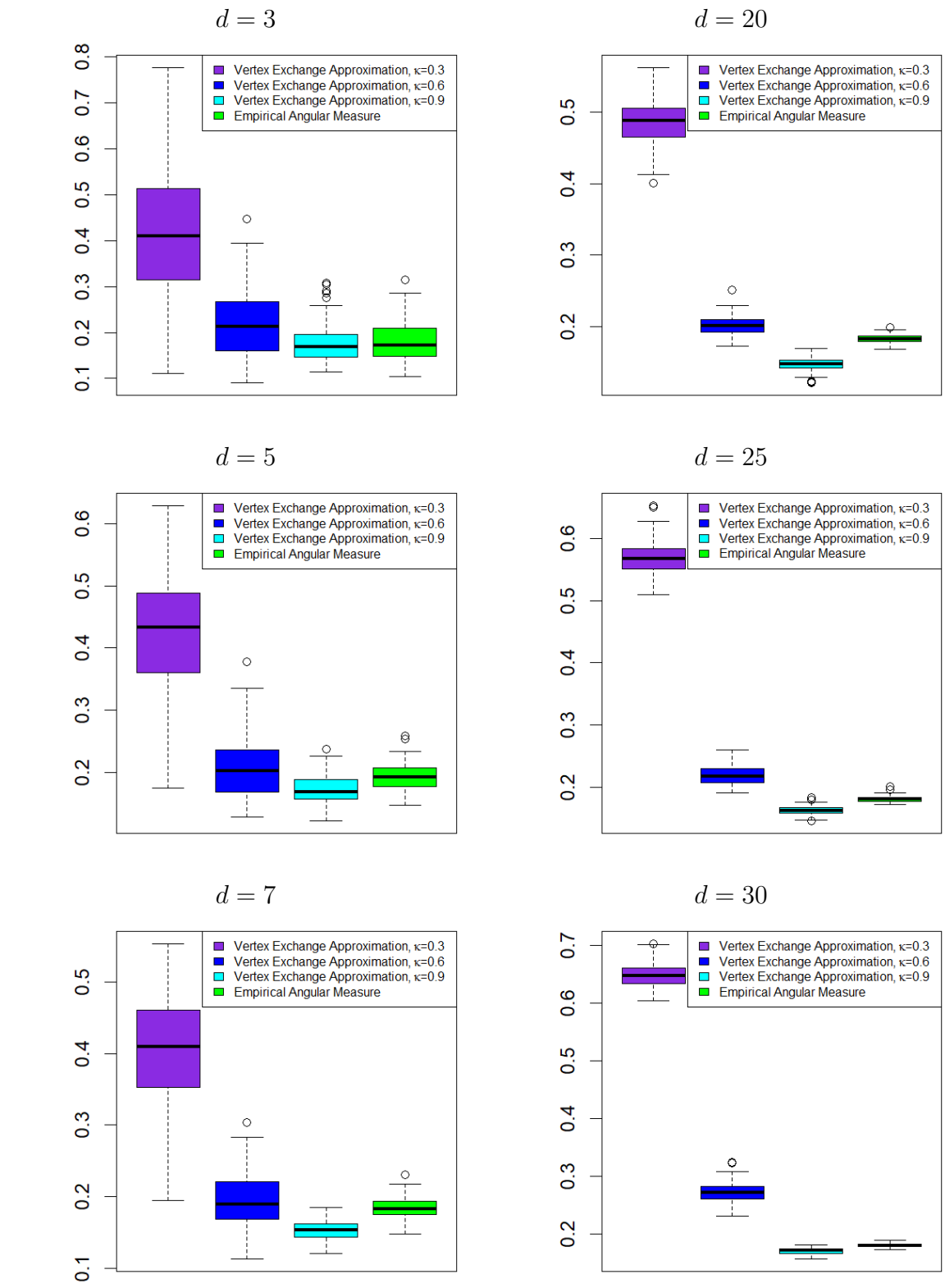

Figure 5.6: Boxplots of the Wasserstein distances for the empirical angular measure (green) and the vertex exchange algorithm approximations (purple, blue, cyan) with $\kappa = \{0.3, 0.6, 0.9\}$ in dimensions $d \in \{3, 5, 7, 20, 25, 30\}$.

Whilst specific compression rates and errors may vary, this general pattern does not seem to depend on various choices. For instance, this behaviour is replicated for different choices of kernels, cf. Figure 5.7, or scale parameters b, cf. Figure 5.8. Figures 5.9 and 5.10 display the kernel functions used. However, we also see that as the curves of the kernels functions decay slower, the approximations $\widehat{\sigma}_{n,k}^{(\kappa)}$ become less sparse. Moreover, when using a higher scale parameter, we obtain sparser approximations $\widehat{\sigma}_{n,k}^{(\kappa)}$, at the expense of a bigger Wasserstein distance to the original spectral measure σ.

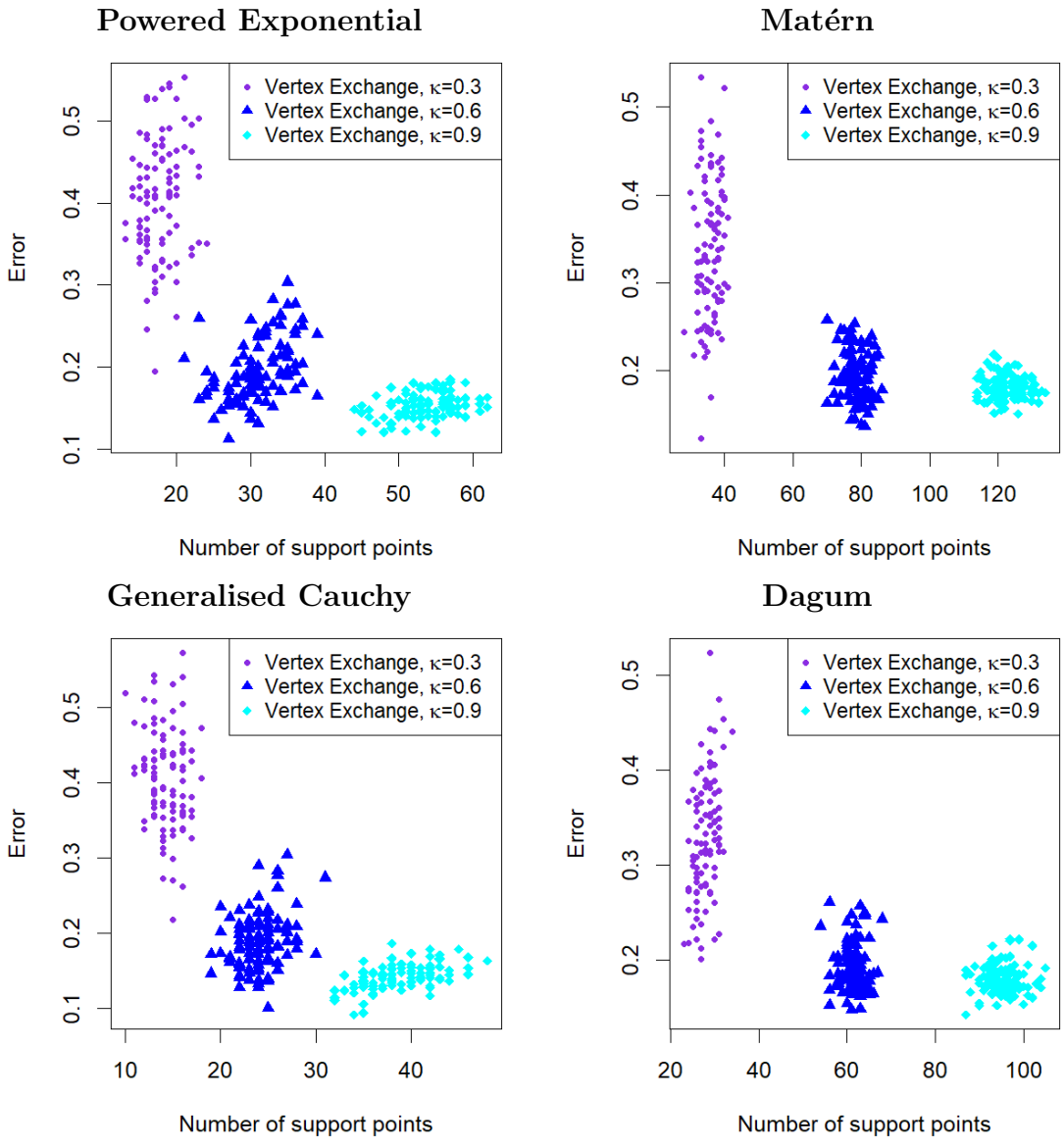

Figure 5.7: Number of support points (sparsity) plotted against the error (Wasserstein distance) of the vertex exchange approximations with $\kappa = \{0.3, 0.6, 0.9\}$ for the powered exponential kernel with $\alpha = 1$ (top left), Matérn kernel with $\nu = 0.25$ (top right), generalised Cauchy kernel with $\alpha = \tau = 1$ (bottom left) and Dagum kernel with $\alpha = 0.5$ and $\tau = 1$ (bottom right). Here, the dimension is $d = 7$ and we fix the scale parameter $b = 1$.

