UNIVERSITÀ DI TRENTO
AND
UNIVERSIDAD DE CANTABRIA

DOCTORAL THESIS

Local depth functions and applications to clustering

Author:
Giacomo FRANCISCI

Supervisors:
Full Prof. Claudio AGOSTINELLI
Assoc. Prof. Alicia NIETO-REYES
Full Prof. Anand N. VIDYASHANKAR

A thesis submitted in fulfillment of the requirements for the degree of

Dottore di ricerca in Matematica, Doctoral Program in Mathematics

and

Doctor por la Universidad de Cantabria, Programa de Doctorado en Ciencia y Tecnología.

January 26, 2022

Abstract

Local depth functions (LDFs) are used for describing the local geometric features and mode(s) in multidimensional distributions. In this thesis, we undertake a rigorous systematic study of LDFs and establish several analytical and statistical properties. First, we show that, when the underlying probability distribution is absolutely continuous, scaled versions of LDFs (referred to as τ-approximation) converge, uniformly and in L^q, to the density, when τ converges to zero. Second, we establish that, as the sample size diverges to infinity the centered and scaled sample LDFs converge in distribution to a centered Gaussian process uniformly in the space of bounded functions on $\mathcal{H}_G$, a class of functions yielding LDFs. Third, using the sample version of the τ-approximation and the gradient system analysis, we develop a new clustering algorithm. The validity of this algorithm requires several results concerning the uniform finite difference approximation of the gradient system associated with the sample τ-approximation. For this reason, we establish a Bernstein-type inequality for deviations between the centered and scaled sample LDFs. Finally, invoking the above results, we establish consistency of the clustering algorithm. Applications of the proposed methods to mode estimation and upper level set estimation are also provided.

Acknowledgements

I thank Alicia and Claudio for proposing many challenging projects and collaborations with other researchers as well as for a great amount of help in developing them. Next, I thank Anand who always tried to push me further and suggested several ideas that have been developed in this work. He also took the time to explain to me how to properly write in English and present the outcome of our research work. Finally, I thank Alicia for taking care of much of the paperwork at the University of Cantabria and helping me translate the summary of the thesis from English into Spanish.

Summary

This thesis is divided into three main chapters. Chapter 1 contains a review of depth functions and multidimensional quantiles. Results from the literature are complemented with new mathematical insights into the general theory of statistical depth functions (Liu, 1990; Zuo and Serfling, 2000a). Depth functions are divided in three main types: (i) simplicial depth, spherical depth, lens depth, and β-skeleton depth are *Type A* depth functions, (ii) L^q-depth and simplicial volume depth are *Type B* depth functions, and (iii) Mahalanobis depth and projection depth are *Type C* depth functions. We investigate the properties of these depth functions and the resulting quantiles. We also study convergence of sample depth and quantiles. Since sample *Type A* and *Type B* depths take the form of a U-statistics, we dedicate the last part of Chapter 1 to review the theory of U-statistics and the corresponding empirical processes, namely, U-processes. Although depth functions have been used in various applications, we focus here on the properties of the depth functions themselves and explain which depth functions yield meaningful multidimensional quantiles.

Chapters 2 and 3 are taken from the paper Francisci et al. (2020). While depth functions are used to describe the global features of multidimensional distributions, local depth functions (LDFs) can also detect local features such as modes and regions with low probability mass (Agostinelli and Romanazzi, 2011). In Chapter 2, we introduce a general class of *Type A* local depth functions and study its properties. Specifically, we show that, as the localizing parameter τ diverges to infinity, LDFs converge to *Type A* depth functions, whereas, as $\tau \to 0^+$ and under appropriate scaling, LDFs converge to a power of the underlying density f. Thus, the root of scaled LDFs (referred to as τ-approximation and denoted by f_τ) becomes arbitrarily close to the density f. Indeed, we show that, under appropriate differentiability assumptions, f_τ and its derivatives converge uniformly to f and the corresponding derivatives. Next, we replace depth by sample depth and obtain an estimator $f_{\tau,n}$ for the density f. Using the theory of U-processes mentioned above, we develop a Bernstein-type inequality for LDFs and show that, for a suitable sequence $\{\tau_n\}_{n=1}^{\infty}$, $f_{\tau_n,n}$ converges almost surely to f, uniformly over compact sets. This opens the door to a series of applications including clustering, mode estimation, and upper level set estimation, which are investigated in Chapter 3. Specifically, we propose a new clustering algorithm via sample τ-approximation. At population level, clusters are defined as the stable manifolds induced by a mode of f via a gradient system (Chacón, 2015). First, we verify that the resulting clusters are well-defined and non-trivial. Next, we show that the clusters of f_τ converge to those of f. Towards this end, we show that the stationary points and modes of f_τ converge to those of f. Turning to sample τ-approximation and empirical clusters, we use the aforementioned Bernstein-type inequality to establish uniform convergence of finite difference approximations of the derivatives of $f_{\tau,n}$ to the appropriate derivatives of f. We use this and a density of data points argument to obtain convergence of empirical clusters. As a by-product, we obtain convergence of the last iterate in the clustering algorithm to the mode. Finally, we illustrate the finite sample behavior of the proposed methods via numerical experiments and data analyses.

Resumen

Esta tesis esta dividida en tres capítulos principales. El capítulo 1 contiene una revisión de las funciones de profundidad y de los cuantiles multidimensionales. Los resultados existentes en la literatura se complementan con nuevos conocimientos matemáticos sobre la teoría general de las funciones estadísticas de profundidad (Liu, 1990; Zuo and Serfling, 2000a). Dividimos las funciones de profundidad en tres grupos: (i) la profundidad simplicial, profundidad esférica, profundidad de lente y profundidad β-esqueleto son funciones de profundidad *Tipo A*, (ii) la profundidad L^q y la profundidad de volumen simplicial son funciones de profundidad *Tipo B* y (iii) la profundidad de Mahalanobis y la profundidad de proyección son funciones de profundidad *Tipo C*. En el capítulo, investigamos las propiedades de estas funciones de profundidad y de los cuantiles resultantes. También estudiamos la convergencia de la profundidad muestral y de los cuantiles muestrales. Dado que las profundidades muestrales del *Tipo A* y *Tipo B* son U-estadísticos, dedicamos la última parte del Capítulo 1 a revisar la teoría de U-estadísticos y de los procesos empíricos correspondientes, conocidos como U-procesos. Aunque las funciones de profundidad se han utilizado en diferentes aplicaciones, aquí nos centramos en las propiedades de las funciones de profundidad y explicamos por qué las funciones de profundidad resultan en cuantiles multidimensionales con sentido.

Los capítulos 2 y 3 se han extraído del manuscrito Francisci et al. (2020). Mientras que las funciones de profundidad se utilizan para describir las características globales de distribuciones multidimensionales, las funciones de profundidad local (FPL) también detectan características locales como las modas y las zonas de baja probabilidad (Agostinelli and Romanazzi, 2011). En el capítulo 2, presentamos una clase general de funciones de profundidad local del *Tipo A* y estudiamos sus propiedades. En particular, mostramos que, a medida que el parámetro de localización τ diverge hacia el infinito, las FPL convergen a las funciones de profundidad *Tipo A*, mientras que, cuando $\tau \to 0^+$, bajo la escala adecuada, las FPL convergen a una potencia de la función de densidad densidad correspondiente, f. Por lo tanto, la raíz de los FPL escalados (a la que llamamos τ-aproximación y denotamos f_τ) se hace arbitrariamente cercana a la densidad f. De hecho, demostramos que, bajo ciertas hipótesis de diferenciabilidad, f_τ y sus derivadas convergen uniformemente en f y sus derivadas, respectivamente. En el siguiente paso, reemplazamos la profundidad por la correspondiente profundidad muestral y obtenemos un estimador $f_{\tau,n}$ de la función de densidad f. Usando la teoría de los U-procesos mencionada anteriormente, desarrollamos una desigualdad de tipo Bernstein para FPL y demostramos que, para una secuencia $\{\tau_n\}_{n=1}^{\infty}$ apropiada, $f_{\tau_n,n}$ converge casi seguro a f, uniformemente sobre conjuntos compactos. Esto abre la puerta a una serie de aplicaciones que incluyen la clasificación no supervisada, estimación de modas y de conjuntos de nivel, que se investigan en el capítulo 3. En particular, proponemos un nuevo algoritmo de clasificación no supervisada haciendo uso de la τ-aproximación muestral. A nivel poblacional, los grupos, o clusters, se definen como las variedades estables inducidas por una moda de f a través de un sistema de gradientes (Chacón, 2015). Primero, verificamos que los grupos que resultan están bien definidos y no son triviales. A

continuación, mostramos que los grupos dados por f_τ convergen a los dados por f. Con este fin, demostramos que los puntos de estacionariedad y las modas de f_τ convergen a las de f. Para la τ-aproximación muestral y los clusters empíricos, usamos la desigualdad de tipo Bernstein antes mencionada y así obtener convergencia uniforme de aproximaciones finito-diferenciales de las derivadas de $f_{\tau,n}$ a las derivadas correspondientes de f. Usamos este argumento y el de la densidad de los puntos para obtener la convergencia de clusters empíricos. Como consecuencia, obtenemos la convergencia a la moda en la la última iteración del algoritmo de clasificación no supervisada. Finalmente, ilustramos el comportamiento muestral, finito, de las metodologías propuestas mediante experimentos numéricos y análisis de datos.

Contents

Chapter 1

Depth functions

1.1 Introduction

The lack of a natural order on $\mathbb{R}^d$ has pushed researchers to look for different approaches for defining multidimensional median and quantiles. One of the most successful approaches uses depth functions to assign a value to every point in $\mathbb{R}^d$ based on a probability distribution P on $\mathbb{R}^d$. Then, points in $\mathbb{R}^d$ are ordered based on these values. Clearly, not every function provides a reasonable ordering. Liu (1990) and Zuo and Serfling (2000a) state a series of properties that depth functions should satisfy. These properties are formalized in the concept of *statistical* depth function. Specifically, the depth median for P is defined as the point in $\mathbb{R}^d$ with highest depth value and depth values are supposed to be non-increasing moving away from the median. If the probability measure P is symmetric, then the depth median must coincide with the point of symmetry. It is conventionally assumed that depth functions are non-negative and decrease to zero as the distance from the median tends to infinity. Finally, it is assumed that depth functions are invariant with respect to (w.r.t.) affine transformations applied to both points in $\mathbb{R}^d$ and the probability distribution P.

In this chapter, we review a series of results concerning multivariate quantiles and depth functions and evaluate them based on the properties of the resulting quantiles. A comparison between depth functions and other approaches to define multivariate quantiles is given in Serfling (2002). As we will see, not all depth functions in the literature satisfy the above axioms. However, it is worth mentioning that depth functions have been used in many other contexts and their performance depends on the specific task at hand. Among the many applications of depth functions we shall mention location and scale estimation, classification and clustering, test for symmetry, and outliers detection. Accordingly, Mosler and Mozharovskyi (2020) evaluates the performance of depth functions based on invariance and uniqueness of the resulting quantiles as well as robustness and computational feasibility.

In Section 1.3 we introduce two of the most popular depth functions, namely halfspace and simplicial depth, and explain the link between these depth functions and quantiles in $\mathbb{R}$. Before studying in detail the properties of these depth functions we briefly discuss other approaches to define quantiles on $\mathbb{R}^d$. One of the simplest approaches is given by componentwise quantiles, but has the drawback of not being equivariant w.r.t. orthogonal transformations. More recent and advanced techniques involve tools from set optimization and transportation theory. Specifically, Hamel and Kostner (2018) define multidimensional quantiles using the partial order on $\mathbb{R}^d$ induced by convex cones, whereas Chernozhukov et al. (2017) and Hallin et al. (2021) obtain multidimensional quantiles using a transportation map to a reference distribution.

In Section 1.5 we formally define the notion of *statistical* depth function. The exposition is slightly different from Zuo and Serfling (2000a) in that depth functions are additionally assumed to be upper semicontinuous. This assumption ensures that a depth median always exists. Apart from this, a weaker and more general definition of *statistical* depth function is used allowing for invariance under a general class of transformations and monotonicity w.r.t. the deepest point is allowed to hold only for a smaller class of probability measures.

We study symmetric distributions in Section 1.4 and provide several examples of depth functions in Sections 1.6-1.8, where we also study their properties. Following Zuo and Serfling (2000a) depth functions are divided in three main types. *Type A* depth functions include simplicial depth and β-skeleton depths, whereas L^q-depth and simplicial volume depth are examples of *Type B* depth functions. Finally, Mahalanobis depth and projection depth are *Type C* depth functions. Specifically, we see that halfspace depth, Mahalanobis depth, projection depth, and L^q-depth are *statistical* depth functions. However, we notice that other depth functions such as simplicial depth, simplicial volume depth, and β-skeleton depths do not satisfy one or more of the above requirements. For the simplicial depth, this was shown already by Zuo and Serfling (2000a). Results for simplicial volume and β-skeleton depths are partially new. In particular, we unexpectedly conclude that the simplicial volume depth does not decrease to zero for certain distributions, whereas β-skeleton depths fail to be non-increasing even for centrally symmetric probability measures.

In Section 1.9 we study consistency properties of depth quantiles. To this end, we use suitable estimators for the above depth functions and show that, if they converge uniformly to the corresponding depth functions, then also the depth quantiles converge. (Uniform) convergence of these estimators is studied in Sections 1.10-1.11 and Chapter 2. Specifically, in Section 1.10, we derive several properties and results for U-statistics, which can be used to obtain consistency and asymptotic normality for *Type A* and *Type B* depth functions. In Section 1.11 we extensively analyze empirical processes and U-processes, that is, empirical processes that take the form of U-statistics. There we summarize without proofs some of the most important results concerning uniform law of large numbers and uniform central limit theorems for U-statistics. To this end, we need to verify several measurability conditions. We differ this analysis to Appendix A. These results are then used in Chapter 2 to prove the uniform law of large numbers and uniform central limit theorem for several *Type A* depth functions.

Section 1.2 below may be skipped at first reading as it contains some mathematical background on probability spaces, independence, convergence of random variables etc. that is used in Sections 1.9-1.11 but, as noticed below, it is not strictly necessary for Sections 1.3-1.8.

1.2 Mathematical background

In this section, we show that many properties of a random variable on an abstract probability space can be checked directly on the space where it assumes its values, thus simplifying the analysis. A probability space is a triple $(\Omega, \Sigma, \mathbb{P})$, where Ω is a non-empty set, Σ is a σ-algebra on Ω,[1] and $\mathbb{P}$ is a probability measure on (Ω, Σ).

[1] Σ is a σ-algebra on Ω if $\Sigma \subset \wp(\Omega)$, where $\wp(\Omega)$ is the power set of Ω, and satisfies the following properties: (i) $\Omega \in \Sigma$, (ii) if $A \in \Sigma$ then $\Omega \setminus A \in \Sigma$, and (iii) if $A_1, A_2, \dots \in \Sigma$ then $\cup_{i=1}^{\infty} A_i \in \Sigma$. Using (i) and (ii) we also have that $\emptyset = \Omega \setminus \Omega \in \Sigma$ and using (ii) and (iii) we obtain that $\cap_{i=1}^{\infty} A_i = \Omega \setminus (\cup_{i=1}^{\infty} (\Omega \setminus A_i)) \in \Sigma$.

Let E be a topological space and $\mathcal{B}(E)$ be the Borel σ-algebra on E.[2] We denote by $\mathring{A}$, $\overline{A}$, and ∂A be the interior, closure, and boundary of a set $A \subset E$. An E-valued random variable is a measurable function $X : (\Omega, \Sigma) \to (E, \mathcal{B}(E))$.[3] X induces a probability measure P on E, where $P = \mathbb{P}_X$ is the push-forward measure given by $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$, for all $B \in \mathcal{B}(E)$). As it is defined on $(E, \mathcal{B}(E))$, P is a Borel probability measure and it is referred to as the probability distribution (or just distribution) of X. Most of the discussion of Sections 1.3-1.8 about multivariate quantiles and depth functions is phrased in terms of Borel probability measures. The underlying concept is that to each random variable $X : \Omega \to E$ corresponds a Borel probability measure P (and vice versa). We denote by $\mathcal{M}(E)$ the set of all Borel measures on E and by $\mathcal{P}(E)$ the set of all probability measures. When $E = \mathbb{R}^d$, we will also write $\mathcal{P}_d$ for $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{B}_d$ for $\mathcal{B}(\mathbb{R}^d)$. Next, let $H : E \to \mathbb{R}$ be Borel measurable. The expectation of $H(X)$, if it exists,[4] is given by

$$\mathbb{E}[H(X)] = \int_\Omega H(X(\omega)) d\mathbb{P}(\omega).$$

By a change of variable, the above expectation can be directly computed as an integral over E. Indeed, by Theorem 16.13 of Billingsley (2012), it holds that

$$\int_\Omega H(X(\omega)) d\mathbb{P}(\omega) = \int_E H(x) dP(x),$$

where $P = \mathbb{P}_X$. This allows to compute mean and variance of a $\mathbb{R}^d$-valued random variable directly on $\mathbb{R}^d$ (cf. Definition 1.4.3). For instance, if $X : \Omega \to \mathbb{R}$ and $H : \mathbb{R} \to \mathbb{R}$, then

$$\mathbb{V}ar[H(X)] = \mathbb{E}[(H(X) - \mathbb{E}[H(X)])^2] = \int \left(H(x) - \int H(y) dP(y) \right)^2 dP(x).$$

Next, we define convergence of random variables on a metric space (F, d_F). We begin by defining weak convergence of a sequence of probability measures $\{P_n\}_{n=1}^\infty$ in $\mathcal{P}(E)$. We say that P_n converges weakly to $P \in \mathcal{P}(E)$ if $\lim_{n\to\infty} \int H(x) P_n(x) = \int H(x) dP(x)$ for all bounded and continuous functions $H : E \to \mathbb{R}$. In this case, we write $P_n \xrightarrow{w} P$. Now, let $(\Omega_n, \Sigma_n, \mathbb{P}_n)$ and $(\Omega, \Sigma, \mathbb{P})$ be probability spaces and $\{X_n\}_{n=1}^\infty$ be a sequence of random variables $X_n : \Omega_n \to F$. We say that X_n converges in distribution to a random variable $X : \Omega \to F$ if $\mathbb{P}_{X_n} \xrightarrow{w} \mathbb{P}_X$. In this case we write $X_n \xrightarrow{d} X$. Next, suppose that $(\Omega_n, \Sigma_n, \mathbb{P}_n) = (\Omega, \Sigma, \mathbb{P})$. We write $X_n \xrightarrow{p} X$ for X_n converges in probability to X, that is, $\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : d_F(X_n(\omega), X(\omega)) \geq \epsilon\}) = 0$, for all $\epsilon > 0$. Finally, $X_n \xrightarrow{a.s.} X$ means that X_n converges almost surely to X, namely, $\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} d_F(X_n(\omega), X(\omega)) = 0\}) = 1$. Almost sure convergence implies convergence in probability, which in turn implies convergence in distribution.[5] Finally, if $c \in F$ then $X_n \xrightarrow{d} c$ if and only if $X_n \xrightarrow{p} c$.[6]

[2] $\mathcal{B}(E)$ is the smallest σ-algebra containing all open sets.

[3] This means that $X : \Omega \to E$ and $X^{-1}(B) \in \Sigma$ for all $B \in \mathcal{B}(E)$.

[4] We refer to Sections 15 and 16 of Billingsley (2012) for a precise definition of the integral w.r.t. a general measure and its properties. Existence of the integral reduces to the condition that at least one of $\int_\Omega H_+(X(\omega)) d\mathbb{P}(\omega)$ and $\int_\Omega H_-(X(\omega)) d\mathbb{P}(\omega)$ is finite, where $H_+ = \max(H, 0)$ and $H_- = \max(-H, 0)$ are the positive and negative part of H, respectively.

[5] See for instance Lemma 1.9.2 (i) and Lemma 1.10.2 (ii) of van der Vaart and Wellner (1996).

[6] See Lemma 1.10.3 (iii) of van der Vaart and Wellner (1996).

We introduce next the product of sets, σ-algebras, and measures. Specifically, for $n \in \mathbb{N} \cup \{\infty\}$,[7] we denote by $\prod_{i=1}^n \Omega_i$ the n-fold Cartesian product of sets $\Omega_1, \dots, \Omega_n$. Next, if Σ_i are σ-algebras on Ω_i, then the product σ-algebra $\otimes_{i=1}^n \Sigma_i$ is the smallest σ-algebra including $\{\prod_{j=1}^n B_j : B_j \in \Sigma_j\}$.[8] Then $(\prod_{i=1}^n \Omega_i, \otimes_{i=1}^n \Sigma_i)$ is a measure space and, for all finite n, the n-fold product $\prod_{i=1}^n \mathrm{m}_i$ of measures m_i on (Ω_i, Σ_i) is a well-defined measure on this space (see e.g. Section 18 of Billingsley (2012)). Also, the infinite product $\prod_{i=1}^\infty \mathrm{m}_i$ is well-defined if $\mathrm{m}_i = \mathbb{P}_i$ are probability measures (see Section 8.2 of Dudley (2018)). When $\Omega_i = \Omega$ (resp. $\Sigma_i = \Sigma$ or $\mathrm{m}_i = \mathrm{m}$) does not depended on i, we also write Ω^n for $\prod_{i=1}^n \Omega_i$ (resp. $\Sigma^{\otimes n}$ for $\otimes_{i=1}^n \Sigma_i$ or m^n for $\prod_{i=1}^n \mathrm{m}_i$). In particular, if $(\Omega_i, \Sigma_i, \mathbb{P}_i)$ are probability spaces, then $(\prod_{i=1}^n \Omega_i, \otimes_{i=1}^n \Sigma_i, \prod_{i=1}^n \mathbb{P}_i)$ is also a probability space (cf. Definition 1.11.2 in Section 1.11).

We now turn to the definition of independent and identically distributed (i.i.d.) random variables. To this end, let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Random variables $X_i : (\Omega, \Sigma_i) \to (E, \mathcal{B}(E))$, where $\Sigma_i \subset \Sigma$, are independent if, for all $n \in \mathbb{N}$ and $A_i \in \mathcal{B}(E)$,

$$\mathbb{P}(\cap_{i=1}^n X_i^{-1}(A_i)) = \prod_{i=1}^n \mathbb{P}(X_i^{-1}(A_i)).$$

Notice that independence means that, for all $n \in \mathbb{N}$, the probability distribution of $(X_1, \dots, X_n) : \Omega \to E^n$ is the product measure $\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}$ on $(E^n, (\mathcal{B}(E))^{\otimes n})$. On the other hand, X_i are identically distributed if $\mathbb{P}_{X_i} = \mathbb{P}_{X_1}$ for all i. In particular, if X_i are both independent and identically distributed, then $(X_1, \dots, X_n)$ has probability distribution P^n, where $P = \mathbb{P}_{X_1}$. Much of the asymptotic properties of depth functions, U-statistics and empirical processes in Sections 1.9, 1.10 and 1.11 are based on a sequence $\{X_i\}_{i=1}^\infty$ of independent and identically distributed (i.i.d.) random variables $X_i : \Omega \to \mathbb{R}^d$. In particular, for the (uniform) asymptotic results in Section 1.11, the probability space $(\Omega, \Sigma, \mathbb{P})$ is conveniently expressed as a product $(E^\infty, (\mathcal{B}(E))^{\otimes\infty}, P^\infty)$, where $P \in \mathcal{P}(E)$, and the random variables $X_i : \Omega \to E$ are defined as the projections into the i component. Also, notice the implicit independence assumption in the definition of *Type A* and *Type B* depth functions (see Definitions 1.6.1 and Definition 1.7.1) such as simplicial depth (see Definition 1.3.6), simplicial volume depth (see Example 1.7.2), and β-skeleton depth (see Example 1.6.1). Indeed, *Type A* and *Type B* depth functions are the expectation of a U-statistics (see Section 1.10).

1.3 Multidimensional median and quantiles

We begin with the definition of quantiles for unidimensional probability measures. To this end, for $P \in \mathcal{P}_1$, we define the cumulative distribution function $F_P : \mathbb{R} \to [0,1]$ by $F_P(x) = P((-\infty, x])$. Clearly, F_P is right-continuous and non-decreasing with $\lim_{x\to-\infty} F_P(x) = 0$ and $\lim_{x\to\infty} F_P(x) = 1$. If P is continuous, i.e. $P(\{x\}) = 0$ for all $x \in \mathbb{R}$, then F_P is continuous. Finding the quantiles of P amounts at inverting the cumulative distribution function. However, this is not always possible since F_P may be constant on part of its domain, e.g. if P is a discrete measure. Let $F_P(x^-) = \lim_{y\to x^-} F_P(x)$ be the left limit of F_P at x.

---

[7] $n = \infty$ is used to denote countable union.

[8] Equivalently, $\otimes_{i=1}^n \Sigma_i$ is the smallest σ-algebra containing the one-dimensional cylinder sets $\prod_{i=1}^{j-1} \Omega_i \times B_j \times \prod_{i=j+1}^n \Omega_i$, where $B_j \in \Sigma_j$ and $j = 1, \dots, n$ (cf. Kallenberg (1997) page 2). Indeed, $\prod_{j=1}^n B_j$, is the intersection of $\prod_{i=1}^{j-1} \Omega_i \times B_j \times \prod_{i=j+1}^n \Omega_i$ for $j = 1, \dots, n$.

Definition 1.3.1 *A quantile of order p for $P \in \mathcal{P}_1$, $0 < p < 1$, is any number $q_p(P) \in Q_p(P) = \{x \in \mathbb{R} : F_P(x^-) \leq p \leq F_P(x)\}$. A median for P is a quantile of order $1/2$ $q_{1/2}(P) \in Q_{1/2}(P)$.*

Notice that $q_p(P)$ is unique if F_P is strictly increasing in a neighborhood of $q_p(P)$ or $F_P(q_p(P)^-) < p < F_P(q_p(P))$.

The definition of quantiles heavily relies on the order "$\leq$" in $\mathbb{R}$. The lack of a natural order in $\mathbb{R}^d$, $d > 1$, prevents a straightforward generalization of the notion of quantiles and median to the multidimensional setting. In the last forty-fifty years, several attempts have been made to provide a notion of multidimensional quantiles that retains the same properties of unidimensional quantiles. An important application of medians and quantiles is robust and non-parametric location estimation. Indeed, unlike the mean, the median exists for all probability measures $P \in \mathcal{P}_1$ and it is not affected much by perturbations, such as replacing P by $P_{x,\epsilon} = (1-\epsilon)P + \epsilon\delta_x$, for some $0 < \epsilon < 1$ and $x \in \mathbb{R}$, where δ_x is the Dirac measure at x.

A natural extension of quantiles and median to $\mathbb{R}^d$ is given by componentwise quantiles and median, which are defined below. First, we provide some notation. Let S^{d-1} be the unit sphere in $\mathbb{R}^d$ and $\{e_j\}_{j=1}^d$ be the standard basis of $\mathbb{R}^d$. The orthogonal projection onto the direction $u \in S^{d-1}$ is the function $\pi_u : \mathbb{R}^d \to \mathbb{R}$ given by $\pi_u(x) = \langle x, u \rangle$. In particular, the j^{th}-coordinate of x is the projection onto e_j $\pi_{e_j}(x) = \langle x, e_j \rangle$. Finally, for a Borel measurable function $T : \mathbb{R}^d \to \mathbb{R}^{d^*}$ and $P \in \mathcal{P}_d$, let $P_T = P \circ T^{-1}$ be the push-forward measure of P w.r.t. T, that is $P_T(A) = P(T^{-1}(A))$, for all $A \in \mathcal{B}_{d^*}$. For ease of notation, we also write P_u for P_{π_u}, $u \in S^{d-1}$.

Definition 1.3.2 *A componentwise quantile of order p for $P \in \mathcal{P}_d$, $0 < p < 1$, is any vector $q_p(P) \in Q_p(P) = Q_p(P_{e_1}) \times \cdots \times Q_p(P_{e_d})$. A componentwise median for P is a componentwise quantile of order $1/2$ $q_{1/2}(P) \in Q_{1/2}(P)$.*

Reich (1980) uses componentwise median as a robust location estimate. In Galpin and Hawkins (1987) (resp. Tyler (1987)), the componentwise median is used as an intermediate estimate for a robust (resp. non-parametric) estimate of scatter. However, the componentwise median is not equivariant w.r.t. affine transformations. Indeed, as the next example shows, it is not even equivariant w.r.t. orthogonal transformations. We begin with the definition of affine and orthogonal equivariance. To this end, let $\mathcal{T}$ be the class of all Borel measurable functions $T : \mathbb{R}^d \to \mathbb{R}^d$ and $\mathcal{A}$ (resp. $\mathcal{U}$) be the class of functions $S : \mathbb{R}^d \to \mathbb{R}^d$ given by $S(x) = Mx + b$, for some invertible (resp. orthogonal) matrix M and $b \in \mathbb{R}^d$.

Definition 1.3.3 *Let $\mathcal{S} \subset \mathcal{T}$. An estimator $\zeta(P)$ is said to be $\mathcal{S}$-equivariant if $S(\zeta(P)) = \zeta(P_S)$ for all $S \in \mathcal{S}$. In particular, if $\mathcal{S} = \mathcal{A}$ (resp. $\mathcal{S} = \mathcal{U}$), then $\zeta(P)$ is said to be affine-equivariant (resp. orthogonal-equivariant).*

Example 1.3.1 $P = 9^{-1}(\delta_{(0,0)^\top} + 2\delta_{(9,9)^\top} + 3\delta_{(0,10)^\top} + 3\delta_{(10,0)^\top}) \in \mathcal{P}_2$ *has a unique componentwise median $q_{1/2}(P) = (9,9)^\top$. For $\gamma \in [0, 2\pi)$, let*

$$U_\gamma = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) \\ \sin(\gamma) & \cos(\gamma) \end{pmatrix}$$

be the rotation matrix corresponding to a clockwise rotation by angle γ about the origin. Then, $P_{U_{\pi/4}} = 9^{-1}(\delta_{(0,0)^\top} + 2\delta_{(0,9\sqrt{2})^\top} + 3\delta_{(-5\sqrt{2},5\sqrt{2})^\top} + 3\delta_{(5\sqrt{2},5\sqrt{2})^\top})$ has componentwise median $q_{1/2}(P_{U_{\pi/4}}) = (0, 5\sqrt{2})^\top$ whereas $U_{\pi/4} q_{1/2}(p) = (0, 9\sqrt{2})^\top$.

Also, notice that $q_{1/2}(P)$ is a vertex of the quadrilateral with the point masses of P as vertices. Hence, for $d > 1$, the componentwise median is not a good measure of centrality. For $d > 2$, things are even worse. Indeed, the componentwise median does not even belong, in general, to the convex hull of the measure support.

Example 1.3.2 $P = 5^{-1}(\delta_{(0,0,0)^\top} + \delta_{(5,0,0)^\top} + \delta_{(0,5,0)^\top} + \delta_{(5,10,10)^\top} + \delta_{(10,5,10)^\top}) \in \mathcal{P}_3$ *has a unique componentwise median* $q_{1/2}(P) = (5,5,0)^\top$.

An advantage of componentwise median is its fairly low computational complexity, and, despite these drawbacks, it is still used in applications involving high-dimensional spaces, such as signal processing, gene expressions and functional data analysis (Astola et al., 1990; Ohm et al., 2012; Kenne Pagui et al., 2017; Makinde, 2019; Ojo et al., 2019). Also, affine-equivariance does not play a role in sequence and functional spaces.

Tukey (1975) proposes an alternative way to define multidimensional quantiles. Instead of computing the *left* probability $F_{P_{e_j}}(\pi_{e_j}(x)) = P_{e_j}((-\infty, \pi_{e_j}(x)])$ along each coordinate projection $\pi_{e_j}(x)$, $x \in \mathbb{R}^d$, he suggested to compute the *left* probability over all projections $\pi_u(x)$ and then take the infimum over $u \in S^{d-1}$. Notice that, by taking the direction $-u$, the *right* probability is also taken into account. This lead to the concept of halfspace depth (see also Donoho and Gasko (1992), Chen (1995), Massé (2002), Massé (2004), Arcones et al. (2006), Dutta et al. (2011), and Kuelbs and Zinn (2016)).

Definition 1.3.4 (Halfspace depth) *The halfspace depth of* $x \in \mathbb{R}^d$ *w.r.t.* $P \in \mathcal{P}_d$ *is given by*

$$\tilde{D}_H(x,P) = \inf_{u \in S^{d-1}} F_{P_u}(\pi_u(x)).$$

Notice that $F_{P_u}(\pi_u(x)) = P_u((-\infty, \pi_u(x)]) = P(H_{x,u})$, where $H_{x,u} = \{y \in \mathbb{R}^d : \pi_u(x) \geq \pi_u(y)\}$ is the closed halfspace with outer normal u and x as boundary point. Then, a (halfspace-)depth median for P is a point $\mu \in \mathbb{R}^d$ satisfying $\tilde{D}_H(\mu,P) = \max_{x \in \mathbb{R}^d} \tilde{D}_H(x,P)$ and (halfspace-)depth quantile sets are the regions $R_{\tilde{D}_H,\alpha}(P) = \{x \in \mathbb{R}^d : \tilde{D}_H(x,P) \geq \alpha\}$, $0 \leq \alpha \leq \tilde{D}_H(\mu,P)$. In Section 1.5 we provide a precise definition of depth functions D, and provide further examples. Furthermore, we give conditions on P that ensure that median and quantiles based on depth functions are well-defined and enjoy reasonable properties. As we shall see an important role is played by symmetry, which is further studied in Section 1.4. The broadest known notion of symmetry is probably halfspace symmetry, which is naturally linked with halfspace depth.

Definition 1.3.5 $P \in \mathcal{P}_d$ *is said to be halfspace symmetric about* $\mu \in \mathbb{R}^d$ *if* $P(H_{\mu,u}) \geq 1/2$ *for all* $u \in S^{d-1}$. *The subclass of* $\mathcal{P}_d$ *consisting of halfspace symmetric measures is denoted by* $\mathcal{P}_{d,H}$.

Notice that P is halfspace symmetric about $\mu \in \mathbb{R}^d$ if and only if $\tilde{D}_H(\mu,P) \geq 1/2$. The constant $1/2$ is, in general, optimal. Indeed, if P_u is continuous, implying that $P_u(\{\pi_u(\mu)\}) = 0$, and $F_{P_u}(\pi_u(\mu)) = P(H_{\mu,u}) \geq 1/2$, then $P(H_{\mu,-u}) = 1 - F_{P_u}(\pi_u(\mu)) \leq 1/2$. It follows that $\tilde{D}_H(\mu,P) = 1/2$. For the same reason, $\tilde{D}_H(x,P) \leq 1/2$, for all $x \in \mathbb{R}^d$, implying that $\mu \in R_{\tilde{D}_H,1/2}(P) = \{y \in \mathbb{R}^d : \tilde{D}_H(y,P) = \max_{x \in \mathbb{R}} \tilde{D}_H(x,P)\}$. Finally, we notice that for $P \in \mathcal{P}_1$,

$$\tilde{D}_H(x,P) = \min(F_P(x), 1 - F_P(x^-)).$$

The next proposition shows that $\mu \in \mathbb{R}$ is a median for P if and only if it is a halfspace-depth median.

Proposition 1.3.1 *Let $P \in \mathcal{P}_1$ and $\mu \in \mathbb{R}$. The following statements are equivalent:*
(i) P is halfspace symmetric about μ,
(ii) $\tilde{D}_H(\mu, P) \geq 1/2$,
(iii) $\mu \in Q_{1/2}(P)$.

Proof of Proposition 1.3.1. It is enough to show the equivalence between (i) and (iii). To this end, notice that P is halfspace symmetric about μ if and only if both $F_P(\mu) = P((-\infty, \mu]) \geq 1/2$ and $1 - F_P(\mu^-) = P([\mu, \infty)) \geq 1/2$. In turn, this holds if and only if $\mu \in Q_{1/2}(P)$. ■

Let $\mathcal{P}_{d,\ll \mathrm{m}} \subset \mathcal{P}_d$ be the subclass of absolutely continuous distributions w.r.t. a Borel measure m on $\mathbb{R}^p$, and $\mathcal{P}_{d,hp} \subset \mathcal{P}_d$ be the subclass of probability measures that assign probability zero to all hyperplanes in $\mathbb{R}^d$. Notice that $\mathcal{P}_{d,\ll \mathrm{m}} \subset \mathcal{P}_{d,hp}$. Furthermore, λ is the Lebesgue measure on $\mathbb{R}^d$. Finally, for a set $A \subset \mathbb{R}^{d^*}$, $d^* \geq 1$, the indicator function $\mathbf{I}_A : \mathbb{R}^{d^*} \to \mathbb{R}$ is given by $\mathbf{I}_A(y) = 1$ if $y \in A$ and $\mathbf{I}_A(y) = 0$ otherwise. Liu (1990) proposes another notion of depth function, namely, simplicial depth.

Definition 1.3.6 (Simplicial depth) *The simplicial depth of $x \in \mathbb{R}^d$ w.r.t. $P \in \mathcal{P}_d$ is*

$$D_S(x, P) = \int_{(\mathbb{R}^d)^{d+1}} \mathbf{I}_{Z_{S,x,\infty}}(x_1, \ldots, x_{d+1}) dP(x_1) \ldots dP(x_{d+1}),$$

where $Z_{S,x,\infty} = \{(y_1, \ldots, y_{d+1}) \in (\mathbb{R}^d)^{d+1} : x \in \Delta[y_1, \ldots, y_{d+1}]\}$ and $\Delta[x_1, \ldots, x_{d+1}] \subset \mathbb{R}^d$ is the closed simplex with vertices $x_1, \ldots, x_{d+1} \in \mathbb{R}^d$.[9]

Then, a (simplicial-)depth median for P is a point $\mu \in \mathbb{R}^d$ satisfying $D_S(\mu, P) = \max_{x \in \mathbb{R}^d} D_S(x, P)$ and quantiles sets are the regions $R_{D_S,\alpha}(P) = \{x \in \mathbb{R}^d : D_S(x, P) \geq \alpha\}$, $0 \leq \alpha \leq D_S(\mu, P)$. The simplicial depth function is related to the notion of angular symmetry.

Definition 1.3.7 *$P \in \mathcal{P}_d$ is said to be angularly symmetric about $\mu \in \mathbb{R}^d$ if $P_{T_{A,\mu}} = P_{-T_{A,\mu}}$, where $T_{A,\mu} : \mathbb{R}^d \to \mathbb{R}^d$ is given by $T_{A,\mu}(x) = (x - \mu)/\|x - \mu\|_2$, if $x \neq \mu$ and $T_{A,\mu}(x) = 0$, if $x = \mu$. The subclass of $\mathcal{P}_d$ consisting of angularly symmetric measures is denoted by $\mathcal{P}_{d,A}$.*

We show in Section 1.4 that angular symmetry implies halfspace symmetry and that if $P \in \mathcal{P}_{d,hp}$ is angularly symmetric, then $P(H_{\mu,u}) = 1/2$ for all $u \in S^{d-1}$. Also, notice that, when $d = 1$, $Z_{S,x,\infty} = \{(y_1, y_2) \in \mathbb{R}^2 : y_1 \leq x \leq y_2$ or $y_2 \leq x \leq y_1\}$ implying that

$$D_S(x, P) = 2F_P(x)(1 - F_P(x^-)).$$

Corollary 1.3.1 *Suppose that $P \in \mathcal{P}_1$ is angularly symmetric about $\mu \in \mathbb{R}$. It holds that:*
(i) $\tilde{D}_H(\mu, P) \geq 1/2$,
(ii) $D_S(\mu, P) \geq 1/2$,
(iii) $\mu \in Q_{1/2}(P)$.

Proof of Corollary 1.3.1. Since P is halfspace symmetric, (i) and (iii) follow from Proposition 1.3.1. Also, halfspace symmetric implies that $F_P(\mu) = P((-\infty, \mu]) \geq 1/2$ and $1 - F_P(\mu^-) = P([\mu, \infty)) \geq 1/2$ yielding (ii). ■

Let $\mathcal{P}_{d,c} \subset \mathcal{P}_d$ be the subclass of continuous probability measures. Notice that $\mathcal{P}_{d,\ll\lambda} \subset \mathcal{P}_{d,hp} \subset \mathcal{P}_{d,c}$ and $\mathcal{P}_{1,c} = \mathcal{P}_{1,hp}$.

[9] $\Delta[x_1, \ldots, x_{d+1}]$ is defined as the convex hull of the points $x_1, \ldots, x_{d+1}$. If $x_1, \ldots, x_{d+1}$ are in general position, then $\Delta[x_1, \ldots, x_{d+1}]$ is a simplex. For $P \in \mathcal{P}_{d,hp}$ this happens with probability one.

Proposition 1.3.2 *Let $P \in \mathcal{P}_{1,c}$ and $\mu \in \mathbb{R}$. The following statements are equivalent:*
(i) P is angularly symmetric about μ,
(ii) $\tilde{D}_H(\mu, P) = 1/2$
(iii) $D_S(\mu, P) = 1/2$,
(iv) $\mu \in Q_{1/2}(P)$.

Proof of Proposition 1.3.2. $P \in \mathcal{P}_{1,c}$ is angularly symmetric about μ if and only if $P((-\infty, \mu]) = P([\mu, \infty)) = 1/2$, that is,

$$F_P(\mu) = 1 - F_P(\mu^-) = 1/2. \tag{1.3.1}$$

Therefore, (i) implies (ii), (iii), and (iv). Next, notice that $F_P(x^-) = F_P(x)$, for all $x \in \mathbb{R}$. This shows that (1.3.1) is equivalent to $F_P(\mu) = 1/2$. The latter equality is implied by either (ii), (iii), or (iv). ∎

Before concluding this section, we show alternative approaches for generalizing the notion of quantiles and median to the multidimensional setting. We denote by $\|x\|_q = (\sum_{j=1}^d |\pi_{e_j}(x)|)^{1/q}$, $q \geq 1$, the L^q-norm of $x \in \mathbb{R}^d$. Notice that, for $d = 1$, the quantile set of P of order p is $Q_p(P) = Q_p^-(P) \cap Q_p^+(P)$, where $Q_p^-(P) = \{x \in \mathbb{R} : F_P(x) \geq p\}$ and $Q_p^+(P) = \{x \in \mathbb{R} : F_P(x^-) \leq p\}$. $Q_p^-(P)$ is the set of all points with distribution function at least p and it is referred to as lower quantile set. Similarly, $Q_p^+(P)$ is the set of points with (left limit of) the cumulative distribution function at most p and it is referred to as upper quantile set. Hamel and Kostner (2018) uses tools from set optimization theory and defines lower and upper quantile sets based on closed convex cones $C \subset \mathbb{R}^d$. We recall that $C \subset \mathbb{R}^d$ is a convex cone if $\alpha z \in C$, for all $\alpha > 0$ and $z \in C$, and $x + y \in C$, for all $x, y \in C$. Suppose further that $0 \in C$ and $C \setminus \{0\} \neq \emptyset$. Lower C-quantile sets[10] of $P \in \mathcal{P}_d$ of order p are are given by $Q_{p,C}^-(P) = \cap_{w \in C\setminus\{0\}} Q_{p,w}^-(P)$, where $Q_{p,w}^-(P) = \{z \in \mathbb{R}^d : F_{P_{w/\|w\|_2}}(\pi_{w/\|w\|_2}(z)) \geq p\}$. Similarly, upper C-quantile sets are given by $Q_{p,C}^+(P) = \cap_{w \in C\setminus\{0\}} Q_{p,w}^+(P)$, where $Q_{p,w}^+(P) = \{z \in \mathbb{R}^d : F_{P_{w/\|w\|_2}}(\pi_{w/\|w\|_2}(z)^-) \leq p\}$. Notice that $z \in Q_{p,w}^\pm(P)$ if and only if $\pi_{w/\|w\|_2}(z) \in Q_p^\pm(P_{w/\|w\|_2})$. Thus, the sets $Q_{p,C}^\pm(P)$ are the intersection of all unidimensional upper (resp. lower) unidimensional quantiles sets along every projection $\pi_{w/\|w\|_2}$ with $w \in C \setminus \{0\}$. Lower and upper C-quantiles are related by

$$Q_{p,C}^-(P) = Q_{1-p,-C}^+(P). \tag{1.3.2}$$

This follows from the fact that $F_{P_{w/\|w\|_2}}(\pi_{w/\|w\|_2}(z)) = P(H_{w/\|w\|_2})$ is bigger or equal to p if and only if $F_{P_{-w/\|w\|_2}}(\pi_{-w/\|w\|_2}(z)^-) = P(\mathring{H}_{-w/\|w\|_2})$ is smaller or equal to $1 - p$. In the case of unidimensional quantiles, it holds that

$$Q_{p,1}^-(P) \cap Q_{p,1}^+(P) = Q_p(P) \text{ and } Q_{p,-1}^-(P) \cap Q_{p,-1}^+(P) = Q_{1-p}(P), \tag{1.3.3}$$

implying corresponding equalities for the cones $C = [0, \infty)$ and $C = (-\infty, 0]$. Despite this, in general, the intersection $Q_{p,C}^-(P) \cap Q_{p,C}^+(P)$ can be empty (see Example 3 in Hamel and Kostner (2018)). This is the case even for $d = 1$ and $C = \mathbb{R}$. Indeed, using (1.3.3), we have that $Q_{p,C}^-(P) \cap Q_{p,C}^+(P) = Q_p(P) \cap Q_{1-p}(P)$. Thus, lower and upper C-quantile sets have to be considered separately. An interesting property of C-quantile sets is that they are affine equivariant, in a certain sense. Namely, for

[10]For simplicity, we define quantile sets directly via the dual cone of Hamel and Kostner (2018).

$S \in \mathcal{A}$ of the form $S(x) = Mx + b$, it holds that

$$Q^{\pm}_{p,(M^{-1})^{\top}C}(P_S) = S(Q^{\pm}_{p,C}(P)). \tag{1.3.4}$$

Ideally, one would like to have the same cone on both sides of (1.3.4). By (1.3.2), it is enough to show (1.3.4) for lower C-quantiles. To this end, we denote by $g \circ h$ the compositions of functions g and h and notice that

$$F_{P_{\pi_{\frac{w}{\|w\|_2}}\circ S}}(\pi_{\frac{w}{\|w\|_2}}(z)) = P_{\pi_{\frac{w}{\|w\|_2}}\circ S}((-\infty, \pi_{\frac{w}{\|w\|_2}}(z)]) = P_S(H_{z,\frac{w}{\|w\|_2}}) = P(S^{-1}(H_{z,\frac{w}{\|w\|_2}})),$$

implying that

$$\begin{aligned} Q^{-}_{p,(M^{-1})^{\top}C}(P) &= \{z \in \mathbb{R}^d : F_{P_{\pi_{w/\|w\|_2}\circ S}}(\pi_{w/\|w\|_2}(z)) \geq p \quad \forall w \in (M^{-1})^{\top}C \setminus \{0\}\} \\ &= \{z \in \mathbb{R}^d : P(S^{-1}(H_{z,w/\|w\|_2})) \geq p \quad \forall w \in (M^{-1})^{\top}C \setminus \{0\}\} \\ &= S(\{z \in \mathbb{R}^d : P(S^{-1}(H_{S(z),((M^{-1})^{\top}w)/(\|(M^{-1})^{\top}w\|_2)})) \geq p \quad \forall w \in C \setminus \{0\}\}). \end{aligned}$$

Next, we observe that $S^{-1}(H_{S(z),((M^{-1})^{\top}w)/(\|(M^{-1})^{\top}w\|_2)})$ is equal to

$$\begin{aligned} &S^{-1}(\{y \in \mathbb{R}^d : \pi_{((M^{-1})^{\top}w)/(\|(M^{-1})^{\top}w\|_2)}(Sz) \geq \pi_{((M^{-1})^{\top}w)/(\|(M^{-1})^{\top}w\|_2)}(y)\}) \\ =&\{y \in \mathbb{R}^d : \pi_{((M^{-1})^{\top}w)/(\|(M^{-1})^{\top}w\|_2)}(Sz) \geq \pi_{((M^{-1})^{\top}w)/(\|(M^{-1})^{\top}w\|_2)}(S(y))\} \\ =&\{y \in \mathbb{R}^d : \pi_{(w)/(\|w\|_2)}(z) \geq \pi_{(w)/(\|w\|_2)}(y)\} = H_{z,w/\|w\|_2}. \end{aligned}$$

(1.3.4) follows. We show next that lower C-quantile sets are closed and convex. By (1.3.2), the same holds for upper C-quantiles. To this end, notice that the function $z \mapsto P(H_{z,w/\|w\|_2})$ is upper semicontinuous since $H_{z,w/\|w\|_2}$ is closed. It follows that $Q^{-}_{p,w}(P)$ is closed and $Q^{-}_{p,C}(P)$ is closed because it is the intersection of closed sets. We now turn to convexity: if $x, y \in \mathbb{R}^d$ satisfy $P(H_{x,w/\|w\|_2}), P(H_{y,w/\|w\|_2}) \geq p$, then, for all $\alpha \in [0,1]$,

$$\begin{aligned} P(H_{\alpha x+(1-\alpha)y,\frac{w}{\|w\|_2}}) &= P(\{z \in \mathbb{R}^d : \pi_{\frac{w}{\|w\|_2}}(z) \geq \pi_{\frac{w}{\|w\|_2}}(\alpha x + (1-\alpha)y)\} \\ &\geq P(\{z \in \mathbb{R}^d : \pi_{\frac{w}{\|w\|_2}}(z) \geq \max(\pi_{\frac{w}{\|w\|_2}}(x), \pi_{\frac{w}{\|w\|_2}}(y))\} \\ &\geq \min(P(H_{x,\frac{w}{\|w\|_2}}, P(H_{y,\frac{w}{\|w\|_2}}) \geq p. \end{aligned}$$

This also shows that halfspace-depth quantiles $R_{\tilde{D}_H,\alpha}(P)$ are closed and convex. Indeed, halfspace-depth quantiles are a special case of lower cone quantile sets when the cone is $C = \mathbb{R}^d$. Specifically, $Q^{-}_{\alpha,\mathbb{R}^d}(P) = R_{\tilde{D}_H,\alpha}(P)$.

Let $S_r(x)$ (or $S^{d-1}_r(x)$), $B_r(x)$ and $\overline{B}_r(x)$ be the sphere, and open and closed ball in $\mathbb{R}^d$ with radius $r \geq 0$ and center $x \in \mathbb{R}^d$. Chernozhukov et al. (2017) (see also Hallin et al. (2021)) notice that convexity of quantile sets is sometimes a drawback. This is the case for the banana-shaped distribution in that paper. To address, this issue, they propose to map an arbitrary probability measure $P \in \mathcal{P}_d$ to the spherical uniform measure $\mathbb{U}$ on $\overline{B}_1(0)$[11] and define quantiles of P as the image of the halfspace-depth quantiles of $\mathbb{U}$ via this map. Indeed, McCann (1995) explains that, for all $P \in \mathcal{P}_d$, there exists a convex function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that $P = \mathbb{U}_{\nabla\psi}$. Although ψ may not be unique, $\nabla\psi$ is unique $\mathbb{U}$-almost everywhere ($\mathbb{U}$-a.e.).[12] These results are due to Cuesta and Matran (1989), Rüschendorf and Rachev (1990), and Brenier (1991). Since such quantiles are based on a transportation map, we call them transportation quantiles. They are defined by

$$Q_\alpha^{(T)}(P) = \nabla\psi(R_{\tilde{D}_H,\alpha}(\mathbb{U})), \quad \alpha \in [0, 1/2].$$

Notice that the function $\nabla\psi$ can be interpreted as the solution of the optimal transport problem from $\mathbb{U}$ to P when the cost function is the squared Euclidean distance (Villani, 2009). Indeed, suppose that $P \in \mathcal{P}_d^{(2)}$, where, for $j \geq 0$, $\mathcal{P}_d^{(j)} = \{P \in \mathcal{P}_d : \int \|x\|_2^j dP(x) < \infty\}$ is the subclass of $\mathcal{P}_d$ of measures with finite j^{th}-moment. Then, by Theorem 9.4 in Villani (2009) (see also (2) in McCann (1995) and Knott and Smith (1984)),

$$\nabla\psi = \inf_{T \in \mathcal{T}: \mathbb{U}_T = P} \int \|x - T(x)\|_2^2 d\mathbb{U}(x) \quad \mathbb{U}\text{-a.e.}$$

For $d = 1$ the optimal transport problem from $\mathbb{U}$ to P is solved by $F_P^{-1} \circ F_{\mathbb{U}}$, where $F_{\mathbb{U}}$[13] is the cumulative distribution function of $\mathbb{U}$ and $F_P^{-1} : [0, 1] \to \mathbb{R}$, given by

$$F_P^{-1}(p) = \min\{x \in [-\infty, \infty] : p \leq F_P(x)\},$$

is the generalized inverse of the cumulative distribution function of P that gives the minimal quantile in $Q_p(P)$, $p \in (0, 1)$.[14] This result is due to Major (1978). See also Section 1.2 of Bonnotte (2013).

For the rest of the chapter, we deal with depth functions, which itself are defined in Section 1.5. Since the concept of symmetry is key when defining median and quantiles via depth functions, in the next section, we study properties of angular and halfspace symmetry, and define other notions of multidimensional symmetry, such as spherical and central symmetry.

[11] Even though $\mathbb{U}$ is taken to be absolutely continuous w.r.t. the Lebesgue measure and invariant w.r.t. orthogonal matrices, it is not assumed that it has constant density $(\lambda(\overline{B}_1(0)))^{-1}\mathbf{I}_{\overline{B}_1(0)}(x)$. Instead, $\mathbb{U}$ assigns the same probability to spherical shells with the same difference in radii: $\mathbb{U}(\overline{B}_{r_2}(0) \setminus \overline{B}_{r_1}(0)) = r_2 - r_1$, for all $0 \leq r_1 \leq r_2 \leq 1$. Specifically, $\mathbb{U}$ has density

$$f_{\mathbb{U}}(x) = \begin{cases} 2^{-1}\mathbf{I}_{\overline{B}_1(0)}(x), & \text{if } d = 1 \\ (H^{d-1}(S^{d-1}_{\|x\|_2}(0)))^{-1}\mathbf{I}_{\overline{B}_1(0)}(x), & \text{if } d \geq 2, \end{cases}$$

where, for $s \geq 0$, H^s is the s-dimensional Hausdorff measure on $\mathbb{R}^d$. Finally, recall that, for all $r \geq 0$ and $y \in \mathbb{R}^d$, $\lambda(\overline{B}_r(y)) = \frac{\pi^{d/2} r^d}{\Gamma(d/2+1)}$ and $H^{d-1}(S_r^{d-1}(y)) = \frac{2\pi^{d/2} r^d}{\Gamma(d/2)}$, where Γ is the gamma function.

