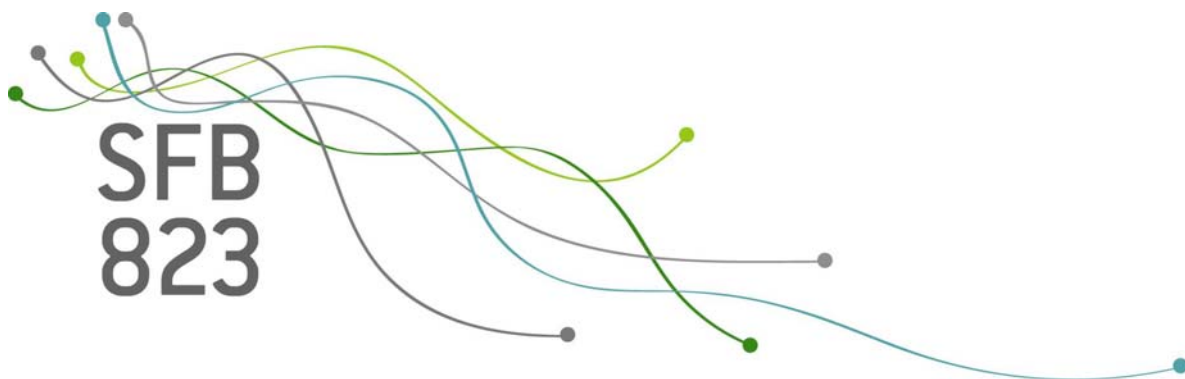# Multiscale inference for a multivariate density with applications to X-ray astronomy

Konstantin Eckle, Nicolai Bissantz,
Holger Dette, Katharina Proksch,
Sabrina Einecke

Multiscale inference for a multivariate density with applications to X-ray astronomy

Konstantin Eckle, Nicolai Bissantz, Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

Katharina Proksch

Institut für Mathematische Stochastik

Georg-August-Universität Göttingen

37077 Göttingen, Germany

Sabrina Einecke

Technische Universität Dortmund

Fakultät Physik

44221 Dortmund, Germany

April 5, 2016

Abstract

In this paper we propose methods for inference of the geometric features of a multivariate density. Our approach uses multiscale tests for the monotonicity of the density at arbitrary points in arbitrary directions. In particular, a significance test for a mode at a specific point is constructed. Moreover, we develop multiscale methods for identifying regions of monotonicity and a general procedure for detecting the modes of a multivariate density. It is is shown that the latter method localizes the modes with an effectively optimal rate. The theoretical results are illustrated by means of a simulation study and a data example. The new method is applied to and motivated by the determination and verification of the position of high-energy sources from X-ray observations by the Swift satellite which is important for a multiwavelength analysis of objects such as Active Galactic Nuclei.

1

1   Introduction

This work is concerned with the development of a statistical toolbox which is useful for data analysis in many problems of applied sciences. As a specific example we consider a problem from X-ray astronomy, namely the determination of the positions of objects of BL Lacertae type with statistical significance. Those objects form a specific subclass of blazars and hence active Galactic Nuclei (AGN), where high-energy relativistic jets arise perpendicular to the accretion disc and (in this case) point in the general direction of the Earth. They are among the brightest extragalactic sources in X-rays and gamma-rays in the sky. Determination and verification of the position of such objects from an observed distribution of origin positions of X-ray photons from the object is of paramount importance for a multiwavelength analysis of the object to understand and compare the appearance of photons of different energies in the object.

From a statistical point-of-view, problems of this type are fundamental and refer to a proper understanding of the shape of a density f based on a sample of multivariate observations. Numerous authors have worked on the detection of qualitative features, such as modes and regions of monotonicity of a density, in particular on tests for the existence and the localization of modes. For example, it was pointed out by Chan and Tong (2004) that the presence of modes can yield to a less precise forecasting. Similarly, a precise localization of modes can be used for non-parametric clustering [see for example Pollard (1981) for an early reference and Li et al. (2007), Chacón and Duong (2013) and Chaudhuri et al. (2014) for more recent work].

As pointed out by Romano (1988) and Grund and Hall (1995), estimation of modes of a density is a very complex problem, even more difficult than the estimation of the density itself. In fact, the problem is closely related to the estimation of the first derivative of the density. There exists a large amount of literature about statistical inference on modes of a density in the univariate setting, which can be roughly divided into four different categories: tests on the number of modes, the localization of modes, significance testing of candidate modes and tests that allow for inference about monotonicity. Donoho and Liu (1991) provide the minimax rate for estimating a single mode. The problem of estimating the number of modes is considered in Silverman (1981) and Hall and York (2001). These authors investigate a test that uses bootstrap methods based on the so-called critical bandwidth of a kernel density estimator [see also Mammen et al. (1991) and Chan and Tong (2004) for an asymptotic analysis and an extension to the dependent case]. In Hartigan and Hartigan (1985), the distance of the empirical distribution function to the best-fitting unimodal density is used as test statistic. Hartigan (1987) and Müller and Sawitzki (1991) propose the excess mass approach for statistical inference of (multi-) modality, which is also used by Polonik (1995); Minnotte (1997) and Fisher and Marron (2001) to construct nonparametric tests for the existence of modes. Chaudhuri and Marron (1999) introduce the SiZer-map as a graphical tool for the analysis of the local monotonicity properties of a density. In this paper, the derivative of a kernel density estimator is tested locally for a significant deviation from zero. A particular characteristic of the SiZer map is that

these tests are performed simultaneously over a fixed range of bandwidths. A multicsale test for the monotonicity of a univariate density, which allows simultaneous confidence statements about regions of increase and decrease, can be found in Dümbgen and Walther (2008). In the univariate deconvolution model, Schmidt-Hieber et al. (2013) propose a multiscale test for qualitative features of a density such as regions of monotonicity.

On the other hand, for multivariate densities there are just a few results on modality and even less on monotonicity. Tsybakov (1990) proves that the optimal minimax rate for mode detection over a β-Hölder class is $n^{-(\beta-1)/(2\beta+d)}$ ($\beta > 0$). The excess mass approach can also be used in the multivariate case, but most authors concentrate on one dimension because - as pointed out by Burman and Polonik (2009) - there is usually a trade-off between practical feasibility and theoretical justification. Abraham et al. (2004) use kernel smoothing to construct consistent estimators of the single mode of a multivariate density, while Klemelä (2005) suggests an adaptive estimate which achieves the optimal rate. Burman and Polonik (2009) do not pre-specify the total number of modes and propose a method for locating modal regions by means of formal testing for the presence of anti-modes. A rate-optimal algorithm for the localization of the modes of a multivariate density based on a k-nearest neighbour estimator of the density can be found in a recent paper of Dasgupta and Kpotufe (2014).

A test about local monotonicity properties of a bivariate density can be found in Godtliebsen et al. (2002) generalizing the SiZer-map. In a multivariate setting Duong et al. (2008) test locally whether the norm of the gradient of the density vanishes using a kernel density estimate with a fixed bandwidth. In a recent paper Genovese et al. (2015) suggest an algorithm for mode estimation of a d-dimensional density. These authors construct non-parametric confidence intervals for the eigenvalues of the Hessian at modes of a density estimate, which can be used for the construction of a significance test. The method is based on a sample splitting, where the first half of the data is used to localize the modes by means of the mean-shift algorithm and the second half of the data is used for the significance test. Genovese et al. (2015) also point out that the multiscale approach of Dümbgen and Walther (2008) for constructing confidence intervals for modes is only applicable to one-dimensional densities.

The goal of the present paper is to fill this gap by providing a multiscale method to identify regions of monotonicity of a multivariate density. In Section 2 we briefly review the approach of Dümbgen and Walther (2008). We also define a concept of monotonicity in the multivariate case and introduce a multiscale test for this property at a pre-specified point $x_0 \in \mathbb{R}^d$. The main idea is to investigate monotonicity properties of the density in "various" directions $e \in \mathbb{R}^d$ by projecting observations from a wedge centered at x_0 onto the line $\{x_0 + te \mid t \geq 0\}$. A multiscale test is provided that allows for a simultaneous inference of the monotonicity properties at a given confidence level α. Section 3 extends the approach to the situation where no prior information regarding the location and the number of the modes is available. The theoretical results of this paper establish the consistency of this approach and show that modes can be detected with the optimal rate (up to a logarithmic factor). The finite sample properties of the multiscale test are

3

investigated in Section 4 and in Section 5 we apply our proposed method to the determination and verification of the positions of the blazars Markarian 501 and S3 0218+35. Finally, all technical details and proofs are deferred to Section 6.

2 Local testing for a mode

In this Section, we present a test for the presence of a mode of the density f at a pre-specified candidate point $x_0 \in \mathbb{R}^d$ based on a sample of independent random variables $X_1, \ldots, X_n$ with density f. We begin with a brief review of the work of Dümbgen and Walther (2008), who investigate regions of monotonicity of a univariate density (that is $d = 1$).

2.1 Multiscale inference about a univariate density revisited

For one-dimensional independent identically distributed random variables $X_1, \ldots, X_n$ with density f let $X_{(1)} \leq \ldots \leq X_{(n)}$ denote the corresponding order statistics and consider the associated local spacings

$$X_{(i;j,k)} = \frac{X_{(i)} - X_{(j)}}{X_{(k)} - X_{(j)}}, \quad j \leq i \leq k.$$

Dümbgen and Walther (2008) propose to use the local spacings

$$T_{jk}(\boldsymbol{X}) = \sum_{i=j+1}^{k-1} \beta(X_{(i;j,k)})$$

to construct a test statistic for (local) monotonicity of the density f on the interval $(X_{(j)}, X_{(k)})$, where the function β is defined by $\beta(x) := (2x - 1)\mathbb{1}_{(0,1)}(x)$. Note that $T_{jk}(\boldsymbol{X})$ has mean zero if f is constant on $(X_{(j)}, X_{(k)})$. Let X denote a random variable with density f independent of $X_1, \ldots, X_n$, and denote by

$$\tilde{F}(x) = \mathbb{P}\big(X \leq x \mid X \in [X_{(1)}, X_{(n)}]\big)$$

the conditional distribution function of X given $X \in [X_{(1)}, X_{(n)}]$. Define $U_{(i)} = \tilde{F}(X_{(i)})$, then $U_{(2)}, \ldots, U_{(n-1)}$ correspond in distribution to the order statistics of a sample of $(n - 2)$ independent uniformly distributed random variables on the interval $[0, 1]$ (note that $U_{(1)} = 0$ and $U_{(n)} = 1$). It can be shown that the statistic

$$T_{jk}(\boldsymbol{U}) = \sum_{i=j+1}^{k-1} \beta(U_{(i;j,k)}) \text{ for } 1 \leq j < k \leq n, k - j > 1, \tag{2.1}$$

satisfies

$$T_{jk}(\boldsymbol{X}) \begin{cases} \geq T_{jk}(\boldsymbol{U}), & \text{if } f \text{ is increasing on } (X_{(j)}, X_{(k)}), \\ \leq T_{jk}(\boldsymbol{U}), & \text{if } f \text{ is decreasing on } (X_{(j)}, X_{(k)}). \end{cases}$$

Define $\Gamma(\delta) := \sqrt{2\log(\frac{\exp(1)}{\delta})}$,

$$T_n(\boldsymbol{U}) = \max_{1 \leq j < k \leq n, k-j>1} \left(\sqrt{\frac{3}{k-j-1}} |T_{jk}(\boldsymbol{U})| - \Gamma\left(\frac{k-j}{n-1}\right) \right), \tag{2.2}$$

and denote by $\kappa_n(\alpha)$ the $(1-\alpha)$-quantile of the statistic $T_n(\boldsymbol{U})$. The multiscale test for monotonicity proposed by Dümbgen and Walther (2008) now concludes that the density f is not increasing on every interval $(X_{(j)}, X_{(k)})$ with

$$T_{jk}(\boldsymbol{X}) < -c_{jk}(\alpha) := \sqrt{\frac{k-j-1}{3}} \left(\Gamma\left(\frac{k-j}{n-1}\right) + \kappa_n(\alpha) \right), \quad 1 \leq j < k \leq n, \ k-j > 1$$

and that f is not decreasing on every interval $(X_{(j)}, X_{(k)})$ with $T_{jk}(\boldsymbol{X}) > c_{jk}(\alpha)$. The overall risk of at least one false-positive decision within the simultaneous tests on all scales (i.e. for $1 \leq j < k \leq n, k-j > 1$) is at most α.

2.2 Assumptions and geometrical preparations

Throughout this paper $\|x\|$ denotes the Euclidean norm of a vector $x \in \mathbb{R}^d$. The function $f : \mathbb{R}^d \to \mathbb{R}$ has a mode at the point x_0, if for every vector $e \in \mathbb{R}^d$ with $\|e\| = 1$ the function $f_e : t \mapsto f(x_0 + te)$, $t \geq 0$, is strictly decreasing in a neighbourhood of $t = 0$. The aim of the test for the presence of a mode defined below is to investigate the monotonicity of functions of this type in different directions e. The number of directions is determined by the sample size n. As the set $\{x_0 + te \mid t \geq 0\}$ has Lebesgue measure 0, we also have to consider observations in a neighbourhood of this line for inference about monotonicity of the function f_e. For this purpose, we introduce a *signed distance of the projection* of a point $x \in \mathbb{R}^d$ onto the line $\{x_0 + te \mid t \in \mathbb{R}\}$ and introduce so-called *wedges*. For the following discussion we denote by $\{e_1, \ldots, e_{d-1}\}$ an arbitrary but fixed orthonormal basis of $(\mathrm{span}\{e\})^\perp$, $\langle x, y \rangle$ is the standard inner product of the vectors $x, y \in \mathbb{R}^d$ and "$\stackrel{d}{=}$" denotes equality in distribution.

Definition 2.1. Let $x_0 \in \mathbb{R}^d$ and $e \in \mathbb{R}^d$ with $\|e\| = 1$.

(1) The *projected signed distance* of a point $x \in \mathbb{R}^d$ from x_0 in direction e on $\{x_0 + te \mid t \in \mathbb{R}\}$ is defined as
$$P_e x := \langle x - x_0, e \rangle.$$

(2) The *wedge* with vertex x_0, direction e, length $l > 0$ and angle $\varphi \in (0, \frac{\pi}{2})$ is defined as

$$K \equiv K(x_0, e, \varphi) := \Big\{ x \in \mathbb{R}^d \mid 0 < P_e x \leq l \text{ and } \langle x - x_0, e_i \rangle \in [-\tan(\varphi)P_e x, \tan(\varphi)P_e x] $$
$$\text{for } i = 1, \ldots, d-1 \Big\}.$$

(3) For a wedge $K \subset \mathbb{R}^d$ let $X_{(1)}, \ldots, X_{(N)}$ be those random variables among $X_1, \ldots, X_n$ which are located in K, arranged in ascending order with respect to their signed projected distances from x_0, i.e. $X_{(j)} \in K$ for $j = 1, \ldots, N$ and $P_e X_{(1)} \leq \ldots \leq P_e X_{(N)}$. The wedge K_N is defined as $K_N := \{ x \in K : 0 < P_e x \leq P_e X_{(N)} \}$.
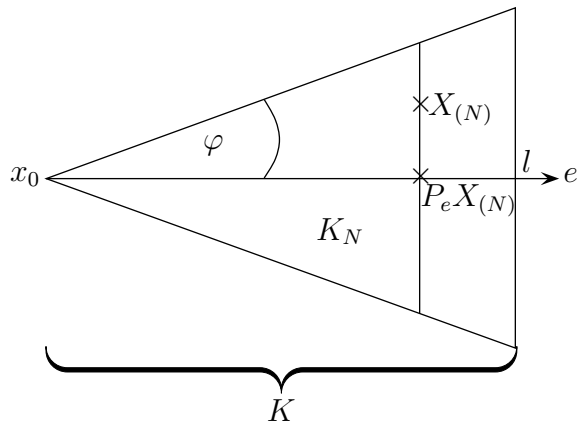

Figure 1: *The wedges K and K_N for $d = 2$.*

A typical wedge is displayed in Figure 1 in the case $d = 2$. We are now able to define monotonicity properties of the function f on the wedge K that will play a crucial role in following discussion.

Definition 2.2.

(i) The function f is called *increasing on the wedge K*, if $f(x_0 + \tilde{t}_2 e_0) \geq f(x_0 + \tilde{t}_1 e_0)$ for all $e_0 \in \mathbb{R}^d$ with $\|e_0\| = 1$ and $\tilde{t}_2 > \tilde{t}_1 \geq 0$ such that $x_0 + \tilde{t}_2 e_0 \in K$.

(ii) The function f is called *decreasing on the wedge K*, if $f(x_0 + \tilde{t}_2 e_0) \leq f(x_0 + \tilde{t}_1 e_0)$ for all $e_0 \in \mathbb{R}^d$ with $\|e_0\| = 1$ and $\tilde{t}_2 > \tilde{t}_1 \geq 0$ such that $x_0 + \tilde{t}_2 e_0 \in K$.

2.3 A local test for modality

Let $x_0 \in \mathbb{R}^d$ denote a candidate position for a mode. The construction of a local test for the hypothesis that f has a mode at x_0 is based on an investigation of the monotonicity properties of f on pairwise disjoint wedges K^i ($i = 1, \ldots, M_n$) with common vertex x_0. We begin with the case $M_n = 1$ and use the notation $K := K^1$ for the sake of simplicity. Throughout this paper $\mathbb{1}_A$ denotes the indicator function of a set A.