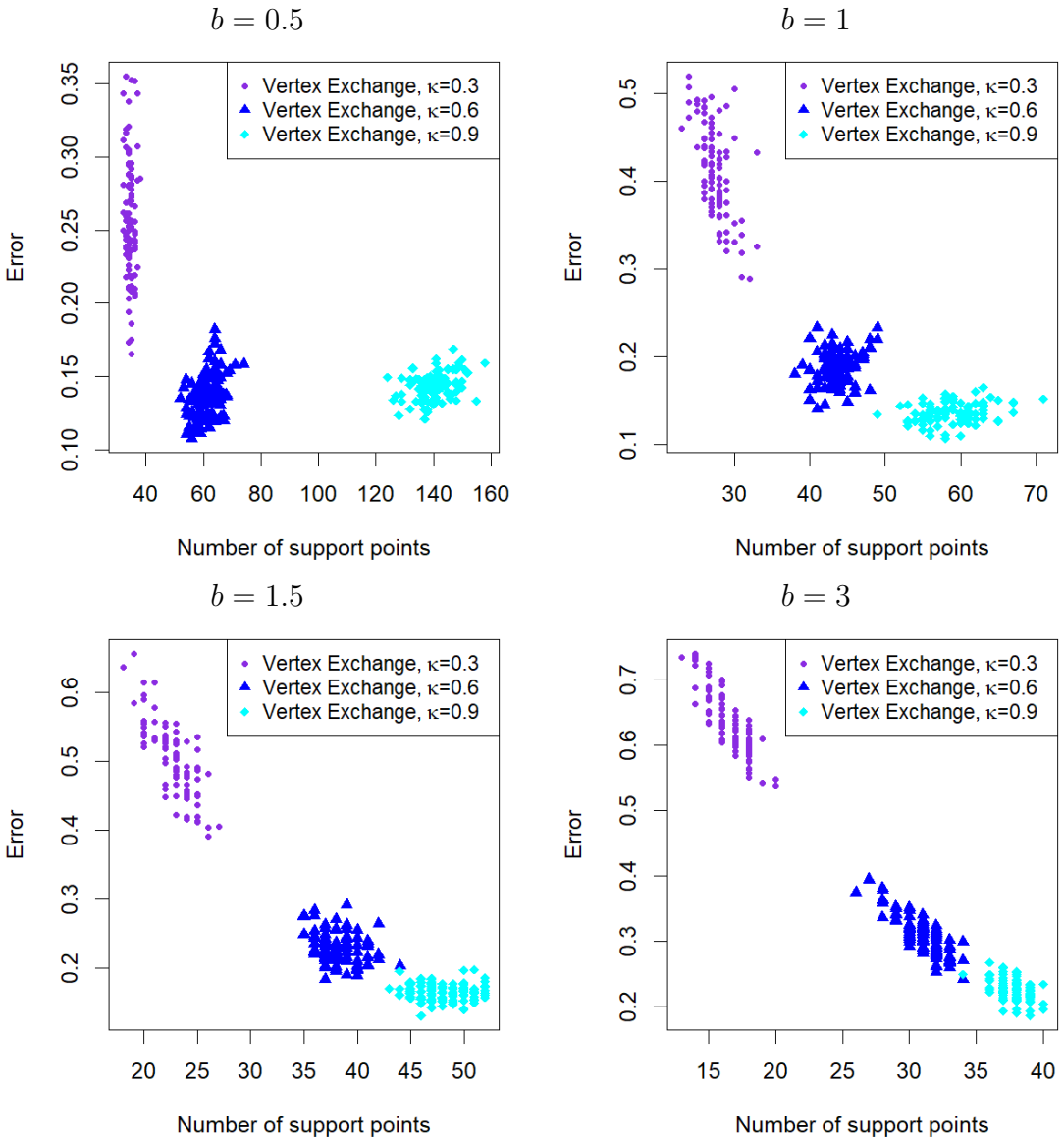

Figure 5.8: Number of support points (sparsity) plotted against the error (Wasserstein distance) of the vertex exchange approximations with $\kappa = \{0.3, 0.6, 0.9\}$ for the powered exponential kernel with smoothness parameter $\alpha = 1$ and scale parameters $b = 0.5$ (top left), $b = 1$ (top right), $b = 1.5$ (bottom left) and $b = 3$ (bottom right). Here, the dimension is $d = 7$.

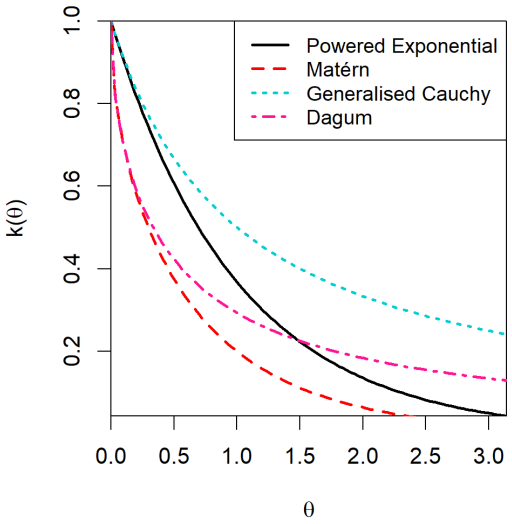

Figure 5.9: The powered exponential kernel with $\alpha = 1$, Matérn kernel with $\nu = 0.25$, generalised Cauchy kernel with $\alpha = \tau = 1$ and Dagum kernel with $\alpha = 0.5$ and $\tau = 1$ for $\theta \in [0, \pi]$ and scale parameter $b = 1$, cf. Table 4.1.

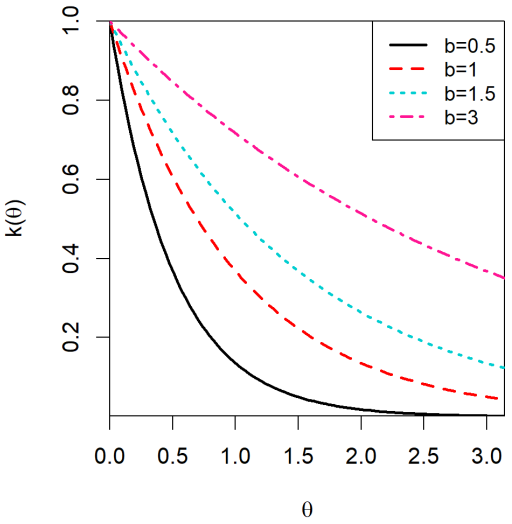

Figure 5.10: The powered exponential kernel, cf. Table 4.1, for $\theta \in [0, \pi]$, smoothness parameter $\alpha = 1$ and different scale parameters $b \in \{0.5, 1, 1.5, 3\}$.

Appendix A

Complete alternation and Bernstein functions

We recall some elementary definitions and facts from Berg et al. (1984), cf. also Molchanov (2017). Let $(S, \circ, e)$ be an *abelian semigroup*, that is, a non-empty set S with a composition $\circ$ that is associative and commutative and has a neutral element e. Three examples are of interest to us:

(i) $S = [0, \infty)$ with $+$ and neutral element 0,

(ii) $S = \mathcal{P}_d$, the power set of $\{1, \dots, d\}$, with the union operation $\cup$ and neutral element $\emptyset$,

(iii) $S = [0, \infty)^d$ with the componentwise maximum operation $\vee$ and neutral element $\mathbf{0}$.

Examples (ii) and (iii) are even *idempotent* semigroups, as $s \circ s = s$ for these operations. We use the standard notation

$$(\Delta_b f)(a) = f(a) - f(a \circ b).$$

Definition A.0.1. A function $f : S \to \mathbb{R}$ is called *completely alternating* if for all $n \geq 1$, $\{s_1, \dots, s_n\} \subset S$ and $s \in S$

$$(\Delta_{s_1} \Delta_{s_2} \dots \Delta_{s_n} f)(s) = \sum_{I \subset \{1,\dots,n\}} (-1)^{|I|} f(s \circ \bigcirc_{i \in I} s_i) \leq 0.$$

For idempotent semigroups (examples (ii) and (iii) above), the complete alternation property coincides with negative definiteness, cf. Berg et al. (1984) 4.4.16 and 4.6.8.

Definition A.0.2. A function $f : S \to \mathbb{R}$ is called *negative definite* if for all $n \geq 2$, $\{s_1, \dots, s_n\} \subset S$, $\{a_1, \dots, a_n\} \subset \mathbb{R}$ with $a_1 + \dots + a_n = 0$

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k f(s_j \circ s_k) \leq 0.$$

In the context of multivariate extremes, max-complete alternation of the stable tail dependence function implies union-complete alternation of the extremal coefficient function. In fact, the following directional version holds true irrespective of whether we take homogeneity or marginal standardisation into account or not.