[12] Notice that, since ψ is convex, it is locally Lipschitz continuous in the interior of the convex set $C_\psi = \mathbb{R}^d \setminus \psi^{-1}(\{\infty\})$. It follows from Rademacher's theorem that ψ is differentiable (hence, $\nabla\psi$ is well-defined) λ-a.e. in $\mathring{C}_\psi$.

[13] For $d = 1$ the cumulative distribution function of $\mathbb{U}$ is given by

$$F_{\mathbb{U}}(x) = \begin{cases} 0 & \text{if } x < -1 \\ (x+1)/2 & \text{if } x \in [-1, 1] \\ 1 & \text{if } x > 1. \end{cases}$$

[14] Here it is used that F_P is non-decreasing and right-continuous.

1.4 Multidimensional symmetry

In Section 1.3 we have seen that for $d = 1$, a point of angular/halfspace symmetry for P is a median of P. We show that angular symmetry implies halfspace symmetry. To this end, let $S_u^{d-1} = \{v \in S^{d-1} : \pi_u(v) \geq 0\}$ be the halfsphere given by the direction $u \in S^{d-1}$ and notice that $P(H_{\mu,u}) = P_{-T_{A,\mu}}(S_u^{d-1} \cup \{0\})$, for all $\mu \in \mathbb{R}^d$ and $u \in S^{d-1}$. If P is angularly symmetric about $\mu \in \mathbb{R}^d$, we have that

$$P(H_{\mu,u}) = P_{-T_{A,\mu}}(S_u^{d-1} \cup \{0\}) = P_{T_{A,\mu}}(S_u^{d-1} \cup \{0\}) = P_{-T_{A,\mu}}(S_{-u}^{d-1} \cup \{0\}) = P(H_{\mu,-u})$$

Since $1 = P(\mathbb{R}^d) = P(H_{\mu,u}) + P(H_{\mu,-u} \setminus H_{\mu,u}) \leq P(H_{\mu,u}) + P(H_{\mu,-u})$, it follows that $P(H_{\mu,u}) \geq 1/2$, for all $u \in S^{d-1}$. Therefore, $\mathcal{P}_{d,A} \subset \mathcal{P}_{d,H}$. As the next examples show this inclusion is strict. Notice that, for $d = 1$,

$$T_{A,\mu}(x) = \begin{cases} -1 & \text{if } x < \mu, \\ 0 & \text{if } x = \mu, \\ 1 & \text{if } x > \mu. \end{cases}$$

Example 1.4.1 *Let $P \in \mathcal{P}_1$ satisfy $P(\{\mu\}) > 0$ and $1/2 - P(\{\mu\}) < P((-\infty, \mu)) < P((\mu, \infty))$, for some $\mu \in \mathbb{R}$. Then, P is halfspace symmetric about μ, but is not angularly symmetric about any $\nu \in \mathbb{R}$. To see this, we use that $P((-\infty, \nu)) = P_{T_{A,\nu}}(\{-1\})$, $P((\nu, \infty)) = P_{T_{A,\nu}}(\{1\})$ and notice that $P((-\infty, \nu)) < P((\nu, \infty))$, for $\nu \leq \mu$, and $P((-\infty, \nu)) > 1/2 > P((\nu, \infty))$, for $\nu > \mu$.*

Nevertheless, it holds that $\mathcal{P}_{d,A} \cap \mathcal{P}_{d,hp} = \mathcal{P}_{d,H} \cap \mathcal{P}_{d,hp}$. Indeed, if $P \in \mathcal{P}_{d,hp}$, then $P(H_{\mu,-u} \cap H_{\mu,u}) = 0$ implying that $P(H_{\mu,-u} \setminus H_{\mu,u}) = P(H_{\mu,-u})$ and $P(H_{\mu,-u}) + P(H_{\mu,u}) = 1$. It follows that, for $P \in \mathcal{P}_{d,hp}$ angular and halfspace symmetry are equivalent, and, if angular/halfspace symmetry holds, $P(H_{\mu,u}) = 1/2$ for all $u \in S^{d-1}$. Next, we define central and spherical symmetry.

Definition 1.4.1 *$P \in \mathcal{P}_d$ is said to be centrally symmetric about $\mu \in \mathbb{R}^d$ if $P_{T_{C,\mu}} = P_{-T_{C,\mu}}$, where $T_{C,\mu} : \mathbb{R}^d \to \mathbb{R}^d$ is given by $T_{C,\mu}(x) = x - \mu$. The subclass of $\mathcal{P}_d$ consisting of centrally symmetric measures is denoted by $\mathcal{P}_{d,C}$.*

Next, we show that central symmetry implies angular symmetry. To this end, let $S : \mathbb{R}^d \to \mathbb{R}^{\tilde{d}}$ be Borel measurable, and P be centrally symmetric about $\mu \in \mathbb{R}^d$. Using the Borel measurability of S, we have that

$$P_{S(T_{C,\mu})}(B) = P_{T_{C,\mu}}(S^{-1}(B)) = P_{-T_{C,\mu}}(S^{-1}(B)) = P_{S(-T_{C,\mu})}(B). \tag{1.4.1}$$

Now, the result follows by taking $S = T_{A,0}$.

Definition 1.4.2 *$P \in \mathcal{P}_d$ is said to be spherically symmetric about $\mu \in \mathbb{R}^d$ if $P_{T_{S,I,\mu}} = P_{T_{S,U,\mu}}$, for any orthogonal matrix U, where $T_{S,U,\mu} : \mathbb{R}^d \to \mathbb{R}^d$ is given by $T_{S,U,\mu}(x) = U(x - \mu)$. The subclass of $\mathcal{P}_d$ consisting of spherically symmetric measures is denoted by $\mathcal{P}_{d,S}$.*

Letting $U = -I$, we see that spherical symmetry implies central symmetry. Hence, $\mathcal{P}_{d,S} \subset \mathcal{P}_{d,C} \subset \mathcal{P}_{d,A} \subset \mathcal{P}_{d,H}$. When $d = 1$, there are only two orthogonal matrices, 1 and -1, yielding $\mathcal{P}_{1,S} = \mathcal{P}_{1,C}$. However, not every median is a point of spherical/central symmetry. Next, we study the uniqueness of the point of symmetry. To this end, let $L_{x,u} = \{x + ru : r \in \mathbb{R}\}$ be the line through $x \in \mathbb{R}^d$ with direction $u \in S^{d-1}$. Page 409 in Liu (1990) and Theorem 2.3.1 in Zuo (1998) show that, if it exists, the point of angular and halfspace symmetry is unique, unless, for some $x \in \mathbb{R}^d$ and $u \in S^{d-1}$, $P(L_{x,u}) = 1$ and, there exists $\mu_1, \mu_2 \in L_{x,u}$, $\mu_1 \neq \mu_2$, with $\pi_u(\mu_1), \pi_u(\mu_2) \in Q_{1/2}(P_u)$. We show that, if it exists, the point of spherical and central symmetry is unique. To this end, let $P \in \mathcal{P}_{1,C}$ with $P(L_{x,u}) = 1$, $x \in \mathbb{R}^d$ and $u \in S^{d-1}$, and suppose that P is centrally symmetric about $\mu_1, \mu_2 \in L_{x,u}$. Letting $S = \pi_u$ and $B = [a, b]$, $a < b$, in (1.4.1), we have that, for $i = 1, 2$,

$$P_{\pi_u}(\pi_u(\mu_i) + [a,b]) = P_{\pi_u(T_{C,\mu_i})}([a,b]) = P_{-\pi_u(T_{C,\mu_i})}([a,b]) = P_{\pi_u}(\pi_u(\mu_i) - [a,b]).$$

By replacing a and b by $a - \pi_u(\mu_i)$ and $b - \pi_u(\mu_i)$, we see that

$$P_{\pi_u}([a,b]) = P_{\pi_u}(2\pi_u(\mu_i) - [a,b]).$$

It follows that

$$P_{\pi_u}(2\pi_u(\mu_1) - [a,b]) = P_{\pi_u}(2\pi_u(\mu_2) - [a,b]),$$

and, replacing a and b by $2\pi_u(\mu_1) - a$ and $2\pi_u(\mu_1) - b$, we conclude that

$$P_{\pi_u}([a,b]) = P_{\pi_u}(2\pi_u(\mu_2 - \mu_1) + [a,b]).$$

Since this holds for all $a < b$, it follows that $\pi_u(\mu_2) = \pi_u(\mu_1)$, and using $\mu_1, \mu_2 \in L_{x,u}$, we have that $\mu_1 = \mu_2$.

An interesting feature of the point of spherical and central symmetry of a measure P is that it coincides with its mean, whenever $P \in \mathcal{P}_d^{(1)}$. We begin with the definition of mean. The notation for covariance matrix and moments is needed in the next section.

Definition 1.4.3 *The mean of $P \in \mathcal{P}_d^{(1)}$ is $\mu(P) = \int x dP(x)$, and the covariance matrix of $P \in \mathcal{P}_d^{(2)}$ is $\Sigma(P) = \int (x - \mu(P))(x - \mu(P))^\top dP(x)$. The j^{th}-moment of $P \in \mathcal{P}_d^{(j)}$, $j \geq 0$, is $\mu_m^{(j)}(P) = \int \|x\|_2^j dP(x)$ and the j^{th} central moment is $\mu_c^{(j)}(P) = \int \|x - \mu(P)\|_2^j dP(x)$.*

Notice that the mean is affine-equivariant, that is, for $S \in \mathcal{A}$ (say, $S(x) = Mx + b$, for some invertible matrix M and $b \in \mathbb{R}^d$), it holds that

$$\mu(P_S) = \int x dP_S(x) = \int M x dP(x) + b = S(\mu(P)). \tag{1.4.2}$$

Using this, for the covariance matrix, we have that

$$\begin{aligned} \Sigma(P_S) &= \int (x - \mu(P_S))(x - \mu(P_S))^\top dP_S(x) \\ &= \int (S(x) - S(\mu(P)))(S(x) - S(\mu(P)))^\top dP(x) = M\Sigma(P)M^\top. \end{aligned} \tag{1.4.3}$$

Finally, we show that the point of central symmetry of $P \in \mathcal{P}_d^{(1)} \cap \mathcal{P}_{d,C}$ is $\mu(P)$. Indeed, suppose that P is symmetric about ν, then, using central symmetry,

$$\begin{aligned} T_{C,\nu}(\mu(P)) &= \int T_{C,\nu}(x) dP(x) = \int x dP_{T_{C,\nu}}(x) \\ &= \int x dP_{-T_{C,\nu}}(x) = \int -T_{C,\nu}(x) dP(x) = -T_{C,\nu}(\mu(P)), \end{aligned}$$

implying that $\nu = \mu(P)$.

1.5 Statistical depth functions

In this section, we define *statistical* depth functions (Liu, 1990; Liu, 1992; Zuo, 1998; Zuo and Serfling, 2000a; Dyckerhoff, 2004). These are functions of points $x \in \mathbb{R}^d$ and measures $P \in \mathcal{P}_{d,1} \subset \mathcal{P}_d$,[15] that satisfy certain properties. First, they should not vary when changes of coordinates or translations occur in both x and P. Thus, the result does not change if a transformation $S : \mathbb{R}^d \to \mathbb{R}^d$ of the type $S(x) = Mx + b$ is applied. Here, M is an invertible matrix and $b \in \mathbb{R}^d$ is a vector. Thus, invariance is w.r.t. all affine transformations $S \in \mathcal{A}$. In some cases, M can only be an orthogonal matrix, thus, $S \in \mathcal{U}$. Therefore, we define depth functions w.r.t. a general class of transformations $\mathcal{S} \subset \mathcal{T}$. $\mathcal{S}$ can be $\mathcal{A}$ or $\mathcal{U}$, but other choices are in principle possible. Second, for all measure in some class of symmetric measures $\mathcal{P}_{d,2}$, the center of symmetry must be the point of maximum depth. Here, $\mathcal{P}_{d,2}$ can be e.g. $\mathcal{P}_{d,S}$, $\mathcal{P}_{d,C}$, $\mathcal{P}_{d,A}$, or $\mathcal{P}_{d,H}$. Clearly, if $\mathcal{P}_{d,2}$ is broader, then depth functions can identify multidimensional medians for a larger class. Third, depth function are non-decreasing along any ray from the point of maximum depth, whenever it exists. Fourth, depth functions are upper semicontinuous. This condition, although it was not assumed by Zuo and Serfling (2000a), ensures that depth quantile sets are closed. Fifth and last, depth functions vanish at infinity, thus ensuring a proper order of points from the point of maximum depth outward. We are now ready to define depth functions.

Definition 1.5.1 *A non-negative and bounded function $D : \mathbb{R}^d \times \mathcal{P}_{d,1} \to \mathbb{R}$ is said to be a* statistical *depth function w.r.t. $(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$, where $\mathcal{S} \subset \mathcal{T}$ is a class of invertible transformations, $\mathcal{P}_{d,1}$ is a subclass of $\mathcal{P}_d$ satisfying $P_S \in \mathcal{P}_{d,1}$ for all $S \in \mathcal{S}$ and $\mathcal{P}_{d,2} \subset \mathcal{P}_{d,1}$ is a class of symmetric measures, if*
(i) $D(S(x), P_S) = D(x, P)$, for all $x \in \mathbb{R}^d$, $S \in \mathcal{S}$ and $P \in \mathcal{P}_{d,1}$,
(ii) $D(\mu, P) = \sup_{x\in\mathbb{R}^d} D(x, P)$, for any $P \in \mathcal{P}_{d,2}$ that is symmetric about $\mu \in \mathbb{R}^d$,
(iii) if $D(\nu, P) = \sup_{x\in\mathbb{R}^d} D(x, P)$, where $\nu \in \mathbb{R}^d$ and $P \in \mathcal{P}_{d,1}$, then $D(\nu + \alpha_1 u, P) \geq D(\nu + \alpha_2 u, P)$, for all $0 \leq \alpha_1 \leq \alpha_2$ and $u \in S^{d-1}$,
(iv) for all $P \in \mathcal{P}_{d,1}$, $D(\cdot, P)$ is upper semicontinuous,
(v) for all $P \in \mathcal{P}_{d,1}$, $\lim_{r\to\infty} \sup_{x\in\mathbb{R}^d\setminus B_r(0)} D(x, P) = 0$.
We denote by $\mathcal{D}(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$ the class of statistical *depth functions w.r.t. $(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$.*

Clearly, $(\tilde{\mathcal{S}}, \tilde{\mathcal{P}}_{d,1}, \tilde{\mathcal{P}}_{d,2}) \subset (\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$ implies that $\mathcal{D}(\tilde{\mathcal{S}}, \tilde{\mathcal{P}}_{d,1}, \tilde{\mathcal{P}}_{d,2}) \supset \mathcal{D}(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$.

Remark 1.5.1 *In some cases it is useful to have* statistical *depth functions that are unbounded below (see Sections 1.6 and 1.7 below). We denote by $\mathcal{D}'(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$ the class of functions $D' : \mathbb{R}^d \times \mathcal{P}_{d,1} \to \mathbb{R} \cup \{-\infty\}$ that are bounded above, satisfy (i)-(iv) of Definition 1.5.1 and the condition*

(v') for all $P \in \mathcal{P}_{d,1}$, $\lim_{r\to\infty} \sup_{x\in\mathbb{R}^d\setminus B_r(0)} D(x, P) = -\infty.$

Notice that the two formulations are equivalent in the sense that every $D \in \mathcal{D}(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$ can be identified with a $D' \in \mathcal{D}'(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$ and vice versa. The identification is done via a continuous, increasing, and surjective function $\psi : [0, \infty) \to \mathbb{R} \cup \{-\infty\}$ with inverse ψ^{-1}. For instance, $\psi(t) = \log(t)$ and $\psi^{-1}(t) = e^t$.

[15] Some depth functions, like Mahalanobis depth, are defined only w.r.t. some subclass $\mathcal{P}_{d,1} \subset \mathcal{P}_d$.

The importance of *statistical* depth functions lies in that they enable the definition of depth median and quantiles.

Definition 1.5.2 *Let D be a* statistical *depth function w.r.t. $(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$ and $P \in \mathcal{P}_{d,2}$. The (D-)depth quantiles of P are the upper level sets $R_{D,\alpha}(P) = \{x \in \mathbb{R}^d : D(x, P) \geq \alpha\}$, where $0 \leq \alpha \leq \alpha_m(D, P)$ and $\alpha_m(D, P) = \max_{x \in \mathbb{R}^d} D(x, P)$. A (D-)depth median of P is a point $\nu \in R_{D,\alpha_m(D,P)}(P)$.*

We show next that the properties (i)-(iv) in Definition 1.5.1 imply analogous properties for the depth quantiles $R_{D,\alpha}(P)$ (see Dyckerhoff (2004)). Specifically, (i) ensures that depth quantiles are $\mathcal{S}$-equivariant. Indeed, for $S \in \mathcal{S}$, it holds that

$$R_{D,\alpha}(P_S) = \{x \in \mathbb{R}^d : D(x, P_S) \geq \alpha\} = \{Sy \in \mathbb{R}^d : D(Sy, P_S) \geq \alpha\} = S(R_{D,\alpha}(P_S)).$$

In particular, this holds for $\alpha = \alpha_m(D, P)$, implying that depth medians are $\mathcal{S}$-equivariant. Definition 1.5.1 (iv)-(v) implies that the sets $R_{D,\alpha}(P)$ are closed and bounded (and thus compact) for all $\alpha > 0$. Then, upper semicontinuity implies the existence of a point of maximum depth, that is, a depth median. We now turn to (iii). If ν is a point of maximum depth and $P \in \mathcal{P}_{d,2}$, then the sets $R_{D,\alpha}(P)$ are star-shaped w.r.t. ν, that is, $(1-t)\nu + tx \in R_{D,\alpha}(P)$, for all $x \in R_{D,\alpha}(P)$ and $t \in [0, 1]$. Finally, (ii) guaranties that the point of symmetry of $P \in \mathcal{P}_{d,2}$ is a $(D$-)depth median.

Many depth functions (e.g. halfspace and simplicial depth) satisfy Definition 1.5.1 (i) with $\mathcal{S} = \mathcal{A}$. In general, the larger the class $\mathcal{P}_{d,2}$ is, the more information is obtained on quantiles of probability measures with a known center of symmetry. Finally, the class $\mathcal{P}_{d,1}$ should be large enough to include finite discrete probability measures (cf. Section 1.9).

Zuo and Serfling (2000a) divides depth functions into four types: *Type A*, *Type B*, *Type C*, and *Type D*. The *Type A* depth of a point $x \in \mathbb{R}^d$ w.r.t. a probability measure $P \in \mathcal{P}_d$ is the integral w.r.t. P^k, $k \geq 1$, of a bounded and non-negative function $h_{x,\infty} : (\mathbb{R}^d)^k \to [0, \infty)$. Similarly, *Type B* depth functions are obtained by applying a function $g : [0, \infty] \to \mathbb{R}$, which is continuous, decreasing, positive on $[0, \infty)$, and zero at infinity, to the integral of an unbounded and non-negative function $i_{x,\infty} : (\mathbb{R}^d)^k \to [0, \infty)$. A typical choice is the function $g(t) = 1/(1 + t)$, $t \in [0, \infty]$. Next, *Type C* depth functions are obtained by applying the function g to a measure of outlyingness $O : \mathbb{R}^d \times \mathcal{P}_{d,1} \to [0, \infty]$, where $\mathcal{P}_{d,1} \subset \mathcal{P}_d$. Finally, halfspace depth $(\tilde{D}_H)$ seems to be the only example of *Type D* depth function. Theorem 2.1 in Zuo and Serfling (2000a) shows that, for the halfspace depth, one can choose $\mathcal{P}_{d,1} = \mathcal{P}_d$ and $\mathcal{P}_{d,2} = \mathcal{P}_{d,H}$ in Definition 1.5.1 yielding $\tilde{D}_H \in \mathcal{D}(\mathcal{A}, \mathcal{P}_d, \mathcal{P}_{d,H})$. In the next three sections we further analyze *Type A*, *Type B*, and *Type C* depth functions and provide additional examples. We denote *Type B* depth functions by $\hat{D}$ (with a possible subscript) and other depth functions such as *Type C* depth functions and halfspace depth by $\tilde{D}$ (with a possible subscript). We reserve the notation D (without any tilde $\tilde{}$ or hat $\hat{}$) to *Type A* depth functions, some of which will be further analyzed in Chapter 2. The subscript ∞ in the functions $h_{x,\infty}$ and $i_{x,\infty}$ refers to depth, in contrast to local depth (cf. Section 2.2).

1.6 Type A depth functions

Definition 1.6.1 *A* Type A *depth is a function $D_G : \mathbb{R}^d \times \mathcal{P}_d \to [0, \infty)$ given by*

$$D_G(x, P) = \int h_{G,x,\infty}(x_1, \ldots, x_{k_G}) dP(x_1) \ldots dP(x_{k_G}),$$

where $h_{G,x,\infty} : (\mathbb{R}^d)^{k_G} \to [0, \infty)$ is a Borel measurable, non-negative and bounded function depending on indeces G and x. $k_G \geq 1$ depends on G.[16]

The simplicial depth is a typical example of *Type A* depth function and is obtained by taking $G = S$, $k_S = d + 1$, and $h_{S,x,\infty} = \mathbf{I}_{Z_{S,x,\infty}}$ (see Definition 1.3.6). Liu (1990) shows that D_S satisfies (i) with $\mathcal{S} = \mathcal{A}$, (iv) and (v), as well as (ii) and (iii) for the subclass $\mathcal{P}_{d,A} \cap \mathcal{P}_{d,hp} \subset \mathcal{P}_d$. We notice that, in that paper, absolute continuity w.r.t. the Lebesgue measure can be replaced by the condition that every hyperplane has probability zero. Upper semicontinuity follows from the fact that $Z_{S,x,\infty}$ is closed for all $x \in \mathbb{R}^d$. However, $D_S \notin \mathcal{D}(\mathcal{A}, \mathcal{P}_d, \mathcal{P}_{d,A} \cap \mathcal{P}_{d,hp})$. Indeed, (iii) needs to hold for all $P \in \mathcal{P}_d$ and not only for $P \in \mathcal{P}_{d,A} \cap \mathcal{P}_{d,hp}$. Counterexamples 1-2 in Section 3.2 of Zuo and Serfling (2000a) show that (iii) fails even for the class $\mathcal{P}_{d,C}$. Finally, Counterexample 3 shows that (ii) does not hold for $\mathcal{P}_{d,2} = \mathcal{P}_{d,H}$.

Example 1.6.1 (β-skeleton depth) *The β-skeleton depth (Bremner and Shahsavarifar, 2018; Yang and Modarres, 2018) is obtained by taking $\beta \in [1, \infty)$, $G = K_\beta$, $k_{K_\beta} = 2$, and $h_{K_\beta,x,\infty} = \mathbf{I}_{Z_{K_\beta,x,\infty}}$, where*

$$Z_{K_\beta,x,\infty} = \{(x_1, x_2) \in (\mathbb{R}^d)^2 : \max_{(i,j)\in\{(1,2),(2,1)\}} \|x_i + (2/\beta - 1)x_j - 2/\beta x\|_2 \leq \|x_1 - x_2\|_2\}.$$

Spherical depth (Elmore et al., 2006) and lens depth (Liu and Modarres, 2011) are obtained by setting in the above definition $\beta = 1$ and $\beta = 2$, respectively. We let $B = K_1$ and $L = K_2$ yielding $D_B = D_{K_1}$ and $D_L = D_{K_2}$. Also, it is worth mentioning that for $d = 1$ β-skeleton depth and simplicial depth coincide. We now turn to the properties of *statistical* depth function. (iv)-(v) of Definition 1.5.1 hold true using that, for fixed $x_1, x_2 \in \mathbb{R}^d$, the function $x \mapsto \mathbf{I}_{Z_{K_\beta,x,\infty}}(x_1, x_2)$ is upper semicontinuous and vanishes when x lies outside a suitable ball.[17] Using invariance of the Euclidean norm w.r.t. orthogonal transformations we see that (i) holds for $\mathcal{S} = \mathcal{U}$. Turning to (ii) and (iii), Kleindessner and Von Luxburg (2017) notice that the proof of (ii) given for D_L is wrong. They also notice that the proofs of (ii) for D_B and D_{K_β} contain the same mistake. In particular, Liu and Modarres (2011) states that (ii) holds with $\mathcal{P}_{d,2} = \mathcal{P}_{d,C}$. A counterexample for (ii) and (iii) and $P \in \mathcal{P}_{d,C}$ is given by Geenens et al. (2021). Specifically, suppose that $P_z \in \mathcal{P}_{d,\ll\lambda}$ has density function given by

$$f_{P_z}(x) = \frac{1}{2}\phi_2(x - z) + \frac{1}{2}\phi_2(x + z),$$

where ϕ_d is the d-variate standard normal density and $z \in \mathbb{R}^d \setminus \{0\}$. Then, P is centrally symmetric about 0,[18], however, $D_{K_\beta}(\cdot, P_z)$ fails to satisfy (ii) and (iii). The function f_{P_z} is plotted for $d = 2$ and $z = (3, 0)^\top$ in Figure 1.1. The corresponding β-skeleton depth for $\beta = 1, 1.5, 2, 4$ is plotted in Figure 1.2. We conclude that D_{K_β} is not a *statistical* depth function as it does not satisfy (iii) of Definition 1.5.1.

[16] The subscript ∞ is used to emphasize that $h_{G,x,\infty}$ yields a depth function in contrast to a local depth function (cf. Chapter 2). The index G helps distinguishing a depth function from another. For instance, when G is replaced by S we obtain simplicial depth. In some cases, for *Type A* depth functions, the index G can be identified with a kernel function $G : (\mathbb{R}^d)^{k_G} \to [0, \infty)$ (see Section 2.2).

[17] Using the triangle inequality $\|x_i + (2/\beta - 1)x_j - 2/\beta x\|_2 \geq |\|x_i + (2/\beta - 1)x_j\|_2 - 2/\beta\|x\|_2|$ yielding that $\mathbf{I}_{Z_{K_\beta,x,\infty}}(x_1, x_2) = 0$ whenever $2/\beta\|x\|_2 > \max_{(i,j)\in\{(1,2),(2,1)\}} \|x_i + (2/\beta - 1)x_j\|_2 + \|x_1 - x_2\|_2$.

[18] Using that $\phi_d(y) = \phi_d(-y)$, we immediately see that $f_{P_z}(T_{C,0}(x)) = f_{P_z}(-T_{C,0}(x))$ yielding that $(P_z)_{T_{C,0}}(A) = P_z(A) = P_z(-A) = (P_z)_{-T_{C,0}}(A)$ for all $A \in \mathcal{B}_d$.

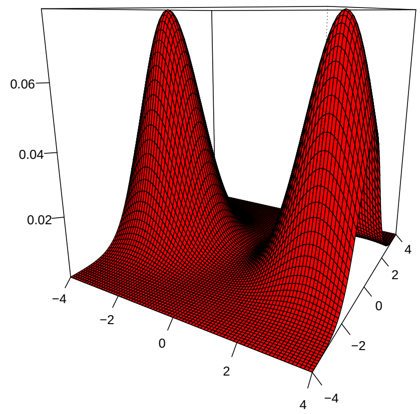

FIGURE 1.1: Density function f_{P_z} for $d = 2$ and $z = (3, 0)^\top$.

One could ask whether D_{K_β} satisfies (ii) and (iii) for $P \in \mathcal{P}_{d,S}$. A first step in this direction is given by the following proposition, which shows that, if P is spherically symmetric about $\mu \in \mathbb{R}^d$, then all points that are equidistant from μ have the same β-skeleton depth. It follows that it is enough to check (ii) and (iii) along a single direction $u \in S^{d-1}$ (for instance $u = e_1$). For an orthogonal matrix U and $\mu \in \mathbb{R}^d$ let $T_{R,U,\mu} : \mathbb{R}^d \to \mathbb{R}^d$ be the rotation matrix given by $T_{R,U,\mu}(x) = \mu + T_{S,U,\mu}(x) = \mu + U(x - \mu)$.

Proposition 1.6.1 *Let $P \in \mathcal{P}_{d,S}$ be spherically symmetric about $\mu \in \mathbb{R}^d$. Then, for any orthogonal matrix U, $D_{K_\beta}(T_{R,U,\mu}(x), P) = D_{K_\beta}(x, P)$.*

Proof of Proposition 1.6.1. We first notice that spherical symmetry is equivalent to $P_{T_{R,U,\mu}} = P$ for all orthogonal matrices U. Therefore,

$$D_{K_\beta}(T_{R,U,\mu}(x), P) = P^2(Z_{K_\beta, T_{R,U,\mu}(x), \infty}) = P^2_{T_{R,U,\mu}}(Z_{K_\beta, T_{R,U,\mu}(x), \infty}).$$

Next, we use the invariance of the Euclidean norm w.r.t. orthogonal matrices and get that

$$\begin{aligned}
&\{(x_1, x_2) \in (\mathbb{R}^d)^2 : \max_{(i,j)\in\{(1,2),(2,1)\}} \|x_i + (2/\beta - 1)x_j - 2/\beta T_{R,U,\mu}(x)\|_2 \le \|x_1 - x_2\|_2\} \\
=&\{(x_1, x_2) \in (\mathbb{R}^d)^2 : \max_{(i,j)\in\{(1,2),(2,1)\}} \|U(T^{-1}_{R,U,\mu}(x_i) + (2/\beta - 1)T^{-1}_{R,U,\mu}(x_j) - 2/\beta x)\|_2 \\
&\le \|U(T^{-1}_{R,U,\mu}(x_1) - T^{-1}_{R,U,\mu}(x_2))\|_2\} \\
=&\{(x_1, x_2) \in (\mathbb{R}^d)^2 : \max_{(i,j)\in\{(1,2),(2,1)\}} \|T^{-1}_{R,U,\mu}(x_i) + (2/\beta - 1)T^{-1}_{R,U,\mu}(x_j) - 2/\beta x\|_2 \\
&\le \|T^{-1}_{R,U,\mu}(x_1) - T^{-1}_{R,U,\mu}(x_2)\|_2\}
\end{aligned}$$

This yields that $Z_{K_\beta, T_{R,U,\mu}(x), \infty} = (T_{R,U,\mu}, T_{R,U,\mu})(Z_{K_\beta, x, \infty})$, where for $S, T \in \mathcal{T}$, $(S, T) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ is the function given by $(S, T)(x, y) = (S(x), T(y))$. We conclude that

$$D_{K_\beta}(T_{R,U,\mu}(x), P) = P^2_{T_{R,U,\mu}}((T_{R,U,\mu}, T_{R,U,\mu})(Z_{K_\beta, x, \infty})) = P^2(Z_{K_\beta, x, \infty}) = D_{K_\beta}(x, P).$$

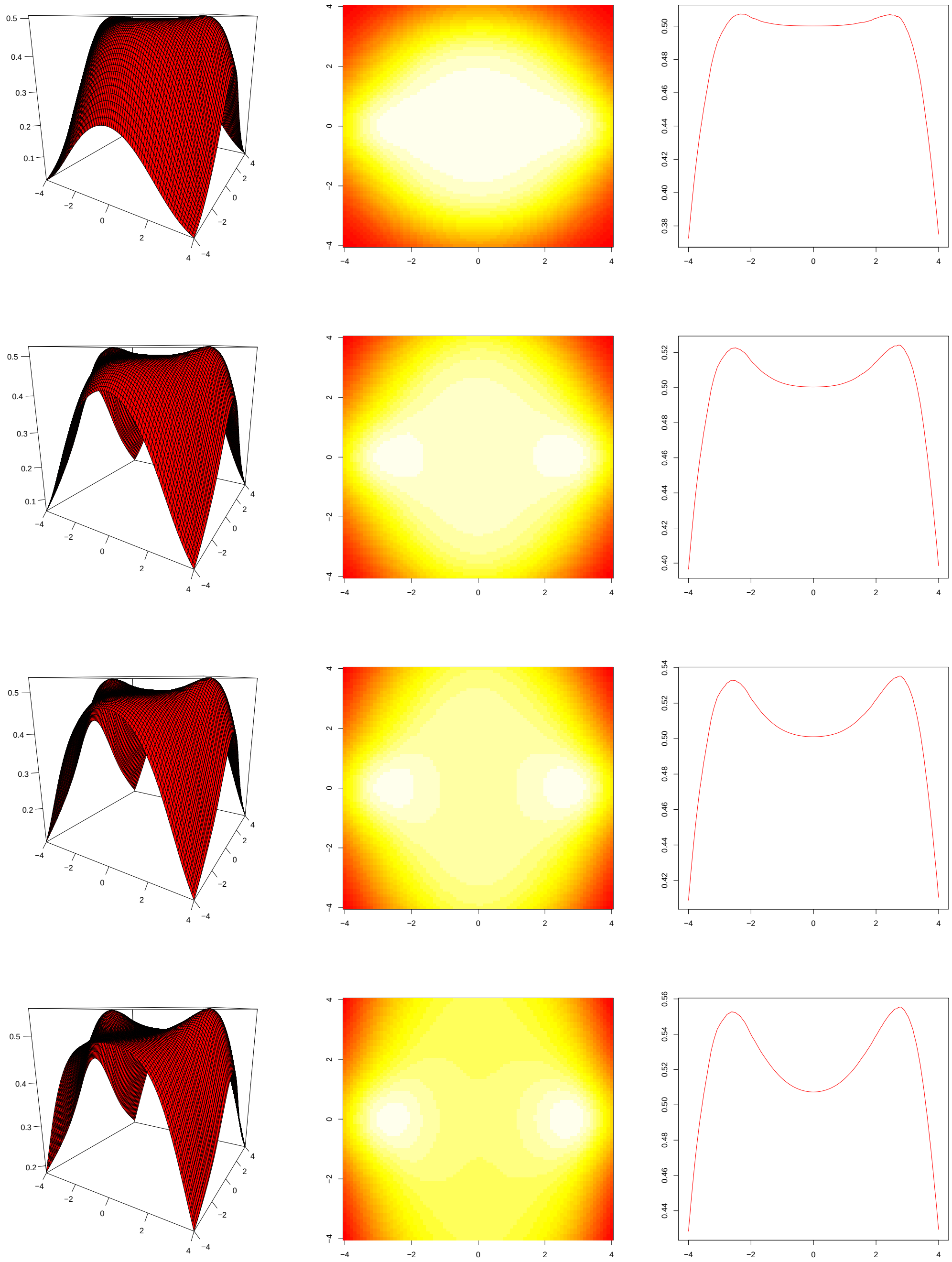

FIGURE 1.2: β-skeleton depth constructed with $n = 5000$ samples from P_z for $d = 2$ and $z = (3, 0)^\top$. From top to bottom, $\beta = 1, 1.5, 2, 4$. From left to right, β-skeleton depth, corresponding heat map, and its section along the line $\{x \in \mathbb{R}^2 : \pi_{e_2}(x) = 0\}$.

■

We now turn to a general discussion on *Type A* depth functions. Zuo (1998) show that (ii) and (iii) of Definition 1.5.1 are satisfied by *Type A* depth functions if the function $x \mapsto h_{G,x,\infty}(x_1,\ldots,x_{k_G})$ is concave (see Theorem 3.3.1 and Theorem 3.3.2 of Zuo (1998)). Unfortunately, the function $x \mapsto h_{G,x,\infty}(x_1,\ldots,x_{k_G})$ is concave if and only if it is constant (see the lemma below). In view of (v) of Definition 1.5.1, the only *Type A statistical* depth function with concave function $x \mapsto h_{G,x,\infty}(x_1,\ldots,x_{k_G})$ is given by $h_{G,x,\infty} = 0$ for all $x \in \mathbb{R}^d$. Indeed, by (v) $D_G(x,P) = 0$ for all $x \in \mathbb{R}^d$, yielding that $h_{G,x,\infty} = 0$ P^k-a.s. Since P is arbitrary, we conclude that $h_{G,x,\infty} = 0$. Accordingly, Zuo and Serfling (2000a) does not provide a systematic study of *Type A* depth functions. We now state and proof the lemma, which applies to the function $x \mapsto h_{G,x,\infty}(x_1,\ldots,x_{k_G})$ since it is bounded below by zero.

Lemma 1.6.1 *Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be either bounded below and concave or bounded above and convex. Then φ is constant.*

Proof of Lemma 1.6.1. Suppose by contradiction that $\varphi(x) < \varphi(y)$ for some $x, y \in \mathbb{R}^d$. If φ is concave, then for all $\alpha > 1$

$$\varphi(x) = \varphi\left(\frac{1}{\alpha}(\alpha(x-y)+y) + \left(1-\frac{1}{\alpha}\right)y\right) \geq \frac{1}{\alpha}\varphi(\alpha(x-y)+y) + \left(1-\frac{1}{\alpha}\right)\varphi(y).$$

It follows that

$$\varphi(\alpha(x-y)+y) \leq \alpha(\varphi(x)-\varphi(y)) + \varphi(y)$$

implying that $\varphi(\alpha(x-y)+y)$ becomes arbitrary small as α increases. Similarly, if φ is convex, we get that

$$\varphi(\alpha(y-x)+x) \geq \alpha(\varphi(y)-\varphi(x)) + \varphi(x)$$

and this becomes arbitrary large as α increases. ■

Thus, to use concavity, the function $x \mapsto h_{G,x,\infty}(x_1,\ldots,x_{k_G})$ must be unbounded below.

Proposition 1.6.2 *Let $D'_G : \mathbb{R}^d \times \mathcal{P}_d \to \mathbb{R} \cup \{-\infty\}$ be given by*

$$D'_G(x,P) = \int h'_{G,x,\infty}(x_1,\ldots,x_{k_G})dP(x_1)\ldots dP(x_{k_G}),$$

where $h'_{G,x,\infty} : (\mathbb{R}^d)^{k_G} \to \mathbb{R} \cup \{-\infty\}$ is a Borel measurable function that is bounded above (uniformly in x). If the function $x \mapsto h'_{G,x,\infty}(x_1,\ldots,x_{k_G})$ is concave for fixed $x_1,\ldots,x_{k_G}$, then D'_G satisfies (iii) of Definition 1.5.1. If, additionally, $P \in \mathcal{P}_{d,C}$ and $D'_G(\cdot,P)$ is symmetric w.r.t. the center of symmetry μ of P,[19] then D'_G satisfies (ii) of Definition 1.5.1 with $\mathcal{P}_{d,2} = \mathcal{P}_{d,C}$.

Proof of Proposition 1.6.2. For (iii) let $P \in \mathcal{P}_d$ and $\nu \in \mathbb{R}^d$ be such that $D'_G(\nu,P) = \sup_{x\in\mathbb{R}^d} D'_G(x,P)$. Also, let $u \in S^{d-1}$ and $0 \leq \alpha_1 < \alpha_2$. Using the concavity of $x \mapsto h'_{G,x,\infty}(x_1,\ldots,x_{k_G})$ and $\nu + \alpha_1 u = (1-\frac{\alpha_1}{\alpha_2})\nu + \frac{\alpha_1}{\alpha_2}(\nu+\alpha_2 u)$ we have that

$$D'_G(\nu+\alpha_1 u, P) \geq \left(1-\frac{\alpha_1}{\alpha_2}\right)D'_G(\nu,P) + \frac{\alpha_1}{\alpha_2}D'_G(\nu+\alpha_2 u) \geq D'_G(\nu+\alpha_2 u).$$

[19] That is, $D'_G(x,P) = D'_G(2\mu - x, P)$ for all $x \in \mathbb{R}^d$.

For (ii) let $P \in \mathcal{P}_{d,C}$ be centrally symmetric about $\mu \in \mathbb{R}^d$. Using the concavity, we have that

$$D'_G(\mu, P) \geq \frac{1}{2} D'_G(x, P) + \frac{1}{2} D'_G(2\mu - x, P) = D'_G(x, P).$$

■

If $P \in \mathcal{P}_{d,C}$ is centrally symmetric about $\mu \in \mathbb{R}^d$, then by Proposition 1.6.1 $D_{K_\beta}(x, P) = D_{K_\beta}(2\mu - x, P)$ for all $x \in \mathbb{R}^d$[20] (cf. Theorem 2 of Yang and Modarres (2018)). However, the function $x \mapsto \mathbf{I}_{Z_{K_\beta,x,\infty}}(x_1, x_2)$ is clearly not concave.

1.7 Type B depth functions

Definition 1.7.1 *A* Type B *depth is a function* $\hat{D}_G : \mathbb{R}^d \times \mathcal{P}_d \to [0, \infty)$ *given by*

$$\hat{D}_G(x, P) = g\left(\int i_{G,x,\infty}(x_1, \ldots, x_{k_G}) dP(x_1) \ldots dP(x_{k_G}) \right),$$

where $i_{G,x,\infty} : (\mathbb{R}^d)^{k_G} \to [0, \infty)$ *is a Borel measurable, non-negative and unbounded function depending on indexes* G *and* x. $g : [0, \infty] \to \mathbb{R}$ *is continuous, decreasing, positive on* $[0, \infty)$*, and zero at infinity.* $k_G \geq 1$ *depends on* G. [21]

Convexity of the function $x \mapsto i_{G,x,\infty}(x_1, \ldots, x_{k_G})$ for fixed $x_1, \ldots, x_{k_G}$ plays an important role in showing that $\hat{D}_G$ satisfies (ii) and (iii) of Definition 1.5.1 (see Theorems 3.3.4 and 3.3.5 of Zuo (1998)).

Proposition 1.7.1 *Let* $\hat{D}_G$ *be a* Type B *depth function. If the function* $x \mapsto i_{G,x,\infty}(x_1, \ldots, x_{k_G})$ *is convex for fixed* $x_1, \ldots, x_{k_G}$*, then* $\hat{D}_G$ *satisfies (iii) of Definition 1.5.1. If, additionally,* $P \in \mathcal{P}_{d,C}$ *and* $\hat{D}_G(\cdot, P)$ *is symmetric w.r.t. the center of symmetry* μ *of* P*, then* $\hat{D}_G$ *satisfies (ii) of Definition 1.5.1 with* $\mathcal{P}_{d,2} = \mathcal{P}_{d,C}$.

Proof of Proposition 1.7.1. For (iii) let $P \in \mathcal{P}_d$ and $\nu \in \mathbb{R}^d$ be such that $\hat{D}_G(\nu, P) = \sup_{x \in \mathbb{R}^d} \hat{D}_G(x, P)$. Fix $u \in S^{d-1}$ and $0 \leq \alpha_1 < \alpha_2$. Using that $\nu + \alpha_1 u = (1 - \frac{\alpha_1}{\alpha_2})\nu + \frac{\alpha_1}{\alpha_2}(\nu + \alpha_2 u)$ and convexity, we have that

$$\begin{aligned}
\hat{D}_G(\nu + \alpha_1 u, P) \geq g\Bigg(\left(1 - \frac{\alpha_1}{\alpha_2}\right) & \int i_{G,\nu,\infty}(x_1, \ldots, x_{k_G}) dP(x_1) \ldots dP(x_{k_G}) \\
& + \frac{\alpha_1}{\alpha_2} \int i_{G,\nu+\alpha_2 u,\infty}(x_1, \ldots, x_{k_G}) dP(x_1) \ldots dP(x_{k_G}) \Bigg) \\
\geq \min_{y \in \{\nu, \nu + \alpha_2 u\}} & \hat{D}_G(y, P) = \hat{D}_G(\nu + \alpha_2 u, P).
\end{aligned}$$

[20] To see this, take $U = -I$ in Proposition 1.6.1 and notice that by central symmetry we have that $P_{T_{\mathbb{R},-I,\mu}} = P$.

[21] As before, the subscript ∞ is used to emphasize that $h_{G,x,\infty}$ yields a depth function in contrast to a local depth function (cf. Chapter 2). The index G helps distinguishing a depth function from another. For instance, when G is replaced by N_q we obtain L^q-depth (see Example 1.7.1 below).

For (ii) let $P \in \mathcal{P}_{d,C}$ be centrally symmetric about $\mu \in \mathbb{R}^d$. Since $\mu = \frac{1}{2}x + \frac{1}{2}(2\mu - x)$, we have that

$$\begin{aligned}\hat{D}_G(\mu, P) &\geq g\Big(\frac{1}{2}\int i_{G,x,\infty}(x_1, \dots, x_{k_G})dP(x_1)\dots dP(x_{k_G}) \\ &\quad + \frac{1}{2}\int i_{G,2\mu-x,\infty}(x_1, \dots, x_{k_G})dP(x_1)\dots dP(x_{k_G})\Big) \\ &\geq \min_{y\in\{x,2\mu-x\}} \hat{D}_G(y, P) = \hat{D}_G(x, P),\end{aligned}$$

where we have used that $\hat{D}_G(x, P) = \hat{D}_G(2\mu - x, P)$. ■

Next, we give some examples of *Type B* depth functions and check if they satisfy (i)-(v) of Definition 1.5.1.

Example 1.7.1 (L^q-depth) *The L^q-depth (Zuo and Serfling, 2000a) is obtained by taking in Definition 1.7.1 $G = N_q$, $k_{N_q} = 1$, and $i_{N_q,x,\infty}$ given by $i_{N_q,x,\infty}(x_1) = \|x - x_1\|_q$.*

Notice that $\int \|x - x_1\|_q dP(x_1) < \infty$ if and only if $P \in \mathcal{P}_d^{(1)}$.[22] Thus, if $P \notin \mathcal{P}_d^{(1)}$, the L^q-depth does not provide any useful information about P although it is still well-defined. As the next proposition shows, it is a *statistical* depth function. To this end, let $\mathcal{R}$ be the class of transformation $R : \mathbb{R}^d \to \mathbb{R}^d$ given by $R(x) = \pm Mx + b$ for some permutation matrix M and $b \in \mathbb{R}^d$.

Proposition 1.7.2 *For all $q \geq 1$, $\hat{D}_{N_q} \in \mathcal{D}(\mathcal{R}, \mathcal{P}_d, \mathcal{P}_{d,C})$ and $\hat{D}_{N_q}(\cdot, P)$ is continuous for all $P \in \mathcal{P}_d$. Additionally, $\hat{D}_{N_2} \in \mathcal{D}(\mathcal{U}, \mathcal{P}_d, \mathcal{P}_{d,C})$.*

Proof of Proposition 1.7.2. We verify (i)-(v) of Definition 1.5.1. Notice that for all $S \in \mathcal{T}$ of the form $S(x) = Mx + b$ for some matrix M and vector b

$$\int \|S(x) - x_1\|_q dP_S(x_1) = \int \|M(x - x_1)\|_q dP(x_1).$$

Now, if M is a permutation matrix, then $\|M(x - x_1)\|_q = \|x - x_1\|_q$. Similarly, if M is an orthogonal matrix, then $\|M(x - x_1)\|_2 = \|x - x_1\|_2$. This gives (i). Now, (ii) and (iii) follow from Proposition 1.7.1 since the function $x \mapsto \|x - x_1\|_q$ is convex for all fixed $x_1 \in \mathbb{R}^d$[23] and $D_G(\cdot, P)$ is symmetric about μ if $P \in \mathcal{P}_{d,C}$ is centrally symmetric about μ. Indeed, since $T_{C,\mu} \in \mathcal{R}$, we have that for all $x \in \mathbb{R}^d$

$$\begin{aligned}\hat{D}_{N_q}(x, P) &= \hat{D}_{N_q}(T_{C,\mu}(x), P_{T_{C,\mu}}) = \hat{D}_{N_q}(T_{C,\mu}(x), P_{-T_{C,\mu}}) \\ &= g\Big(\int \|T_{C,\mu}(x) + T_{C,\mu}(x_1)\|_q dP(x_1)\Big) = \hat{D}_{N_q}(2\mu - x, P).\end{aligned}$$

Next, we notice that $D(\cdot, P)$ is continuous because the functions g and $x \mapsto \int \|x - x_1\|_q dP(x_1)$ are continuous.[24] In particular, (iv) holds true. Turning to (v), notice that, if $\|x_1\|_q \leq 1 \leq r \leq \|x\|_q$, then by the triangle inequality $\|x - x_1\|_q \geq \|x\|_q - \|x_1\|_q \geq$

[22] By the triangle inequality $\int \|x_1\|_q dP(x_1) - \|x\|_q \leq \int \|x - x_1\|_q dP(x_1) \leq \int \|x_1\|_q dP(x_1) + \|x\|_q$ and, since all norms on $\mathbb{R}^d$ are equivalent, $\int \|x_1\|_q dP(x_1) < \infty$ if and only if $\int \|x_1\|_2 dP(x_1) < \infty$.

[23] By the triangle inequality $\|tx + (1-t)y - x_1\|_q \leq t\|x - x_1\|_q + (1-t)\|y - x_1\|_q$ for all $t \in [0, 1]$ and $x, y \in \mathbb{R}$.

[24] If $\int \|x_1\|_q dP(x_1) < \infty$, then $|\int \|x - x_1\|_q dP(x_1) - \int \|y - x_1\|_q dP(x_1)| \leq \|x - y\|_q$ for all $x, y \in \mathbb{R}^d$. Otherwise, for all $x \in \mathbb{R}^d$, $\int \|x - x_1\|_q dP(x_1) \geq \int \|x_1\|_q dP(x_1) - \|x\|_q = \infty$.

$r-1$ yielding that

$$\int \|x-x_1\|_q dP(x_1) \geq \int_{\{y\in\mathbb{R}^d:\|x_1\|_q\leq 1\}} \|x-x_1\|_q dP(x_1) \geq r-1.$$

We conclude that

$$\lim_{r\to\infty} \sup_{x\in\mathbb{R}^d\setminus B_r(0)} \hat{D}_{N_q}(x,P) \leq \lim_{r\to\infty} g(r-1) = 0.$$

■

Example 1.7.2 (Simplicial volume depth) *The simplicial volume depth (Oja, 1983; Zuo and Serfling, 2000a) is obtained by taking in Definition 1.7.1 $G=V$, $k_V=d$, and $i_{V,x,\infty}$ given by $i_{V,x,\infty}(x_1,\dots,x_d)=\lambda(\Delta[x,x_1,\dots,x_d])$.*

Next, we investigate (i)-(v) of Definition 1.5.1. It is known (Stein, 1966) that the volume of the simplex $\Delta[x,x_1,\dots,x_d]$ is

$$\lambda(\Delta[x,x_1,\dots,x_d]) = \frac{1}{d!}|\det\begin{pmatrix}1 & 1 & \dots & 1\\ x & x_1 & \dots & x_d\end{pmatrix}|.$$

Furthermore, subtracting the first column to every other column, one has

$$\det\begin{pmatrix}1 & 1 & \dots & 1\\ x & x_1 & \dots & x_d\end{pmatrix} = \det\begin{pmatrix}1 & 0 & \dots & 0\\ x & x_1-x & \dots & x_d-x\end{pmatrix} = \det\begin{pmatrix}x_1-x & \dots & x_d-x\end{pmatrix}.$$ [25]

For $S\in\mathcal{A}$ of the form $S(y)=My+b$, $i=1,\dots,d$, one analogously has

$$\begin{aligned}\det\begin{pmatrix}1 & 1 & \dots & 1\\ S(x) & S(x_1) & \dots & S(x_d)\end{pmatrix} &= \det\begin{pmatrix}S(x_1)-S(x) & \dots & S(x_d)-S(x)\end{pmatrix}\\ &= \det(M)\det\begin{pmatrix}x_1-x & \dots & x_d-x\end{pmatrix}.\end{aligned}$$

It follows that

$$\lambda(\Delta[S(x),S(x_1),\dots,S(x_d)]) = \det(M)\lambda(\Delta[x,x_1,\dots,x_d]). \tag{1.7.1}$$

In particular, if $S\in\mathcal{U}$ then $\det(M)=1$ yielding

$$\hat{D}_V(S(x),P_S) = g\Big(\int \lambda(\Delta[S(x),S(x_1),\dots,S(x_d)])dP(x_1)\dots dP(x_d)\Big) = \hat{D}_V(x,P).$$

Thus, (i) of Definition 1.5.1 holds with $\mathcal{S}=\mathcal{U}$. (ii) and (iii) are obtained from Proposition 1.7.1 using the convexity of the integrand $x\mapsto\lambda(\Delta[x,x_1,\dots,x_d])$ for fixed $x_1,\dots,x_d$.[26] For (ii) we additionally use that the simplicial volume depth is symmetric about the point $\mu\in\mathbb{R}^d$ of central symmetry of $P\in\mathcal{P}_{d,C}$. Indeed, using $T_{C,\mu}\in\mathcal{U}$,

[25] $|\det(x_1-x \ \dots \ x_d-x)|$ is the volume of the parallelepiped spanned by $x_1-x,\dots,x_d-x$.

