

## Article

# On the Identification of the Riskiest Directional Components from Multivariate Heavy-Tailed Data

Miriam Hägele <sup>†</sup>  and Jaakko Lehtomaa <sup>\*,†</sup>

Department of Mathematics and Statistics, University of Helsinki, 00014 Helsinki, Finland; miriam.hagele@helsinki.fi

\* Correspondence: jaakko.lehtomaa@helsinki.fi

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** In univariate data, there exist standard procedures for identifying dominating features that produce the largest number of observations. However, in the multivariate setting, the situation is quite different. This paper aims to provide tools and methods for detecting dominating directional components in multivariate data. We study general heavy-tailed multivariate random vectors in dimension  $d \geq 2$  and present procedures that can be used to explain why the data are heavy-tailed. This is achieved by identifying the set of the riskiest directional components. The results are of particular interest in insurance when setting reinsurance policies, and in finance when hedging a portfolio of multiple assets.

**Keywords:** multivariate data; heavy-tail distribution; dominating tail behaviour

**MSC:** 60E05; 91B30; 91B28; 62P05



**Citation:** Hägele, Miriam, and Jaakko Lehtomaa. 2023. On the Identification of the Riskiest Directional Components from Multivariate Heavy-Tailed Data. *Risks* 11: 130. <https://doi.org/10.3390/risks11070130>

Academic Editor: Mogens Steffensen

Received: 24 May 2023

Revised: 3 July 2023

Accepted: 5 July 2023

Published: 13 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

It is not uncommon to find heavy-tailed features in multivariate data sets in insurance and finance (Embrechts et al. 1997; Peters and Shevchenko 2015). Since financial entities seek ways to reduce the total risks of their portfolios, it is necessary to understand what the main sources of risks are. Once this is known, one can seek optimal ways of reducing risk. In insurance, where the multivariate observations could consist of losses of different lines of business, the companies are typically interested in finding the best-suited reinsurance policy. In finance, where the data could consist of returns of multiple assets, the aim could be to find an optimal hedging strategy against large losses. Even though multivariate heavy-tailed distributions are encountered frequently in various applications, there does not exist a general framework for analysis.

In earlier research, several models for what a heavy-tailed random vector should mean have been introduced (Cline and Resnick 1992; Omey 2006; Resnick 2004; Samorodnitsky and Sun 2016). One of the simplest ways to model the situation is to present a  $d$ -dimensional random vector  $\mathbf{X}$  in polar coordinates using the length  $R$  of the vector and the directional vector  $\mathbf{U}$  on the unit sphere  $\mathbb{S}^{d-1}$  so that  $\mathbf{X} = R\mathbf{U}$ . Here, perhaps the simplest assumption would be that all directions are equally likely, i.e., the random vector  $\mathbf{U}$  has a uniform distribution on the unit sphere. In realistic models, some directions are more likely than others, specifically, when  $R$  is large. Efforts have been made to capture the heterogeneity in the distribution of  $\mathbf{U}$  by studying, for example, the set of elliptic distributions (Hult and Lindskog 2002; Klüppelberg et al. 2007; Li and Sun 2009). In addition, alternative approaches have been suggested by Feng et al. (2017, 2020). However, there are typical data sets that do not fit well to these assumptions because the support of  $\mathbf{U}$  is concentrated into a cone and does not cover the entire set  $\mathbb{S}^{d-1}$ . There appears to be a need for models that allow the heaviness of the tail to vary in different directions, as in Weng and Zhang (2012),

but which are not restricted by a parametric class of distributions. Ideally, the model should admit the tail of  $R$  to have a more general form than, say, only the form of a power function.

To answer the question of what causes the heavy-tailedness of multivariate data, one usually selects a suitable norm  $\|\cdot\|$  and analyses the resulting one-dimensional distribution of  $\|\mathbf{X}\|$  with standard procedures, such as the mean excess plot (Das and Ghosh 2016; Ghosh and Resnick 2010, 2011), or alternative methods, such as the ones in Asmussen and Lehtomaa (2017). In practice, given the heaviness identified from the distribution of  $\|\mathbf{X}\|$ , the aim is to analyse the distribution of  $\mathbf{U}$  and the dependence structure between  $R$  and  $\mathbf{U}$ . If the conditional distribution exists, we can write

$$\mathbb{P}(R > k \mid \mathbf{U} = \mathbf{x}) = e^{-h(k, \mathbf{x})}, \quad k \geq 0, \quad (1)$$

where  $h(\cdot, \mathbf{x})$  is an increasing function for a fixed  $\mathbf{x} \in \mathbb{S}^{d-1}$ . In this setting, the problem is to identify which set of vectors  $\mathbf{x} \in \mathbb{S}^{d-1}$  produces the heaviest or dominating conditional tails of the form (1). The dominating directions are the vectors  $\mathbf{x} \in \mathbb{S}^{d-1}$ , for which the function  $h(k, \mathbf{x})$  grows at the slowest rate, as  $k \rightarrow \infty$ . Presentation (1) admits the study of a wide class of distributions.

We derive a presentation for the set of directional components that produce the heaviest tails, and present a procedure to be used in practical analysis. The analysis is preliminary in the sense that some of the assumptions could be undoubtedly weakened. However, the simple assumptions admit a clear presentation of the main ideas without many technical difficulties. The method applied to the generated data returns a subset of the space, which gives more information about the distribution than a single number or vector, but demands less data than a full empirical measure. Furthermore, the original data do not have to be pre-processed or transformed, and the method is also applicable to data where, in some directions, the observations are sparse or do not occur at all. This is different from the grid-based approaches, such as those in Lehtomaa and Resnick (2020), where the entire space is first divided into cells, and each cell is studied separately. In such approaches, there can exist empty or sparsely populated cells, which might be problematic for analysis. The presented methods can be applied, in particular, to analyse financial data, where extreme observations typically appear in two opposite directions, which might be unknown.

### Notation

**Notation 1.** For set  $A$ , we denote its closure by  $\overline{A}$  or  $cl(A)$ , its interior by  $int(A)$ , and its complement by  $A^c$ , and  $\emptyset$  denotes an empty set. The symbol  $A \setminus D$  means  $A \cap D^c$ . The term  $A \subset D$  means  $A$  is a subset of  $D$ , whereas  $A \subsetneq D$  implies that  $A$  is a proper subset of  $D$ . The notation  $a(n) = o(b(n))$  refers to the little-o notation, and means  $a(n)/b(n) \rightarrow 0$ , as  $n \rightarrow \infty$ . We use the convention  $\log(0) = -\infty$ .

## 2. Assumptions and Definitions

We write a  $d$ -dimensional random vector  $\mathbf{X}$ , where  $d \geq 2$ , in the form

$$\mathbf{X} = R\mathbf{U}, \quad (2)$$

where  $R = \|\mathbf{X}\|$  and  $\mathbf{U} = \frac{\mathbf{X}}{\|\mathbf{X}\|}$ . Here, we assume that the norm is an  $l_p$  norm with  $1 \leq p < \infty$ . The unit sphere is set  $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$ . Then, we have a geodesic metric on  $\mathbb{S}^{d-1}$ , defined, in detail, in terms of the geodesic distance in Section 2.2. In particular, the space  $\mathbb{S}^{d-1}$  equipped with the geodesic metric is a complete metric space. Open sets, balls, and other topological concepts on  $\mathbb{S}^{d-1}$  are defined using this metric. For example,  $B(\mathbf{x}, \varepsilon) \subset \mathbb{S}^{d-1}$  is an open ball with centre  $\mathbf{x}$  and radius  $\varepsilon > 0$ . The Borel sigma-algebra, the sigma-algebra generated by open sets, is denoted by  $\mathcal{B}$ , e.g.,  $\mathcal{B}(\mathbb{S}^{d-1})$ . In the case of spherical or elliptical distributions, we restrict the norm to be the  $l_2$ -norm denoted by  $\|\cdot\|_2$ . This restriction ensures that ellipsoids have their natural interpretation.

## 2.1. Assumptions

We need the following conditions in the formulation of the results. The required conditions are indicated in the assumptions of each result.

**Assumption 1.**  $R$  is a positive random variable with right-unbounded support. For any  $k \in \mathbb{R}$ , it holds that  $\mathbb{P}(R > k) > 0$  and  $R$  is heavy-tailed in the sense that  $\lim_{k \rightarrow \infty} -\log(\mathbb{P}(R > k))/k = 0$ .

**Assumption 2.**  $\mathbf{U}$  is a random vector on the unit sphere  $\mathbb{S}^{d-1}$  such that the quantity

$$\mathbb{P}(\mathbf{U} \in A | R > k)$$

remains constant for all  $k > k_0$  where  $k_0 > 0$  is a fixed number that does not depend on the Borel set  $A \subset \mathbb{S}^{d-1}$ . In particular, the limiting probability distribution of  $\mathbf{U} | R > k$  exists, as  $k \rightarrow \infty$ . In fact, the limiting distribution on  $\mathbb{S}^{d-1}$  is reached once  $k > k_0$ .

**Assumption 3.** The limit  $\lim_{k \rightarrow \infty} g(k, A)$  exists in  $[0, \infty]$  for all Borel sets  $A \subset \mathbb{S}^{d-1}$ , where  $g : (k_R, \infty) \times \mathcal{B}(\mathbb{S}^{d-1}) \rightarrow \mathbb{R}$  is defined as

$$g(k, A) = \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in A))}{\log(\mathbb{P}(R > k))} \quad (3)$$

and  $k_R = \sup\{k : \mathbb{P}(R > k) = 1\}$ .

**Remark 1.** The condition on heavy-tailedness in Assumption 1 is equivalent to the condition  $\mathbb{E}(e^{sR}) = \infty$  for all  $s > 0$ , which is the usual definition of a (right) heavy-tailed real-valued random variable  $R$ . The distribution of  $\mathbf{U}$  does not have to be uniform on the unit sphere  $\mathbb{S}^{d-1}$ . In fact, the distribution of  $\mathbf{U}$  does not even need to have probability mass in every direction. Assumption 2 provides a simplified setting for the study. This assumption could be alleviated by considering a suitable mode of convergence for the distributions.