Lemma A.0.3. *Let $\ell : [0,\infty)^d \to [0,\infty)$ be max-completely alternating. Let $\boldsymbol{x} \in [0,\infty)^d$. Let $\theta^{(x)} : \mathcal{P}_d \to [0,\infty)$ be defined as $\theta^{(x)}(A) = \ell(\boldsymbol{x}_A)$, where $\boldsymbol{x}_A = \boldsymbol{x} \cdot \boldsymbol{e}_A \in [0,\infty)^d$ is the vector with x_i as i-th coordinate if $i \in A$ and zero else. Then $\theta^{(x)}$ is union-completely alternating.*

Proof. The result follows from the observation that $\boldsymbol{x}_{A\cup B} = \boldsymbol{x}_A \vee \boldsymbol{x}_B$ for $A, B \in \mathcal{P}_d$. Therefore,

$$(\Delta_{A_1} \dots \Delta_{A_n} \theta^{(x)})(A) = (\Delta_{\boldsymbol{x}_{A_1}} \dots \Delta_{\boldsymbol{x}_{A_n}} \ell)(\boldsymbol{x}_A) \leq 0$$

for $A, A_1, \dots, A_n \in \mathcal{P}_d$, where $n \geq 1$. □

Lemma A.0.4 (Independent concatenation)**.** *Let $\theta_1 : \mathcal{P}(M) \to [0,\infty)$ and $\theta_2 : \mathcal{P}(N) \to [0,\infty)$ be union-completely alternating, where $\mathcal{P}(M)$ and $\mathcal{P}(N)$ are the power sets of finite sets M and N, respectively, such that $\theta_1(\emptyset) = \theta_2(\emptyset) = 0$. Then $\theta : \mathcal{P}(M \cup N) \to [0,\infty)$ with $\theta(A) = \theta_1(A \cap M) + \theta_2(A \cap N)$ is union-completely alternating and $\theta(\emptyset) = 0$.*

Proof. By the Choquet theorem (Schneider and Weil, 2008, Theorem 2.3.2) we may express

$$\theta_1(A) = \sum_{K \in \mathcal{P}(M) : K \cap A \neq \emptyset} a_K \qquad \text{and} \qquad \theta_2(B) = \sum_{L \in \mathcal{P}(N) : L \cap B \neq \emptyset} b_L$$

for non-negative coefficients a_K, $K \subset M$, $K \neq \emptyset$ and b_L, $L \subset N$, $L \neq \emptyset$. Define for $A \subset M$, $B \subset N$

$$\theta(A \cup B) = \sum_{(K,L) \in \mathcal{P}(M) \times \mathcal{P}(N) : (K \cup L) \cap (A \cup B) \neq \emptyset} c_{K \cup L},$$

where

$$c_{K\cup L} = \begin{cases} a_K & \text{if } K \neq \emptyset, L = \emptyset, \\ b_L & \text{if } K = \emptyset, L \neq \emptyset, \\ 0 & \text{if } K \neq \emptyset, L \neq \emptyset. \end{cases}$$

Then it is easily seen that $\theta(A \cup B) = \theta_1(A) + \theta_2(B)$, hence the assertion. □

Corollary A.0.5. *Let $\theta : \mathcal{P}_d \to [0,\infty)$ be union-completely alternating with $\theta(\emptyset) = 0$. Then $\theta' : \mathcal{P}_{d+1} \to [0,\infty)$, $\theta'(A) = \theta(A \cap \{1, \dots, d\}) + c\mathbf{1}_{d+1 \in A}$ is union-completely alternating with $\theta'(\emptyset) = 0$ for any $c \geq 0$.*

There are various equivalent definitions for *Bernstein functions*. For us it will be sufficient to consider the following. The equivalence of (i) and (ii) in the following theorem is a consequence from the 2-divisibility of $([0,\infty),+,0)$, cf. Berg et al. (1984) 4.6.8.

Theorem/Definition A.0.6. *A function $g : [0,\infty) \to \mathbb{R}$ is called a* Bernstein function *if one of the following equivalent conditions is satisfied:*

(i) $g \geq 0$, *g is continuous, and g is negative definite with respect to addition.*

(ii) $g \geq 0$, *g is continuous, and g is completely alternating with respect to addition.*

(iii) *g can be expressed as*

$$g(r) = a + br + \int_0^\infty (1 - \mathrm{e}^{-tr})\nu(\mathrm{d}t), \quad r \geq 0,$$

where $a, b \geq 0$ and ν is a non-negative Radon measure on $(0,\infty)$ satisfying the integrability condition $\int_0^\infty \min(t,1)\nu(\mathrm{d}t) < \infty$.

An important property of Berstein functions is that they act on negative definite kernels with non-negative diagonal, cf. Berg et al. (1984) 4.4.3.

Corollary A.0.7. *Let S be an idempotent semigroup and $f : S \to [0,\infty)$ be completely alternating and g a Bernstein function. Then the composition map $g \circ f : S \to [0,\infty)$ is completely alternating.*

Corollary A.0.8. *Let $\theta : \mathcal{P}_d \to [0,\infty)$ be union-completely alternating with $\theta(\emptyset) = 0$ and g be a Bernstein function. Let $A^* \subset \{1,\dots,d\}$ and $c > 0$. Then*

$$\sum_{J \subset \{1,\dots,d\} \setminus A^*} (-1)^{|J|} g\big(\theta(A^* \cup J)\big) \leq \sum_{J \subset \{1,\dots,d\} \setminus A^*} (-1)^{|J|} g\big(\theta(A^* \cup J) + c\big).$$

Proof. By Corollary A.0.5, the function $\theta' : \mathcal{P}_{d+1} \to [0,\infty)$, $\theta'(A) = \theta(A \cap \{1,\dots,d\}) + c\mathbf{1}_{d+1 \in A}$ is union-completely alternating with $\theta'(\emptyset) = 0$. Hence, Corollary A.0.7 implies that $g \circ \theta'$ is again union-completely alternating. Hence, by Definition A.0.1 and since $\{1,\dots,d,d+1\} \setminus A^*$ is not empty (it contains at least the element $d+1$)

$$\sum_{J' \subset \{1,\dots,d,d+1\} \setminus A^*} (-1)^{|J'|} g\big(\theta'(A^* \cup J')\big) \leq 0.$$

Now each J' above is either a subset J of $\{1,\dots,d\} \setminus A^*$ or it is of the form $J \cup \{d+1\}$, where J is a subset of $\{1,\dots,d\} \setminus A^*$. Separating the summands accordingly gives the assertion. □

The following proposition is the key argument to establish the implication $\Lambda \leq_{\text{uo}} \widetilde{\Lambda} \Rightarrow G \leq_{\text{uo}} \widetilde{G}$ in Theorem 3.2.1.

Proposition A.0.9. *Let $\theta : \mathcal{P}_d \to [0,\infty)$ and $\widetilde{\theta} : \mathcal{P}_d \to [0,\infty)$ be union-completely alternating with $\theta(\emptyset) = \widetilde{\theta}(\emptyset) = 0$. For $A \subset \{1,\dots,d\}$, $A \neq \emptyset$ set*

$$\chi(A) = \sum_{I \subset A,\, I \neq \emptyset} (-1)^{|I|+1} \theta(I) \qquad \text{and} \qquad \widetilde{\chi}(A) = \sum_{I \subset A,\, I \neq \emptyset} (-1)^{|I|+1} \widetilde{\theta}(I).$$

Suppose

$$\chi(A) \leq \widetilde{\chi}(A) \quad \text{for all} \quad A \subset \{1,\dots,d\},\, A \neq \emptyset.$$

Let $g : [0,\infty) \to [0,\infty)$ be a Bernstein function. Then

$$\sum_{I \subset \{1,\dots,d\}} (-1)^{|I|+1} g\big(\theta(I)\big) \leq \sum_{I \subset \{1,\dots,d\}} (-1)^{|I|+1} g\big(\widetilde{\theta}(I)\big).$$

Remark A.0.10. Under the assumptions of Proposition A.0.9 we have also

$$\sum_{I \subset A} (-1)^{|I|+1} g\big(\theta(I)\big) \leq \sum_{I \subset A} (-1)^{|I|+1} g\big(\widetilde{\theta}(I)\big)$$

for any non-empty subset A of $\{1,\dots,d\}$. This follows directly from the proposition as we may restrict θ and $\widetilde{\theta}$ to the respective subset A and all assumptions that were previously made for $\{1,\dots,d\}$ will be valid for the restrictions to A, too.