[26] Using Laplace expansion w.r.t. the first column we see that, for all $t\in[0,1]$ and $x,y\in\mathbb{R}^d$,

$$\det\begin{pmatrix}1 & 1 & \dots & 1\\ tx+(1-t)y & x_1 & \dots & x_d\end{pmatrix} = t\det\begin{pmatrix}1 & 1 & \dots & 1\\ x & x_1 & \dots & x_d\end{pmatrix} + (1-t)\det\begin{pmatrix}1 & 1 & \dots & 1\\ y & x_1 & \dots & x_d\end{pmatrix}$$

yielding $\lambda(\Delta[sx+ty,x_1,\dots,x_d]) \leq t\lambda(\Delta[x,x_1,\dots,x_d]) + (1-t)\lambda(\Delta[y,x_1,\dots,x_d])$.

central symmetry, and (1.7.1), we obtain that

$$\begin{aligned}\hat{D}_V(x,P) &= \hat{D}_V(T_{C,\mu}(x), P_{T_{C,\mu}}) = \hat{D}_V(T_{C,\mu}(x), P_{-T_{C,\mu}})\\ &= g\Big(\int \lambda(\Delta[-T_{C,\mu}(x), T_{C,\mu}(x_1),\dots,T_{C,\mu}(x_d)])dP(x_1)\dots dP(x_d)\Big)\\ &= \hat{D}_V(2\mu - x, P).\end{aligned}$$

Next, we show (iv). Since g is continuous and decreasing, it is enough to show that the function $x \mapsto \int \lambda(\Delta[x,x_1,\dots,x_d])dP(x_1)\dots dP(x_d)$ is lower semicontinuous. To this end, we first notice that the function $x \mapsto \lambda(\Delta[x,x_1,\dots,x_d])$ is continuous. Indeed, Hadamard's inequality (Theorem 14.1.1 in Garling (2007)) gives

$$|\det\begin{pmatrix}1 & 1 & \dots & 1\\ x & x_1 & \dots & x_d\end{pmatrix}| \le (1+\|x\|_2)\prod_{j=1}^{d}(1+\|x_j\|_2), \tag{1.7.2}$$

which implies that

$$|\lambda(\Delta[x,x_1,\dots,x_d]) - \lambda(\Delta[y,x_1,\dots,x_d])| \le d!\prod_{j=1}^{d}(1+\|x_j\|_2)\|x-y\|_2.$$

Finally, using Fatou's lemma we have that, for every sequence $\{y_j\}_{j=1}^{\infty}$ converging to $x \in \mathbb{R}^d$,

$$\begin{aligned}&\liminf_{j\to\infty}\int \lambda(\Delta[y_j,x_1,\dots,x_d])dP(x_1)\dots dP(x_d)\\ &\ge \int \liminf_{j\to\infty}\lambda(\Delta[y_j,x_1,\dots,x_d])dP(x_1)\dots dP(x_d)\\ &= \int \lambda(\Delta[x,x_1,\dots,x_d])dP(x_1)\dots dP(x_d).\end{aligned}$$

We now turn to (v). Surprisingly, $\hat{D}_V$ does not satisfy this property. To see this, suppose that the support of $P \in \mathcal{P}_d$ is contained in a hyperplane, that is, $P(\partial H_{z,u}) = 1$ for some $z \in \mathbb{R}^d$ and $u \in S^{d-1}$. Since for all $r > 0$ there exists $x \in (\mathbb{R}^d \setminus B_r(0)) \cap \partial H_{z,u}$ yielding $\lambda(\Delta[x,\cdot,\dots,\cdot])) = 0$ P^d-a.s., we conclude that

$$\lim_{r\to\infty}\sup_{x\in\mathbb{R}^d\setminus B_r(0)}\hat{D}_V(x,P) = \lim_{r\to\infty}\sup_{x\in(\mathbb{R}^d\setminus B_r(0))\cap\partial H_{z,u}}\hat{D}_V(x,P) = g(0) > 0.$$

Thus, even though simplicial volume satisfies (i) of Definition 1.5.1 with $\mathcal{S} = \mathcal{A}$, (ii) with $\mathcal{P}_{d,2} = \mathcal{P}_{d,C}$, (iii), and (iv), it fails to satisfies (v) and therefore it is not a *statistical* depth function. Also, we notice that $\int \lambda(\Delta[x,x_1,\dots,x_d])dP(x_1)\dots dP(x_d)$ may be infinite, in which case, very little can be deduced from $\hat{D}_V(x,P)$ about P. However, (1.7.2) ensures that $\int \lambda(\Delta[x,x_1,\dots,x_d])dP(x_1)\dots dP(x_d)$ is finite for $P \in \mathcal{P}_d^{(1)}$.

Finally, we notice that a *Type B* depth functions can be transformed into a *Type A* depth function by applying the function g directly to $i_{G,x,\infty}$; thus, computing the integral of $h_{G,x,\infty} = g \circ i_{G,x,\infty}$ w.r.t. P^k (cf. Remark 3.3.2 of Zuo (1998)). Of course, the *Type A* depth function obtained in this way is different from the original *Type B* depth function. Moreover, because of the lack of concavity of $x \mapsto h_{G,x,\infty}(x_1,\dots,x_{k_G})$, $G = V, N_q$, (ii) and (iii) of Definition 1.5.1 do not hold in general. To address this issue, one can replace g by a continuous, decreasing and concave function $g' : [0,\infty] \to \mathbb{R} \cup \{-\infty\}$ and let D'_{N_q} and D'_V be as in Proposition 1.6.2 with $h'_{G,x,\infty} = g' \circ i_{G,x,\infty}$, $G =$

N_q, V. Since for fixed $x_1, \dots, x_{k_G}$ the functions $x \mapsto i_{G,x,\infty}(x_1, \dots, x_{k_G})$ are convex and g' is decreasing and concave, the functions $x \mapsto h'_{G,x,\infty}(x_1, \dots, x_{k_G})$ are also concave and (ii) and (iii) hold by Proposition 1.6.2. For (ii) we additionally use that $D'_{N_q}(\cdot, P)$ and $D'_V(\cdot, P)$ are symmetric about the center of symmetry μ of $P \in \mathcal{P}_{d,C}$, which can be obtained as before. (i) can be shown as before for both D'_{N_q} and D'_V. We now turn to (iv) and (v) for the L^q-depth. Since the function $x \mapsto h'_{N_q,x,\infty}(x_1)$ is bounded by $g'(0)$ and continuous, by reverse Fatou's lemma, we have that, for every sequence $\{y_j\}_{j=1}^{\infty}$ converging to $x \in \mathbb{R}^d$,

$$\limsup_{j\to\infty} D'_{N_q}(y_j, P) \leq \int \limsup_{j\to\infty} g(\|y_j - x_1\|_q) dP(x_1) = D_{N_q}(x, P).$$

Finally, using that g' is decreasing and the triangle inequality, we have that

$$\begin{aligned}\sup_{x\in\mathbb{R}^d\setminus B_r(0)} \int g'(\|x - x_1\|_q) dP(x_1) &\leq \int \sup_{x\in\mathbb{R}^d\setminus B_r(0)} g'(|\|x\|_q - \|x_1\|_q|) dP(x_1) \\ &\leq g'(0)P(\mathbb{R}^d \setminus \overline{B}_{r/2}(0)) + g'(r/2)P(\overline{B}_{r/2}(0)),\end{aligned}$$

where the last term converges to $-\infty$ as $r \to \infty$. We conclude that $D'_{N_q} \in \mathcal{D}'(\mathcal{R}, \mathcal{P}_d, \mathcal{P}_{d,C})$, $q \geq 1$, and $D'_{N_2} \in \mathcal{D}'(\mathcal{U}, \mathcal{P}_d, \mathcal{P}_{d,C})$.

1.8 Type C depth functions

Definition 1.8.1 *A* Type C *depth function is a function $\tilde{D}_G : \mathbb{R}^d \times \mathcal{P}_{d,1} \to [0, \infty)$ given by*

$$\tilde{D}_G(x, P) = g(O_G(x, P)),$$

where $\mathcal{P}_{d,1} \subset \mathcal{P}_d$, $O_G : \mathbb{R}^d \times \mathcal{P}_{d,1} \to [0, \infty]$ is Borel measurable, and $g : [0, \infty] \to \mathbb{R}$ is continuous, decreasing, positive on $[0, \infty)$, and zero at infinity.

As for *Type B* depth functions, (ii) and (iii) of Definition 1.5.1 follow from convexity (see Theorems 3.3.8 and 3.3.9 of Zuo (1998)).

Proposition 1.8.1 *Let $\tilde{D}_G$ be a* Type C *depth function and $P \in \mathcal{P}_{d,1}$. If the function $x \mapsto O_G(x, P)$ is convex, then $\hat{D}_G$ satisfies (iii) of Definition 1.5.1. If, additionally, $P \in \mathcal{P}_{d,C}$ and $O_G(\cdot, P)$ is symmetric w.r.t. the center of symmetry μ of P, then $\tilde{D}_G$ satisfies (ii) of Definition 1.5.1 with $\mathcal{P}_{d,2} = \mathcal{P}_{d,C}$.*

Proof of Proposition 1.8.1. The proof is similar to that of Proposition 1.7.1. ∎

The class $\mathcal{P}_d^{(2)}$ of probability distributions admitting second moment can be partitioned into $\mathcal{P}_d^{(2,i)} \cup \mathcal{P}_d^{(2,s)}$, where $P \in \mathcal{P}_d^{(2,i)}$ has invertible covariance matrix and $\mathcal{P}_d^{(2,s)} = \mathcal{P}_d^{(2)} \setminus \mathcal{P}_d^{(2,i)}$.

Example 1.8.1 (Mahalanobis depth) *The Mahalanobis distance (Mahalanobis, 1936) on $\mathbb{R}^d$ w.r.t. $P \in \mathcal{P}_d^{(2,i)}$ is $d_{\Sigma(P)} : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ given by*

$$d^2_{\Sigma(P)}(x, y) = (x - y)^\top (\Sigma(P))^{-1}(x - y).$$[27]

The Mahalanobis depth (Liu, 1992; Liu and Singh, 1993) is obtained by taking in Definition 1.8.1 $G = M$, $\mathcal{P}_{d,1} = \mathcal{P}_d^{(2,i)}$, *and* $O_M(x,P) = d^2_{\Sigma(P)}(x, \mu(P))$.

$\mu(P)$ and $\Sigma(P)$ can be replaced by any other affine-equivariant location estimator and covariance measure satisfying (1.4.2) and (1.4.3) (Zuo and Serfling, 2000a).

Proposition 1.8.2 $\tilde{D}_M \in \mathcal{D}(\mathcal{A}, \mathcal{P}_d^{(2,i)}, \mathcal{P}_{d,C} \cap \mathcal{P}_d^{(2,i)})$ *and* $\tilde{D}_M(\cdot, P)$ *is continuous for all* $P \in \mathcal{P}_d^{(2,i)}$.

Proof of Proposition 1.8.2. We verify (i)-(v) of Definition 1.5.1 for $\tilde{D}_M$. (i) follows from (1.4.2) and (1.4.3). For (ii) and (iii) we notice that $O(\cdot, P)$ is convex and symmetric about $\mu(P)$, which is the center of symmetry of $P \in \mathcal{P}_{d,C} \cap \mathcal{P}_d^{(2,i)}$, and apply Proposition 1.8.1. Next, $\tilde{D}_M(\cdot, P)$ is continuous because g and the function $x \mapsto d^2_{\Sigma(P)}(x, \mu(P))$ are continuous. In particular, this gives (iv). Finally, (v) follows from $\lim_{r\to\infty} \inf_{x \in \mathbb{R}^d \setminus B_r(0)} d^2_{\Sigma(P)}(x, \mu(P)) = \infty$ and $\lim_{t\to\infty} g(t) = 0$. ∎

We now turn to the projection depth (Stahel, 1981; Donoho and Gasko, 1992; Liu, 1992; Zuo and Serfling, 2000a; Zuo, 2003; Dyckerhoff, 2004).

Example 1.8.2 (Projection depth) *The projection depth is obtained by taking in Definition 1.8.1* $G = J$, $\mathcal{P}_{d,1} = \mathcal{P}_d^{(2)}$, *and* $O_J(x,P) = \sup_{u \in S^{d-1}} (\Sigma(P_u))^{-1/2} |\pi_u(x) - \mu(P_u)|$.[28]

Proposition 1.8.3 $\tilde{D}_J \in \mathcal{D}(\mathcal{A}, \mathcal{P}_d^{(2)}, \mathcal{P}_{d,C} \cap \mathcal{P}_d^{(2)})$.

Proof of Proposition 1.8.3. Let $P \in \mathcal{P}_d^{(2)}$. For (i), notice that, for $S \in \mathcal{A}$ given by $S(x) = Mx + b$,

$$\mu((P_S)_u) = \int t d(P_{\pi_u \circ S})(t) = \int \pi_u(S(x)) dP(x) = \pi_u(S(\mu(P)))$$

and

$$\Sigma((P_S)_u) = \int (t - \mu((P_S)_u))^2 d(P_{\pi_u \circ S})(t) = \int (\pi_u(S(x - \mu(P))))^2 dP(x)$$

implying that, replacing $M^\top u / \|M^\top u\|_2$ by v,

$$\begin{aligned}
\tilde{D}_J(S(x), P_S) &= g\Bigg(\sup_{u \in S^{d-1}} \left(\int (\pi_u(S(x - \mu(P))))^2 dP(x) \right)^{-1/2} |\pi_u(S(x - \mu(P)))| \Bigg) \\
&= g\Bigg(\sup_{u \in S^{d-1}} \left(\int (\pi_u(M(x - \mu(P))))^2 dP(x) \right)^{-1/2} |\pi_u(M(x - \mu(P)))| \Bigg) \\
&= g\Bigg(\sup_{v \in S^{d-1}} \left(\int (\pi_v(x - \mu(P)))^2 dP(x) \right)^{-1/2} |\pi_v(x - \mu(P))| \Bigg) \\
&= \tilde{D}_J(x, P).
\end{aligned}$$

[27] If M is a positive definite matrix, then $\langle x, y \rangle_M = x^\top M^{-1} y$ is a scalar product on $\mathbb{R}^d$ yielding that $\|x\|_M = \langle x, x \rangle_M$ is a norm. Then, $d_{\Sigma(P)}(x,y) = \|x - y\|_{\Sigma(P)}$ is the distance induced by this norm.

[28] Here, it is used the convention that

$$0^{-1} a = \begin{cases} \infty & \text{if } a \neq 0, \\ 0 & \text{if } a = 0. \end{cases}$$

(ii) and (iii) follow from Proposition 1.8.1 and the convexity of $O_J(\cdot, P)$. For (ii) we also use that $P \in \mathcal{P}_{d,C} \cap \mathcal{P}_d^{(2)}$ is centrally symmetric about $\mu(P)$ and

$$\pi_u(\mu(P)) = \pi_u\Big(\int x dP(x)\Big) = \int \pi_u(x) dP(x) = \mu(P_u),$$

which implies that $O_J(\cdot, P)$ is symmetric about $\mu(P)$. Next, (iv) follows from the fact that $O_J(\cdot, P)$ is lower semicontinuous (as it is a supremum of lower semicontinuous functions) and g is decreasing and continuous.[29] We now turn to (v). Notice that, by Cauchy–Schwarz inequality, for all $u \in S^{d-1}$,

$$|\mu(P_u)| \leq \int |t| dP_u(t) = \int |\pi_u(x)| dP(x) \leq \mu_m^{(1)}(P)$$

and

$$\Sigma(P_u) = \int (t - \mu(P_u))^2 dP_u(t) = \int (\pi_u(x - \mu(P)))^2 dP(x) \leq \mu_c^{(2)}(P)$$

implying that

$$O_J(x, P) \geq (\mu_c^{(2)}(P))^{-1/2}(\sup_{u \in S^{d-1}} |\pi_u(x)| - \mu_m^{(1)}(P)) = (\mu_c^{(2)}(P))^{-1/2}(\|x\|_2 - \mu_m^{(1)}(P)).$$

It follows from the continuity of g that

$$\lim_{r \to \infty} \sup_{x \in \mathbb{R}^d \setminus B_r(0)} \tilde{D}(x, P) \leq \lim_{r \to \infty} g((\mu_c^{(2)}(P))^{-1/2}[r - \mu_m^{(1)}(P)]) = g(\infty) = 0.$$

■

It is usually preferred to replace mean and variance by other location and dispersion estimators such as the median and median absolute deviation. However, they are not, in general, unique. For instance, Zuo and Serfling (2000a) claims that, if $P \in \mathcal{P}_{d,H}$ is halfspace symmetric about a unique point $v \in \mathbb{R}^d$, then the projection depth based on median and median absolute deviation is maximized at v. However, for this, it is used that P has a unique unidimensional median (namely, $\pi_u(v)$) for all $u \in S^{d-1}$. Part (ii) of Theorem 2.4 in Zuo and Serfling (2000b) (which is used in the proof) gives only $\pi_u(v) \in Q_{1/2}(P_u)$ and not $Q_{1/2}(P_u) = \{\pi_u(v)\}$. The following example shows that uniqueness of unidimensional medians can easily fail even for centrally symmetric probability measures.

Example 1.8.3 *Let $P^{(1)} \in \mathcal{P}_2$ be the discrete measure assigning probability $1/4$ to each of the points $(1,1)^\top, (-1,1)^\top, (-1,-1)^\top, (1,-1)^\top \in \mathbb{R}^2$. Then $P^{(1)} \in \mathcal{P}_{2,C} \cap \mathcal{P}_2^{(2)}$ is centrally symmetric about $v = (0,0)^\top$ implying that $\pi_u(v) = 0$ for all $u \in S^{d-1}$. However,*

$$Q_{1/2}((P^{(1)})_u) = \begin{cases} [\bar{u}_1 - \bar{u}_2, \bar{u}_2 - \bar{u}_1] & \text{if } \bar{u}_1 < \bar{u}_2, \\ [\bar{u}_2 - \bar{u}_1, \bar{u}_1 - \bar{u}_2] & \text{if } \bar{u}_1 \geq \bar{u}_2, \end{cases}$$

where $\bar{u}_i = |\pi_{e_i}(u)|$, $i = 1,2$. Thus, $Q_{1/2}((P^{(1)})_u) = \{\pi_u(v)\}$ if and only if

$$2u \in \{(\sqrt{2}, \sqrt{2})^\top, (-\sqrt{2}, \sqrt{2})^\top, (-\sqrt{2}, -\sqrt{2})^\top, (\sqrt{2}, -\sqrt{2})^\top\}.$$

[29] Let for simplicity $\varphi = O_J(\cdot, P)$. The lower semicontinuity of φ implies that, for all $a \in [0, g(0)]$, the set $(g \circ \varphi)^{-1}([a, \infty)) = \varphi^{-1}(g^{-1}([a, \infty))) = \varphi^{-1}([0, g^{-1}(a)])$ is closed.

For $P \in \mathcal{P}_{d,C}$ it seems that one can avoid the issue of non-uniqueness by defining the median as the midpoint of the interval $Q_{1/2}(P_u)$. However, as the following example shows, $\pi_u(v)$ is, in general, no longer the midpoint of $Q_{1/2}(P_u)$ for $P \in \mathcal{P}_{d,H} \setminus \mathcal{P}_{d,C}$.

Example 1.8.4 *Let $P^{(\alpha)} \in \mathcal{P}_2$, where $\alpha \in (0,1) \cup (1,\infty)$, be the discrete measure assigning probability $1/4$ to each of the points $(\alpha,\alpha)^\top, (-1,1)^\top, (-1,-1)^\top, (\alpha,-\alpha)^\top \in \mathbb{R}^2$. Then $P^{(\alpha)} \in (\mathcal{P}_{2,A} \setminus \mathcal{P}_{2,C}) \cap \mathcal{P}_2^{(2)}$ is angularly symmetric about $v = (0,0)^\top$. A careful calculation yields*

$$Q_{1/2}((P^{(\alpha)})_u) = \begin{cases} E_{1,u} & \text{if } \overline{u}_1 < \overline{u}_2, \pi_{e_1}(u) \geq 0, \\ E_{2,u} & \text{if } \overline{u}_1 < \overline{u}_2, \pi_{e_1}(u) < 0, \\ [\overline{u}_2 - \overline{u}_1, \alpha(\overline{u}_1 - \overline{u}_2)] & \text{if } \overline{u}_1 \geq \overline{u}_2, \pi_{e_1}(u) \geq 0, \\ [\alpha(\overline{u}_2 - \overline{u}_1), \overline{u}_1 - \overline{u}_2] & \text{if } \overline{u}_1 \geq \overline{u}_2, \pi_{e_1}(u) < 0, \end{cases}$$

where $\overline{u}_i = |\pi_{e_i}(u)|$ for $i = 1,2$,

$$E_{1,u} = \begin{cases} [\alpha(\overline{u}_1 - \overline{u}_2), \alpha(\overline{u}_1 + \overline{u}_2)] & \text{if } \overline{u}_2 \geq \frac{1+\alpha}{1-\alpha}\overline{u}_1, 0 < \alpha < 1, \\ [\alpha(\overline{u}_1 - \overline{u}_2), \overline{u}_2 - \overline{u}_1] & \text{if } \overline{u}_2 < \frac{1+\alpha}{|\alpha-1|}\overline{u}_1, \\ [-\overline{u}_1 - \overline{u}_2, \overline{u}_2 - \overline{u}_1] & \text{if } \overline{u}_2 \geq \frac{1+\alpha}{\alpha-1}\overline{u}_1, \alpha > 1, \end{cases}$$

for $u \in S^{d-1}$ satisfying $\overline{u}_1 < \overline{u}_2, \pi_{e_1}(u) \geq 0$, and

$$E_{2,u} = \begin{cases} [\alpha(-\overline{u}_1 - \overline{u}_2), \alpha(\overline{u}_2 - \overline{u}_1)] & \text{if } \overline{u}_2 \geq \frac{1+\alpha}{1-\alpha}\overline{u}_1, 0 < \alpha < 1, \\ [\overline{u}_1 - \overline{u}_2, \alpha(\overline{u}_2 - \overline{u}_1)] & \text{if } \overline{u}_2 < \frac{1+\alpha}{|\alpha-1|}\overline{u}_1, \\ [\overline{u}_1 - \overline{u}_2, \overline{u}_1 + \overline{u}_2] & \text{if } \overline{u}_2 \geq \frac{1+\alpha}{\alpha-1}\overline{u}_1, \alpha > 1, \end{cases}$$

for $u \in S^{d-1}$ satisfying $\overline{u}_1 < \overline{u}_2, \pi_{e_1}(u) < 0$.

Notice that the above examples do not show that the projection depth with unidimensional median as location estimator is maximized at the center of symmetry, but they raise some questions about the definition of median and the proof.

1.9 Sample depth and quantiles

We now turn to the estimation of a depth function $D \in \mathcal{D}(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$. To this end, let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with probability distribution $P \in \mathcal{P}_{d,1}$. A natural estimator of $D(x,P)$, where $x \in \mathbb{R}^d$ and $P \in \mathcal{P}_{d,1}$, is obtained by replacing P with the empirical probability distribution $P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$. The corresponding estimator, $D(x,P_n)$, is referred to as sample depth. We denote by $\mathcal{P}_{d,d}$ (resp. $\mathcal{P}_{d,fd}$) the subclass of $\mathcal{P}_d$ of discrete (resp. finitely discrete) probability measures. It is desirable that $\mathcal{P}_{d,1}$ is large enough to contain $\mathcal{P}_{d,fd}$. Indeed, this implies that $D(\cdot, P_n)$ satisfies (i), (iii), (iv), and (v) of Definition 1.5.1. From Section 1.5 we deduce that P_n possesses well-defined depth quantiles and median for all $n \in \mathbb{N}$. This condition is satisfied by many of the *statistical* depth functions of Sections 1.5-1.8, namely, halfspace depth, projection depth, and L^q-depth. However, since discrete distribution can have singular covariance matrix, it is not satisfied by Mahalanobis depth. Furthermore, it is not satisfied by simplicial volume depth, which does not satisfy (v), simplicial depth and β-skeleton depths, which do not satisfy (iii). Of course, even if P is symmetric, P_n is not in general symmetric. Thus, the depth median[30] of

[30]Notice that by (i) and (iii)-(v) a depth median always exists even though it is not in general unique.

P_n does not in general coincide with the depth median of P. Then, it is of interest to study (uniform) convergence of $D(x, P_n)$ to $D(x, P)$, the corresponding rate, and convergence of quantiles and medians. The last question is addressed by Theorem 4.1 in Zuo and Serfling (2000c),[31] which we now state. $D_n(x, P)$ denotes a general estimator of $D(x, P)$ defined on a probability space $(\Omega, \Sigma, \mathbb{P})$ such that $P = \mathbb{P}_X$ is the probability distribution of a random variable $X : \Omega \to \mathbb{R}^d$. We use the abbreviation $\mathbb{P}$-a.s. (or a.s.) for almost surely w.r.t. $\mathbb{P}$.

Theorem 1.9.1 *For $P \in \mathcal{P}_d$, let $D(\cdot, P) : \mathbb{R}^d \to [0, \infty)$ be continuous and $D_n(\cdot, P) : \mathbb{R}^d \times \Omega \to [0, \infty)$ be such that $\sup_{x \in \mathbb{R}^d} |D(x, P) - D_n(x, P)| \xrightarrow{a.s.} 0$. For every sequence $\{\alpha_n\}_{n=1}^{\infty}$ of positive scalars converging to $\alpha > 0$, it holds that*

$$\mathring{R}_{D,\alpha}(P) \subset \liminf_{n\to\infty} R_{D_n,\alpha_n}(P) \subset \limsup_{n\to\infty} R_{D_n,\alpha_n}(P) \subset R_{D,\alpha}(P) \text{ a.s.}$$

and, if $P(\partial R_{D,\alpha}(P)) = 0$, then

$$\lim_{n\to\infty} R_{D_n,\alpha_n}(P) = R_{D,\alpha}(P) \text{ a.s.}^{32}$$

Proof of Theorem 1.9.1. For $k, n \in \mathbb{N}$, let $A_{n,k} = \{x \in \mathbb{R}^d : |D(x, P) - D_n(x, P)| < 1/k\}$. We first show that $\lim_{n\to\infty} A_{n,k} = \mathbb{R}^d$ a.s. To this end, notice that, for $k \in \mathbb{N}$ and almost all $\omega \in \Omega$, there exists $n_0(k, \omega) \in \mathbb{N}$ such that $|D(x, P) - D_n(x, P)| < 1/k$ for all $x \in \mathbb{R}^d$. It follows that

$$\liminf_{n\to\infty} A_{n,k}(\omega) = \lim_{n\to\infty} \cap_{l=n}^{\infty} A_{n,k}(\omega) \supset \cap_{l=n_0(k,\omega)}^{\infty} A_{l,k} = \mathbb{R}^d.$$

Next, let $n_1 \in \mathbb{N}$ such that $|\alpha_n - \alpha| < 1/k$ for all $n \geq n_1$. Then, for $n \geq n_1$, $R_{D_n,\alpha+1/k}(P) \subset R_{D_n,\alpha_n}(P) \subset R_{D_n,\alpha-1/k}(P)$. Using Corollary B.1 (v), we have that

$$\begin{aligned}\liminf_{n\to\infty} R_{D_n,\alpha_n}(P) &\supset \liminf_{n\to\infty} (R_{D_n,\alpha+1/k}(P) \cap A_{n,k}) \\ &\supset \lim_{n\to\infty} (R_{D,\alpha+2/k}(P) \cap A_{n,k}) = R_{D,\alpha+2/k}(P) \text{ a.s.}\end{aligned}$$

and

$$\begin{aligned}\limsup_{n\to\infty} R_{D_n,\alpha_n}(P) &\subset^{33} \limsup_{n\to\infty} (R_{D_n,\alpha-1/k}(P) \cap A_{n,k}) \\ &\subset \lim_{n\to\infty} (R_{D,\alpha-2/k}(P) \cap A_{n,k}) = R_{D,\alpha-2/k}(P) \text{ a.s.}\end{aligned}$$

Using the continuity of $D(\cdot, P)$ we conclude that

$$\begin{aligned}\mathring{R}_{D,\alpha}(P) = \cup_{k=1}^{\infty} R_{D,\alpha+2/k}(P) &\subset \liminf_{n\to\infty} R_{D_n,\alpha_n}(P) \subset \limsup_{n\to\infty} R_{D_n,\alpha_n}(P) \\ &\subset \cap_{k=1}^{\infty} R_{D,\alpha-2/k}(P) = R_{D,\alpha}(P) \text{ a.s.}\end{aligned}$$

[31] One needs to take some extra care in the statement and proof of this result. For instance, the authors use that, if a sequence of random variables $\{X_i\}_{i=1}^{\infty}$ converges almost surely to X then, for all $\epsilon > 0$, there exists $n_\epsilon \in \mathbb{N}$ such that $|X_i - X| \leq \epsilon$ almost surely for all $i \geq n_\epsilon$. However, this is not in general true because pointwise convergence does not imply uniform convergence. We provide here a simplified argument.

[32] That is, $\mathbb{P}$-a.s. the sets $\liminf_{n\to\infty} R_{D_n,\alpha_n}(P)$, $\limsup_{n\to\infty} R_{D_n,\alpha_n}(P)$, and $R_{D,\alpha}(P)$ differ at most by a set of probability zero.

[33] Here, it is crucial that $\lim_{n\to\infty} A_{n,k}$ exists a.s. (c.f. Lemma B.1 (ii)).

The last part follows from $R_{D,\alpha}(P) = \mathring{R}_{D,\alpha}(P) \cup \partial R_{D,\alpha}(P)$ and $\mathbb{P}(X^{-1}(\partial R_{D,\alpha}(P))) = P(\partial R_{D,\alpha}(P)) = 0$. ∎

Thus, for convergence of depth quantiles it is enough to check continuity of the depth function and almost sure uniform convergence of $D(\cdot, P_n)$ to $D(\cdot, P)$, which is also referred to as uniform consistency. By (iv) of Definition 1.5.1 for continuity it is enough to check lower semicontinuity. For instance, halfspace and simplicial depth are continuous if $P \in \mathcal{P}_{d,hp}$ (see Lemma 6.1 in Donoho and Gasko (1992), Theorem 2 in Liu (1990), and Proposition 12 of Francisci et al. (2019)). Furthermore, L^q-depth and Mahalanobis depth are continuous by Proposition 1.7.2 and Proposition 1.8.2, respectively. Uniform consistency can be often obtained using tools from empirical processes theory (see Section 1.11 below). For instance, it is well-known that the uniform law of large numbers holds for VC classes of sets (see Definition 1.11.3). Since closed halfspaces on $\mathbb{R}^d$ have this property, we obtain that

$$\sup_{x\in\mathbb{R}^d} |\tilde{D}_H(x, P_n) - \tilde{D}_H(x, P)| \leq \sup_{x\in\mathbb{R}^d, u\in S^{d-1}} |P(H_{x,u}) - P_n(H_{x,u})| \xrightarrow[n\to\infty]{a.s.} 0,$$

which is given by (6.2) and (6.6) of Donoho and Gasko (1992). Uniform convergence of sample projection depth is studied in Theorem 2.2 of Zuo (2003) under some conditions on the unidimensional location and dispersion estimators. In the next proposition we study uniform consistency of Mahalanobis depth (see Remark 2.2 of Liu and Singh (1993)).

Proposition 1.9.1 *For $P \in \mathcal{P}_d^{(2,i)}$ it holds that $\sup_{x\in\mathbb{R}^d} |\tilde{D}_M(x, P) - \tilde{D}_M(x, P_n| \xrightarrow[n\to\infty]{a.s.} 0$.*

Proof of Proposition 1.9.1. We can assume without loss of generality (w.l.o.g.) that $\det(\Sigma(P_n)) > 0$[34] implying that $d^2_{\Sigma(P_n)}(x, \mu(P_n))$ and $\tilde{D}_M(x, P_n)$ are well-defined. We need to show that, for all $\epsilon > 0$, $\mathbb{P}(\limsup_{n\to\infty} A_{n,\epsilon}) = 0$, where $A_{n,\epsilon} = \{\omega \in \Omega : \sup_{x\in\mathbb{R}^d} |\tilde{D}_M(x, P) - \tilde{D}_M(x, P_n)(\omega)| > \epsilon\}$. We first notice that, for all $r > 0$, $A_{n,\epsilon} \subset B_{n,\frac{\epsilon}{3},r} \cup C_{n,\frac{\epsilon}{3},r} \cup D_{\frac{\epsilon}{3},r}$, where

$$\begin{aligned}
B_{n,\frac{\epsilon}{3},r} &= \{\omega \in \Omega : \sup_{x\in B_r(0)} |\tilde{D}_M(x, P) - \tilde{D}_M(x, P_n)(\omega)| > \frac{\epsilon}{3}\}, \\
C_{n,\frac{\epsilon}{3},r} &= \{\omega \in \Omega : \sup_{x\in\mathbb{R}^d \setminus B_r(0)} \tilde{D}_M(x, P_n)(\omega) > \frac{\epsilon}{3}\}, \\
D_{\frac{\epsilon}{3},r} &= \{\omega \in \Omega : \sup_{x\in\mathbb{R}^d \setminus B_r(0)} \tilde{D}_M(x, P) > \frac{\epsilon}{3}\}.
\end{aligned}$$

Now, by Proposition 1.8.2, there exists $r_1 > 0$ such that $D_{\frac{\epsilon}{3},r} = \emptyset$ for all $r \geq r_1$. By Lemma B.1 (iv) we have that, for $r \geq r_1$,

$$\begin{aligned}
\mathbb{P}(\limsup_{n\to\infty}(B_{n,\frac{\epsilon}{3},r} \cup C_{n,\frac{\epsilon}{3},r})) &= \mathbb{P}((\limsup_{n\to\infty} B_{n,\frac{\epsilon}{3},r}) \cup (\limsup_{n\to\infty} C_{n,\frac{\epsilon}{3},r}))) \\
&\leq \mathbb{P}(\limsup_{n\to\infty} B_{n,\frac{\epsilon}{3},r}) + \mathbb{P}(\limsup_{n\to\infty} C_{n,\frac{\epsilon}{3},r}).
\end{aligned}$$

Thus, it is enough to show that, for some $r_2 \geq r_1$, $\mathbb{P}(\limsup_{n\to\infty} B_{n,\frac{\epsilon}{3},r_2}) = 0$ and $\mathbb{P}(\limsup_{n\to\infty} C_{n,\frac{\epsilon}{3},r_2}) = 0$. To this end, we notice that, by the strong law of large

[34] Notice that by (1.9.2) below $\{\omega \in \Omega : \det(\Sigma(P_n)(\omega)) = 0\} \subset F_{n,\delta}$ and by (1.9.1) $\mathbb{P}(\limsup_{n\to\infty} F_{n,\delta}) = 0$.

numbers, $\mu(P_n) \xrightarrow{a.s.} \mu(P)$ and

$$\Sigma(P_n) = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top - (\mu(P_n))(\mu(P_n))^\top \xrightarrow{a.s.} \int xx^\top dP(x) - (\mu(P))(\mu(P))^\top = \Sigma(P).$$

By continuous mapping theorem, we also have that $(\Sigma(P_n))^{-1} \xrightarrow{a.s.} (\Sigma(P))^{-1}$.[35] Now, for $\delta > 0$, let $E_{n,\delta} = \{\omega \in \Omega : \|\mu(P_n)(\omega) - \mu(P)\|_2 > \delta\}$ and $F_{n,\delta} = \{\omega \in \Omega : \|(\Sigma(P_n))^{-1}(\omega) - (\Sigma(P))^{-1}\|_{\mathcal{M},2} > \delta\}$, where $\|M\|_{\mathcal{M},q} = \sup_{y\in\mathbb{R}^d\setminus\{0\}} \frac{\|My\|_q}{\|y\|_q}$ is the L^q-matrix-norm of a $d \times d$ matrix M. By the above results, we have that, for all $\delta > 0$,

$$\mathbb{P}(\limsup_{n\to\infty}(E_{n,\delta} \cup F_{n,\delta})) \leq \mathbb{P}(\limsup_{n\to\infty} E_{n,\delta}) + \mathbb{P}(\limsup_{n\to\infty} F_{n,\delta}) = 0. \tag{1.9.1}$$

Let $\hat{\sigma}(M) = \{\lambda_1(M), \dots, \lambda_d(M)\}$ be the spectrum of a diagonalizable matrix M.[36] Since $\Sigma(P_n)$ and $\Sigma(P)$ are symmetric, they are diagonalizable by orthogonal matrices. By Bauer–Fike theorem (Theorem 2.1 in Eisenstat and Ipsen (1998)), for all $j = 1, \dots, d$, we have that

$$\min_{i=1,\dots,p} |(\lambda_j(\Sigma(P_n)))^{-1} - (\lambda_i(\Sigma(P)))^{-1}| \leq \|(\Sigma(P_n))^{-1} - (\Sigma(P))^{-1}\|_{\mathcal{M},2}. \tag{1.9.2}$$

Now, let $2\delta = \min_{i=1,\dots,d}(\lambda_i(\Sigma(P)))^{-1}$. There exist orthogonal matrices U_n such that $\Sigma(P_n) = U_n^{-1}D(\Sigma(P_n))U_n$, where $D(\Sigma(P_n))$ is the diagonal matrix with the eigenvalues $\lambda_1(\Sigma(P_n)), \dots, \lambda_d(\Sigma(P_n))$ on the diagonal. Now, if $\|(\Sigma(P_n))^{-1} - (\Sigma(P))^{-1}\|_{\mathcal{M},2} \leq \delta$, then $\min_{j=1,\dots,d}(\lambda_j(\Sigma(P_n)))^{-1} \geq \delta$,[37] yielding that

$$\begin{aligned} d^2_{\Sigma(P_n)}(x, \mu(P_n)) &= (x - \mu(P_n))^\top (\Sigma(P_n))^{-1}(x - \mu(P_n))^\top \\ &= (U_n(x - \mu(P_n)))^\top (D(\Sigma(P_n)))^{-1}(U_n(x - \mu(P_n))) \\ &\geq \delta \|U_n(x - \mu(P_n))\|_2^2 \\ &= \delta \|x - \mu(P_n)\|_2^2. \end{aligned}$$

If also $\|\mu(P_n) - \mu(P)\|_2 \leq \delta$ and $\|x\|_2 \geq \|\mu(P)\|_2 + \delta$, then

$$d^2_{\Sigma(P_n)}(x, \mu(P_n)) \geq \delta(\|x\|_2 - \|\mu(P_n)\|_2)^2 \geq \delta(\|x\|_2 - \delta - \|\mu(P)\|_2)^2. \tag{1.9.3}$$

Using that $\lim_{t\to\infty} g(t) = 0$, let $r_2 > 0$ be larger that r_1 and $\|\mu(P)\|_2 + \delta$, and such that $g(\delta(r_2 - \delta - \|\mu(P)\|_2)^2) \leq \frac{\epsilon}{3}$. Then, using (1.9.3), we have that $C_{n,\frac{\epsilon}{3},r_2} \subset E_{n,\delta} \cup F_{n,\delta}$ and, by (1.9.1), $\mathbb{P}(\limsup_{n\to\infty} C_{n,\frac{\epsilon}{3},r_2}) = 0$.

We are left to show that $\mathbb{P}(\limsup_{n\to\infty} B_{n,\frac{\epsilon}{3},r_2}) = 0$. Since g is uniformly continuous (because it is bounded, continuous, and decreasing), there exists $\eta > 0$ such that, for all $s, t \geq 0$ with $|s - t| \leq \eta$, $|g(s) - g(t)| \leq \frac{\epsilon}{3}$, yielding that $B_{n,\frac{\epsilon}{3},r_2} \subset G_{n,\eta,r_2}$, where

$$G_{n,\eta,r_2} = \{\omega \in \Omega : \sup_{x\in B_{r_2}(0)} |d^2_{\Sigma(P)}(x, \mu(P)) - d^2_{\Sigma(P_n)}(x, \mu(P_n))| > \eta\}.$$

[35] Here it is used that the function sending invertible matrices to their inverse is continuous. This follows from the continuity of the functions sending a matrix to its determinant and adjugate, respectively.

[36] That is, the set of its eigenvalues.

[37] This ensures that the covariance matrix $\Sigma(P_n)$ is invertible and the distance $d_{\Sigma(P_n)}$ is well-defined.

By adding and subtracting $d^2_{\Sigma(P)}(x, \mu(P_n))$, we see that $|d^2_{\Sigma(P)}(x, \mu(P)) - d^2_{\Sigma(P_n)}(x, \mu(P_n))|$ is bounded above by

$$|d^2_{\Sigma(P)}(x, \mu(P)) - d^2_{\Sigma(P)}(x, \mu(P_n))| + |d^2_{\Sigma(P)}(x, \mu(P_n)) - d^2_{\Sigma(P_n)}(x, \mu(P_n))|,$$

where the first term is[38]

$$\begin{aligned} &|\|x - \mu(P)\|^2_{\Sigma(P)} - \|x - \mu(P_n)\|^2_{\Sigma(P)}| \\ \leq &\|\mu(P) - \mu(P_n)\|_{\Sigma(P)}(\|x - \mu(P)\|_{\Sigma(P)} + \|x - \mu(P_n)\|_{\Sigma(P)}) \end{aligned}$$

and the second term is

$$\begin{aligned} &|(x - \mu(P_n))^\top((\Sigma(P))^{-1} - (\Sigma(P_n))^{-1})(x - \mu(P_n))| \\ \leq &\|x - \mu(P_n)\|_2 \|((\Sigma(P))^{-1} - (\Sigma(P_n))^{-1})(x - \mu(P_n))\|_2 \\ \leq &\|x - \mu(P_n)\|^2_2 \|(\Sigma(P))^{-1} - (\Sigma(P_n))^{-1}\|_{\mathcal{M},2} \end{aligned}$$

Using $\|y\|_{\Sigma(P)} \leq c\|y\|_2$ for some $c > 0$ and all $y \in \mathbb{R}^d$,[39] $\|x - \mu(P_n)\|_2 \leq \|x - \mu(P)\|_2 + \|\mu(P) - \mu(P_n)\|_2$, and $\|x\|_2 < r_2$ in the above inequalities, we see that there exists $\delta > 0$ such that, if $\|\mu(P_n) - \mu(P)\|_2 \leq \delta$ and $\|(\Sigma(P_n))^{-1} - (\Sigma(P))^{-1}\|_{\mathcal{M},2} \leq \delta$, then

$$\sup_{x \in B_{r_2}(0)} |d^2_{\Sigma(P)}(x, \mu(P)) - d^2_{\Sigma(P_n)}(x, \mu(P_n))| \leq \eta.$$

As before, it follows that $G_{n,\eta,r_2} \subset E_{n,\delta} \cup F_{n,\delta}$, yielding $\mathbb{P}(\limsup_{n\to\infty} G_{n,\eta,r_2}) = 0$. ∎

We now turn to the (uniform) consistency of *Type A* and *Type B* depth functions. To this end, notice that they are given by (a function of)

$$U_k(h, P) = \int h(x_1, \ldots, x_k) dP(x_1) \ldots dP(x_k),$$

where $k \geq 1$ and $h : (\mathbb{R}^d)^k \to \mathbb{R}$ is Borel measurable and bounded below. A natural estimate for $U_k(h, P)$ is

$$U_k(h, P_n) = \int h(x_1, \ldots, x_k) dP_n(x_1) \ldots dP_n(x_k) = \frac{1}{n^k} \sum_{i_1,\ldots,i_k=1}^{n} h(X_{i_1}, \ldots, X_{i_k}).$$

However, when $k > 1$, the appearance of the same random variable multiple times in the function h can be detrimental. Therefore, $U_k(h, P_n)$ is replaced by

$$U_{k,n}(h, P) = \frac{(n-k)!}{n!} \sum_{\substack{i_1,\ldots,i_k=1 \\ i_1,\ldots,i_k \text{ all different}}}^{n} h_{G,x,\infty}(X_{i_1}, \ldots, X_{i_k}).$$

Now, using the independence of $X_1, \ldots, X_n$, we see that $U_{k,n}(h, P)$ is an unbiased estimator of $U_k(h, P)$, thus, a U-statistics for the estimation of $U_k(h, P)$.[40] We study U-statistics in Section 1.10. We notice that if h is bounded, as it is the case for *Type A* depth functions, then $U_k(h, P_n)$ and $U_{k,n}(h, P)$ are asymptotically equivalent because

[38] See 27.

[39] This is because all norms on $\mathbb{R}^d$ are equivalent.

[40] $U_k(h, P_n)$ is called a V-statistics for the estimation of $U_k(h, P)$.

$\frac{n!}{n^k(n-k)!} = 1 + O(\frac{1}{n})$ and

$$|U_k(h, P_n) - \frac{n!}{n^k(n-k)!} U_{k,n}(h, P)| \leq \left(1 - \frac{n!}{n^k(n-k)!}\right) l,$$

where $0 \leq l < \infty$ is the constant bounding $|h|$. Thus, results for $U_{k,n}(h, P)$ carry over to $U_k(h, P_n)$. Uniform consistency results for a class $\mathcal{H}$ of Borel measurable functions $h : (\mathbb{R}^d)^k \to \mathbb{R}$, such as,

$$\sup_{h \in \mathcal{H}} |U_{k,n}(h, P) - U_k(h, P)| \xrightarrow{a.s.} 0, \tag{1.9.4}$$

are studied in Section 1.11. For $k = k_G$ and $\mathcal{H} = \{h_{G,x,\infty} : x \in \mathbb{R}^d\}$, (1.9.4) entails uniform consistency for *Type A* depth functions (see also Theorem 2.4.1 in Chapter 2). Similarly, for $\mathcal{H} = \{i_{G,x,\infty} : x \in \mathbb{R}^d\}$, using that g in Definition 1.7.1 is uniformly continuous over $[0, \infty)$, (1.9.4) gives uniform consistency for *Type B* depth functions with $\sup_{x \in \mathbb{R}^d} |U_{k_G}(i_{G,x,\infty}, P)| < \infty$.

1.10 U-statistics

In this section we study U-statistics on a Hausdorff[41] topological space E (Hoeffding, 1961; Hoeffding, 1963; Arcones and Giné, 1993; Arcones, 1995; Serfling, 2009; Korolyuk and Borovskich, 2013). For *Type A* and *Type B* depth functions as well as local depth functions (see Chapter 2) we take $E = \mathbb{R}^d$. We begin with some notation. For $k \geq 1$, we denote by $\mathscr{H}_k$ the space of all Borel measurable functions $h : E^k \to \mathbb{R}$ and by $\mathcal{M}_\pm(E^k)$ the set of all finite signed Borel measures on E^k.[42,43] We will need to integrate functions $h \in \mathscr{H}_k$ w.r.t. signed measures $Q \in \mathcal{M}_\pm(E^k)$. To this end, we let

$$\mathscr{D}_{E^k} = \{(h, Q) \in \mathscr{H}_k \times \mathcal{M}_\pm(E^k) : \int h(x_1, \ldots, x_k) dQ(x_1, \ldots, x_k) \text{ exists and is finite}\}$$[44]

[41] A topological space E is Hausdorff if for all $x, y \in E$, $x \neq y$, there are (open) neighborhoods U_x and U_y of x and y that are disjoint, i.e. $U_x \cap U_y = \emptyset$. It follows that all points $\{x\} \subset E = \{x\} \cup \cup_{y \in E, y \neq x} U_y$ are closed. In particular, they are Borel sets (cf. Harley and McNulty (1979)) and $\delta_x \neq \delta_y$ for all $x, y \in E$, $x \neq y$, which prevents degeneracy of empirical measures.

[42] $\mathcal{M}_\pm(E^k) = \{P_+ - P_- : P_+, P_- \in \mathcal{M}_f(E^k)\}$, where $\mathcal{M}_f(E^k) \subset \mathcal{M}(E^k)$ is the set of all finite Borel measures on E^k.

[43] A natural question is whether the equality $(\mathcal{B}(E))^{\otimes k} = \mathcal{B}(E^k)$ holds. We show that $(\mathcal{B}(E))^{\otimes k} \subset \mathcal{B}(E^k)$. In particular, this implies that $P^k \in \mathcal{P}(E^k)$ for all $P \in \mathcal{P}(E)$. Moreover, $(\mathcal{B}(E))^{\otimes k} = \mathcal{B}(E^k)$ provided that E is second countable, that is, there are countable many open set $\{U_i\}_{i=1}^\infty$ in E such that every open set $V \subset E$ can be written as the union of elements from $\{U_i\}_{i=1}^\infty$. More generally, for topological spaces $\{E_i\}_{i=1}^\infty$ we have that $\otimes_{i=1}^n \mathcal{B}(E_i) \subset \mathcal{B}(\prod_{i=1}^n E_i)$ for all $n \in \mathbb{N} \cup \{\infty\}$. Indeed, by definition, $\otimes_{i=1}^n \mathcal{B}(E_i)$ is the smallest σ-algebra including $\{\prod_{i=1}^n B_i : B_i \in \mathcal{B}(E_i)\}$, whereas $\mathcal{B}(\prod_{i=1}^n E_i)$ is a σ-algebra including $\{\prod_{i=1}^n B_i : B_i \in \mathcal{B}(E_i)\}$. In particular, this implies that $\prod_{i=1}^n \mathrm{m}_i \in \mathcal{M}(E^k)$ for all $\mathrm{m}_1, \ldots, \mathrm{m}_n \in \mathcal{M}(E)$. Suppose now that $\{E_i\}_{i=1}^\infty$ are second countable. Using the proof of Lemma 1.2 of Kallenberg (1997), we see that $\otimes_{i=1}^n \mathcal{B}(E_i) = \mathcal{B}(\prod_{i=1}^n E_i)$ for all $n \in \mathbb{N} \cup \{\infty\}$. Next, we notice that a second countable topological space E is separable (see Proposition 4.5 of Folland (1999)), that is, there exists $\{x_n\}_{n=1}^\infty$ in E such that $U \cap \{x_n\}_{n=1}^\infty \neq \emptyset$ for every non-empty open set $U \subset E$. Finally, a metric space is second countable if and only if it is separable (see Proposition 2.1.4 of Dudley (2018)).

and define the function $\mathcal{J}_k : \mathscr{D}_{E^k} \to \mathbb{R}$ by

$$\mathcal{J}_k(h, Q) = \int h(x_1, \ldots, x_k) dQ(x_1, \ldots, x_k).$$

We will also need some moment assumptions on the function h. For $Q \in \mathcal{P}(E^k)$ and $q \geq 1$, let

$$\mathscr{L}^q(\mathscr{H}_k, Q) = \{h \in \mathscr{H}_k : \|h\|_{\mathscr{L}^q(\mathscr{H}_k,Q)} < \infty\}, \text{ where } \|h\|_{\mathscr{L}^q(\mathscr{H}_k,Q)} = (J_k(|h|^q, Q))^{1/q},$$

be the subclass of $\mathscr{H}_k$ of functions with finite q^{th}-moment w.r.t. the distribution Q. The seminorm $\|\cdot\|_{\mathscr{L}^q(\mathscr{H}_k,Q)}$ induces a pseudometric[45] $\tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,Q)}$ on $\mathscr{L}^q(\mathscr{H}_k, Q)$ via $\tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,Q)}(h_1, h_2) = \|h_1 - h_2\|_{\mathscr{L}^q(\mathscr{H}_k,Q)}$. Notice that $\mathscr{L}^1(\mathscr{H}_k, Q)$ is precisely the subspace of $\mathscr{H}_k$ where the integral $\mathcal{J}_k(h, Q)$ is well-defined and finite, that is, for $Q \in \mathcal{P}(E^k)$, $\mathscr{L}^1(\mathscr{H}_k, Q) = \{h \in \mathscr{H}_k : (h, Q) \in \mathscr{D}_{E^k}\}$.[46] We are now ready to give a precise definition of U-statistics.

Definition 1.10.1 *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with probability distribution $P \in \mathcal{P}(E)$ and $h \in \mathscr{L}^1(\mathscr{H}_k, P^k)$. A U-statistics for the estimation of $U_k(h, P) = \mathcal{J}_k(h, P^k)$ is $U_{k,n}(h, P) = \mathcal{J}_k(h, \hat{P}_n^k)$, where*

$$\hat{P}_n^k = \frac{(n-k)!}{n!} \sum_{\substack{i_1,\ldots,i_k=1 \\ i_1,\ldots,i_k \text{ all different}}}^{n} \prod_{j=1}^{k} \delta_{X_{i_j}}.$$

The function h is the kernel of the U-statistics and k is its order.

Thus, when estimating the integral $\mathcal{J}_k(h, P^k)$, the product measure P^k is replaced by the empirical measure $\hat{P}_n^k$.

Remark 1.10.1 *In the following, we will restrict our attention to the class of Borel measurable functions that are symmetric w.r.t. their arguments. There is no loss of generality in doing so since every function is associated to a unique symmetric function via the map $\tilde{\sigma} : \mathscr{H}_k \to \mathscr{H}_k$ [47] given by*

$$\tilde{\sigma}(h)(x_1, \ldots, x_k) = \frac{1}{k!} \sum_{(i_1,\ldots,i_k) \in \mathrm{P}_k} h(x_{i_1}, \ldots, x_{i_k}),\text{[48]}$$

where P_k is the set of all permutations of $(1, \ldots, k)$. From now on, we use the notation $\mathscr{H}_k$ for the class of Borel measurable, symmetric functions $h : E^k \to \mathbb{R}$. This implies a simplification in the empirical measure $\hat{P}_n^k$. Indeed, since the order of the elements in $(X_{i_1}, \ldots, X_{i_k})$

[44] Write $Q = Q_+ - Q_-$, where $Q_+, Q_- \in \mathcal{M}_f(E^k)$, and $h = h_+ - h_-$, where $h_+ = \max(h, 0)$ and $h_- = \max(-h, 0)$. The condition $(h, Q) \in \mathscr{D}_{E^k}$ is equivalent to the fact that $\int h_{*_1}(x_1, \ldots, x_k) dQ_{*_2}(x_1, \ldots, x_k)$ exist and are finite for all $*_1, *_2 \in \{+, -\}$ and is automatically satisfied if the function h is bounded.

[45] $\tilde{d}_F : F \times F \to [0, \infty)$ is a pseudometric on F if, for all $x, y, z \in F$, (i) $\tilde{d}_F(x, x) = 0$, (ii) $\tilde{d}_F(x, y) = \tilde{d}_F(y, x)$ for all $x, y \in F$, and (iii) $\tilde{d}_F(x, z) \leq \tilde{d}_F(x, y) + \tilde{d}_F(y, z)$ for all $x, y, z \in F$. In this case, the pair $(F, \tilde{d}_F)$ is said to be a pseudometric space.

[46] Use [44] with $Q = Q^+$.

[47] Notice that if $h : E^k \to \mathbb{R}$ is Borel measurable, then also $\tilde{\sigma}(h)$ is Borel measurable, yielding $\tilde{\sigma}(\mathscr{H}_k) = \{\tilde{\sigma}(h) : h \in \mathscr{H}_k\} \subset \mathscr{H}_k$.

[48] Using that P^k is a product measure and the linearity of the integral we see that $\tilde{\sigma}(h) \in \mathscr{L}^1(\mathscr{H}_k, P^k)$ if and only if $h \in \mathscr{L}^1(\mathscr{H}_k, P^k)$.

does not matter, we take

$$\hat{P}_n^k = \binom{n}{k}^{-1} \sum_{1\le i_1<\cdots<i_k\le n} \prod_{j=1}^{k} \delta_{X_{i_j}}.$$

In this section, we derive the law of large numbers (LLN) and central limit theorem (CLT) for U-statistics. To this end, we use Hoeffding decomposition of U-statistics, which we now state (see Lemma 1 of Hoeffding (1961) and (1.8) of Arcones and Giné (1993)).