Distributions that satisfy the assumptions include random vectors where all the tail distributions in a given direction remain unchanged for sufficiently large values. The values close to origin do not affect the analysis. In particular, the proposed model admits the study of typical asymptotic dependence structures, such as full dependence, strong dependence, and asymptotic independence, as defined in [Lehtomaa and Resnick \(2020\)](#). Some concrete examples are provided in Section 5.1.

The risk function of  $R = \|\mathbf{X}\|$  is defined as

$$h(k) = -\log(\mathbb{P}(R > k)).$$

The function  $h$  provides a benchmark against which risk functions calculated from subsets can be compared. In this sense, the function  $h$  indicates what the heaviness of the tail will be in the set of the riskiest directions  $S$ . The logarithmic transformation is used widely in asymptotic analysis. One reason for the use of this particular transformation of the tail function instead of some other transformation is that, roughly speaking, the function

$$x \mapsto -\log \mathbb{P}(R > x)$$

is concave for heavy-tailed  $R$  and convex for light-tailed  $R$ . Assumption 1 indicates that we operate in the heavy-tailed regime.

Here,  $S$  is the set

$$\left\{ \mathbf{x} \in \mathbb{S}^{d-1} : \lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k | \mathbf{U} \in B(\mathbf{x}, \varepsilon)))}{\log(\mathbb{P}(R > k))} = 1, \right. \\ \left. \mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon)) > 0, \forall \varepsilon > 0 \right\}. \quad (4)$$

The limit in (4) can be written in equivalent forms.

**Lemma 1.** Let Assumptions 1–3 hold and  $\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon)) > 0$ . Then,

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k | \mathbf{U} \in B(\mathbf{x}, \varepsilon)))}{\log(\mathbb{P}(R > k))} = \lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in B(\mathbf{x}, \varepsilon)))}{\log(\mathbb{P}(R > k))}. \quad (5)$$

In addition, the quantity in (5) can be written as

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon) | R > k))}{\log(\mathbb{P}(R > k))} + 1. \quad (6)$$

**Proof.** The proof is based on the properties of the logarithmic function. Since  $\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon)) > 0$  is a constant, it holds that

$$\begin{aligned} \log(\mathbb{P}(R > k | \mathbf{U} \in B(\mathbf{x}, \varepsilon))) &= \log(\mathbb{P}(R > k, \mathbf{U} \in B(\mathbf{x}, \varepsilon))) - \log(\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon))) \\ &\sim \log(\mathbb{P}(R > k, \mathbf{U} \in B(\mathbf{x}, \varepsilon))), \end{aligned}$$

as  $k \rightarrow \infty$ , and the claim of (5) is proven. The other claim follows if we multiply and divide the probability in the numerator of the latter presentation of (5) by  $\mathbb{P}(R > k)$ .  $\square$

It is seen that the quantities in (5) are equal. However, the interpretations of the two forms are slightly different. The form on the left compares the decay rates of two tail functions. The form on the right does not have a tail function in its numerator, but this form makes monotonicity arguments and situations where  $\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon)) = 0$  easier to handle.

If the quantities in (5) equal 1, as they do in the definition of set  $S$ , the quantity in (6) equals 0. So, in principle, there are two ways in which a point can belong to set  $S$ . A point belongs to  $S$  if  $B(\mathbf{x}, \varepsilon)$  has a positive probability under the limiting measure of Assumption 2 or if the limiting probability is 0 but the function  $-\log(\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon) | R > k))$  grows to infinity slowly enough. Since we assume in Assumption 2 that the limit distribution is obtained after some  $k_0$ , the latter possibility is excluded by the assumptions.

Next, we prove general properties for the function  $g$  defined in (3). To simplify notation, we denote, in short, that

$$G(A) = \lim_{k \rightarrow \infty} g(k, A) = \lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in A))}{\log(\mathbb{P}(R > k))}, \quad (7)$$

where  $A \subset \mathbb{S}^{d-1}$  is a Borel set.

**Lemma 2.** Suppose Assumptions 1–3 hold. Then, the following properties hold for the function  $g$  defined in (3) and for the function  $G$  defined in (7). In the statements below, we assume that  $k > k_R$  and  $A \subset \mathbb{S}^{d-1}$  is a Borel set.

- (i)  $g(k, \mathbb{S}^{d-1}) = 1$ ,  $G(\mathbb{S}^{d-1}) = 1$ ,  $g(k, \emptyset) = \infty$  and  $G(\emptyset) = \infty$ .
- (ii)  $g(k, A) \geq 1$  and  $G(A) \geq 1$ .
- (iii) Function  $g$  is monotone in the sense that if  $A \subset D \subset \mathbb{S}^{d-1}$ , where  $D$  is a Borel set, then  $g(k, D) \leq g(k, A)$ . In addition,  $G(D) \leq G(A)$ .
- (iv) Suppose  $A_1, \dots, A_n \subset \mathbb{S}^{d-1}$  are Borel sets. Then,  $G(A_1 \cup \dots \cup A_n) = \min(G(A_1), \dots, G(A_n))$ . In particular,  $\min(G(A), G(A^c)) = 1$ .

**Proof.**

- (i) Since  $\mathbb{P}(R > k, \mathbf{U} \in \emptyset) = 0$ , it follows that  $g(k, \emptyset) = \infty$  and  $G(\emptyset) = \infty$ . The statements for  $\mathbb{S}^{d-1}$  follow from the definition of  $g$ .
- (ii) Since  $A \subset \mathbb{S}^{d-1}$ , the statements follow from (iii) and (i).
- (iii)  $A \subset D$  implies  $\log(\mathbb{P}(R > k, \mathbf{U} \in A)) \leq \log(\mathbb{P}(R > k, \mathbf{U} \in D))$ . Dividing by the negative term  $\log(\mathbb{P}(R > k))$  yields the claims.

(iv) Lemma 1.2.15 in Dembo and Zeitouni (1993) combined with Assumption 3 yields

$$\begin{aligned}
 & -G(A_1 \cup \dots \cup A_n) \\
 \leq & \limsup_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in A_1) + \dots + \mathbb{P}(R > k, \mathbf{U} \in A_n))}{-\log(\mathbb{P}(R > k))} \\
 = & \max \left( \limsup_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in A_1))}{-\log(\mathbb{P}(R > k))}, \dots, \limsup_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in A_n))}{-\log(\mathbb{P}(R > k))} \right) \\
 = & -\min \left( \lim_{k \rightarrow \infty} g(k, A_1), \dots, \lim_{k \rightarrow \infty} g(k, A_n) \right) \\
 = & -\min(G(A_1), \dots, G(A_n)).
 \end{aligned}$$

On the other hand,  $G(A_1 \cup \dots \cup A_n) \leq G(A_i)$  for all  $i = 1, \dots, n$  due to (iii). Thus,  $G(A_1 \cup \dots \cup A_n) = \min(G(A_1), \dots, G(A_n))$ . The fact that  $\min(G(A), G(A^c)) = 1$  follows from (i) because the sets  $A$  and  $A^c$  partition set  $\mathbb{S}^{d-1}$ .

□

In practical applications, the aim is to estimate  $S$  from data. Set  $S$  is a subset of the support of  $\mathbf{U}$  in  $\mathbb{S}^{d-1}$ . The following results show that set  $S$  is not empty.

**Lemma 3.** Suppose Assumptions 1–3 hold and  $G(D) = 1$  for a closed set  $D \subset \mathbb{S}^{d-1}$ . Then,  $D \cap S \neq \emptyset$ .

**Proof.** Suppose  $\mathcal{P}_1, \mathcal{P}_2, \dots$  is a sequence of finite partitions of  $D$ . Assume further that all sets of the partitions are Borel sets and, for  $n \geq 1$ ,  $\mathcal{P}_{n+1}$  is a refinement of  $\mathcal{P}_n$  such that the maximal diameter of the sets in  $\mathcal{P}_n$  converges to 0, as  $n \rightarrow \infty$ . For example, we could use dyadic partitions intersected with  $D$ .

Suppose  $n \geq 1$  is fixed and consider the partition  $\mathcal{P}_n = \{P_{n,1}, P_{n,2}, \dots, P_{n,m_n}\}$  where the number of sets in  $\mathcal{P}_n$  is denoted by  $m_n$ . Based on Part (iv) of Lemma 2, we know that  $\min(G(P_{n,1}), \dots, G(P_{n,m_n})) = G(D) = 1$ . That is, there is  $i_n \in \{1, \dots, m_n\}$ , such that  $G(P_{n,i_n}) = 1$ .

The partition  $\mathcal{P}_{n+1}$  is assumed to be a refinement of  $\mathcal{P}_n$ . So, set  $P_{n,i_n}$  is possibly partitioned into smaller sets, and there is a subset, say  $P_{n+1,i_{n+1}} \subset P_{n,i_n}$ , where  $i_{n+1} \in \{1, \dots, m_{n+1}\}$ , which satisfies  $G(P_{n+1,i_{n+1}}) = 1$ . Recall that the maximal diameters of the partitioning sets are assumed to converge to 0. We see that there is a sequence of sets  $P_{1,i_1} \supset P_{2,i_2} \supset \dots$ , where  $P_{j,i_j} \in \mathcal{P}_j$  and  $G(P_{j,i_j}) = 1$ , for all  $j = 1, 2, \dots$ . Because  $\mathbb{S}^{d-1}$  equipped with the geodesic metric is a complete metric space, there must be a limit point in the sequence of the sets. Let us denote the limit point by  $\mathbf{x}$ . Because set  $D$  is closed, the limit point  $\mathbf{x} \in D$ .