Theorem 2.3. *Let X be a d-dimensional random variable with density f independent of $X_1, \ldots, X_n$ and denote by*

$$\tilde{F}(z) = \mathbb{P}(P_e X \leq z | N, X \in K_N, X_{(N)}) \quad (0 < z \leq P_e X_{(N)})$$

the distribution function of $P_e X$ conditional on $N = \sum_{i=1}^{n} \mathbb{1}_K(X_i)$, $\{X \in K_N\}$ and $X_{(N)}$. Then, conditionally on N, $\tilde{F}(P_e X_{(1)}), \ldots, \tilde{F}(P_e X_{(N-1)})$ are distributed as the order statistics of $N-1$ independent uniformly distributed random variables on the interval $[0,1]$.

The first step in the construction of a test for a mode at the point x_0 is to investigate monotonicity in the sense of Definition 2.2. For this purpose, we use a comparison of the projected distances $P_e X_{(1)}, \ldots, P_e X_{(N-1)}$ with the distribution of projected distances of random variables $U_1, \ldots, U_{N-1}$ which are uniformly distributed on the wedge K_N. For a random variable U which is uniformly distributed on K and independent of $X_1, \ldots, X_n, U_1, \ldots, U_{N-1}$, we have

$$\tilde{F}_U(z) = \mathbb{P}(P_e U \leq z | N, U \in K_N, X_{(N)}) = \frac{z^d}{(P_e X_{(N)})^d} \quad (0 < z \leq P_e X_{(N)}),$$

and by Theorem 2.3, the random variable $\tilde{F}_U(P_e U_j) = \left(\frac{P_e U_j}{P_e X_{(N)}}\right)^d$, $j = 1, \ldots, N-1$, has a uniform distribution on the interval $[0,1]$, conditionally on N and the event $\{U_j \in K_N\}$. Consequently, we propose the test statistic

$$T_K = \sum_{j=1}^{N-1} \beta\left(\frac{(P_e X_{(j)})^d}{(P_e X_{(N)})^d}\right) \tag{2.3}$$

for testing monotonicity properties of the density f on the wedge K, where $\beta(z) = (2z - 1)\mathbb{1}_{(0,1)}(z)$. If f is constant on K, we have $\mathbb{E}[T_K] = 0$ as $\frac{(P_e X_{(1)})^d}{(P_e X_{(N)})^d}, \ldots, \frac{(P_e X_{(N-1)})^d}{(P_e X_{(N)})^d}$ have the same distribution as an order statistic of uniformly distributed random variables on the interval $[0,1]$. On the other hand, if f is increasing on the wedge K, the observations in K_N tend to have large projected distances from x_0, which results in positive values of the test statistic T_K. Similarly, if f is decreasing on K, it is more likely that the test statistic is negative.

Theorem 2.4. *Let $\tilde{F}$ denote the conditional distribution function defined in Theorem 2.3, T_K be defined in (2.3) and $T_K^U := \sum_{j=1}^{N-1} \beta(\tilde{F}(P_e X_{(j)}))$.*

(i) If f is increasing on K, then $T_K^U \leq T_K$ (a.s.) conditionally on N.

(ii) If f is decreasing on K, then $T_K^U \geq T_K$ (a.s.) conditionally on N.

By Theorem 2.3, conditionally on N, the statistic T_K^U has the same distribution as the random variable $\sum_{j=1}^{N-1} \beta(U_{(j)})$, where $U_1, \ldots, U_{N-1}$ are independent uniformly distributed random variables on the interval $[0,1]$. Therefore, Theorem 2.4 is the key result to obtain critical values for a multiscale test.

In the second step, we combine test statistics of the form T_K for different wedges to construct a test for a mode at the point x_0. For this purpose, define

$$l_n := \left(\frac{\log(n)}{n}\right)^{\frac{1}{d+4}} \tag{2.4}$$

7

and construct a family $\mathcal{K}_n$ of M_n pairwise disjoint wedges $K^1, \ldots, K^{M_n}$ with common vertex x_0, length $C_1 \log(n)^{\frac{d-1}{d+4}} l_n$ and angle $\varphi_n := \frac{C_2}{2} \log(n)^{-1}$ (for some constants $C_1, C_2 > 0$) and by specifying the central directions $\{e_n^1, \ldots, e_n^{M_n}\}$ as follows

(1) Choose a direction e_n^1 with $\|e_n^1\| = 1$

(2) If $e_n^1, \ldots, e_n^{i-1}$ have been specified, then - whenever possible - choose a vector e_n^i with $\|e_n^i\| = 1$ such that for some $\varepsilon > 0$

$$|\text{angle}(e_n^i, e_n^j)| \geq (2 + \varepsilon) \arctan\left(\sqrt{d-1} \tan(\varphi_n)\right) \text{ for all } j < i.$$

Note that this procedure does not define $\mathcal{K}_n$ in a unique way. However, if $\mathcal{K}_n$ has been fixed for the central directions $e_n^1, \ldots, e_n^{M_n}$, then the following property holds. For any normalized vector $e \in \mathbb{R}^d$ there exists a direction e_n^i such that

$$|\text{angle}(e, e_n^i)| < (2 + \varepsilon) \arctan\left(\sqrt{d-1} \tan(\varphi_n)\right) = O\left(\log(n)^{-1}\right).$$

This can be seen easily by deriving a contradiction from the opposite assertion using the expansion $\tan(z) = z(1 + o(1)) = \arctan(z)$ for $z \to 0$.
Now define for each $K^i \in \mathcal{K}_n$

$$N^i := \sum_{j=1}^{n} \mathbb{1}_{K^i}(X_j)$$

as the number of observations in the wedge K^i and consider the corresponding statistics T_{K^i} and $T_{K^i}^U$ defined in (2.3) and Theorem 2.4, respectively. An application of Theorem 2.3 on each wedge K^i shows that, conditionally on $N^1, \ldots, N^{M_n}$,

$$T_{K^i}^U \overset{d}{=} \sum_{j=1}^{N^i-1} \beta(U_{(j)}^i) \quad (i = 1, \ldots, M_n), \tag{2.5}$$

where $\{U_j^i \mid j = 1, \ldots, N^i - 1, i = 1, \ldots, M_n\}$ are independent uniformly distributed random variables on the interval $[0, 1]$. In particular, the statistics $T_{K^1}^U, \ldots, T_{K^{M_n}}^U$ are conditionally independent.
By means of the representation (2.5), the quantile $\tilde{\kappa}_n(\alpha)$ defined by the condition

$$\mathbb{P}\left(\max_{i=1,\ldots,M_n}\left(\sqrt{\frac{3}{N^i-1}}|T_{K^i}^U| - \Gamma\left(\frac{N^i}{n-1}\right)\right) \leq \tilde{\kappa}_n(\alpha) \,\Big|\, N^1, \ldots, N^{M_n}\right) = 1 - \alpha \tag{2.6}$$

can be obtained by numerical simulation, as soon as the numbers of observations $N^1, \ldots, N^{M_n}$ in the wedges $K^1, \ldots, K^{M_n}$ have been specified. We note that a calibration by the term $\Gamma\left(\frac{N^i}{n-1}\right)$ for various scales (i.e. different values of N^i) is necessary to show that the quantile $\tilde{\kappa}_n(\alpha)$ is asymptotically bounded [see Section 6 for details].

In a third step, we consider on each of the wedges $K^1, \ldots, K^{M_n}$ two hypotheses, that is

$$H_{0,i}^{incr}: \quad f \text{ is increasing on } K^i \qquad versus \qquad H_{1,i}^{incr}: \quad f \text{ is not increasing on } K^i \qquad (2.7)$$

$(i = 1, \ldots, M_n)$, and

$$H_{0,i}^{decr}: \quad f \text{ is decreasing on } K^i \qquad versus \qquad H_{1,i}^{decr}: \quad f \text{ is not decreasing on } K^i \qquad (2.8)$$

$(i = 1, \ldots, M_n)$, where the the notation of an increasing (decreasing) function on the wedge K^i is introduced in Definition 2.2. The ith hypothesis in (2.7) is rejected, whenever

$$T_{K^i} < -\tilde{c}_{K^i}(\alpha) \qquad (2.9)$$

$(i = 1, \ldots, M_n)$, where the quantile $\tilde{c}_{K^i}(\alpha)$ is defined by $\tilde{c}_{K^i}(\alpha) = \sqrt{\frac{N^i-1}{3}}\big(\tilde{\kappa}_n(\alpha) + \Gamma(\frac{N^i}{n-1})\big)$ $(i = 1, \ldots, M_n)$. Similarly, the ith hypothesis in (2.8) is rejected, whenever

$$T_{K^i} > \tilde{c}_{K^i}(\alpha) \qquad (2.10)$$

$(i = 1, \ldots, M_n)$. The final result of this section specifies the error of at least one false decision among these $2M_n$ local level α-tests on monotonicity.

Theorem 2.5. *Assume that all tests (2.9) and (2.10) for the hypotheses (2.7) and (2.8) are performed $(i = 1, \ldots, M_n)$. The probability of at least one false rejection of any of the tests is at most α.*

We conclude this section by showing that the results presented so far can be used to obtain a consistent multiscale test for the hypothesis that the density f has a mode at a given point $x_0 \in \mathbb{R}^d$. The test decides for the presence of a mode at x_0 if every test (2.9) for the null hypothesis that f is increasing on the wedge $K^i, i = 1, \ldots, M_n$, rejects the null. Note that in this case we use the one-sided quantiles $\tilde{c}'_{K^i}(\alpha) = \sqrt{\frac{N^i-1}{3}}\big(\tilde{\kappa}'_n(\alpha) + \Gamma(\frac{N^i}{n-1})\big)$ in (2.9), where $\tilde{\kappa}'_n(\alpha)$ is defined by the condition

$$\mathbb{P}\Big(\max_{i=1,\ldots,M_n}\Big(-\sqrt{\frac{3}{N^i-1}}T_{K^i}^U - \Gamma\Big(\frac{N^i}{n-1}\Big)\Big) \leq \tilde{\kappa}'_n(\alpha)\Big|N^1, \ldots, N^{M_n}\Big) = 1 - \alpha. \qquad (2.11)$$

Theorem 2.6. *Assume that the density f is twice continuously differentiable in a neighbourhood of x_0 with $f(x_0) \neq 0$, gradient $\nabla f(x_0) = 0$ and a Hessian $H_f(x_0)$ satisfying $e_0^\top H_f(x_0)e_0 \leq -c < 0$ for all $e_0 \in \mathbb{R}^d$ with $\|e_0\| = 1$. Consider the family of wedges $\mathcal{K}_n$ defined in Section 2.3 with constants C_1, C_2 satisfying*

$$C_1^{d+4}C_2^{d-1} > \frac{4D^2}{c^2}\frac{f(x_0)}{d+4}, \qquad (2.12)$$

where

$$D = \frac{\sqrt{2}(2d+2)(d+2)}{\big(1-\frac{d}{d+2}\big)^{(d+2)/d}\Big[1 - \frac{d^2}{2(2d+2)^2}\big(-1 + \{1 + 4(\frac{2d+2}{d})^2\}^{1/2}\big)\Big]^{1/2}}. \qquad (2.13)$$

Then, all M_n tests defined in (2.9) (using the quantiles $\tilde{c}'_{K^i}(\alpha)$ instead of $\tilde{c}_{K^i}(\alpha)$ $(i = 1, \ldots, M_n)$) reject the null hypothesis with asymptotic probability one as $n \to \infty$.

9

Note that the constant D in (2.13) depends only on the dimension d. Hence, the lower bound on the constants C_1 and C_2 is determined by the shape of the modal region (more precisely the largest eigenvalue of the Hessian $H_f(x_0)$ at x_0) as well as by the value of the density at the point x_0.

3 Global inference on monotonicity

In this section we extend the local inference on modality at a fixed point to the situation where no specific candidate position for the mode can be defined in advance. This is particularly important since there exist several applications where at most approximate information about the position of the modes is available. As in the previous section, let $X_1, \dots, X_n$ denote independent d-dimensional random variables with density f. The proposed test for the detection of modes proceeds in several steps.

The first step consists in a selection of the candidate modes. Here, we choose these as the vertices of an equidistant grid in $\mathbb{R}^d$. Secondly, we introduce a generalization of the multiscale test on monotonicity presented in Section 2, where we divide the wedges in subsections that are determined by the data. The latter approach can be very useful in settings without a priori knowledge about the modes, as a true mode obviously has not to be located at the vertex of a wedge. Figure 2 provides a graphical representation of the results of the global test on modality in the bivariate case where the multiscale generalization has been omitted. Here, on every dotted wedge K, the test has rejected that f is decreasing on K. Accordingly, the cross-hatches refer to a rejection that f is increasing on K. Non-marked wedges indicate that no significant result has been found. For a detailed description of the settings used to provide Figure 2 and an analysis of the results, we refer to the end of this section.

3.1 Geometrical preparations

Throughout this section let b_j denote the jth unit vector in $\mathbb{R}^d$ ($j = 1, \dots, d$) and define $\lceil x \rceil := \inf\{z \in \mathbb{Z} \,|\, z \geq x\}$. Recall the definition of l_n in (2.4) and denote (for given constants $C_1, \varepsilon > 0$) by $\mathcal{G}_n$ the grid consisting of the vertices

$$\sum_{j=1}^d (i_j(2+\varepsilon)C_1 \log(n) l_n - n) b_j$$

$\left(i_1 \dots i_d \in \left\{ 0, \dots, \lceil \frac{2n}{(2+\varepsilon)C_1 \log(n) l_n} \rceil \right\} \right)$. Note that the grid $\mathcal{G}_n$ covers the cube $[-n, n]^d$ and that the sequence $\log(n) l_n$ (which determines the order of the mesh size) is chosen such that the test of modality defined below is consistent.

We now define by $\mathcal{K}_n$ a family of wedges (cf. Definition 2.1) with length $l = C_1 \log(n) l_n$, an angle $\varphi = \varphi_n = \frac{C_2}{2} \log(n)^{-1}$ for a given constant $C_2 > 0$, vertex in $\mathcal{G}_n$, and a direction contained in the set of given directions $\{e_n^1, \dots, e_n^{M_n}\}$ (cf. Section 2.3). For an arbitrary but fixed element K
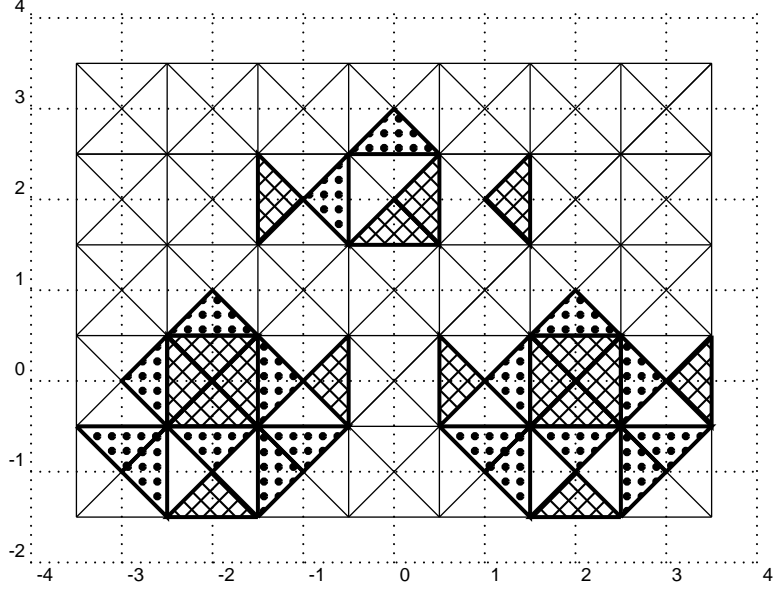
Figure 2: *Example of a global map for monotonicity of a density.*

of $\mathcal{K}_n$ let $X_{(1)}, \ldots, X_{(N)}$ denote those random variables among $\{X_1, \ldots, X_n\}$ which are located in K and ordered with respect to their signed projected distances from the vertex x_0 of K. For $0 \leq j < k \leq N$ with $k - j > 1$ we define

$$K(j,k) := \{x \in \mathbb{R}^d \mid x \in K \text{ and } P_e X_{(j)} < P_e x \leq P_e X_{(k)}\}$$

as a *subsection* of the wedge K, where $X_0 := x_0$ and x_0 denotes the vertex of K. A typical set is depicted in Figure 3. We conclude this section with a definition of a concept of monotonicity on subsections of a wedge.