Proof. The inverse linear operation to recover θ from χ is given by

$$\theta(A) = \sum_{I \subset A,\, I \neq \emptyset} (-1)^{|I|+1} \chi(I)$$

(and likewise for $\widetilde{\theta}$ and $\widetilde{\chi}$), so that both quantities contain the same information. If $\chi = \widetilde{\chi}$ and hence $\theta = \widetilde{\theta}$, the statement is trivially true. Otherwise, we will show the proposition in two steps. First, we will establish its validity in the situation when $\chi(A) < \widetilde{\chi}(A)$ only for one $A^* \subset \{1,\dots,d\}$, $A^* \neq \emptyset$ and $\chi(A) = \widetilde{\chi}(A)$ for all other $A \subset \{1,\dots,d\}$, $A \neq \emptyset$. Second, we will show how this allows us to derive the proposition using convexity and continuity arguments.

Step 1: Let $\chi(A) < \widetilde{\chi}(A)$ only for one $A^* \subset \{1,\dots,d\}$, $A^* \neq \emptyset$ and $\chi(A) = \widetilde{\chi}(A)$ for all other $A \subset \{1,\dots,d\}$, $A \neq \emptyset$. Then $c = \widetilde{\chi}(A^*) - \chi(A^*) > 0$ and

$$\widetilde{\theta}(A) = \begin{cases} \theta(A) + c & \text{if } A^* \subset A, \\ \theta(A) & \text{else} \end{cases}$$

if $|A^*|$ is odd, and

$$\theta(A) = \begin{cases} \widetilde{\theta}(A) + c & \text{if } A^* \subset A, \\ \widetilde{\theta}(A) & \text{else} \end{cases}$$

if $|A^*|$ is even, and in both situations it suffices to show that

$$\sum_{I\subset\{1,\dots,d\}:A^*\subset I}(-1)^{|I|+1}g\big(\theta(I)\big)\le\sum_{I\subset\{1,\dots,d\}:A^*\subset I}(-1)^{|I|+1}g\big(\widetilde{\theta}(I)\big),$$

which is equivalent to

$$\sum_{J\subset\{1,\dots,d\}\setminus A^*}(-1)^{|J|+|A^*|+1}g\big(\theta(A^*\cup J)\big)\le\sum_{J\subset\{1,\dots,d\}\setminus A^*}(-1)^{|J|+|A^*|+1}g\big(\widetilde{\theta}(A^*\cup J)\big).$$

Hence, if $|A^*|$ is odd, we need to establish

$$\sum_{J\subset\{1,\dots,d\}\setminus A^*}(-1)^{|J|}g\big(\theta(A^*\cup J)\big)\le\sum_{J\subset\{1,\dots,d\}\setminus A^*}(-1)^{|J|}g\big(\theta(A^*\cup J)+c\big),$$

and, if $|A^*|$ is even, we need to establish

$$\sum_{J\subset\{1,\dots,d\}\setminus A^*}(-1)^{|J|}g\big(\widetilde{\theta}(A^*\cup J)\big)\le\sum_{J\subset\{1,\dots,d\}\setminus A^*}(-1)^{|J|}g\big(\widetilde{\theta}(A^*\cup J)+c\big).$$

Both inequalities now follow directly from Corollary A.0.8.

Step 2: Let C_d be the set of points $x=(x_A)_{A\in\mathcal{P}_d\setminus\{\emptyset\}}$ in $\mathbb{R}^{\mathcal{P}_d\setminus\{\emptyset\}}$ such that the mapping $A\mapsto x_A$ becomes union-completely alternating when setting $x_\emptyset=0$. Then C_d is a convex cone with non-empty interior and $C_d\subset[0,\infty)^{\mathcal{P}_d\setminus\{\emptyset\}}$ with $(0,0,\dots,0)\in C_d$. Let $T:\mathbb{R}^{\mathcal{P}_d\setminus\{\emptyset\}}\to\mathbb{R}^{\mathcal{P}_d\setminus\{\emptyset\}}$ be the linear map, such that

$$(Tx)_A=\sum_{I\subset A,\,I\neq\emptyset}(-1)^{|I|+1}x_I.$$

Then $T\circ T$ is the identity mapping, hence T is invertible. In particular $D_d=\{Tx:x\in C_d\}$ is also a convex cone with non-empty interior and $C_d=\{Tx:x\in D_d\}$. We also note that $D_d\subset[0,\infty)^{\mathcal{P}_d\setminus\{\emptyset\}}$, cf. (**??**) and that $(0,0,\dots,0)\in D_d$. Within the setting of the proposition, we have $\theta,\widetilde{\theta}\in C_d$ and $\chi,\widetilde{\chi}\in D_d$ with $\theta=T(\chi)$, $\widetilde{\theta}=T(\widetilde{\chi})$ and $\chi=T(\theta)$, $\widetilde{\chi}=T(\widetilde{\theta})$.

If both θ and $\widetilde{\theta}$ are points in the interior of C_d, then χ and $\widetilde{\chi}$ are in the interior of D_d. Therefore, there exists $\varepsilon>0$ such that the Minkowski sum of the line segment between χ and $\widetilde{\chi}$ and an (e.g. Euclidean) ε-ball centered at $(0,0,\dots,0)\in\mathbb{R}^{\mathcal{P}_d\setminus\{\emptyset\}}$ is completely contained in D_d. Within this set we can find a chain $\chi=\chi^{(0)}\le\chi^{(1)}\le\chi^{(2)}\le\cdots\le\chi^{(n)}=\widetilde{\chi}$, such that for each $i=0,\dots,n-1$ we have that $\chi^{(i)}$ and $\chi^{(i+1)}$ differ only in one component. By construction, we also have that $\theta^{(i)}=T(\chi^{(i)})\in C_d$ and $\theta^{(i+1)}=T(\chi^{(i+1)})\in C_d$, so that we are in the situation of *Step 1* and we may conclude that

$$\sum_{I\subset\{1,\dots,d\}}(-1)^{|I|+1}g\big(\theta^{(i)}(I)\big)\le\sum_{I\subset\{1,\dots,d\}}(-1)^{|I|+1}g\big(\theta^{(i+1)}(I)\big)$$

for all $i = 0, \ldots, n-1$, hence the assertion (which does not depend on the choice of ε or the choice of the chain). In other words, we have established the assertion of the proposition if both θ and $\widetilde{\theta}$ are points in the interior of C_d.

To complete the argument, note that the mapping $f : C_d \to \mathbb{R}$ with

$$f(x) = g(0) + \sum_{I \subset \{1,\ldots,d\}, I \neq \emptyset} (-1)^{|I|+1} g\big(x_I\big)$$

is continuous. Let $v \in C_d$ be a vector in the interior of C_d. Then, for any $\delta > 0$ both $\theta + \delta v$ and $\widetilde{\theta} + \delta v$ are in the interior of C_d, whereas $\chi + \delta T(v) = T(\theta + \delta v)$ and $\widetilde{\chi} + \delta T(v) = T(\widetilde{\theta} + \delta v)$ are in the interior of D_d and still satisfy $\chi + \delta T(v) \leq \widetilde{\chi} + \delta T(v)$. Therefore, $f(\theta + \delta v) \leq f(\widetilde{\theta} + \delta v)$. Finally, since f is continuous, we can find for given $\varepsilon > 0$ a corresponding $\delta > 0$, such that $f(\theta + \delta v)$ is ε-close to $f(\theta)$, while $f(\widetilde{\theta} + \delta v)$ is ε-close to $f(\widetilde{\theta})$. The assertion of the proposition $f(\theta) \leq f(\widetilde{\theta})$ follows as we may choose ε arbitrarily close to zero. $\square$

Appendix B

Calculation of the max-zonoid envelope

Let K be the max-zonoid (or dependency set) associated with a stable tail dependence function ℓ of a simple max-stable random vector, that is,

$$K = \big\{\boldsymbol{k} \in [0,\infty)^d \,:\, \langle \boldsymbol{k}, \boldsymbol{u}\rangle \le \ell(\boldsymbol{u}) \quad \text{for all } \boldsymbol{u} \in \mathbb{S}_+^{(d-1)}\big\},$$

and, conversely,

$$\ell(\boldsymbol{x}) = \sup\{\langle \boldsymbol{x}, \boldsymbol{k}\rangle \,:\, \boldsymbol{k} \in K\}, \quad \boldsymbol{x} \in [0,\infty)^d,$$

cf. Molchanov (2008). Here, $\mathbb{S}_+^{(d-1)} = \{\boldsymbol{u} \in [0,\infty)^d \,:\, \|\boldsymbol{u}\|_2 = 1\}$ denotes the $(d-1)$-dimensional Euclidean unit sphere in $\mathbb{R}^d$ restricted to the upper orthant $[0,\infty)^d$. It is well-known that

$$\Delta^d \subset K \subset [0,1]^d,$$

where the cross-polytope $\Delta^d = \{\boldsymbol{x} \in [0,\infty)^d \,:\, \langle \boldsymbol{x}, \boldsymbol{1}\rangle \le 1\}$ corresponds to perfect dependence, whereas the cube $[0,1]^d$ corresponds to independence. In particular, in the direction along the i-th axis the set K contains precisely the set $\{t\boldsymbol{e}_i \,:\, t \in [0,1]\}$.