Proposition 1.10.1 *For $j = 1, \dots, k$, define on E^k the product measure, $P_{k,j}(x_1, \dots, x_j) = \prod_{i=1}^{j}(\delta_{x_i} - P) \times P^{k-j}$, where $x_1, \dots, x_j \in E$ and $P \in \mathcal{P}(E)$, and let $p_{k,j} : \mathcal{L}^1(\mathcal{H}_k, P^k) \to \mathcal{L}^1(\mathcal{H}_j, P^j)$ be given by*

$$(p_{k,j}h)(x_1, \dots, x_j) = \mathcal{J}_k(h, P_{k,j}(x_1, \dots, x_j)).$$

Then, for all $h \in \mathcal{L}^1(\mathcal{H}_k, P^k)$,

$$\mathcal{J}_k(h, \hat{P}_n^k) = \mathcal{J}_k(h, P^k) + \sum_{j=1}^{k} \binom{k}{j} \mathcal{J}_j(p_{k,j}h, \hat{P}_n^j).$$

Notice that the terms $\mathcal{J}_j(p_{k,j}h, \hat{P}_n^j) = U_{j,n}(p_{k,j}h, P)$ are U-statistics with kernel function $p_{k,j}h$ and order j. Thus, Proposition 1.10.1 shows that

$$U_{k,n}(h, P) = U_k(h, P) + \sum_{j=1}^{k} \binom{k}{j} U_{j,n}(p_{k,j}h, P).$$

Proof of Proposition 1.10.1. We show that, for $s = 1, \dots, k$ and $h \in \mathcal{L}^1(\mathcal{H}_k, P^k)$,

$$\begin{aligned} \mathcal{J}_k(h, P_{k,s}(x_1, \dots, x_s)) &= \mathcal{J}_k(h, \prod_{i=1}^{s} \delta_{x_i} \times P^{k-s} - P^k) \\ &\quad - \sum_{j=1}^{s-1} \sum_{1\le l_1<\cdots<l_j\le s} \mathcal{J}_k(h, P_{k,j}(x_{l_1}, \dots, x_{l_j})). \end{aligned} \tag{1.10.1}$$

It follows from (1.10.1) with $s = k$ that

$$\mathcal{J}_k(h, \prod_{i=1}^{k} \delta_{x_i}) = \mathcal{J}_k(h, P^k) + \sum_{j=1}^{k} \sum_{1\le l_1<\cdots<l_j\le k} \mathcal{J}_k(h, P_{k,j}(x_{l_1}, \dots, x_{l_j}))$$

implying that

$$\begin{aligned} \mathcal{J}_k(h, \hat{P}_n^k) &= \mathcal{J}_k(h, P^k) + \sum_{j=1}^{k} \binom{n}{k}^{-1} \sum_{1\le i_1<\cdots<i_k\le n} \sum_{\substack{1\le l_1<\cdots<l_j\le k: \\ l_1,\dots,l_j\in\{i_1,\dots,i_k\}}} \mathcal{J}_k(h, P_{k,j}(X_{l_1}, \dots, X_{l_j})) \\ &= \mathcal{J}_k(h, P^k) + \sum_{j=1}^{k} \binom{k}{j} \binom{n}{j}^{-1} \sum_{1\le i_1<\cdots<i_j\le n} \mathcal{J}_k(h, P_{k,j}(X_{i_1}, \dots, X_{i_j})). \end{aligned}$$

We now show (1.10.1). Equality holds for all $k \in \mathbb{N}$ and $s = 1$. We show that if it holds for s and all $k \ge s$ then it holds for $s + 1$ and all $k \ge s + 1$. To this end, let

$1 \leq s \leq k-1$ be fixed. Using the symmetry of $h \in \mathscr{L}^1(\mathscr{H}_k, P^k)$ and (1.10.1) for s and $k-1$, we have that

$$\begin{aligned}\mathcal{J}_k(h, P_{k,s+1}(x_1,\dots,x_{s+1})) &= \mathcal{J}_k(h, (\delta_{x_{s+1}} - P) \times P_{k-1,s}(x_1,\dots,x_s)) \\ &= \mathcal{J}_k(h, (\delta_{x_{s+1}} - P) \times (\prod_{i=1}^{s} \delta_{x_i} \times P^{k-1-s} - P^{k-1})) \\ &- \sum_{j=1}^{s-1} \sum_{1 \leq l_1 < \cdots < l_j \leq s} \mathcal{J}_k(h, (\delta_{x_{s+1}} - P) \times P_{k-1,j}(x_{l_1},\dots,x_{l_j})).\end{aligned}$$

The first term in the last equation is equal to

$$\mathcal{J}_k(h, \prod_{i=1}^{s+1} \delta_{x_i} \times P^{k-(s+1)} - P^k) - \mathcal{J}_k(h, \prod_{i=1}^{s} \delta_{x_i} \times P^{k-s} - P^k) - \mathcal{J}_k(h, P_{k,1}(x_{s+1})),$$

where the term in the middle is

$$-\mathcal{J}_k(h, \prod_{i=1}^{s} \delta_{x_i} \times P^{k-s} - P^k) = -\sum_{j=1}^{s} \sum_{1 \leq l_1 < \cdots < l_j \leq s} \mathcal{J}_k(h, P_{k,j}(x_{l_1},\dots,x_{l_j})).$$

Using symmetry we see that the second term is equal to

$$-\sum_{j=2}^{s} \sum_{1 \leq l_1 < \cdots < l_{j-1} \leq s} \mathcal{J}_k(h, P_{k,j+1}(x_{l_1},\dots,x_{l_j}, x_{s+1})).$$

(1.10.1) for $s+1$ and k follows. ■

We now turn to the definition of rank of a U-statistics.

Definition 1.10.2 *Let $P \in \mathcal{P}(E)$ and $0 \neq h \in \mathscr{L}^1(\mathscr{H}_k, P^k)$. The rank of the U-statistics $U_{k,n}(h, P)$ is*

$$r = \min(\{j \in \{1,\dots,k\} : p_{k,j}h \text{ is not equal to } 0 \text{ a.s.}\} \cup \{k\}).\text{[49]}$$

$U_{k,n}(h, P)$ is non-degenerate if $r = 1$ and degenerate if $r > 1$. If $r = k$, then $U_{k,n}(h, P)$ is completely degenerate.

Thus, if $U_{k,n}(h, P)$ has rank r, then Proposition 1.10.1 yields

$$U_{k,n}(h, P) = U_k(h, P) + \sum_{j=r}^{k} \binom{k}{j} U_{j,n}(p_{k,j}h, P) \text{ a.s.} \tag{1.10.2}$$

Notice that the U-statistics $U_{j,n}(p_{k,j}h, P)$ are completely degenerate since $p_{j,i}p_{k,j}h = 0$ a.s. for all $i = 1,\dots,j-1$. To see this, use that $P_{k,j}(x_1,\dots,x_j) = \prod_{i=1}^{j}(\delta_{x_i} - P) \times P^{k-j}$ is a product measure and for all Borel measurable functions $g : E \to \mathbb{R}$,

$$\int \left[\int g(y) d(\delta_{x_i} - P)(y)\right] dP(x_i) = \int g(x_i) dP(x_i) - \int g(y) dP(y) = 0. \tag{1.10.3}$$

Hoeffding decomposition can be used to compute the (asymptotic) variance of $U_{k,n}(h, P)$ (see Lemma B, page 184, of Serfling (2009)).

[49] Equivalently, $r = \min(\{j \in \{1,\dots,k\} : \mathbb{V}ar[p_{k,j}h(X_1,\dots,X_j)] > 0\} \cup \{k\})$.

Proposition 1.10.2 *For all $P \in \mathcal{P}(E)$ and $h \in \mathscr{L}^2(\mathscr{H}_k, P^k)$,*

$$\begin{aligned}\mathbb{V}ar[U_{k,n}(h,P)] &= \sum_{j=r}^{k} \binom{k}{j}^2 \binom{n}{j}^{-1} \mathbb{V}ar[p_{k,j}h(X_1,\dots,X_j)] \\ &= \binom{k}{r}^2 r! n^{-r} \mathbb{V}ar[p_{k,r}h(X_1,\dots,X_r)] + O(n^{-(r+1)})^{50}.\end{aligned}$$

Proof of Proposition 1.10.2. Using (1.10.2) we have that

$$\begin{aligned}\mathbb{V}ar[U_{k,n}(h,P)] &= \mathbb{E}[(U_{k,n}(h,P) - U_k(h,P))^2] \\ &= \sum_{j=r}^{k}\sum_{l=r}^{k} \binom{k}{j}\binom{k}{l} \mathbb{E}[U_{j,n}(p_{k,j}h,P)U_{l,n}(p_{k,l}h,P)].\end{aligned}$$

Next, using (1.10.3) we see that $\mathbb{E}[U_{j,n}(p_{k,j}h,P)U_{l,n}(p_{k,l}h,P)] = 0$ for $j \neq l$, yielding

$$\mathbb{V}ar[U_{k,n}(h,P)] = \sum_{j=r}^{k} \binom{k}{j}^2 \mathbb{V}ar[U_{j,n}(p_{k,j}h,P)].$$

Using again (1.10.3) we have that

$$\begin{aligned}\mathbb{V}ar[U_{j,n}(p_{k,j}h,P)] &= \binom{n}{j}^{-2} \mathbb{E}\left[\left(\sum_{1\le i_1<\dots<i_j\le k} p_{k,j}h(X_{i_1},\dots,X_{i_j})\right)^2\right] \\ &= \binom{n}{j}^{-2} \sum_{1\le i_1<\dots<i_j\le k} \mathbb{E}\left[\left(p_{k,j}h(X_{i_1},\dots,X_{i_j})\right)^2\right] \\ &= \binom{n}{j}^{-1} \mathbb{E}\left[\left(p_{k,j}h(X_1,\dots,X_j)\right)^2\right],\end{aligned}$$

which gives the first equality. For the second equality, notice that $\binom{n}{j}^{-1} = O(n^{-j})$ for all $j = r+1,\dots,k$ and

$$\frac{r!}{n^r} \le \binom{n}{r}^{-1} = \frac{r!}{n^r}\frac{n^r}{n\dots(n-r+1)} \le \frac{r!}{n^r}\left(1+\frac{r-1}{n}\right)^{r-1} = \frac{r!}{n^r} + O(n^{-(r+1)}),$$

where the last equality follows from binomial theorem. ■

We are now ready to state the central limit theorem for non-degenerate U-statistics (Hoeffding, 1948). The degenerate case is more complicated and it is studied in Section 4.4 of Korolyuk and Borovskich (2013).

Proposition 1.10.3 *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with probability distribution $P \in \mathcal{P}(E)$ and $h \in \mathscr{L}^2(\mathscr{H}_k, P^k)$ with $\mathbb{V}ar[p_{k,1}h(X_1)] > 0$. Then,*

$$\sqrt{n}(U_{k,n}(h,P) - U_k(h,P)) \xrightarrow{d} (W(P))(h),$$

where $(W(P))(h)$ is normally distributed with mean zero and variance $k^2\mathbb{V}ar[p_{k,1}h(X_1)]$.

[50] $g(n) = O(f(n))$ if there exists $M, N > 0$ such that $|g(n)| \le M|f(n)|$ for all $n \ge N$.

Proof of Proposition 1.10.3. By Hoeffding decomposition (Proposition 1.10.1), we have that

$$U_{k,n}(h,P) - U_k(h,P) = \frac{k}{n}\sum_{i=1}^{n} p_{k,1}h(X_i) + \sum_{j=2}^{k}\binom{k}{j} r_{n,j},$$

where

$$r_{n,j} = \mathcal{J}_j(p_{k,j}h, \hat{P}_n^j) = \binom{n}{j}^{-1} \sum_{1\le i_1<\cdots<i_j\le n} p_{k,j}h(X_{i_1},\ldots,X_{i_j}).$$

We show that $\sqrt{n}r_{n,j}$ converges in probability to zero. First notice that by Markov inequality for all $\epsilon > 0$

$$\mathbb{P}(\sqrt{n}r_{n,j} \ge \epsilon)^{51} \le \frac{n}{\epsilon^2}\mathbb{E}[r_{n,j}^2]. \tag{1.10.4}$$

Next, using (1.10.3) we have that

$$\begin{aligned}
\mathbb{E}[r_{n,j}^2] &= \binom{n}{j}^{-2} \sum_{1\le i_1<\cdots<i_j\le n}\ \sum_{1\le l_1<\cdots<l_j\le n} \mathbb{E}[p_{k,j}h(X_{i_1},\ldots,X_{i_j})p_{k,j}h(X_{l_1},\ldots,X_{l_j})] \\
&= \binom{n}{j}^{-2} \sum_{1\le i_1<\cdots<i_j\le n} \mathbb{E}[(p_{k,j}h(X_{i_1},\ldots,X_{i_j}))^2] \\
&= \binom{n}{j}^{-1} \mathbb{E}[(p_{k,j}h(X_1,\ldots,X_j))^2].
\end{aligned}$$

Now, we apply Jensen inequality and, using that $h \in \mathscr{L}^2(\mathscr{H}_k, P^k)$, we obtain that

$$\mathbb{E}[(p_{k,j}h(X_1,\ldots,X_j))^2] \le \mathbb{E}[(h(X_1,\ldots,X_k))^2] < \infty,$$

Since $j \ge 2$, we conclude that $\sqrt{n}r_{n,j} \xrightarrow[n\to\infty]{p} 0$. Finally, we notice that $\frac{k}{n}\sum_{i=1}^{n} p_{k,1}h(X_i)$ is an average of i.i.d. random variables with mean 0 and variance $k^2\mathbb{V}ar[p_{k,1}h(X_1)]$. The result now follows from the central limit theorem for i.i.d. random variables. ∎

Remark 1.10.2 *By Slutsky theorem,*[52] *the CLT in Proposition 1.10.3 is equivalent to*

$$\sqrt{\frac{n}{k^2\mathbb{V}ar[p_{k,1}h(X_1)]}}(U_{k.n}(h,P) - U_k(h,P)) \xrightarrow{d} N(0,1).$$

Since by Proposition 1.10.2

$$\mathbb{V}ar[U_{k,n}(h,P)] = k^2\mathbb{V}ar[p_{k,1}h(X_1)]n^{-1} + O(n^{-2}),$$

another application of Slutsky theorem yields that the CLT is equivalent to

$$(\mathbb{V}ar[U_{k,n}(h,P)])^{-1/2}(U_{k,n}(h,P) - U_k(h,P)) \xrightarrow{d} N(0,1).$$

A straightforward consequence of the above CLT is the (weak) law of large numbers: $U_{k,n}(h,P) - U_k(h,P) \xrightarrow{p} 0$. Indeed, by Slutsky theorem $U_{k,n}(h,P) - U_k(h,P) \xrightarrow{d} 0$

[51] $\mathbb{P}(\sqrt{n}r_{n,j} \ge \epsilon)$ is a short notation for $\mathbb{P}(\{\omega \in \Omega : \sqrt{n}r_{n,j}(\omega) \ge \epsilon\})$. Using a common convention, here and in the following, the full definition of subsets of Ω is shortened.

[52] See for instance Theorem 1.3.6 and Example 1.4.7 of van der Vaart and Wellner (1996).

and convergence in distribution and in probability to a constant are equivalent (see e.g. Section 1.2). The strong law of large numbers was proved by Hoeffding (1961) using Hoeffding decomposition (Proposition 1.10.1). Of course, the conditions $h \in \mathscr{L}^2(\mathscr{H}_k, P^k)$ and $\mathbb{V}ar[p_{k,1}h(X_1)] > 0$ are not required.

Proposition 1.10.4 *For all $P \in \mathcal{P}(E)$ and $h \in \mathscr{L}^1(\mathscr{H}_k, P^k)$, $U_{k,n}(h,P) - U_k(h,P) \xrightarrow{a.s.} 0$.*

For a simplified proof, which makes use of the backward martingale convergence theorem, see Theorem 3.1.1 of Korolyuk and Borovskich (2013).

Next, we provide (without proof) some bounds on the deviations probabilities of U-statistics from their means. For $x \in \mathbb{R}$, $\lfloor x \rfloor$ is the greatest integer less than or equal to x.

Proposition 1.10.5 *Let $P \in \mathcal{P}(E)$ and $h \in \mathscr{H}_k$ be bounded (say $a \le h \le b$ and let $l = b - a$). Then, for all $n \ge k$ and $t > 0$,*

(i) $\mathbb{P}(|U_{k,n}(h,P) - U_k(h,P)| \ge t) \le 2\exp\left(-2\frac{\lfloor n/k \rfloor t^2}{l^2}\right)$,

(ii) $\mathbb{P}(|U_{k,n}(h,P) - U_k(h,P)| \ge t) \le 2\exp\left(-\frac{\lfloor n/k \rfloor t^2}{2\mathbb{V}ar[h(X_1,\dots,X_k)] + 2/3lt}\right)$.

If additionally $U_{k,n}(h,P)$ has rank $r = 1$, then

(iii) $\mathbb{P}\left(\sqrt{\frac{n}{k^2\mathbb{V}ar[p_{k,1}h(X_1)]}}|U_{k,n}(h,P) - U_k(h,P)| \ge t\sqrt{\log(n)}\right) = O\left(\frac{1}{\sqrt{\log(n)}n^{\frac{t^2}{2}}}\right)$,

(iv) $\mathbb{P}(\sqrt{n}|U_{k,n}(h,P) - U_k(h,P)| \ge t) \le 4\exp\left(-\frac{t^2}{2k^2\mathbb{V}ar[p_{k,1}h(X_1)] + c_n lt}\right)$,

where $c_n = \frac{2^{k+2}k^k}{\sqrt{n-1}} + \frac{2}{3k\sqrt{n}}$.

(i) and (ii) provide upper bounds for the probabilities of large deviations of the U-statistics $U_{k,n}(h,P)$ from its mean $U_k(h,P)$. Notice that (i) gives a better bound for large t.[53] These are two-sided versions[54] of Theorem 8.1.1 of Korolyuk and Borovskich (2013) and Theorem A, page 201, of Serfling (2009) (see also Hoeffding (1963)). (iii) is concerned with deviations of order $\sqrt{\log(n)}$ for the term

$$\sqrt{\frac{n}{k^2\mathbb{V}ar[p_{k,1}h(X_1)]}}(U_{k,n}(h,P) - U_k(h,P)),$$

which converges by Remark 1.10.2 to a standard normal distribution. Such deviations are called moderate. See Corollary 8.2.1 of Korolyuk and Borovskich (2013). Finally, (iv) is a Bernstein-type inequality for U-statistics and its given in Theorem 2 of Arcones (1995).
The next result gives an upper bound on the moments of a U-statistics (see (2.1.16) of Korolyuk and Borovskich (2013) and Theorem 1.5.1 of Lee (1990)).

Proposition 1.10.6 *Let $P \in \mathcal{P}(E)$ and $h \in \mathscr{L}^q(\mathscr{H}_k, P^k)$, where $q \ge 2$. Then, there is a constant $0 \le c < \infty$ depending on h and q such that for all $n \ge 2k$*

$$\mathbb{E}[|U_{k,n}(h,P) - U_k(h,P)|^q] \le cn^{-q/2}.$$

[53] This happens whenever $t > \frac{3}{l}(\frac{l^2}{4} - \mathbb{V}ar[h(X_1,\dots,X_k)])$.

[54] For (ii) it is used that both $U_k(h,P) - a$ and $b - U_k(h,P)$ are non-negative and bounded above by l.

Furthermore, if $|h| \leq l$ for some $0 \leq l < \infty$, then c can be chosen so that it depends on l and q only.

Proof of Proposition 1.10.6. Let $s = \lfloor n/k \rfloor$ be the largest integer less than or equal to n/k and define $\tilde{h}_n : E^n \to \mathbb{R}$ by $\tilde{h}_n(x_1, \dots, x_n) = \frac{1}{s} \sum_{i=0}^{s-1} h(x_{ki+1}, \dots, x_{ki+k})$. It holds that

$$\mathcal{J}_n(\tilde{h}_n, \frac{1}{n!} \sum_{(i_1,\dots,i_n)\in \mathrm{P}_n} \prod_{j=1}^{n} \delta_{X_{i_j}}) = \mathcal{J}_k(h, \hat{P}_n^k) = U_{k,n}(h, P).$$

Using Jensen's inequality we obtain that

$$\begin{aligned}|U_{k,n}(h,P) - U_k(h,P)|^q &= |\mathcal{J}_n(\tilde{h}_n - U_k(h,P), \frac{1}{n!} \sum_{(i_1,\dots,i_n)\in \mathrm{P}_n} \prod_{j=1}^{n} \delta_{X_{i_j}})|^q \\ &\leq \mathcal{J}_n(|\tilde{h}_n - U_k(h,P)|^q, \frac{1}{n!} \sum_{(i_1,\dots,i_n)\in \mathrm{P}_n} \prod_{j=1}^{n} \delta_{X_{i_j}}).\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}[|U_{k,n}(h,P) - U_k(h,P)|^q] &\leq \mathbb{E}[|\tilde{h}_n(X_1,\dots,X_n) - U_k(h,P)|^q] \\ &= s^{-q}\mathbb{E}[|\sum_{i=0}^{s-1}[h(X_{ki+1},\dots,X_{ki+k}) - U_k(h,P)]|^q].\end{aligned}$$

Next, we apply Theorem 10.3.2 of Chow and Teicher (1997) and obtain that

$$\mathbb{E}[|\sum_{i=0}^{s-1}[h(X_{ki+1},\dots,X_{ki+k}) - U_k(h,P)]|^q] \leq \tilde{c}_q \mathbb{E}[|\sum_{i=0}^{s-1}[h(X_{ki+1},\dots,X_{ki+k}) - U_k(h,P)]^2|^{q/2}]$$

for some constant $\tilde{c}_q$ depending on q only. Another application of Jensen's inequality yields that

$$|\sum_{i=0}^{s-1}[h(X_{ki+1},\dots,X_{ki+k}) - U_k(h,P)]^2|^{q/2} \leq s^{q/2}\frac{1}{s}\sum_{i=0}^{s-1}|h(X_{ki+1},\dots,X_{ki+k}) - U_k(h,P)|^q.$$

It follows that

$$\mathbb{E}[|U_{k,n}(h,P) - U_k(h,P)|^q] \leq s^{-q/2}\tilde{c}_q\mathbb{E}[|h(X_1,\dots,X_k) - U_k(h,P)|^q].$$

Since $s \geq n/k - 1$ and $n \geq 2k$, it holds that $s^{-q/2} \leq (n/k-1)^{-q/2} \leq (n/(2k))^{-q/2} = (2k)^{q/2}n^{-q/2}$. The result follows by taking $c = (2k)^{q/2}\tilde{c}_q\mathbb{E}[|h(X_1,\dots,X_k) - U_k(h,P)|^q]$. Finally, if $|h| \leq l$, then $\mathbb{E}[|h(X_1,\dots,X_k) - U_k(h,P)|^q] \leq (2l)^q$ and we can take $c = (2k)^{q/2}\tilde{c}_q(2l)^q$. ■

We conclude this section by generalizing the CLT in Proposition 1.10.3 to multidimensional U-statistics. To this end, let $\mathscr{H}_k^m$ be the space of all Borel measurable functions $h : E^k \to \mathbb{R}^m$ that are symmetric w.r.t. their arguments and

$$\tilde{\mathscr{D}}_{E^k}^m = \{(h,Q) \in \mathscr{H}_k^m \times \mathcal{M}_\pm(E^k) : \int h(x_1,\dots,x_k)dQ(x_1,\dots,x_k) \text{ exists and is finite}\},$$

where the integral of a vector is the vector of the integrals and it exists and is finite if and only if each of its components exists and is finite. Next, we define the function

$\tilde{\mathcal{J}}_k^m : \tilde{\mathscr{D}}_{E^k}^m \to \mathbb{R}$ by

$$\tilde{\mathcal{J}}_k^m(h, Q) = \int h(x_1, \ldots, x_k) dQ(x_1, \ldots, x_k).$$

The subclass of $\tilde{\mathscr{H}}_k^m$ of functions with finite q^{th}-moment w.r.t. $Q \in \mathcal{P}(E^k)$ is given by

$\mathscr{L}^q(\tilde{\mathscr{H}}_k^m, Q) = \{h \in \tilde{\mathscr{H}}_k^m : \|h\|_{\mathscr{L}^q(\tilde{\mathscr{H}}_k^m, Q)} < \infty\}$, where $\|h\|_{\mathscr{L}^q(\tilde{\mathscr{H}}_k^m, Q)} = (J_k(\|h\|_2^q, Q))^{1/q}$.

Then, for $P \in \mathcal{P}(E)$ and $h \in \mathscr{L}^1(\tilde{\mathscr{H}}_k^m, P^k)$, $\tilde{U}_{k,n}^m(h, P) = \tilde{\mathcal{J}}_k^m(h, \hat{P}_n^k)$ is a (multidimensional) U-statistics for the estimation of $\tilde{U}_k^m(h, P) = \tilde{\mathcal{J}}_k^m(h, P^k)$.

Proposition 1.10.7 *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with probability distribution $P \in \mathcal{P}(E)$ and $h \in \mathscr{L}^2(\tilde{\mathscr{H}}_k^m, P^k)$ with $\mathbb{V}ar[p_{k,1}h_l(X)] > 0$ for all $l = 1, \ldots, m$, where $h_l = \pi_{e_l} \circ h$ is the l^{th}-component of h. Then,*

$$\sqrt{n}(\tilde{U}_{k,n}^m(h, P) - \tilde{U}_k^m(h, P)) \xrightarrow{d} (\tilde{W}^m(P))(h),$$

where $(\tilde{W}^m(P))(h) = ((W(P))(h_1), \ldots, (W(P))(h_m))^\top$ is normally distributed with mean zero and covariance matrix whose $(l_1, l_2)^{th}$-element is given by $k^2\mathbb{E}[p_{k,1}h_{l_1}(X)p_{k,1}h_{l_2}(X)]$, where $l_1, l_2 = 1, \ldots, m$.

Proof of Proposition 1.10.7. Let $\tilde{p}_{k,j}^m : \mathscr{L}^1(\tilde{\mathscr{H}}_k^m, P^k) \to \mathscr{L}^1(\tilde{\mathscr{H}}_j^m, P^j)$ be given by

$$(\tilde{p}_{k,j}^m h)(x_1, \ldots, x_j) = \tilde{\mathcal{J}}_k^m(h, P_{k,j}(x_1, \ldots, x_j)).$$

By Hoeffding decomposition (Proposition 1.10.1) applied to every component,

$$\tilde{U}_{k,n}^m(h, P) - \tilde{U}_k^m(h, P) = \frac{k}{n}\sum_{i=1}^{n} \tilde{p}_{k,1}^m h(X_i) + \sum_{j=2}^{k} \binom{k}{j} \tilde{r}_{n,j}^m,$$

where $\tilde{r}_{n,j}^m = \tilde{\mathcal{J}}_j^m(\tilde{p}_{k,j}^m h, \hat{P}_n^j)$. Next, we notice that $\sqrt{n}\tilde{r}_{n,j}^m \xrightarrow[n\to\infty]{p} 0$. Indeed, the l^{th}-component of $\tilde{r}_{n,j}^m$ is $\pi_{e_l}(\tilde{r}_{n,j}^m) = \mathcal{J}_j(p_{k,j}h_l, \hat{P}_n^j)$ and, as in the proof of Proposition 1.10.3, we see that $\sqrt{n}\pi_{e_l}(\tilde{r}_{n,j}^m) \xrightarrow[n\to\infty]{p} 0$. Finally, observe that $\frac{k}{n}\sum_{i=1}^{n} \tilde{p}_{k,1}^m h(X_i)$ is an average of i.i.d. random variables with mean 0 and covariance matrix whose $(l_1, l_2)^{\text{th}}$-element is given by

$$k^2\mathbb{E}[\pi_{e_{l_1}}(\tilde{p}_{k,1}^m h(X))\pi_{e_{l_2}}(\tilde{p}_{k,1}^m h(X))] = k^2\mathbb{E}[p_{k,1}h_{l_1}(X)p_{k,1}h_{l_2}(X)].$$

Therefore, the result follows from the multivariate central limit theorem. ∎

1.11 U-processes

In this section, we use tools from empirical processes theory and make the convergence results of the previous section (Propositions 1.10.4 and 1.10.3) uniform over a class of functions $\mathcal{H}$. This results are better understood for the case of sums of i.i.d. random variables, that is, U-statistics of order one (van der Vaart and Wellner, 1996; Dudley, 2014; Giné and Nickl, 2016). Extension to arbitrary order is due to (among others) Arcones and Giné (1993). For most of the results in this section we do not provide proofs as they are often quite lengthy and involved. Nevertheless, proofs are provided in the above references. Throughout this section, we denote by E a Hausdorff topological space. First, we extend the definition of expectation and probability to non-measurable functions and sets. Specifically, given a probability space $(\Omega, \Sigma, \mathbb{P})$, we define the outer expectation of $X : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ by

$$\mathbb{E}^*[X] = \inf\Big\{ \mathbb{E}[Y] : Y \geq X, Y : \Omega \to \mathbb{R} \cup \{\pm\infty\} \text{ is measurable and } \mathbb{E}[Y] \text{ exists} \Big\}.$$

Similarly, the outer probability of an arbitrary subset $A \subset \Omega$ is given by

$$\mathbb{P}^*(A) = \inf\{\mathbb{P}(B) : A \subset B, B \in \Sigma\}.$$

We notice that the above infima are achieved. Indeed, by Lemma 1.2.1 of van der Vaart and Wellner (1996), there exists a measurable function $X^* : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ such that (i) $X^* \geq X$, (ii) $X^* \leq Y$ a.s. for every measurable function $Y : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ with $Y \geq X$ a.s. Thus, if $\mathbb{E}[X^*]$ exists, then $E^*[X] = \mathbb{E}[X^*]$. For the existence of $\mathbb{E}[X^*]$ it is enough that $E^*[X] < \infty$. Next, Lemma 1.2.3 of van der Vaart and Wellner (1996) shows that (i) the equality $\mathbb{P}^*(A) = \mathbb{E}^*[\mathbf{I}_A]$, where $A \subset \Omega$, continues to hold for outer expectation and (ii) there exists $A \subset A^* \in \Sigma$ such that $\mathbb{P}^*(A) = P(A^*)$ and $\mathbf{I}_{A^*} = (\mathbf{I}_A)^*$. We refer to Section 1.2 of van der Vaart and Wellner (1996) for further properties of outer expectation and probability. Next, we extend the notions of convergence in distribution, convergence in probability, and almost sure convergence to non-measurable maps (c.f. Section 1.2). Let $(\Omega_n, \Sigma_n, \mathbb{P}_n)$ and $(\Omega, \Sigma, \mathbb{P})$ be probability spaces and $\{X_n\}_{n=1}^{\infty}$ be a sequence of functions $X_n : \Omega_n \to F$, where (F, d_F) is a metric space. We say that X_n converges in distribution* to a random variable $X : \Omega \to F$ if $\lim_{n\to\infty} \mathbb{E}^*[H(X_n)] = \mathbb{E}[H(X)]$ for all bounded and continuous functions $H : F \to \mathbb{R}$ (see Definition 1.3.3 of van der Vaart and Wellner (1996)). In this case, we write $X_n \xrightarrow{d^*} X$. Notice that if X_n are measurable, then $X_n \xrightarrow{d^*} X$ if and only if $\mathbb{P}_{X_n} \xrightarrow{w} \mathbb{P}_X$, that is, $X_n \xrightarrow{d} X$. Next, suppose that $(\Omega_n, \Sigma_n, \mathbb{P}_n) = (\Omega, \Sigma, \mathbb{P})$. We say that X_n converges in outer probability (or in probability*) if $\lim_{n\to\infty} \mathbb{P}^*(\{\omega \in \Omega : d_F(X_n(\omega), X(\omega)) \geq \epsilon\}) = 0$ for all $\epsilon > 0$ (see Definition 1.9.1 of van der Vaart and Wellner (1996)). In this case, we write $X_n \xrightarrow{p^*} X$. Finally, X_n converges outer almost surely (or almost surely*) if $\mathbb{P}^*(\{\omega \in \Omega : \lim_{n\to\infty} d_F(X_n(\omega), X(\omega)) = 0\}) = 1$. In this case, we write $X_n \xrightarrow{a.s.^*} X$. We are now ready to define uniform convergence for a class of functions $\mathcal{H} \subset \mathscr{H}_k$.

Definition 1.11.1 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^1(\mathscr{H}_k, P^k)$. $\mathcal{H}$ is a (P-)Glivenko-Cantelli class if* $\sup_{h\in\mathcal{H}} |U_{k,n}(h, P) - U_k(h, P)| \xrightarrow{a.s.^*} 0$.

Following Arcones and Giné (1993) we give necessary and sufficient conditions for $\mathcal{H}$ to be Glivenko-Cantelli. To this end, we need some conditions allowing for measurability and finiteness of suprema of functions $h \in \mathcal{H}$ and their projections $p_{k,j}h$, $j = 0, \dots, k$, where $p_{k,0} : \mathscr{L}^1(\mathscr{H}_k, P^k) \to \mathbb{R}$ is given by $p_{k,0}h = J_k(h, P^k)$ (see Arcones and Giné (1993) page 1497). For this purpose, it is convenient to assume that the underlying probability space is the countable product of probability spaces $(E, \mathcal{B}(E), P)$, where $P \in \mathcal{P}(E)$. Then, it is sufficient that the above measurability conditions hold for the completion[55] of this probability space (c.f. van der Vaart and

[55]The completion of a probability space $(\Omega, \Sigma, \mathbb{P})$ (w.r.t. $\mathbb{P}$) is the probability space $(\Omega, \overline{\Sigma}, \overline{\mathbb{P}})$, where $\overline{\Sigma}$ consists of all the sets $B \cup N$, where $B \in \Sigma$ and $N \subset \Omega$ with $\mathbb{P}^*(N) = 0$, and $\overline{\mathbb{P}}$ is given by $\overline{\mathbb{P}}(A \cup N) = \mathbb{P}(A)$ (see Exercise 1.2.10 of van der Vaart and Wellner (1996)).

Wellner (1996) pages 108-110). A sufficient condition for this is that the class $\mathcal{H}$ is image admissible Suslin (see Dudley (2014) page 186). We summarize these conditions in the following definition.

Definition 1.11.2 *For $P \in \mathcal{P}(E)$, let $(\Omega, \Sigma, \mathbb{P}) = (E^\infty, (\mathcal{B}(E))^{\otimes\infty}, P^\infty)$ be the product of countably many identical probability spaces $(E, \mathcal{B}(E), P)$ and $\{X_n\}_{n=1}^\infty$ be the sequence of random variables $X_n : \Omega \to E$ that map $\omega \in \Omega$ into its n^{th}-component. We say that a class of functions $\mathcal{H} \subset \mathscr{L}^1(\mathscr{H}_k, P^k)$ is measurable in the sense of Arcones and Giné (1993) if (i) $\sup_{h\in\mathcal{H}}|p_{k,j}h(x_1,\ldots,x_j)| < \infty$ for all $x_1,\ldots,x_k \in E$ and $j = 0,\ldots,k$ and (ii) $\mathcal{H}$ is image admissible Suslin, that is, there exists a Suslin measurable space $(Y, \mathcal{Y})$[56] and a surjective function $T : Y \to \mathcal{H}$ such that the map $(x_1,\ldots,x_k,y) \mapsto (T(y))(x_1,\ldots,x_k)$ is measurable in $(E^k \times Y, (\mathcal{B}(E))^{\otimes k} \otimes \mathcal{Y})$.*

Several steps are required to verify that the desired measurability conditions (cf. page 1497 of Arcones and Giné (1993)) are satisfied by the class $\mathcal{H}$ in Definition 1.11.2. We refer to Appendix A for a thorough analysis. In particular, it is shown there that $\sup_{h\in\mathcal{H}}|U_{k,n}(h,P) - U_k(h,P)|$, the envelope function $h_\mathcal{H}$, and several related functions are measurable in $(\Omega, \overline{\Sigma}, \overline{\mathbb{P}})$, where $(\Omega, \overline{\Sigma}, \overline{\mathbb{P}})$ is the completion of $(\Omega, \Sigma, \mathbb{P})$. Finally, Proposition A.5 shows that the outer expectation of functions that are completion measurable is equal to the expectation w.r.t. the completion. Using Proposition A.5, we suppress in the following the superscript * for sets and functions that are completion measurable. Under the assumptions in Definition 1.11.2, necessary and sufficient conditions for $\mathcal{H}$ (endowed with some pseudometric $\tilde{d}_\mathcal{H}$) to be Glivenko-Cantelli can be given in terms of covering numbers. The ϵ-covering number of a (non-empty) pseudometric space $(F, \tilde{d}_F)$ is the minimum number of $\tilde{d}_F$-balls with radius at most ϵ needed to cover F, that is,

$$N(F, \tilde{d}_F, \epsilon) = \inf\{j \in \mathbb{N} : \exists z_1,\ldots,z_j \in F : \sup_{y\in F} \min_{i=1,\ldots,j} \tilde{d}_F(y, z_i) \le \epsilon\}.$$ [57]

Examples of (random) pseudodistances on a class of functions $\mathcal{H} \subset \mathscr{H}_k$ are $\tilde{d}^{(k,0)}_{\mathcal{H},q,\hat{P}_n^k} : \mathcal{H} \times \mathcal{H} \to [0,\infty)$ and $\tilde{d}^{(1,k-1)}_{\mathcal{H},q,\hat{P}_n^k} : \mathcal{H} \times \mathcal{H} \to [0,\infty)$, where for $q \ge 1$

$$\begin{aligned}
\tilde{d}^{(k,0)}_{\mathcal{H},q,\hat{P}_n^k}(h_1,h_2) &= \tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,\hat{P}_n^k)}(h_1,h_2) \\
\tilde{d}^{(1,k-1)}_{\mathcal{H},q,\hat{P}_n^k}(h_1,h_2) &= \|\mathcal{J}_k(h_1 - h_2, k\delta_{(\cdot)} \times \tilde{P}_n^{k-1}(\cdot))\|_{\mathscr{L}^q(\mathscr{H}_1,P_n)},
\end{aligned}$$

and, for $y \in \{X_1,\ldots,X_n\}$, $\tilde{P}_n^{k-1}(y)$ is the empirical measure $\hat{P}_{n-1}^{k-1}$ of Remark 1.10.1 based on the sample $\{X_1,\ldots,X_n\} \setminus \{y\}$ of size $n-1$.

We introduce some more notation and define the envelope of a class $\mathcal{H} \subset \mathscr{H}_k$ as the function $h_\mathcal{H} : E^k \to [0,\infty)$ given by $h_\mathcal{H}(x_1,\ldots,x_k) = \sup_{h\in\mathcal{H}}|h(x_1,\ldots,x_k)|$. We are now ready to state Theorem 3.1 of Arcones and Giné (1993).

Theorem 1.11.1 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^1(\mathscr{H}_k, P^k)$ be measurable in the sense of Arcones and Giné (1993). If $\mathcal{J}_k(h_\mathcal{H}, P^k) < \infty$, then $\mathcal{H}$ is P-Glivenko-Cantelli if and only if $\frac{1}{n}\log(N(\mathcal{H}, \tilde{d}^{(1,k-1)}_{\mathcal{H},1,\hat{P}_n^k}, \epsilon)) \xrightarrow{p^*} 0$ for all $\epsilon > 0$.*

[56] A Polish space X is a separable, completely metrizable topological space, that is, there is a metric d_X on X such that (X, d_X) is a complete (every Cauchy sequence of points in X has a limit in X) metric space and the topology on X is generated by d_X. A Suslin measurable space is a measurable space $(Y, \mathcal{Y})$ such that (i) $\mathcal{Y}$ is generated by a countable subclass $\mathcal{Z} \subset \mathcal{Y}$ and $\{y\} \in \mathcal{Y}$ for all $y \in Y$ and (ii) there is a Polish space X and a measurable, surjective map $S : X \to Y$ (see Dudley (2014) pages 179-180 and 185).

[57] As usual, the infimum over the empty set is infinity.

We notice that by Jensen's inequality, for all $h_1, h_2 \in \mathcal{H}$,

$$\tilde{d}^{(1,k-1)}_{\mathcal{H},1,\hat{P}^k_n}(h_1, h_2) \leq \tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}^k_n}(h_1, h_2)$$

yielding that, for all $\epsilon > 0$,

$$N(\mathcal{H}, \tilde{d}^{(1,k-1)}_{\mathcal{H},1,\hat{P}^k_n}, \epsilon) \leq N(\mathcal{H}, \tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}^k_n}, \epsilon).$$

This immediately gives the following corollary (Corollary 3.2 in Arcones and Giné (1993)).

Corollary 1.11.1 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^1(\mathscr{H}_k, P^k)$ be measurable in the sense of Arcones and Giné (1993). If $\mathcal{J}_k(h_{\mathcal{H}}, P^k) < \infty$ and $\frac{1}{n}\log(N(\mathcal{H}, \tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}^k_n}, \epsilon)) \xrightarrow{p^*} 0$ for all $\epsilon > 0$, then $\mathcal{H}$ is P-Glivenko-Cantelli.*

For $k = 1$, $\tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}^k_n} = \tilde{d}^{(1,k-1)}_{\mathcal{H},1,\hat{P}^k_n}$[58] yielding that the condition $\frac{1}{n}\log(N(\mathcal{H}, \tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}^k_n}, \epsilon)) \xrightarrow{p^*} 0$, for all $\epsilon > 0$, is necessary (see also Theorem 3.7.14 and Corollary 3.7.15 in Giné and Nickl (2016)). On the other hand, Arcones and Giné (1993), pages 1511-1512, show that, for $k > 1$, the condition $\frac{1}{n}\log(N(\mathcal{H}, \tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}^k_n}, \epsilon)) \xrightarrow{p^*} 0$, for all $\epsilon > 0$, is, in general, only sufficient.

As we will see, a typical sufficient condition for $\frac{1}{n}\log(N(\mathcal{H}, \tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}^k_n}, \epsilon)) \xrightarrow{p^*} 0$ is that $\mathcal{H}$ is a *VC subgraph* class of functions, that is, a class of functions whose subgraphs form a VC class of sets (see Section 2.6 of van der Vaart and Wellner (1996) and Section 3.6 of Giné and Nickl (2016)). We give now a precise definition of VC subgraph class. To this end, we denote by $\#A$ the cardinality of a set A.

Definition 1.11.3 *Let $\mathcal{C}$ be a collection of subsets of a set A. The VC index of $\mathcal{C}$ is*

$$V(\mathcal{C}) = \inf\{j \in \mathbb{N} : \max_{z_1,\ldots,z_j \in A} \#(\mathcal{C} \cap \{\{z_1\}, \ldots, \{z_j\}\}) < 2^j\}.$$[59]

$\mathcal{C}$ is said to be a VC class if $V(\mathcal{C}) < \infty$. The subgraph of a function $h : A \to \mathbb{R}$ is

$$G_h = \{(z,t) \in A \times \mathbb{R} : h(z) \geq t\}.$$

A collection $\mathcal{H}$ of functions $h : A \to \mathbb{R}$ is called a VC subgraph class if the collection $\mathcal{C}_{\mathcal{H}} = \{G_h : h \in \mathcal{H}\}$ of all subgraphs of functions in $\mathcal{H}$ is a VC class of sets in $A \times \mathbb{R}$.

We give several examples of VC subgraph classes of functions in Section 2.7. In many of those examples it is used that the collection of indicators of a VC class of sets is a VC subgraph class of functions. To see this, let $\mathcal{H} = \{\mathbf{I}_C : C \in \mathcal{C}\}$, where $\mathcal{C}$ is a VC class of sets in a set A. Then, for $C \in \mathcal{C}$,

$$G_{\mathbf{I}_C} = \{(z,t) \in A \times \mathbb{R} : \mathbf{I}_C \geq t\} = A \times (-\infty, 0] \cup C \times (0, 1],$$

yielding that

$$\mathcal{C}_H = \{G_{\mathbf{I}_C} : C \in \mathcal{C}\} = A \times (-\infty, 0] \cup \mathcal{C} \times (0, 1].$$

[58]For $k = 1$, $k\delta_{(\cdot)} \times \hat{P}^{k-1}_n(\cdot)$ reduces to $\delta_{(\cdot)}$ yielding that $\mathcal{J}_k(h_1 - h_2, k\delta_{(\cdot)} \times \hat{P}^{k-1}_n(\cdot)) = h_1 - h_2$. Also, notice that $\hat{P}^1_n = P_n$.

[59]For two collections of subsets of a set A, $\mathcal{C}$ and $\mathcal{D}$, $\mathcal{C} \cap \mathcal{D} = \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$.

It follows from Proposition 1.11.1 below that $\mathcal{C}_H$ is a VC class of sets. We summarize therein several permanence properties of VC classes of sets (see Lemma 2.6.17 in van der Vaart and Wellner (1996) and Proposition 3.6.7 in Giné and Nickl (2016)).

Proposition 1.11.1 *Let $\mathcal{C}$ and $\mathcal{D}$ be VC classes of sets in a set A, $\varphi : A \to B$ be a bijective function and $\psi : Z \to A$ be an arbitrary function. Then, the following classes of sets are VC:*

(i) $A \setminus \mathcal{C} = \{A \setminus C : C \in \mathcal{C}\}$,

(ii) $\mathcal{C} \cap \mathcal{D} = \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$,

(iii) $\mathcal{C} \cup \mathcal{D} = \{C \cup D : C \in \mathcal{C}, D \in \mathcal{D}\}$,

(iv) $\varphi(\mathcal{C}) = \{\varphi(C) : C \in \mathcal{C}\}$,

(v) $\psi^{-1}(\mathcal{C}) = \{\psi^{-1}(C) : C \in \mathcal{C}\}$,

(vi) $\tilde{\mathcal{C}}$, *where* $\tilde{\mathcal{C}} \subset \mathcal{C}$.

Finally, if $\mathcal{C}$ and $\mathcal{D}$ are VC classes in sets A and B, then the following class is VC:

(vii) $\mathcal{C} \times \mathcal{D} = \{C \times D : C \in \mathcal{C}, D \in \mathcal{D}\}$.

For VC subgraph classes of functions we have the following result (see Lemma 2.6.18 in van der Vaart and Wellner (1996)).

Proposition 1.11.2 *Let $\mathcal{H}$ and $\mathcal{I}$ be VC subgraph classes of functions on a set A, $g : \mathbb{R} \to \mathbb{R}$ be a monotone function, and $\varphi : A \to \mathbb{R}$ and $\psi : Z \to A$ be arbitrary functions. Then, the following classes of functions are VC subgraph:*

(i) $\min(\mathcal{H}, \mathcal{I}) = \{\min(h, i) : h \in \mathcal{H}, i \in \mathcal{I}\}$,

(ii) $\max(\mathcal{H}, \mathcal{I}) = \{\max(h, i) : h \in \mathcal{H}, i \in \mathcal{I}\}$,

(iii) $\varphi + \mathcal{H} = \{\varphi + h : h \in \mathcal{H}\}$,

(iv) $\varphi\mathcal{H} = \{\varphi h : h \in \mathcal{H}\}$,

(v) $\mathcal{H} \circ \psi = \{h \circ \psi : h \in \mathcal{H}\}$,

(vi) $g \circ \mathcal{H} = \{g \circ h : h \in \mathcal{H}\}$.

The importance of VC subgraph classes lies in that their ϵ-covering numbers are bounded above by a polynomial in ϵ (see Theorem 2.6.7 of van der Vaart and Wellner (1996) or Theorem 3.6.9 of Giné and Nickl (2016)).

Theorem 1.11.2 *For $Q \in \mathcal{P}(E^k)$ and $q \geq 1$, let $\mathcal{H} \subset \mathscr{L}^q(\mathscr{H}_k, Q)$ be a VC subgraph class of functions with envelope $h_{\mathcal{H}}$ that is measurable for the completion of $(E^k, (\mathcal{B}(E))^{\otimes k}, Q)$ and satisfies $\|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k, Q)} > 0$.*[60] *Then, for all $0 < \epsilon < 1$,*

$$N(\mathcal{H}, \tilde{d}_{\mathscr{L}^q(\mathscr{H}_k, Q)}, \epsilon \|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k, Q)}) \leq c_1 \epsilon^{-q c_2}$$

for some constants $c_1 \geq 1$ and $c_2 \geq 0$ that depend on $V(\mathcal{C}_{\mathcal{H}})$ but not on Q.

[60] $\|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k, Q)} = \infty$ is possible, in which case, by definition of covering number, $N(\mathcal{H}, \tilde{d}_{\mathscr{L}^q(\mathscr{H}_k, Q)}, \epsilon \|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k, Q)}) = 1$.

Remark 1.11.1 *Under the assumptions of Theorem 1.11.2, we have that*

$$N(\mathcal{H}, \tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,Q)}, \epsilon \|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k,Q)}) \leq \begin{cases} c_1 \epsilon^{-qc_2} & \text{if } 0 < \epsilon < 1, \\ c_1 & \text{if } 1 \leq \epsilon < 2, \\ 1 & \text{if } \epsilon \geq 2. \end{cases}$$

Indeed, since, for every pseudometric space $(F, \tilde{d}_F)$, the function $\epsilon \mapsto N(F, \tilde{d}_F, \epsilon)$ is non-increasing, we have that, for all $\epsilon \geq 1$,

$$N(\mathcal{H}, \tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,Q)}, \epsilon \|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k,Q)}) \leq \lim_{\eta \to 1^-} c_1 \eta^{-qc_2} = c_1.$$

Next, for all $h_1, h_2 \in \mathcal{H}$, we have that

$$\sup_{h \in \mathcal{H}} \min_{i=1,2} \tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,Q)}(h, h_i) \leq \sup_{h \in \mathcal{H}} \|h\|_{\mathscr{L}^q(\mathscr{H}_k,Q)} + \min_{i=1,2} \|h_i\|_{\mathscr{L}^q(\mathscr{H}_k,Q)} \leq 2\|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k,Q)},$$

implying that $N(\mathcal{H}, \tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,Q)}, \epsilon \|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k,Q)}) = 1$ for all $\epsilon \geq 2$. We finally notice that, if $\|h_{\mathcal{H}}\|_{\mathscr{L}^q(\mathscr{H}_k,Q)} = 0$, then $N(\mathcal{H}, \tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,Q)}, \epsilon) = 1$ for all $\epsilon > 0$.

By combining Corollary 1.11.1 and Theorem 1.11.2, we obtain that VC subgraph classes of functions are Glivenko-Cantelli (Corollary 3.3 in Arcones and Giné (1993)).

Corollary 1.11.2 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^1(\mathscr{H}_k, P^k)$ be measurable in the sense of Arcones and Giné (1993). If $\mathcal{J}_k(h_{\mathcal{H}}, P^k) < \infty$ and $\mathcal{H}$ is VC subgraph, then $\mathcal{H}$ is P-Glivenko-Cantelli.*

Proof of Corollary 1.11.2. By Corollary 1.11.1 it is enough to show that for all $\epsilon > 0$ $\frac{1}{n} \log(N(\mathcal{H}, \tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}_n^k}, \epsilon)) \xrightarrow{p^*} 0$. Recall that $\tilde{d}^{(k,0)}_{\mathcal{H},1,\hat{P}_n^k} = \tilde{d}_{\mathscr{L}^1(\mathscr{H}_k,\hat{P}_n^k)}$. Using Theorem 1.11.2 with $q = 1$ and $Q = \hat{P}_n^k$ and Remark 1.11.1 we see that for all $\epsilon > 0$

$$N(\mathcal{H}, \tilde{d}_{\mathscr{L}^1(\mathscr{H}_k,\hat{P}_n^k)}, \epsilon) \leq c_1 \max\left(1, \left(\frac{\|h_{\mathcal{H}}\|_{\mathscr{L}^1(\mathscr{H}_k,\hat{P}_n^k)}}{\epsilon}\right)^{c_2}\right). \tag{1.11.1}$$

Since, by Proposition 1.10.4, $\|h_{\mathcal{H}}\|_{\mathscr{L}^1(\mathscr{H}_k,\hat{P}_n^k)} \xrightarrow{a.s.} \|h_{\mathcal{H}}\|_{\mathscr{L}^1(\mathscr{H}_k,P^k)}$, (1.11.1) yields that

$$\frac{1}{n} \log(N(\mathcal{H}, \tilde{d}_{\mathscr{L}^1(\mathscr{H}_k,\hat{P}_n^k)}, \epsilon)) \xrightarrow{a.s.^*} 0.$$

■

We now turn to the uniform CLT over a class $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_k, P^k)$, where $P \in \mathcal{P}(E)$. We study only the non-degenerate case, that is, $U_{n,k}(h)$ is non-degenerate for all $h \in \mathcal{H}$ (Section 4 of Arcones and Giné (1993)). The degenerate case is studied in Section 5 of Arcones and Giné (1993). We begin with the case $k = 1$ and define the Gaussian limit process. We denote by $\mathfrak{f}(A)$ the set of all real-valued functions from a set A and by $\ell^\infty(A) \subset \mathfrak{f}(A)$ the subset of all bounded functions. We endow $\ell^\infty(A)$ with the supremum norm.

Definition 1.11.4 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_1, P)$. A (P-)Brownian bridge is an $\mathfrak{f}(\mathcal{H})$-valued Gaussian process $\{(B(P))(h)\}_{h\in\mathcal{H}}$ with mean $0 \in \mathfrak{f}(\mathcal{H})$ and covariance function*[61,62] $\gamma_{B(P)} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ *given by*

$$\gamma_{B(P)}(h_1, h_2) = J_1((p_{1,1}h_1)(p_{1,1}h_2), P).$$

The name P-Brownian bridge is due to the fact that if $P = \lambda_{|E}$ is the Lebesgue measure on $E = [0,1]$ and $\mathcal{H} = \{\mathbf{I}_{[0,t]} : t \in [0,1]\}$, then $\gamma_{B(P)}(\mathbf{I}_{[0,s]}, \mathbf{I}_{[0,t]}) = \min(s,t) - st$. By identifying $\mathcal{H}$ with $[0,1]$ via the bijection $\mathbf{I}_{[0,t]} \longleftrightarrow t$, we see that $\{(B(P))(h)\}_{h\in\mathcal{H}}$ is indeed a Brownian bridge.