Suppose  $\varepsilon > 0$  is fixed. Suppose  $n$  is so large that the maximal diameter of the sets in Partition  $\mathcal{P}_n$  is less than  $\varepsilon$ . Then, by construction, there exists set  $P_{n,i_n} \in \mathcal{P}_n$ , such that  $P_{n,i_n} \subset B(\mathbf{x}, \varepsilon)$ . Then, by the monotonicity property (iii) of Lemma 2, we reveal that

$$G(B(\mathbf{x}, \varepsilon)) \leq G(P_{n,i_n}) = 1,$$

and the claim is proven because  $\mathbf{x}$  belongs to set  $S$  by the definition of  $S$ . □

**Corollary 1.** Set  $S$  defined in (4) is not empty.

**Proof.** Taking  $D = \mathbb{S}^{d-1}$ , it holds that  $G(D) = 1$  by Part (i) in Lemma 2, and thus, set  $S$  is not empty by Lemma 3. □

We make an assumption on the form of set  $S$  to rule out technically challenging cases that have little impact on practical applications.

**Assumption 4.** Firstly, we assume that there exists an open set in  $S^c$ . Secondly, we assume that  $S$  in (4) can be written as

$$S = \overline{T_1} \cup T_2,$$

where  $T_1$  is an open subset (possibly empty) of  $\mathbb{S}^{d-1}$  and  $T_2$  is a finite collection of individual points (possibly empty) of  $\mathbb{S}^{d-1}$ . We assume that each point in  $T_2$  contains positive probability mass of the limit distribution of  $\mathbf{U} \mid R > k$ , as  $k \rightarrow \infty$ .

**Remark 2.** Assumption 4 implies that not all directions have the same riskiness. The assumption also ensures that  $S$  does not contain continuous subsets in lower dimensions than  $d - 1$ . It admits directly, e.g., distributions where the riskiest direction is concentrated to a cone or a single vector. Even if the original distribution of  $\mathbf{X}$  does not satisfy Assumption 4, it is possible to construct a new approximating distribution by adding a small independent continuous perturbation to vector  $\mathbf{U}$  to obtain a new distribution that satisfies Assumption 4. The perturbation could be, for example, a random variable that has the uniform distribution on a small ball.

If there exists a joint density of  $(R, \mathbf{U})$ , or if  $\mathbf{U}$  is discrete, we can write the conditional risk or hazard function as

$$h(k, \mathbf{x}) = -\log(\mathbb{P}(R > k \mid \mathbf{U} = \mathbf{x})),$$

where  $h(k, \mathbf{x})$  is a positive increasing function for fixed  $\mathbf{x} \in \mathbb{S}^{d-1}$ . The notation admits presenting the conditional risk function in the following simplified way in typical cases.

### Example 1.

1. If the random vector  $\mathbf{X}$  is elliptically distributed with  $R = \|\mathbf{X}\|_2$  and  $\mathbf{U} = \frac{\mathbf{X}}{\|\mathbf{X}\|_2}$ , its conditional risk function can be written as

$$\log(\mathbb{P}(R > k \mid \mathbf{U} = \mathbf{x})) = c(\mathbf{x})h(k).$$

Here,  $c: \mathbb{S}^{d-1} \rightarrow \mathbb{R}$  is a function such that  $c(\mathbf{v}) = c(-\mathbf{v})$  holds for all  $\mathbf{v} \in \mathbb{S}^{d-1}$ . The function  $c$  is a continuous map of  $\mathbb{S}^{d-1}$  to an interval. In general, we set  $h(k)$  such that  $\min c(\mathbf{x}) = 1$ , so  $h(k)$  is the risk function in the riskiest direction. In the special case, where the distribution of  $\mathbf{X}$  is spherical  $c(\mathbf{x})$  is constant.

2. There can be different tail behaviour in different subsets of  $\mathbb{S}^{d-1}$ . If there exists a finite partition  $A_1, \dots, A_m$  of  $\mathbb{S}^{d-1}$ , such that the risk function does change given  $\mathbf{x} \in A_j$ , we can write

$$h(k, \mathbf{x}) = \sum_{j=1}^m h_j(k) \mathbb{1}(\mathbf{x} \in A_j),$$

where  $h_j(k)$  refers to the risk function in the direction of set  $A_j$ .

### 2.2. Subsets of the Unit Sphere

For  $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{d-1}$ , we define the geodesic metric  $\text{dist}(\mathbf{x}, \mathbf{y})$  as

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \{\text{length of the shortest geodesic connecting } \mathbf{x} \text{ and } \mathbf{y}\}.$$

In this metric, open balls are subsets of  $\mathbb{S}^{d-1}$  denoted by  $B(\mathbf{x}, r)$ , where vector  $\mathbf{x} \in \mathbb{S}^{d-1}$  is the centre of the open ball and  $r > 0$ . So,  $B(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{S}^{d-1} : \text{dist}(\mathbf{x}, \mathbf{y}) < r\}$ . The corresponding closed ball is denoted by  $\overline{B}(\mathbf{x}, r)$ . Note that the shape of the ball  $B(\mathbf{x}, r)$  depends on the used norm.

**Definition 1.** For any set  $A \subset \mathbb{S}^{d-1}$ , we call set

$$A^\delta = \{\mathbf{x} \in \mathbb{S}^{d-1} : \text{dist}(\mathbf{x}, A) < \delta\}$$



the geodesic  $\delta$ -swelling of set  $A$ . Here,  $\text{dist}(\mathbf{x}, A)$  is the geodesic distance of  $\mathbf{x}$  to set  $A$ ,  $\text{dist}(\mathbf{x}, A) = \inf_{\mathbf{a} \in A} \text{dist}(\mathbf{x}, \mathbf{a})$ .

By

$$\text{dist}_H(A, B) = \max \left\{ \sup_{\mathbf{a} \in A} \text{dist}(\mathbf{a}, B), \sup_{\mathbf{b} \in B} \text{dist}(\mathbf{b}, A) \right\},$$

we denote the Hausdorff distance, where  $\text{dist}(\mathbf{a}, B)$  is defined as in Definition 1.

### 3. Minimal Set of Riskiest Directions

Our aim is to find the riskiest directions. We search for the minimal set that dominates the tail behaviour of the random vector  $\mathbf{X}$  in the sense of (4). To this end, we need to identify sets  $A \subset \mathbb{S}^{d-1}$  for which, given  $\delta > 0$ , the inequality

$$\log(\mathbb{P}(R > k, \mathbf{U} \in A)) > \log(\mathbb{P}(R > k, \mathbf{U} \in (A^\delta)^c))$$

holds for all values of  $k$  large enough. The inequality demands a positive probability measure of  $A$ .

For the next result, we define the collection  $\mathcal{A}$  of testing sets  $A \subset \mathbb{S}^{d-1}$  as follows. Set  $A$  is an element of  $\mathcal{A}$  if  $A$  is a finite union of open balls, such that for all  $\mathbf{x} \in A^c$  and for all  $\varepsilon > 0$ , the open ball  $B(\mathbf{x}, \varepsilon)$  contains an open ball  $B$  that belongs to  $A^c$ . Note that point  $\mathbf{x}$  does not have to be in set  $B$ . In particular, this guarantees that  $A^c$  does not contain any isolated points.

**Theorem 1.** Let  $\mathbf{X} = R\mathbf{U} \in \mathbb{R}^d, d \geq 2$ , be such that Assumptions 1–4 hold.

$$\tilde{S} = \cap \left\{ A \in \mathcal{A} : \lim_{k \rightarrow \infty} g(k, A) < \lim_{k \rightarrow \infty} g(k, A^c) \right\}. \quad (8)$$

Then,  $S = \text{cl}(\tilde{S})$ , where  $S$  is as in (4), and  $\text{cl}$  refers to the closure of the set. Furthermore, for all  $\delta > 0$ ,

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in \tilde{S}^\delta))}{\log(\mathbb{P}(R > k))} = 1 \quad (9)$$

and

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in (\tilde{S}^\delta)^c))}{\log(\mathbb{P}(R > k))} > 1. \quad (10)$$

**Proof.** The proof of the theorem is performed in steps.

1. WE CLAIM  $S \subset \text{cl}(\tilde{S})$ .

Recall that under Assumption 4, set  $S$  can be written as the union of  $\overline{\text{int}(S)}$  and a finite number of individual points. We have two cases to cover.

First, we study the case where set  $T_2$  of Assumption 4 contains a point. Suppose there is an individual point  $\mathbf{x} \in \mathbb{S}^{d-1}$ , such that

$$1 = \lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k | \mathbf{U} = \mathbf{x}))}{\log(\mathbb{P}(R > k))} = \lim_{k \rightarrow \infty} g(k, \{\mathbf{x}\}), \quad (11)$$

where the latter equation follows from Lemma 1. Then,  $\mathbf{x}$  cannot belong to  $A^c$  for any set  $A \in \mathcal{A}$  that satisfies the inequality in (8). To see this, assume, on the contrary, that  $\mathbf{x}$  belongs to  $A^c$  for some set  $A \in \mathcal{A}$ . Then, due to monotonicity mentioned in (iii) of Lemma 2,

$$\lim_{k \rightarrow \infty} g(k, A^c) \leq \lim_{k \rightarrow \infty} g(k, \{\mathbf{x}\}) = 1$$

by Equality (11). Since  $\lim_{k \rightarrow \infty} g(k, A)$  cannot be strictly less than 1 by Part (ii) of Lemma 2, the inequality cannot be true if  $\mathbf{x}$  is in the complement of a testing set  $A \in \mathcal{A}$ .

We conclude that  $\mathbf{x}$  must belong to all testing sets  $A \in \mathcal{A}$  that satisfy the inequality in (8). So,  $\mathbf{x} \in \tilde{S}$ .