Definition 3.1. *Let K be a wedge with vertex x_0 and $K(j,k) \subseteq K$ be a subsection. The function $f : \mathbb{R}^d \to \mathbb{R}$ is*

(i) *increasing on $K(j,k)$, if $f(x_0 + \tilde{t}_2 e_0) \geq f(x_0 + \tilde{t}_1 e_0)$ for all $e_0 \in \mathbb{R}^d$ with $\|e_0\| = 1$ and all $\tilde{t}_2 > \tilde{t}_1 \geq 0$ such that $x_0 + \tilde{t}_\ell e_0 \in K(j,k)$ $(\ell = 1, 2)$.*

(i) *decreasing on $K(j,k)$, if $f(x_0 + \tilde{t}_2 e_0) \leq f(x_0 + \tilde{t}_1 e_0)$ for all $e_0 \in \mathbb{R}^d$ with $\|e_0\| = 1$ and all $\tilde{t}_2 > \tilde{t}_1 \geq 0$ such that $x_0 + \tilde{t}_\ell e_0 \in K(j,k)$ $(\ell = 1, 2)$.*

3.2   Regions of monotonicity and mode detection

The approach proposed here consists of simultaneous tests for monotonicity of the density f on every subsection of every wedge in $\mathcal{K}_n$. For the definition of these tests we will proceed similarly
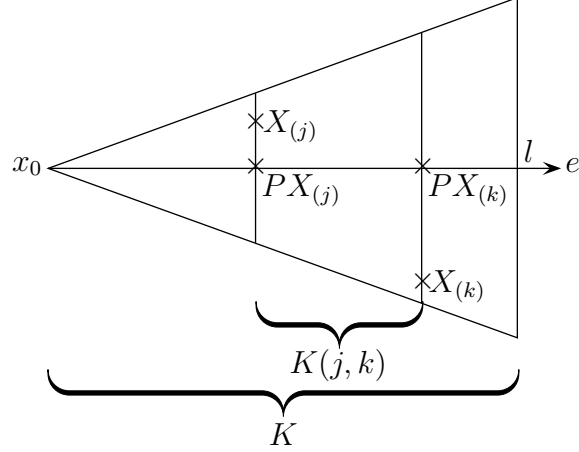
Figure 3: *The subsection $K(j,k)$ for $d = 2$.*

as in Section 2.3. We begin by introducing a multiscale test statistic on the subsection $K(j,k)$ of a wedge $K \in \mathcal{K}_n$ which is defined by

$$T_{K(j,k)} := \sum_{l=j+1}^{k-1} \beta \left(\frac{(P_e X_{(l)})^d - (P_e X_{(j)})^d}{(P_e X_{(k)})^d - (P_e X_{(j)})^d} \right), \tag{3.1}$$

where $0 \leq j < k \leq N$, $k - j > 1$. Note that $T_{K(0,N)} = T_K$, where T_K is the test statistic defined in (2.3).

Now, let $\mathcal{K}_n = \{K^i \mid i = 1, \ldots, L_n\}$ denote the family of wedges defined in Section 3.1. For the multiscale approach, we use for each subsection $K^i(j_i, k_i)$ of the wedge K^i the test statistic $T_{K^i(j_i,k_i)}$ defined by (3.1) ($0 \leq j_i < k_i \leq N^i$, $k_i - j_i > 1$, $i = 1, \ldots, L_n$) and consider

$$T_{K^i(j_i,k_i)}^U = \sum_{l=j_i+1}^{k_i-1} \beta \left(\frac{\tilde{F}^i(P_e^i X_{(l)}) - \tilde{F}^i(P_e^i X_{(j_i)})}{\tilde{F}^i(P_e^i X_{(k_i)}) - \tilde{F}^i(P_e^i X_{(j_i)})} \right),$$

where P_e^i denotes the signed projected distance and $\tilde{F}^i$ denotes the conditional distribution function with respect to K^i ($0 \leq j_i < k_i \leq N^i$, $k_i - j_i > 1$, $i = 1, \ldots, L_n$). Using similar arguments as in Section 2.3, it follows that (conditionally on N^i)

$$T_{K^i(j_i,k_i)}^U \leq T_{K^i(j_i,k_i)} \ (a.s.) \quad \text{if } f \text{ is increasing on } K(j_i,k_i),$$
$$T_{K^i(j_i,k_i)}^U \geq T_{K^i(j_i,k_i)} \ (a.s.) \quad \text{if } f \text{ is decreasing on } K(j_i,k_i)$$

($0 \leq j_i < k_i \leq N^i$, $k_i - j_i > 1$, $i = 1, \ldots, L_n$). Moreover,

$$T_{K^i(j_i,k_i)}^U \overset{d}{=} \sum_{l=j_i+1}^{k_i-1} \beta \left(\frac{U_{(l)}^i - U_{(j_i)}^i}{U_{(k_i)}^i - U_{(j_i)}^i} \right) \quad \text{conditional on } N^1, \ldots, N^{L_n},$$

where $\{U_{j_i}^i \mid j_i = 1, \ldots, N^i - 1, i = 1, \ldots, L_n\}$ are independent random variables which are uniformly distributed on the interval $[0,1]$, and $U_{(1)}^i \leq \ldots \leq U_{(N^i-1)}^i$ is the order statistics of

12

$U_1^i, \ldots, U_{N^i-1}^i$ ($i = 1, \ldots, L_n$). Finally, let $\overline{\kappa}_n(\alpha)$ denote the $(1-\alpha)$-quantile of the conditional distribution of the random variable

$$\max_{i=1,\ldots,L_n} \max_{0 \le j_i < k_i \le N^i,\, k_i - j_i > 1} \left(\sqrt{\frac{3}{k_i - j_i - 1}} |T_{K^i(j_i,k_i)}^U| - \Gamma\left(\frac{k_i - j_i}{n-1}\right) \right) \tag{3.2}$$

given $N^1, \ldots, N^{L_n}$. We consider on each subsection $K^i(j_i, k_i)$ of the wedge K^i the hypotheses

$$H_{0,i,j_i,k_i}^{incr} : f \text{ is increasing on } K^i(j_i,k_i) \text{ versus } H_{1,i,j_i,k_i}^{incr} : f \text{ is not increasing on } K^i(j_i,k_i), \tag{3.3}$$
$$H_{0,i,j_i,k_i}^{decr} : f \text{ is decreasing on } K^i(j_i,k_i) \text{ versus } H_{1,i,j_i,k_i}^{decr} : f \text{ is not decreasing on } K^i(j_i,k_i)$$

($0 \le j_i < k_i \le N^i,\ k_i - j_i > 1,\ i = 1, \ldots, L_n$). The hypothesis H_{0,i,j_i,k_i}^{incr} is rejected if

$$T_{K^i(j_i,k_i)} < -\overline{c}_{K^i(j_i,k_i)}(\alpha) \tag{3.4}$$

($0 \le j_i < k_i \le N^i,\ k_i - j_i > 1,\ i = 1, \ldots, L_n$), where $\overline{c}_{K^i(j_i,k_i)}(\alpha) := \sqrt{\frac{k_i - j_i - 1}{3}}\left(\overline{\kappa}_n(\alpha) + \Gamma(\frac{k_i-j_i}{n-1})\right)$. Similarly, H_{0,i,j_i,k_i}^{decr} is rejected if

$$T_{K^i(j_i,k_i)} > \overline{c}_{K^i(j_i,k_i)}(\alpha) \tag{3.5}$$

($0 \le j_i < k_i \le N^i,\ k_i - j_i > 1\ i = 1, \ldots, L_n$). Following the line of arguments in the proof of Theorem 2.5, we obtain the following result.

Theorem 3.2. *If all tests (3.4) and (3.5) are performed simultaneously ($0 \le j_i < k_i \le N^i,\ k_i - j_i > 1, i = 1, \ldots, L_n$), then the probability of at least one false rejection is at most α.*

3.3 Mode detection

We consider the following asymptotic regime. For $n \in \mathbb{N}$ let $\mathcal{K}_n$ denote the family of wedges defined in Section 3.1 and define $\mathcal{I}_n$ as the set of indices i corresponding to the wedges $K_n^i \in \mathcal{K}_n$ whose vertices $x_{0,n}$ fulfill $C_1 \log(n) l_n \le \|x_0^n - x_0\| \le m_n C_1 \log(n) l_n$ for a mode x_0 of f and $m_n = (\log(n))^{\frac{5}{2}}$ and whose direction e_n fulfills $\text{angle}(x_0^n - x_0, e_0^n) = O(\log(n)^{-1})$. Then, every test (3.4) for the hypothesis $H_{0,i,0,N^i}^{incr}$ defined (3.3) (i.e. f is increasing on K_n^i) with $i \in \mathcal{I}_n \subseteq \{1, \ldots, L_n\}$, rejects the null with asymptotic probability one.

Theorem 3.3. *Let*

$$D = \frac{2\sqrt{2}(2d+1)(d+1)}{\left(1 - \frac{d}{d+1}\right)^{(d+1)/d} \left[1 - \frac{d^2}{2(2d+1)^2}\left(-1 + \left\{1 + 4(\frac{2d+1}{d})^2\right\}^{\frac{1}{2}}\right)\right]^{\frac{1}{2}}}.$$

Assume that for any mode $x_0 \in \mathbb{R}^d$ the density f satisfies $c_1 \ge f(x_0) > 0$ and that there exist functions $g_{x_0} : \mathbb{R}^d \to \mathbb{R}$, $\tilde{f}_{x_0} : \mathbb{R} \to \mathbb{R}$ such that the density f has a representation of the form

$$f(x) \equiv (1 + g_{x_0}(x)) \tilde{f}_{x_0}(\|x - x_0\|) \tag{3.6}$$

(in a neighbourhood of x_0). Furthermore, let g_{x_0} be differentiable in a neighbourhood of x_0 with $g_{x_0}(x) = o(1)$ and $\langle \nabla g_{x_0}(x), e_0 \rangle = o(\|x - x_0\|^{1+\gamma})$ (for some $\gamma > 0$) if $x \to x_0$ and all $e_0 \in \mathbb{R}^d$ with $\|e_0\| = 1$. In addition, let $\tilde{f}_{x_0}$ be differentiable in a neighbourhood of 0 with $\tilde{f}'_{x_0}(h) \leq -ch(1 + o(1))$ for $h \to 0$. If $\mathcal{K}_n$ is the family of wedges defined in Section 3.1 with

$$C_1^{d+4} C_2^{d-1} > \frac{D^2}{c^2} \frac{c_1}{d+4}, \tag{3.7}$$

then every mode x_0 of f will be detected with asymptotic probability one as $n \to \infty$.

Theorem 3.3 shows that the proposed procedure can find all modes with (up to the logarithmic factor) optimal rate. Note that we proceed in two steps: the verification of the presence of a mode and its localization. With probability one the presence of every mode will be detected (by means of the asymptotic regime introduced at the beginning of this section). The rate for the localization of a mode is given by the mesh size of the grid $\mathcal{G}_n$, which is determined by the length of the wedges.

4   Finite sample properties

In this section we illustrate the finite sample properties of the proposed multiscale inference. In particular, we study the power of the local test for a mode at a given point $x_0 \in \mathbb{R}^d$. We also present an example illustrating how the results of Section 3 can be used to obtain a graphical representation of the local monotonicity properties of the density.

4.1   Local test for modality

Here, we investigate the finite sample properties of the local test for a two-dimensional density, where the level is given by $\alpha = 0.05$. The corresponding quantiles $\tilde{\kappa}'_n(0.05)$ defined in (2.11) are determined from 1000 simulation runs based on independent and uniformly distributed random variables on the interval $[0, 1]$ and are listed in Table 1 for the sample sizes $n = 100, 500, 5000$ in the situation considered in Table 2 (note that $\tilde{\kappa}'_n(0.05)$ depends on the number of observations in every wedge and hence both on the number and on the size of the wedges). By its construction,

observations	$\tilde{\kappa}'_n(0.05)$
100	0.126
500	-0.319
5000	-0.854

Table 1: *Simulated quantiles $\tilde{\kappa}'_n(0.05)$ in the situation considered in Table 2.*

the local test is conservative, and therefore we also investigate a calibrated version of the new

14

test. The quantiles of the calibrated test are chosen such that the level of the test coincides with $\alpha = 0.05$ for the data obtained from a uniform distribution on the set $[-2.5, 2.5]^2$. Note that this calibration does not require any knowledge about the unknown density f. However, the procedure requires the choice of the length and the angle of the wedges and according to Theorem 2.6 we used

$$l_n = C_1 \Big(\frac{\log(n)}{n}\Big)^{\frac{1}{d+4}} \log(n)^{\frac{d-1}{d+4}} \ , \quad \varphi_n = \frac{C_2}{2} \log(n)^{-1},$$

where $C_1, C_2 > 0$ are constants. In the following, the power and level of the test with respect to different choices of C_1 and C_2 is investigated. We also consider different numbers M_n of wedges in our study. Recall from the discussion in Section 2.3 that the constants C_1 and C_2 have to satisfy (2.12) in order to guarantee consistency of the test. All results presented below are based on 1000 simulation runs.

We begin with a comparison of the test introduced in Section 2 (based on the critical values $\tilde{\kappa}'_n(0.05)$) and a calibrated version of this test. In Table 2 we present the simulated level and power of the local test for a mode at the point $x_0 = (0,0)^\top$ for different sample sizes. The constants in the definition of the length and the angle are chosen as $C_1 = 2$ and $C_2 = 9.65$. For the investigation of the level we consider a uniform distribution on the square $[-2.5, 2.5]^2$, since it represents a "worst" case scenario. For the calculation of the power, we sample from the standard normal distribution. We observe that the test proposed in Section 2 is conservative but it has reasonable power with increasing sample size. On the other hand, the calibrated version of the multiscale test keeps its nominal level and rejects the null hypothesis of no mode at x_0 in nearly all cases.

observations	l_n	M_n	level	power	level (cal.)	power (cal.)
100	1.54	3	0.0	36.8	4.8	97.6
500	1.31	4	0.0	50.0	4.5	98.4
5000	0.99	5	0.0	72.7	5.0	100

Table 2: *Simulated level and power of the local test for a mode at $x_0 = (0,0)^\top$ of a 2-dimensional density.*

Next we investigate the influence of the shape of the modal region on the power of the local test. To this end, we sample from normal distributions with expectation $(0,0)^\top$ and covariance matrix $\Sigma \neq I_2$. The results for

$$\Sigma_1 = \begin{pmatrix} 0 & 0.5 \\ -1 & 1.5 \end{pmatrix} \text{ and } \Sigma_2 = \begin{pmatrix} -0.5 & 1 \\ -2 & 2.5 \end{pmatrix} \tag{4.1}$$

are presented in Table 3. We conclude that the shape of the modal region (determined by the absolute values of the eigenvalues of the covariance matrix) has a strong influence on the power of the test. In the case Σ_1 (eigenvalues 0.5 and 1), the absolute values of both eigenvalues

observations	l_n	M_n	Σ_1		Σ_2	
			power	power (cal.)	power	power (cal.)
100	1.54	3	65.4	98.7	38.8	94.3
500	1.31	4	95.6	100	80.1	99.6
5000	0.99	5	97.8	100	92.1	99.7

Table 3: *Influence of the shape of the modal region on the power of the local test. The matrices Σ_1 and Σ_2 are given in (4.1).*

are smaller than one. For Σ_2 the eigenvalues are given by 0.5 and 1.5. Hence, we observe a (slight) decrease in power in comparison to the first case. However, due to the existence of an eigenvalue with absolute value smaller than one, the test still performs better as in the case of a standard normal distribution.

As the local test requires the specification of the point x_0, we next investigate the influence of its incorrect specification on the power of the test. For this purpose, we consider the same data (two-dimensional standard normal distribution) and perform the tests under the assumption that the modes are given by $x_0 = (0.2, 0.2)^\top$ and $x_0 = (0.7, 0.7)^\top$, respectively (which has to be compared to the true position of the mode at $(0,0)^\top$). The corresponding results are shown in Table 4, and we conclude that a "small" deviation of the candidate mode from the true mode has a very small effect on the power of the tests. In the case $x_0 = (0.7, 0.7)^\top$, the distance between the candidate and the true mode is very large in comparison to the length of the wedges. For $n = 100$ observations the length of the wedges is still substantially larger than the distance between the candidate and the true mode. Hence, the test detects the presence of a mode, but we observe a decrease in its power. However, for $n = 5000$ observations the distance between the candidate position and the true mode is approximately equal to the length of the wedges. As a consequence, the multiscale test is performed with a finer triangulation and (correctly) does not indicate the existence of a mode at the point $x_0 = (0.7, 0.7)^\top$.

observations	l_n	M_n	$x_0 = (0.2, 0.2)^\top$		$x_0 = (0.7, 0.7)^\top$	
			power	power (cal.)	power	power (cal.)
100	1.54	3	32.4	96.6	2.8	75.6
500	1.31	4	43.1	97.8	1.2	57.1
5000	0.99	5	47.2	98.3	0.1	10.8

Table 4: *Influence of a misspecification of the mode on the power of the local test.*

In the remaining part of this section we investigate the influence of the choice of the parameters C_1 and C_2 on the power of the test. Note that the volume of every wedge is proportional to l_n^d, where l_n is the length of the wedge. This means that dividing the length in half yields a wedge with a volume which is 2^{-d} times smaller than the volume of the original wedge. Thus, the number of observations in the smaller wedge is substantially smaller than the number of observations in the larger wedge. Therefore, we expect that the constant C_1 has an impact on the power of the test. These theoretical considerations are reflected by the numerical results

in Table 5, which show the power for a fixed sample size $n = 500$, different choices of C_1 (represented by the different lengths) and fixed parameter C_2. We observe a loss of power of both tests with decreasing length of the wedge.