In general, the stochastic process $\{(B(P))(h)\}_{h\in\mathcal{H}}$ is not continuous nor bounded. However, it is useful that $\{(B(P))(h)\}_{h\in\mathcal{H}}$ admits a continuous and bounded version (see Definition 3.7.26 of Giné and Nickl (2016)). We recall that $\{\tilde{Y}(a)\}_{a\in A}$ is a version of the stochastic process $\{Y(a)\}_{a\in A}$ if $\mathbb{P}(\tilde{Y}(a) = Y(a)) = 1$ for all $a \in A$ (see page 116 of van der Vaart and Wellner (1996)). For $k \in \mathbb{N}$, $q \geq 1$, and $\mathcal{H} \subset \mathscr{H}_k$ we define the pseudodistance $\tilde{d}^{(1,k-1)}_{\mathcal{H},q,P^k} : \mathcal{H} \times \mathcal{H} \to [0, \infty)$ by $\tilde{d}^{(1,k-1)}_{\mathcal{H},q,P^k}(h_1, h_2) = \|p_{k,1}h_1 - p_{k,1}h_2\|_{\mathscr{L}^q(\mathscr{H}_1,P)}$.[63]

Definition 1.11.5 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_1, P)$. We say that $\mathcal{H}$ is (P-)pre-Gaussian if the P-Brownian bridge $\{(B(P))(h)\}_{h\in\mathcal{H}}$ admits a version whose sample paths are all bounded and uniformly continuous w.r.t. the pseudodistance $\tilde{d}^{(1,0)}_{\mathcal{H},2,P}$.*

Thus, if $\mathcal{H}$ is P-pre-Gaussian, then $\{(B(P))(h)\}_{h\in\mathcal{H}}$ has a version that takes values in $\ell^\infty(\mathcal{H})$. We also suppose that the empirical processes $\{B_n(P)(h)\}_{h\in\mathcal{H}}$, where $B_n(P)(h) = \sqrt{n}(U_{1,n}(h,P) - U_1(h,P))$, are bounded so that $\{B_n(P)(h)\}_{h\in\mathcal{H}}$ assumes values in $\ell^\infty(\mathcal{H})$. For this, it is sufficient that (i) of Definition 1.11.2 holds (here $k = 1$). Accordingly, the uniform CLT is given as convergence in distribution* in $\ell^\infty(\mathcal{H})$ (see Definition 3.7.29 of Giné and Nickl (2016)).

Definition 1.11.6 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_1, P)$ satisfying (i) of Definition 1.11.2. We say that $\mathcal{H}$ is (P-)Donsker if it is P-pre-Gaussian and*

$$B_n(P) \xrightarrow{d^*} \tilde{B}(P),$$

where $\tilde{B}(P)$ is the bounded and uniformly continuous version of $B(P)$.

Using Remark 3.7.27 of Giné and Nickl (2016), we see that $\mathcal{H}$ is P-pre-Gaussian if and only if $\{(B(P))(h)\}_{h\in\mathcal{H}}$ admits a version whose sample paths are all uniformly continuous w.r.t. the pseudodistance $\tilde{d}^{(1,0)}_{\mathcal{H},2,P}$ and $(\mathcal{H}, \tilde{d}^{(1,0)}_{\mathcal{H},2,P})$ is totally bounded (cf. page 89 of van der Vaart and Wellner (1996)). We recall that a pseudometric space $(F, \tilde{d}_F)$ is totally bounded if and only if for all $\epsilon > 0$ there exists $x_1, \ldots, x_m \in F$ such that $F \subset \cup_{i=1}^m \{y \in F : \tilde{d}_F(y, x_i) < \epsilon\}$. Next, we notice that, by the continuous mapping theorem (Theorem 1.3.6 of van der Vaart and Wellner (1996)), $B_n(P) \xrightarrow{d^*} \tilde{B}(P)$ implies convergence of the corresponding finite-dimensional marginal distributions. Theorem 1.5.4 and 1.5.7 of van der Vaart and Wellner (1996) yield the converse of

[61] $\gamma : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is a covariance function if, for all $m \in \mathbb{N}$ and $h_1, \ldots, h_m \in \mathcal{H}$, the $m \times m$ matrix whose $(i,j)^{\text{th}}$-element is given by $\gamma(h_i, h_j)$ is symmetric and positive semi-definite.

[62] Existence of a Gaussian process with given mean and covariance function on some probability space $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}})$ is guaranteed by Kolmogorov existence theorem (see Theorem 2.1.5 in Oksendal (2003) or Theorem 2.4.3 in Tao (2011)).

[63] Notice the similarity between $\tilde{d}^{(1,k-1)}_{\mathcal{H},q,P^k}$ and $\tilde{d}^{(1,k-1)}_{\mathcal{H},q,\hat{P}^k_n}$. However, $\hat{P}^k_n$ is not a product measure, which makes its expression seemingly more involved.

this statement under an additional condition (see (iii) below), which entails that, with high probability, $B_n(P)(h_1)$ and $B_n(P)(h_2)$ are close to each other whenever $\tilde{d}^{(1,0)}_{\mathcal{H},2,P}(h_1, h_2)$ is small and n is large. We thus obtain Example 1.5.10 of van der Vaart and Wellner (1996) (see also Theorem 3.7.31 of Giné and Nickl (2016)).

Theorem 1.11.3 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_1, P)$ satisfying (i) of Definition 1.11.2. $\mathcal{H}$ is P-Donsker if and only if the following three conditions hold:*
(i) the finite-dimensional marginal distributions of $\{(B_n(P))(h)\}_{h\in\mathcal{H}}$ converge weakly to the corresponding marginal distributions of $\{(B(P))(h)\}_{h\in\mathcal{H}}$,
(ii) $(\mathcal{H}, \tilde{d}^{(1,0)}_{\mathcal{H},2,P})$ is totally bounded, and
(iii) $\{(B_n(P))(h)\}_{h\in\mathcal{H}}$ is asymptotically equicontinuous in probability w.r.t. $\tilde{d}^{(1,0)}_{\mathcal{H},2,P}$, that is, for all $\epsilon > 0$,

$$\lim_{\delta\to 0^+} \limsup_{n\to\infty} \mathbb{P}^*\Big(\sup_{h_1-h_2 \in \mathcal{H}_{\tilde{d}^{(1,0)}_{\mathcal{H},2,P},\delta}(P)} |(B_n)(P)(h_1 - h_2)| \geq \epsilon\Big) = 0,$$

where

$$\mathcal{H}_{\tilde{d}^{(1,0)}_{\mathcal{H},2,P},\delta}(P) = \{h_1 - h_2 : h_1, h_2 \in \mathcal{H} \text{ and } \tilde{d}^{(1,0)}_{\mathcal{H},2,P}(h_1, h_2) \leq \delta\}.$$

Actually, Example 1.5.10 shows that it is enough that (ii) and (iii) in Theorem 1.11.3 hold for any pseudodistance on $\mathcal{H}$. As we will see below, a convenient choice is the $\mathscr{L}^2(\mathscr{H}_1, P)$-pseudometric $\tilde{\tilde{d}}^{(1,0)}_{\mathcal{H},2,P}$, which is defined as the restriction of $\tilde{d}_{\mathscr{L}^2(\mathscr{H}_1,P)}$ to $\mathcal{H}$. Similarly, if (i) of Definition 1.11.2 holds, then $\tilde{d}^{(1,0)}_{\mathcal{H},2,P}$ can be replaced in Definition 1.11.7 by $\tilde{\tilde{d}}^{(1,0)}_{\mathcal{H},2,P}$ (see page 89 of van der Vaart and Wellner (1996)). Finally, we notice that, by the CLT (Proposition 1.10.7 with $k = 1$), (i) of Theorem 1.11.3 is satisfied if and only if $\mathbb{V}ar[p_{1,1}h(X_1)] > 0$ for all $h \in \mathcal{H}$, that is, $p_{1,1}h \neq 0$ P-a.s. and h is non-constant on a set of positive probability. In this case, we say that $\mathcal{H}$ is (P-)non-degenerate. Using that E is a Hausdorff topological space and footnote [41], we see that in this case $U_{1,n}(h, P) = \mathcal{J}_1(h, P_n) = \frac{1}{n}\sum_{i=1}^n h(X_i)$ is non-degenerate for all $h \in \mathcal{H}$.

As before, necessary conditions for a class to be Donsker can be given in terms of covering numbers. To this end, we assume that E is second countable, which is used alongside Definition 1.11.2 to obtain the desired measurability conditions (see Proposition A.3 in Appendix A). We begin with Theorem 3.7.36 of Giné and Nickl (2016).

Theorem 1.11.4 *Suppose that E is second countable. Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_1, P)$ be measurable in the sense of Arcones and Giné (1993) and P-non-degenerate. If $\mathcal{J}_1(h^2_{\mathcal{H}}, P) < \infty$ and*

$$\lim_{\delta\to 0^+} \limsup_{n\to\infty} E^*\Big[\int_0^\delta \sqrt{\log(N(\mathcal{H}, \tilde{d}_{\mathscr{L}^2(\mathscr{H}_1,\hat{P}^1_n)}, \epsilon)} d\epsilon\Big] = 0, \tag{1.11.2}$$

then $\mathcal{H}$ is P-Donsker.

Remark 1.11.2 *Let $\mathcal{P}_{df}(E)$ be the subset of $\mathcal{P}(E)$ consisting of all finitely discrete probability measures. By Theorem 3.7.37 of Giné and Nickl (2016) (see also Theorem 2.5.2 of van der Vaart and Wellner (1996) and Theorem 6.3.1 of Dudley (2014)), a sufficient condition for (1.11.2) is that*

$$\int_0^\infty \sup_{Q\in\mathcal{P}_{df}(E): \|h_{\mathcal{H}}\|_{\mathscr{L}^2(\mathscr{H}_1,Q)}>0} \sqrt{\log(N(\mathcal{H}, \tilde{d}_{\mathscr{L}^2(\mathscr{H}_1,Q)}, \epsilon\|h_{\mathcal{H}}\|_{\mathscr{L}^2(\mathscr{H}_1,Q)})} d\epsilon < \infty. \tag{1.11.3}$$

Notice that (1.11.3) holds if $\mathcal{H}$ is a VC subgraph class. Indeed, using Remark 1.11.1, we see that the left hand side (LHS) of (1.11.3) is bounded above by

$$\int_0^1 \sqrt{\log(c_1) + 2c_2 \log(1/\epsilon)} d\epsilon + \sqrt{\log(c_1)} < \infty.$$[64]

Now, we extend the above results to general k and class $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_k, P^k)$. To this end, we need some more notation. For $j = 1, \ldots, k$, we define the product measure on E^k, $\overline{P}_{k,j}(x_1, \ldots, x_j) = \prod_{i=1}^{j} \delta_{x_i} \times P^{k-j}$, where $x_1, \ldots, x_j \in E$, and the function $\overline{p}_{k,j} : \mathscr{L}^1(\mathscr{H}_k, P^k) \to \mathscr{L}^1(\mathscr{H}_j, P^j)$ by $(\overline{p}_{k,j}h)(x_1, \ldots, x_j) = \mathcal{J}_k(h, \overline{P}_{k,j}(x_1, \ldots, x_j))$. The limit process is the $\mathfrak{f}(\mathcal{H})$-valued Gaussian process $\{(W_k(P))(h)\}_{h \in \mathcal{H}}$, where $W_k(P)(h) = kB(P)(\overline{p}_{k,1}h)$ and $\{(B(P))(\overline{p}_{k,1}h)\}_{h \in \mathcal{H}}$ is the P-Brownian bridge process indexed by $\overline{p}_{k,1}\mathcal{H}$. Next, the empirical process $\{B_n(P)(h)\}_{h \in \mathcal{H}}$ is replaced, for general k, by the U-process $\{(W_{k,n}(P))(h)\}_{h \in \mathcal{H}}$, where $(W_{k,n}(P))(h) = \sqrt{n}(U_{k,n}(h, P) - U_k(h, P))$. Finally, for general k, the pseudodistances $\tilde{d}^{(1,0)}_{\mathcal{H},2,P}$ and $\stackrel{\approx}{d}{}^{(1,0)}_{\mathcal{H},2,P}$ are replaced by $\tilde{d}^{(1,k-1)}_{\mathcal{H},2,P^k}$ and $\stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},2,P^k}$, respectively, where, for $k \in \mathbb{N}$ and $q \geq 1$, $\tilde{d}^{(1,k-1)}_{\mathcal{H},q,P^k}$ is defined before Definition 1.11 and $\stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},q,P^k} : \mathcal{H} \times \mathcal{H} \to [0, \infty)$ is given by $\stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},q,P^k}(h_1, h_2) = \|\overline{p}_{k,1}h_1 - \overline{p}_{k,1}h_2\|_{\mathscr{L}^q(\mathscr{H}_1, P)}$. We begin with the definition of Donsker class for general k.

Definition 1.11.7 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_k, P^k)$ satisfying (i) of Definition 1.11.2. We say that $\mathcal{H}$ is a (P-)Donsker class if $\overline{p}_{k,1}\mathcal{H}$ is P-pre-Gaussian and*

$$W_{k,n}(P) \xrightarrow{d^*} \tilde{W}_k(P),$$

where $\tilde{W}_k(P)$ is the bounded and uniformly continuous version of $W_k(P)$.

The condition that $\overline{p}_{k,1}\mathcal{H}$ is P-pre-Gaussian entails that $\{(W_k(P))(h)\}_{h \in \mathcal{H}}$ admits a version whose sample paths are all bounded and uniformly continuous w.r.t. the pseudodistance $\stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},2,P^k}$. Indeed, for all $h_1, h_2 \in \mathcal{H}$ and $q \geq 1$

$$\stackrel{\approx}{d}{}^{(1,0)}_{\overline{p}_{k,1}\mathcal{H},q,P}(\overline{p}_{k,1}h_1, \overline{p}_{k,1}h_2) = \|\overline{p}_{k,1}h_1 - \overline{p}_{k,1}h_2\|_{\mathscr{L}^q(\mathscr{H}_1,P)} = \stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},q,P^k}(h_1, h_2).$$

As mentioned before, boundedness of the sample paths is equivalent to total boundedness of $(\mathcal{H}, \stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},2,P^k})$. Theorem 1.11.3 now takes the following form (see pages 1496-1497 of Arcones and Giné (1993)).

Theorem 1.11.5 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_k, P^k)$ satisfying (i) of Definition 1.11.2. $\mathcal{H}$ is P-Donsker if and only if the following three conditions hold:*
(i) the finite-dimensional marginal distributions of $\{(W_{k,n}(P))(h)\}_{h \in \mathcal{H}}$ converge weakly to the corresponding marginal distributions of $\{(W_k(P))(h)\}_{h \in \mathcal{H}}$,
(ii) $(\mathcal{H}, \stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},2,P^k})$ is totally bounded, and
(iii) $\{(W_{k,n}(P))(h)\}_{h \in \mathcal{H}}$ is asymptotically equicontinuous in probability w.r.t. $\stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},2,P^k}$, that is, for all $\epsilon > 0$,

$$\lim_{\delta \to 0^+} \limsup_{n \to \infty} \mathbb{P}^*\Big(\sup_{h_1 - h_2 \in \mathcal{H}_{\stackrel{\approx}{d}{}^{(1,k-1)}_{\mathcal{H},2,P^k}, \delta}(P)} |(W_{k,n})(P)(h_1 - h_2)| \geq \epsilon \Big) = 0, \tag{1.11.4}$$

[64] Set $t = \frac{1}{\epsilon}$ and use that $\log(t) \leq t$ for all $t \geq 1$.

where

$$\mathcal{H}_{\tilde{\tilde{d}}^{(1,k-1)}_{\mathcal{H},2,P^k},\delta}(P) = \{h_1 - h_2 : h_1, h_2 \in \mathcal{H} \text{ and } \tilde{\tilde{d}}^{(1,k-1)}_{\mathcal{H},2,P^k}(h_1,h_2) \leq \delta\}.$$

Using Proposition 1.10.7, we see that condition (i) is satisfied if and only if $\mathbb{V}ar[p_{k,1}h(X_1)] > 0$ for all $h \in \mathcal{H}$, that is, $p_{k,1}h \neq 0$ P-a.s. As before, in this case, we say that $\mathcal{H}$ is (P-)non-degenerate. Thus, non-degenerate classes are Donsker if and only if (ii) and (iii) of Theorem 1.11.5 hold.

Now, we briefly discuss convergence of finite dimensional distributions for depth functions, which, by Theorem 1.11.5, is necessary for the uniform CLT. Specifically, Proposition 1.10.7 yields convergence of the finite dimensional distributions of *Type A* and *Type B* depth functions that, for a given $P \in \mathcal{P}_d$, are (functions of) non-degenerate U-statistics with finite second moment. In particular, *Type A* depth functions possess finite second moments because they are bounded (see Definition 1.6.1). Finiteness of the second moment for *Type B* depth functions depends on the distribution P. Turning to the halfspace depth, uniform CLT is studied by Massé (2004). In this case, convergence of finite dimensional distributions does not hold for a large class of probability distributions (see Theorem 2.4 and Remark 2.1 of Massé (2004)). Nevertheless, Theorem 2.1 and Corollary 2.3 therein shows that the uniform CLT holds under some additional conditions. Specifically, it is supposed that $P \in \mathcal{P}_{d,hp}$. Then, the function $u \mapsto P(H_{x,u})$ is continuous by Proposition 4.5 (i) and, since S^{d-1} is compact, it assumes a minimal value. Thus, for all $x \in \mathbb{R}^d$ there exists $u_x^* \in S^{d-1}$ such that $\tilde{D}_H(x,P) = P(H_{x,u_x^*})$. The CLT holds uniformly on a closed subset $A \subset \mathbb{R}^d$ such that $\tilde{D}(x,P) > 0$ and u_x^* is unique for all $x \in A$.

We now return to the general theory of U-processes and provide in the following necessary and sufficient conditions for a class to be Donsker. We begin with Theorem 4.1 of Arcones and Giné (1993).

Theorem 1.11.6 *Suppose that E is second countable. Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_k, P^k)$ be measurable in the sense of Arcones and Giné (1993), P-non-degenerate, and such that $\lim_{t\to\infty} t^2 P^k(h_{\mathcal{H}} > t) = 0$. $\mathcal{H}$ is P-Donsker if and only if $(\mathcal{H}, \tilde{\tilde{d}}^{(1,k-1)}_{\mathcal{H},2,P^k})$ is totally bounded and for some (resp. all) $0 < q < 2$*

$$\lim_{\delta\to 0^+} \limsup_{n\to\infty} E[\sup_{h_1-h_2 \in \mathcal{H}_{\tilde{\tilde{d}}^{(1,k-1)}_{\mathcal{H},2,P^k},\delta}(P)} |(W_{k,n})(P)(h_1-h_2)|^q] = 0. \tag{1.11.5}$$

If $\lim_{t\to\infty} t^2 P^k(h_{\mathcal{H}} > t) = 0$ we say that the envelope function $h_{\mathcal{H}}$ possesses a weak second moment condition. This condition is necessary for $\mathcal{H}$ to be Donsker (see pages 129-130 of van der Vaart and Wellner (1996) and Proposition 3.7.32 of Giné and Nickl (2016)) and, since $t^2 P^k(h_{\mathcal{H}} > t) \leq \mathcal{J}_k(h^2_{\mathcal{H}} \mathbf{I}_{[h_{\mathcal{H}}>t]}, P^k)$, it is satisfied whenever $\mathcal{J}_k(h^2_{\mathcal{H}}, P^k) < \infty$. Under the conditions of Definition 1.11.2, we can remove outer probability in Theorem 1.11.5. Thus, the "if" part in Theorem 1.11.6 follows immediately from Theorem 1.11.5 and Markov inequality. The term "resp. all" in Theorem 1.11.6 means that if $\mathcal{H}$ is P-Donsker, then (1.11.5) already holds for all $0 < q < 2$.

We know from Hoeffding decomposition (Proposition 1.10.1) that for all $h \in \mathcal{H}$

$$W_{k,n}(P)(h) = \sqrt{n} \sum_{j=1}^{k} \binom{k}{j} U_{j,n}(p_{k,j}h, P). \tag{1.11.6}$$

Since $\mathcal{H}$ is P-non-degenerate, the finite dimensional distributions of $\{\sqrt{n}kU_{1,n}(p_{k,1}h,P)\}_{h\in\mathcal{H}}$ converge to those of $\{W_k(P)(h)\}_{h\in\mathcal{H}}$, whereas, for fixed h, the other terms in (1.11.6) converge to zero (cf. Proposition 1.10.7). Therefore, we see that $\mathcal{H}$ is P-Donsker if $p_{k,1}\mathcal{H}$ is P-Donsker and the remainder terms in (1.11.6) satisfy the asymptotic equicontinuity condition (1.11.4) or (1.11.5). This shows that (ii) and (iii) of the following corollary imply (i) (see Corollary 4.2 of Arcones and Giné (1993)).

Corollary 1.11.3 *Suppose that E is second countable. Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_k, P^k)$ be measurable in the sense of Arcones and Giné (1993), P-non-degenerate, and such that $\lim_{t\to\infty} t^2 P^k(h_{\mathcal{H}} > t) = 0$. The following are equivalent:*
(i) $\mathcal{H}$ is P-Donsker,
(ii) $\overline{p}_{k,1}\mathcal{H}$ is P-Donsker and $\sqrt{n}\sup_{h\in\mathcal{H}}|U_{j,n}(p_{k,j}h, P)| \xrightarrow{P} 0$ for all $j = 2, \ldots, m$, and
(iii) $\overline{p}_{k,1}\mathcal{H}$ is P-Donsker and $\sqrt{n}E[\sup_{h\in\mathcal{H}}|U_{j,n}(p_{k,j}h, P)|^q] \to 0$ for all $j = 2, \ldots, m$ and some (resp. all) $0 < q < 2$.

The next result gives a simple necessary condition for convergence in (ii) and (iii) of Corollary 1.11.3 in terms of covering numbers (see Theorem 4.4 of Arcones and Giné (1993)). We say that $\mathcal{H} \subset \mathscr{H}_K$ is uniformly bounded if for some $0 < l_{\mathcal{H}} < \infty$ and all $h \in \mathcal{H}$ $|h| \le l_{\mathcal{H}}$.

Theorem 1.11.7 *Suppose that E is second countable. Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{H}_k$ be measurable in the sense of Arcones and Giné (1993), P-non-degenerate, and uniformly bounded. Then, $\mathcal{H}$ is P-Donsker if $\overline{p}_{k,1}\mathcal{H}$ is P-Donsker and for all $\epsilon > 0$*

$$\lim_{n\to\infty} E^*[n^{-1/2}\log(N(\mathcal{H}, \tilde{d}_{\mathscr{L}^1(\mathscr{H}_k, \hat{P}_n^k)}, n^{-1/2}\epsilon))] = 0. \tag{1.11.7}$$

Notice that (1.11.7) is satisfied by a uniformly bounded VC subgraph class of functions $\mathcal{H}$. Indeed, using (1.11.1) and $\|h_{\mathcal{H}}\|_{\mathscr{L}^1(\mathscr{H}_k, \hat{P}_n^k)} \le l_{\mathcal{H}}$, for some $0 < l_{\mathcal{H}} < \infty$, we obtain that for $n \ge \frac{\epsilon^2}{l_{\mathcal{H}}^2}$

$$n^{-1/2}\log(N(\mathcal{H}, \tilde{d}_{\mathscr{L}^1(\mathscr{H}_k, \hat{P}_n^k)}, n^{-1/2}\epsilon)) \le n^{-1/2}\log(c_1) + n^{-1/2}c_2\log\left(\frac{l_{\mathcal{H}}}{n^{-1/2}\epsilon}\right),$$

which converges to zero as $n \to \infty$. We thus obtain the following corollary.

Corollary 1.11.4 *Suppose that E is second countable. Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{H}_k$ be measurable in the sense of Arcones and Giné (1993), P-non-degenerate, and uniformly bounded. If $\mathcal{H}$ is a VC subgraph class, then $\mathcal{H}$ is P-Donsker.*

Proof of Corollary 1.11.4. By Theorem 1.11.7, it is enough to show that $\overline{p}_{k,1}\mathcal{H}$ is P-Donsker. This follows from Proposition 1.11.3 below. ■

Proposition 1.11.3 *Suppose that E is second countable. Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_k, P^k)$ be measurable in the sense of Arcones and Giné (1993), P-non-degenerate, and such that $\mathcal{J}_k(h_{\mathcal{H}}^2, P^k) < \infty$. If $\mathcal{H}$ is a VC subgraph class, then $\overline{p}_{k,1}\mathcal{H}$ is P-Donsker.*

Proof of Proposition 1.11.3. We apply Theorem 1.11.4 to the class $\overline{p}_{k,1}\mathcal{H}$. To this end, notice that by Jensen's inequality $\overline{p}_{k,1}\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_1, P)$ and $\mathcal{J}_1(h^2_{\overline{p}_{k,1}\mathcal{H}}, P) \le \mathcal{J}_k(h_{\mathcal{H}}^2, P^k) < \infty$. Next, using that $\mathcal{H}$ is measurable in the sense of Arcones and Giné (1993) and $p_{k,j}h = p_{1,j}\overline{p}_{k,1}h$ for $h \in \mathcal{H}$ and $j = 0, 1$, we obtain that $\overline{p}_{k,1}\mathcal{H}$ satisfies (i) of Definition 1.11.2. $\overline{p}_{k,1}\mathcal{H}$ satisfies (ii) of Definition 1.11.2 by Proposition A.1 in Appendix A. To conclude the proof, it is enough to show that

$$\lim_{\delta\to 0^+}\limsup_{n\to\infty} E^*\left[\int_0^\delta \sqrt{\log(N(\overline{p}_{k,1}\mathcal{H}, \tilde{d}_{\mathscr{L}^2(\mathscr{H}_1, \hat{P}_n^1)}, \epsilon)}d\epsilon\right] = 0.$$

Now, by Jensen's inequality we have that for all $h_1, h_2 \in \mathcal{H}$

$$\tilde{d}_{\mathscr{L}^2(\mathscr{H}_1, \hat{P}_n^1)}(\overline{p}_{k,1}h_1, \overline{p}_{k,1}h_2) \leq \tilde{d}_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}(h_1, h_2),$$

which implies that for all $\epsilon > 0$

$$N(\overline{p}_{k,1}\mathcal{H}, \tilde{d}_{\mathscr{L}^2(\mathscr{H}_1, \hat{P}_n^1)}, \epsilon) \leq N(\mathcal{H}, \tilde{d}_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}, \epsilon).$$

Using Theorem 1.11.2 with $q = 2$ and $Q = \hat{P}_n^1 \times P^{k-1}$ and Remark 1.11.1 we see that for all $\epsilon > 0$

$$N(\mathcal{H}, \tilde{d}_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}, \epsilon) \leq c_1 \max\left(1, \left(\frac{\|h_{\mathcal{H}}\|_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}}{\epsilon}\right)^{2c_2}\right).$$

Using that $\log(s) \leq s$ and $\sqrt{c_1 + t} \leq c_1 + \sqrt{t}$ for all $s > 0$ and $t \geq 0$, we obtain that

$$\begin{aligned}\sqrt{\log(N(\overline{p}_{k,1}\mathcal{H}, \tilde{d}_{\mathscr{L}^2(\mathscr{H}_1, \hat{P}_n^1)}, \epsilon)} &\leq \sqrt{c_1 + 2c_2\|h_{\mathcal{H}}\|_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}\epsilon^{-1}} \\ &\leq c_1 + \sqrt{2c_2\|h_{\mathcal{H}}\|_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}}\epsilon^{-1/2}.\end{aligned}$$

Therefore, for all $\delta > 0$,

$$E^*\left[\int_0^\delta \sqrt{\log(N(\overline{p}_{k,1}\mathcal{H}, \tilde{d}_{\mathscr{L}^2(\mathscr{H}_1, \hat{P}_n^1)}, \epsilon)}d\epsilon\right] \leq c_1\delta + \sqrt{8c_2\delta}E[\|h_{\mathcal{H}}\|^{1/2}_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}].$$

We finally notice that by Hölder inequality

$$E[\|h_{\mathcal{H}}\|^{1/2}_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}] \leq (E[\|h_{\mathcal{H}}\|^2_{\mathscr{L}^2(\mathscr{H}_k, \hat{P}_n^1 \times P^{k-1})}])^{1/4} = (\mathcal{J}_k(h^2_{\mathcal{H}}, P^k))^{1/4} < \infty.$$

∎

Theorem 4.9 of Arcones and Giné (1993) provides a version of Theorem 1.11.7 for unbounded classes, which we now state.

Theorem 1.11.8 *Suppose that E is second countable. Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{L}^2(\mathscr{H}_k, P^k)$ be measurable in the sense of Arcones and Giné (1993), P-non-degenerate, and such that $\mathcal{J}_k(h^2_{\mathcal{H}}, P^k) < \infty$. If $\mathcal{H}$ is a VC subgraph class and $\lim_{t\to\infty} tP(\overline{p}_{k,1}h^2_{\mathcal{H}} > t) = 0$, then $\mathcal{H}$ is P-Donsker.*

Notice that, under the assumptions of Theorem 1.11.8, Proposition 1.11.3 yields that $\overline{p}_{k,1}\mathcal{H}$ is P-Donsker. The moment condition $\lim_{t\to\infty} tP(\overline{p}_{k,1}h^2_{\mathcal{H}} > t) = 0$ is used by Arcones and Giné (1993) to obtain (ii) and (iii) of Corollary 1.11.3 yielding that $\mathcal{H}$ is P-Donsker. Assumption (i) of Theorem 4.9 seems unnecessary as $\overline{p}_{k,1}\mathcal{H}$ is already P-pre-Gaussian by Proposition 1.11.3.

We conclude this section with a Bernstein-type inequality for the process $\{W_k(P)(h)\}_{h\in\mathcal{H}}$. This is a uniform version of (iv) of Proposition 1.10.5 and it is given in Theorem 5 of Arcones (1995). Notice that, if $\mathcal{H}$ is uniformly bounded by $0 < l_{\mathcal{H}} < \infty$, then, by Jensen's inequality,

$$\sigma^2_{\mathcal{H}} = \sup_{h\in\mathcal{H}} \mathbb{V}ar[p_{k,1}h(X_1)] \leq \sup_{h\in\mathcal{H}} \mathbb{V}ar[h(X_1, \ldots, X_k)] \leq \sup_{h\in\mathcal{H}} \mathbb{E}[h^2(X_1, \ldots, X_k)] \leq l^2_{\mathcal{H}}.$$

Theorem 1.11.9 *Let $P \in \mathcal{P}(E)$ and $\mathcal{H} \subset \mathscr{H}_k$ be measurable in the sense of Arcones and Giné (1993) and uniformly bounded. If $\mathcal{H}$ is a VC subgraph class, then there are constants*

$1 < c_{\mathcal{H},0}, c_{\mathcal{H},1}, c_{\mathcal{H},2} < \infty$ *such that for all* $t \geq \max(2^3\sigma_{\mathcal{H}}, 2^4 c_{\mathcal{H},0})$

$$\mathbb{P}(\sup_{h\in\mathcal{H}} W_{k,n}(P)(h) \geq t) \leq \sum_{j=1}^{3} M_{\mathcal{H},j}(n,t)$$

where

$$M_{\mathcal{H},1}(n,t) = 8\exp\left(-\frac{\sqrt{n}t^2}{2^{15}k^2(\sqrt{n}\sigma_{\mathcal{H}}^2 + tl_{\mathcal{H}})}\right),$$
$$M_{\mathcal{H},2}(n,t) = 8c_{\mathcal{H},1}^{2c_{\mathcal{H},2}}\left(\sigma_{\mathcal{H}}^2 + \frac{2tl_{\mathcal{H}}}{\sqrt{n}}\right)^{-c_{\mathcal{H},2}}\exp\left(-\left(\frac{n\sigma_{\mathcal{H}}^2}{2l_{\mathcal{H}}^2} + \frac{\sqrt{n}t}{4l_{\mathcal{H}}}\right)\right), \textit{ and}$$
$$M_{\mathcal{H},3}(n,t) = 2\exp\left(-\frac{\sqrt{n}t^2}{2^{6+k}k^{k+1}l_{\mathcal{H}}c_{\mathcal{H},0}(\sqrt{n}\sigma_{\mathcal{H}}^2 + tl_{\mathcal{H}})}\right).$$

Chapter 2

Local depth functions

2.1 Introduction

The notion of local depth, as first described by Agostinelli and Romanazzi (2011), provides a framework to describe the local features of multidimensional distributions. Recently, Francisci et al. (2020) show that several *Type A* depth functions originate from a larger class of functions, which is referred to as *Type A* local depth functions. Interestingly, this class not only contains depth functions but also kernel density estimators. Therefore, it can be used for density estimation and many related applications such as clustering, mode estimation, and upper level set estimation (see Chapter 3 for a through analysis). In this chapter, we provide a thorough analysis of *Type A* local depth functions as contained in Section 2 and Appendices A, B, C, and H of Francisci et al. (2020).

We recall from Definition 1.6.1 that *Type A* depth functions take the form

$$D_G(x,P) = \int h_{G,x,\infty}(x_1,\dots,x_{k_G})dP(x_1)\dots dP(x_{k_G}) \tag{2.1.1}$$

for some Borel measurable, non-negative, and bounded function $h_{G,x,\infty}$ depending on G and x. As we will see below, *Type A* local depth functions are obtained by replacing $h_{G,x,\infty}$ in (2.1.1) with a function $h_{G,x,\tau}$, where $\tau \in [0,\infty]$. Specifically, *Type A* local depth functions take the form

$$L_G(x,\tau,P) = \int h_{G,x,\tau}(x_1,\dots,x_{k_G})dP(x_1)\dots dP(x_{k_G}). \tag{2.1.2}$$

We denote by $\mathcal{H}_G = \{h_{G,x,\tau} : x \in \mathbb{R}^d, \tau \in [0,\infty]\}$ the class of functions yielding L_G. Notice that for $\tau = \infty$ local depth functions coincide with *Type A* depth functions. On the other hand, we will show that, if P is absolutely continuous w.r.t. the Lebesgue measure and τ is small, then the local depth, under appropriate scaling, is close to the density f of P. For this reason, τ is called the localizing parameter. Actually, we restrict our attention to specific classes of functions $\mathcal{H}_G$ that are generated by an appropriate kernel function G. Thus, the index G in (2.1.1) and (2.1.2) can be identified with this kernel function. However, not all *Type A* (local) depth functions can be obtained from such a kernel function (see Examples 2.7.3 and 2.7.4 below).

For simplicial depth and β-skeleton depths, $\mathcal{H}_G$ is a class of indicators of appropriate Borel sets. Specifically, we recall from Definition 1.3.6 that for the simplicial depth $G = S$ and $h_{S,x,\infty} = \mathbf{I}_{Z_{S,x,\infty}}$, where $Z_{S,x,\infty} = \{(x_1,\dots,x_{d+1}) \in (\mathbb{R}^d)^{d+1} : x \in \Delta[x_1,\dots,x_{d+1}]\}$. The local simplicial depth is obtained by replacing $Z_{S,x,\infty}$ with

$$Z_{S,x,\tau} = \{(x_1,\dots,x_{d+1}) \in (\mathbb{R}^d)^{d+1} : x \in \Delta[x_1,\dots,x_{d+1}], \max_{1\le i<j\le d+1} \|x_i - x_j\|_2 \le \tau\}.$$

Similarly, for the local β-skeleton depth (see Example 1.6.1), we take $G = K_\beta$ and $h_{K_\beta,x,\tau} = \mathbf{I}_{Z_{K_\beta,x,\tau}}$, where $\beta \geq 1$ and

$$Z_{K_\beta,x,\tau} = \{(x_1,x_2) \in (\mathbb{R}^d)^2 : \max_{(i,j)\in\{(1,2),(2,1)\}} \|x_i + (2/\beta - 1)x_j - 2/\beta x\|_2 \leq \|x_1 - x_2\|_2 \leq \tau\}.$$

In particular, local spherical depth and local lens depth are obtained by taking $B = K_1$ and $L = K_2$. Thus, local simplicial and β-skeleton depths are obtained by limiting the sets $Z_{G,x,\infty}$, $G = S, K_\beta$ to a smaller region, which is obtained by imposing that the norms $\|x_i - x_j\|_2$, $i,j = 1,\ldots,k_G$, are bounded by τ. Clearly, if $\tau = \infty$ no bound applies and we retrieve simplicial depth and β-skeleton depths. We finally notice that when $d = 1$ local simplicial depth and local β-skeleton depths coincide.

In Section 2.2 we define *Type A* local depth functions (LDFs) and study their properties as the localizing parameter varies. Specifically, we show that, if the localizing parameter tends to infinity, LDFs converge to *Type A* depth functions. On the other hand, if the localizing parameter tends to zero, then, under appropriate scaling, LDFs converge to the k^{th} power of the underlying density f. This suggests that the k^{th} root of scaled LDFs (referred to as τ-approximation) can be used to approximate f and it is therefore called τ-approximation. Indeed, we show in Section 2.3 that the τ-approximation and its derivatives converge uniformly to f. In Section 2.4 we define an appropriate estimator for *Type A* local depth functions, called sample local depth, which takes the form of a U-statistics. Using the theory developed in Sections 1.10 and 1.11 we show that the sample local depth is a uniformly consistent and asymptotically normal estimator of *Type A* LDFs. We also obtain a Bernstein-type inequality for *Type A* LDFs. Next, we use sample local depth functions to obtain an estimate of the τ-approximation (referred to as sample τ-approximation). Using the aforementioned Bernstein-type inequality, we show in Section 2.5 that the sample τ-approximation is a uniformly consistent estimator of the density. In Section 2.6 we determine the correct centering and scaling for the sample τ-approximation to be asymptotically normal. Several examples of *Type A* local depth functions arise both from the depth and kernel density estimator literature and are summarized in Section 2.7. Finally, in Section 2.8 we derive a method for choosing the localizing parameter τ.

2.2 Local depth functions

In this section, we describe in detail *Type A* local depth functions. We begin with some notation. First, the support of a function $g : A \mapsto \mathbb{R}$ is $S_g = \{y \in A : g(y) \neq 0\}$. Next, $L^q((\mathbb{R}^d)^k) = L^q((\mathbb{R}^d)^k, \lambda^k)$, $1 \leq q < \infty$, denote the space of Lebesgue measurable functions $g : (\mathbb{R}^d)^k \to \mathbb{R}$ for which g^q is absolutely integrable,[1] and $L^\infty((\mathbb{R}^d)^k) = L^\infty((\mathbb{R}^d)^k, \lambda^k)$ be the space of Lebesgue measurable functions $g : (\mathbb{R}^d)^k \to \mathbb{R}$ that are essentially bounded. First, we describe the kernel function G. Specifically, we assume that $G : (\mathbb{R}^d)^{k_G} \to [0,\infty)$ and satisfies the following properties:

(P1) G is Borel measurable and $\Lambda_{G,1} = \int G(x_1,\ldots,x_{k_G})dx_1\ldots dx_{k_G} < \infty$,

(P2) $G(v) = G(-v)$ for all $v \in (\mathbb{R}^d)^{k_G}$ and $t \mapsto G(tv)$ is non-increasing on $[0,\infty)$ for all $v \in (\mathbb{R}^d)^{k_G}$,

[1] That is, g is Lebesgue integrable and $\int |g|^q d\lambda^k < \infty$.

(P3) $\sup_{(x_1,\ldots,x_{k_G})\in\prod_{i=1}^{k_G}(\mathbb{R}^d\setminus B_{r_i}(0))} G(x_1,\ldots,x_{k_G}) \to 0$ as $\max_{i=1,\ldots,k_G} r_i \to \infty$, and

(P4) For any $\epsilon > 0$, there exist $0 < \delta \le \epsilon$ and $c_G > 0$ such that $\lambda((\overline{B}_\delta(0))^{k_G} \cap S_G) > 0$ and $G \ge c_G$ in $(\overline{B}_\delta(0))^{k_G} \cap S_G$.

We denote by $\mathcal{G}$ the class of kernel functions $G : (\mathbb{R}^d)^{k_G} \to [0,\infty)$ satisfying the properties **(P1)**-**(P4)**. Some comments are in order. First, **(P1)** ensures that G is absolutely integrable. Second, G is symmetric by the first part of **(P2)**. Third, **(P4)** says that G is positive in a region including the origin of $(\mathbb{R}^d)^{k_G}$ and having positive Lebesgue measure. Fourth, G is non-decreasing along any ray from the origin by the second part of **(P2)**. Finally, **(P3)** entails that G tends to zero as the distance from the origin increases to infinity. In typical examples, such as simplicial and β-skeleton depths, G will have bounded support implying **(P3)**; *i.e.*, for some $\rho > 0$,

$$\overline{S}_G \subset (\overline{B}_\rho(0))^k. \tag{2.2.1}$$

Additionally we assume, without loss of generality (w.l.o.g.), that functions in $\mathcal{G}$ are permutation invariant. If not, apply to $G \in \mathcal{G}$ the symmetrization function $\tilde{\sigma}$ of Remark 1.10.1. We are now ready to give a precise definition of *Type A* local depth function.

Definition 2.2.1 *Let $G \in \mathcal{G}$, $x \in \mathbb{R}^d$, $\tau \in [0,\infty]$, and let $h_{G,x,\tau} : (\mathbb{R}^d)^{k_G} \to [0,\infty)$ be given by $h_{G,x,\tau}(x_1,\ldots,x_{k_G}) = G(\frac{x_1-x}{\tau},\ldots,\frac{x_{k_G}-x}{\tau})$ if $\tau \in (0,\infty)$, $h_{G,x,0}(x_1,\ldots,x_{k_G}) = \lim_{\tau\to 0^+} h_{G,x,\tau}(x_1,\ldots,x_{k_G})$, and $h_{G,x,\infty}(x_1,\ldots,x_{k_G}) = \lim_{\tau\to\infty} h_{G,x,\tau}(x_1,\ldots,x_{k_G})$.
The* Type A *local depth at localization level $\tau \in [0,\infty]$ of a point $x \in \mathbb{R}^d$ with respect to $P \in \mathcal{P}_d$ is given by*

$$L_G(x,\tau,P) = \int h_{G,x,\tau}(x_1,\ldots,x_{k_G})dP(x_1)\ldots dP(x_{k_G}). \tag{2.2.2}$$

We obtain a *Type A* depth function by setting $\tau = \infty$ in (2.2.2), that is, D_G given by $D_G(x,P) = L_G(x,\infty,P)$ is a *Type A* depth function. The kernel functions yielding simplicial depth and β-skeleton depths are given in Examples 2.7.1 and 2.7.2 below. It is worth mentioning that not all $G \in \mathcal{G}$ yield useful depth functions. For instance, if G is continuous at $(0,\ldots,0) \in (\mathbb{R}^d)^{k_G}$, then $h_{G,x,\infty}(x_1,\ldots,x_{k_G}) = G(0,\ldots,0)$ for all $x \in \mathbb{R}^d$ and $(x_1,\ldots,x_{k_G}) \in (\mathbb{R}^d)^{k_G}$. It follows that $D_G(x,P) = G(0,\ldots,0)$ for all $x \in \mathbb{R}^d$ and $P \in \mathcal{P}_d$. In particular, D_G does not satisfy (v) of Definition 1.5.1 and it is not a *statistical* depth function. Typical examples in which G is continuous at the origin of $(\mathbb{R}^d)^{k_G}$ arise from kernel density techniques (see Examples 2.7.5, 2.7.6, and 2.7.7 below).

We discuss now some consequences of Definition 2.2.1 and properties **(P1)**-**(P4)**. We first notice that, by **(P2)**, G is maximized at $(0,\ldots,0) \in (\mathbb{R}^d)^{k_G}$ and let $l_G = G(0,\ldots,0)$. Next, using **(P3)**, we obtain that $h_{G,x,0} = l_G \mathbf{I}_{\{(x,\ldots,x)\}}$ is zero unless all its k_G components equal to $x \in \mathbb{R}^d$. Indeed, if $x_i = x$ for all i, then $G(\frac{x_1-x}{\tau},\ldots,\frac{x_{k_G}-x}{\tau}) = l_G$ for all $\tau > 0$ and $h_{G,x,0}(x_1,\ldots,x_{k_G}) = l_G$. On the other hand, if $x_i \neq x$ for some i, then $r_i = \|(x_i - x)/\tau\|_2 \to \infty$ as $\tau \to 0^+$ and by **(P3)** $h_{G,x,0}(x_1,\ldots,x_{k_G}) = 0$. Also, $h_{G,0,1} = G$, and using **(P1)** and **(P2)**, we see that

$$0 \le h_{G,x,\tau} \le l_G, \tag{2.2.3}$$

for all $x \in \mathbb{R}^d$ and $\tau \in [0,\infty]$, that is, the class $\mathcal{H}_G = \{h_{G,x,\tau} : x \in \mathbb{R}^d, \tau \in [0,\infty]\}$ is uniformly bounded by l_G. Furthermore, **(P4)** ensures that $\Lambda_{G,1} > 0$ and $h_{G,0,\tau}$ is non-trivial for all $\tau > 0$. Indeed, there is a region including $(0,\ldots,0) \in (\mathbb{R}^d)^{k_G}$ where

$h_{G,0,\tau}$ is positive. Since G is permutation invariant, we also have that

$$h_{G,x,\tau}(x_1,\dots,x_{k_G}) = h_{G,x,\tau}(x_{i_1},\dots,x_{i_{k_G}}) \tag{2.2.4}$$

for every permutation $(i_1,\dots,i_{k_G})$ of $(1,\dots,k_G)$. Furthermore, notice that

$$h_{G,x+b,\tau}(x_1+b,\dots,x_{k_G}+b) = h_{G,x,\tau}(x_1,\dots,x_{k_G}), \quad b\in\mathbb{R}^d, \tag{2.2.5}$$

and

$$h_{G,-x,\tau}(-x_1,\dots,-x_{k_G}) = h_{G,x,\tau}(x_1,\dots,x_{k_G}). \tag{2.2.6}$$

If P is absolutely continuous w.r.t. λ with density $f = f_P$, then, by (2.2.5) and (2.2.6), for all $x\in\mathbb{R}^d$ and $\tau\in[0,\infty]$, it holds that

$$\begin{aligned} L_G(x,\tau,P) &= \int h_{G,0,\tau}(x-x_1,\dots,x-x_{k_G})f(x_1)\dots f(x_{k_G})dx_1\dots dx_{k_G} \\ &= (h_{G,0,\tau} * f^{\times k_G})(x,\dots,x), \end{aligned} \tag{2.2.7}$$

where $*$ is the convolution operator and $f^{\times k_G}:(\mathbb{R}^d)^{k_G}\to[0,\infty)$ is the k_G-fold product of identical functions f, that is, $f^{\times k_G}(x_1,\dots x_{k_G}) = f(x_1)\dots f(x_{k_G})$.

When there is no scope for confusion we suppress the subscript G. Hence, we also write e.g. k for k_G, $h_{x,\tau}$ for $h_{G,x,\tau}$, Λ_1 for $\Lambda_{G,1}$. Since P is fixed in the following, we write $D_G(x)$ for $D_G(x,P)$ and $L_G(x,\tau)$ for $L_G(x,\tau,P)$. Also, for $j=1,\dots,d$, we denote by $\partial_j g$ the partial derivative of the function $g:\mathbb{R}^d\to\mathbb{R}$ with respect to its j^{th}-component. Our first proposition summarizes several continuity and differentiability properties of the LDFs. Specifically, the behavior of the LDFs when $\tau\to 0^+$ and $\tau\to\infty$ are provided.

Proposition 2.2.1 *(i) For all $x\in\mathbb{R}^d$, $L_G(x,\cdot)$ is monotonically non-decreasing with*

$$\lim_{\tau\to 0^+} L_G(x,\tau) = l_G(P(\{x\}))^k \textit{ and } \lim_{\tau\to\infty} L_G(x,\tau) = D_G(x).$$

(ii) For $\tau\in[0,\infty)$, $\lim_{r\to\infty}\sup_{x\in\mathbb{R}^d\setminus B_r(0)} L_G(x,\tau) = 0$.
(iii) If $P\in\mathcal{P}_{d,\ll\lambda}$, then, for all $\tau\in[0,\infty)$, $L_G(\cdot,\tau)$ is bounded and continuous.
(iv) Under assumption (2.2.1), if $P\in\mathcal{P}_{d,\ll\lambda}$ with m-times continuously differentiable density f, then, for all $\tau\in[0,\infty)$, $L_G(\cdot,\tau)$ is m-times continuously differentiable and, for $i_1,\dots,i_m\in\{1,\dots,d\}$,

$$\partial_{i_m}\dots\partial_{i_1}L_G(x,\tau) = (h_{0,\tau} * (\partial_{i_m}\dots\partial_{i_1} f^{\times k}))(x,\dots,x). \tag{2.2.8}$$

When $\tau=\infty$, part (ii) does not hold in general. Indeed, D_G does not always satisfy (v) of Definition 1.5.1.

Proof of Proposition 2.2.1. We start by proving (i). For the monotonicity, observe that, by Definition 2.2.1 and **(P2)**, for all $x\in\mathbb{R}^d$, $(x_1,\dots,x_k)\in(\mathbb{R}^d)^k$ and $0\le\tau_1\le\tau_2\le\infty$, $h_{x,\tau_1}(x_1,\dots,x_k)\le h_{x,\tau_2}(x_1,\dots,x_k)$ and therefore $L_G(x,\tau_1)\le L_G(x,\tau_2)$. Using Lebesgue dominated convergence Theorem (LDCT) and Definition 2.2.1, we get that

$$\lim_{\tau\to 0^+} L_G(x,\tau) = \int \lim_{\tau\to 0^+} h_{x,\tau}(x_1,\dots,x_k)dP(x_1)\dots dP(x_k) = l_G(P(\{x\}))^k$$

and

$$\lim_{\tau\to\infty} L_G(x,\tau) = \int \lim_{\tau\to\infty} h_{x,\tau}(x_1,\dots,x_k)dP(x_1)\dots dP(x_k) = D_G(x).$$

For (ii) let $\tau \in [0,\infty)$, $0 < \epsilon < 1$, and $r_1 > 0$ such that $P(\overline{B}_{r_1}(0)) \geq 1-\epsilon$. By **(P3)**, there exists $\tilde{r} > 0$ such that, if $x_1 \in \mathbb{R}^d \setminus \overline{B}_{\tau\tilde{r}}(x)$, then $h_{x,\tau}(x_1,\dots,x_k) \leq \epsilon$ for all $(x_2,\dots,x_k) \in (\mathbb{R}^d)^{k-1}$. Since, for $r_2 > \tau\tilde{r}$ and $x \in \mathbb{R}^d \setminus B_{r_1+r_2}(0)$, it holds that $\overline{B}_{\tau\tilde{r}}(x) \subset \mathbb{R}^d \setminus \overline{B}_{r_1}(0)$, using (2.2.3), we conclude that, for all $r \geq r_1 + r_2$,

$$\begin{aligned}
\sup_{x\in\mathbb{R}^d\setminus B_r(0)} L_G(x,\tau) \leq & \sup_{x\in\mathbb{R}^d\setminus B_{r_1+r_2}(0)} \int_{\overline{B}_{\tau\tilde{r}}(x)\times(\mathbb{R}^d)^{k-1}} h_{x,\tau}(x_1,\dots,x_k)dP(x_1)\dots dP(x_k) \\
& + \sup_{x\in\mathbb{R}^d\setminus B_{r_1+r_2}(0)} \int_{(\mathbb{R}^d\setminus\overline{B}_{\tau\tilde{r}}(x))\times(\mathbb{R}^d)^{k-1}} h_{x,\tau}(x_1,\dots,x_k)dP(x_1)\dots dP(x_k) \\
& \leq l_G \sup_{x\in\mathbb{R}^d\setminus B_{r_1+r_2}(0)} P(\overline{B}_{\tau\tilde{r}}(x)) + \epsilon \\
& \leq l_G P(\mathbb{R}^d \setminus \overline{B}_{r_1}(0)) + \epsilon \\
& \leq (l_G + 1)\epsilon.
\end{aligned}$$

We now prove (iii). Let f be the density function of P with respect to λ. By (2.2.3), we have that

$$L_G(x,\tau) \leq l_G \int f(x_1),\dots,f(x_k)dx_1\dots dx_k = l_G,$$

which shows that $L_G(\cdot,\tau)$ is bounded. Furthermore, by (2.2.7) and (2.2.5), it holds that

$$\begin{aligned}
|L_G(y,\tau) - L_G(x,\tau)| = & \Bigg| \int h_{0,\tau}(x_1,\dots,x_k)\prod_{j=1}^k f(y-x_j)dx_1\dots dx_k \\
& - \int h_{0,\tau}(x_1,\dots,x_k)\prod_{j=1}^k f(x-x_j)dx_1\dots dx_k \Bigg| \\
& \leq l_G \int \Bigg| \prod_{j=1}^k f(y-x_j) - \prod_{j=1}^k f(x-x_j) \Bigg| dx_1\dots dx_k.
\end{aligned}$$

By Theorem 8.19 in Wheeden and Zygmund (2015), it follows that $|L_G(y,\tau) - L_G(x,\tau)|$ converges to 0 as $\|y-x\|_2 \to 0$.