Next, we consider the case where set  $T_1$  of Assumption 4 contains a point. Suppose  $\mathbf{x} \in \text{int}(S)$ . Then, there is a number  $\varepsilon > 0$  such that  $B(\mathbf{x}, \varepsilon) \subset \text{int}(S)$ . Let  $A \in \mathcal{A}$ . We show that  $\mathbf{x}$  cannot be in  $A^c$  for a set  $A$  that satisfies the inequality in (8). Assume, on the contrary, that  $\mathbf{x} \in A^c$ . In this situation, we can find a small ball that is entirely in the intersection  $B(\mathbf{x}, \varepsilon) \cap A^c$ . To see this, let  $\varepsilon' = \varepsilon/2$ . By the definition of the testing sets in  $\mathcal{A}$ , the ball  $B(\mathbf{x}, \varepsilon')$  contains another open ball, say,  $B(\mathbf{x}', \varepsilon'')$ , which is contained in  $A^c$ . Since  $B(\mathbf{x}, \varepsilon') \subset B(\mathbf{x}, \varepsilon)$ , we see that  $B(\mathbf{x}', \varepsilon'') \subset B(\mathbf{x}, \varepsilon) \cap A^c \subset \text{int}(S) \cap A^c$ . So, because  $\mathbf{x}' \in S$ , the limit in (4) applied for radius  $\varepsilon''$  states that

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k | \mathbf{U} \in B(\mathbf{x}', \varepsilon''))) }{\log(\mathbb{P}(R > k))} = \lim_{k \rightarrow \infty} g(k, B(\mathbf{x}', \varepsilon'')) = 1.$$

So, due to the monotonicity mentioned in (iii) of Lemma 2,

$$\lim_{k \rightarrow \infty} g(k, A^c) \leq \lim_{k \rightarrow \infty} g(k, B(\mathbf{x}', \varepsilon'')) = 1.$$

In conclusion, any testing set  $A \in \mathcal{A}$  that does not contain  $\mathbf{x}$  cannot satisfy the inequality in (8). So, all testing sets that satisfy the inequality must contain  $\mathbf{x}$ . So,  $\mathbf{x} \in \tilde{S}$ .

The above deductions imply, using the notation of Assumption 4, that  $\text{int}(S) \cup T_2 \subset \tilde{S}$ , which implies  $\text{cl}(\text{int}(S) \cup T_2) = S \subset \text{cl}(\tilde{S})$ , and the claim is proven.

2. WE CLAIM  $\text{cl}(\tilde{S}) \subset S$ .

The claim is equivalent to  $S^c \subset \text{cl}(\tilde{S})^c$ . Let  $\mathbf{x} \in S^c$ . Then, by the definition of  $S$  in (4), either there exists  $\varepsilon > 0$ , such that  $\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon)) = 0$ , or there exists  $\varepsilon > 0$ , such that

$$1 < \lim_{k \rightarrow \infty} g(k, B(\mathbf{x}, \varepsilon)). \quad (12)$$

In the latter case, Inequality (12) also holds when  $\varepsilon$  is replaced by any  $\varepsilon' < \varepsilon$  due to the monotonicity; see (iii) in Lemma 2.

In order to show that  $\mathbf{x} \in \tilde{S}^c$ , it suffices to find one testing set  $A \in \mathcal{A}$ , such that  $\mathbf{x} \in A^c$  and  $A$  satisfies the inequality of (8). This is because set  $\tilde{S}$  is the intersection of all testing sets that satisfy the inequality.

Let the number  $\varepsilon$  be such that  $\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon)) = 0$  or (12) holds. Now, formally setting  $A' = \mathbb{S}^{d-1} \setminus B(\mathbf{x}, \varepsilon)$  fulfils the inequality of (8), but this  $A'$  is not a member of the collection  $\mathcal{A}$ . However, we can construct a set  $A \in \mathcal{A}$  using a finite number of open balls, such that  $A$  covers set  $\mathbb{S}^{d-1} \setminus B(\mathbf{x}, \varepsilon)$  but does not intersect set  $B(\mathbf{x}, \varepsilon/2)$ .

With this set  $A$ , we have

$$\lim_{k \rightarrow \infty} g(k, A) < \lim_{k \rightarrow \infty} g(k, A^c) \quad (13)$$

because  $A^c \subset B(\mathbf{x}, \varepsilon)$ . More precisely, by monotonicity,

$$\lim_{k \rightarrow \infty} g(k, A^c) \geq \lim_{k \rightarrow \infty} g(k, B(\mathbf{x}, \varepsilon)) > 1.$$

The limit  $\lim_{k \rightarrow \infty} g(k, A)$  equals 1 according to (iv) of Lemma 2 and so Inequality (13) holds. In conclusion,  $\mathbf{x}$  belongs to the complement of this  $A$  and consequently  $\mathbf{x} \in \tilde{S}^c$ . We have shown that  $S^c \subset \tilde{S}^c$ , which is equivalent to  $\tilde{S} \subset S$ , which implies  $\text{cl}(\tilde{S}) \subset \text{cl}(S) = S$ .

3. WE CLAIM THAT (9) AND (10) HOLD.

Let  $\delta > 0$  be fixed. Through Parts 1–2 of the proof, we know that

$$S \subset \tilde{S}^\delta. \quad (14)$$



Assume, on the contrary, to Claim (10) that

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k, \mathbf{U} \in (\tilde{S}^\delta)^c))}{\log(\mathbb{P}(R > k))} = 1. \quad (15)$$

In particular, set  $(\tilde{S}^\delta)^c$  is closed, so it must contain a point of  $S$  by Corollary 1. This is impossible by (14), and thus (15) cannot hold as an equality. So, Inequality (10) and, consequently, by (iv) of Lemma 2, Equality (9) holds.  $\square$

In the case of elliptically distributed random vectors, the minimal set that dominates the tail behaviour of the random vector might consist only of singletons that do not have probability mass. One can still approximate the distribution using the method of Remark 2. In this example, we use the  $l_2$ -norm.

**Example 2.** Let the random vector  $\mathbf{X}$  be elliptically distributed, such that  $h(k, \mathbf{u}) = c(\mathbf{u})h(k)$ , where  $c$  is a continuous function on  $\mathbb{S}^{d-1}$  that achieves its minimum only at  $\mathbf{x}$  and  $-\mathbf{x} \in \mathbb{S}^{d-1}$ , and assume that  $\mathbf{x}$  and  $-\mathbf{x}$  are points in set  $T_2$  of Assumption 4. Then, for all  $\varepsilon > 0$ , it holds that

$$\lim_{k \rightarrow \infty} g(k, (B(\mathbf{x}, \varepsilon) \cup B(-\mathbf{x}, \varepsilon)) \cap \mathbb{S}^{d-1}) < \lim_{k \rightarrow \infty} g(k, \mathbb{S}^{d-1} \setminus (B(\mathbf{x}, \varepsilon) \cup B(-\mathbf{x}, \varepsilon))).$$

So,  $S = \{\mathbf{x}, -\mathbf{x}\}$ . Choosing the risk function of  $R$  to be  $h(k)$ , the minimum  $\min c(\mathbf{v})$  equals one.

**Corollary 2.** Let  $\mathbf{X} = R\mathbf{U}$  be such that Assumptions 1–4 hold and  $A \subset \mathbb{S}^{d-1}$  is a Borel set. If there exists  $\delta > 0$  such that  $\tilde{S}^\delta \subset A$ , it holds that

$$\lim_{k \rightarrow \infty} g(k, A) < \lim_{k \rightarrow \infty} g(k, A^c). \quad (16)$$

**Proof.** The claim follows from (9) and (10) together, with the monotonicity of the function  $g$ .  $\square$

#### 4. Towards Estimators

In this section, we introduce procedures that can be used to form estimators for set  $S$  based on data. We do not present such estimators explicitly here, but the results can be used as a theoretical basis for this work. Throughout the section, we use the  $l_2$ -norm, so the geodesic distance or great ball distance on the unit sphere  $\mathbb{S}^{d-1}$  is defined as  $\text{dist}(\mathbf{x}, \mathbf{y}) = \arccos(\mathbf{x} \cdot \mathbf{y})$ , where  $\mathbf{x} \cdot \mathbf{y}$  is the dot product of  $\mathbf{x}$  and  $\mathbf{y}$ . The metric  $\text{dist}$  is discussed in detail, for instance, in Proposition 2.1 of [Bridson and Haefliger \(2013\)](#).

The following lemmas are auxiliary results for Theorem 2.

**Lemma 4.** Let  $\mathbb{S}^{d-1}$  be the unit sphere of  $\mathbb{R}^d$  equipped with the  $l_2$ -norm, and for  $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{d-1}$ , let  $\text{dist}(\mathbf{x}, \mathbf{y}) = \arccos(\mathbf{x} \cdot \mathbf{y})$  be the geodesic distance or the great ball distance on  $\mathbb{S}^{d-1}$ . We define  $\bar{B}(\mathbf{x}, 0) = \{\mathbf{x}\}$  for  $\mathbf{x} \in \mathbb{S}^{d-1}$ .

Then, for fixed  $\mathbf{x} \in \mathbb{S}^{d-1}$  and  $0 < r \leq \pi$ , the balls  $B(\mathbf{x}, r)$  and  $\bar{B}(-\mathbf{x}, \pi - r)$  partition the unit sphere.