On the other hand, decreasing the constant C_2 such that the number of wedges doubles has the effect that the number of observations in every wedge decreases approximately by 50%. In Table 6 we show the power for a fixed sample size $n = 500$, a fixed constant $C_1 = 2$ and varying values of C_2 (represented by the different number of wedges). Here, the picture is not so clear. While we observe a loss in power of the non-calibrated tests with an increasing number of wedges, the power of the calibrated test changes only slightly. In both cases, the calibrated version still performs rather well opposite to its uncalibrated version.

observations	l_n	M_n	level	power	level (cal.)	power (cal.)
500	1.31	4	0.0	50.0	4.5	98.4
500	0.98	4	0.0	1.5	4.9	74.4
500	0.65	4	0.0	0.0	5.4	37.2

Table 5: *Influence of the length of the wedges on the power.*

observations	l_n	M_n	level	power	level (cal.)	power (cal.)
500	1.31	4	0.0	50.0	4.5	98.4
500	1.31	6	0.0	0.7	5.3	92.7
500	1.31	8	0.0	0.0	4.9	89.8

Table 6: *Influence of the number of directions tested on the power.*

4.2   Identifying local monotonicity of a multivariate density

In this section we demonstrate how the results of Section 3 can be used to obtain a graphical representation of the local monotonicity behaviour of the density (in the case $d = 2$). We conduct the procedure to detect regions of monotonicity as proposed in Section 3.2. For the sake of convenience, we use only the largest scales in the test statistic (3.1) (i.e. we test on the entire wedges and not on the subsections introduced in Section 3.1). The significance level is $\alpha = 0.05$. We chose an equidistant grid covering $[-3, 3] \times [-1, 3]$ with points $(i, j)^\top$, $i = -3, \ldots, 3, j = -1, \ldots, 3$, the length of any wedge is $l = \frac{1}{2}$ and all angles are given by $\varphi = \frac{\pi}{4}$. Figure 2 presents the map of the local monotonicity properties on the basis of $n = 100000$ observations from a mixture of three normal distributions (i.e. f has three modes of different shape) [see Figure 4]. Here, a cross-hatched wedge indicates that the local test rejected the hypothesis that the density is increasing on the respective wedge. Similarly, a dotted wedge implies that the test rejected the hypothesis that the density is decreasing. Non-significant wedges are not marked.

The map indicates the existence of modes close to the grid points $(-2, 0)^\top$ and $(2, 0)^\top$ and in a weaker sense indicates a mode close to the grid point at $(0, 2)^\top$. The marked geometrical

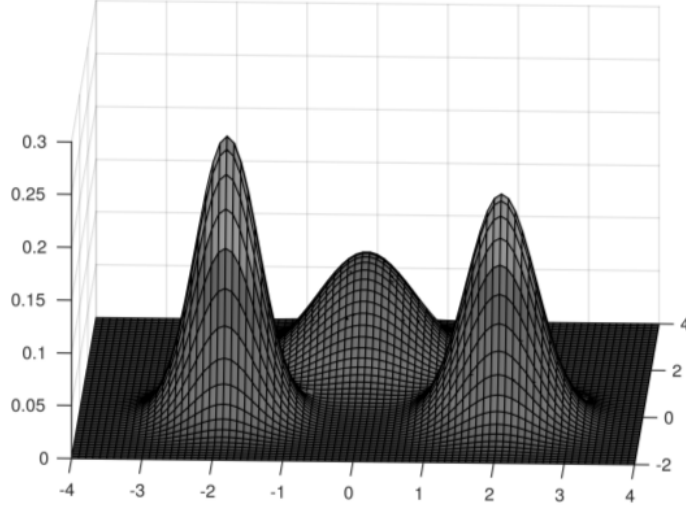Figure 4: *The density of a (uniform) mixture of a $\mathcal{N}((-0.05, 2.1)^\top, 0.5I)$, $\mathcal{N}((-1.9, -0.07)^\top, 0.2I)$ and $\mathcal{N}((2, -0.1)^\top, 0.25I)$ distribution.*

objects around these grid points are shown in Figure 5. In the grid point at $(0, 2)^\top$ we obtain not so many significant rejections as in the wedges with vertex $(2, 0)^\top$. Still, the dotted wedges show that there is a significant increase towards the mode which gives an indication for the presence of a mode as well. An improved procedure with a direct focus on the modes will be discussed in the following section.
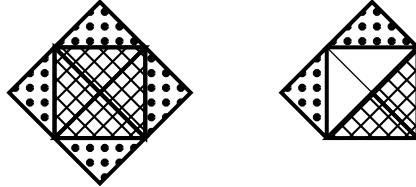

Figure 5: *Indications for the presence of a mode. Left panel: grid points $(-2, 0)^\top$ and $(2, 0)^\top$. Right panel: grid point $(0, 2)^\top$.*

4.3   Mode detection

In this section we demonstrate how the multiscale test can be successfully used for the localization of modes if the inference on the local monotonicity behaviour of the density is not included in the test statistic. More precisely, we consider the grid introduced in Section 4.2. Similar to the local test on modality, we conclude that the density has a mode close to a grid point if every test on every wedge whose vertex is given by the grid point rejects that the density is increasing on the respective wedge. We again recommend a calibrated version of the global

test, where the quantiles are chosen such that the probability of the discovery of a non-existing mode is approximately 5% if the data comes from a two-dimensional uniform distribution. The following results are based on $n = 2500$ observations and 1000 simulation runs.

We have investigated two densities, a constant density on the square $[-3.5, 3.5] \times [-1.5, 3.5]$ and the density with three modes presented in Figure 4. For the uniform distribution the test found a mode in 4.6% of the simulation runs. For the the tri-modal density the test detected in 78.9% cases a mode in the point $(-2, 0)^\top$, in 53.8% cases a mode in $(2, 0)^\top$ and in 7.4% cases a mode in the point $(0, 2)^\top$.

5 Real data example

Active Galactic Nuclei (AGN) consist of a supermassive black hole at their centre surrounded by an accretion disc. For some AGN classes, high-energy relativistic jets perpendicular to the disc are produced. In case this jet is pointing in the general direction of the Earth, the object is referred to as a blazar. The BL Lacertae type object Markarian (Mrk) 501 is such a blazar and one of the closest (in a distance of 4.7×10^8 light years) and brightest extragalactic sources in X-rays and gamma-rays. It is known as one of the most extreme blazars and features very strong and fast variability, making Mrk 501 a perfect candidate for probing AGN. Due to the strong emission over the entire electromagnetic spectrum, correlation studies between different energy bands (parts of the electromagnetic spectrum) are particularly interesting and will give insights into the processes inside an AGN, e.g. the emission models or the particle populations, since different spatial regions of the object may emit radiation of diverse energies. Therefore, the determination of the position of the radiation in a certain energy regime is of paramount importance.

Here, 19 individual observations of Mrk 501 in the year 2015, performed in photon counting (PC) mode by the *Swift*-XRT on board the *Swift* satellite, are analysed. The *Swift* satellite was launched in 2004 and is a multiwavelength space observatory with three instruments on board, the X-ray Telescope (XRT) being one of them, which is capable to observe X-rays in the $0.3 - 10\,\text{keV}$ energy regime. For each considered observation, the High Energy Astrophysics Science Archive Research Center (HEASARC[1]) provides an image, based on Level 2 event files that have been calibrated and screened by a standard pipeline. These images with a size of 1000×1000 pixels contain the information how many photons (i.e. which X-ray flux) have been recorded in each pixel during the exposure time. The exposure times of the analysed images range from about $100\,\text{s}$ to $1000\,\text{s}$. Due to different positions of the satellite in space and different alignments of its main axis, each image shows a slightly different region of the sky. Figure 6 provides an illustration of the data obtained from one observation.

By a combination of the 19 individual observations, we have at our disposal 49248 observations of X-ray photons with known positions of origin on the sky. Our aim is the precise localiza-

---

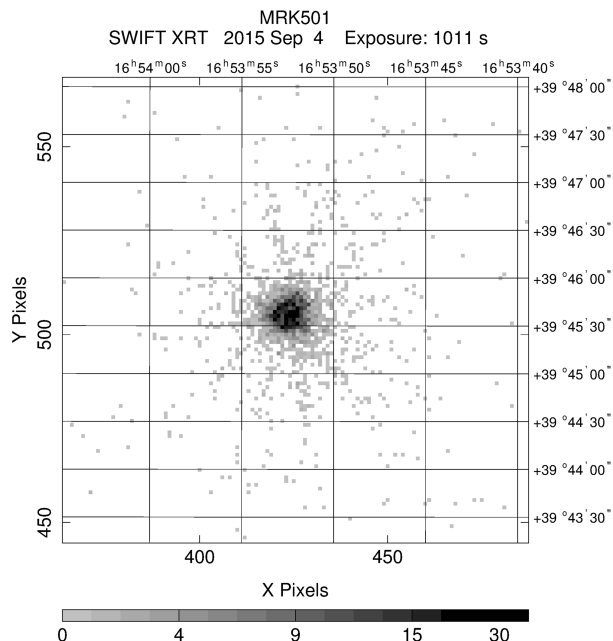[1]NASA/Goddard Space Flight Center, https://heasarc.gsfc.nasa.gov/.

Figure 6: *Observation of Mrk 501 from Sep. 4, 2015.*[2]

tion of the mode of the distribution. To this end, we conduct the test presented in Section 4.3 for a significance level of 0.01. We chose an equidistant grid covering $[253.446°, 253.586°] \times [39.64°, 39.88°]$ (corresponding to $[16\,\text{h}\,53\,\text{min}\,47.04\,\text{s}, 16\,\text{h}\,54\,\text{min}\,20.64\,\text{s}] \times [39°\,38'\,24'', 39°\,52'\,48''])$ consisting of 961 grid points with mesh size $0.008°$. The length of any wedge is $l = 0.004°$ and all angles are given by $\varphi = \frac{\pi}{4}$. Again, we used a calibrated version of the test where the quantiles are chosen such that a non-existing mode for a uniform distribution on $[253°, 253.8°] \times [39.5°, 40.1°]$ has been found in less than 1% of the simulation runs (based on 1000 simulation runs). Our test detected the mode of the distribution in $(253.466°, 39.760°)$ (corresponding to $(16\,\text{h}\,53\,\text{min}\,51.84\,\text{s}, 39°\,45'\,36''))$. The precision regarding the location of this mode is given by the mesh size $0.008°$.

In contrast to the high confidence detection possible within short exposure times for the brightest extragalactic sources, such as Mrk 501, the detection of fainter point sources is more challenging. Often only few photons reach the detector due to, e.g., the large distance to the source and the absorption of photons. Within the sky region of one *Swift*-XRT image, there might be multiple point sources in X-rays, but even background fluctuations can look like faint point sources. The study of the population of these point sources, the correlation to other energy bands, and variability studies contribute enormously to the understanding of the X-ray sky. This requires reliable methods for the detection and the determination of the position, including the confidence of a given calculation. In the following, the capability to determine multiple modes of faint point sources in *Swift*-XRT images is demonstrated. 18 images provided

---

[2]This picture has been created using HEAsoft, http://heasarc.nasa.gov/docs/software/lheasoft/.
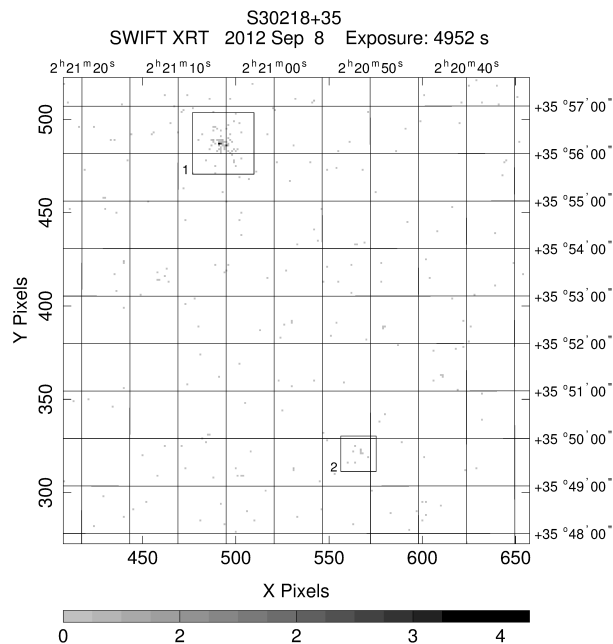
20

Figure 7: *Observation of the sky region around the blazar S3 0218+35 from Sep. 8, 2012.*[3]

by HEASARC of individual observations of the sky region around the blazar S3 0218+35 in the years 2005, 2012, 2014 and 2016 performed in PC mode by the *Swift*-XRT are analysed. The exposure times of the images provided by HEASARC range from 3000 s to 5000 s. Figure 7 shows one of these images. The two point sources are marked with a square. Figure 8 provides detailed images of the two point sources.

For the detection and the localization of the two point sources we conduct the test presented in Section 4.3 at a significance level of 0.01. Here, we have at our disposal 18061 observations. For this application, we chose an equidistant grid covering $[35.2°, 35.32°] \times [35.825°, 35.945°]$ (corresponding to $[2\,\mathrm{h}\,20\,\mathrm{min}\,48\,\mathrm{s}, 2\,\mathrm{h}\,21\,\mathrm{min}\,16.8\,\mathrm{s}] \times [35°\ 49'\ 30'', 35°\ 56'\ 42'']$) consisting of 961 grid points with mesh size 0.004°. The length of any wedge is $l = 0.002°$ and all angles are given by $\varphi = \frac{\pi}{4}$. The quantiles are chosen such that a non-existing mode for a uniform distribution on $[35°, 35.5°] \times [35.7°, 36.1°]$ has been found in less than 1% of the simulation runs (based on 1000 simulation runs). Our test detected the two point sources at $(35.212°, 35.829°)$ (corresponding to $(2\,\mathrm{h}\,20\,\mathrm{min}\,50.88\,\mathrm{s}, 35°\ 49'\ 44.4'')$) and at $(35.272°, 35.937°)$ (corresponding to $(2\,\mathrm{h}\,21\,\mathrm{min}\,5.28\,\mathrm{s}, 35°\ 56'\ 13.2'')$) at a precision of 0.004°.

---

[3]This picture has been created using HEAsoft, http://heasarc.nasa.gov/docs/software/lheasoft/.
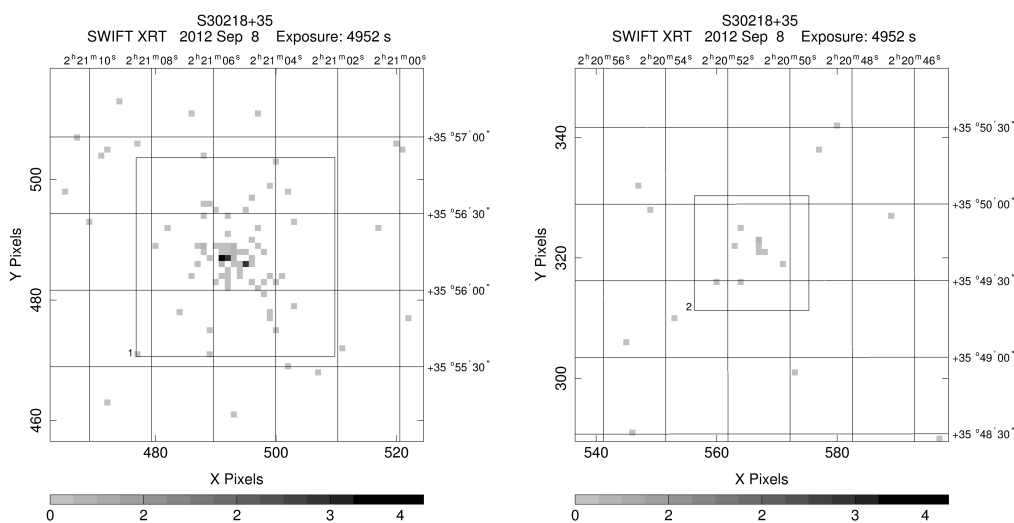
Figure 8: *Observations of the two points sources around S3 0218+35 from Sep. 8, 2012.*[4]

manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C1, C4) of the German Research Foundation (DFG).

References

Abraham, C., Biau, G., and Cadre, B. (2004). On the asymptotic properties of a simple estimate of the mode. *ESAIM Probab. Stat.*, 8:1–11 (electronic).

Burman, P. and Polonik, W. (2009). Multivariate mode hunting: data analytic tools with measures of significance. *J. Multivariate Anal.*, 100(6):1198–1218.

Chacón, J. and Duong, T. (2013). Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics*, 7:499–532.

Chan, K. S. and Tong, H. (2004). Testing for multimodality with dependent data. *Biometrika*, 91(1):113–123.

Chaudhuri, K., Dasgupta, S., Kpotufe, S., and von Luxburg, U. (2014). Consistent procedures for cluster tree estimation and pruning. *IEEE Trans. Inform. Theory*, 60(12):7900–7912.

Chaudhuri, P. and Marron, J. S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.*, 94(447):807–823.

Dasgupta, S. and Kpotufe, S. (2014). Optimal rates for k-NN density and mode estimation. *Advances in Neural Information Processing Systems*, 27:2555–2563.

Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence. III. *Ann. Statist.*, 19(2):633–667, 668–701.

---

[4]These pictures have been created using HEAsoft, http://heasarc.nasa.gov/docs/software/lheasoft/.

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.*, 36(4):1758–1785.

Duong, T., Cowling, A., Koch, I., and Wand, M. P. (2008). Feature significance for multivariate kernel density estimation. *Comput. Statist. Data Anal.*, 52(9):4225–4242.

Fisher, N. I. and Marron, J. S. (2001). Mode testing via the excess mass estimate. *Biometrika*, 88(2):499–517.