We turn to the proof of (iv). We first observe that, by (iii) and (2.2.7), (iv) holds when $m = 0$. Also, if $\tau = 0$ then $L_G(x,\tau) = 0$ for all $x \in \mathbb{R}^d$ and the statement is trivial. Let $\tau > 0$ and $m \geq 1$. We will show that, for all $0 \leq j \leq m$, the partial derivatives of $L_G(\cdot,\tau)$ up to order j exist and are given by

$$\partial_{i_j}\dots\partial_{i_1} L_G(x,\tau) = (h_{0,\tau} * g_{i_j,\dots,i_1})(x,\dots,x), \tag{2.2.9}$$

where, for $(x_1,\dots,x_k) \in (\mathbb{R}^d)^k$, $g_{i_j,\dots,i_1}(x_1,\dots,x_k) = \partial_{i_j}\dots\partial_{i_1} f(x_1)\dots f(x_k)$. In particular, since f is m-times continuously differentiable, $g_{i_j,\dots,i_1}$ is $(m-j)$-times continuously differentiable. For $h > 0$ and $i \in \{1,\dots,d\}$, we define the i^{th} partial finite difference of a function $\tilde{g} : \mathbb{R}^d \to \mathbb{R}$ by

$$\partial_i^h \tilde{g}(x) = \frac{\tilde{g}(x+he_i) - \tilde{g}(x)}{h}.$$

Suppose by induction that the partial derivatives of the local depth up to order $j-1$ $(1 \leq j \leq m)$ exist and, for some choice of indices $i_1,\dots,i_{j-1} \in \{1,\dots,d\}$ are given by (2.2.9). Let $i_j \in \{1,\dots,d\}$. Then by (2.2.9) and the mean value theorem, there exists

$0 \le c \le 1$, such that

$$\begin{aligned}
&\partial^h_{i_j}\partial_{i_{j-1}}\dots\partial_{i_1}L_G(x,\tau)\\
&=\int h_{0,\tau}(x_1,\dots,x_k)\partial^h_{i_j}g_{i_{j-1},\dots,i_1}(x-x_1,\dots,x-x_k)dx_1\dots dx_k\\
&=\int h_{0,\tau}(x_1,\dots,x_k)\partial_{i_j}g_{i_{j-1},\dots,i_1}(x-x_1,\dots,x+che_{i_j}-x_{i_j},\dots,x-x_k)dx_1\dots dx_k.
\end{aligned} \tag{2.2.10}$$

Notice that, by (2.2.1), $\overline{S}_{h_{0,\tau}} \subset (\overline{B}_{\tau\rho}(0))^k$; by (2.2.3), $h_{0,\tau}$ is bounded; and finally $\partial_{i_j}g_{i_{j-1},\dots,i_1} = g_{i_j,\dots,i_1}$ is $(m-j)$-times continuously differentiable (in particular, continuous). By taking the limit for $h \to 0^+$ in (2.2.10) and using LDCT, we get (2.2.9). By induction, (2.2.8) follows. To conclude, we show that $\partial_{i_m}\dots\partial_{i_1}L_G(\cdot,\tau)$ is continuous. By (2.2.5) and (2.2.3), we have that, for $x, y \in \mathbb{R}^d$,

$$\begin{aligned}
&|\partial_{i_m}\dots\partial_{i_1}L_G(y,\tau)-\partial_{i_m}\dots\partial_{i_1}L_G(x,\tau)|\\
=&\Big|\int h_{0,\tau}(x_1,\dots,x_k)g_{i_m,\dots,i_1}(y-x_1,\dots,y-x_k)dx_1\dots dx_k\\
&-\int h_{0,\tau}(x_1,\dots,x_k)g_{i_m,\dots,i_1}(x-x_1,\dots,x-x_k)dx_1\dots dx_k\Big|\\
\le& l_G\int_{\overline{S}_{h_{0,\tau}}}|g_{i_m,\dots,i_1}(y-x_1,\dots,y-x_k)-g_{i_m,\dots,i_1}(x-x_1,\dots,x-x_k)|dx_1\dots dx_k.
\end{aligned}$$

Since $\overline{S}_{h_{0,\tau}}$ is compact by (2.2.1), the continuity follows from the uniform continuity of $g_{i_m,\dots,i_1}$ over compact sets. ■

If $P \in \mathcal{P}_{d,\ll\lambda}$ with density f, then Proposition 2.2.1 (i) entails that

$$\lim_{\tau\to 0^+} L_G(x,\tau) = 0.$$

A natural question is whether a suitable scaling yields a non-trivial limit and how it relates with f. The behavior of L_G for $\tau \to 0^+$, that is, under extreme localization, is investigate in our next result. Specifically, we show there that, under appropriate conditions, scaled versions of LDFs converge pointwise, uniformly, and in $L^d(\mathbb{R}^d)$ to the k^{th} power of the density f. Also, part (iii) of the following theorem is concerned with the rate of convergence of scaled LDFs. It is worth noticing that, under the assumption (2.2.1), for all $x \in \mathbb{R}^d \setminus \overline{S}_f$, $f^k(x) = 0$ and $\tau^{-kd}L_G(x,\tau) = 0$ for small values of τ.

Theorem 2.2.1 *Let $P \in \mathcal{P}_{d,\ll\lambda}$ with density f.*
(i) Under assumption (2.2.1) at every point x of continuity of f, it holds that

$$\lim_{\tau\to 0^+} \tau^{-kd}\Lambda_1^{-1}L_G(x,\tau) = f^k(x). \tag{2.2.11}$$

Furthermore, (2.2.11) holds uniformly on any set where f is uniformly continuous.
(ii) If $f \in L^\infty(\mathbb{R}^d)$, then (2.2.11) holds at every point of continuity of f and the convergence in (2.2.11) is uniform on any set where f is uniformly continuous.
(iii) Let f be twice continuously differentiable. Then, under assumption (2.2.1), for all $x \in S_f$, it holds that

$$\lim_{\tau\to 0^+} \tau^{-2}\left(\tau^{-kd}L_G(x,\tau)-\Lambda_1 f^k(x)\right) = R_1(x)+R_2(x),$$

where

$$R_1(x) = \frac{k}{2} f^{k-1}(x) \int h_{0,1}(x_1, \ldots, x_k) x_1^\top H_f(x) x_1 dx_1 \ldots dx_k \quad \textit{and}$$
$$R_2(x) = \frac{k(k-1)}{2} f^{k-2}(x) \int h_{0,1}(x_1, \ldots, x_k) \langle \nabla f(x), x_1 \rangle \langle \nabla f(x), x_2 \rangle dx_1 \ldots dx_k.$$

(iv) If $f^k \in L^q(\mathbb{R}^d)$, $1 \le q < \infty$, then $\tau^{-kd} \Lambda_1^{-1} L_G(\cdot, \tau)$ converges in $L^q(\mathbb{R}^d)$ to f^k.

Before we prove Theorem 2.2.1 we state a couple of lemmas concerning the approximation of the identity for the function G (see Section 9.2 in Wheeden and Zygmund (2015) and Section XIII.2 in Torchinsky (1995)).

Lemma 2.2.1 *Let $\tilde{G}_\tau = \tau^{-kd} \Lambda_1^{-1} h_{0,\tau}$. Then the following hold:*

(i) $\displaystyle\int \tilde{G}_\tau(x_1, \ldots, x_k) dx_1 \ldots dx_k = 1.$

(ii) For all $\delta > 0$, $\displaystyle\lim_{\tau \to 0^+} \int_{(\mathbb{R}^d)^k \setminus (\overline{B}_\delta(0))^k} \tilde{G}_\tau(y_1, \ldots, y_k) dy_1 \ldots y_k = 0.$

Proof of Lemma 2.2.1. (i) follows from a change of variable. Similarly, for (ii), we have that

$$\begin{aligned} \int_{(\mathbb{R}^d)^k \setminus (\overline{B}_\delta(0))^k} \tilde{G}_\tau(y_1, \ldots, y_k) dy_1 \ldots dy_k &= \tau^{-kd} \Lambda_1^{-1} \int_{(\mathbb{R}^d)^k \setminus (\overline{B}_\delta(0))^k} G\left(\frac{y_1}{\tau}, \ldots, \frac{y_k}{\tau}\right) dy_1 \ldots dy_k \\ &= \Lambda_1^{-1} \int_{(\mathbb{R}^d)^k \setminus (\overline{B}_{\delta/\tau}(0))^k} G(y_1, \ldots, y_k) dy_1 \ldots dy_k, \end{aligned}$$

and the last term converges to zero as $\tau \to 0^+$ by (**P1**). ■

Lemma 2.2.2 *Let $\tilde{f} : (\mathbb{R}^d)^k \to \mathbb{R}^d$ be Lebesgue measurable and suppose that assumption (2.2.1) holds true. Then, at every point $(x_1, \ldots, x_k) \in (\mathbb{R}^d)^k$ of continuity of $\tilde{f}$*

$$\lim_{\tau \to 0^+} (\tilde{G}_\tau * \tilde{f})(x_1, \ldots, x_k) = \tilde{f}(x_1, \ldots, x_k). \tag{2.2.12}$$

Furthermore, (2.2.12) holds uniformly on any set $A \subset (\mathbb{R}^d)^k$ where $\tilde{f}$ is uniformly continuous.

Proof of Lemma 2.2.2. Using Lemma 2.2.1 we obtain that

$$\begin{aligned} &|(\tilde{G}_\tau * \tilde{f})(x_1, \ldots, x_k) - \tilde{f}(x_1, \ldots, x_k)| \\ &\le \int |\tilde{f}(x_1 - y_1, \ldots, x_k - y_k) - \tilde{f}(x_1, \ldots, x_k)| \tilde{G}_\tau(y_1, \ldots, y_k) dy_1 \ldots dy_k. \end{aligned}$$

Now, (2.2.1) yields that $\overline{S}_{\tilde{G}_\tau} \subset (\overline{B}_{\tau\rho}(0))^k$ and, if $\tilde{f}$ is continuous at $(x_1, \ldots, x_k) \in (\mathbb{R}^d)^k$, then, for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$|\tilde{f}(x_1 - y_1, \ldots, x_k - y_k) - \tilde{f}(x_1, \ldots, x_k)| \le \epsilon, \tag{2.2.13}$$

for all $(y_1, \ldots, y_k) \in (\overline{B}_\delta(0))^k$. Therefore, for $0 < \tau \le \delta/\rho$, it holds that

$$|(\tilde{G}_\tau * \tilde{f})(x_1, \ldots, x_k) - \tilde{f}(x_1, \ldots, x_k)| \le \epsilon. \tag{2.2.14}$$

Finally, if $\tilde{f}$ is uniformly continuous on A, then (2.2.13) and (2.2.14) hold for all $(x_1,\dots,x_k) \in A$. ∎

Proof of Theorem 2.2.1. We begin by proving (i). To this end, notice that, if f is continuous at $x \in \mathbb{R}^d$, then $f^{\times k}$ is continuous at $(x,\dots,x) \in (\mathbb{R}^d)^k$. Similarly, if f is uniformly continuous in $A \subset \mathbb{R}^d$, then $f^{\times k}$ is uniformly continuous in $(A)^k \subset (\mathbb{R}^d)^k$. Now, the result follows from (2.2.7) and Lemma 2.2.2 with $\tilde{f} = f^{\times k}$.

We now prove (ii). We first notice that, since $f \in L^\infty(\mathbb{R}^d)$, there exists a constant $1 \le c < \infty$ such that $f \le c$ λ-a.e. In particular, for all $1 \le q < \infty$, it holds that $f^q \le c^q$ λ-a.e., implying that $f^q \in L^\infty(\mathbb{R}^d)$. Then, we compute

$$\begin{aligned} |\frac{L_G(x,\tau)}{\Lambda_1\tau^{kd}} - f^k(x)| &= \left|\int \prod_{j=1}^{k} f(x-x_j)\tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k - f^k(x)\right| \\ &\le \int \left|\prod_{j=1}^{k} f(x-x_j) - f^k(x)\right| \tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k. \end{aligned} \tag{2.2.15}$$

Next, we recursively apply the triangle inequality and obtain

$$\begin{aligned} \left|\prod_{j=1}^{k} f(x-x_j) - f^k(x)\right| &= |\sum_{i=1}^{k}\prod_{j=1}^{i-1} f(x-x_j)(f(x-x_i)-f(x))f^{k-i}(x)| \\ &\le \sum_{i=1}^{k}\prod_{j=1}^{i-1} f(x-x_j)|f(x-x_i)-f(x)|f^{k-i}(x), \end{aligned} \tag{2.2.16}$$

thus implying that the right hand side (RHS) of (2.2.15) is bounded above by

$$\begin{aligned} &\sum_{i=1}^{k}\int \prod_{j=1}^{i-1} f(x-x_j)|f(x-x_i)-f(x)|f^{k-i}(x)\tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k \\ \le & c^{k-1}\sum_{i=1}^{k}\int |f(x-x_i)-f(x)|\tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k. \end{aligned}$$

Now, by Lemma 2.2.1 (ii), for all $\delta > 0$ there exists $\tilde{\tau}(\delta) > 0$ such that, for all $0 < \tau \le \tilde{\tau}(\delta)$,

$$\int_{(\mathbb{R}^d)^k \setminus (\overline{B}_\delta(0))^k} \tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots x_k \le \epsilon. \tag{2.2.17}$$

If $x \in \mathbb{R}^d$ is a continuity point for f, then for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$|f(x-y) - f(x)| \le \epsilon, \tag{2.2.18}$$

for all $y \in \overline{B}_\delta(0)$. Using Lemma 2.2.1(i), (2.2.17), and (2.2.18), we conclude that, for all $0 < \tau \le \tilde{\tau}(\delta)$,

$$\begin{aligned} |\frac{L_G(x,\tau)}{\Lambda_1\tau^{kd}} - f^k(x)| \le {} & c^{k-1}\sum_{i=1}^{k}\int_{(\overline{B}_\delta(0))^k} |f(x-x_i)-f(x)|\tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k \\ & + c^{k-1}\sum_{i=1}^{k}\int_{(\mathbb{R}^d)^k \setminus (\overline{B}_\delta(0))^k} |f(x-x_i)-f(x)|\tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k \\ & \le kc^{k-1}(1+2c)\epsilon. \end{aligned} \tag{2.2.19}$$

Finally, if f is uniformly continuous on $A \subset \mathbb{R}^d$, then (2.2.18) and (2.2.19) hold for all $x \in A$.

For (iii), notice that, by (2.2.6) and a change of variable in (2.2.7),

$$\tau^{-kd} L_G(x,\tau) - \Lambda_1 f^k(x) = \int h_{0,1}(x_1,\dots,x_k)\left[\prod_{j=1}^{k} f(x-\tau x_j) - f^k(x)\right] dx_1 \dots dx_k. \tag{2.2.20}$$

Since f is twice continuously differentiable, by multivariate Taylor's theorem with integral remainder, for $i = 1,\dots,k$,

$$f(x+\tau x_i) = f(x) + \tau\langle \nabla f(x), x_i\rangle + \tau^2 \int_0^1 (1-z) x_i^\top H_f(x+\tau z x_i) x_i dz.$$

Therefore,

$$\begin{aligned} f(x+\tau x_1)\dots f(x+\tau x_k) = f^k(x) &+ \tau f^{k-1}(x)\langle \nabla f(x), \sum_{i=1}^{k} x_i\rangle \\ &+ \tau^2 f^{k-1}(x) \sum_{i=1}^{k} \int_0^1 (1-z) x_i^\top H_f(x+z\tau x_i) x_i dz \\ &+ \tau^2 f^{k-2}(x) \sum_{i=1}^{k} \sum_{j=i+1}^{k} \langle \nabla f(x), x_i\rangle\langle \nabla f(x), x_j\rangle + O(\tau^2). \end{aligned} \tag{2.2.21}$$

Since S_f is open, there exist $\tilde{\tau} > 0$ such that, for all $\tau \in [0,\tilde{\tau}]$, $(x_1,\dots,x_k) \in \overline{S}_{h_{0,1}}$ and $z \in [0,1]$, $x + z\tau x_i \in S_f$. The continuity of the second order partial derivatives implies that, for $\tau \in [0,\tilde{\tau}]$, the functions

$$(x_1,\dots,x_k) \mapsto \int_0^1 (1-z) x_i^\top H_f(x+z\tau x_i) x_i dz$$

are continuous (and uniformly bounded for all $\tau \in [0,\tilde{\tau}]$) with

$$\lim_{\tau\to 0^+} \int_0^1 (1-z) x_i^\top H_f(x+z\tau x_i) x_i dz = \frac{1}{2} x_i^\top H_f(x) x_i. \tag{2.2.22}$$

Similarly, the remainder $O(\tau^2)$ is uniformly bounded for all $\tau \in [0,\tilde{\tau}]$ and continuous with respect to $(x_1,\dots,x_k)$. By substituting (2.2.21) in (2.2.20), we see that

$$\begin{aligned} \tau^{-kd} L_G(x,\tau) - \Lambda_1 f^k(x) &= \tau f^{k-1}(x) \int h_{0,1}(x_1,\dots,x_k)\langle \nabla f(x), \sum_{i=1}^{k} x_i\rangle dx_1 \dots dx_k \\ &+ \tau^2 f^{k-1}(x) \int h_1(0;x_1,\dots,x_k)\left[\sum_{i=1}^{k} \int_0^1 (1-z) x_i^\top H_f(x+z\tau x_i) x_i dz\right] dx_1 \dots dx_k \\ &+ \tau^2 f^{k-2}(x) \int h_1(0;x_1,\dots,x_k)\left[\sum_{i=1}^{k} \sum_{j=i+1}^{k} \langle \nabla f(x), x_i\rangle\langle \nabla f(x), x_j\rangle\right] dx_1 \dots dx_k + O(\tau^2). \end{aligned}$$

By (2.2.6) and the change of variable $-(x_1,\dots,x_k)$ for $(x_1,\dots,x_k)$, it follows that

$$\int h_{0,1}(x_1,\dots,x_k)\langle \nabla f(x), \sum_{i=1}^{k} x_i\rangle dx_1\dots dx_k$$
$$= -\int h_{0,1}(x_1,\dots,x_k)\langle \nabla f(x), \sum_{i=1}^{k} x_i\rangle dx_1\dots dx_k.$$

Therefore,

$$\int h_{0,1}(x_1,\dots,x_k)\langle \nabla f(x), \sum_{i=1}^{k} x_i\rangle dx_1\dots dx_k = 0.$$

Now, using (2.2.4), we obtain that

$$\int h_{0,1}(x_1,\dots,x_k)\left[\sum_{i=1}^{k}\int_0^1 (1-z)x_i^{\top} H_f(x+z\tau x_i)x_i dz\right] dx_1\dots dx_k$$
$$=k\int h_{0,1}(x_1,\dots,x_k)\left[\int_0^1 (1-z)x_1^{\top} H_f(x+z\tau x_1)x_1 dz\right] dx_1\dots dx_k$$

and

$$\int h_{0,1}(x_1,\dots,x_k)\left[\sum_{i=1}^{k}\sum_{j=i+1}^{k}\langle\nabla f(x),x_i\rangle\langle\nabla f(x),x_j\rangle\right] dx_1\dots dx_k$$
$$=\frac{k(k-1)}{2}f^{k-2}(x)\int h_{0,1}(x_1,\dots,x_k)\langle\nabla f(x),x_1\rangle\langle\nabla f(x),x_2\rangle dx_1\dots dx_k.$$

By (2.2.22) and LDCT, we conclude that

$$\lim_{\tau\to 0^+}\tau^{-2}(\tau^{-kd}L_G(x,\tau)-\Lambda_1 f^k(x)) = R_1(x)+R_2(x).$$

We now prove (iv). We first notice that, since $f\in L^1(\mathbb{R}^d)\cap L^{kq}(\mathbb{R}^d)$, then $f\in L^q(\mathbb{R}^d)$, for all $1\le q\le kd$. Indeed, it holds that

$$\int_{\mathbb{R}^d} f^q(x)dx = \int_{\{y\in\mathbb{R}^d: f(y)<1\}} f^q(x)dx + \int_{\{y\in\mathbb{R}^d: f(y)\ge 1\}} f^q(x)dx$$
$$\le \int_{\{y\in\mathbb{R}^d: f(y)<1\}} f(x)dx + \int_{\{y\in\mathbb{R}^d: f(y)\ge 1\}} f^{kd}(x)dx < \infty.$$

Next, we write in (2.2.15) $\tilde{G}_\tau(x_1,\dots,x_k) = \tilde{G}_\tau^{1/q}(x_1,\dots,x_k)\tilde{G}_\tau^{1/\tilde{q}}(x_1,\dots,x_k)$, where $\tilde{q}\in(1,\infty]$ satisfies $1/q+1/\tilde{q}=1$ ($\tilde{q}=\infty$ if $q=1$), apply Hölder inequality with exponents q and $\tilde{q}$ and Lemma 2.2.1 (i), thus obtaining

$$|\frac{L_G(x,\tau)}{\Lambda_1\tau^{kd}} - f^k(x)|^q \le \int\left|\prod_{j=1}^{k} f(x-x_j) - f^k(x)\right|^q \tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k.$$

Now, Jensen's inequality yields that $(\sum_{i=1}^k a_i)^q \le k^{q-1}\sum_{i=1}^k a_i^q$ for $a_i\ge 0$. Using this and (2.2.16), we obtain that

$$\int|\frac{L_G(x,\tau)}{\Lambda_1\tau^{kd}} - f^k(x)|^q dx \le k^{q-1}\sum_{i=1}^{k} I_{\tau,i},$$

where

$$I_{\tau,i} = \int \left(\int \left(\prod_{j=1}^{i-1} f^q(x-x_j)|f(x-x_i)-f(x)|^q f^{(k-i)q}(x) \right) \tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k \right) dx.$$

Notice that $I_{\tau,i}$ are finite since $f \in L^q(\mathbb{R}^d)$, $1 \le q \le kd$, and, by (2.2.3), $0 \le \tilde{G}_\tau \le l_G/\Lambda_1\tau^{kd}$. By Fubini's theorem, we have that

$$I_{\tau,i} = \int J_{\tau,i}(x_1,\dots,x_k)\tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k,$$

where

$$J_{\tau,i}(x_1,\dots,x_k) = \int \prod_{j=1}^{i-1} f^q(x-x_j)|f(x-x_i)-f(x)|^q f^{(k-i)q}(x)dx.$$

Now, we apply again Hölder inequality with exponents $s = k/(k-1)$ and $t = k$, and see that

$$\begin{aligned} J_{\tau,i}(x_1,\dots,x_k) &\le \left[\int \prod_{j=1}^{i-1} f^{sq}(x-x_j)f^{(k-i)sq}(x)dx \right]^{1/s} \left[\int |f(x-x_i)-f(x)|^{tq}dx \right]^{1/t} \\ &\le c_1 K(x_1,\dots,x_k), \end{aligned}$$

where

$$c_1 = \max_{i=1,\dots,q} \left[\int \prod_{j=1}^{i-1} f^{sq}(x-x_j)f^{(k-i)sq}(x)dx \right]^{1/s}$$

and

$$K(x_1,\dots,x_k) = \max_{i=1,\dots,q} \left[\int |f(x-x_i)-f(x)|^{tq}dx \right]^{1/t}.$$

Notice that

$$K(x_1,\dots,x_k) \le c_2 = 2^q \left[\int f(x)^{tq}dx \right]^{1/t} < \infty, \tag{2.2.23}$$

and, for all $\epsilon > 0$, by Theorem 8.19 in Wheeden and Zygmund (2015), there exists $\delta > 0$ such that, for all $(x_1,\dots,x_k) \in (\overline{B}_\delta(0))^k$,

$$K(x_1,\dots,x_k) \le \epsilon. \tag{2.2.24}$$

Using Lemma 2.2.1 (i), (2.2.23), (2.2.24), and (2.2.17), we conclude that, for all $0 < \tau \le \tilde{\tau}(\delta)$,

$$\begin{aligned} \int \left|\frac{L_G(x,\tau)}{\Lambda_1\tau^{kd}} - f^k(x)\right|^q dx &\le c_1 k^q \int K(x_1,\dots,x_k)\tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k \\ &\le c_1(1+c_2)k^q\epsilon. \end{aligned}$$

■

When $d = 1$ certain simplifications occur in Theorem 2.2.1. Specifically, for $G = L, S, B, K_\beta$, $\Lambda_{G,1} = 1$ and the functions R_1 and R_2 in (iii) take a simpler form. This is summarized in the following corollary.

Corollary 2.2.1 *Let $G = L, S, B, K_\beta$, $d = 1$, and $P \in \mathcal{P}_{1,\ll\lambda}$ with density f. It holds that*

$$L_G(x,\tau) = 2\int_{T_{\tau,++}} f(x+x_1)f(x-x_2)dx_1dx_2, \tag{2.2.25}$$

where $T_{\tau,++} = \{(x_1,x_2) : x_1 \geq 0, x_2 \geq 0, x_1+x_2 \leq \tau\}$. Furthermore, we have that (i) at every point x of continuity of f

$$\lim_{\tau\to 0^+} \tau^{-2}L_G(x,\tau) = f^2(x), \tag{2.2.26}$$

and (2.2.26) holds uniformly on any set where f is uniformly continuous.
(ii) If $f \in L^\infty(\mathbb{R})$, then (2.2.26) holds at every point of continuity of f and the convergence in (2.2.26) is uniform on any set where f is uniformly continuous.
(iii) If f is twice continuously differentiable, then

$$\lim_{\tau\to 0^+} \tau^{-2}(\tau^{-2}L_G(x,\tau) - f^2(x)) = \frac{1}{12}\left[2f(x)f''(x) + [f'(x)]^2\right].$$

(iv) If $f^2 \in L^q(\mathbb{R})$, $1 \leq q < \infty$, then $\tau^{-2}L_G(\cdot,\tau)$ converges in $L^q(\mathbb{R})$ to f^2.

Proof of Corollary 2.2.1. By a change of variable, it follows that

$$\begin{aligned} L_G(x,\tau) &= \int_{Z_{G,x,\tau}} f(x_1)f(x_2)dx_1dx_2 \\ &= \int_{Z_{G,0,\tau}} f(x+x_1)f(x+x_2)dx_1dx_2. \end{aligned} \tag{2.2.27}$$

In two dimensions $Z_{G,0,\tau}$ can be expressed as the union of two triangles $T_{\tau,-+}$ and $T_{\tau,+-}$; that is,

$$\begin{aligned} T_{\tau,-+} &= \{(x_1,x_2) \in \mathbb{R}^2 : x_1 \leq 0, x_2 \geq 0, x_2 - x_1 \leq \tau\} \\ T_{\tau,+-} &= \{(x_1,x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \leq 0, x_1 - x_2 \leq \tau\}. \end{aligned}$$

Now, by a change of variables in the integrals over the triangles it follows that

$$\begin{aligned} L_G(x,\tau) &= \int_{T_{\tau,-+}} f(x+x_1)f(x+x_2) + f(x-x_1)f(x-x_2)dx_1dx_2 \\ &= \int_{T_{\tau,++}} f(x-x_1)f(x+x_2) + f(x+x_1)f(x-x_2)dx_1dx_2 \\ &= 2\int_{T_{\tau,++}} f(x+x_1)f(x-x_2)dx_1dx_2. \end{aligned}$$

(i), (ii) and (iv) follows directly from Theorem 2.2.1 (i), (ii) and (iv) and the fact that $Z_{G,0,1}$ is bounded with area $\Lambda_{G,1} = 1$. Finally, (iii) follows from Theorem 2.2.1 (iii), where

$$\begin{aligned} R_1(x) &= 2f(x)f''(x)\int_0^1\int_0^{1-x_1} x_1^2 dx_1dx_2 = \frac{1}{6}f(x)f''(x), \quad \text{and} \\ R_2(x) &= 2f'(x)]^2\int_0^1\int_0^{1-x_1} x_1x_2dx_1dx_2 = \frac{1}{12}[f'(x)]^2. \end{aligned}$$

∎

In the next section, we will use Theorem 2.2.1 to approximate the density f. Specifically, scaled LDFs converge to f^k and, by taking the k^{th} root, we obtain an approximation of the density. This leads to the concept of τ-approximation.

2.3 τ-approximation

Using (2.2.11) one can express f in terms of the limit of LDFs, for a given choice of G. This leads to an important idea, namely the τ-approximation. This approximation is useful since in applications it enables one to provide alternate approaches for density estimation. In Chapter 3, we will illustrate this idea in three distinct but related contexts; *viz.* clustering, estimation of mode, and estimation of upper level sets.

Definition 2.3.1 (τ-approximation) *For any $\tau > 0$,*

$$f_{G,\tau}(x) = \left(\frac{L_G(x,\tau)}{\tau^{kd}\Lambda_1}\right)^{1/k}. \tag{2.3.1}$$

Remark 2.3.1 *From Proposition 2.2.1 (iii), it follows that when P has a density f then, $f_{G,\tau}$ is continuous. Additionally, under the assumptions of Proposition 2.2.1 (iv) we have that $f_{G,\tau}$ is m-times continuously differentiable in $S_{f_{G,\tau}}$.*

We begin by studying the properties of $S_{f_{G,\tau}}$. First, we recall the definition of limits of sets below. The limit inferior and superior of a sequence of sets $\{A_n\}_{n=1}^{\infty}$ are $\liminf_{n\to\infty} A_n = \cup_{n=1}^{\infty} \cap_{l=n}^{\infty} A_l$ and $\limsup_{n\to\infty} A_n = \cap_{n=1}^{\infty} \cup_{l=n}^{\infty} A_l$. If they are equal we say that the sequence $\{A_n\}_{n=1}^{\infty}$ converges and write $A = \lim_{n\to\infty} A_n$, where $A = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$. We summarize in Appendix B various properties concerning the limit of sets.

Proposition 2.3.1 *For all $0 < \tau_1 \leq \tau_2$, we have that $S_{f_{G,\tau_1}} \subset S_{f_{G,\tau_2}}$. Additionally, if f is continuous, then, for all $\tau > 0$, $S_f \subset S_{f_{G,\tau}}$ and $\lim_{\tau\to 0^+} S_{f_{G,\tau}} \supset S_f$. Under assumption (2.2.1), $\lim_{\tau\to 0^+} S_{f_{G,\tau}} \subset \overline{S}_f$.*

We observe that the assumption (2.2.1) is essential in the last part of Proposition 2.3.1. Indeed, if $S_G = \mathbb{R}^d$, then $S_{f_{G,\tau}} = \mathbb{R}^d$, for all $\tau > 0$, implying $\lim_{\tau\to 0^+} S_{f_{G,\tau}} = \mathbb{R}^d$. Also, since ∂S_f and S_G have arbitrary shape, it is unclear if $x \in \partial S_f$ belongs to $\lim_{\tau\to 0^+} S_{f_{G,\tau}}$ or not.

Proof of Proposition 2.3.1. We first observe that $x \in S_{f_{G,\tau}}$ if and only if $f_{G,\tau}(x) > 0$ if and only if $L_G(x,\tau) > 0$. Proposition 2.2.1 (i) implies that for $x \in \mathbb{R}^d$, $L_G(x,\tau_1) \leq L_G(x,\tau_2)$, from which it follows that $S_{f_{G,\tau_1}} \subset S_{f_{G,\tau_2}}$. Next, suppose that f is continuous and let $x \in S_f$ and $\tau > 0$. Since f is continuous, S_f is open and there exists $\epsilon > 0$ such that $\overline{B}_{\tau\epsilon}(x) \subset S_f$. By **(P4)**, there exist $0 < \delta \leq \tau\epsilon$ and $c > 0$ such that $\lambda((\overline{B}_\delta(x))^k \cap S_{h_{x,\tau}}) > 0$ and $h_{x,\tau} \geq c$ in $(\overline{B}_\delta(x))^k \cap S_{h_{x,\tau}}$. It follows that

$$\begin{aligned} L_G(x,\tau) &= \int h_{x,\tau}(x_1,\ldots,x_k) f(x_1)\ldots f(x_k) dx_1 \ldots dx_k \\ &\geq c \int_{(\overline{B}_\delta(x))^k \cap S_{h_{x,\tau}}} f(x_1)\ldots f(x_k) dx_1 \ldots dx_k > 0. \end{aligned}$$

Thus, $x \in S_{f_{G,\tau}}$ and $S_f \subset S_{f_{G,\tau}}$. Since the sets $\{S_{f_{G,\tau}}\}_{\tau>0}$ are monotonically decreasing with τ, we have that $\lim_{\tau\to 0^+} S_{f_{G,\tau}} = \cap_{\tau>0} S_{f_{G,\tau}} \supset S_f$. For the last part, let $x \in \mathbb{R}^d \setminus \overline{S}_f$.

Since $\mathbb{R}^d \setminus \overline{S}_f$ is open, there exists $\epsilon > 0$ such that $\overline{B}_\epsilon(x) \subset \mathbb{R}^d \setminus \overline{S}_f$. Let $0 < \tau \leq \epsilon/\rho$. By (2.2.1) it follows that $S_{h_{x,\tau}} \subset (\overline{B}_{\rho\tau}(x))^k \subset (\overline{B}_\epsilon(x))^k$ implying that $L_G(x,\tau) = 0$. Therefore, $x \notin \cap_{\tau>0} S_{f_{G,\tau}}$ and $\cap_{\tau>0} S_{f_{G,\tau}} \subset \overline{S}_f$. ■

Our next result establishes that the analytical properties of L_G are inherited by its τ-approximation $f_{G,\tau}$. Specifically, $f_{G,\tau}$ and its derivatives converge uniformly to those of f. This plays a critical role in the properties of clustering investigated in Chapter 3.

Proposition 2.3.2 *Let $P \in \mathcal{P}_{d,\ll\lambda}$ with density f. Then the following hold:*
(i) If f is uniformly continuous, then

$$\lim_{\tau\to 0^+} \sup_{x\in\mathbb{R}^d} |f_{G,\tau}(x) - f(x)| = 0. \tag{2.3.2}$$

(ii) If f is continuous, then for all compact sets $K \subset \mathbb{R}^d$

$$\lim_{\tau\to 0^+} \sup_{x\in K} |f_{G,\tau}(x) - f(x)| = 0.$$

In particular, for all $x \in \mathbb{R}^d$, $\lim_{\tau,\,\epsilon\to 0^+} \sup_{y\in\overline{B}_\epsilon(x)} |f_{G,\tau}(y) - f(x)| = 0$.
(iii) If $f \in L^{kq}(\mathbb{R}^d)$, $q \geq 1$, then $f_{G,\tau}$ converges in $L^{kq}(\mathbb{R}^d)$ to f.
(iv) Suppose (2.2.1) holds and f is m-times continuously differentiable, then, for all compact sets $K \subset S_f$ and $i_1, \ldots, i_m \in \{1, \ldots, d\}$,

$$\lim_{\tau\to 0^+} \sup_{x\in K} |\partial_{i_m} \ldots \partial_{i_1} f_{G,\tau}(x) - \partial_{i_m} \ldots \partial_{i_1} f(x)| = 0.$$

Notice that, if f is uniformly continuous then it is bounded. Indeed, suppose by contradiction that f is unbounded. Let $y_0 = 0$. Then, for all $n \in \mathbb{N}$ there exists $y_n \in \mathbb{R}^d$ such that $f(y_n) \geq n$ and $\|y_n\|_2 \geq \|y_{n-1}\|_2 + 1$. Since f is uniformly continuous, there exists $\delta \in (0, 1/2)$ such that $\sup_{z\in\overline{B}_\delta(x)} |f(x) - f(z)| \leq 1$ for all $x \in \mathbb{R}^d$. It follows that $\inf_{z\in\overline{B}_\delta(y_n)} f(z) \geq n - 1$ and

$$\int f(x)dx \geq \sum_{n=2}^{\infty} \int_{\overline{B}_\delta(y_n)} f(x)dx \geq \sum_{n=2}^{\infty} \lambda(\overline{B}_\delta(y_n)) = \infty.$$

Before proving Proposition 2.3.2 we establish useful inequalities in the following lemma.

Lemma 2.3.1 *Let $s, t \geq 0$. The following hold: (i) $|t^a - s^a| \leq |t - s|^a$, for all $0 < a \leq 1$, and (ii) $|t^a - s^a| \geq |t - s|^a$, for all $a > 1$.*

Proof. It is enough to prove the statement for $0 < s < t$. Let $\varphi : (0, \infty) \to \mathbb{R}$ be given by $\varphi(a) = (1 - s/t)^a - 1 + (s/t)^a$. Notice that, $\lim_{a\to 0^+} \varphi(a) = 1$, $\lim_{a\to\infty} \varphi(a) = -1$ and

$$\varphi'(a) = \log(1 - s/t)(1 - s/t)^a + \log(s/t)(s/t)^a < 0.$$

Then, the equality $\varphi(1) = 0$ shows that $\varphi(a) \geq 0$, for $0 < a \leq 1$, and $\varphi(a) < 0$, for $a > 1$. The same inequalities hold for $t^a\varphi(a)$ and the result follows. ■

Proof of Proposition 2.3.2. We start by proving (i). By Lemma 2.3.1, for $\tau > 0$,

$$\begin{aligned}\sup_{x\in\mathbb{R}^d}|f_{G,\tau}(x) - f(x)| &= \sup_{x\in\mathbb{R}^d}\left|\left(\frac{L_G(x,\tau)}{\tau^{kd}\Lambda_1}\right)^{1/k} - (f^k(x))^{1/k}\right| \\ &\leq \sup_{x\in\mathbb{R}^d}\left|\frac{L_G(x,\tau)}{\tau^{kd}\Lambda_1} - f^k(x)\right|^{1/k} \\ &= \sup_{x\in\mathbb{R}^d}(F_{G,\tau}(x))^{1/k},\end{aligned}$$

where

$$F_{G,\tau}(x) = \left|\frac{1}{\Lambda_1}\int h_{0,1}(x_1,\ldots,x_k)\left[\prod_{j=1}^{k} f(x+\tau x_j) - f^k(x)\right]dx_1\ldots dx_k\right|.$$

Since the k^{th} root is a continuous and increasing function, we have that

$$\sup_{x\in\mathbb{R}^d}(F_{G,\tau}(x))^{1/k} = (\sup_{x\in\mathbb{R}^d} F_{G,\tau}(x))^{1/k}. \tag{2.3.3}$$

Hence, (2.3.2) follows, if we show that

$$\lim_{\tau\to 0^+}\sup_{x\in\mathbb{R}^d} F_{G,\tau}(x) = 0. \tag{2.3.4}$$

For this, observe that $\sup_{x\in\mathbb{R}^d} F_{G,\tau}(x)$ is bounded above by

$$\int h_{0,1}(x_1,\ldots,x_k)\sup_{x\in\mathbb{R}^d}\left|\prod_{j=1}^{k} f(x+\tau x_j) - f^k(x)\right|dx_1\ldots dx_k.$$

Since f is uniformly continuous, for all $(x_1,\ldots,x_k)\in(\mathbb{R}^d)^k$, it holds that

$$\lim_{\tau\to 0^+}\sup_{x\in\mathbb{R}^d}\left|\prod_{j=1}^{k} f(x+\tau x_j) - f^k(x)\right| = 0.$$

(2.3.4) now follows from LDCT, since $h_{0,1}\in L^1((\mathbb{R}^d)^k)$ and the supremum is bounded because f is bounded.
Since a continuous function is uniformly continuous on a compact set, the proof of the first part of (ii) follows from the proof of (i) with $\mathbb{R}^d$ replaced by K. For the second part of (ii), notice that

$$\sup_{y\in\overline{B}_\epsilon(x)}|f_{G,\tau}(y) - f(x)| \leq \sup_{y\in\overline{B}_\epsilon(x)}|f_{G,\tau}(y) - f(y)| + \sup_{y\in\overline{B}_\epsilon(x)}|f(y) - f(x)|.$$

The result now follows from the first part of (ii) and continuity of f. Finally, for (iii), notice that, by Lemma 2.3.1 and Theorem 2.2.1 (iv),

$$\int|f_{G,\tau}(y) - f(y)|^{kd}dy \leq \int|f^k_{G,\tau}(y) - f^k(y)|^d dy \xrightarrow[\tau\to 0^+]{} 0.$$

Before we prove Proposition 2.3.2 (iv) we state without proof a result concerning the partial derivatives of the composition of two functions (Proposition 1 in Hardy (2006)). For any set R, we denote by $\#R$ the cardinality of R.

Claim 2.3.1 *Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ and $\psi : \mathbb{R} \to \mathbb{R}$ be m-times continuously differentiable in $A \subset \mathbb{R}^d$ and $\varphi(A) \subset \mathbb{R}$, respectively. Then, $\psi \circ \varphi$ is m-times continuously differentiable in A and, for $x \in A$ and $i_1, \ldots, i_m \in \{1, \ldots, d\}$, it holds that*

$$\partial_{i_m} \ldots \partial_{i_1} \psi(\varphi(x)) = \sum_{R \in \mathcal{R}_m} [\partial^{[\#R]} \psi](\varphi(x)) \prod_{\{i_{j_l}, \ldots, i_{j_1}\} \in R} \partial_{i_{j_l}} \ldots \partial_{i_{j_1}} \varphi(x),$$

where $\mathcal{R}_m$ is the set of all partitions of $\{1, \ldots, m\}$ and $\partial^{[n]}$ denotes the (unidimensional) n^{th} derivative.

The following lemma is also required for completing the proof of the proposition.

Lemma 2.3.2 *Let $\varphi_n : \mathbb{R}^d \to \mathbb{R}$, $\psi_n : \mathbb{R}^d \to \mathbb{R}$ and $A \subset \mathbb{R}^d$. Suppose that φ_n and ψ_n converge uniformly on A to φ and ψ, respectively. It holds that (i) $\varphi_n + \psi_n$ converges uniformly on A to $\varphi + \psi$ and (ii) if φ and ψ are bounded on A, then $\varphi_n \psi_n$ converges uniformly on A to $\varphi\psi$.*

Proof of Lemma 2.3.2. Proof of (i) is trivial. For (ii) let $c_\varphi = \sup_{x \in A} \varphi(x)$ and $c_\psi = \sup_{x \in A} \psi(x)$. For $0 < \epsilon \leq 1$, there exists $\tilde{n} \in \mathbb{N}$ such that for all $n \geq \tilde{n}$ $\sup_{x \in A} |\varphi_n(x) - \varphi(x)| \leq \epsilon$ and $\sup_{x \in A} |\psi_n(x) - \psi(x)| \leq \epsilon$. It follows that

$$\begin{aligned}
\sup_{x \in A} |\varphi_n(x)\psi_n(x) - \varphi(x)\psi(x)| &\leq \sup_{x \in A} |\varphi_n(x)||\psi_n(x) - \psi(x)| + \sup_{x \in A} |\psi(x)||\varphi_n(x) - \varphi(x)| \\
&\leq (1 + c_\varphi + c_\psi)\epsilon.
\end{aligned}$$

■

We now turn to the proof of (iv). We first notice that, by Proposition 2.2.1 (iv), Remark 2.3.1, and Proposition 2.3.1, $L_G(\cdot, \tau)$ and $f_{G,\tau}$ are m-times continuously differentiable in S_f. Since $K \subset S_f$, $c_1 = \min_{x \in K} f^k(x) > 0$ and $c_2 = \max_{x \in K} f^k(x) < \infty$. By Theorem 2.2.1 (i), there exists $\tau_1 > 0$ such that, for all $0 < \tau \leq \tau_1$, $\sup_{x \in K} |f^k_{G,\tau}(x) - f^k(x)| \leq \frac{c_1}{2}$, implying that $f^k_{G,\tau}(x) \in [c_3, c_4]$, where $c_3 = \frac{c_1}{2}$ and $c_4 = c_2 + \frac{c_1}{2}$. Next, we apply Lemma 2.3.1 with $\varphi = f^k$ and $\psi = (\cdot)^{1/k}$, and obtain that

$$\partial_{i_m} \ldots \partial_{i_1} f(x) = \partial_{i_m} \ldots \partial_{i_1} \psi(\varphi(x)) = \sum_{R \in \mathcal{R}_m} [\partial^{[\#R]} \psi](\varphi(x)) \prod_{\{i_{j_l}, \ldots, i_{j_1}\} \in R} \partial_{i_{j_l}} \ldots \partial_{i_{j_1}} \varphi(x). \tag{2.3.5}$$

Similarly, with $\varphi_\tau = f^k_{G,\tau}$, we have that

$$\partial_{i_m} \ldots \partial_{i_1} f_{G,\tau}(x) = \partial_{i_m} \ldots \partial_{i_1} \psi(\varphi_\tau(x)) = \sum_{R \in \mathcal{R}_m} [\partial^{[\#R]} \psi](\varphi_\tau(x)) \prod_{\{i_{j_l}, \ldots, i_{j_1}\} \in R} \partial_{i_{j_l}} \ldots \partial_{i_{j_1}} \varphi_\tau(x). \tag{2.3.6}$$

By Proposition 2.2.1 (iv), it holds that $\partial_{i_{j_l}} \ldots \partial_{i_{j_1}} \varphi_\tau(x) = (\tilde{G}_\tau * (\partial_{i_{j_l}} \ldots \partial_{i_{j_1}} f^{\times k}))(x, \ldots, x)$. We apply Lemma 2.2.2 with $\tilde{f} = \partial_{i_{j_l}} \ldots \partial_{i_{j_1}} f^{\times k}$ and $A = (K)^k$, and obtain that $\partial_{i_{j_l}} \ldots \partial_{i_{j_1}} \varphi_\tau$ converges uniformly on K to $\partial_{i_{j_l}} \ldots \partial_{i_{j_1}} \varphi$. Next, notice that, for all $j \in \{1, \ldots, m\}$, $\partial^{[j]} \psi$ is uniformly continuous on $[c_3, c_4]$: for all $\epsilon > 0$, there exists $\delta > 0$ such that $\sup_{s,t \in [c_3, c_4] : |s-t| \leq \delta} |\partial^{[j]} \psi(s) - \partial^{[j]} \psi(t)| \leq \epsilon$. By Theorem 2.2.1 (i), there exists $0 < \tau_2 \leq \tau_1$, such that, for all $0 < \tau \leq \tau_2$, $\sup_{x \in K} |f^k_{G,\tau}(x) - f^k(x)| \leq \delta$. Therefore, we have that, for all $0 < \tau \leq \tau_2$, $\sup_{x \in K} |[\partial^{[j]} \psi](\varphi_\tau(x)) - [\partial^{[j]} \psi](\varphi(x))| \leq \epsilon$; that is, $[\partial^{[j]} \psi](\varphi_\tau(\cdot))$ converges uniformly on K to $[\partial^{[j]} \psi](\varphi(\cdot))$. Now, the result follows from (2.3.5), (2.3.6), and Lemma 2.3.2 with $A = K$. ■

We show that continuity is not enough in Proposition 2.3.2 (i). To this end, let for simplicity $d = 1$ and $G = L, S, B, K_\beta$. Consider the function $\tilde{f} : \mathbb{R} \to \mathbb{R}$ with support $S_{\tilde{f}} = \cup_{n=1}^{\infty} \left(n - \frac{1}{2n^3}, n + \frac{1}{2n^3}\right)$ defined by

$$\tilde{f}(n + x_1) = \tilde{f}(n - x_1) = \begin{cases} 2n^4 \left(\frac{1}{2n^3} - x_1\right) & \text{if } 0 \leq x_1 < \frac{1}{2n^3} \\ 0 & \text{if } \frac{1}{2n^3} \leq x_1 \leq \frac{1}{2}. \end{cases}$$

Notice that, since $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$,

$$\int \tilde{f}(x)dx = \sum_{n=1}^{\infty} \int_{\left(n - \frac{1}{2n^3}, n + \frac{1}{2n^3}\right)} \tilde{f}(x)dx = \sum_{n=1}^{\infty} \frac{1}{2n^2} = \frac{\pi^2}{12}.$$

Let $c = \frac{\pi^2}{12}$. Then $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \frac{1}{c}\tilde{f}(x)$ is an unbounded, continuous density function. The τ-approximation of f is given by $f_{G,\tau}(x) = \frac{1}{\tau}\sqrt{L_G(x,\tau)}$, where, by Corollary 2.2.1,

$$\begin{aligned} L_G(x,\tau) &= 2\int_{T_{\tau,++}} f(x+x_1)f(x-x_2)dx_1dx_2 \\ &= 2\int_0^{\tau} \left[\int_0^{\tau - x_1} f(x-x_2)dx_2\right] f(x+x_1)dx_1. \end{aligned}$$

Notice that f is symmetric about $n \in \mathbb{N}$ and for $\frac{1}{n^3} \leq \tau \leq \frac{1}{2}$

$$\begin{aligned} L_G(n,\tau) &= 2\int_0^{\min\left(\tau, \frac{1}{2n^3}\right)} \left[\int_0^{\min\left(\tau - x_1, \frac{1}{2n^3}\right)} \frac{2n^4}{c}\left(\frac{1}{2n^3} - x_2\right)dx_2\right] \frac{2n^4}{c}\left(\frac{1}{2n^3} - x_1\right)dx_1 \\ &= \frac{2}{c^2}\int_0^{\frac{1}{2n^3}} \left[\int_0^{\frac{1}{2n^3}} 2n^4\left(\frac{1}{2n^3} - x_2\right)dx_2\right] 2n^4\left(\frac{1}{2n^3} - x_1\right)dx_1 \\ &= \frac{2}{c^2}\int_0^{\frac{1}{2n^3}} \frac{n^2}{2}\left(\frac{1}{2n^3} - x_1\right)dx_1 = \frac{1}{8c^2n^4}. \end{aligned}$$

For all $0 < \tau \leq \frac{1}{2}$ fixed there exists $n \in \mathbb{N}$ such that $\frac{1}{n^2} \leq \tau$, and therefore

$$\begin{aligned} \sup_{x \in \mathbb{R}} |f_{G,\tau}(x) - f(x)| &\geq \sup_{n \in \mathbb{N}: \frac{1}{n^2} \leq \tau} |f_{G,\tau}(n) - f(n)| = \sup_{n \in \mathbb{N}: \frac{1}{n^2} \leq \tau} \left|\frac{1}{\tau}\frac{1}{2\sqrt{2}cn^2} - n\right| \\ &\geq \sup_{n \in \mathbb{N}: \frac{1}{n^2} \leq \tau} \left|\frac{1}{2\sqrt{2}c} - n\right| = \infty. \end{aligned}$$

Uniform continuity in Proposition 2.3.2 (i) prevents f to become arbitrarily large implying that the above supremum is bounded. Also, it ensures that the supremum converges to zero, thus allowing to use LDCT and obtain the statement.
In the next proposition, we use Theorem 2.2.1 (iii) to obtain a uniform approximation of $f_\tau = f_{G,\tau}$ in compact sets. We need the following notation: for $A \subset \mathbb{R}^d$ and $\delta > 0$, $(A)^{+\delta} = \{x \in \mathbb{R}^d : \inf_{y \in A}\|x - y\|_2 \leq \delta\}$ and $(A)^{-\delta} = \mathbb{R}^d \setminus \left(\mathbb{R}^d \setminus A\right)^{+\delta} = \{x \in \mathbb{R}^d : \inf_{y \in \mathbb{R}^d \setminus A}\|x - y\|_2 > \delta\}$.