For illustrative purposes we restrict our attention to $d = 2$, where we seek to calculate a parametrisation of the boundary curve of a general dependency set K. To this end, we parametrise the upper unit circle via $\boldsymbol{u} = (\cos(\alpha), \sin(\alpha))^T \in \mathbb{S}_+^1$ for $\alpha \in [0, \pi/2]$ and we assume that ℓ is differentiable. For $\alpha \in (0, \pi/2)$ a point (x_1, x_2) on the desired envelope curve will then satisfy the two conditions

$$\begin{aligned} \left\langle \begin{pmatrix}\cos(\alpha)\\ \sin(\alpha)\end{pmatrix}, \begin{pmatrix}x_1\\ x_2\end{pmatrix} \right\rangle - \ell\begin{pmatrix}\cos(\alpha)\\ \sin(\alpha)\end{pmatrix} &= 0, \\ \frac{\partial}{\partial\alpha}\left[\left\langle \begin{pmatrix}\cos(\alpha)\\ \sin(\alpha)\end{pmatrix}, \begin{pmatrix}x_1\\ x_2\end{pmatrix} \right\rangle - \ell\begin{pmatrix}\cos(\alpha)\\ \sin(\alpha)\end{pmatrix}\right] &= 0, \end{aligned}$$

which can be seen by a standard calculus of variations argument (European Mathematical Society, 2020). Let $\partial_1\ell$ and $\partial_2\ell$ denote the partial derivatives of ℓ with respect to first and second component. The two conditions can be then be expressed as

$$\begin{aligned} x_1\cos(\alpha) + x_2\sin(\alpha) &= \ell(\cos(\alpha),\sin(\alpha)) \\ -x_1\sin(\alpha) + x_2\cos(\alpha) &= -\sin(\alpha)\partial_1\ell(\cos(\alpha),\sin(\alpha)) + \cos(\alpha)\partial_2\ell(\cos(\alpha),\sin(\alpha)). \end{aligned}$$

Solving the system for x_1 and x_2 (while taking into account $\sin^2(\alpha)+\cos^2(\alpha)=1$) gives

$$x_1 = \cos(\alpha)L(\alpha) + \sin^2(\alpha)L_1(\alpha) - \sin(\alpha)\cos(\alpha)L_2(\alpha), \tag{B.1}$$

$$x_2 = \sin(\alpha)L(\alpha) - \sin(\alpha)\cos(\alpha)L_1(\alpha) + \cos^2(\alpha)L_2(\alpha), \tag{B.2}$$

where

$$L(\alpha) = \ell(\cos(\alpha),\sin(\alpha)) \qquad \text{and} \qquad L_i(\alpha) = \partial_i\ell(\cos(\alpha),\sin(\alpha)), \quad i=1,2.$$

The parametrisation of the boundary curve of K as given by (B.1) and (B.2) is the basis for all our plots in this text.

Example B.0.1 (Hüsler-Reiß distribution)**.** For the bivariate Hüsler-Reiß family with stable tail dependence function

$$\ell(x_1,x_2) = x_1\Phi\left(\frac{\eta}{2} + \frac{\log(x_1/x_2)}{\eta}\right) + x_2\Phi\left(\frac{\eta}{2} + \frac{\log(x_2/x_1)}{\eta}\right), \tag{B.3}$$

where $\eta^2 = \gamma_{12}$, straightforward calculations show that

$$L_1(\alpha) = \widetilde{L}(\cot(\alpha)) \qquad \text{and} \qquad L_2(\alpha) = \widetilde{L}(\tan(\alpha)),$$

with

$$\widetilde{L}(t) = \Phi\left(\frac{\eta}{2} + \frac{\log(t)}{\eta}\right) + \frac{1}{\eta}\varphi\left(\frac{\eta}{2} + \frac{\log(t)}{\eta}\right) - \frac{1}{\eta t}\varphi\left(\frac{\eta}{2} - \frac{\log(t)}{\eta}\right).$$

In other situations the spectral density h of ℓ may be known, such that

$$\ell(x_1,x_2) = \int_0^1 \max(\omega x_1, (1-\omega)x_2)h(\omega)d\omega. \tag{B.4}$$

Example B.0.2 (Dirichlet model)**.** The spectral density of the bivariate Dirichlet model with parameter vector $(\alpha_1,\alpha_2)\in(0,\infty)^2$ is given by

$$h(\omega) = \frac{\Gamma(\alpha_1+\alpha_2+1)}{(\alpha_1\omega+\alpha_2(1-\omega))^{(\alpha_1+\alpha_2+1)}}\frac{\alpha_1^{\alpha_1}}{\Gamma(\alpha_1)}\frac{\alpha_2^{\alpha_2}}{\Gamma(\alpha_2)}\omega^{\alpha_1-1}(1-\omega)^{\alpha_2-1}. \tag{B.5}$$

Let us abbreviate

$$H(t) = \int_0^t h(\omega)d\omega \qquad \text{and} \qquad \widetilde{H}(t) = \int_0^t \omega h(\omega)d\omega.$$

Taking into account the identities $H(1) = 2$ and $\widetilde{H}(1) = 1$ (due to marginal standardisation) straightforward calculations yield

$$\begin{aligned}
\ell(x_1, x_2) &= x_1 - (x_1 + x_2)\widetilde{H}\left(\frac{x_2}{x_1 + x_2}\right) + x_2 H\left(\frac{x_2}{x_1 + x_2}\right), \\
\partial_1\ell(x_1, x_2) &= 1 - \widetilde{H}\left(\frac{x_2}{x_1 + x_2}\right), \\
\partial_2\ell(x_1, x_2) &= H\left(\frac{x_2}{x_1 + x_2}\right) - \widetilde{H}\left(\frac{x_2}{x_1 + x_2}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
L(\alpha) &= \cos(\alpha) - (\sin(\alpha) + \cos(\alpha))\widetilde{H}\left(\frac{1}{1 + \cot(\alpha)}\right) + \sin(\alpha) H\left(\frac{1}{1 + \cot(\alpha)}\right), \\
L_1(\alpha) &= 1 - \widetilde{H}\left(\frac{1}{1 + \cot(\alpha)}\right), \\
L_2(\alpha) &= H\left(\frac{1}{1 + \cot(\alpha)}\right) - \widetilde{H}\left(\frac{1}{1 + \cot(\alpha)}\right).
\end{aligned}$$

Appendix C

Supplementary code

The following code represents the core pieces for the numerical experiments in Section 5.2.

For low dimensions, one could in principle still determine k from p in the numerical setup of Section 5.2 using actual quantiles of the multinomial distribution.

```
StarsBars <- function(n,k){
  ## stopifnot(is.integer(n) & is.integer(k))
  stopifnot(n >= 1 & k>=0)
  nr_J <- k+1
  nc_J <- n-k+1
  J_matrix <- matrix(0:k,nr_J,nc_J)
  New_list <- list(matrix(NA,nr=0,nc=0))
  for (i in 1:n){
    Old_list <- New_list
    New_list <- vector("list", i+1)
    v <- cbind(1:(i+1),rev(1:(i+1)))
    J_i <- J_matrix[v[v[,1] <= nr_J & v[,2] <= nc_J,,drop=F]]+1
    for (j in J_i) {
      if (j==1) {
        res <- matrix(0,1,i) }
      else {
        if (j==i+1) {
          res <- matrix(1,1,i)
        }
        else {
          A <- cbind(1,Old_list[[j-1]])
          B <- cbind(0,Old_list[[j]])
          res <-rbind(A,B)
        }
      }
      New_list[[j]] <- res
    }
  }
  return(New_list[[k+1]])
}

allocation_from_SB <- function(sb_row){
  diff(c(0,which(sb_row==1),length(sb_row)+1))-1
}
```

```
all_cluster_allocations <- function(no_clusters,no_points){
  t(apply(StarsBars(no_clusters+no_points-1,no_clusters-1),
          MARGIN=1, FUN=allocation_from_SB))
}

probab_lowerbound_event <- function(lower_bound,no_clusters,no_points,
                                    cluster_prob=NULL){
  if (!is.null(cluster_prob)) {stopifnot(length(cluster_prob)==no_clusters)}
  stopifnot(lower_bound*no_clusters <= no_points)
  all_alloc <- all_cluster_allocations(no_clusters,no_points)
  new_alloc <- all_alloc[apply(all_alloc,1,
                               function(row){all(row>=lower_bound)}),,drop=F]
  if (is.null(cluster_prob)){cluster_prob=rep(1/no_clusters,no_clusters)}
  sum(apply(new_alloc,1,function(row){dmultinom(row, prob = cluster_prob)}))
}
```