**Proof.** Since  $\mathbf{x} \in \mathbb{S}^{d-1}$ , it holds  $1 = \|\mathbf{x}\|_2 = \|\mathbf{x}\|_2$  so  $-\mathbf{x} \in \mathbb{S}^{d-1}$ . To prove the claim, we show that  $\bar{B}(-\mathbf{x}, \pi - r)$  is the complement of  $B(\mathbf{x}, r)$  in  $\mathbb{S}^{d-1}$ . Since  $B(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{S}^{d-1} : \text{dist}(\mathbf{x}, \mathbf{y}) < r\}$ , its complement is set  $\{\mathbf{y} \in \mathbb{S}^{d-1} : \text{dist}(\mathbf{x}, \mathbf{y}) \geq r\}$ . The condition  $\text{dist}(\mathbf{x}, \mathbf{y}) \geq r$  is, by definition,

$$\arccos(\mathbf{x} \cdot \mathbf{y}) \geq r. \quad (17)$$

Due to the fact that for all  $x \in [-1, 1]$ ,  $\arccos(x) + \arccos(-x) = \pi$ , the condition (17) is equivalent to  $\arccos(-\mathbf{x} \cdot \mathbf{y}) \leq \pi - r$  because  $\mathbf{x}$  and  $\mathbf{y}$  are unit vectors. The last expression can be written as  $\text{dist}(-\mathbf{x}, \mathbf{y}) \leq \pi - r$ . So  $\bar{B}(-\mathbf{x}, \pi - r)$  is the complement of  $B(\mathbf{x}, r)$  in  $\mathbb{S}^{d-1}$ .  $\square$

**Lemma 5.** Let  $\mathbb{S}^{d-1}$  be the unit sphere of  $\mathbb{R}^d$  equipped with the  $l_2$ -norm, and for  $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{d-1}$ , let  $\text{dist}(\mathbf{x}, \mathbf{y}) = \arccos(\mathbf{x} \cdot \mathbf{y})$  be the geodesic distance.

Then, for  $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{d-1}, \mathbf{x} \neq \mathbf{y}, \mathbf{x} \neq -\mathbf{y}$  there exists  $\delta > 0$  such that the intersection  $B(\mathbf{y}, \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2) \cap B(\mathbf{x}, \delta)$  contains an open set of  $\mathbb{S}^{d-1}$  and  $\overline{B}(\mathbf{y}, \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2)^c \cap B(\mathbf{x}, \delta)$  contains an open set of  $\mathbb{S}^{d-1}$ .

**Proof.** Since  $\mathbf{y} \neq -\mathbf{x}$  and  $\mathbf{y} \neq \mathbf{x}$ , it holds that  $0 < \text{dist}(\mathbf{x}, \mathbf{y}) < \pi$  due to the definition of the great ball distance. Take  $\delta < \text{dist}(-\mathbf{x}, \mathbf{y})$ . By the choice of  $\delta$ ,  $B(\mathbf{y}, \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2)$ ,  $B(\mathbf{x}, \delta)$  and  $\overline{B}(\mathbf{y}, \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2)^c$  are proper subsets of the unit sphere. Since all three sets are open, it remains to show that there is a point in each of the intersections of the claim. We recover the points explicitly.

Since  $\arccos(x) + \arccos(-x) = \pi$ , it holds  $\pi = \text{dist}(\mathbf{x}, -\mathbf{x}) = \text{dist}(\mathbf{x}, \mathbf{v}) + \text{dist}(-\mathbf{x}, \mathbf{v})$  for any  $\mathbf{v} \in \mathbb{S}^{d-1}$ . For  $\mathbf{y} \neq \mathbf{x}, \mathbf{y} \neq -\mathbf{x}$  there exists a unique minimising geodesic between  $\mathbf{x}$  and  $\mathbf{y}$ . For all points  $\mathbf{z}$  that lie on this minimising geodesic between  $\mathbf{x}$  and  $-\mathbf{y}$  we have that  $\text{dist}(\mathbf{x}, -\mathbf{y}) = \text{dist}(\mathbf{x}, \mathbf{z}) + \text{dist}(\mathbf{z}, -\mathbf{y})$ .

Taking  $\mathbf{z}$  on this geodesic such that  $\text{dist}(\mathbf{z}, \mathbf{x}) = \frac{1}{4}\delta$ , it holds, by definition, that  $\mathbf{z} \in B(\mathbf{x}, \delta)$ . Because

$$\begin{aligned} \text{dist}(\mathbf{y}, \mathbf{z}) &= \pi - \text{dist}(\mathbf{z}, -\mathbf{y}) = \pi - (\text{dist}(\mathbf{x}, -\mathbf{y}) - \text{dist}(\mathbf{x}, \mathbf{z})) \\ &= \text{dist}(\mathbf{x}, \mathbf{y}) + \frac{1}{4}\delta < \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2, \end{aligned}$$

we also have  $\mathbf{z} \in B(\mathbf{y}, \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2)$ . In conclusion,  $\mathbf{z} \in B(\mathbf{y}, \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2) \cap B(\mathbf{x}, \delta)$ .

On the other hand, taking  $\mathbf{z}'$  on the same minimising geodesic between  $\mathbf{x}$  and  $-\mathbf{y}$ , such that  $\text{dist}(\mathbf{z}', \mathbf{x}) = \frac{3}{4}\delta$ , it holds by definition that  $\mathbf{z}' \in B(\mathbf{x}, \delta)$ . By similar calculations as above, it holds that

$$\text{dist}(\mathbf{y}, \mathbf{z}') = \text{dist}(\mathbf{x}, \mathbf{y}) + \text{dist}(\mathbf{x}, \mathbf{z}') > \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2,$$

so  $\mathbf{z}' \in \overline{B}(\mathbf{y}, \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2)^c$  and hence  $\mathbf{z}' \in \overline{B}(\mathbf{y}, \text{dist}(\mathbf{x}, \mathbf{y}) + \delta/2)^c \cap B(\mathbf{x}, \delta)$ . Thus, both intersections contain a point, and therefore, an open set.  $\square$

### Algorithm

In this section, we present an algorithm to find the minimal set  $S$  that dominates the tail behaviour of the studied random vectors under suitable assumptions. To this end, we study the function  $G$  defined in (7) for different sets. Due to Theorem 1, the minimal set  $S$  in (4) that dominates the tail behaviour of the random vector is contained in the intersection of all testing sets that fulfil the inequality in Condition (8). We present a theoretical procedure for finding the set of the riskiest directions.

We define for Algorithm 1 below a map  $\mathbf{v} \mapsto A_{\mathbf{v}}$ , where  $A_{\mathbf{v}}$  is an open set and  $\mathbf{v}$  is an element on the unit sphere  $\mathbb{S}^{d-1}$ . The algorithm for finding the minimal set  $S$  that dominates the tail behaviour of the random vectors is presented in two steps. In the theoretical setup, we perform Step 1 for all  $\mathbf{v} \in \mathbb{S}^{d-1}$ . In practice, we must use a finite collection of vectors  $\mathbf{v} \in \mathbb{S}^{d-1}$ .

**Algorithm 1.** The algorithm has two steps.

1. Let  $\mathbf{v} \in \mathbb{S}^{d-1}$ .

If, for some  $r < \pi$ , it holds that  $G(B(\mathbf{v}, r)^c) > 1$  and  $G(B(\mathbf{v}, r)) = 1$  then define  $A_{\mathbf{v}} = B(\mathbf{v}, r_{\mathbf{v}})$ , where  $r_{\mathbf{v}}$  is the smallest radius fulfilling the condition

$$\lim_{k \rightarrow \infty} g(k, B(\mathbf{v}, r_{\mathbf{v}})^c) > 1. \quad (18)$$

In other words,

$$r_{\mathbf{v}} = \inf\{r > 0 : \lim_{k \rightarrow \infty} g(k, B(\mathbf{v}, r)^c) > 1\}$$

and  $B(\mathbf{v}, r_{\mathbf{v}})$  is the smallest ball centred around  $\mathbf{v}$  that contains  $S$ .

If  $G(B(\mathbf{v}, r)^c) = 1$ , for all  $0 < r < \pi$ , set  $A_{\mathbf{v}} = \mathbb{S}^{d-1}$  and if  $G(B(\mathbf{v}, r)^c) > 1$  for all  $0 < r < \pi$ , set  $A_{\mathbf{v}} = \emptyset$ .

2. Set

$$\hat{S} = \bigcap_{\mathbf{v} \in \mathbb{S}^{d-1}} A_{\mathbf{v}}.$$

The sets  $A_{\mathbf{v}} = B(\mathbf{v}, r_{\mathbf{v}})$  are open balls and belong by definition to the set of testing sets  $\mathcal{A}$ . Furthermore, the entire unit sphere and the empty set are open sets as well and thus belong to  $\mathcal{A}$ . Due to Lemma 4,  $B(\mathbf{v}, r_{\mathbf{v}})^c = \overline{B}(-\mathbf{v}, \pi - r_{\mathbf{v}})$ , Equation (18) can be rewritten in equivalent form

$$\lim_{k \rightarrow \infty} g(k, \overline{B}(-\mathbf{v}, \pi - r_{\mathbf{v}})) > 1.$$

So,  $r_{\mathbf{v}}$  is the smallest radius such that  $G(\overline{B}(-\mathbf{v}, \pi - r_{\mathbf{v}})) > 1$ . The condition that  $G(B(\mathbf{v}, r)^c) > 1$  for all  $0 < r < \pi$  implies that the neighbourhood of  $\mathbf{v}$  does not belong to  $S$  and  $S$  contains only of the point  $\mathbf{v}$ .

The following lemma shows the connection between the choice of set  $A_{\mathbf{v}}$  and vector  $-\mathbf{v}$ , which points to the opposite direction of  $\mathbf{v}$ .

**Lemma 6.** Suppose Assumptions 1–4 hold. Then, it holds for any  $\mathbf{v} \in \mathbb{S}^{d-1}$  and its corresponding set  $A_{\mathbf{v}}$  defined in Algorithm 1 that  $A_{\mathbf{v}} = \mathbb{S}^{d-1}$  if, and only if,  $-\mathbf{v} \in S$ .

**Proof.** To prove  $A_{\mathbf{v}} = \mathbb{S}^{d-1}$  is equivalent to  $-\mathbf{v} \in S$ , we show that  $-\mathbf{v} \in S$  implies  $A_{\mathbf{v}} = \mathbb{S}^{d-1}$  and  $-\mathbf{v} \notin S$  results in  $A_{\mathbf{v}} \subsetneq \mathbb{S}^{d-1}$ .