Proposition 2.3.3 *Suppose (2.2.1) holds true and f is three times continuously differentiable. Let K be a compact subset of S_f. Then, there are constants $\tilde{\tau}(K), \tilde{c}_1(K), \tilde{c}_2(K) > 0$ and a continuously differentiable function $\tilde{R}_\tau : K \to \mathbb{R}$ such that, for all $x \in K$ and*

$0 < \tau \leq \tilde{\tau}(K)$, $|\tilde{R}_\tau(x)| \leq \tilde{c}_1(K)$, $\|\nabla \tilde{R}_\tau(x)\|_2 \leq \tilde{c}_2(K)$, *and*

$$f_\tau(x) = f(x) + \tilde{R}_\tau(x)\tau^2.$$

Proof of Proposition 2.3.3. Notice that, since $K \subset S_f$, K is closed and S_f is open, there is $\delta, \tilde{h} > 0$ such that $(K)^{+(\delta+\tilde{h})} \subset S_f$. Let $\tau_1 = \delta/\rho$ and $\tilde{K} = K^{+\tilde{h}}$. Then, for $\tau \in (0, \tau_1]$, we have that $(\tilde{K})^{+\rho\tau} \subset (K)^{+(\delta+\tilde{h})}$ and, by Remark 2.3.1, f_τ is three times continuously differentiable in $\tilde{K}$. Since f is three times continuously differentiable, by Theorem 2.2.1 (iii), we have that, for $x \in \tilde{K}$,

$$f_\tau^k(x) - f^k(x) = Q_\tau(x)\tau^2,$$

where, for all $\tau \in [0, \tau_1]$, $Q_\tau(x) = (R_1(x) + R_2(x))/\Lambda_1 + o(\tau)$ is well-defined and continuously differentiable with uniformly bounded derivatives in $\tilde{K}$. Let

$$\tilde{c}_3(\tilde{K}) = \sup_{y\in\tilde{K}} \sup_{\tau\in[0,\tau_1]} |Q_\tau(y)/f^k(y)|$$

and $\tilde{\tau}(K) \in (0, \min(1, \tau_1, \tilde{c}_3^{-1/2}(\tilde{K}), \tilde{c}_3^{-2}(\tilde{K})))$. It follows from Newton's generalized binomial theorem that, for $\tau \in (0, \tilde{\tau}(K)]$,

$$\begin{aligned} f_\tau(x) &= f(x)\left(1 + \tau^{1/2} Q_\tau(x)/f^k(x)\tau^2\right)^{1/k} \\ &= f(x) + 1/kQ_\tau(x)/f^{k-1}(x)\tau^2 + \tau^3\tilde{Q}_\tau(x), \end{aligned}$$

where $\tilde{Q}_\tau(x) = 1/\tau^3 f(x)\sum_{j=2}^{\infty}\binom{1/k}{j}\left(Q_\tau(x)/f^k(x)\tau^2\right)^j$ and $\binom{1/k}{j} = (1/k\dots(1/k - j+1))/j!$. Now, since $\tau \leq \tilde{\tau}(K) < 1$, we obtain that

$$\begin{aligned} |\tilde{Q}_\tau(x)| &\leq f(x)|\sum_{l=2}^{\infty}\binom{1/k}{l}\left(\tau^{1/2}Q_\tau(x)/f^k(x)\right)^l| \\ &= f(x)(1 + 1/k\tau^{1/2}Q_\tau(x)/f^k(x) - (1+\tau^{1/2}Q_\tau(x)/f^k(x))^{1/k}). \end{aligned}$$

Hence, $\tilde{c}_4(\tilde{K}) = \sup_{y\in\tilde{K}}\sup_{\tau\in[0,\tilde{\tau}(K)]}|\tilde{Q}_\tau(y)| < \infty$. Let $\tilde{R}_\tau(x) = 1/kQ_\tau(x)/f^{k-1}(x) + \tau\tilde{Q}_\tau(x)$. We need to show that, for all $\tau \in (0, \tilde{\tau}(K)]$, $\tilde{Q}_\tau$ is continuously differentiable in K with uniformly bounded derivatives. To this end, let

$$\begin{aligned} T_{\tau,l}(x) &= \binom{1/k}{l}(Q_\tau(x)/f^k(x)\tau^2)^l, \\ \tilde{T}_{\tau,l}^{(i)}(x) &= \binom{1/k}{l} l(Q_\tau(x)/f^k(x)\tau^2)^{l-1}\partial_i(Q_\tau(x)/f^k(x)\tau^2), \end{aligned}$$

$S_{\tau,j}(x) = \sum_{l=2}^{j} T_{\tau,l}(x)$ and $\tilde{S}_{\tau,j}^{(i)}(x) = \sum_{l=2}^{j}\tilde{T}_{\tau,l}^{(i)}(x)$. Notice that $\partial_i S_{\tau,j}(x) = \tilde{S}_{\tau,j}^{(i)}(x)$ and let

$$\tilde{c}_5(\tilde{K}) = \sup_{y\in\tilde{K}}\sup_{\tau\in[0,\tilde{\tau}(K)]}\|\nabla Q_\tau(y)/f^k(y)\|_2.$$

We compute

$$\sup_{y\in\tilde{K}}\sup_{\tau\in[0,\tilde{\tau}(K)]}|S_{\tau,\infty}(y) - S_{\tau,j}(y)| \leq \sum_{l=j+1}^{\infty}(\tilde{c}_3(\tilde{K})\tilde{\tau}^2(K))^l = \frac{(\tilde{c}_3(\tilde{K})\tilde{\tau}^2(K))^{j+1}}{1 - \tilde{c}_3(\tilde{K})\tilde{\tau}^2(K)} \xrightarrow[j\to\infty]{} 0$$

and

$$\sup_{y\in\tilde{K}}\sup_{\tau\in[0,\tilde{\tau}(K)]}|\tilde{S}^{(i)}_{\tau,\infty}(y)-\tilde{S}^{(i)}_{\tau,j}(y)|\leq \tilde{c}_5(\tilde{K})\tilde{\tau}^2(K)\sum_{l=j}^{\infty}(\tilde{c}_3(\tilde{K})\tilde{\tau}^2(K))^l$$
$$=\tilde{c}_5(\tilde{K})\tilde{\tau}^2(K)\frac{(\tilde{c}_3(\tilde{K})\tilde{\tau}^2(K))^j}{1-\tilde{c}_3(\tilde{K})\tilde{\tau}^2(K)}\xrightarrow[j\to\infty]{}0.$$

Hence, the series $S_{\tau,j}$ and $\tilde{S}^{(i)}_{\tau,j}$ converge uniformly to $S_{\tau,\infty}$ and $\tilde{S}^{(i)}_{\tau,\infty}$. By the fundamental theorem of calculus (FTC), the uniform convergence of $\tilde{S}^{(i)}_{\tau,\infty}$, and LDCT, we have that, for all $x\in K$ and $0<h\leq\tilde{h}$,

$$S_{\tau,\infty}(x+he_i)-S_{\tau,\infty}(x)=\sum_{l=2}^{\infty}(T_{\tau,l}(x+he_i)-T_{\tau,l}(x))$$
$$=\sum_{l=2}^{\infty}\int_0^h\tilde{T}^{(i)}_{\tau,l}(x+te_i)dt=\int_0^h\tilde{S}^{(i)}_{\tau,\infty}(x+te_i)dt.$$

Now, using again FTC, we have that

$$\partial_iS_{\tau,\infty}(x)=\lim_{h\to0^+}\frac{S_{\tau,\infty}(x+he_i)-S_{\tau,\infty}(x)}{h}=\lim_{h\to0^+}\frac{1}{h}\int_0^h\tilde{S}^{(i)}_{\tau,\infty}(x+te_i)dt=\tilde{S}^{(i)}_{\tau,\infty}(x)$$

and

$$\nabla S_{\tau,\infty}(x)=(\tilde{S}^{(1)}_{\tau,\infty}(x),\ldots,\tilde{S}^{(d)}_{\tau,\infty}(x))^\top$$
$$=\sum_{l=2}^{\infty}\binom{1/k}{l}l(Q_\tau(x)/f^k(x)\tau^2)^{l-1}\nabla(Q_\tau(x)/f^k(x)\tau^2).$$

It follows that $\tilde{Q}_\tau$ is continuously differentiable in K with

$$\nabla\tilde{Q}_\tau(x)=\nabla f(x)(S_{\tau,\infty}(x)/\tau^3)+f(x)\nabla S_{\tau,\infty}(x)/\tau^3.$$

Since $\tilde{c}_3(\tilde{K})\tilde{\tau}^{1/2}(K)<1$ we obtain that

$$\sup_{y\in K}\sup_{\tau\in[0,\tilde{\tau}(K)]}|S_{\tau,\infty}(x)/\tau^3|\leq\sum_{l=2}^{\infty}(\tilde{c}_3(\tilde{K})\tilde{\tau}^{1/2}(K))^l<\infty,$$

$$\sup_{y\in K}\sup_{\tau\in[0,\tilde{\tau}(K)]}\|\nabla S_{\tau,\infty}(x)/\tau^3\|_2\leq\tilde{c}_5(\tilde{K})\sum_{l=1}^{\infty}(\tilde{c}_3(\tilde{K})\tilde{\tau}^{1/2}(K))^l<\infty,$$

and $\tilde{c}_6(K)=\sup_{y\in K}\sup_{\tau\in[0,\tilde{\tau}(K)]}\|\nabla\tilde{Q}_\tau(x)\|_2<\infty$. Let

$$\tilde{c}_7(K)=\sup_{y\in K}\sup_{\tau\in[0,\tilde{\tau}(K)]}|(Q_\tau(y)/f^{k-1}(y))|$$

and

$$\tilde{c}_8(K)=\sup_{y\in\tilde{K}}\sup_{\tau\in[0,\tilde{\tau}(K)]}\|\nabla(Q_\tau(y)/f^{k-1}(y))\|_2.$$

Then, we conclude that

$$\sup_{x\in K}\sup_{\tau\in[0,\tilde{\tau}(K)]}|\tilde{R}_\tau(x)|\leq\tilde{c}_1(K)=\tilde{c}_7(K)/k+\tilde{\tau}(K)\tilde{c}_4(\tilde{K})<\infty$$

and

$$\sup_{x \in K} \sup_{\tau \in [0,\tilde{\tau}(K)]} \|\nabla \tilde{R}_\tau(x)\|_2 \le \tilde{c}_2(K) = \tilde{c}_8(K)/k + \tilde{\tau}(K)\tilde{c}_6(K) < \infty.$$

■

2.4 Sample local depth functions

As noticed at the end of Section 1.9, typical estimator of *Type A* depth functions are U-statistics. Since *Type A* LDFs have the same integral form (2.2.2), the same considerations apply to them. Specifically, let $\{X_i\}_{i=1}^\infty$ be a sequence of i.i.d. random variables from $P \in \mathcal{P}_d$. The estimate of L_G, called sample local depth, is the U-statistics of order k

$$L_{G,n}(x,\tau,P) = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \cdots < i_k \le n} h_{x,\tau}(X_{i_1},\ldots,X_{i_k}), \tag{2.4.1}$$

where $x \in \mathbb{R}^d$ and $\tau \in [0,\infty]$. In particular, (2.4.1) entails that the sample depth $D_{G,n}(x,P) = L_{G,n}(x,\infty,P)$ is an estimator of $D_G(x,P)$. When there is no scope of confusion, we will also write $L_{G,n}(x,\tau)$ and $D_{G,n}(x)$ for $L_{G,n}(x,\tau,P)$ and $D_{G,n}(x,P)$. In the next proposition, we study the rank of the U-statistics $L_{G,n}(x,\tau,P)$ (see Definition 1.10.2). To this end, let $S_P = \cap_{A \subset \mathbb{R}^d : A \text{ is open and } P(A)=0} (\mathbb{R}^d \setminus A)$[2] be the support of a probability measure $P \in \mathcal{P}_d$.

Proposition 2.4.1 *Let $P \in \mathcal{P}_d$ and $x \in \mathbb{R}^d$. The following holds:*
(i) If S_P is unbounded and $x \in S_P$, then, for all $\tau \in (0,\infty)$, $L_{G,n}(x,\tau,P)$ has rank 1.
(ii) Suppose (2.2.1) holds and $x \notin S_P$, then there exists $\tilde{\tau} > 0$ such that $L_{G,n}(x,\tau,P) = 0$ P^n-a.s. and $p_{k,j}h_{x,\tau} = 0$ P^j-a.s. for all $\tau \in [0,\tilde{\tau}]$ and $j = 1,\ldots,k$.

Notice that, if $P \in \mathcal{P}_{d,\ll\lambda}$ with continuous density f, then $x \in S_P$ if and only if $x \in \overline{S}_f$. The considerations after Definition 2.2.1 shows that (i) does not hold in general for $\tau = \infty$.

Proof of Proposition 2.4.1. We begin by proving (i). We need to show that $p_{k,1}h_{x,\tau}$ is not equal to 0 P-a.s, that is, $\overline{p}_{k,1}h_{x,\tau}$ is not equal to $L(x,\tau,P)$ P-a.s. Using that $x \in S_P$ and **(P4)** we obtain that $L(x,\tau,P) > 0$. Next, let $\{r_n\}_{n=1}^\infty$ be a sequence of positive scalars with $\lim_{n\to\infty} r_n = \infty$. Using **(P3)** we obtain that

$$\lim_{n\to\infty} \sup_{y \in \mathbb{R}^d \setminus B_{r_n}(0)} \sup_{(x_2,\ldots,x_k) \in (\mathbb{R}^d)^{k-1}} h_{x,\tau}(y,x_2,\ldots,x_k) = 0.$$

It follows that

$$\lim_{n\to\infty} \sup_{y \in \mathbb{R}^d \setminus B_{r_n}(0)} \overline{p}_{k,1}h_{x,\tau}(y) = 0.$$

Now, let $\tilde{n} \in \mathbb{N}$ be such that

$$\sup_{y \in \mathbb{R}^d \setminus B_{r_n}(0)} \overline{p}_{k,1}h_{x,\tau}(y) < L(x,\tau,P)$$

[2]It holds that $S_P = \mathbb{R}^d \setminus (\cup_{A \subset \mathbb{R}^d : A \text{ is open and } P(A)=0} A) = \{x \in \mathbb{R}^d : x \in \mathbb{R}^d : P(U_x) > 0$ for every (open) neighborhood U_x of $x\}$. In particular, S_P is closed and $x \in S_P$ if and only if every neighborhood of x has positive probability.

for all $n \geq \tilde{n}$. Next, we use that S_P is unbounded and let $y \in S_P \cap (\mathbb{R}^d \setminus B_{r_{\tilde{n}}+1}(0))$. We conclude that $P(B_1(y)) > 0$ and $\overline{p}_{k,1}h_{x,\tau}(z) < L(x,\tau,P)$ for all $z \in B_1(y)$.

We now prove (ii). Using (2.2.1), we obtain that $\overline{S}_{h_{x,\tau}} \subset (\overline{B}_{\tau\rho}(x))^k$. Next, notice that since $x \notin S_P$ there exists $r > 0$ such that $P(\overline{B}_r(x)) = 0$. Let $\tilde{\tau} = r/\rho$. It follows that, for all $\tau \in [0,\tilde{\tau}]$ and $x_2,\dots,x_k \in \mathbb{R}^d$, $h_{x,\tau}(\cdot,x_2,\dots,x_k) = 0$ in $\mathbb{R}^d \setminus \overline{B}_r(x)$. Since $P(\overline{B}_r(x)) = 0$, we conclude that $h_{x,\tau} = 0$ P^k-a.s. yielding that $L_{G,n}(x,\tau,P) = 0$ P^n-a.s. and $p_{k,j}h_{x,\tau} = 0$ P^j-a.s. for all $j = 1,\dots,k$. ■

Several properties of $L_{G,n}$ can be deduced from the properties of U-statistics and U-processes derived in Sections 1.10 and 1.11. The fact that $\mathcal{H}_G$ is uniformly bounded yields some simplification in the assumptions. In particular, the variance of $L_{G,n}(x,\tau,P)$ is given by Proposition 1.10.2 and it is equal to

$$\begin{aligned}\mathbb{V}ar[L_{G,n}(x,\tau,P)] &= \sum_{j=1}^{k}\binom{k}{j}^2\binom{n}{j}^{-1}\mathbb{V}ar[p_{k,j}h_{x,\tau}(X_1,\dots,X_j)] \\ &= \frac{k^2}{n}\mathbb{V}ar[p_{k,1}h_{x,\tau}(X_1)] + O(n^{-2}).\end{aligned} \tag{2.4.2}$$

Each term on the RHS is in general non-zero, although it may be zero in some cases (cf. Proposition 2.4.1). Importantly, Proposition 1.10.4 yields that $L_{G,n}(x,\tau,P)$ is a consistent estimator of $L_G(x,\tau,P)$ for all $P \in \mathcal{P}_d$. However, in typical applications, the choice of x, τ, and G varies and different choices of x, τ and G may be investigated. Using Corollary 1.11.2, we show that $L_{G,n}$ is uniformly consistent over x and τ. To this end, notice that $\mathbb{R}^d \times [0,\infty]$ can be identified with $\mathcal{H}_G$ via the bijection $(x,\tau) \leftrightarrow h_{x,\tau}$. We assume that $\mathcal{H}_G$ is a VC subgraph class. We show in Section 2.7 that this assumption holds for local simplicial depth, local β-skeleton depth, and other examples of LDFs.

Theorem 2.4.1 *Let $P \in \mathcal{P}_d$ and $\mathcal{H}_G$ be VC subgraph. Then,*

$$\sup_{(x,\tau)\in\mathbb{R}^d\times[0,\infty]} |L_{G,n}(x,\tau,P) - L_G(x,\tau,P)| \xrightarrow{a.s.} 0.$$

Proof of Theorem 2.4.1. Since $\mathcal{H}_G$ is uniformly bounded, using Corollary 1.11.2, it is enough to show that $\mathcal{H}_G$ is image admissible Suslin (see Definition 1.11.2). To this end, we show that the function $S : (\mathbb{R}^d)^k \times \mathbb{R}^d \times [0,\infty] \to \mathbb{R}$ given by $S(x_1,\dots,x_k,x,\tau) = h_{x,\tau}(x_1,\dots,x_k)$ is Borel measurable. To see this, for $\tau \in [0,\infty]$, let $h_\tau : (\mathbb{R}^d)^k \times \mathbb{R}^d \to \mathbb{R}$ be given by $\tilde{h}_\tau(x_1,\dots,x_k,x) = h_{x,\tau}(x_1,\dots,x_k)$ and $R : (\mathbb{R}^d)^k \times \mathbb{R}^d \times [0,\infty] \to \mathbb{R}$ be given by $R(x_1,\dots,x_k,x,\tau) = \left(\frac{x_1-x}{\tau},\dots,\frac{x_k-x}{\tau}\right)^\top$. Since G is Borel measurable and R is continuous, $\tilde{h}_{(\cdot)}(\cdot) = G(R(\cdot))$ is Borel measurable. In particular, $\tilde{h}_\tau$ is Borel measurable for all $\tau \in (0,\infty)$ and $\tilde{h}_0$ and $\tilde{h}_\infty$ are Borel measurable because they are limit of Borel measurable functions. It follows that, for all $A \in \mathcal{B}_1$,

$$\begin{aligned}S^{-1}(A) =&(R^{-1}(G^{-1}(A)) \cup (\tilde{h}_0^{-1}(A) \times \{0\}) \cup (\tilde{h}_\infty^{-1}(A) \times \{\infty\}) \\ \in&\mathcal{B}((\mathbb{R}^d)^k \times \mathbb{R}^d \times [0,\infty]).\end{aligned}$$

Hence, $\mathcal{H}_G$ is image admissible Suslin via the onto Borel measurable map $T : \mathbb{R}^d \times [0,\infty] \to \mathcal{H}_G$ given by $T(x,\tau) = h_{x,\tau}$. ■

In some examples, it is possible that $G = G_\theta \in \mathcal{G}$ is indexed by a parameter $\theta \in \Theta \subset \mathbb{R}$, as is the case for β-skeletons. In such cases, one can strengthen the above Theorem 2.4.1 to obtain uniformity in the indexing parameter, that is,

$$\sup_{(\theta,x,\tau)\in\Theta\times\mathbb{R}^d\times[0,\infty]} |L_{G_\theta,n}(x,\tau,P) - L_{G_\theta}(x,\tau,P)| \xrightarrow{a.s.} 0. \tag{2.4.3}$$

To this end, we make the following assumptions:

(A1) G_θ satisfies **(P1)**-**(P4)**, where $k = k^*_\Theta = k_{G_\theta}$ is independent of θ.

(A2) $\mathcal{H}^*_\Theta = \cup_{\theta\in\Theta}\mathcal{H}_{G_\theta}$ is a VC subgraph class.

(A3) $\sup_{\theta\in\Theta} G_\theta \leq l^*_\Theta$ for some $0 < l^*_\Theta < \infty$.

(A4) $G_{(\cdot)}(\cdot)$ is jointly Borel measurable.

The details for the β-skeleton are provided in Section 2.7.

Proof of (2.4.3). Using **(A3)**, we obtain that $\mathcal{H}^*_\Theta$ is uniformly bounded. Since $\mathcal{H}^*_\Theta$ is VC subgraph by **(A2)**, the statement follows from Corollary 1.11.2 if we show that $\mathcal{H}^*_\Theta$ is image admissible Suslin. To this end, let $R^*_\Theta, S^*_\Theta : (\mathbb{R}^d)^k \times \mathbb{R}^d \times \Theta \times [0,\infty] \to \mathbb{R}$ be given by $R^*_\Theta(x_1,\ldots,x_k,x,\theta,\tau) = \left(\frac{x_1-x}{\tau},\ldots,\frac{x_k-x}{\tau},\theta\right)^\top$ and $S^*_\Theta(x_1,\ldots,x_k,x,\theta,\tau) = h_{G_\theta,x,\tau}(x_1,\ldots,x_k)$. Also, for $\tau \in [0,\infty]$, let $h^*_{\Theta,\tau} : (\mathbb{R}^d)^k \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ be given by $h^*_{\Theta,\tau}(x_1,\ldots,x_k,x,\theta) = h_{G_\theta,x,\tau}(x_1,\ldots,x_k)$ and $G^*_\Theta : (\mathbb{R}^d)^k \times \Theta \to \mathbb{R}$ be given by $G^*_\Theta(x_1,\ldots,x_k,\theta) = G_\theta(x_1,\ldots,x_k)$. Since, for $\theta \in \Theta$ and $\tau \in (0,\infty)$, $h^*_{\Theta,\tau}(\cdot,\cdot) = G^*_\Theta(R^*_\Theta(\cdot,\cdot,\theta,\tau)$, where G^*_Θ is Borel measurable by **(A4)** and R^*_Θ is continuous, we obtain that $G^*_\Theta \circ R^*_\Theta$, $h^*_{\Theta,0}$, and $h^*_{\Theta,\infty}$ are Borel measurable. Therefore, for all $A \in \mathcal{B}_1$,

$$\begin{aligned}(S^*_\Theta)^{-1}(A) =& ((R^*_\Theta)^{-1}((G^*_\Theta)^{-1}(A)) \cup ((h^*_{\Theta,0})^{-1}(A) \times \{0\}) \cup ((h^*_{\Theta,\infty})^{-1}(A) \times \{\infty\}) \\ &\in \mathcal{B}((\mathbb{R}^d)^k \times \mathbb{R}^d \times \Theta \times [0,\infty]).\end{aligned}$$

We conclude that $\hat{S}_\Theta$ is Borel measurable and the class $\hat{\mathcal{H}}_\Theta$ is image admissible Suslin via $\hat{T}_\Theta : \mathbb{R}^d \times \Theta \times [0,\infty] \to \hat{\mathcal{H}}_\Theta$ given by $\hat{T}_\Theta(x,\theta,\tau) = h_{G_\theta,x,\tau}$. ■

We now turn to the uniform central limit theorem for $L_{G,n}$ over a suitable subset A of $\mathbb{R}^d \times [0,\infty]$. Specifically, for $P \in \mathcal{P}_d$, A is chosen so that that $L_{G,n}(x,\tau,P)$ is non-degenerate for all $(x,\tau) \in A$ (cf. Proposition 2.4.1). Notice that there is a one-to-one correspondence between A and the class $\hat{H}_{G,A} = \{h_{x,\tau} \in \mathcal{H}_G : (x,\tau) \in A\}$. Also, if $\mathcal{H}_G$ is VC subgraph, then $\hat{H}_{G,A}$ is VC subgraph by Proposition 1.11.1 (vi). We are now ready to state the uniform CLT for sample LDFs.

Theorem 2.4.2 *Let $P \in \mathcal{P}_d$ and $A \subset \mathbb{R}^d \times [0,\infty]$ such that $\hat{H}_{G,A}$ is P-non-degenerate. If $\hat{H}_{G,A}$ is VC subgraph, then*

$$\sqrt{n}(L_{G,n}(\cdot,\cdot,P) - L_G(\cdot,\cdot,P)) \xrightarrow{d^*} (\hat{W}_k(P))(\cdot,\cdot),$$

where $\{(\hat{W}(P))(x,\tau)\}_{(x,\tau)\in A}$ is a Gaussian process with mean zero and covariance function $\gamma_{\hat{W}(P)} : A \times A \to \mathbb{R}$ given by

$$\gamma_{\hat{W}(P)}((x,\tau),(y,\nu)) = k^2\left(\int(\overline{p}_{k,1}h_{x,\tau})(x_1)(\overline{p}_{k,1}h_{y,\nu})(x_1)dP(x_1) - L_G(x,\tau,P)L_G(y,\nu,P)\right). \tag{2.4.4}$$

Proof of Theorem 2.4.2. Let $\hat{S}_A : (\mathbb{R}^d)^k \times A \to \mathbb{R}$ be given by $\hat{S}_A(x_1, \ldots, x_k, x, \tau) = S(x_1, \ldots, x_k, x, \tau)$, where S is defined in the proof of Theorem 2.4.1 and $(x, \tau) \in A$. Since S is Borel measurable, $\hat{S}_A$ is Borel measurable on $(\mathbb{R}^d)^k \times A$ and $\hat{H}_{G,A}$ is image admissible Suslin via the onto Borel measurable map $\hat{T}_A : A \to \hat{H}_{G,A}$ given by $\hat{T}_A(x, \tau) = h_{x,\tau}$. Since $\hat{H}_{G,A}$ is VC subgraph and uniformly bounded, Corollary 1.11.4 yields that

$$\sqrt{n}(L_{G,n}(\cdot, \cdot, P) - L_G(\cdot, \cdot, P)) \xrightarrow{d^*} (\hat{W}_k(P))(\cdot, \cdot),$$

where $(\hat{W}_k(P))(x, \tau) = (\tilde{W}(P))(h_{x,\tau}) = k(\tilde{B}(P))(\overline{p}_{k,1} h_{x,\tau})$. For $(x, \tau), (y, \nu) \in A$, we obtain that

$$\begin{aligned}
\gamma_{\hat{W}(P)}((x, \tau), (y, \nu)) &= k^2 \gamma_{\tilde{B}(P)}(\overline{p}_{k,1} h_{x,\tau}, \overline{p}_{k,1} h_{y,\nu}) \\
&= k^2 \mathcal{J}_1((p_{1,1}\overline{p}_{k,1} h_{x,\tau})(p_{1,1}\overline{p}_{k,1} h_{y,\nu}), P) \\
&= k^2 \int (p_{1,1}\overline{p}_{k,1} h_{x,\tau})(x_1)(p_{1,1}\overline{p}_{k,1} h_{y,\nu})(x_1) dP(x_1) \\
&= k^2 \left(\int (\overline{p}_{k,1} h_{x,\tau})(x_1)(\overline{p}_{k,1} h_{y,\nu})(x_1) dP(x_1) - L_G(x, \tau, P) L_G(y, \nu, P) \right).
\end{aligned}$$

■

In the clustering applications discussed below, we will establish the consistency of the sample clustering algorithm. This will involve approximating the τ-approximations of the depth functions and their derivatives via their sample versions. The quality of this approximation will play a critical role in the consistency arguments. Our next result enables this study by establishing a Bernstein-type inequality for local depth functions.

Theorem 2.4.3 *Let $P \in \mathcal{P}_d$ and $\sigma_G^2 = \sup_{(x,\tau) \in \mathbb{R}^d \times [0,\infty]} \mathbb{V}ar[p_{k,1} h_{G,x,\tau}(X_1)]$. If $\mathcal{H}_G$ is VC subgraph, then there are constants $1 < c_{G,0}, c_{G,1}, c_{G,2} < \infty$ such that, for all $t \geq \max(2^3 \sigma_G, 2^4 c_{G,0})$,*

$$\mathbb{P}(\sqrt{n} \sup_{(x,\tau) \in \mathbb{R}^d \times [0,\infty]} |L_{G,n}(x, \tau, P) - L_G(x, \tau, P)| \geq t) \leq M_G(n, t) = \sum_{j=1}^{3} M_{G,j}(n, t), \tag{2.4.5}$$

where

$$\begin{aligned}
M_{G,1}(n, t) &= 8 \exp\left(-\frac{\sqrt{n} t^2}{2^{15} k^2 (\sqrt{n} \sigma_G^2 + t l_G)} \right), \\
M_{G,2}(n, t) &= 8 c_{G,1}^{2 c_{G,2}} \left(\sigma_G^2 + \frac{2 t l_G}{\sqrt{n}} \right)^{-c_{G,2}} \exp\left(-\left(\frac{n \sigma_G^2}{2 l_G^2} + \frac{\sqrt{n} t}{4 l_G} \right) \right), \quad \text{and} \\
M_{G,3}(n, t) &= 2 \exp\left(-\frac{\sqrt{n} t^2}{2^{6+k} k^{k+1} l_G c_{G,0} (\sqrt{n} \sigma_G^2 + t l_G)} \right).
\end{aligned}$$

Proof of Theorem 2.4.3. We apply Theorem 1.11.9 with $E = \mathbb{R}^d$ and $\mathcal{H} = \mathcal{H}_G$. For ease of notation we use the subscript G in place of $\mathcal{H}_G$. ■

2.5 Sample τ-approximation

Sample LDFs can be used to estimate the τ-approximation via a plug-in approach. The resulting estimator is given by

$$f_{G,\tau,n}(x) = \left(\frac{L_{G,n}(x,\tau)}{\tau^{kd}\Lambda_1}\right)^{1/k} \tag{2.5.1}$$

and is called sample τ-approximation. In this section, we use Proposition 2.3.2 and Theorem 2.4.3 and show that, if P is absolutely continuous with respect to the Lebesgue measure with density f, then the sample τ-approximation is a uniform consistent estimator of f. The precise statement reads as follows.

Proposition 2.5.1 *Let $P \in \mathcal{P}_{d,\ll\lambda}$ with density f and suppose that $\mathcal{H}_G$ is VC subgraph. Let $\{\tau_n\}_{n=1}^{\infty}$ and $\{\epsilon_n\}_{n=1}^{\infty}$ be sequences of positive scalars converging to zero with $\lim_{n\to\infty}\frac{n}{\log(n)}\tau_n^{2kd} = \infty$. Then the following hold:*
(i) If f is uniformly continuous, then

$$\sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f(x)| \xrightarrow{a.s.} 0.$$

(ii) If f is continuous, then for all compact sets $K \subset \mathbb{R}^p$

$$\sup_{x\in K}|f_{\tau_n,n}(x) - f(x)| = 0 \xrightarrow{a.s.} 0.$$

In particular, for all $x \in \mathbb{R}^d$, $\sup_{y\in\overline{B}_{\epsilon_n}(x)}|f_{\tau_n,n}(y) - f(x)| \xrightarrow{a.s.} 0$.

In the remaining of this section we prove Proposition 2.5.1. In Section 2.6, we study the asymptotic limit distribution of the sample τ-approximation. In Section 2.7, we provide several examples of LDFs and verify that they satisfy the VC subgraph property. We begin with a lemma concerning the quantity M_G in (2.4.5).

Lemma 2.5.1 *Let M_G, σ_G, and $c_{G,0}$ be as in Theorem 2.4.3, $\{a_n\}_{n=1}^{\infty}$ be a sequence of positive scalars converging to zero with $\lim_{n\to\infty}\frac{na_n^2}{\log(n)} = \infty$, $b > 0$, and $t_n = \sqrt{n}a_n b$. Then, there are constants $0 < \tilde{c}_G < \infty$ and $\tilde{n} \in \mathbb{N}$ such that, for all $n \geq \tilde{n}$ and $t_n \geq \max(2^3\sigma_G, 2^4 c_{G,0})$,*

$$M_G(n,t_n) \leq \frac{\tilde{c}_G}{n^2}.$$

Proof of Lemma 2.5.1. Since $\lim_{n\to\infty} t_n = \infty$ and $\lim_{n\to\infty} a_n = 0$, there is $n_1 \in \mathbb{N}$, such that, for all $n \geq n_1$, $t_n \geq \max(2^3\sigma_G, 2^4 c_{G,0})$ and $t_n/\sqrt{n} = a_n b \leq 1$. Then, for all $n \geq n_1$, it holds that

$$\begin{aligned} M_G(n,t_n) &\leq 8\exp\left(-\frac{t_n^2}{2^{15}k^2(\sigma_G^2 + l_G)}\right) + 2\exp\left(-\frac{t_n^2}{2^{6+k}k^{k+1}l_G c_{G,0}(\sigma_G^2 + l_G)}\right) \\ &\quad + 8c_{G,1}^{2c_{G,2}}(\sigma_G^2 + 2a_n b l_G)^{-c_{G,2}}\exp\left(-\left(\frac{n\sigma_G^2}{2l_G^2} + \frac{\sqrt{n}t_n}{4l_G}\right)\right) \\ &\leq 16\exp\left(-\frac{t_n^2}{c_{G,3}}\right) + c_{G,4}a_n^{-c_{G,2}}\exp\left(-\frac{\sqrt{n}t_n}{c_{G,5}}\right), \end{aligned}$$

where $c_{G,3} = (\sigma_G^2 + l_G)\max(2^{15}k^2, 2^{6+k}k^{k+1}l_G c_{G,0})$, $c_{G,4} = 8c_{G,1}^{2c_{G,2}}(2bl_G)^{-c_{G,2}}$, and $c_{G,5} = 4l_G$. Next, we use that $\lim_{n\to\infty}\frac{na_n^2}{\log(n)} = \infty$ and obtain that

$$\lim_{n\to\infty} n^2 \exp\left(-\frac{t_n^2}{c_{G,3}}\right) = \lim_{n\to\infty} \exp\left(-\left(\frac{\log(n)}{c_{G,3}}\right)\left(\frac{t_n^2}{\log(n)} - 2c_{G,3}\right)\right) = 0.$$

In particular, there is $n_2 \in \mathbb{N}$, such that, for all $n \geq n_2$, $\exp\left(-\frac{t_n^2}{c_{G,3}}\right) \leq \frac{1}{n^2}$. Next, notice that

$$\begin{aligned} n^2 a_n^{-c_{G,2}} \exp\left(-\frac{\sqrt{n}t_n}{c_{G,5}}\right) &= \exp\left(2\log(n) - \frac{na_n b}{2c_{G,5}}\right)\exp\left(-c_{G,2}\log(a_n) - \frac{na_n b}{2c_{G,5}}\right) \\ &= \exp\left(-\log(n)\left(\frac{b}{2c_{G,5}}\frac{na_n}{\log(n)} - 2\right)\right) \\ &\quad \exp\left(-\frac{b}{2c_{G,5}}na_n\left(1 + \frac{2c_{G,2}c_{G,5}}{b}\frac{\log(a_n)}{na_n}\right)\right). \end{aligned}$$

Now, $\lim_{n\to\infty}\frac{na_n^2}{\log(n)} = \infty$ implies that $\lim_{n\to\infty}\frac{na_n}{\log(n)} = \infty$ and $\lim_{n\to\infty} na_n = \infty$ yielding that

$$\lim_{n\to\infty} n^2 a_n^{-c_{G,2}} \exp\left(-\frac{\sqrt{n}t_n}{c_{G,5}}\right) = 0.$$

This show that there is $n_3 \in \mathbb{N}$ such that $a_n^{-c_{G,2}}\exp\left(-\frac{\sqrt{n}t_n}{c_{G,5}}\right) \leq \frac{1}{n^2}$ for all $n \geq n_3$. Let $\tilde{n} = \max_{i=1,\ldots,3} n_i$. Then, for all $n \geq \tilde{n}$, it holds that

$$M_G(n, t_n) \leq \frac{16 + c_{G,4}}{n^2},$$

and the result follows by setting $\tilde{c}_G = 16 + c_{G,4}$. ∎

Proof of Proposition 2.5.1. For (i), observe that

$$\sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f(x)| \leq \sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f_{\tau_n}(x)| + \sup_{x\in\mathbb{R}^d}|f_{\tau_n}(x) - f(x)|$$

and, by Proposition 2.3.2 (i), it is enough to show that

$$\sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f_{\tau_n}(x)| \xrightarrow{a.s.} 0. \tag{2.5.2}$$

Now, using Lemma 2.3.1, we see that

$$\sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f_{\tau_n}(x)| \leq \sup_{(x,\tau)\in\mathbb{R}^d\times[0,\infty]}\left|\frac{L_{G,n}(x,\tau) - L_G(x,\tau)}{\Lambda_1\tau_n^{kd}}\right|^{1/k}.$$

Let $\epsilon > 0$, $t_n = \sqrt{n}\tau_n^{kd}\Lambda_1\epsilon^k$ and notice that, since $\lim_{n\to\infty} n\tau_n^{2kd} = \infty$, $\lim_{n\to\infty} t_n = \infty$. It follows from Theorem 2.4.3 and Lemma 2.5.1 with $a_n = \tau_n^{kd}$ and $b = \Lambda_1\epsilon^k$ that there are constants $1 < c_{G,0} < \infty$, $0 < \tilde{c}_G < \infty$, and $\tilde{n} \in \mathbb{N}$ such that, for all $n \geq \tilde{n}$,

$t_n \geq \max(2^3\sigma_G, 2^4 c_{G,0})$ and

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f_{\tau_n}(x)| \geq \epsilon\right) = \mathbb{P}\left(\sqrt{n}\sup_{(x,\tau)\in\mathbb{R}^d\times[0,\infty]}|L_{G,n}(x,\tau) - L_G(x,\tau)| \geq t_n\right)$$
$$\leq M_G(n,t_n) \leq \frac{\tilde{c}_G}{n^2}.$$

Therefore, we obtain that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f_{\tau_n}(x)| \geq \epsilon\right) \leq \tilde{n} - 1 + \sum_{n=\tilde{n}}^{\infty} M_G(n,t_n) \leq \tilde{n} - 1 + \sum_{n=\tilde{n}}^{\infty}\frac{\tilde{c}_G}{n^2} < \infty.$$

Now, (2.5.2) follows from Borel-Cantelli lemma. The proof of the first part of (ii) follows from the inequality

$$\sup_{x\in K}|f_{\tau_n,n}(x) - f(x)| \leq \sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f_{\tau_n}(x)| + \sup_{x\in K}|f_{\tau_n}(x) - f(x)|,$$

(2.5.2), and Proposition 2.3.2 (ii). For the second part of (ii), let $\epsilon^* > 0$ and $n^* \in \mathbb{N}$ such that $\epsilon_n \leq \epsilon^*$ for all $n \geq n^*$. Then, for all $n \geq n^*$ and $x \in \mathbb{R}^d$,

$$\sup_{y\in\overline{B}_{\epsilon_n}(x)}|f_{\tau_n,n}(y) - f(x)| \leq \sup_{y\in\overline{B}_{\epsilon^*}(x)}|f_{\tau_n,n}(y) - f_{\tau_n}(y)| + \sup_{y\in\overline{B}_{\epsilon_n}(x)}|f_{\tau_n}(y) - f(x)|.$$

Now, using the compactness of $\overline{B}_{\epsilon^*}(x)$ and the first part of (ii), we have that

$$\lim_{n\to\infty}\sup_{y\in\overline{B}_{\epsilon^*}(x)}|f_{\tau_n,n}(y) - f_{\tau_n}(y)| \xrightarrow{a.s.} 0.$$

Finally, Proposition 2.3.2 (ii) implies that

$$\lim_{n\to\infty}\sup_{y\in\overline{B}_{\epsilon_n}(x)}|f_{\tau_n}(y) - f(x)| = 0.$$

■

2.6 Central limit results for sample τ-approximations

It is well known that extreme localization is an important concept in depth analysis, however, the fluctuations of $f_{\tau,n}$ are unknown. Our main result in this section characterizes the asymptotic variance and establishes a related limit distribution. To this end, let

$$\hat{\Lambda}_1^2 = \int \tilde{\Lambda}_1^2(x_1)dx_1,$$

where, for $x_1 \in \mathbb{R}^d$,

$$\tilde{\Lambda}_1(x_1) = \int G(x_1,\ldots,x_d)dx_2\ldots dx_d.$$

Theorem 2.6.1 *Let $P \in \mathcal{P}_{d,\ll\lambda}$ with continuous density f and suppose (2.2.1) holds true. Let $x \in S_f$ and $\{\tau_n\}_{n=1}^{\infty}$ be a sequence of positive scalars converging to zero.*

If $\lim_{n\to\infty}\sqrt{n}\tau_n^{((2k-1)d)/2}=\infty$, *then*

$$\sqrt{n}\tau_n^{d/2}\left(f_{\tau_n,n}(x)-f_{\tau_n}(x)\right)\xrightarrow{d}N\left(0,\frac{\hat{\Lambda}_1^2}{\Lambda_1^2}f(x)\right).$$

Remark 2.6.1 *We notice that, for* $k>1$, *the limit distribution in Theorem 2.6.1 with* f_{τ_n} *replaced by* f *cannot hold. In fact, the deterministic term* $f_{\tau_n}(x)-f(x)$ *is, by Proposition 2.3.3, of order* $O(\tau_n^2)$, *while the term* $f_{\tau_n,n}(x)-f_{\tau_n}(x)$ *converges to a normal distribution at rate* $1/(\sqrt{n}\tau_n^{d/2})$. *Since, necessarily,* $\lim_{n\to\infty}\sqrt{n}\tau_n^{((2k-1)d)/2}=\infty$, $f_{\tau_n}(x)-f(x)$ *is the dominant term. On the other hand, if* $k=1$, $\lim_{n\to\infty}\sqrt{n}\tau_n^{d/2}=\infty$ *and* $\lim_{n\to\infty}\sqrt{n}\tau_n^{d/2+2}=0$, *then, by Proposition 2.3.3,* $\lim_{n\to\infty}\sqrt{n}\tau_n^{d/2}\left(f_{\tau_n}(x)-f(x)\right)=0$. *Hence,*

$$\sqrt{n}\tau_n^{d/2}\left(f_{\tau_n,n}(x)-f(x)\right)\xrightarrow{d}N\left(0,\frac{\hat{\Lambda}_1^2}{\Lambda_1^2}f(x)\right).$$

In the examples with uniform kernel, the constant Λ_1 appearing in the limiting variance in Theorem 2.6.1 can be calculated numerically using (2.2.1) by computing the percentage of uniformly distributed random points in $(\overline{B}_\rho(0))^k$ that lie in $Z_{G,0,1}$ (e.g., for $G=K_\beta$, $k=2$ and $\rho=\sqrt{2}\min(1,\beta/2)$) and multiplying the result by its volume

$$\lambda^k((\overline{B}_\rho(0))^k)=(\lambda(\overline{B}_\rho(0)))^k=\left(\frac{\pi^{d/2}\rho^d}{\Gamma(d/2+1)}\right)^k,$$

where Γ is the gamma function. Similarly, the constant $\hat{\Lambda}_1^2$ can be calculated by approximating the integral with a sum. An alternative form for Theorem 2.6.1 without the factor $f(x)$ in the variance term is given in the following corollary.

Corollary 2.6.1 *Under the hypothesis of Theorem 2.6.1,*

$$\sqrt{n}\tau_n^{d/2}\left(\sqrt{f_{\tau_n,n}(x)}-\sqrt{f_{\tau_n}(x)}\right)\xrightarrow{d}N\left(0,\frac{\hat{\Lambda}_1}{4\Lambda_1^2}\right).$$

We now turn to the proof of Theorem 2.6.1 and Corollary 2.6.1. To this end, notice that, by Hoeffding's decomposition of U-statistics, the limit behavior of $L_G(x,\tau)$ for $\tau\to 0^+$ critically depends on the limit behavior of $\frac{1}{n}\sum_{i=1}^n p_{k,1}h_{x,\tau}(X_i)$ for $\tau\to 0^+$, which has variance $\frac{1}{n}\mathbb{V}ar[p_{k,1}h_{x,\tau}(X_1)]$. Under the hypothesis of Theorem 2.6.1, Proposition 2.4.1 yields that $\mathbb{V}ar[p_{k,1}h_{x,\tau}(X_1)]>0$ for all $\tau\in(0,\infty)$. Using that $P\in\mathcal{P}_{d,\ll\lambda}$, we also see that $p_{k,1}h_{x,0}=0$ P-a.s. and $\mathbb{V}ar[p_{k,1}h_{x,0}(X_1)]=0$. We begin by studying the order of convergence of $\mathbb{V}ar[p_{k,1}h_{x,\tau}(X_1)]$ to 0, as $\tau\to 0^+$.

Lemma 2.6.1 *Suppose (2.2.1) holds true. If f is continuous, then*

$$\lim_{\tau\to 0^+}\frac{\mathbb{V}ar[p_{k,1}h_{x,\tau}(X_1)]}{\tau^{(2k-1)d}}=\hat{\Lambda}_1^2 f^{2k-1}(x).$$

Proof of Lemma 2.6.1. Let $\tau>0$. We compute

$$\frac{\mathbb{V}ar[p_{k,1}h_{x,\tau}(X_1)]}{\tau^{(2k-1)d}}=\frac{\mathbb{V}ar[\overline{p}_{k,1}h_{x,\tau}(X_1)]}{\tau^{(2k-1)d}}-\left(\frac{L_G(x,\tau)}{\tau^{(k-1/2)d}}\right)^2, \tag{2.6.1}$$

where, by Theorem 2.2.1 (i),

$$\lim_{\tau\to 0^+}\frac{L_G(x,\tau)}{\tau^{kd}}=\Lambda_1 f^k(x)\quad\text{and}\quad\lim_{\tau\to 0^+}\frac{L_G(x,\tau)}{\tau^{(k-1/2)d}}=0.$$

We now focus on the first term in (2.6.1). By changing variables twice, we note that

$$\frac{\mathbb{V}ar[\overline{p}_{k,1}h_{x,\tau}(X_1)]}{\tau^{(2k-1)d}} = \frac{1}{\tau^{(2k-1)d}}\int\left(\int h_{x,\tau}(x_1,\ldots,x_k)\prod_{j=2}^{k}f(x_j)dx_2\ldots dx_k\right)^2 f(x_1)dx_1$$
$$= \frac{1}{\tau^{2(k-1)d}}\int\left(\int G\left(x_1,\frac{x_2-x}{\tau},\ldots,\frac{x_k-x}{\tau}\right)\prod_{j=2}^{k}f(x_j)dx_2\ldots dx_k\right)^2 f(x+\tau x_1)dx_1$$
$$= \int\left(\int G(x_1,\ldots,x_k)\prod_{j=2}^{k}f(x+\tau x_j)dx_2\ldots dx_k\right)^2 f(x+\tau x_1)dx_1.$$

Since f is continuous, it follows from the boundedness of G, (2.2.1) and LDCT that

$$\lim_{\tau\to 0^+}\int G(x_1,\ldots,x_k)\prod_{j=2}^{k}f(x+\tau x_j)dx_2\ldots dx_k$$
$$=f^{k-1}(x)\int G(x_1,\ldots,x_k)dx_2\ldots dx_k$$

and

$$\lim_{\tau\to 0^+}\int\left(\int G(x_1,\ldots,x_k)\prod_{j=2}^{k}f(x+\tau x_j)dx_2\ldots dx_k\right)^2 f(x+\tau x_1)dx_1$$
$$=f^{2k-1}(x)\int\left(\int G(x_1,\ldots,x_k)dx_2\ldots dx_k\right)^2 dx_1$$

■

Proof of Theorem 2.6.1. Using Hoeffding's decomposition of U-statistics (Proposition 1.10.1 with $E=\mathbb{R}^d$ and $h=h_{x,\tau_n}$), it follows that

$$L_{G,n}(x,\tau_n)-L_G(x,\tau_n) = \frac{k}{n}\sum_{i=1}^{n}p_{k,1}h_{x,\tau_n}(X_i)+\sum_{j=2}^{k}\binom{k}{j}\tilde{r}_{n,j}, \tag{2.6.2}$$

where

$$\tilde{r}_{n,j}=\binom{n}{j}^{-1}\sum_{1\le i_1<\cdots<i_j\le n}p_{k,j}h_{x,\tau_n}(X_{i_1},\ldots,X_{i_j}).$$

Now, applying Lindeberg-Levy Theorem for triangular arrays (Billingsley, 2012, Theorem 27.2) with

$$r_n=n,\ s_n=\sqrt{n}\sqrt{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]},\ \text{and}\ S_n=\sum_{i=1}^{n}p_{k,1}h_{x,\tau_n}(X_i),$$

it follows that

$$\sqrt{\frac{n}{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}\left(\frac{1}{n}\sum_{i=1}^{n}p_{k,1}h_{x,\tau_n}(X_i)\right)\xrightarrow{d}N(0,1), \tag{2.6.3}$$

provided the Lindeberg condition (Billingsley, 2012, Equation (27.8))

$$\lim_{n\to\infty}\frac{1}{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}\int_{A_{n,\epsilon}}(p_{k,1}h_{x,\tau_n}(x_1))^2 f(x_1)dx_1=0 \tag{2.6.4}$$

holds for all $\epsilon > 0$, where

$$A_{n,\epsilon} = \{x_1 \in \mathbb{R}^d : (p_{k,1}h_{x,\tau_n}(x_1))^2 \geq \epsilon^2 n \mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]\}.$$

Using (2.2.3), it holds that $(p_{k,1}h_{x,\tau_n}(x_1))^2 \leq l_G^2$, for all $x, x_1 \in \mathbb{R}^d$. Also, due to $x \in S_f$ and $\lim_{n\to\infty} n\tau_n^{(2k-1)d} = \infty$, Lemma 2.6.1 implies that $\lim_{n\to\infty} n\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)] = \infty$. Let $\tilde{n} \in \mathbb{N}$ be such that $l_G^2 < \epsilon^2 n \mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]$ for all $n \geq \tilde{n}$. It follows that $A_{n,\epsilon} = \emptyset$ for all $n \geq \tilde{n}$. Thus, (2.6.4) holds true and we obtain (2.6.3). Next, we use (1.10.4) with $h, r_{n,j}$ and n replaced by h_{x,τ_n}, $\tilde{r}_{n,j}$, and $\frac{n}{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}$, and obtain that, for all $\epsilon > 0$,

$$\mathbb{P}\left(\sqrt{\frac{n}{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}\tilde{r}_{n,j} \geq \epsilon\right) \leq \frac{n\binom{n}{j}^{-1}}{\epsilon^2 \mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}\mathbb{E}[(p_{k,j}h_{x,\tau_n}(X_1,\ldots,X_j))^2].$$

Since $j \geq 2$ and $\lim_{n\to\infty} n\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)] = \infty$, it follows that

$$\lim_{n\to\infty} \mathbb{P}\left(\sqrt{\frac{n}{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}\tilde{r}_{n,j} \geq \epsilon\right) = 0. \tag{2.6.5}$$

From (2.6.2), (2.6.3), and (2.6.5), we conclude that

$$\sqrt{\frac{n}{k^2\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}(L_{G,n}(x,\tau_n) - L_G(x,\tau_n)) \xrightarrow{d} N(0,1). \tag{2.6.6}$$

Now, using the delta method we obtain

$$\sqrt{\frac{n}{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}(L_G(x,\tau_n))^{1-1/k}\left((L_{G,n}(x,\tau_n))^{1/k} - (L_G(x,\tau_n))^{1/k}\right) \xrightarrow{d} N(0,1);$$

equivalently,

$$Z_n = \sqrt{\frac{n}{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}(\tau_n^{kd}\Lambda_1 f_{\tau_n}^{k-1}(x))\,(f_{\tau_n,n}(x) - f_{\tau_n}(x)) \xrightarrow{d} N(0,1). \tag{2.6.7}$$

To complete the proof, since $x \in S_f$ and $\tau_n > 0$, it holds, by Theorem 2.2.1 (i), that

$$\lim_{n\to\infty}\frac{f_{\tau_n}^k(x)}{f^k(x)} = \lim_{n\to\infty}\frac{L_G(x,\tau_n)}{\Lambda_1\tau_n^{kd}f^k(x)} = 1 \tag{2.6.8}$$

and, by Lemma 2.6.1,

$$\lim_{n\to\infty}\frac{\sqrt{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}{\tau_n^{(k-1/2)d}} = \hat{\Lambda}_1 f^{k-1/2}(x) > 0. \tag{2.6.9}$$

(2.6.8) and (2.6.9) imply that

$$\begin{aligned}
Y_n &= \frac{\sqrt{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}{\tau_n^{(k-1/2)d}f_{\tau_n}^{k-1}(x)} \cdot \frac{1}{\hat{\Lambda}_1 f^{\frac{1}{2}}(x)} \\
&= \frac{\sqrt{\mathbb{V}ar[p_{k,1}h_{x,\tau_n}(X_1)]}}{\tau_n^{(k-1/2)d}} \cdot \frac{1}{\hat{\Lambda}_1 f^{k-1/2}(x)} \cdot \frac{f^{k-1}(x)}{f_{\tau_n}^{k-1}(x)} \xrightarrow[n\to\infty]{} 1.
\end{aligned}$$

From (2.6.7) and Slutsky's Theorem it follows that

$$Y_n Z_n \xrightarrow{d} N(0,1),$$

completing the proof. ■

Proof of Corollary 2.6.1. Using that $\lim_{n\to\infty} \sqrt{n}\tau_n^{d/2} = \infty$ and Theorem 2.6.1 we obtain that $f_{\tau_n,n}(x) - f_{\tau_n}(x) \xrightarrow{d} 0$. Also, (ii) of Proposition 2.3.2 yields that $\lim_{n\to\infty}|f_{\tau_n}(x) - f(x)| = 0$. It follows that

$$\frac{f_{\tau_n,n}(x)}{f(x)} - 1 = \frac{(f_{\tau_n,n}(x) - f_{\tau_n}(x)) + (f_{\tau_n}(x) - f(x))}{f(x)} \xrightarrow{d} 0 \tag{2.6.10}$$

By Theorem 2.6.1 it holds that

$$\tilde{Z}_n = \sqrt{n}\tau_n^{d/2}\tilde{Y}_n\left(\sqrt{f_{\tau_n,n}(x)} - \sqrt{f_{\tau_n}(x)}\right) \xrightarrow{d} N\left(0, \frac{\hat{\Lambda}_1^2}{\Lambda_1^2}\right),$$

where

$$\tilde{Y}_n = \frac{\sqrt{f_{\tau_n,n}(x)} + \sqrt{f_{\tau_n}(x)}}{\sqrt{f(x)}}.$$

Using (2.6.8) and (2.6.10) we obtain that

$$\tilde{Y}_n = \sqrt{\frac{f_{\tau_n,n}(x)}{f(x)}} + \sqrt{\frac{f_{\tau_n}(x)}{f(x)}} \xrightarrow{d} 2,$$

and, by Slutsky's Theorem, we conclude that

$$\frac{\tilde{Z}_n}{\tilde{Y}_n} \xrightarrow{d} N\left(0, \frac{\hat{\Lambda}_1^2}{4\Lambda_1^2}\right).$$

■

An extension of Theorem 2.6.1 uniformly over S_f, namely,

$$\sqrt{n}\tau_n^{d/2}(f_{\tau_n,n}(\cdot) - f_{\tau_n}(\cdot)) \xrightarrow{d^*} \frac{\hat{\Lambda}_1}{\Lambda_1}(W_k^*(f))(\cdot),$$

where $\{(W_k^*(f))(x)\}_{x\in S_f}$ is a centered Gaussian process with the covariance function $\gamma : S_f \times S_f \to \mathbb{R}$ given by $\gamma_{W_k^*(f)}(x,y) = \sqrt{f(x)f(y)}$, requires an extension of the results of Arcones and Giné (1993) to triangular arrays and it is beyond the scope of the present work. A result in this direction, when the kernel is uniform, is given by Schneemeier (1989), but this is not sufficient in this context since the sets $\{Z_{G,x,\tau_n}\}_{n=1}^{\infty}$ not only depend on x but also on n.

2.7 Examples

In this section, we provide additional examples of LDFs and verify that they satisfy the VC subgraph property. We begin with local simplicial depth (Agostinelli and Romanazzi, 2008).

Example 2.7.1 (Local simplicial depth) *$G = S = \mathbf{I}_{Z_{S,0,1}}$, where for $x \in \mathbb{R}^d$ and $\tau \in [0,\infty]$*

$$Z_{S,x,\tau} = \{(x_1,\dots,x_{d+1}) \in (\mathbb{R}^d)^{d+1} : x \in \Delta[x_1,\dots,x_{d+1}], \max_{1\le i<j\le d+1} \|x_i - x_j\|_2 \le \tau\}.$$

Notice that that the last constraint $\max_{1\le i<j\le d+1} \|x_i - x_j\|_2 \le \tau$ restricts the indicator function $h_{S,x,\tau} = \mathbf{I}_{Z_{S,x,\tau}}$ to simplices $\Delta[x_1,\dots,x_{d+1}]$ with side lengths smaller than τ. Next, observe that the class of simplices in $\mathbb{R}^d$ is VC, since they are given by the intersections of $d+1$ halfspaces (see Lemma 6.6 and Corollary 6.7 of Arcones and Giné (1993)). Finally, using that the set $Z_{S,0,1}$ is closed, we see that $\mathbf{I}_{Z_{S,0,1}}$ is Borel measurable.

Example 2.7.2 (Local β-skeleton depth) *For some $\beta \ge 1$, $G = K_\beta = \mathbf{I}_{Z_{K_\beta,0,1}}$, where for $x \in \mathbb{R}^d$ and $\tau \in [0,\infty]$*

$$Z_{K_\beta,x,\tau} = \{(x_1,x_2) \in (\mathbb{R}^d)^2 : \max_{(i,j)\in\{(1,2),(2,1)\}} \|x_i + (2/\beta - 1)x_j - 2/\beta x\|_2 \le \|x_1 - x_2\|_2 \le \tau\}.$$

By taking $\beta = 1$ and $\beta = 2$ in the above example, we obtain local spherical depth and local lens depth (Kleindessner and Von Luxburg, 2017), respectively. We will now verify that this class of depth functions satisfy the VC subgraph property. Let $\mathrm{B} = \{\overline{B}_r(x) : x \in \mathbb{R}^d, r > 0\}$ be the class of balls in $\mathbb{R}^d$ and, for $\beta \ge 1$, $\mathrm{K}_\beta = \{\overline{B}_{\frac{\beta}{2}\|x_1-x_2\|_2}(\frac{\beta}{2}x_1 + (1-\frac{\beta}{2})x_2) \cap \overline{B}_{\frac{\beta}{2}\|x_1-x_2\|_2}((1-\frac{\beta}{2})x_1 + \frac{\beta}{2}x_2) : x_1, x_2 \in \mathbb{R}^d\}$ be the class of all β-skeleton sets. By Theorem 1 in Dudley (1979), B is a VC class of sets. Applying Proposition 1.11.1 (ii), it follows that also the intersection $\mathrm{B} \cap \mathrm{B}$ is a VC class of sets. Proposition 1.11.1 (vi) implies that $\mathrm{K}^*_{[1,\infty)} = \cup_{\beta\in[1,\infty)} \mathrm{K}_\beta \subset \mathrm{B} \cap \mathrm{B}$ is a VC class. In particular, K_β is a VC class for all $\beta \ge 1$. We finally notice that the function $\mathbf{I}_{Z_{K_{(\cdot)},0,1}}(\cdot)$ is jointly Borel measurable. Thus, assumptions **(A1)**-**(A4)** hold true.