In all other cases, we use the central limit theorem approximation of the multinomial distribution in order to determine k from p in the numerical setup of Section 5.2.

```
probab_lowerbound_event_clt <- function(lower_bound,no_clusters,no_points,
                                        cluster_prob=NULL){
  require(mvtnorm)
  if (!is.null(cluster_prob)) {stopifnot(length(cluster_prob)==no_clusters)}
  remaining_points <- no_points - lower_bound*no_clusters
  stopifnot(remaining_points >=0)
  if (is.null(cluster_prob)){cluster_prob=rep(1/no_clusters,no_clusters)}
  upper_mvnorm <- (no_points * cluster_prob - lower_bound)/
                      sqrt(no_points * cluster_prob * (1-cluster_prob))
  a <- matrix(cluster_prob/(1-cluster_prob),nc=1)
  corr_matrix <-  - sqrt(a %*% t(a))
  diag(corr_matrix) <- rep(1,no_clusters)
  pmvnorm(lower=-Inf, upper=upper_mvnorm, corr=corr_matrix)
}
```

The following code is used to simulate data according to the model (5.3) detailed in Section 5.1.

```
# Generate uniform points on d-dimensional sphere

unif.sphere <- function (r,d,nsim){
  x <-matrix(rnorm(d*nsim),ncol=d,nrow=nsim)
  norm <- sqrt(rowSums(x^2))
  (r*x)/norm
}

# Generating regularly varying data, spectral measure given by cluster centers

model.1.simu <- function(nsim, clusters, weights, h=function(x){1/(x^0.5)}){

  d <- ncol(clusters)
  cl <- nrow(clusters)

  stopifnot(abs(sum(weights)-1)<1e-12 & all(weights>=0))
  stopifnot(cl == length(weights))
```

```
  S0 <- clusters[sample(1:cl,nsim,replace=T,prob=weights),]
  R0 <- 1/runif(nsim)

  R1 <- h(R0)
  S1 <- unif.sphere(1,d,nsim)

  return(R0*S0 + R1*S1)
}
```

The following is the code used for the vertex exchange algorithm (Algorithm 1) from Section 4.3.

```
 VEX_Solver <- function(n_iter, Mat_K, w_vec = 1, d_vec = 1, kappa, stop = T){

  N <- nrow(Mat_K)

  if (identical(w_vec,1)){w_vec <- rep(1/N,N)}
  if (identical(d_vec,1)){d_vec <- rep(1,N)}

  Kw_vec <- Mat_K %*% w_vec
  wKw <- sum(Kw_vec * w_vec)
  d_vec_inv <- kappa / d_vec

  x_vec <- rep(0,N)
  x_supp <- c()

  ## Initialisation (best vertex alone, for sparsity)
  vec_ini <- diag(Mat_K) * (d_vec_inv^2) - 2 * Kw_vec * d_vec_inv
  b <- which.min(vec_ini)
  x_supp <- b
  x_vec[b] <- d_vec_inv[b]

  Kx_vec <- Mat_K[,b] * x_vec[b]
  grad <- Kx_vec -  Kw_vec # up to 2

  ## Record evolution of D (useful for stopping)
  Rec_D <- rep(0, n_iter)
  Rec_D[1] <- wKw + vec_ini[b]
  xKx <- Mat_K[b,b] * d_vec_inv[b]^2
  wKx <- Kw_vec[b] * d_vec_inv[b]


  if (n_iter>2){
    for (i in 2:n_iter){

      u <- which.min(grad * d_vec_inv)
      d <- x_supp[which.max(grad[x_supp] * d_vec_inv[x_supp])]
      if (u==d){ i <- i-1 ; break } # can only happen if x is the true solution

      ## Optimal stepsize
      term_1 <- grad[u] * d_vec_inv[u] - grad[d] * d_vec_inv[d]

      term_2 <- Mat_K[u,u] * d_vec_inv[u]^2 +
        Mat_K[d,d] * d_vec_inv[d]^2 -
        2 * Mat_K[u,d] * d_vec_inv[u] * d_vec_inv[d]
```

```
      oss <- - term_1 / term_2
      r <- min(oss, x_vec[d]/d_vec_inv[d])

      ## Update
      xKx <- xKx + (r^2) * term_2 +
        2 * r * (Kx_vec[u] * d_vec_inv[u] - Kx_vec[d] * d_vec_inv[d])
      wKx <- wKx + r * (Kw_vec[u] * d_vec_inv[u] - Kw_vec[d] * d_vec_inv[d])

      x_vec[u] <- x_vec[u] + r * d_vec_inv[u]
      x_vec[d] <- x_vec[d] - r * d_vec_inv[d]
      x_supp <- union(x_supp, u)

      Kx_vec <- Kx_vec + r * (Mat_K[,u] * d_vec_inv[u] - Mat_K[,d] * d_vec_inv[d])
      grad <- Kx_vec -  Kw_vec

      ## Record evolution of D
      Rec_D[i] <- wKw + xKx - 2 * wKx

      if(stop==T){## Stopping based on improvement
      eps <- 1e-4 # sensitivity parameter
      if ((Rec_D[i-1]-Rec_D[i])/wKw < eps){ break } }
    }
  }
  return(list(x_vec = x_vec,
              x_supp = x_supp,
              Rec_D = Rec_D[1:i]))

}
```

References

Aronszajn, N. (1950), 'Theory of reproducing kernels', *Trans. Amer. Math. Soc.* **68**, 337–404.
URL: https://doi.org/10.2307/1990404

Aulbach, S., Falk, M. and Zott, M. (2015), 'The space of D-norms revisited', *Extremes* **18**(1), 85–97.
URL: https://doi.org/10.1007/s10687-014-0204-y

Avella-Medina, M., Davis, R. A. and Samorodnitsky, G. (2022), 'Kernel PCA for multivariate extremes'.
URL: https://arxiv.org/abs/2211.13172

Avella-Medina, M., Davis, R. A. and Samorodnitsky, G. (2024), 'Spectral learning of multivariate extremes', *J. Mach. Learn. Res.* **25**, Paper No. 124, 36.
URL: http://jmlr.org/papers/v25/21-1367.html

Ballani, F. and Schlather, M. (2011), 'A construction principle for multivariate extreme value distributions', *Biometrika* **98**(3), 633–645.
URL: https://doi.org/10.1093/biomet/asr034

Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004), *Statistics of Extremes*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester. Theory and Applications, With contributions from Daniel De Waal and Chris Ferro.
URL: https://doi.org/10.1002/0470012382

Berg, C., Christensen, J. P. R. and Ressel, P. (1984), *Harmonic analysis on semigroups*, Vol. 100 of *Graduate Texts in Mathematics*, Springer-Verlag, New York. Theory of positive definite and related functions.
URL: https://doi.org/10.1007/978-1-4612-1128-0

Berlinet, A. and Thomas-Agnan, C. (2004), *Reproducing kernel Hilbert spaces in probability and statistics*, Kluwer Academic Publishers, Boston, MA. With a preface by Persi Diaconis.
URL: https://doi.org/10.1007/978-1-4419-9096-9

Billingsley, P. (1999), *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics, second edn, John Wiley & Sons, Inc., New York. A Wiley-Interscience Publication.
URL: https://doi.org/10.1002/9780470316962

Böhning, D. (1986), 'A vertex-exchange-method in D-optimal design theory', *Metrika* **33**(6), 337–347.
URL: https://doi.org/10.1007/BF01894766

Boldi, M.-O. and Davison, A. C. (2007), 'A mixture model for multivariate extremes', *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69**(2), 217–229.
URL: https://doi.org/10.1111/j.1467-9868.2007.00585.x

Brockett, P. L. and Golden, L. L. (1987), 'A class of utility functions containing all the common utility functions', *Management Sci.* **33**(8), 955–964.
URL: https://doi.org/10.1287/mnsc.33.8.955

Brown, B. M. and Resnick, S. I. (1977), 'Extreme values of independent stochastic processes', *J. Appl. Probability* **14**(4), 732–739.
URL: https://doi.org/10.2307/3213346

Buhmann, M. D. (2000), Radial basis functions, *in* 'Acta numerica, 2000', Vol. 9 of *Acta Numer.*, Cambridge Univ. Press, Cambridge, pp. 1–38.
URL: https://doi.org/10.1017/S0962492900000015

Caballé, J. and Pomansky, A. (1996), 'Mixed risk aversion', *J. Econom. Theory* **71**(2), 485–513.
URL: https://doi.org/10.1006/jeth.1996.0130

Chautru, E. (2015), 'Dimension reduction in multivariate extreme value analysis', *Electron. J. Stat.* **9**(1), 383–418.
URL: https://doi.org/10.1214/15-EJS1002