Let  $-\mathbf{v} \in S$ . Then, by the definition of  $S$  for all  $\varepsilon > 0$ , it holds that  $\mathbb{P}(\mathbf{U} \in B(-\mathbf{v}, \varepsilon)) > 0$ , and

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k | \mathbf{U} \in B(-\mathbf{v}, \varepsilon)))}{\log(\mathbb{P}(R > k))} = 1,$$

which is equivalent to  $G(B(-\mathbf{v}, \varepsilon)) = 1$  due to Lemma 1. As a consequence, the algorithm chooses  $A_{\mathbf{v}} = \mathbb{S}^{d-1}$  because  $\varepsilon > 0$  does not exist, such that  $G(B(-\mathbf{v}, \varepsilon)) > 1$  and  $B(-\mathbf{v}, \varepsilon) = \overline{B}(\mathbf{v}, \pi - \varepsilon)^c$  by Lemma 4.

On the other hand, let  $-\mathbf{v} \notin S$ . Then, either  $\mathbb{P}(\mathbf{U} \in B(-\mathbf{v}, \varepsilon)) = 0$  for some  $\varepsilon > 0$ , or for all  $\varepsilon > 0$  it holds that  $\mathbb{P}(\mathbf{U} \in B(-\mathbf{v}, \varepsilon)) > 0$  and there exists  $\varepsilon' > 0$  such that

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k | \mathbf{U} \in B(-\mathbf{v}, \varepsilon')))}{\log(\mathbb{P}(R > k))} > 1. \quad (19)$$

In the first case,  $\mathbb{P}(\mathbf{U} \in B(-\mathbf{v}, \varepsilon)) = 0$  implies  $G(B(-\mathbf{v}, \varepsilon)) > 1$ , and in the second case, Inequality (19) is equivalent to  $G(B(-\mathbf{v}, \varepsilon')) > 1$  by Lemma 1. Due to Lemma 4,  $B(-\mathbf{v}, \varepsilon') = \overline{B}(\mathbf{v}, \pi - \varepsilon')^c$  so  $G(\overline{B}(\mathbf{v}, \pi - \varepsilon')^c) > 1$  and by Lemma 2 it holds that  $G(\overline{B}(\mathbf{v}, \pi - \varepsilon')) = 1$ . Hence,  $r_{\mathbf{v}} \leq \pi - \varepsilon'$  and  $A_{\mathbf{v}} = B(\mathbf{v}, r_{\mathbf{v}}) \subsetneq \mathbb{S}^{d-1}$  or  $A_{\mathbf{v}} = \emptyset$ .  $\square$

The estimator does not detect all possible sets. Recall that, in general, we assume in Assumption 4 that  $S = \overline{T}_1 \cup T_2$ , where  $T_1$  is an open subset of  $\mathbb{S}^{d-1}$  and  $T_2$  is a finite collection of individual points. For example, if  $S = T_2$  or  $S = \overline{T}_1 \cup T_2$ , and  $T_2$  is not empty, Algorithm 1 will not detect set  $T_2$ . If  $S$  contains only a singleton, so  $S = \{\mathbf{v}\}$  for some  $\mathbf{v} \in \mathbb{S}^{d-1}$ , it holds that  $G(B(\mathbf{v}, \varepsilon)) = 1$  and  $G(B(\mathbf{v}, \varepsilon)^c) > 1$  for all  $\varepsilon > 0$ . Then, the algorithm sets  $A_{\mathbf{v}} = \emptyset$  so the estimator is empty. In general, it can be seen that the algorithm does not detect any finite number of individual points in  $S$ . However, the procedure described in Remark 2 can be used to modify data sets in order to avoid problems in practice.

Estimator  $\hat{S}$  has the capacity to detect sets  $S$  that are not necessarily convex or even connected. Set  $S$  can be, for example, a disjointed union of open sets.

**Example 3.** A classical football is made of 12 black pentagons and 20 white hexagons. Assume that the directions  $\mathbf{U}$  of the random vectors are uniformly distributed on the surface of the football. Furthermore, assume that random variables  $R$  connected with open black pentagons have a much heavier tail than random variables  $R$  connected with closed white hexagons, such that  $G(\text{white part of football}) > 1$  and  $G(\text{black part of the football}) = 1$ . If we choose  $\mathbf{v} \in \mathbb{S}^2$ , such that  $-\mathbf{v}$  points towards the centre of a black pentagon, it holds that  $G(B(-\mathbf{v}, \varepsilon)) = 1$  for any  $\varepsilon > 0$ , so the algorithm sets  $A_{\mathbf{v}} = \mathbb{S}^2$ . If we choose  $\mathbf{v} \in \mathbb{S}^2$ , such that  $-\mathbf{v}$  points in the centre of a white hexagon, it holds for its closed inscribed circle  $\bar{B}(-\mathbf{v}, r)$  that  $G(\bar{B}(-\mathbf{v}, r)) > 1$ , so  $A_{\mathbf{v}}$  does not contain this closed inscribed circle, and thus  $\hat{S}$  does not contain it. If we choose  $\mathbf{v} \in \mathbb{S}^2$ , such that  $-\mathbf{v}$  is the centre of an edge of two white hexagons, it holds that  $G(\bar{B}(-\mathbf{v}, r)) > 1$ , where  $r$  is half of the length of the edge. Therefore, the edge is not included in set  $A_{\mathbf{v}}$ , and thus, the edge is not included in  $\hat{S}$ . All in all, the intersection of sets  $A_{\mathbf{v}}$ , where  $\mathbf{v}$  are such that  $-\mathbf{v}$  points either in the direction of the centre of a white hexagon or in the direction of the centre of an edge between two white hexagons, returns a set that is not connected. Regarding the intersection over all  $\mathbf{v} \in \mathbb{S}^2$ , Algorithm 1 would return the union of the closed black pentagons as  $\hat{S}$ .

To avoid the problem with individual points, we propose a simplifying assumption for  $S$ .

**Theorem 2.** Suppose Assumptions 1–4 with  $S = \overline{T_1}$ . In particular, set  $S$  does not contain any individual points.

Then, it holds that

$$S = \text{cl}(\hat{S}),$$

where  $\hat{S}$  is as in Algorithm 1.

**Proof.** The proof of the theorem is performed in steps.

1. We claim  $S \subset \text{cl}(\hat{S})$ .

We show that  $\mathbf{x} \in \text{int}(S)$  implies  $\mathbf{x} \in \text{cl}(\hat{S})$ , and then take the closure of the sets to prove the claim.

Note first that, by Assumption 4,  $S$  is a proper subset of  $\mathbb{S}^{d-1}$ . Let  $\mathbf{x}$  be in the interior of  $S$ . Then, there exists some  $\delta > 0$  such that  $B(\mathbf{x}, \delta) \subset S$ . Furthermore, for all  $\varepsilon > 0$ , it holds that

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k | \mathbf{U} \in B(\mathbf{x}, \varepsilon)))}{\log(\mathbb{P}(R > k))} = 1,$$

which is equivalent to  $G(B(\mathbf{x}, \varepsilon)) = 1$  due to Lemma 1.

We need to show that  $\mathbf{x} \in A_{\mathbf{v}}$  for all  $\mathbf{v} \in \mathbb{S}^{d-1}$ . In the algorithm, set  $A_{\mathbf{v}}$  can be the empty set, the unit sphere, or an open ball with centre  $\mathbf{v}$ . If  $A_{\mathbf{v}} = \mathbb{S}^{d-1}$ , it contains  $\mathbf{x}$  by default.

Let  $\mathbf{v} \in \mathbb{S}^{d-1}$  be fixed. We show that  $\mathbf{x} \in A_{\mathbf{v}}$ . There are different cases to consider.

(a) If  $\mathbf{v} = \mathbf{x}$ , set  $B(\mathbf{x}, \delta/2)^c$  contains an open subset of  $S$  since

$$\emptyset \neq B(\mathbf{x}, \delta/2)^c \cap B(\mathbf{x}, \delta) \subset B(\mathbf{x}, \delta) \subset S,$$

so  $G(B(\mathbf{x}, \delta/2)^c) = 1$ , and thus,  $r_{\mathbf{x}} \geq \delta/2$ . It follows that  $A_{\mathbf{x}}$  cannot be empty and  $\mathbf{x} \in A_{\mathbf{x}}$ .

(b) If  $\mathbf{v} = -\mathbf{x}$ , it holds that  $A_{\mathbf{v}} = \mathbb{S}^{d-1}$  according to Lemma 6, so  $\mathbf{x} \in A_{-\mathbf{x}}$ .

(c) If  $\mathbf{v} \in B(\mathbf{x}, \delta)$ ,  $\mathbf{v} \neq \mathbf{x}$ , there exists some  $\delta' > 0$  such that  $B(\mathbf{v}, \delta') \subset B(\mathbf{x}, \delta)$ , so both sets  $B(\mathbf{v}, \delta')$  and its complement contain an open subset of  $S$ , and thus,  $r_{\mathbf{v}} \geq \delta'$ , so  $A_{\mathbf{v}} \neq \emptyset$ .

It holds, according to Lemma 5, that both sets  $B(\mathbf{x}, \delta) \cap B(\mathbf{v}, \text{dist}(\mathbf{x}, \mathbf{v}) + \delta/2)$  and  $B(\mathbf{x}, \delta) \cap B(\mathbf{v}, \text{dist}(\mathbf{x}, \mathbf{v}) + \delta/2)^c$  are not empty. By the monotonicity of  $G$ , it holds that  $G(B(\mathbf{x}, \delta) \cap B(\mathbf{v}, \text{dist}(\mathbf{x}, \mathbf{v}) + \delta/2)^c) \leq G(B(\mathbf{x}, \delta)) = 1$ , so  $r_{\mathbf{v}} \geq \text{dist}(\mathbf{x}, \mathbf{v}) + \delta/2$  and  $\mathbf{x} \in A_{\mathbf{v}}$ .