Genovese, C., Perone-Pacifico, M., Verdinelli, I., and Wasserman, L. A. (2015). Nonparametric inference for density modes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*

Godtliebsen, F., Marron, J. S., and Chaudhuri, P. (2002). Significance in scale space for bivariate density estimation. *J. Comput. Graph. Statist.*, 11(1):1–21.

Grund, B. and Hall, P. (1995). On the minimisation of L^p error in mode estimation. *Annals of Statistics*, 23:2264–2284.

Hall, P. and York, M. (2001). On the calibration of Silverman's test for multimodality. *Statistica Sinica*, 11:515–536.

Hartigan, J. A. (1987). Estimation of a convex density contour in two dimensions. *Journal of the American Statistical Association*, 82(397):267–270.

Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *Ann. Statist.*, 13(1):70–84.

Klemelä, J. (2005). Adaptive estimation of the mode of a multivariate density. *J. Nonparametr. Stat.*, 17(1):83–105.

Li, J., Ray, S., and Lindsay, B. G. (2007). A nonparametric statistical approach to clustering via mode identification. *J. Mach. Learn. Res.*, 8:1687–1723.

Mammen, E., Marron, J. S., and Fisher, N. (1991). Some asymptotics for multimodality tests based on kernel density estimates. *Probab. Theory Relat. Fields*, 91:115–132.

Minnotte, M. C. (1997). Nonparametric testing of the existence of modes. *Ann. Statist.*, 25(4):1646–1660.

Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.*, 86(415):738–746.

Pollard, D. (1981). Strong consistency of k-means clustering. *Ann. Statist.*, 9(1):135–140.

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.*, 23(3):855–881.

Romano, J. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Annals of Statistics*, 16:629–647.

Schmidt-Hieber, J., Munk, A., and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: confidence statements for qualitative features. *Ann. Statist.*, 41(3):1299–1328.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B*, 43(1):97–99.

Tsybakov, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution. *Problemy Peredachi Informatsii*, 26(1):38–45.

6 Proofs

The assertions of most of our results are split up in two parts, one for densities that are increasing and one for decreasing densities. Often, the proof for one case can be transferred in a straightforward way to the other one. In this situation, we only prove one case as the other case follows immediately by similar arguments. Only if this transfer is not obvious, we give details for both situations.

6.1 Proofs of the results of Section 2.3

Proof of Theorem 2.3: It is well-known that, given $X_{(N)}$ and N, the random variables $\tilde{F}(P_e X_{(1)}), \ldots, \tilde{F}(P_e X_{(N-1)})$ have the same distribution as the order statistics of $N - 1$ uniformly distributed and independent random variables on the interval $[0, 1]$. By an application of the law of iterated expectations, the assertion follows. □

Proof of Theorem 2.4: We only consider the case where the density f is increasing on K. As $\tilde{F}(P_e X_{(j)}) \neq 0$ almost surely for $j = 1, \ldots, N - 1$ (cf. Theorem 2.3) and $P_e X_{(j)} = P_e X_{(N)}$ implies $\tilde{F}(P_e X_{(j)}) = 1$, it is sufficient to prove

$$\tilde{F}(z) \leq \tilde{F}_U(z) = \frac{z^d}{(P_e X_{(N)})^d} \tag{5.1}$$

for all $z \in (0, P_e X_{(N)}]$. For this purpose, notice that the distribution function $\tilde{F}$ in Theorem 2.3 is given by

$$\tilde{F}(z) = \frac{\mathbb{P}(P_e X \leq z, X \in K_N | N, X_{(N)})}{\mathbb{P}(X \in K_N | N, X_{(N)})} = \frac{G(z)}{G(P_e X_{(N)})} \ ,$$

where the function G is defined by

$$G(z) = \int_0^z \int_{-t\tan(\varphi)}^{t\tan(\varphi)} \ldots \int_{-t\tan(\varphi)}^{t\tan(\varphi)} f(x_0 + te + s_1 e_1 + \ldots + s_{d-1} e_{d-1}) \, \mathrm{d}s_1 \ldots \, \mathrm{d}s_{d-1} \, \mathrm{d}t \ .$$

We now prove (5.1) by contradiction and assume that there exits $z^* \in (0, P_e X_{(N)}]$ such that

$$\int_0^{z^*} \tilde{f}(t) \, \mathrm{d}t = \frac{G(z^*)}{G(P_e X_{(N)})} > \frac{(z^*)^d}{(P_e X_{(N)})^d} = \frac{d}{(P_e X_{(N)})^d} \int_0^{z^*} t^{d-1} \, \mathrm{d}t, \tag{5.2}$$

where $\tilde{f}(t) := \frac{G'(t)}{G(P_e X_{(N)})}$ is the density of $\tilde{F}$. From (5.2) and the monotonicity of the integral it follows that there exists a point $t_0 \in (0, z^*]$ with

$$\tilde{f}(t_0) > \frac{dt_0^{d-1}}{(P_e X_{(N)})^d}. \tag{5.3}$$

For the following discussion, we introduce an alternative parametrization of the wedge K. As $\{e, e_1, \ldots, e_{d-1}\}$ is an orthonormal basis of $\mathbb{R}^d$, every $x \in K$ can be represented as $x = x_0 + t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1}$ for some $t_1 \in (0, l]$ and $s_i \in [-t_1 \tan(\varphi), t_1 \tan(\varphi)]$ for $i = 1, \ldots, d-1$. With the notation

$$e_0 := \frac{t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1}}{\|t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1}\|} \quad \text{and} \quad \tilde{t}_1 := \|t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1}\|, \tag{5.4}$$

we have $x = x_0 + \tilde{t}_1 e_0$, and the mapping of $(t_1, s_1, \ldots, s_{d-1})$ to $(\tilde{t}_1, e_0)$ defines a bijection. Hence, any $x \in K$ can also be uniquely represented by the vector e_0 and the scalar $\tilde{t}_1$ (see Figure 9).
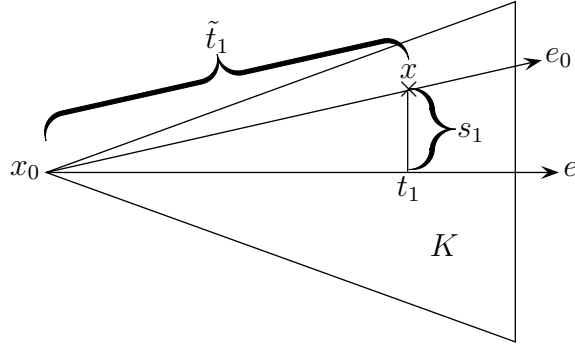

Figure 9: Representation of $x \in K$ for $d = 2$.

Let $t_1 \in [t_0, P_e X_{(N)}]$ and consider a point

$$x = x_0 + t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1} = x_0 + \tilde{t}_1 e_0 \in K_N,$$

where e_0 and $\tilde{t}_1$ are defined in (5.4). Let $\tilde{t}_0 = \frac{t_0}{t_1} \|t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1}\| \leq \tilde{t}_1$ and define

$$y := x_0 + \tilde{t}_0 e_0 = x_0 + \tilde{t}_0 \frac{t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1}}{\|t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1}\|}.$$

A straightforward calculation shows that $y = x_0 + t_0 e + \tilde{s}_1 e_1 + \ldots + \tilde{s}_{d-1} e_{d-1}$, where $\tilde{s}_i = \frac{t_0}{t_1} s_i$, $(i = 1, \ldots, d-1)$. Note that $P_e y = t_0$ and that $f(y) \leq f(x)$, as f is increasing on the wedge K. We therefore obtain

$$\tilde{f}(t_1) = \int_{-t_1 \tan(\varphi)}^{t_1 \tan(\varphi)} \cdots \int_{-t_1 \tan(\varphi)}^{t_1 \tan(\varphi)} \frac{f(x_0 + t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1})}{G(P_e X_{(N)})} \, \mathrm{d}s_1 \ldots \mathrm{d}s_{d-1}$$

$$\geq \int_{-t_1 \tan(\varphi)}^{t_1 \tan(\varphi)} \cdots \int_{-t_1 \tan(\varphi)}^{t_1 \tan(\varphi)} \frac{f(x_0 + t_0 e + \frac{t_0}{t_1} s_1 e_1 + \ldots + \frac{t_0}{t_1} s_{d-1} e_{d-1})}{G(P_e X_{(N)})} \, \mathrm{d}s_1 \ldots \mathrm{d}s_{d-1}.$$

$$= \left(\frac{t_1}{t_0}\right)^{d-1} \int_{-t_0 \tan(\varphi)}^{t_0 \tan(\varphi)} \cdots \int_{-t_0 \tan(\varphi)}^{t_0 \tan(\varphi)} \frac{f(x_0 + t_0 e + \tilde{s}_1 e_1 + \ldots + \tilde{s}_{d-1} e_{d-1})}{G(P_e X_{(N)})} \, \mathrm{d}\tilde{s}_1 \ldots \mathrm{d}\tilde{s}_{d-1}$$

$$= \left(\frac{t_1}{t_0}\right)^{d-1} \tilde{f}(t_0).$$

25

Using (5.3) we have $\tilde{f}(t_1) > \left(\frac{t_1}{t_0}\right)^{d-1} \frac{dt_0^{d-1}}{(P_e X_{(N)})^d} = \frac{dt_1^{d-1}}{(P_e X_{(N)})^d}$ for any point $t_1 \in [t_0, P_e X_{(N)}]$, and from (5.2) we conclude

$$1 = \int_0^{P_e X_{(N)}} \tilde{f}(t) \, dt > \int_0^{z^*} \frac{dt^{d-1}}{(P_e X_{(N)})^d} \, dt + \int_{z^*}^{P_e X_{(N)}} \frac{dt^{d-1}}{(P_e X_{(N)})^d} \, dt = \int_0^{P_e X_{(N)}} \frac{dt^{d-1}}{(P_e X_{(N)})^d} \, dt = 1,$$

which is a contradiction and proves the first assertion of Theorem 2.4. $\qquad\square$

Proof of the representation (2.5): It follows from Theorem 2.3 that for fixed i, the random variables $\tilde{F}^i(P_{e_n^i} X_1^i), \ldots, \tilde{F}^i(P_{e_n^i} X_{N^i-1}^i)$ are independent and uniformly distributed on the interval $[0,1]$, given $\{X_j^i \in K_{N^i}^i\}$ $(j = 1, \ldots, N^i - 1)$ and N^i. Here, $\tilde{F}^i$ denotes the conditional distribution function of $P_{e_n^i} X$ on K^i. Recall that the wedges $K_{N_i}^i$, $i = 1, \ldots, M_n$, are disjoint. Standard arguments and the law of iterated expectations yield stochastic independence of observations lying in different wedges. $\qquad\square$

Proof of Theorem 2.5: Let $\boldsymbol{a} \in \{-1, 1, 0\}^{M_n}$ be the vector determined by

$$a_i = \begin{cases} 1, & \text{if } f \text{ is increasing on } K^i, \\ -1, & \text{if } f \text{ is decreasing on } K^i, \\ 0, & \text{else.} \end{cases}$$

The probability of at least one false rejection among all tests in (2.9) and (2.10) can be estimated by Theorem 2.4, that is

$$\begin{aligned} p := \ & \mathbb{P}_f\big(\exists i \in \{i \mid a_i \neq 0\} : a_i T_{K^i} < -\tilde{c}_{K^i}(\alpha) \big| N^1, \ldots, N^{M_n}\big) \\ \leq \ & \mathbb{P}_f\big(\exists i \in \{i \mid a_i \neq 0\} : a_i T_{K^i}^U < -\tilde{c}_{K^i}(\alpha) \big| N^1, \ldots, N^{M_n}\big) \\ \leq \ & 1 - \mathbb{P}\big(|T_{K^i}^U| \leq \tilde{c}_{K^i}(\alpha) \text{ for all } i = 1, \ldots, M_n \big| N^1, \ldots, N^{M_n}\big). \end{aligned}$$

Using (2.6) we further deduce

$$\begin{aligned} p \leq \ & 1 - \mathbb{P}\Big(\Big(\sqrt{\frac{3}{N^i - 1}}\,|T_{K^i}^U| - \Gamma\Big(\frac{N^i}{n-1}\Big)\Big) \leq \tilde{\kappa}_n(\alpha) \text{ for all } i = 1, \ldots, M_n \Big| N^1, \ldots, N^{M_n}\Big) \\ = \ & 1 - \mathbb{P}\Big(\max_{i=1,\ldots,M_n}\Big(\sqrt{\frac{3}{N^i - 1}}\,|T_{K^i}^U| - \Gamma\Big(\frac{N^i}{n-1}\Big)\Big) \leq \tilde{\kappa}_n(\alpha)\big| N^1, \ldots, N^{M_n}\Big) = \alpha. \end{aligned}$$

$\qquad\square$

6.2   Further Results

This section provides a general consistency result which is the main ingredient for the proof of Theorems 2.6 and 3.3. The consistency result stated in Theorem 6.3 below is a more general result. The following notation is used throughout this section for the precise statement of the assumptions on the density f.

Definition 6.1. *Let K be a wedge with vertex x_0 and $j \in \{1, 2\}$. We define*

$$
\begin{aligned}
\inf_K f^{(j)} &:= \inf_{x_0 + \tilde{t}e_0 \in K, 0 \leq \tilde{s} < \tilde{t}} \frac{f(x_0 + \tilde{t}e_0) - f(x_0 + \tilde{s}e_0)}{\tilde{t}^j - \tilde{s}^j}, \\
\sup_K f^{(j)} &:= \sup_{x_0 + \tilde{t}e_0 \in K, 0 \leq \tilde{s} < \tilde{t}} \frac{f(x_0 + \tilde{t}e_0) - f(x_0 + \tilde{s}e_0)}{\tilde{t}^j - \tilde{s}^j}.
\end{aligned}
\tag{5.5}
$$

For a better interpretation of Definition 6.1, let f be twice continuously differentiable in a neighbourhood of x_0. A straightforward application of the mean value theorem shows

$$
\inf_K f^{(1)} = \inf_{x_0 + \tilde{t}e_0 \in K} \langle \nabla f(x_0 + \tilde{t}e_0), e_0 \rangle, \quad \sup_K f^{(1)} = \sup_{x_0 + \tilde{t}e_0 \in K} \langle \nabla f(x_0 + \tilde{t}e_0), e_0 \rangle.
$$

If we have a mode in x_0, the gradient $\nabla f(x_0)$ vanishes and $\sup_K f^{(1)}$ vanishes as well. Similarly, a Taylor expansion of order 2 yields

$$
\inf_K f^{(2)} = \frac{1}{2} \inf \left\{ \langle e_0, H_f(x_0)e_0 \rangle \big| \{ x_0 + \tilde{t}e_0, \tilde{t} \geq 0 \} \cap K \neq \emptyset \big| \right\} + o(1)
$$

$$
\sup_K f^{(2)} = \frac{1}{2} \sup \left\{ \langle e_0, H_f(x_0)e_0 \rangle \big| \{ x_0 + \tilde{t}e_0, \tilde{t} \geq 0 \} \cap K \neq \emptyset \big| \right\} + o(1)
$$

for $l \to 0$, where $H_f(x_0)$ denotes the Hessian of f in x_0. In the situation of Theorem 2.6, the condition $e_0^\top H_f(x_0)e_0 \leq -c < 0$ holds for all $e_0 \in \mathbb{R}^d$ with $\|e_0\| = 1$. Thus, only the case $j = 2$ is relevant for its proof. However, if the assumption is violated and the Hessian is indefinite, the results can be generalized in a straightforward manner by considering orders $j \geq 3$ as well. For the proof of Theorem 3.3, the case $j = 1$ will be used.
To simplify notation, let

$$
\tilde{\varphi} = \tan(\varphi) \text{ for } \varphi \in \left(0, \frac{\pi}{2} \right) \quad \text{and} \quad F(K) = \int_K f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.
$$

The quantities

$$
H_+^j(f, K) := \frac{(2\tilde{\varphi})^{d-1} l^{d+j} \inf_K f^{(j)}}{\sqrt{F(K)}} \ , \quad H_-^j(f, K) := \frac{(2\tilde{\varphi})^{d-1} l^{d+j} \sup_K f^{(j)}}{\sqrt{F(K)}} \quad (j = 1, 2)
$$

depend on the size of the wedge K through its length l and angle φ and on the monotonicity of f on K and are the key objects in the following discussion. We begin by showing that the quantiles $\tilde{\kappa}_n(\alpha)$ and $\overline{\kappa}_n(\alpha)$ defined in (2.6) and (3.2) are bounded from above by a constant independent of $M_n, N^i, i = 1, \ldots, M_n$, and n. As a consequence, the same holds for the quantiles $\tilde{\kappa}_n'(\alpha)$ defined in (2.11).