Chen, Y., Liu, P., Liu, Y. and Wang, R. (2022), 'Ordering and inequalities for mixtures on risk aggregation', *Math. Finance* **32**(1), 421–451.
URL: https://doi.org/10.1111/mafi.12323

Coles, S. G. and Tawn, J. A. (1991), 'Modelling extreme multivariate events', *J. Roy. Statist. Soc. Ser. B Stat. Methodol.* **53**(2), 377–392.
URL: https://doi.org/10.1111/j.2517-6161.1991.tb01830.x

Cooley, D. and Thibaud, E. (2019), 'Decompositions of dependence for high-dimensional extremes', *Biometrika* **106**(3), 587–604.
URL: https://doi.org/10.1093/biomet/asz028

Corradini, M. and Strokorb, K. (2024), 'Stochastic ordering in multivariate extremes', *Extremes* pp. 1–40.
URL: https://doi.org/10.1007/s10687-024-00486-0

Davison, A., Huser, R. and Thibaud, E. (2019), Spatial extremes, *in* 'Handbook of environmental and ecological statistics', Chapman & Hall/CRC Handb. Mod. Stat. Methods, CRC Press, Boca Raton, FL, pp. 711–744.
URL: https://doi.org/10.1201/9781315152509-35

Davydov, Y., Molchanov, I. and Zuyev, S. (2008), 'Strictly stable distributions on convex cones', *Electron. J. Probab.* **13**(11), 259–321.
URL: https://doi.org/10.1214/EJP.v13-487

de Haan, L. (1984), 'A spectral representation for max-stable processes', *Ann. Probab.* **12**(4), 1194–1204.
URL: https://www.jstor.org/stable/2243357

de Haan, L. and Ferreira, A. (2006), *Extreme value theory*, Springer Series in Operations Research and Financial Engineering, Springer, New York. An introduction.
URL: https://doi.org/10.1007/0-387-34471-3

de Haan, L., Klein Tank, A. and Neves, C. (2015), 'On tail trend detection: modeling relative risk', *Extremes* **18**(2), 141–178.
URL: https://doi.org/10.1007/s10687-014-0207-8

Dombry, C. and Ribatet, M. (2015), 'Functional regular variations, Pareto processes and peaks over threshold', *Stat. Interface* **8**(1), 9–17.
URL: https://doi.org/10.4310/SII.2015.v8.n1.a2

Drees, H. and Sabourin, A. (2021), 'Principal component analysis for multivariate extremes', *Electron. J. Stat.* **15**(1), 908–943.
URL: https://doi.org/10.1214/21-ejs1803

Einmahl, J. H. J. and Segers, J. (2009), 'Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution', *Ann. Statist.* **37**(5B), 2953–2989.
URL: https://doi.org/10.1214/08-AOS677

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling extremal events*, Vol. 33 of *Applications of Mathematics (New York)*, Springer-Verlag, Berlin. For insurance and finance.
URL: https://doi.org/10.1007/978-3-642-33483-2

Engelke, S. and Hitz, A. S. (2020), 'Graphical models for extremes', *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82**(4), 871–932. With discussions.
URL: https://doi.org/10.1111/rssb.12355

Engelke, S. and Ivanovs, J. (2021), 'Sparse structures for multivariate extremes', *Annu. Rev. Stat. Appl.* **8**, 241–270.
URL: https://doi.org/10.1146/annurev-statistics-040620-041554

European Mathematical Society (2020), Envelope, *in* 'Encyclopedia of Mathematics', EMS Press. Accessed on 2022-08-29.
URL: https://encyclopediaofmath.org/wiki/Envelope

Falk, M. (2019), *Multivariate Extreme Value theory and D-Norms*, Springer Series in Operations Research and Financial Engineering, Springer, Cham.
URL: https://doi.org/10.1007/978-3-030-03819-9

Falk, M., Hüsler, J. and Reiß, R.-D. (2004), *Laws of Small Numbers: Extremes and Rare Events*, extended edn, Birkhäuser Verlag, Basel.
URL: https://doi.org/10.1007/978-3-0348-7791-6

Fallat, S., Lauritzen, S., Sadeghi, K., Uhler, C., Wermuth, N. and Zwiernik, P. (2017), 'Total positivity in Markov structures', *Ann. Statist.* **45**(3), 1152–1184.
URL: https://doi.org/10.1214/16-AOS1478

Fisher, R. A. and Tippett, L. H. C. (1928), 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Proc. Cambridge Philos. Soc.* **24**(2), 180–190.
URL: https://doi.org/10.1017/s0305004100015681

Fomichov, V. and Ivanovs, J. (2023), 'Spherical clustering in detection of groups of concomitant extremes', *Biometrika* **110**(1), 135–153.
URL: https://doi.org/10.1093/biomet/asac020

Foss, S., Korshunov, D. and Zachary, S. (2013), *An introduction to heavy-tailed and subexponential distributions*, Springer Series in Operations Research and Financial Engineering, second edn, Springer, New York.
URL: https://doi.org/10.1007/978-1-4614-7101-1

Gauthier, B. and Suykens, J. A. K. (2018), 'Optimal quadrature-sparsification for integral operator approximation', *SIAM J. Sci. Comput.* **40**(5), A3636–A3674.
URL: https://doi.org/10.1137/17M1123614

Genton, M. G. (2002), 'Classes of kernels for machine learning: a statistics perspective', *J. Mach. Learn. Res.* **2**(2), 293–312.
URL: https://doi.org/10.1162/15324430260185637

Gissibl, N. and Klüppelberg, C. (2018), 'Max-linear models on directed acyclic graphs', *Bernoulli* **24**(4A), 2693–2720.
URL: https://doi.org/10.3150/17-BEJ941

Gneiting, T. (2013), 'Strictly and non-strictly positive definite functions on spheres', *Bernoulli* **19**(4), 1327–1349.
URL: https://doi.org/10.3150/12-BEJSP06

Goix, N., Sabourin, A. and Clémençon, S. (2017), 'Sparse representation of multivariate extremes with applications to anomaly detection', *J. Multivariate Anal.* **161**, 12–31.
URL: https://doi.org/10.1016/j.jmva.2017.06.010

Gudendorf, G. and Segers, J. (2010), Extreme-value copulas, *in* 'Copula Theory and Its Applications', Vol. 198 of *Lect. Notes Stat. Proc.*, Springer, Heidelberg, pp. 127–145.
URL: https://doi.org/10.1007/978-3-642-12465-5_6

Huang, X. (1992), *Statistics of Bivariate Extremes*, PhD Thesis, Erasmus University, Rotterdam, Tinbergen Institute Research series No. 22.

Hüsler, J. and Reiß, R.-D. (1989), 'Maxima of normal random vectors: between independence and complete dependence', *Statist. Probab. Lett.* **7**(4), 283–286.
URL: https://doi.org/10.1016/0167-7152(89)90106-5

Janßen, A. and Wan, P. (2020), 'k-means clustering of extremes', *Electron. J. Stat.* **14**(1), 1211–1233.
URL: https://doi.org/10.1214/20-EJS1689

Kabluchko, Z. (2011), 'Extremes of independent Gaussian processes', *Extremes* **14**(3), 285–310.
URL: https://doi.org/10.1007/s10687-010-0110-x

Kabluchko, Z., Schlather, M. and de Haan, L. (2009), 'Stationary max-stable fields associated to negative definite functions', *Ann. Probab.* **37**(5), 2042–2065.
URL: https://doi.org/10.1214/09-AOP455

Karlin, S. and Rinott, Y. (1980), 'Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions', *J. Multivariate Anal.* **10**(4), 467–498.
URL: https://doi.org/10.1016/0047-259X(80)90065-2

Kulik, R. and Soulier, P. (2020), *Heavy-tailed time series*, Springer Series in Operations Research and Financial Engineering, Springer, New York.
URL: https://doi.org/10.1007/978-1-0716-0737-4

Li, H. (2013), Dependence comparison of multivariate extremes via stochastic tail orders, *in* 'Stochastic orders in reliability and risk', Vol. 208 of *Lect. Notes Stat.*, Springer, New York, pp. 363–387.
URL: https://doi.org/10.1007/978-1-4614-6892-9_19

Li, H. and Li, X., eds (2013), *Stochastic Orders in Reliability and Risk*, Vol. 208 of *Lecture Notes in Statistics*, Springer, New York. In honor of Professor Moshe Shaked, Papers from the International Workshop (SORR2011) held in Xiamen, June 27–29, 2011.
URL: https://doi.org/10.1007/978-1-4614-6892-9

Mainik, G. and Rüschendorf, L. (2012), 'Ordering of multivariate risk models with respect to extreme portfolio losses', *Stat. Risk Model.* **29**(1), 73–105.
URL: https://doi.org/10.1524/strm.2012.1103

Marshall, A. W. and Olkin, I. (1983), 'Domains of attraction of multivariate extreme value distributions', *Ann. Probab.* **11**(1), 168–177.
URL: https://www.jstor.org/stable/2243583