- (d) If  $\mathbf{v} \notin B(\mathbf{x}, \delta)$ ,  $\mathbf{v} \neq -\mathbf{x}$ , there exists, according to Lemma 5, a number  $\delta' < \min(\delta, \text{dist}(-\mathbf{x}, \mathbf{v}))$  such that the intersections

$$B(\mathbf{v}, \text{dist}(\mathbf{x}, \mathbf{v}) + \delta'/2) \cap B(\mathbf{x}, \delta)$$

and

$$B(\mathbf{v}, \text{dist}(\mathbf{x}, \mathbf{v}) + \delta'/2)^c \cap B(\mathbf{x}, \delta)$$

are not empty, and, in particular, both contain an open subset. Since  $B(\mathbf{x}, \delta) \subset S$ , it holds that  $A_{\mathbf{v}}$  cannot be empty. With a similar deduction as in Part 1c, it follows that  $\mathbf{x} \in A_{\mathbf{v}}$ .

2. We claim  $\text{cl}(\hat{S}) \subset S$ .

Let  $\mathbf{x} \notin S$ . We show that there exists  $\mathbf{v}$ , such that  $\mathbf{x} \notin A_{\mathbf{v}}$ . More specifically, we can choose  $\mathbf{v} = -\mathbf{x}$ .

If  $\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon)) = 0$  for some  $\varepsilon > 0$ , it holds that  $G(B(\mathbf{x}, \varepsilon)) > 1$ . Additionally, if  $\mathbb{P}(\mathbf{U} \in B(\mathbf{x}, \varepsilon)) > 0$  for all  $\varepsilon > 0$ , it holds, by the definition of  $S$  in (4), that there exists some  $\varepsilon > 0$  such that

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(R > k | \mathbf{U} \in B(\mathbf{x}, \varepsilon)))}{\log(\mathbb{P}(R > k))} > 1. \quad (20)$$

Inequality (20) is also equivalent to  $G(B(\mathbf{x}, \varepsilon)) > 1$ . According to Lemma 2, it holds, in both cases, that  $G(\bar{B}(\mathbf{x}, \frac{\varepsilon}{2})) > 1$ . Set  $A_{-\mathbf{x}}$  is of the form  $B(-\mathbf{x}, r_{-\mathbf{x}})$  with  $r_{-\mathbf{x}} \leq \pi - \varepsilon/2$ , and it is not set  $\mathbb{S}^{d-1}$ . Thus,  $\mathbf{x} \notin A_{\mathbf{v}}$  when  $\mathbf{v} = -\mathbf{x}$ .

□

## 5. Applications and Examples

In practical applications, one has only a finite number of observations. Thus, an approximation based on a theoretical algorithm is required. Next, we present a starting point for the formulation of estimators, and study how they perform with data. The algorithm can be performed with general multidimensional data. Here, we consider two-dimensional data to keep the results easily presentable.

We define an empirical version of the function  $g$  defined in (3). Given observations  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  in  $\mathbb{R}^d$ , the empirical version of  $g$  is denoted by

$$\hat{g}(k, A) = \frac{\log\left(\#\{i : \|\mathbf{x}_i\| > k, \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|} \in A\}/n\right)}{\log(\#\{i : \|\mathbf{x}_i\| > k\}/n)}, \quad (21)$$

where  $k > 0$  is the user-defined threshold and  $A$  is a Borel set on the unit sphere.

In general, if

$$\hat{g}(k, A) > 1 + c \quad (22)$$

holds for set  $A$  and some  $c > 0$ , it provides evidence for  $A$  being in the complement of  $S$ . If

$$\hat{g}(k, A) < 1 + c, \quad (23)$$

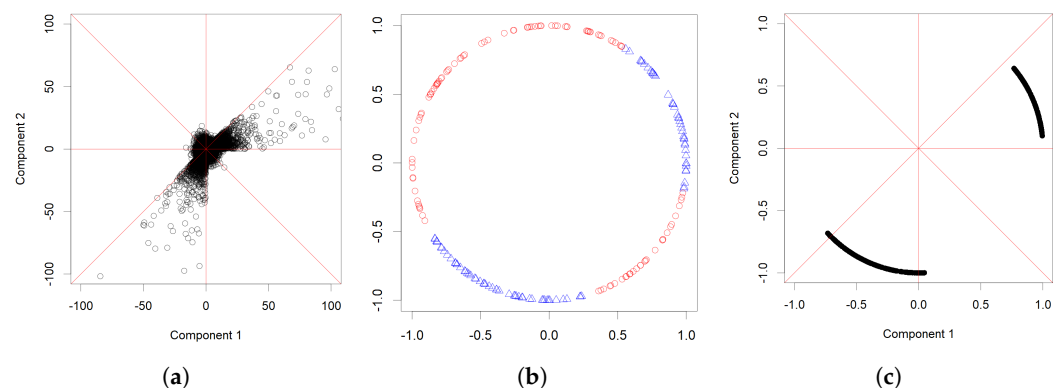
we gain evidence for  $A$  containing at least some subset of  $S$ . Since we can calculate the values of  $\hat{g}$  for any set, the challenge is to perform the calculations in a systematic way and combine the results to form an estimate for  $S$ . The most practical choices for sets  $A$  appear to be open balls centred around a given point on the unit sphere.

### 5.1. Simulation Study

A simulation study with  $n = 800,000$  observations was performed. Assumptions 1–4 are valid by construction. A two-dimensional data set was produced where heavy-tailed observations are possible in all directions, but some directions are heavier than others. The space was split into 8 equally sized cones. More specifically, the directional components

were uniformly sampled within each cone, and each direction was assigned a radius vector independently. The distribution of the radius was allowed to depend on the cone, but within each cone, all the radii are i.i.d. random variables. The idea is that, in some of the cones, the radii are heavier than others, and the objective is to identify such cones.

The original data set is presented in Figure 1a. The red lines indicate how the space is split. Two of the sectors have heavier Pareto-distributed radial components and the rest have lighter Weibull-distributed components. In Figure 1a, the cones with Pareto-distributed radial components are the ones with the largest number of observations, measured in the  $l_2$ -norm.



**Figure 1.** Illustration of simulated data. The original data set is presented in (a); (b) illustrates which directions are accepted into or rejected from the final estimate; (c) illustrates the preliminary estimate for  $S$  based on numerical data. The estimate is a projection of two cones onto the unit sphere, even though the numerically produced image consists of finitely many points.

The idea presented in Inequalities (22) and (23) was studied numerically. The testing sets  $D$  were chosen to be open balls of the form  $B(\mathbf{v}, s_{\mathbf{v}})$ . The radius  $s_{\mathbf{v}}$  was selected to be the smallest number such that 10% of all observations had directional components in  $B(\mathbf{v}, s_{\mathbf{v}})$ . Figure 1b presents the tested directions on the unit sphere. Each point corresponds to a fixed value of  $\mathbf{v}$ . We selected a fixed collection of approximately uniformly distributed points from the unit sphere to be used in place of vector  $\mathbf{v}$  in the first step of the algorithm. The red dots are the directions that were rejected, i.e., the value  $\hat{g}(k, B(\mathbf{v}, s_{\mathbf{v}}))$  of (21) is too high given the tolerance  $c$ , so that (22) holds. The blue triangles are the accepted centres of cones from which the final estimate is formed. In the final estimate, we have removed all open balls that were rejected for some direction  $\mathbf{v}$ . The values of the parameters were set to be  $c = 0.5$ , and  $k$  was such that the top 0.5% of the observations were chosen in  $l_2$  norm.

In Figure 1c, the preliminary estimate for  $S$  is presented on the unit sphere. The method correctly identifies the heaviest directions. The estimate seems to be most accurate near the centres of the cones, and less accurate near the edges between heavier and lighter radial components.

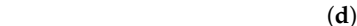
## 5.2. On the Detection Accuracy with Pareto Tails

We used the same algorithm as in Section 5.1 to analyse a similar data set, except that all the radial components have a Pareto distribution. The Pareto index is the heaviest in the same cones as earlier. The remaining directions have lighter Pareto tails.

The heaviest tail has an index of 2 and the lighter tails have tail index values of 2.3, 2.4, 2.5, and 3 in Figure 2a–d, respectively. The same random seed is used in all simulations.

When the lighter tail parameter is very close to the value of the heavier parameter 2, the algorithm cannot recover the set of riskiest directions with the given sample size and selected threshold  $c$ , see Figure 3. When the lighter tail parameter is 2.5 or larger, the produced estimate is rather accurate.