Theorem 6.2. *There exists a constant $A > 0$, such that $\max\{\tilde{\kappa}_n(\alpha), \overline{\kappa}_n(\alpha)\} \leq A$.*

Proof: We only consider $\tilde{\kappa}_n$ as the result for $\overline{\kappa}_n$ can be shown similarly. From the discussion in Section 2.3, it follows that

$$T^U_{K^i} \stackrel{d}{=} \sum_{j=1}^{N^i-1} \beta(U^i_j) \ \text{for} \ i=1,\ldots,M_n$$

(conditionally on $N^1,\ldots,N^{M_n}$), where U^i_j, $j=1,\ldots,N^i-1, i=1,\ldots,M_n$, are independent uniformly distributed random variables on the interval $[0,1]$. Recall the definition of $T_{jk}(\boldsymbol{U})$ in (2.1), then we will show a the end of the proof that, conditionally on $N^1,\ldots,N^{M_n}$,

$$T_{\sum_{k=1}^{i-1} N^k, \sum_{k=1}^i N^k}(\boldsymbol{U}) = \sum_{\ell=\sum_{k=1}^{i-1} N^k + 1}^{\sum_{k=1}^i N^k - 1} \beta\big(U_{(\ell; \sum_{k=1}^{i-1} N^k, \sum_{k=1}^i N^k)}\big) \stackrel{d}{=} \sum_{j=1}^{N^i-1} \beta(U^i_j) \tag{5.6}$$

$(i=1,\ldots,M_n)$. As the statistic $T_n(\boldsymbol{U})$ defined in (2.2) calculates the maximum over more scales than the statistic defined in (2.6), we obtain $\tilde{\kappa}_n \le \kappa_n$, where κ_n is the $(1-\alpha)$-quantile of the statistic $T_n(\boldsymbol{U})$ defined in (2.2). By Theorem 3.1 in Dümbgen and Walther (2008), there exists an upper bound $A > 0$ for κ_n, which is independent of n. This completes the proof.

For a proof of (5.6), we finally note that for a sample of independent uniformly distributed random variables $U_1,\ldots,U_n$ on the interval $[0,1]$ and fixed $1 \le j < k \le n$ with $k-j > 1$, the random variables $U_{(l;j,k)} = (U_{(l)} - U_{(j)})/(U_{(k)} - U_{(j)})$ $(l = j+1,\ldots,k-1)$ have the same distribution as the order statistic of $k-j-1$ independent uniformly distributed random variables on the interval $[0,1]$. $\square$

Following the notation from Section 2.3 (recall that $\Gamma(\delta) = \sqrt{2\log(\frac{\exp(1)}{\delta})}$), we define the quantiles

$$c_{K^i_n}(\alpha) = \sqrt{\frac{N^i-1}{3}}\Big(A + \Gamma\Big(\frac{N^i}{n-1}\Big)\Big)$$

$(i=1,\ldots,M_n)$ and provide a general consistency result for locally increasing and decreasing densities.

Theorem 6.3. *Let $j \in \{1,2\}$ and $\mathcal{K}_n = \{K^1_n,\ldots,K^{M_n}_n\}$ be a family of M_n pairwise disjoint wedges with length $l_n > 0$ and angle $\varphi_n \in (0, \frac{\pi}{2})$.*

   (i) *If (for all $i=1,\ldots,M_n$) the condition $H^j_+(f, K^i_n) \ge D\big(\frac{\Gamma(F(K^i_n))}{\sqrt{2}} + b_n\big)\frac{1}{\sqrt{n}}$ holds for some constant*

$$D > \frac{j(d+j)2\sqrt{2}}{(2d+j)\Big(-1 + \{1 + \frac{2j^2}{(2d+j)^2}\}^{\frac{1}{2}}\Big)\Big(1 - \frac{j^2}{2(2d+j)^2}\{-1 + [1 + 4\frac{(2d+j)^2}{j^2}]^{\frac{1}{2}}\}\Big)^{\frac{1}{2}}} \tag{5.7}$$

   and a positive sequence b_n satisfying $b_n \to \infty$ and $b_n = o(\sqrt{\log(n)})$ as $n \to \infty$, then

$$\mathbb{P}\big(T_{K^i_n} > c_{K^i_n}(\alpha) \ \text{for all} \ i=1,\ldots,M_n\big) \to 1.$$

(ii) If (for all $i = 1 \ldots, M_n$) the condition $H_-^j(f, K_n^i) \le -D\left(\frac{\Gamma(F(K_n^i))}{\sqrt{2}} + b_n\right)\frac{1}{\sqrt{n}}$ holds for some constant

$$D > \frac{2\sqrt{2}(2d+j)(d+j)}{j\left(1 - \frac{d}{d+j}\right)^{(d+j)/d}\left[1 - \frac{d^2}{2(2d+j)^2}\left(-1 + \left\{1 + 4\left(\frac{2d+j}{d}\right)^2\right\}^{1/2}\right)\right]^{1/2}}$$

and a positive sequence b_n satisfying $b_n \to \infty$ and $b_n = o(\sqrt{\log(n)})$ as $n \to \infty$, then

$$\mathbb{P}\big(T_{K_n} < -c_{K_n}(\alpha) \text{ for all } i = 1, \ldots, M_n\big) \to 1.$$

Remark 6.4. It follows from the proof of Theorem 6.3 below, that the bounds on $H_+^j(f, K_n)$ resp. $H_-^j(f, K_n)$ imply in particular that $M_n = o(n)$ as $n \to \infty$.

The proof of Theorem 6.3 is divided into eight parts: seven technical lemmas (Lemma 6.5 - Lemma 6.11) and the main part of the proof. We first state and prove the technical lemmas and finally combine the results in order to complete the proof of Theorem 6.3. For each of the lemmas, we assume that the conditions of Theorem 6.3 hold. At first, we consider only one fixed wedge K with length $l_n = l$ and angle $\varphi_n = \varphi$. To simplify notation, let

$$\overline{F}_\pm^j(z) := z + c_\pm \left(\frac{2^{d-1}}{d+j}z^{\frac{d+j}{d}} - \frac{2^{d-1}}{d+j}z\right)$$

for $j \in \{1, 2\}$ and $z \in [0, 1]$ with

$$c_+^j := \frac{(\tan(\varphi))^{d-1}(P_e X_{(N)})^{d+j}\inf_K f^{(j)}}{F(K_N)} \quad \text{and} \quad c_-^j := \frac{(\tan(\varphi))^{d-1}(P_e X_{(N)})^{d+j}\sup_K f^{(j)}}{F(K_N)}.$$

Lemma 6.5.

(i) If $\inf_K f^{(j)} > 0$, then $\tilde{F}(z) \le \overline{F}_+^j\left(\frac{z^d}{(P_e X_{(N)})^d}\right)$ for $z \in (0, P_e X_{(N)}]$.

(ii) If $\sup_K f^{(j)} < 0$, then $\tilde{F}(z) \ge \overline{F}_-^j\left(\frac{z^d}{(P_e X_{(N)})^d}\right)$ for $z \in (0, P_e X_{(N)}]$.

Proof: We only prove the first part and define the auxiliary function

$$\overline{f}_+^j(s_1, \ldots, s_{d-1}, t) := \mathbb{1}_{K_N}(x_0 + te + s_1 e_1 + \ldots + s_{d-1}e_{d-1})$$

$$\cdot \left(\frac{1}{|K_N|} + \left(\frac{\inf_K f^{(j)}}{F(K_N)}\right)\left(t^j - \frac{d}{d+j}(P_e X_{(N)})^j\right)\right),$$

where $|K_N| := \int_{K_N} 1\, d\boldsymbol{x}$ denotes the volume of K_N. Note that

$$\int_0^z \int_{-t\tan(\varphi)}^{t\tan(\varphi)} \cdots \int_{-t\tan(\varphi)}^{t\tan(\varphi)} \overline{f}_+^j(s_1, \ldots, s_{d-1}, t)\, ds_1 \ldots\, ds_{d-1}\, dt$$

$$= \frac{z^d}{(P_e X_{(N)})^d} + \int_0^z (2\tan(\varphi)t)^{d-1}\left(\frac{\inf_K f^{(j)}}{F(K_N)}\right)\left(t^j - \frac{d}{d+j}(P_e X_{(N)})^j\right) dt \tag{5.8}$$

$$= \frac{z^d}{(P_e X_{(N)})^d} + c_+\left(\frac{2^{d-1}}{d+j}\frac{z^{d+j}}{(P_e X_{(N)})^{d+j}} - \frac{2^{d-1}}{d+j}\frac{z^d}{(P_e X_{(N)})^d}\right) = \overline{F}_+^j\left(\frac{z^d}{(P_e X_{(N)})^d}\right).$$

29

In particular, as $\overline{F}^j_+(1) = 1$, the function $\overline{f}^j_+$ defines a density on K_N. We now prove assertion *(i)* by contradiction and assume that there exists $z^* \in (0, P_e X_{(N)}]$, such that

$$\tilde{F}(z^*) = \int_0^{z^*} \tilde{f}(t)dt = \frac{G(z^*)}{G(P_e X_{(N)})} > \overline{F}^j_+ \left(\frac{(z^*)^d}{(P_e X_{(N)})^d} \right) = \int_0^{z^*} h^j_+(t)dt, \tag{5.9}$$

where $\tilde{f}$ and

$$h^j_+(t) := \int_{-t\tan(\varphi)}^{t\tan(\varphi)} \ldots \int_{-t\tan(\varphi)}^{t\tan(\varphi)} \overline{f}^j_+(s_1, \ldots, s_{d-1}, t) \, \mathrm{d}s_1 \ldots \mathrm{d}s_{d-1}$$

denote the density of $\tilde{F}$ and $\overline{F}^j_+$, respectively. Due to the monotonicity of the integral, there exists a point $t_0 \in (0, z^*]$ with

$$\tilde{f}(t_0) > h^j_+(t_0), \tag{5.10}$$

which implies

$$\tilde{f}(t_1) > h^j_+(t_1) \text{ for all } t_1 \in [t_0, P_e X_{(N)}]. \tag{5.11}$$

For a proof of (5.11), let $t_1 > t_0$ and $x = x_0 + t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1} \in K_N$. As in the proof of Theorem 2.4, we use the representation $x = x_0 + \tilde{t}_1 e_0$, where $\tilde{t}_1$ and e_0 are defined in (5.4). For $\tilde{t}_0 := \frac{t_0}{t_1} \tilde{t}_1 < \tilde{t}_1$, let

$$y := x_0 + \tilde{t}_0 e_0 = x_0 + t_0 e_0 + \frac{t_0}{t_1} s_1 e_1 + \ldots + \frac{t_0}{t_1} s_{d-1} e_{d-1} \in K_N.$$

Using $\inf_K f^{(j)} > 0$, we find

$$
\begin{aligned}
f(x) - f(y) &= f(x_0 + \tilde{t}_1 e_0) - f(x_0 + \tilde{t}_0 e_0) = (\tilde{t}^j_1 - \tilde{t}^j_0) \frac{f(x_0 + \tilde{t}_1 e_0) - f(x_0 + \tilde{t}_0 e_0)}{\tilde{t}^j_1 - \tilde{t}^j_0} \tag{5.12}\\
&\geq (\tilde{t}^j_1 - \tilde{t}^j_0) \inf_K f^{(j)} = \tilde{t}^j_1 \left(1 - \frac{t^j_0}{t^j_1} \right) \inf_K f^{(j)} \geq (t^j_1 - t^j_0) \inf_K f^{(j)},
\end{aligned}
$$

where the last estimate follows since $\tilde{t}_1 \geq t_1$. Recall that $G(P_e X_{(N)}) = F(K_N)$, then we obtain

$$
\begin{aligned}
\tilde{f}(t_1) &= \int_{-t_1\tan(\varphi)}^{t_1\tan(\varphi)} \ldots \int_{-t_1\tan(\varphi)}^{t_1\tan(\varphi)} \frac{f(x_0 + t_1 e + s_1 e_1 + \ldots + s_{d-1} e_{d-1})}{G(P_e X_{(N)})} \, \mathrm{d}s_1 \ldots \mathrm{d}s_{d-1}\\
&\geq \int_{-t_1\tan(\varphi)}^{t_1\tan(\varphi)} \ldots \int_{-t_1\tan(\varphi)}^{t_1\tan(\varphi)} \frac{1}{F(K_N)} \Big(f(x_0 + t_0 e + \frac{t_0}{t_1} s_1 e_1 + \ldots + \frac{t_0}{t_1} s_{d-1} e_{d-1})\\
&\qquad\qquad\qquad + (t^j_1 - t^j_0) \inf_K f^{(j)} \Big) \, \mathrm{d}s_1 \ldots \mathrm{d}s_{d-1}.
\end{aligned}
$$

30

A change of variables yields

$$\tilde{f}(t_1) \geq \left(\frac{t_1}{t_0}\right)^{d-1} \int_{-t_0 \tan(\varphi)}^{t_0 \tan(\varphi)} \cdots \int_{-t_0 \tan(\varphi)}^{t_0 \tan(\varphi)} \frac{f(x_0 + t_0 e + \tilde{s}_1 e_1 + \ldots + \tilde{s}_{d-1} e_{d-1})}{F(K_N)} \, d\tilde{s}_1 \ldots d\tilde{s}_{d-1}$$

$$+ (2\tan(\varphi)t_1)^{d-1}(t_1^j - t_0^j)\frac{\inf_K f^{(j)}}{F(K_N)}$$

$$= \left(\frac{t_1}{t_0}\right)^{d-1} \tilde{f}(t_0) + (2\tan(\varphi)t_1)^{d-1}(t_1^j - t_0^j)\frac{\inf_K f^{(j)}}{F(K_N)},$$

and straightforward calculations show that

$$h_+^j(t_1) = \left(\frac{t_1}{t_0}\right)^{d-1} h_+^j(t_0) + (2\tan(\varphi)t_1)^{d-1}(t_1^j - t_0^j)\frac{\inf_K f^{(j)}}{F(K_N)}$$

$$< \left(\frac{t_1}{t_0}\right)^{d-1} \tilde{f}(t_0) + (2\tan(\varphi)t_1)^{d-1}(t_1^j - t_0^j)\frac{\inf_K f^{(j)}}{F(K_N)} \leq \tilde{f}(t_1),$$

where we used (5.10) to obtain the strict inequality. From (5.9) and (5.11) we also get

$$\int_0^{P_e X_{(N)}} \tilde{f}(t) \, dt > \int_0^{z^*} h_+^j(t) \, dt + \int_{z^*}^{P_e X_{(N)}} h_+^j(t) \, dt = \overline{F}_+^j(1) = 1,$$

which contradicts the condition $1 = \tilde{F}(P_e X_{(N)}) = \int_0^{P_e X_{(N)}} \tilde{f}(t) \, dt$. This completes the proof of Lemma 6.5. $\qquad\square$

Lemma 6.6.

(i) *If* $\inf_K f^{(j)} > 0$, *we have* $\sum_{i=1}^{N-1} \beta\big((\overline{F}_+^j)^{-1}(\tilde{F}(P_e X_{(i)}))\big) \leq T_K$.

(ii) *If* $\sup_K f^{(j)} < 0$, *we have* $\sum_{i=1}^{N-1} \beta\big((\overline{F}_-^j)^{-1}(\tilde{F}(P_e X_{(i)}))\big) \geq T_K$.

Proof: We only prove the first part and begin showing that the function $z \mapsto \overline{F}_+^j(z)$ is strictly increasing for $z \in [0,1]$. Recalling the representation (5.8), it is sufficient to prove that the inequality

$$\overline{f}_+^j(s_1, \ldots, s_{d-1}, t) > 0 \quad \text{for } t \in (0, P_e X_{(N)}]$$

holds for all $(s_1, \ldots, s_{d-1}) \in [-t \tan\varphi, t \tan\varphi]$. For the sake of simplicity, we suppress the dependence of $\overline{f}_+^j$ on $(s_1, \ldots, s_{d-1})$ and note that the function $t \mapsto \overline{f}_+^j(t) = \overline{f}_+^j(t, s_1, \ldots, s_{d-1})$ is strictly increasing. Therefore, it remains to show $\overline{f}_+^j(0) \geq 0$. We prove this inequality by contradiction and assume that $\overline{f}_+^j(0) < 0$. For $x = x_0 + te + s_1 e_1 + \ldots + s_{d-1} e_{d-1} \in K_N$, it follows (using $\tilde{t}_0 = 0$ in (5.12))

$$\frac{f(x)}{F(K_N)} \geq \frac{f(x) - f(x_0)}{F(K_N)} \geq t^j \frac{\inf_K f^{(j)}}{F(K_N)} = \overline{f}_+^j(t) - \overline{f}_+^j(0) > \overline{f}_+^j(t). \tag{5.13}$$

31

Integrating both sides of (5.13) leads to a contradiction. Consequently, the map $z \mapsto \bar{F}^j_+(z)$ is strictly increasing on $[0, 1]$, which implies (using Lemma 6.5 and the monotonicity of the function β) that $\beta((\bar{F}^j_+)^{-1}(\tilde{F}(P_e X_{(j)}))) \leq \beta\big(\frac{(P_e X_{(j)})^d}{P_e X_{(N)})^d}\big)$, whenever $\tilde{F}(P_e X_{(j)}) \neq 0$ and $P_e X_{(j)} \neq P_e X_{(N)}$. However, it is easy to see that these cases correspond to $P_e X_{(j)} = 0$ and $\tilde{F}(P_e X_{(j)}) = 1$, where there is in fact equality. Thus the proof of the first part is completed. $\qquad\square$

The conditional expectation considered in the following lemma is used to derive a bound on the corresponding conditional probability via the Hoeffding inequality in Lemma 6.11 below.

Lemma 6.7.