Marshall, A. W., Olkin, I. and Arnold, B. C. (2011), *Inequalities: Theory of Majorization and Its Applications*, Springer Series in Statistics, second edn, Springer, New York.
URL: https://doi.org/10.1007/978-0-387-68276-1

Marshall, A. W. and Proschan, F. (1965), 'An inequality for convex functions involving majorization', *J. Math. Anal. Appl.* **12**, 87–90.
URL: https://doi.org/10.1016/0022-247X(65)90056-9

Meyer, N. and Wintenberger, O. (2023), 'Multivariate sparse clustering for extremes', *Journal of the American Statistical Association* pp. 1–12.
URL: https://doi.org/10.1080/01621459.2023.2224517

Molchanov, I. (2008), 'Convex geometry of max-stable distributions', *Extremes* **11**(3), 235–259.
URL: https://doi.org/10.1007/s10687-008-0055-5

Molchanov, I. (2017), *Theory of random sets*, Vol. 87 of *Probability Theory and Stochastic Modelling*, second edn, Springer-Verlag, London.
URL: https://doi.org/10.1007/978-1-4471-7349-6

Molchanov, I. and Strokorb, K. (2016), 'Max-stable random sup-measures with comonotonic tail dependence', *Stochastic Process. Appl.* **126**(9), 2835–2859.
URL: https://doi.org/10.1016/j.spa.2016.03.004

Muandet, K., Fukumizu, K., Sriperumbudur, B., Schölkopf, B. et al. (2017), 'Kernel mean embedding of distributions: A review and beyond', *Foundations and Trends® in Machine Learning* **10**(1-2), 1–141.
URL: https://doi.org/10.1561/9781680832891

Müller, A. (1997), 'Integral probability metrics and their generating classes of functions', *Adv. in Appl. Probab.* **29**(2), 429–443.
URL: https://doi.org/10.2307/1428011

Müller, A. and Stoyan, D. (2002), *Comparison Methods for Stochastic Models and Risks*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester.
URL: https://www.wiley.com/-p-9780471494461

Papastathopoulos, I. and Strokorb, K. (2016), 'Conditional independence among max-stable laws', *Statist. Probab. Lett.* **108**, 9–15.
URL: https://doi.org/10.1016/j.spl.2015.08.008

Papastathopoulos, I. and Tawn, J. A. (2015), 'Stochastic ordering under conditional modelling of extreme values: drug-induced liver injury', *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64**(2), 299–317.
URL: https://doi.org/10.1111/rssc.12074

Paulsen, V. I. and Raghupathi, M. (2016), *An introduction to the theory of reproducing kernel Hilbert spaces*, Vol. 152 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge.
URL: https://doi.org/10.1017/CBO9781316219232

Pronzato, L. and Pázman, A. (2013), *Design of experiments in nonlinear models*, Vol. 212 of *Lecture Notes in Statistics*, Springer, New York. Asymptotic normality, optimality criteria and small-sample properties.
URL: https://doi.org/10.1007/978-1-4614-6363-4

R Core Team (2023), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. R version 4.3.2.
URL: https://www.R-project.org/

Resnick, S. I. (1987), *Extreme Values, Regular Variation, and Point Processes*, Vol. 4 of *Applied Probability. A Series of the Applied Probability Trust*, Springer-Verlag, New York.
URL: https://doi.org/10.1007/978-0-387-75953-1

Resnick, S. I. (2007), *Heavy-tail phenomena*, Springer Series in Operations Research and Financial Engineering, Springer, New York. Probabilistic and statistical modeling.
URL: https://doi.org/10.1007/978-0-387-45024-7

Ressel, P. (2013), 'Homogeneous distributions—and a spectral representation of classical mean values and stable tail dependence functions', *J. Multivariate Anal.* **117**, 246–256.
URL: https://doi.org/10.1016/j.jmva.2013.02.013

Ressel, P. (2022), 'Stable tail dependence functions—some basic properties', *Depend. Model.* **10**(1), 225–235.
URL: *https://doi.org/10.1515/demo-2022-0114*

Rohrbeck, C. and Cooley, D. (2023), 'Simulating flood event sets using extremal principal components', *Ann. Appl. Stat.* **17**(2), 1333–1352.
URL: https://doi.org/10.1214/22-aoas1672

Röttger, F., Engelke, S. and Zwiernik, P. (2023), 'Total positivity in multivariate extremes', *Ann. Statist.* **51**(3), 962–1004.
URL: https://doi.org/10.1214/23-aos2272

Sabourin, A. and Naveau, P. (2014), 'Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization', *Comput. Statist. Data Anal.* **71**, 542–567.
URL: https://doi.org/10.1016/j.csda.2013.04.021

Saitoh, S. and Sawano, Y. (2016), *Theory of reproducing kernels and applications*, Vol. 44 of *Developments in Mathematics*, Springer, Singapore.
URL: https://doi.org/10.1007/978-981-10-0530-5

Schlather, M. and Tawn, J. (2002), 'Inequalities for the extremal coefficients of multivariate extreme value distributions', *Extremes* **5**(1), 87–102.
URL: https://doi.org/10.1023/A:1020938210765

Schneider, R. and Weil, W. (2008), *Stochastic and integral geometry*, Probability and its Applications (New York), Springer-Verlag, Berlin.
URL: https://doi.org/10.1007/978-3-540-78859-1

Schuhmacher, D., Bähre, B., Bonneel, N., Gottschlich, C., Hartmann, V., Heinemann, F., Schmitzer, B. and Schrieber, J. (2024), *transport: Computation of Optimal Transport Plans and Wasserstein Distances*. R package version 0.15-2.
URL: *https://cran.r-project.org/package=transport*

Shaked, M. and Shanthikumar, J. G. (2007), *Stochastic Orders*, Springer Series in Statistics, Springer, New York.
URL: *https://doi.org/10.1007/978-0-387-34675-5*

Sibuya, M. (1960), 'Bivariate extreme statistics. I', *Ann. Inst. Statist. Math. Tokyo* **11**, 195–210.
URL: *https://doi.org/10.1007/bf01682329*

Simon-Gabriel, C.-J., Barp, A., Schölkopf, B. and Mackey, L. (2023), 'Metrizing weak convergence with maximum mean discrepancies', *J. Mach. Learn. Res.* **24**, Paper No. [184], 20.
URL: *http://jmlr.org/papers/v24/21-0599.html*

Simon-Gabriel, C.-J. and Schölkopf, B. (2018), 'Kernel distribution embeddings: universal kernels, characteristic kernels and kernel metrics on distributions', *J. Mach. Learn. Res.* **19**, Paper No. 44, 29.
URL: *http://jmlr.org/papers/v19/16-291.html*

Slepian, D. (1962), 'The one-sided barrier problem for Gaussian noise', *Bell System Tech. J.* **41**, 463–501.
URL: *https://doi.org/10.1002/j.1538-7305.1962.tb02419.x*

Sriperumbudur, B. K., Fukumizu, K. and Lanckriet, G. R. G. (2011), 'Universality, characteristic kernels and RKHS embedding of measures', *J. Mach. Learn. Res.* **12**, 2389–2410.
URL: *http://jmlr.org/papers/v12/sriperumbudur11a.html*

Sriperumbudur, B. K., Gretton, A., Fukumizu, K., Schölkopf, B. and Lanckriet, G. R. G. (2010), 'Hilbert space embeddings and metrics on probability measures', *J. Mach. Learn. Res.* **11**, 1517–1561.
URL: *http://jmlr.org/papers/v11/sriperumbudur10a.html*

Steinwart, I. and Ziegel, J. F. (2021), 'Strictly proper kernel scores and characteristic kernels on compact spaces', *Appl. Comput. Harmon. Anal.* **51**, 510–542.
URL: *https://doi.org/10.1016/j.acha.2019.11.005*

Strokorb, K. and Schlather, M. (2015), 'An exceptional max-stable process fully parameterized by its extremal coefficients', *Bernoulli* **21**(1), 276–302.
URL: *https://doi.org/10.3150/13-BEJ567*

Tong, Y. L. (1980), *Probability inequalities in multivariate distributions*, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London-Toronto, Ont.
URL: https://doi.org/10.1016/C2013-0-11609-1

Wendland, H. (2005), *Scattered data approximation*, Vol. 17 of *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge.
URL: https://doi.org/10.1017/CBO9780511617539

Yuen, R. and Stoev, S. (2014), 'Upper bounds on value-at-risk for the maximum portfolio loss', *Extremes* **17**(4), 585–614.
URL: https://doi.org/10.1007/s10687-014-0198-5

Yuen, R., Stoev, S. and Cooley, D. (2020), 'Distributionally robust inference for extreme Value-at-Risk', *Insurance Math. Econom.* **92**, 70–89.
URL: https://doi.org/10.1016/j.insmatheco.2020.03.003