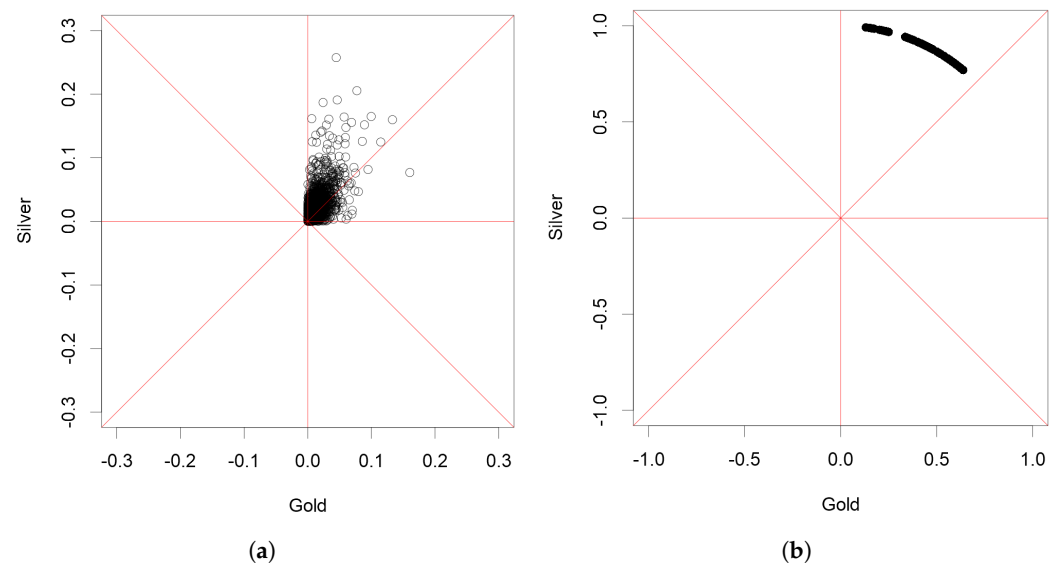




We studied only the negative changes in daily prices, i.e., when both the prices of silver and gold declined. Since only the days when both prices moved down are selected from the data set, the observations are concentrated into two cones. Thus, the coordinates indicate the size of the daily price drop. This decision was made in order to make the results comparable with the earlier results. In fact, the data set studied here is the data set pictured in Figure 9a of [Lehtomaa and Resnick \(2020\)](#), except it contains a few more observations. For background, the data of [Lehtomaa and Resnick \(2020\)](#) were gathered from the London Bullion Market Association. In the data set, the price of one ounce of gold

or silver is recorded each day during the time period from 1973 to 2014. Only complete cases where the price information was available for both gold and silver were accepted as part of the data set.

Figure 4b shows the estimate for the riskiest directions. The estimate is consistent with the earlier result obtained in Lehtomaa and Resnick (2020) in the sense that the riskiest observations appear to concentrate on a cone, and the riskiest observations are more concentrated above the diagonal than below it. It should be noted that the analysis here was performed using Euclidean distance, while the analysis of Lehtomaa and Resnick (2020) used the  $l_1$  distance and diamond plots.



**Figure 4.** A preliminary estimate for  $S$  in a real data set based on daily changes in the prices of gold and silver is presented on the right. The left subfigure (a) is a plot of the original data set.

In conclusion, a result that is consistent with the earlier study was obtained, but without the need to verify the assumptions of the multivariate, regularly varying distributions.

## 6. Conclusions

### 6.1. On the Interpretation of Estimates

In practice, we only have access to a finite amount of data. In addition, the user must select suitable values for the parameters  $c$  and  $k$  in (22) and (23). These seem to be the main challenges in the accurate detection of the directional components with heaviest tails. Typically, the parameter values are found by experimentation. One can fix  $k$  and increase the tolerance until some directions are accepted into the final estimate of  $S$ .

We state that tail function  $\overline{F}_1$  is heavier than  $\overline{F}_2$  if there exists a number  $x_0$  such that

$$\overline{F}_1(x) > \overline{F}_2(x) \quad (24)$$

for all  $x \geq x_0$ . The problem with real data is that the number  $x_0$  can be very large. Consequently, the largest observed data points might not be produced by the heaviest tails if the size of the data set is not sufficiently large. For example, a lognormal distribution has a heavier tail than a Weibull distribution with parameter  $\beta \in (0, 1)$ . If  $\beta$  is close to 0, a typical i.i.d. sample from these distributions could produce data where the points from a Weibull distribution appear to be larger. In the multidimensional setting, the direction can affect the heaviness of the  $R$ -variable. For this reason, the interpretation of the estimates is the following. *The estimator recovers the heaviest directional components with respect to the size of the data set. It does not exclude the possibility that there exist even heavier directional components than what is detected, and which remain undetected due to the limited number of data points.* To summarise, the estimator detects the directions in which heaviness comparable to that of one-dimensional data produced by normed observations is obtained.

## 6.2. Further Remarks

It seems plausible that the theoretical result in Theorem 1 could be used to create statistical estimators for set  $S$ . For example, given an estimate for the set of riskiest directions, an insurance company with heavy-tailed total losses could find the root cause of heavy-tailedness. Here, the cause is found by identifying individual components or interactions between multiple components that add heavy-tailedness to the total loss. If the company understands the set of riskiest directions, it is possible to formulate hedging or reinsurance strategies that potentially mitigate the largest risks. In this sense, understanding the set of riskiest directions tells us why the entire vector is heavy-tailed, and offers a strategy for reducing risk.

The presented results offer a starting point for creating rigorous statistical estimators with a clear workflow that can be implemented as a computer algorithm. At first, one checks for the heavy-tailedness of the observations by calculating the empirical hazard function of the normed observations that produce a one-dimensional data set. Once the heaviness of this one-dimensional data set has been established and the empirical hazard function turns out to be concave, we can search the directions where the heaviness of the observations corresponds to the heaviness of the one-dimensional data. A way to implement the method is to study cones around fixed points and determine the size of each cone based on a given portion of the total observations, e.g., we can find the smallest cone that contains 10% of the observations, as in the presented examples. This avoids the problem encountered with grid-based methods, where the space is divided into cells of equal size, and where it is possible that some of the cells remain sparsely populated with observations.

As the examples with simulated and real data show, an algorithm can be also implemented in the case where there are directions with only few observations or no observations at all. Furthermore, the data do not have to be transformed or pre-processed before applying the method, but it gives an idea of where the riskiest directions are in the case where, for instance, some components have different Pareto indices than other components.

Detecting small differences requires a large amount of data. To us, this means that there must be more data if the tails associated with different directions are almost equally heavy. The presented idea works best if there exists one or more directions where the tail is substantially heavier than in other directions. As in earlier models, the practical application of the method requires some parameters to be set by the user. In particular, choosing the threshold  $k$  is not easy, but this is a well-known problem that exists in different forms in most heavy-tailed modelling strategies (Nguyen and Samorodnitsky 2012).

**Author Contributions:** Conceptualization and investigation, M.H. and J.L.; writing—original draft preparation, M.H. and J.L.; writing—review and editing, M.H. and J.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** Open access funding is provided by University of Helsinki.

**Data Availability Statement:** The data used in the article can be requested from the corresponding author via email.

**Acknowledgments:** Suggestions made by the reviewers helped to improve the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Asmussen, Søren, and Jaakko Lehtomaa. 2017. Distinguishing log-concavity from heavy tails. *Risks* 5: 10. [\[CrossRef\]](#)
- Bridson, Martin R., and André Haeffliger. 2013. *Metric Spaces of Non-Positive Curvature*. Berlin and Heidelberg: Springer Science & Business Media, vol. 319.
- Cline, Daren BH, and Sidney I. Resnick. 1992. Multivariate subexponential distributions. *Stochastic Processes and Their Applications* 42: 49–72. [\[CrossRef\]](#)
- Das, Bikramjit, and Souvik Ghosh. 2016. Detecting tail behavior: Mean excess plots with confidence bounds. *Extremes* 19: 325–49. [\[CrossRef\]](#)

- Dembo, Amir, and Ofer Zeitouni. 1993. *Large Deviations Techniques and Applications*. Boston: Jones and Bartlett Publishers.
- Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. Modelling Extremal Events. In *Applications of Mathematics: For Insurance and Finance*. Berlin and Heidelberg: Springer, vol. 33. [\[CrossRef\]](#)
- Feng, Minyu, Hong Qu, Zhang Yi, and Jürgen Kurths. 2017. Subnormal distribution derived from evolving networks with variable elements. *IEEE Transactions on Cybernetics* 48: 2556–68. [\[CrossRef\]](#) [\[PubMed\]](#)
- Feng, Minyu, Liang-Jian Deng, Feng Chen, Matjaž Perc, and Jürgen Kurths. 2020. The accumulative law and its probability model: An extension of the pareto distribution and the log-normal distribution. *Proceedings of the Royal Society A* 476: 20200019. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ghosh, Souvik, and Sidney Resnick. 2010. A discussion on mean excess plots. *Stochastic Processes and their Applications* 120: 1492–517. [\[CrossRef\]](#)
- Ghosh, Souvik, and Sidney I. Resnick. 2011. When does the mean excess plot look linear? *Stochastic Models* 27: 705–22. [\[CrossRef\]](#)
- Hult, Henrik, and Filip Lindskog. 2002. Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability* 34: 587–608. [\[CrossRef\]](#)
- Klüppelberg, Claudia, Gabriel Kuhn, and Liang Peng. 2007. Estimating the tail dependence function of an elliptical distribution. *Bernoulli* 13: 229–51. [\[CrossRef\]](#)
- Lehtomaa, Jaakko, and Sidney I. Resnick. 2020. Asymptotic independence and support detection techniques for heavy-tailed multivariate data. *Insurance: Mathematics and Economics* 93: 262–77. [\[CrossRef\]](#)
- Li, Haijun, and Yannan Sun. 2009. Tail dependence for heavy-tailed scale mixtures of multivariate distributions. *Journal of Applied Probability* 46: 925–37. [\[CrossRef\]](#)
- Nguyen, Tilo, and Gennady Samorodnitsky. 2012. Tail inference: Where does the tail begin? *Extremes* 15: 437–61. [\[CrossRef\]](#)
- Omey, E. A. M. 2006. Subexponential distribution functions in  $\mathbf{R}^d$ . *Journal of Mathematical Sciences* 138: 5434–49. [\[CrossRef\]](#)
- Peters, Gareth W., and Pavel V. Shevchenko. 2015. *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*. Wiley Handbook in Financial Engineering and Econometrics. Hoboken: John Wiley & Sons, Inc. [\[CrossRef\]](#)
- Resnick, Sidney I. 2004. On the foundations of multivariate heavy-tail analysis. *Journal of Applied Probability* 41A: 191–212. [\[CrossRef\]](#)
- Samorodnitsky, Gennady, and Julian Sun. 2016. Multivariate subexponential distributions and their applications. *Extremes* 19: 171–96. [\[CrossRef\]](#)
- Weng, Chengguo, and Yi Zhang. 2012. Characterization of multivariate heavy-tailed distribution families via copula. *Journal of Multivariate Analysis* 106: 178–86. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.