(i) $\qquad\qquad \mathbb{E}\Big(\sum_{i=1}^{N-1} \beta\big((\bar{F}^j_+)^{-1}(\tilde{F}(P_e X_{(i)}))\big) | N, X_{(N)} \Big) = \frac{2^{d-1}j(N-1)}{(2d+j)(d+j)} c^j_+.$

(ii) $\qquad\qquad \mathbb{E}\Big(\sum_{i=1}^{N-1} \beta\big((\bar{F}^j_-)^{-1}(\tilde{F}(P_e X_{(i)}))\big) | N, X_{(N)} \Big) = \frac{2^{d-1}j(N-1)}{(2d+j)(d+j)} \tilde{c}^j_-.$

Proof: We only prove the first part. Let $U_1, \ldots, U_{N-1}$ be independent uniformly distributed random variables on the interval $[0, 1]$. Theorem 2.3 yields

$$\sum_{i=1}^{N-1} \beta\left((\bar{F}^j_+)^{-1}(\tilde{F}(P_e X_{(i)})) \right) \overset{d}{=} \sum_{i=1}^{N-1} \beta\left((\bar{F}^j_+)^{-1}(U_i) \right),$$

given N and $X_{(N)}$. The assertion now follows from

$$\mathbb{E}\big(\beta((\bar{F}^j_+)^{-1}(U_1)) | N, X_{(N)}\big) = \int_0^1 \beta(x) g^j_+(x) \, dx = \frac{2^{d-1}j}{(2d+j)(d+j)} c^j_+,$$

where $g^j_+(x) := \frac{d}{dx} \bar{F}^j_+(x) = 1 + c^j_+\big(\frac{2^{d-1}}{d} x^{\frac{d+j}{d}-1} - \frac{2^{d-1}}{d+j} \big).$ $\qquad\square$

In the following, we consider a sequence of wedges $(K_n)_{n\in\mathbb{N}}$ given by vertices x^n_0, directions e^n, lengths $l_n > 0$ and angles $\varphi_n \in (0, \frac{\pi}{2})$. Furthermore, we denote by $e^n_1, \ldots, e^n_{d-1}$ the orthonormal basis of $(\text{span}\{e^n\})^\perp$ and let $\delta_n = F(K_n)$, $\delta_{N_n} = F(K_{N_n})$. Lemma 6.8 and Lemma 6.9 below ensure the feasibility of our procedure in an asymptotic sense. They show that the random wedge K_{N_n} is of similar size than the deterministic, predefined wedge K_n, i.e. that its complement $K_n \backslash K_{N_n}$ is small. Note that the test procedure can only be consistent if the wedge K_n contains a sufficiently large number of observations. Therefore, we introduce for $\gamma \in (0, \frac{1}{2}]$ and $0 < \varepsilon < 1$ the conditional probability $\mathbb{P}_{\mathcal{N}_n}$ given the event

$$\mathcal{N}_n = \Big\{ N_n \geq (1-\gamma)n\delta_n, \; \frac{|K_{N_n}|}{|K_n|} \geq 1 - \varepsilon \Big\}. \tag{5.14}$$

The results of Lemma 6.11 below are only shown for $\mathbb{P}_{\mathcal{N}_n}$. However, the following Lemmas 6.8 and 6.9 demonstrate that these conditions are asymptotically negligible. For example, Lemma 6.8 shows that, with increasing n, the wedge $K_{N_n} \subseteq K_n$ approximates K_n in probability at an exponential rate.
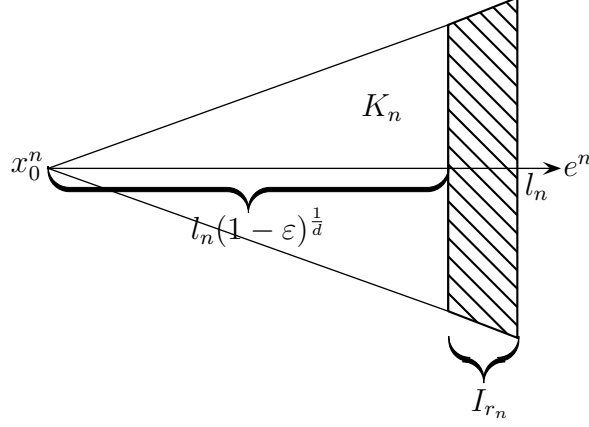
Figure 10: *The section I_{r_n}.*

Lemma 6.8. *Let $j \in \{1, 2\}$ and $0 < \varepsilon < 1$.*

(i) If the assumptions of Theorem 6.3 (i) are satisfied, then

$$\mathbb{P}\left(N_n = 0 \ \text{ or } \ \frac{|K_{N_n}|}{|K_n|} < 1 - \varepsilon\right) \leq \exp\left(-\frac{D}{d+j}\sqrt{\frac{n\delta_n}{2}}\Gamma(\delta_n)\left(1 - (1-\varepsilon)^{\frac{d+j}{d}}\right)\right).$$

(ii) If the assumptions of Theorem 6.3 (ii) are satisfied, then

$$\mathbb{P}\left(N_n = 0 \ \text{ or } \ \frac{|K_{N_n}|}{|K_n|} < 1 - \varepsilon\right) \leq \exp\left(-\frac{D}{d+j}\sqrt{\frac{n\delta_n}{2}}\Gamma(\delta_n)\left((1-\varepsilon)^{\frac{d+j}{d}} + \frac{d+j}{d}\varepsilon - 1\right)\right).$$

Proof: We only prove the first part. Since $\frac{|K_{N_n}|}{|K_n|} = \frac{(P_{e^n}X_{(N_n)})^d}{l_n^d} < 1 - \varepsilon$ if and only if $P_{e^n}X_{(N_n)} < l_n(1-\varepsilon)^{\frac{1}{d}}$, we obtain

$$\left\{N_n = 0 \ \text{ or } \ \frac{|K_{N_n}|}{|K_n|} < 1 - \varepsilon\right\} = \left\{N_n = 0 \ \text{ or } \ P_{e^n}X_{(N_n)} < l_n(1-\varepsilon)^{\frac{1}{d}}\right\}.$$

Define

$$I_{r_n} := K_n \cap \left\{x \in \mathbb{R}^d : \ P_{e^n}x \geq l_n(1-\varepsilon)^{\frac{1}{d}}\right\}$$

(see Figure 10), then $\{N_n = 0 \ \text{ or } \ \frac{|K_{N_n}|}{|K_n|} < 1 - \varepsilon\} \subseteq \{\text{ no observation in } I_{r_n}\}$. Now, recall for a proof of part *(i)* from (5.12) that (with $t_0 = 0$ and $t_1 = t$) $f(x_0^n + te^n + s_1e_1^n + \ldots + s_{d-1}e_{d-1}^n) \geq t^j \inf_{K_n} f^{(j)}$. Thus,

$$
\begin{aligned}
p_{f,I_{r_n}} &:= \int_{I_{r_n}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int_{l_n(1-\varepsilon)^{\frac{1}{d}}}^{l_n} \int_{-t\tilde{\varphi}_n}^{t\tilde{\varphi}_n} \cdots \int_{-t\tilde{\varphi}_n}^{t\tilde{\varphi}_n} f(x_0^n + te^n + s_1e_1^n + \ldots + s_{d-1}e_{d-1}^n)\,\mathrm{d}s_1 \ldots \mathrm{d}s_{d-1}\,\mathrm{d}t \\
&\geq \int_{l_n(1-\varepsilon)^{\frac{1}{d}}}^{l_n} \int_{-t\tilde{\varphi}_n}^{t\tilde{\varphi}_n} \cdots \int_{-t\tilde{\varphi}_n}^{t\tilde{\varphi}_n} t^j \inf_{K_n} f^{(j)}\,\mathrm{d}s_1 \ldots \mathrm{d}s_{d-1}\,\mathrm{d}t = \frac{(2\tilde{\varphi}_n)^{d-1}}{d+j}l_n^{d+j}\left(1 - (1-\varepsilon)^{\frac{d+j}{d}}\right)\inf_{K_n} f^{(j)}.
\end{aligned}
$$

33

The assumptions of Theorem 6.3 *(i)* imply that $H_+^j(f, K_n) \geq D\frac{\Gamma(\delta_n)}{\sqrt{2n}}$ and therefore $p_{f, I_{r_n}} \geq \frac{D}{d+j}\sqrt{\frac{\delta_n}{2n}}\Gamma(\delta_n)(1 - (1 - \varepsilon)^{\frac{d+j}{d}})$. As the variables $Z_i = \mathbb{1}_{I_{r_n}}(X_i)$ are Bernoulli distributed with parameters $p_{f, I_{r_n}}$, we have

$$\mathbb{P}(\text{ no observation in } I_{r_n}) \leq \exp\Big(-\frac{D}{d+j}\sqrt{\frac{n\delta_n}{2}}\Gamma(\delta_n)\big(1 - (1 - \varepsilon)^{\frac{d+j}{d}}\big)\Big).$$

$\square$

As the number N_n of observations in K_n is $\text{Bin}(n, \delta_n)$-distributed, we obtain the following result from Chernoff's Inequality.

Lemma 6.9. $\mathbb{P}(N_n \leq (1 - \gamma)n\delta_n) \leq \exp\big(-n\delta_n\frac{\gamma^2}{2}\big)$ *for any* $\gamma \in (0, \frac{1}{2}]$.

Lemma 6.10. *Let* $j \in \{1, 2\}$.

(i) *If the assumptions of Theorem 6.3 (i) are satisfied, then*

$$(a) \quad n\delta_n \geq \frac{D^2}{(d+j)^2}\frac{\Gamma(\delta_n)^2}{2} =: \tilde{c}_+^j\frac{\Gamma(\delta_n)^2}{2},$$

$$(b) \quad n\delta_n \geq L_n\tilde{c}_+^j\log(\exp(1)n), \text{ where } L_n \geq 1 - o(1) \text{ for } n \to \infty.$$

(ii) *If the assumptions of Theorem 6.3 (ii) are satisfied, then*

$$(a') \quad n\delta_n \geq \frac{j^2 D^2}{(d(d+j))^2}\frac{\Gamma(\delta_n)^2}{2} =: \tilde{c}_-^j\frac{\Gamma(\delta_n)^2}{2}.$$

$$(b') \quad n\delta_n \geq L_n\tilde{c}_-^j\log(\exp(1)n), \text{ where } L_n \geq 1 - o(1) \text{ for } n \to \infty.$$

Proof: We only prove the first part. As in the proof of Lemma 6.8, we obtain

$$\delta_n = \int_{K_n} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \geq \frac{(2\tilde{\varphi}_n)^{d-1}}{d+j}l_n^{d+j}\inf_{K_n} f^{(j)}.$$

Hence,

$$H_+^j(f, K_n) = \frac{(2\tilde{\varphi}_n)^{d-1}l_n^{d+j}\inf_{K_n} f^{(j)}}{\sqrt{\delta_n}} \leq \frac{(2\tilde{\varphi}_n)^{d-1}l_n^{d+j}\inf_{K_n} f^{(j)}}{\frac{(2\tilde{\varphi}_n)^{d-1}}{d+j}l_n^{d+j}\inf_{K_n} f^{(j)}}\sqrt{\delta_n} = (d+j)\sqrt{\delta_n}.$$

Therefore, it follows from the assumption $H_+^j(f, K_n)$ that $n\delta_n \geq \frac{D^2}{(d+j)^2}\frac{\Gamma(\delta_n)^2}{2}$. Part *(b)* is a consequence of Lemma 7.5 in Dümbgen and Walther (2008). $\square$

The following Lemma provides the key to prove consistency. Note that for the construction of the test statistic T_{K_n}, it is necessary that at least two observations are contained in the wedge K_n. Given the event $\mathcal{N}_n$, we have that $N_n \geq 2$ if $n\delta_n \geq 4$. If the assumptions of Theorem 6.3 *(i)* hold, it follows from *(a)* that $n\delta_n \geq 4$ is fulfilled for $D \geq 2(d+j)$. Similarly, if the assumptions of Theorem 6.3 *(ii)* hold, then *(a')* yields the condition $D \geq \frac{2d(d+j)}{j}$.

Lemma 6.11. *Let $j \in \{1,2\}$ and $\mathbb{P}_{\mathcal{N}_n}$ denote the probability conditional on the event $\mathcal{N}_n$ defined in (5.14), and define for $\gamma \in (0, \frac{1}{2}]$, $0 < \varepsilon < 1$ and $\eta > 0$ the constant*

$$\mathbf{D}^j(\eta, \delta_n) := \frac{\frac{(2d+j)(d+j)}{j\sqrt{3}}}{(1-\varepsilon)^{\frac{d+j}{d}}\sqrt{(1-\gamma) - \frac{1}{n\delta_n}}} \left(\sqrt{2} + \sqrt{2}\frac{\kappa_n(\alpha) + \eta}{\Gamma(\delta_n)} - \frac{2\sqrt{2}\log\left((1-\gamma) - \frac{1}{n\delta_n}\right)}{\Gamma(\delta_n)^2} \right).$$

$$(5.15)$$

(i) If $H_+^j(f, K_n) \geq D\frac{\Gamma(\delta_n)}{\sqrt{2n}}$ for some constant $D \geq \mathbf{D}^j(n, \delta_n) \vee 2(d+j)$, then

$$\mathbb{P}_{\mathcal{N}_n}\left(T_{K_n} \leq c_{K_n}(\alpha)\big| N_n\right) \leq \exp\left(-\frac{\eta^2}{6}\right).$$

(ii) If $H_-^j(f, K_n) \leq -D\frac{\Gamma(\delta_n)}{\sqrt{2n}}$ for some constant $D \geq \mathbf{D}^j(n, \delta_n) \vee \frac{2d(d+j)}{j}$, then

$$\mathbb{P}_{\mathcal{N}_n}\left(T_{K_n} \geq -c_{K_n}(\alpha)\big| N_n\right) \leq \exp\left(-\frac{\eta^2}{6}\right).$$

Proof: We only prove the first part and define

$$c_{n,+}^j := \frac{\tilde{\varphi}_n^{d-1}(P_{e^n}X_{(N_n)})^{d+j}\inf_{K_n} f^{(j)}}{\delta_{N_n}}.$$

Then a tedious but straightforward calculation shows that the inequality

$$\frac{2^{d-1}jc_{n,+}^j(N_n - 1)}{(2d+j)(d+j)} - \eta\sqrt{\frac{N_n - 1}{3}} \geq c_{K_n}(\alpha)$$

holds for $\frac{|K_{N_n}|}{|K_n|} \geq 1 - \varepsilon$ and $N_n \geq (1-\gamma)n\delta_n$. This implies

$$\mathbb{P}_{\mathcal{N}_n}\left(T_{K_n} \leq c_{K_n}(\alpha)\big| N_n\right) = \mathbb{E}\left[\mathbb{P}_{\mathcal{N}_n}\left(T_{K_n} \leq c_{K_n}(\alpha)|N_n, X_{(N_n)}\right)\right]$$

$$\leq \mathbb{E}\left[\mathbb{P}_{\mathcal{N}_n}\left(T_{K_n} \leq \frac{2^{d-1}jc_{n,+}^j(N_n - 1)}{(2d+j)(d+j)} - \eta\sqrt{\frac{N_n - 1}{3}}\,\bigg|\, N_n, X_{(N_n)}\right)\right]$$

$$\leq \mathbb{E}\left[\mathbb{P}_{\mathcal{N}_n}\left(R_{K_n} \leq \frac{2^{d-1}jc_{n,+}^j(N_n - 1)}{(2d+j)(d+j)} - \eta\sqrt{\frac{N_n - 1}{3}}\,\bigg|\, N_n, X_{(N_n)}\right)\right],$$

where we used Lemma 6.6 and the notation $R_{K_n} = \sum_{i=1}^{N_n-1}\beta((\overline{F}_+^j)^{-1}(\tilde{F}_n(P_{e^n}X_{(i)})))$. Therefore, the assertion follows from Lemma 6.7 and Hoeffding's inequality. $\square$

Proof of Theorem 6.3: For a proof of the first part we proceed in two steps: firstly, we will find an upper bound for the probability that the test will not reject for one single wedge. Secondly, we will consider the probability for simultaneous rejection on every wedge in $\mathcal{K}_n$. For a fixed wedge $K_n \in \mathcal{K}_n$, $0 < \varepsilon < 1$ and $\gamma \in (0, \frac{1}{2}]$, we have

$$\mathbb{P}\left(T_{K_n} \leq c_{K_n}(\alpha) \text{ for any single } K_n \in \mathcal{K}_n\right) \leq \mathbb{P}_{\mathcal{N}_n}\left(T_{K_n} \leq c_{K_n}(\alpha) \text{ for any single } K_n \in \mathcal{K}_n\right) + \mathbb{P}(\mathcal{N}_n^c),$$

where the event $\mathcal{N}_n$ is defined in (5.14). Notice that the assumptions of Theorem 6.3 imply those of Lemma 6.11 and recall that $F(K_n) = \delta_n$. We have from the assumption in Theorem 6.3 *(i)* with $K_n = K_n^i$

$$H_+^j(f, K_n) \geq D\big(1 + b_n \tfrac{\sqrt{2}}{\Gamma(\delta_n)}\big) \tfrac{\Gamma(\delta_n)}{\sqrt{2n}},$$

which relaxes the assumption on the constant D in Lemma 6.11 as follows. Let $\eta_{n,K_n} > 0$ and

$$D\big(1 + b_n \tfrac{\sqrt{2}}{\Gamma(\delta_n)}\big) \geq \max\Big\{\mathbf{D}^j(\eta_{n,K_n}, \delta_n), 2(d+j)\Big\}, \tag{5.16}$$

where $\mathbf{D}^j$ is defined in (5.15). Therefore, it follows from Lemma 6.11, Lemma 6.8 and Lemma 6.9 that the probability under consideration can be bounded by

$$\exp\big(-\tfrac{\eta_{n,K_n}^2}{6}\big) + \exp\big(-\tfrac{D}{d+j}\sqrt{\tfrac{n\delta_n}{2}}\Gamma(\delta_n)(1 - (1-\varepsilon)^{\frac{d+j}{d}})\big) + \exp\big(-n\delta_n\tfrac{\gamma^2}{2}\big), \tag{5.17}$$

which concludes the proof for any single wedge.

We now consider the union of all wedges of $\mathcal{K}_n$ and define $\tilde{\delta}_n := \inf_{K_n \in \mathcal{K}_n} F(K_n)$. As the wedges in $\mathcal{K}_n$ are pairwise disjoint, it follows from Lemma 6.10 that

$$M_n = \big(\#\{K_n : K_n \in \mathcal{K}_n\}\big) \leq \frac{1}{\tilde{\delta}_n} = o(n). \tag{5.18}$$

Therefore, $\mathcal{K}_n$ consists of a finite number of wedges. From (5.17) and the monotonicity of the function $\delta \mapsto \delta \log\big(\tfrac{\exp(1)}{\delta}\big)$, we obtain the estimate

$$\mathbb{P}\big(T_{K_n} \leq c_{K_n}(\alpha) \text{ for at least one } K_n \in \mathcal{K}_n\big) \tag{5.19}$$

$$\leq \sum_{K_n \in \mathcal{K}_n} \exp\big(-\tfrac{\eta_{n,K_n}^2}{6}\big) + M_n\Big(\exp\big(-\tfrac{D}{d+j}\sqrt{\tfrac{n\tilde{\delta}_n}{2}}\Gamma(\tilde{\delta}_n)(1 - (1-\varepsilon)^{\frac{d+j}{d}})\big) + \exp\big(-\tfrac{n\tilde{\delta}_n\gamma^2}{2}\big)\Big),$$

if condition (5.16) is fulfilled for every $K_n \in \mathcal{K}_n$. We now show that the right-hand side of (5.19) vanishes as $n \to \infty$ by investigating the asymptotic behaviour of every summand. For the first summand, let $\eta_{n,K_n} := \big(6\log\big(\tfrac{1}{\delta_n}\big) + b_n\big)^{\frac{1}{2}}$, then

$$\sum_{K_n \in \mathcal{K}_n} \exp\big(-\tfrac{\eta_{n,K_n}^2}{6}\big) = \sum_{K_n \in \mathcal{K}_n} \delta_n \exp\big(-\tfrac{b_n}{6}\big) = o(1),$$

because $\sum_{K_n \in \mathcal{K}_n} \delta_n \leq 1$ and $b_n \to \infty$ as $n \to \infty$. Next, we consider the second summand in (5.19). An application of Lemma 6.10 *(b)* gives

$$\sqrt{\tfrac{n\tilde{\delta}_n}{2}}\Gamma(\tilde{\delta}_n) \geq \sqrt{\tilde{c}_+^j}\log(\exp(1)n)(1 - o(1)).$$

Hence, by (5.18), if

$$\tfrac{D}{d+j}\sqrt{\tfrac{D^2}{(d+j)^2}}\big(1 - (1-\varepsilon)^{\frac{d+j}{d}}\big) > 1, \tag{5.20}$$

36

$$M_n \exp\left(-\frac{D}{d+j}\sqrt{\tilde{c}_+^j}\log(\exp(1)n)\left(1-(1-\varepsilon)^{\frac{d+j}{d}}\right)(1-o(1))\right) = o(1).$$

Finally, by (5.18) and Lemma 6.10 *(b)*, we have

$$o(1)\,\exp\left(-n\tilde{\delta}_n\frac{\gamma^2}{2}+\log(n)\right) \leq o(1)\,\exp\left(-(1-o(1))\tilde{c}_+^j\log(\exp(1)n)\frac{\gamma^2}{2}+\log(n)\right)$$

$$\leq o(1)\,\exp\left(-\log(n)\left(\tilde{c}_+^j\frac{\gamma^2}{2}-(1+o(1))\right)\right) = o(1),$$

if

$$\frac{D^2}{(d+j)^2}\frac{\gamma^2}{2} > 1. \tag{5.21}$$

In this case, the third term vanishes as well as $n \to \infty$.

It remains to show that condition (5.16) is fulfilled for every $K_n \in \mathcal{K}_n$. With $\kappa_n(\alpha) \leq A$, we have to prove that

$$D\left(1+b_n\frac{\sqrt{2}}{\Gamma(\delta_n)}\right) \tag{5.22}$$

$$\geq \frac{(2d+j)(d+j)}{\sqrt{3}j(1-\varepsilon)^{\frac{d+j}{d}}\sqrt{(1-\gamma)-\frac{1}{n\delta_n}}}\left(\sqrt{2}+\sqrt{2}\frac{A+\eta_{n,K_n}}{\Gamma(\delta_n)}-\frac{2\sqrt{2}\log\left(1-\gamma-\frac{1}{n\delta_n}\right)}{\Gamma(\delta_n)^2}\right).$$

From Lemma 6.10 *(b)* it follows that $n\delta_n \to \infty$ for $n \to \infty$ for all $K_n \in \mathcal{K}_n$. Thus, using $\eta_{n,K_n} := \left(6\log\left(\frac{1}{\delta_n}\right)+b_n\right)^{\frac{1}{2}}$, we find that for sufficiently large n an upper bound for the right hand side of (5.22) is given by

$$\frac{(2d+j)(d+j)}{\sqrt{3}j(1-\varepsilon)^{\frac{d+j}{d}}\sqrt{(1-\gamma)-o(1)}}\left(\sqrt{2}+\frac{A+\left(\{6\log\left(\frac{1}{\delta_n}\right)\}^{1/2}+\sqrt{b_n}\right)}{\left\{\log\left(\frac{1}{\delta_n}\right)\right\}^{1/2}}\right)$$

$$\leq \frac{(\sqrt{2}+\sqrt{6})(2d+j)(d+j)}{\sqrt{3}j(1-\varepsilon)^{\frac{d+j}{d}}\sqrt{1-\gamma-o(1)}}+\frac{(2d+j)(d+j)(A+\sqrt{b_n})}{\sqrt{3}j(1-\varepsilon)^{\frac{d+j}{d}}\left\{\log\left(\frac{1}{\delta_n}\right)\right\}^{1/2}\sqrt{1-\gamma-o(1)}} \tag{5.23}$$

$$\leq \left(1+\frac{o(b_n)}{\{\log\left(\frac{1}{\delta_n}\right)\}^{1/2}}\right)\frac{2\sqrt{2}(2d+j)(d+j)}{j(1-\varepsilon)^{\frac{d+j}{d}}\sqrt{1-\gamma-o(1)}}.$$

Combining (5.20), (5.21) and (5.23), we obtain the following condition

$$\frac{D}{d+j} > \max\left\{\left(1-(1-\varepsilon)^{\frac{d+j}{d}}\right)^{-1/2},\ \frac{\sqrt{2}}{\gamma},\ \frac{2\sqrt{2}(2d+j)}{j(1-\varepsilon)^{\frac{d+j}{d}}\sqrt{1-\gamma}}\right\}. \tag{5.24}$$

In order to minimize the restrictions imposed by condition (5.24), we now determine $0 < \varepsilon < 1$ and $\gamma \in (0,\frac{1}{2}]$, such that the lower bound on D is as small as possible. Balancing the second and third terms in (5.24) we obtain

$$\gamma = \frac{-j^2+\sqrt{j^4+4j^2(2d+j)^2}}{2(2d+j)^2} < \frac{j}{2d+j} \leq \frac{1}{2},$$

37

where we used $(2d+j)^2 \geq 9$ (note that $d \geq 1$) for the first inequality. For the choice of ε we introduce the notation $a := (1-\varepsilon)^{\frac{d+j}{d}}$ and balance the first and third expression in (5.24) and obtain

$$a = \frac{(2d+j)^2}{j^2}\left(-1 + \sqrt{1 + \frac{2j^2}{(2d+j)^2}}\right).$$

Finally, inserting our choice of ε and γ in (5.24), we find the condition (5.7) since our calculations also show that D is larger than all three terms of (5.24) simultaneously in this case. $\qquad\square$

6.3 Proof of Theorem 2.6 and 3.3

For the sake of simplicity, we prove both results for the case $C_2 = 1$ and $C := C_1$. The general case follows by exactly the same arguments with an additional amount of notation.

Proof of Theorem 2.6: We note that it follows from Theorem 6.2 that $c_{K^i}(\alpha) \geq \tilde{c}_{K^i}(\alpha)$ for $i = 1, \ldots, M_n$. Hence, it remains to show that the assumptions for Theorem 6.3 *(ii)* are satisfied. By assumption on f, we have $\sup_{K^i} f^{(2)} \leq -\frac{c}{2} + o(1)$ for $n \to \infty$ $(i = 1, \ldots, M_n)$. Moreover, from the approximation $\tan(x) = x(1 - o(1))$ for $x \to 0$, we have

$$F(K^i) = f(x_0)\frac{1}{d}C^d \log(n)^{-d+1+d\frac{d-1}{d+4}}\left(\frac{\log(n)}{n}\right)^{\frac{d}{d+4}}(1+o(1)) \quad (i = 1, \ldots, M_n). \tag{5.25}$$

Hence,

$$-H_-^2(f, K^i) \geq \frac{\log(n)^{-d+1+(d+2)\frac{d-1}{d+4}}C^{d+2}\left(\frac{\log(n)}{n}\right)^{\frac{d+2}{d+4}}(c/2 - o(1))}{\sqrt{f(x_0)\frac{1}{d}C^{\frac{d}{2}}\log(n)^{\frac{-d+1}{2}+d\frac{d-1}{2(d+4)}}\left(\frac{\log(n)}{n}\right)^{\frac{d}{2(d+4)}}(1+o(1))}} = C^{\frac{d}{2}+2}\left(\frac{\log(n)}{n}\right)^{\frac{1}{2}}\frac{c/2 - o(1)}{\sqrt{f(x_0)\frac{1}{d}}}$$

$(i = 1, \ldots, M_n)$. Furthermore, we obtain from (5.25) and the assumption $b_n = o(\sqrt{\log(n)})$ that

$$D\left(\frac{\Gamma(F(K^i))}{\sqrt{2n}} + \frac{b_n}{\sqrt{n}}\right) = D\sqrt{\frac{d}{d+4}\frac{\log n}{n}}(1+o(1)).$$

Therefore, the assumptions of Theorem 6.3 *(ii)* (for $j = 2$) are fulfilled as the constant C satisfies $C > \left(\frac{2D}{c}\sqrt{\frac{f(x_0)}{d+4}}\right)^{\frac{2}{d+4}}$ by (2.12) and hence the assertion follows by an application of Theorem 6.3. $\qquad\square$

Proof of Theorem 3.3: Note that it is sufficient to prove consistency for the largest scale. By Theorem 6.2, we have to prove that the assumptions for Theorem 6.3 *(ii)* are satisfied for the family of wedges $\{K_n^i \mid i \in \mathcal{I}_n\}$ introduced in Section 3.3. Let $K_n \in \{K_n^i \mid i \in \mathcal{I}_n\}$. We begin with the determination of an upper bound for the quantity $\sup_{K_n} f^{(1)}$ defined in (5.5). For this purpose, consider a point $x_0^n + \tilde{t}e_0 \in K_n$ with $e_0 \in \mathbb{R}^d$ ($\|e_0\| = 1$) and $\tilde{t} > 0$. Now, the representation (3.6) and an application of the mean value theorem yields for $0 \leq \tilde{s} < \tilde{t}$

$$\begin{aligned} \Delta_n &= \frac{f(x_0^n + \tilde{t}e_0) - f(x_0^n + \tilde{s}e_0)}{\tilde{t} - \tilde{s}} \\ &= \tilde{f}_{x_0}(\|x_0^n + \tilde{s}e_0 - x_0\|)\langle \nabla g_{x_0}(\xi_2), e_0\rangle + (1 + g_{x_0}(x_0^n + \tilde{t}e_0))\tilde{f}_{x_0}'(\xi_1)R_n \end{aligned} \tag{5.26}$$

for some $\|x_0^n + \tilde{s}e_0 - x_0\| \le \xi_1 \le \|x_0^n + \tilde{t}e_0 - x_0\|$ and $\xi_2 \in [x_0^n + \tilde{s}e_0, x_0^n + \tilde{t}e_0]$, where

$$R_n = \frac{\|x_0^n + \tilde{t}e_0 - x_0\| - \|x_0^n + \tilde{s}e_0 - x_0\|}{\tilde{t} - \tilde{s}}.$$

A further application of the mean value theorem gives

$$
\begin{aligned}
R_n &= \frac{\sum_{j=1}^d (x_{0,j}^n + \tilde{t}e_{0,j} - x_{0,j})^2 - \sum_{j=1}^d (x_{0,j}^n + \tilde{s}e_{0,j} - x_{0,j})^2}{2\sqrt{\xi}(\tilde{t} - \tilde{s})} \\
&= \frac{\sum_{j=1}^d 2(x_{0,j}^n - x_{0,j})e_{0,j}(\tilde{t} - \tilde{s}) + \tilde{t}^2 - \tilde{s}^2}{2\sqrt{\xi}(\tilde{t} - \tilde{s})} \ge \frac{\sum_{j=1}^d (x_{0,j}^n - x_{0,j})e_{0,j}}{\sqrt{\xi}} \\
&= \frac{\cos(\text{angle}(x_0^n - x_0, e_0))\|x_0^n - x_0\|}{\sqrt{\xi}} \ge C \log(n) l_n (1 - o(1)) \frac{1}{\sqrt{\xi}}
\end{aligned}
$$

for some $\|x_0^n + \tilde{s}e_0 - x_0\|^2 \le \xi \le \|x_0^n + \tilde{t}e_0 - x_0\|^2$. Moreover, we have

$$\xi \le \|x_0^n + \tilde{t}e_0 - x_0\|^2 \le \big(\|x_0^n - x_0\| + \tilde{t}\big)^2 \le l_n^2 (m_n C \log(n))^2 (1 + o(1)).$$

Hence, $R_n \ge \frac{1-o(1)}{m_n}$. With the same arguments as used before, one shows that $\xi_1 \ge C \log(n) l_n$ and $\|\xi_2 - x_0\| \le C m_n \log(n) l_n (1 + o(1))$. Finally, by assumption on g_{x_0} and $\tilde{f}_{x_0}$ and (5.26), this yields

$$
\begin{aligned}
\Delta_n &\le (1 + o(1))(-c) C \log(n) l_n R_n + f(x_0) o((m_n \log(n) l_n)^{1+\gamma}) \\
&\le (1 - o(1))(-c) C \log(n) \frac{l_n}{m_n} + o((m_n \log(n) l_n)^{1+\gamma}) = -\frac{cC}{m_n} \log(n) l_n (1 - o(1)),
\end{aligned}
$$

as $\frac{m_n}{l_n} o((m_n l_n)^{1+\gamma}) = o(m_n^{2+\gamma} l_n^\gamma) = o(1)$ as $n \to \infty$ by the choice of m_n and l_n. Consequently, $\sup_{K_n} f' \le -\frac{cC}{m_n} \log(n) l_n (1 - o(1))$. As $\tan(x) = x(1 - o(1))$ $(x \to 0)$, we have

$$
\begin{aligned}
F(K_n) &= f(x_0^n) \frac{1}{d} C^d (\log(n))^{-d+1+d} \Big(\frac{\log(n)}{n}\Big)^{\frac{d}{d+4}} (1 + o(1)) \\
&= f(x_0) \frac{1}{d} C^d \log(n) \Big(\frac{\log(n)}{n}\Big)^{\frac{d}{d+4}} (1 + o(1)).
\end{aligned}
$$

It follows from the conditions $c_1 \ge f(x_0) \ge 0$

$$
\begin{aligned}
H_-^1(f, K_n) &\le -\frac{C^{d+1}(\log(n))^3 \big(\frac{\log(n)}{n}\big)^{\frac{d+2}{d+4}} \frac{cC}{m_n}(1 - o(1))}{\sqrt{f(x_0) \frac{1}{d} C^{\frac{d}{2}} (\log(n))^{\frac{1}{2}} \big(\frac{\log(n)}{n}\big)^{\frac{d}{2(d+4)}} (1 + o(1))}} \\
&\le -cC^{\frac{d}{2}+2} \Big(d \frac{\log(n)}{n}\Big)^{\frac{1}{2}} \frac{1 - o(1)}{\sqrt{c_1}} \frac{(\log(n))^{\frac{5}{2}}}{m_n} = -cC^{\frac{d}{2}+2} \Big(d \frac{\log(n)}{n}\Big)^{\frac{1}{2}} \frac{1 - o(1)}{\sqrt{c_1}}.
\end{aligned}
$$

If $b_n = o(\sqrt{\log n})$, we have

$$D\Big(\frac{\Gamma(F(K_n))}{\sqrt{2n}} + \frac{b_n}{\sqrt{n}}\Big) = D\sqrt{\frac{d}{d+4} \frac{\log n}{n}}(1 + o(1)),$$

and the assumptions of Theorem 6.3 *(ii)* are fulfilled if the constant C satisfies $C > \Big(\frac{D}{c}\sqrt{\frac{c_1}{d+4}}\Big)^{\frac{2}{d+4}}$, which is a direct consequence of (3.7). $\qquad\square$