At the end of Section 1.7 we notice that *Type B* depth functions can be transformed into *Type A* by applying the function g directly to $i_{G,x,\infty}$ in Definition 1.7.1 and computing the integral of $h_{G,x,\infty} = g \circ i_{G,x,\infty}$ w.r.t. P^k. Of course, the depth function obtained in this way is different from the original depth function and there's no guarantee that it is a *statistical* depth function. However, this observation allows one to define *Type A* local depth versions for L^q-depth and simplicial volume depth (see Examples 1.7.1 and 1.7.2). To this end, let $g : [0,\infty) \to \mathbb{R}$ be continuous, positive, and decreasing with $\lim_{t\to\infty} g(t) = 0$. The resulting local depth functions are not generated by a kernel function $G \in \mathcal{G}$ and G, N_q, and V below are just indices.

Example 2.7.3 (Local L^q-depth) *Let $G = N_q$, $k_{N_q} = 1$ and the function $h_{N_q,x,\tau} : \mathbb{R}^d \to [0,\infty)$ is given by*

$$h_{N_q,x,\tau}(x_1) = g(\|x_1 - x\|_q)\mathbf{I}_{Z_{N_q,x,\tau}}(x_1)$$

and $Z_{N_q,x,\tau} = \{y \in \mathbb{R}^d : \|y - x\|_q \le \tau\}$ is the closed L^q-ball with center x and radius τ.

Notice that the VC dimension of L^q-balls in $\mathbb{R}^d$ is finite for $q = 2,\infty$ (see Dudley (1979) and Despres (2017)). This is true also for $q = 1$, since L^1-balls are given by intersections of halfspaces. Hence, the function $\|\cdot\|_q$ is VC subgraph for $q = 1,2,\infty$. Since g is monotone, using Proposition 1.11.2 (vi), we see that $g(\|\cdot\|_q)$ is VC subgraph for $q = 1,2,\infty$, and hence so is $h_{N_q,x,\tau}$. We now turn to local simplicial volume depth (Agostinelli and Romanazzi, 2008).

Example 2.7.4 (Local simplicial volume depth) *In this case, $G = V$, $k_V = d$ and the function $h_{V,x,\tau} : (\mathbb{R}^d)^d \to [0, \infty)$ is given by*

$$h_{V,x,\tau}(x_1, \ldots, x_d) = g(\lambda(\Delta[x, x_1, \ldots, x_d]))\mathbf{I}_{Z_{V,x,\tau}}(x_1, \ldots, x_d)$$

where for $x \in \mathbb{R}^d$ and $\tau \in [0, \infty]$

$$Z_{V,x,\tau} = \{(x_1, \ldots, x_d) \in (\mathbb{R}^d)^d : \max_{1 \le i < j \le d} \|x_i - x_j\|_2 \le \tau, \max_{i=1,\ldots,d} \|x_i - x\|_2 \le \tau\}.$$

We show next that kernel density techniques can also be developed using *Type A* LDFs. We begin with the uniform kernel (see Devroye and Györfi (1985)).

Example 2.7.5 (Uniform kernel) *Let $k_G = 1$ and*

$$G = \tilde{U} = \mathbf{I}_{\overline{B}_1(0))}.$$

Since closed balls in $\mathbb{R}^d$ form a VC class of sets by Theorem 1 in Dudley (1979), it follows that $h_{\tilde{U},x,\tau} = \mathbf{I}_{\overline{B}_\tau(x))}$ belongs to the VC subgraph class for all $x \in \mathbb{R}^d$ and $\tau \in (0, \infty]$.

Example 2.7.6 (Gaussian kernel) *Set $k = 1$ and*

$$G = \tilde{K} = \exp(-\|\cdot\|_2^2/2).$$

It follows that $\Lambda_{\tilde{K},1} = 1$ and $h_{\tilde{K},x,\tau} = \exp(-\|(\cdot - x)/\tau\|_2^2/2)$. In particular, $\tau^{-d} h_{\tilde{K},x,\tau}$ is a Gaussian kernel with covariance matrix $\tau^2 I$ (see, for instance, Chacón and Duong (2018)). Also, the h-depth (Cuevas et al., 2007), used in functional data, can be obtained by scaling $h_{\tilde{K},x,\tau}$ by τ. As a last example, we consider LDFs generated using continuous bump functions (see e.g. Section 13 in Tu (2011)). These are non-negative, continuous functions $G : \mathbb{R}^d \to \mathbb{R}$ with bounded support. From the continuity and bounded support assumption it follows that G has finite integral. Under the additional assumptions that $G(0) > 0$ and G is non-increasing along any ray from the origin $0 \in \mathbb{R}^d$, we see that **(P2)** and **(P4)** hold.

Example 2.7.7 (Bump functions) *Let $k_G = 1$ and $G = \tilde{B}$, where $\tilde{B} : \mathbb{R}^d \to \mathbb{R}$ is non-negative and continuous with bounded support. Moreover, $\tilde{B}(0) > 0$ and $\tilde{B}$ satisfies* **(P2)**.

Bump functions can be constructed, for instance, by the following procedure. Let $g_1 : \mathbb{R} \to \mathbb{R}$ be positive, continuous and increasing with $\lim_{t\to\infty} g_1(t) = \infty$. Set $g_2(t) = 1/g_1(1/t)$ and

$$g_3(t) = \begin{cases} g_2(1+t)g_2(1-t) & \text{if } |t| < 1 \\ 0 & \text{if } |t| \ge 1. \end{cases}$$

Finally, let $G(x) = g_3(\|x\|_2)$. Alternative, one can let $G(x)$ be the product $\prod_{i=1}^d g_3(\pi_{e_i}(x))$. Additional smoothness can be added by requiring that g_1 has continuous derivatives of all orders and $\lim_{t\to\infty} g_1(t)/t^n = \infty$, for all $n \in \mathbb{N}$. This last assumption ensures that G decays quickly to zero near the boundary of its support. For instance, by taking $g_1(t) = e^{t/2}$, we get the classical bump function

$$G(x) = \begin{cases} e^{-1/(1-\|x\|_2^2)} & \text{if } \|x\|_2 < 1 \\ 0 & \text{if } \|x\|_2 \ge 1, \end{cases}$$

which has continuous derivatives of all orders and decays exponentially fast as $\|x\|_2 \to 1^-$.

2.8 Choice of localizing parameter

A key issue in the use of LDFs is that it requires a method to choose τ. A typical approach, as for kernel density techniques, is to choose τ so as to minimize the squared error. This involves the square of the bias and the variance term. We begin by calculating these terms. To this end, notice that, by the Newton generalized binomial theorem, for $t \in (-1,1)$, $(1+t)^{1/k} = \sum_{j=0}^{\infty} \binom{1/k}{j} t^j$, where $\binom{1/k}{j} = (1/k \dots (1/k - j + 1))/j!$. Setting $t = \frac{f_\tau^k(x) - f^k(x)}{f^k(x)}$ and $t = \frac{f_{\tau,n}^k(x) - f_\tau^k(x)}{f_\tau^k(x)}$, respectively, we see that, for $x \in S_f$, τ small enough, and n large enough,

$$f_\tau(x) = \sum_{j=0}^{\infty} \binom{1/k}{j} f^{1-kj}(x)(f_\tau^k(x) - f^k(x))^j, \quad \text{and} \tag{2.8.1}$$

$$f_{\tau,n}(x) = \sum_{j=0}^{\infty} \binom{1/k}{j} f_\tau^{1-kj}(x)(f_{\tau,n}^k(x) - f_\tau^k(x))^j. \tag{2.8.2}$$

Next, we show that for all $t \in (-1,1)$ and $m \geq 0$

$$0 < (1-|t|)^{1/k} \leq \sum_{j=0}^{m} \binom{1/k}{j} t^j \leq (1+|t|)^{1/k} + |\binom{1/k}{2}||t|^2 < 2^{1/k} + \frac{k-1}{2k^2}. \tag{2.8.3}$$

First, notice that that for all $j \geq 0$

$$\binom{1/k}{j}(-|t|)^j \leq \binom{1/k}{j} t^j \leq \binom{1/k}{j}|t|^j$$

yielding that

$$\sum_{j=0}^{m} \binom{1/k}{j}(-|t|)^j \leq \sum_{j=0}^{m} \binom{1/k}{j} t^j \leq \sum_{j=0}^{m} \binom{1/k}{j}|t|^j.$$

Since $\binom{1/k}{j}(-|t|)^j \leq 0$ for all $j \geq 1$, we obtain the second inequality in (2.8.3). We now turn to the third inequality in (2.8.3). Since $|\binom{1/k}{j}||t|^j \geq |\binom{1/k}{j+1}||t|^{j+1}$ for all $j \geq 0$ and the coefficients $\binom{1/k}{j}$ have alternating signs for $j \geq 1$, we obtain that

$$\sum_{j=m+1}^{\infty} \binom{1/k}{j}|t|^j \geq \begin{cases} 0 & \text{for } m \text{ even} \\ -|\binom{1/k}{m+1}||t|^{m+1} \geq -|\binom{1/k}{2}||t|^2 & \text{for } m \text{ odd.} \end{cases}$$

Therefore, we conclude that

$$\sum_{j=0}^{m} \binom{1/k}{j}|t|^j \leq (1+|t|)^{1/k} + |\binom{1/k}{2}||t|^2.$$

Using (2.8.3) with $t = \frac{f_{\tau,n}^k(x) - f_\tau^k(x)}{f_\tau^k(x)}$, (2.8.2), and LDCT, we obtain that

$$\begin{aligned}\mathbb{E}[f_{\tau,n}(x)] &= \mathbb{E}\left[\lim_{m\to\infty}\sum_{j=0}^{m}\binom{1/k}{j} f_\tau^{1-kj}(x)(f_{\tau,n}^k(x) - f_\tau^k(x))^j\right] \\ &= \lim_{m\to\infty}\mathbb{E}\left[\sum_{j=0}^{m}\binom{1/k}{j} f_\tau^{1-kj}(x)(f_{\tau,n}^k(x) - f_\tau^k(x))^j\right] \\ &= \sum_{j=0}^{\infty}\binom{1/k}{j} f_\tau^{1-kj}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^j]. \end{aligned} \tag{2.8.4}$$

Using Proposition 1.10.6 with $h = h_{x,\tau}$ and $q = j \geq 2$, one obtains that

$$\mathbb{E}[(L_{G,n}(x,\tau) - L_G(x,\tau))^j] = O(n^{-j/2}),$$

implying

$$f_\tau^{-kj}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^j] = O(n^{-j/2}). \tag{2.8.5}$$

(2.8.5) suggests that

$$\sum_{j=3}^{\infty}\binom{1/k}{j} f_\tau^{1-kj}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^j] = o\left(\frac{1}{n}\right). \tag{2.8.6}$$

Indeed, (2.8.6) can be obtained by a careful computation of the LHS of (2.8.5) along the lines of Lemma 5.2.2.B of Serfling (2009) (see also Theorem 1.5.1 of Lee (1990)). Using (2.8.4), the unbiasedness of $f_{\tau,n}^k$, and (2.8.6), we have that

$$\mathbb{E}[f_{\tau,n}(x)] = f_\tau(x) + \frac{1-k}{2k^2} f_\tau^{1-2k}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^2] + o\left(\frac{1}{n}\right). \tag{2.8.7}$$

It follows that the squared bias term is given by

$$\begin{aligned}(\mathbb{E}[f_{\tau,n}(x)] - f(x))^2 &= (f_\tau(x) - f(x))^2 \\ &\quad + (f_\tau(x) - f(x))\frac{1-k}{k^2} f_\tau^{1-2k}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^2] + o\left(\frac{1}{n}\right).\end{aligned}$$

Similarly, one can show that

$$\begin{aligned}\mathbb{E}[f_{\tau,n}^2(x)] &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{1/k}{i}\binom{1/k}{j} f_\tau^{2-k(i+j)}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^{i+j}] \\ &= f_\tau^2(x) + \frac{2-k}{k^2} f_\tau^{2-2k}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^2] + o\left(\frac{1}{n}\right).\end{aligned} \tag{2.8.8}$$

It follows from (2.8.7) and (2.8.8) that

$$\begin{aligned}\mathbb{V}ar[f_{\tau,n}(x)] &= \mathbb{E}[f_{\tau,n}^2(x)] - (\mathbb{E}[f_{\tau,n}(x)])^2 \\ &= \frac{1}{k^2} f_\tau^{2-2k}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^2] + o\left(\frac{1}{n}\right).\end{aligned}$$

Thus, we conclude that

$$(\mathbb{E}[f_{\tau,n}(x)] - f(x))^2 + \mathbb{V}ar[f_{\tau,n}(x)] = (f_\tau(x) - f(x))^2$$
$$+\frac{1}{k^2}\Big(f_\tau(x) + (1-k)(f_\tau(x) - f(x))\Big) f_\tau^{1-2k}(x)\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^2] + o\left(\frac{1}{n}\right).$$

Notice that by (2.4.2)

$$\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^2] = \frac{1}{n}\left(\frac{k}{\Lambda_1 \tau^{kd}}\right)^2 \mathbb{V}ar[p_{k,1}h_{x,\tau}(X_1)] + o\left(\frac{1}{n}\right)$$

and by Lemma 2.6.1

$$\lim_{\tau \to 0^+} \frac{\mathbb{V}ar[p_{k,1}h_{x,\tau}(X_1)]}{\tau^{(2k-1)d}} = \hat{\Lambda}_1^2 f^{2k-1}(x),$$

implying that

$$\mathbb{E}[(f_{\tau,n}^k(x) - f_\tau^k(x))^2] = O(n^{-1}\tau^{-d}).$$

Next, using (2.8.1) and Theorem 2.2.1 (iii), we have that

$$f_\tau(x) = f(x) + \frac{1}{k\Lambda_1} f_\tau^{1-k}(x)(R_1(x) + R_2(x))\tau^2 + o(\tau^2).$$

Therefore, it holds that

$$(\mathbb{E}[f_{\tau,n}(x)] - f(x))^2 + \mathbb{V}ar[f_{\tau,n}(x)] = O(\tau^4) + O(n^{-1}\tau^{-d}).$$

By imposing the same order of convergence on the terms τ_n^4 and $n^{-1}\tau_n^{-d}$, for some sequence $\{\tau_n\}_{n=1}^{\infty}$, we have that τ_n is $\tau_n = O(n^{-1/(d+4)})$, and the rate of convergence is $n^{-4/(d+4)}$. Thus, an optimal choice for τ_n is $\tau_n = O(n^{-1/(d+4)})$

Chapter 3

Applications to clustering, mode estimation, and upper level set estimation

3.1 Introduction

As described in Section 3 and Appendices A, D, E, F, G, I, J, and K of Francisci et al. (2020), the developments in Chapter 2 allow further applications to clustering, mode estimation, and upper level set estimation. This is the main content of this chapter. We suppose throughout that P is absolutely continuous with respect to the Lebesgue measure with density f.

Under appropriate differentiability assumptions, Proposition 2.3.2 (iv) shows that the derivatives of f_τ converge uniformly over compact sets to those of f, which facilitates an inquiry into the modes of the density via a gradient system analysis. This, in turn, allows one to characterize the related *stable manifolds* paving the way for cluster analysis. Related ideas about clustering appear in Chazal et al. (2013), Chen et al. (2016), and Genovese et al. (2016). Our methodology differs from the existing literature in that we take advantage of the local depth notion, specifically the τ-approximation f_τ and its properties, developed in Chapter 2, as an approximation to the density.

We recall from dynamical systems that the stable manifold generated by a mode m of a "smooth" density f is given by

$$C(m) = \{x \in S_f : \lim_{t\to\infty} u_x(t) = m\},$$

where $u_x(t)$ is the solution at time t of the gradient system

$$u'(t) = \nabla f(u(t)) \tag{3.1.1}$$

with initial value $u(0) = x$ and ∇f represents the gradient of f. If $m_1, \cdots m_M$ are the modes of f, then the clusters associated with f are given by $C(m_1), \ldots, C(m_M)$ (Chacón, 2015). We verify in Section 3.3 that these clusters are well-defined, non-trivial and disjoint using Lyapunov's theory in dynamical systems. Additionally, we establish that

$$S_f = \cup_{i=1}^{M} C(m_i) \cup \cup_{l=1}^{L} C(\mu_l),$$

where $\mu_1, \ldots, \mu_L$ are the other stationary points of f. Hence, $C(m_1), \ldots, C(m_M)$, $C(\mu_1), \ldots, C(\mu_L)$ form a partition of S_f. Additionally, we show that the set $S_f \setminus$

$(\cup_{i=1}^{M} C(m_i)) = \cup_{l=1}^{L} C(\mu_l)$ has (topological) dimension smaller than d and the clusters $C(m_1), \ldots, C(m_M)$ are separated in S_f by the lower dimensional stable manifolds $C(\mu_1), \ldots, C(\mu_L)$.

In real applications, f is unknown and is replaced by its estimate $f_{\tau,n}$. For every fixed τ, by the results in Section 2.4, $f_{\tau,n}$ converges to f_τ. This raises the question concerning the convergence of clusters associated with f_τ to that of f. Of course, to get the clusters associated with f we would prefer to replace τ by τ_n so that the sample τ-approximation, $f_{\tau_n,n}$, converges to f (cf. Proposition 2.5.1). Hence, to use the gradient system above, it is natural to replace the derivatives by their finite difference approximations. In Sections 3.5 and 3.6, we execute this strategy wherein we establish the convergence of population clusters and empirical clusters. First, we show in Section 3.4 that the stationary points and modes of f_τ converge to those of f. Next, we replace f by f_τ in (3.1.1) and consider the gradient system

$$u'(t) = \nabla f_\tau(u(t)). \tag{3.1.2}$$

We show that the solution $u_{x,\tau}$ of (3.1.2) with initial value $u_{x,\tau}(0) = x$ converges to the solution u_x of (3.1.1). We exploit this convergence in Section 3.5 to obtain convergence of the clusters of f_τ to those of f. The convergence of empirical clusters requires the uniform convergence of empirical finite difference approximations to the appropriate derivatives which is established using the Bernstein-type inequality described in Theorem 2.4.3. Convergence of empirical finite difference approximations is proved in Section 3.7, which also contains other preliminary results such as a discrete Grönwall lemma. The proof of convergence of empirical clusters itself is contained in Section 3.8. To the best of our knowledge, these results seem to be the first to provide strong theoretical guarantees for clustering in multidimensional problems. We note here that since the clustering described above is based on mode(s) of the density and upper level sets, the sample τ-approximation can also be used for mode estimation and upper level set estimation. While Section 3.6 includes mode estimation, upper level set estimation is studied in Section 3.2. We provide a detailed description of the clustering algorithm and its numerical implementation in Section 3.9. Section 3.10 contains some examples illustrating the clustering algorithm and the role of the localizing parameter τ. Specifically, we explain in detail how the population and estimated clusters are computed and compare the estimated clusters with those obtain via kernel density estimators (cf. Chacón and Duong (2018)). A through analysis of the performance of the clustering algorithm is provided in Sections 3.11 and 3.12.

3.2 Density upper level set estimation

In this section, we provide an application of LDFs to estimate the upper level sets. We begin with the definition of level sets and upper level sets.

Definition 3.2.1 *For $\alpha > 0$, the level sets of f and f_τ are $\mathrm{L}_\alpha = \{x \in \mathbb{R}^d : f(x) = \alpha\}$ and $\mathrm{L}_{\alpha,\tau} = \{x \in \mathbb{R}^d : f_\tau(x) = \alpha\}$, respectively. The upper level sets of f, f_τ and $f_{\tau,n}$ are $\mathrm{R}_\alpha = \{x \in \mathbb{R}^d : f(x) \geq \alpha\}$, $\mathrm{R}_{\alpha,\tau} = \{x \in \mathbb{R}^d : f_\tau(x) \geq \alpha\}$ and $\mathrm{R}_{\alpha,\tau,n} = \{x \in \mathbb{R}^d : f_{\tau,n}(x) \geq \alpha\}$, respectively.*

The next proposition shows that in the limit the upper level sets induced by f_τ and $f_{\tau,n}$ coincide with those induced by f.

Proposition 3.2.1 *Suppose that f is uniformly continuous. Let $\{\alpha_n\}_{n=1}^{\infty}$ and $\{\tau_n\}_{n=1}^{\infty}$ be sequences of positive scalars converging to $\alpha > 0$ and 0, respectively. It holds that*

$$\mathring{\mathrm{R}}_\alpha \subset \liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} \subset \limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} \subset \mathrm{R}_\alpha, \tag{3.2.1}$$

and, if $\lambda(\mathrm{L}_\alpha) = 0$, then

$$\lim_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} = \mathrm{R}_\alpha \ \lambda\text{-a.e.}^{1} \tag{3.2.2}$$

Suppose additionally that $\mathcal{H}_G$ is a VC subgraph class of functions and $\lim_{n\to\infty} \frac{n}{\log(n)} \tau_n^{2kd} = \infty$. It holds that

$$\mathring{\mathrm{R}}_\alpha \subset \liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} \subset \limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} \subset \mathrm{R}_\alpha \ a.s., \tag{3.2.3}$$

and, if $\lambda(\mathrm{L}_\alpha) = 0$, then

$$\lim_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} = \mathrm{R}_\alpha \ a.s.^{2} \tag{3.2.4}$$

Proof of Proposition 3.2.1. Using $\lim_{l\to\infty} \alpha_l = \alpha$ and Proposition 2.3.2 (i), we have that, for all $m \in \mathbb{N}$, there exists a constant $n \in \mathbb{N}$ such that $|\alpha_l - \alpha| < \frac{1}{m}$, for all $l \geq n$, and $|f_{\tau_l}(x) - f(x)| < \frac{1}{m}$, for all $l \geq n$ and $x \in \mathbb{R}^d$. It follows that

$$\begin{aligned}\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} &= \cup_{n=1}^{\infty} \cap_{l=n}^{\infty} \{x \in \mathbb{R}^d : f_{\tau_l}(x) \geq \alpha_l\} \\ &\supset \{x \in \mathbb{R}^d : f(x) > \alpha + \frac{2}{m}\} = \mathring{\mathrm{R}}_{\alpha+\frac{2}{m}} \uparrow_{m\to\infty} \cup_{m=1}^{\infty} \mathring{\mathrm{R}}_{\alpha+\frac{2}{m}} = \mathring{\mathrm{R}}_\alpha\end{aligned}$$

and

$$\begin{aligned}\limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} &= \cap_{n=1}^{\infty} \cup_{l=n}^{\infty} \{x \in \mathbb{R}^d : f_{\tau_l}(x) \geq \alpha_l\} \\ &\subset \{x \in \mathbb{R}^d : f(x) \geq \alpha - \frac{2}{m}\} = \mathrm{R}_{\alpha-\frac{2}{m}} \downarrow_{m\to\infty} \cap_{m=1}^{\infty} \mathrm{R}_{\alpha-\frac{2}{m}} = \mathrm{R}_\alpha,\end{aligned}$$

establishing (3.2.1). For the second part, using $\mathrm{R}_\alpha = \mathrm{L}_\alpha \cup \mathring{\mathrm{R}}_\alpha$ and (3.2.1), it follows that

$$\begin{aligned}\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} &= \mathring{\mathrm{R}}_\alpha \cup (\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} \cap \mathrm{L}_\alpha) \text{ and} \\ \limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} &= \mathring{\mathrm{R}}_\alpha \cup (\limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} \cap \mathrm{L}_\alpha), \text{ where}\end{aligned}$$

$$\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} \cap \mathrm{L}_\alpha \subset \limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} \cap \mathrm{L}_\alpha \subset \mathrm{L}_\alpha$$

are sets of Lebesgue measure 0. Therefore,

$$\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} = \limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n} = \mathrm{R}_\alpha$$

except for a set of Lebesgue measure 0 and we obtain (3.2.2). We now prove (3.2.3). Let $A_{n,m} = \{x \in \mathbb{R}^d : |f_{\tau_n,n}(x) - f(x)| < \frac{1}{m}\}$. We first show that $\lim_{n\to\infty} A_{n,m} = \mathbb{R}^d$ a.s. To this end, we use Proposition 2.5.1 (i) and notice that, almost surely, there exists

[1] That is, the sets $\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n}$, $\limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n}$, and R_α differ at most by a set of Lebesgue measure zero.

[2] That is, $\mathbb{P}$-a.s. the sets $\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n}$, $\limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n}$, and R_α differ at most by a set of probability zero.

$\tilde{n}(m) \in \mathbb{N}$ (in general, different for different samples) such that, for all $n \geq \tilde{n}(m)$, $\sup_{x\in\mathbb{R}^d}|f_{\tau_n,n}(x) - f(x)| < \frac{1}{m}$. It follows that

$$\liminf_{n\to\infty} A_{n,m} = \lim_{n\to\infty} \cap_{l=n}^{\infty} A_{l,m} \supset \cap_{l=\tilde{n}(m)}^{\infty} A_{l,m} = \mathbb{R}^d \text{ a.s.}$$

Next, using Corollary B.1 (v), we have that, for all $m \in \mathbb{N}$,

$$\begin{aligned} \liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} &\supset \liminf_{n\to\infty} \{x \in \mathbb{R}^d : f_{\tau_n,n}(x) > \alpha + \frac{1}{m}, |f_{\tau_n,n}(x) - f(x)| < \frac{1}{m}\} \\ &\supset \lim_{n\to\infty} (\mathring{\mathrm{R}}_{\alpha+\frac{2}{m}} \cap A_{n,m}) = \mathring{\mathrm{R}}_{\alpha+\frac{2}{m}} \text{ a.s.,} \end{aligned} \tag{3.2.5}$$

and

$$\begin{aligned} \limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} &\subset \limsup_{n\to\infty} \{x \in \mathbb{R}^d : f_{\tau_n,n}(x) \geq \alpha - \frac{1}{m}, |f_{\tau_n,n}(x) - f(x)| < \frac{1}{m}\} \\ &\subset \lim_{n\to\infty} (\mathrm{R}_{\alpha-\frac{2}{m}} \cap A_{n,m}) = \mathrm{R}_{\alpha-\frac{2}{m}} \text{ a.s.} \end{aligned} \tag{3.2.6}$$

Using (3.2.5) and (3.2.6), we conclude that

$$\mathring{\mathrm{R}}_{\alpha} = \cup_{m=1}^{\infty} \mathring{\mathrm{R}}_{\alpha+\frac{2}{m}} \subset \liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} \subset \limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} \subset \cap_{m=1}^{\infty} \mathrm{R}_{\alpha-\frac{2}{m}} = \mathrm{R}_{\alpha} \text{ a.s.}$$

Finally, notice that, since $P \in \mathcal{P}_{d,\ll\lambda}$, $\lambda(\mathrm{L}_\alpha) = 0$ implies that

$$\mathbb{P}(\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} \cap \mathrm{L}_\alpha) \leq \mathbb{P}(\limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} \cap \mathrm{L}_\alpha) \leq P(\mathrm{L}_\alpha) = 0.$$

Thus,

$$\liminf_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} = \limsup_{n\to\infty} \mathrm{R}_{\alpha_n,\tau_n,n} = \mathrm{R}_\alpha \text{ a.s.,}$$

except for a set of probability 0 and (3.2.4) holds. ■

Remark 3.2.1 *If we restrict R_α, $\mathrm{R}_{\alpha,\tau}$ and $\mathrm{R}_{\alpha,\tau,n}$ to a compact subset of $\mathbb{R}^d$, then, using Propositions 2.3.2-2.5.1 (ii), we see that Proposition 3.2.1 also holds for continuous f.*

Although we will follow a different approach we notice that one can also define clusters as the connected components of the upper level sets R_α for some $\alpha > 0$ (see Menardi (2016)). Once the connected components are computed, the remaining points may be allocated to one of the clusters by using supervised classification techniques. A common approach is then to study how the clusters change as the parameter α varies, yielding cluster trees.

3.3 Mathematical background on clustering identification

In this section, we give a precise definition of modes and clusters and verify that the latter are non-trivial and disjoint. To this end, we make the following assumption on the density f. We denote by H_g the Hessian matrix associated with a function g and by $\langle \cdot, \cdot \rangle$ the inner product on $\mathbb{R}^d$.

Assumption 3.3.1 *f is a probability density function on $\mathbb{R}^d$ that is twice continuously differentiable with a finite number of stationary points in S_f. Additionally, the Hessian matrix H_f has non-zero eigenvalues at its stationary points. Also, let $\mathrm{R}_\alpha = \{x \in \mathbb{R}^d : f(x) \geq \alpha\}$ be a bounded set for every $\alpha > 0$.*

By continuity of f, R_α is compact. We notice that R_α is bounded if f vanishes at infinity, that is, $\lim_{t\to\infty}\sup_{x\in\mathbb{R}^d:\|x\|_2\geq t} f(x) = 0$, which is satisfied, for example, if S_f is bounded.

Recall that the clusters are defined as the stable manifolds generated by the mode and are obtained using the limiting trajectory of the gradient system. Specifically, for any $\mu \in S_f$, the stable manifold generated by μ is given by

$$C(\mu) = \{x \in S_f : \lim_{t\to\infty} u_x(t) = \mu\}, \tag{3.3.1}$$

where $u_x(t)$ is the solution at time t of the gradient system

$$u'(t) = \nabla f(u(t)) \tag{3.3.2}$$

with initial value $u(0) = x$. For any choice of μ, it is not required for the stable manifold so-defined to be non-trivial; i.e. the Lebesgue measure of $C(\mu)$ can be zero. However, we will see below that if μ is chosen as a mode of f then the resulting manifold has a positive Lebesgue measure. We next turn to define the stationary points type, and, in particular, the mode.

Definition 3.3.1 *A stationary point $\mu \in S_f$ of f is said to be of type l, $0 \leq l \leq d$, if $\mathrm{H}_f(\mu)$ has l negative and $d-l$ positive eigenvalues. In particular, $m \in S_f$ is said to be a mode (resp. an antimode) for f if it is a stationary point of f and $\mathrm{H}_f(m)$ has only negative (resp. positive) eigenvalues, that is, m is a local maximum (resp. minimum) for f. If $m_1,\ldots,m_M$ are the modes of f, then the clusters induced by $m_1,\ldots,m_M$ are the stable manifolds $C(m_1),\ldots,C(m_M)$.*

Since the clusters are obtained as limits of trajectories induced by modes, we now summarize relevant properties of the gradient system (3.3.2) by using results from the theory of ordinary differential equations and dynamical systems (Agarwal and Lakshmikantham, 1993; Hale, 1980; Teschl, 2012; Perko, 2013). We first note that u_x exists and is unique for t in some maximal time interval (a,b) with $a < 0 < b$, where $a = -\infty$ or $b = \infty$ is allowed. To see this, observe that, as f is twice continuously differentiable, for every $x \in \mathbb{R}^d$ there exists a convex neighborhood $\mathrm{U}(x)$ of x in which the second order partial derivatives are bounded. By applying Lemma 3.2.1 in Agarwal and Lakshmikantham (1993) to ∇f, it follows that ∇f is Lipschitz in $\mathrm{U}(x)$, and therefore ∇f is locally Lipschitz in S_f. Now, applying Picard-Lindelöf Theorem (see Theorems 2.2 and 2.5 in Teschl (2012)), it follows that u_x exists in some time interval, which can be chosen to be maximal in view of Theorem 2.13 in Teschl (2012).

We now show that, using the boundedness of R_α, the solution $u_x(t)$ exists for all $t \geq 0$ and all $x \in S_f$. Furthermore, the solution starting in R_α cannot leave the set. To this end, notice that the equilibria of (3.3.2) are the stationary points of f. Furthermore, the gradient computed at each point gives the direction of most rapid increase of f. Hence, the trajectories $\{u_x(t) : t \in \mathbb{R}\}$ for $x \in S_f$ that are not stationary points are curves of steepest ascent for f. More specifically, if $u_x(t) \in \mathrm{L}_\alpha$ for some $x \in S_f$ and $t \in \mathbb{R}$, then any vector v tangent to L_α at $u_x(t)$ satisfies $\langle v, u_x'(t)\rangle = 0$ (see Theorem 9.4.2 in Hirsch et al. (1974) and Lemma 6.4.2. in Jost (2005)). Hence, either $u_x(t) = x$ for all t is an equilibrium of the gradient system (3.3.2) or the trajectory $\{u_x(t) : t \in \mathbb{R}\}$ crosses L_α orthogonally. This also implies that $u_x(t)$ cannot leave R_α for $t \geq 0$. Furthermore, this property shows that, for all $x \in S_f$, the solutions $u_x(t)$ exists for all $t \geq 0$, i.e. the maximal time interval in which u_x is defined is (a,∞) for some $a < 0$, where $a = -\infty$ is possible. To see this, for $x \in S_f$, choose

$\alpha > 0$ such that $x \in \mathrm{R}_\alpha$. Since $u_x(t)$ cannot leave R_α for $t \geq 0$ and R_α is compact by Assumption 3.3.1, the result follows from Corollary 2.15 of Teschl (2012). Recalling that our clusters are the stable manifolds generated by modes, we now link modes to the gradient system. This requires the notion of ω-limit which we now define.

Definition 3.3.2 *The ω-limit of a point $x \in S_f$ is the set of points $y \in S_f$ such that $u_x(t)$ goes to y as $t \to \infty$, in symbols*

$$\omega(x) = \{y \in S_f : \lim_{t\to\infty} u_x(t) = y\}.$$

We use the following definition of Theorem 9.3.1 in Hirsch et al. (1974) and Section 6.6 of Teschl (2012). For any function $\mathrm{W} : \mathrm{U}(\mu) \to \mathbb{R}$, we use the notation $\mathrm{W}'(u_x(t)) = \frac{d}{dt}\mathrm{W}(u_x(t))$.

Definition 3.3.3 *Let $\mu \in S_f$ be an equilibrium point for* (3.3.2) *and $\mathrm{U}(\mu) \subset S_f$ a neighborhood of μ. A differentiable function $\mathrm{W} : \mathrm{U}(\mu) \to \mathbb{R}$ is a strict Lyapunov function if (i) $\mathrm{W}(\mu) = 0$ and $\mathrm{W}(u) > 0$ for $u \neq \mu$, and (ii) $\mathrm{W}'(u_x(t)) < 0$ when $u_x(t) \in \mathrm{U}(\mu) \setminus \{\mu\}$.*

Let $\mathrm{V} = -f$. If m is a mode for f, there exists a neighborhood $\mathrm{U}(m)$ of m such that, for all $u \in \mathrm{U}(m) \setminus \{m\}$, $\mathrm{V}(u) - \mathrm{V}(m) > 0$ and

$$(\mathrm{V}(u) - \mathrm{V}(m))' = -(f(u))' = -\langle \nabla f(u), u' \rangle = -\|\nabla f(u)\|_2^2 < 0.$$

Hence, $\mathrm{V}(\cdot) - \mathrm{V}(m)$, restricted to $\mathrm{U}(m)$, is a strict Lyapunov function. By the Lyapunov stability Theorem (see Theorem 9.3.1 in Hirsch et al. (1974) and Theorem X.1.1 in Hale (1980)) m is asymptotically stable, that is, there is a neighborhood $\tilde{\mathrm{U}}(m) \subset \mathrm{U}(m)$ of m such that each solution starting from a point $x \in \tilde{\mathrm{U}}(m)$ converges to m, i.e., for all $x \in \tilde{\mathrm{U}}(m)$, $\omega(x) = \{m\}$. As we will see below, the set of points that have m as ω-limit (that is, the stable manifold generated by m) is typically much larger than $\tilde{\mathrm{U}}(m)$. For instance, if $0 < \alpha < f(m)$ is such that the connected component of m in R_α contains no equilibria other than m, then, since each solutions of (3.3.2) starting in that component cannot leave it, by LaSalle's invariance principle (see Theorem 9.3.2 in Hirsch et al. (1974) and Theorem 6.14 in Teschl (2012)) applied to the strict Lyapunov function $\mathrm{V}(\cdot) - \mathrm{V}(m)$, all the points in that component have m as an ω-limit point. On the other hand, if m is an antimode for f, then there exists a neighborhood $\mathrm{U}(m)$ of m such that for all $u \in \mathrm{U}(m) \setminus \{m\}$, $\mathrm{V}(m) - \mathrm{V}(u) > 0$ and $(\mathrm{V}(m) - \mathrm{V}(u))' > 0$. This implies that m is unstable (see Theorem X.1.2 in Hale (1980)): for every neighborhood $\tilde{\mathrm{U}}(m) \subset \mathrm{U}(m)$ of m, every solution u_x starting from a point $x \in \tilde{\mathrm{U}}(m)$ eventually leaves $\tilde{\mathrm{U}}(m)$. Furthermore, any ω-limit point of gradient system (3.3.2) is an equilibrium point: that is, a stationary points of f (see Theorem 9.4.4 in Hirsch et al. (1974) and Theorem X.1.3 in Hale (1980), and Lemma 6.4.4 in Jost (2005) in a different context).

For a stationary point μ of f, recall from (3.3.1) that $C(\mu)$ is the stable manifold induced by μ, that is, the set of points with ω-limit μ. The hypothesis that H_f has nonzero eigenvalues at stationary points and Stable Manifold Theorem (see Section 2.7 in Perko (2013), Section 9 in Teschl (2012), and Theorem A, Remark 2.3 in Abbondandolo and Pietro (2006)) indeed imply that the sets $C(\mu)$ are immersed submanifolds of $\mathbb{R}^d$ with (topological) dimension equal to the number of negative eigenvalues of $\mathrm{H}_f(\mu)$. As in Definition 3.3.1, let $m_1, \ldots, m_M$ be the modes and $\mu_1, \ldots, \mu_L$ the other stationary points of f. We are now ready to verify that the clusters $C(m_1), \ldots, C(m_M)$ are non-trivial. We first observe that, by the uniqueness of the limit,

$$C(m_1), \ldots, C(m_M), C(\mu_1), \ldots, C(\mu_L)$$

are disjoint and, since any ω-limit point of gradient system (3.3.2) is an equilibrium point,

$$S_f = \cup_{i=1}^{M} C(m_i) \cup \cup_{l=1}^{L} C(\mu_l). \tag{3.3.3}$$

Hence, $C(m_1), \ldots, C(m_M), C(\mu_1), \ldots, C(\mu_L)$ form a partition of S_f. Also, the set $S_f \setminus (\cup_{i=1}^{M} C(m_i)) = \cup_{l=1}^{L} C(\mu_l)$ has (topological) dimension smaller than d. The next proposition provides a characterization of the boundaries of $C(m_1), \ldots, C(m_M)$. In particular, it shows that the clusters $C(m_1), \ldots, C(m_M)$ are divided in S_f by the lower dimensional stable manifolds $C(\mu_1), \ldots, C(\mu_L)$. We denote by $\operatorname{dist}(A, B) = \inf_{x \in A, y \in B} \|x - y\|_2$ the distance between two sets $A, B \in \mathbb{R}^d$.

Proposition 3.3.1 *Suppose that Assumption 3.3.1 holds true. Then, for all $i = 1, \ldots, M$, $C(m_i)$ is open and $\partial C(m_i) \subset \partial S_f \cup \cup_{l=1}^{L} C(\mu_l)$.*

Proof of Proposition 3.3.1. Suppose by contradiction that $C(m_i)$ is not open. Then, there exists $x \in C(m_i)$ such that $\overline{B}_\epsilon(x) \cap (\mathbb{R}^d \setminus C(m_i)) \neq \emptyset$ for all $\epsilon > 0$. Using (3.3.3), we obtain that

$$(\mathbb{R}^d \setminus C(m_i)) = (\mathbb{R}^d \setminus S_f) \cup \cup_{\substack{j=1 \\ j \neq i}}^{M} C(m_j) \cup \cup_{l=1}^{L} C(\mu_l).$$

Since $x \in S_f$ and S_f is open, there exists $\tilde{\epsilon} > 0$ such that $\overline{B}_{\tilde{\epsilon}}(x) \cap (\mathbb{R}^d \setminus S_f) = \emptyset$. Hence, for all $0 < \epsilon \leq \tilde{\epsilon}$, it holds that

$$\overline{B}_\epsilon(x) \cap (\cup_{\substack{j=1 \\ j \neq i}}^{M} C(m_j) \cup \cup_{l=1}^{L} C(\mu_l)) \neq \emptyset.$$

Therefore, there is a sequence $\{x_l\}_{l=1}^{\infty}$ in $(\cup_{\substack{j=1 \\ i \neq j}}^{M} C(m_j) \cup \cup_{l=1}^{L} C(\mu_l))$ with $\lim_{l \to \infty} x_l = x$ and $f(x_l) \geq \alpha$, where $\alpha = f(x)/2$. Notice that, by Assumption 3.3.1, f is twice continuously differentiable and R_α is compact. In particular, ∇f is locally Lipschitz. Denote by L the Lipschitz constant of ∇f on R_α and let $\delta = \operatorname{dist}(\{m_i\}, \mathbb{R}^d \setminus C(m_i))/3$. Since $C(m_i)$ contains an open neighborhood $\tilde{U}(m_i)$ of m_i,

$$\delta \geq \operatorname{dist}(\{m_i\}, \mathbb{R}^d \setminus \tilde{U}(m_i))/3 > 0.$$

Recall that the solution $u_x(t)$ of (3.3.2) exists for all $t \in (a, \infty)$, $a < 0$. Since $x \in C(m_i)$, there exists $\tilde{t} \geq 0$ such that, for all $t \geq \tilde{t}$, $\|m_i - u_x(t)\|_2 \leq \delta$. Next, we use the continuity of solutions of ordinary differential equations with respect to the initial value (see Theorem 2.8 and (2.43) in Teschl (2012)) and obtain that for all $t \geq 0$

$$\|u_x(t) - u_{x_l}(t)\|_2 \leq \|x_l - x\|_2 e^{Lt}.$$

Now, let $\tilde{l}$ such that $\|x_l - x\|_2 e^{L\tilde{t}} \leq \delta$ for all $l \geq \tilde{l}$. Then, by the triangle inequality, we have that $\|m_i - u_{x_{\tilde{l}}}(\tilde{t})\|_2 \leq 2\delta$. Hence, $u_{x_{\tilde{l}}}(\tilde{t}) \in C(m_i)$. By the flow property of autonomous ordinary differential equations (see (6.10) in Teschl (2012)), it holds that $u_{x_{\tilde{l}}}(t + \tilde{t}) = u_{u_{x_{\tilde{l}}}(\tilde{t})}(t)$, implying that

$$\lim_{t \to \infty} u_{x_{\tilde{l}}}(t + \tilde{t}) = \lim_{t \to \infty} u_{u_{x_{\tilde{l}}}(\tilde{t})}(t) = m_i.$$

It follows that $x_{\tilde{l}} \in C(m_i)$. A contradiction. Hence, $C(m_i)$ are open for all $i = 1, \ldots, M$. Now, using again (3.3.3) and $C(m_i) \subset S_f$, we have that

$$\partial C(m_i) \subset \partial S_f \cup \cup_{\substack{j=1 \\ j\neq i}}^{M} C(m_j) \cup \cup_{l=1}^{L} C(\mu_l).$$

Since $C(m_i)$ and $C(m_j)$ are open and disjoint for $j \neq i$, we obtain that

$$\partial C(m_i) \subset \overline{C(m_i)} \subset \mathbb{R}^d \setminus C(m_j)$$

yielding that $\partial C(m_i) \cap (\cup_{\substack{j=1 \\ j\neq i}}^{M} C(m_j)) = \emptyset$. ∎

Remark 3.3.1 *Since Hausdorff dimension is larger or equal to topological dimension (see Theorem 6.3.10 in Edgar (2007)), the stable manifold $C(\mu_l)$ does not necessarily have Lebesgue measure zero. However, $\lambda(C(\mu_l)) = 0$ whenever topological and Hausdorff dimension coincide; and if they differ, the latter is smaller than n. Osgood curves (Sagan, 1994, Chapter 8) are examples of one-dimensional embedded manifolds in $\mathbb{R}^2$ with positive Lebesgue measure. These examples also show that, in Proposition 3.2.1, the assumption that the level sets of f have zero Lebesgue measure is, in general, necessary.*

3.4 Identification of stationary points

The clusters and boundaries critically depend on the stationary points of f and their type. In this section, we characterize the stationary points of f_τ and show that they converge to stationary points of f. Notice that for small τ, the first and second order derivatives are close (Proposition 2.3.2). Hence, one can pick a hypercube, centered at the stationary point with directions provided by eigenvectors of Hessian matrix, so that f and f_τ share similar properties within the hypercube. This idea is made precise in the following theorem.

Theorem 3.4.1 *Suppose (2.2.1) holds true. The following hold:*
(i) If f is continuously differentiable in $\overline{B}_{\rho\tau}(\mu) \subset S_f$, $\tau > 0$, then $\nabla f_\tau(\mu) = 0$ if and only if

$$\int h_{0,\tau}(x_1, \ldots, x_k) \nabla f(\mu + x_1) f(\mu + x_2) \ldots f(\mu + x_k) dx_1 \ldots dx_k = 0, \tag{3.4.1}$$

where the integral of a vector is the vector of the integrals.
(ii) If f is twice continuously differentiable in $\overline{B}_\delta(\mu) \subset S_f$, $\delta > 0$, and μ is a stationary point of f of type l, then there exists $h^*, \tau^* > 0$ and a closed hypercube $F_{h^*}(\mu) \subset \overline{B}_\delta(\mu)$ with side length $3/2h^*$ such that, for $0 < \tau \leq \tau^*$, f_τ has a unique stationary point μ_τ in $\mathring{F}_{h^*}(\mu)$ and μ_τ is of type l. Moreover, $\lim_{\tau \to 0^+} \|\mu_\tau - \mu\|_2 = 0$.
(iii) If f is three times continuously differentiable, then $\|\mu_\tau - \mu\|_2 = O(\tau^2)$.

Before we prove Theorem 3.4.1, we introduce few additional notations. The L^q-norm of a $d \times d$ matrix M is given by $\|M\|_{\mathcal{M},q} = \sup_{y \in \mathbb{R}^d, y \neq 0} \|My\|_q / \|y\|_q$ and the spectrum of M, that is, the set of all the eigenvalues of M is denoted by $\sigma(M)$. Finally, the sign function $\mathrm{sgn} : \mathbb{R} \to \mathbb{R}$ is given by

$$\mathrm{sgn}(t) = \begin{cases} -1 & \text{if } t < 0 \\ 0 & \text{if } t = 0 \\ 1 & \text{if } t > 0. \end{cases}$$

Before the proof, we provide a brief description of the idea. Proof of (i) is standard and allows for a characterization of the stationary points of f_τ (see Theorem 3.4.2 below). As for the proof of part (ii), note that for each stationary point μ of f, first a closed hypercube centered at μ with directions given by the orthogonal eigenvectors of $\mathrm{H}_f(\mu)$ is constructed. The side lengths of the hypercube are such that, for small enough τ and all points in the hypercube (i) the eigenvalues of H_{f_τ} and H_f corresponding to the same eigenvector have the same sign and (ii) points on opposite "hyperfaces" have directional derivatives (w.r.t. the eigenvector that is orthogonal to the two "hyperfaces") of opposite sign. This follows using the convergence of first and second order derivatives of f_τ to those of f. Now, (ii) implies that every straight line connecting the two "hyperfaces" contains a point having zero directional derivative. Thus, by intersecting all such sets of points along every direction, we find a point μ_τ having zero directional derivative w.r.t. all eigenvectors. Since these are orthogonal, the gradient of μ_τ is zero, that is, μ_τ is a stationary point of f_τ. Next, using (i), we conclude that μ_τ and μ are of the same type. Finally, the convergence $\mu_\tau \to \mu$ follows by letting the side length of the hypercube converge to zero. For part (iii), we use Proposition 2.3.3 to show that, in a compact set, $|\nabla f_\tau(\cdot) - \nabla f(\cdot)| = o(\tau^2)$. We then infer the same order of convergence for μ_τ to μ.

Proof of Theorem 3.4.1. We start by proving (i). Notice that, if f is continuously differentiable in $\overline{B}_{\rho\tau}(x) \subset S_f$, then, for $j = 1, \ldots, d$,

$$\partial_j f_\tau(x) = \frac{1}{k}(f_\tau(x))^{1-k}\frac{\partial_j L_G(x,\tau)}{\tau^{kd}\Lambda_1}, \tag{3.4.2}$$

where, by Proposition 2.2.1, (2.2.4), and (2.2.6),

$$\begin{aligned}\partial_j L_G(x,\tau) &= \int h_{0,\tau}(x_1,\ldots,x_k)\partial_j(f(x-x_1)\ldots f(x-x_k))dx_1\ldots dx_k \\ &= k\int h_{0,\tau}(x_1,\ldots,x_k)\partial_j f(x+x_1)f(x+x_2)\ldots f(x+x_k)dx_1\ldots dx_k.\end{aligned}$$

Hence, $\partial_j f_\tau(\mu) = 0$ if and only if

$$\int h_{0,\tau}(x_1,\ldots,x_k)\partial_j f(\mu+x_1)f(\mu+x_2)\ldots f(\mu+x_k)dx_1\ldots dx_k = 0, \tag{3.4.3}$$

and hence (3.4.1) holds. We next turn to the proof of (ii). Since $\mathrm{H}_f(\mu)$ is symmetric, it has orthonormal eigenvectors v_i associated with eigenvalues λ_i, $i = 1, \ldots, d$. Notice that, since μ is of type l, l eigenvalues are negative and $d-l$ are positive. In particular,

$$\min_{i=1,\ldots,d}|\lambda_i| > 0. \tag{3.4.4}$$

Let $0 < \tau \le \tau_1$, where $\tau_1 = \delta/(2(1+\rho))$, and $x \in \overline{B}_{\tau_1}(\mu)$. Since $x \in \overline{B}_{\delta/2}(\mu)$, $(\overline{B}_{\tau_1}(x))^{+\rho\tau} \subset \overline{B}_{\delta/2}(x) \subset \overline{B}_\delta(\mu)$. It follows that f_τ is twice continuously differentiable in $\overline{B}_{\tau_1}(x)$ and its first order partial derivatives are given by (3.4.2). By uniform continuity of the second order partial derivatives of f in $\overline{B}_\delta(\mu)$ and Proposition 2.3.2 (iv), it follows that, for $i, j = 1, \ldots, d$,

$$\sup_{y\in\overline{B}_\delta(\mu)} |\partial_i\partial_j f(y) - \partial_i\partial_j f(\mu)| \xrightarrow[\delta\to 0^+]{} 0. \tag{3.4.5}$$

and, for $0 < \tilde{\tau}_1, \tilde{\tau}_2 \leq \tau_1$,

$$\begin{aligned}\sup_{0<\tau\leq\tilde{\tau}_2} \sup_{y\in\overline{B}_{\tilde{\tau}_1}(\mu)} |\partial_i\partial_j f_\tau(y) - \partial_i\partial_j f(y)| \leq &\sup_{0<\tau\leq\tilde{\tau}_2} \sup_{y\in\overline{B}_{\tau_1}(\mu)} |\partial_i\partial_j f_\tau(y) - \partial_i\partial_j f(y)| \\ &+ \sup_{y\in\overline{B}_{\tilde{\tau}_1}(\mu)} |\partial_i\partial_j f(y) - \partial_i\partial_j f(y)| \xrightarrow[\tilde{\tau}_1,\tilde{\tau}_2\to 0^+]{} 0.\end{aligned} \tag{3.4.6}$$

For $y_1, \ldots, y_d \in \overline{B}_\delta(0)$, let

$$\mathrm{H}_f(x; y_1, \ldots, y_d) = \begin{pmatrix} (\nabla\partial_1 f(x+y_1)))^\top \\ \vdots \\ (\nabla\partial_d f(x+y_d))^\top \end{pmatrix}^\top$$

and, for $y_1, \ldots, y_d \in \overline{B}_{\tau_1}(0)$,

$$\mathrm{H}_{f_\tau}(x; y_1, \ldots, y_d) = \begin{pmatrix} (\nabla\partial_1 f_\tau(x+y_1)))^\top \\ \vdots \\ (\nabla\partial_d f_\tau(x+y_d))^\top \end{pmatrix}^\top.$$

(3.4.5) and (3.4.6) show that,

$$\sup_{y_1,\ldots,y_d\in\overline{B}_\delta(0)} \|\mathrm{H}_f(\mu; y_1, \ldots, y_d) - \mathrm{H}_f(\mu)\|_{\mathcal{M},2} \xrightarrow[\delta\to 0^+]{} 0 \tag{3.4.7}$$

and

$$\sup_{0<\tau\leq\tilde{\tau}_2} \sup_{y_1,\ldots,y_d\in\overline{B}_{\tilde{\tau}_1}(0)} \|\mathrm{H}_{f_\tau}(\mu; y_1, \ldots, y_d) - \mathrm{H}_f(\mu)\|_{\mathcal{M},2} \xrightarrow[\tilde{\tau}_1,\tilde{\tau}_2\to 0^+]{} 0. \tag{3.4.8}$$

In particular, (3.4.7) implies that, for $i = 1, \ldots, d$,

$$\sup_{y_1,\ldots,y_d\in\overline{B}_\delta(0)} \|\mathrm{H}_f(\mu; y_1, \ldots, y_d)v_i - \lambda_i v_i\|_2 \xrightarrow[\delta\to 0^+]{} 0.$$

and, for $t_i \in \mathbb{R}$,

$$\sup_{y_1,\ldots,y_d\in\overline{B}_\delta(0)} \left| \langle \mathrm{H}_f(\mu; y_1, \ldots, y_d) \left(v_i + \sum_{j=1, j\neq i}^{d} t_j v_j \right), v_i \rangle - \lambda_i \right| \xrightarrow[\delta\to 0^+]{} 0.$$

By (3.4.4), there exists $0 < \delta_2 \leq \delta$ such that, for $i = 1, \ldots, d$,

$$\operatorname{sgn}\left(\langle \mathrm{H}_f(\mu; y_1, \ldots, y_d) \left(v_i + \sum_{j=1, j\neq i}^{d} t_j v_j \right), v_i \rangle \right) = \operatorname{sgn}(\lambda_i), \tag{3.4.9}$$

for all $y_1, \ldots, y_d \in \overline{B}_{\delta_2}(0)$. Similarly, by (3.4.8), we see that

$$\sup_{0<\tau\leq\tilde{\tau}_2} \sup_{y_1,\ldots,y_d\in\overline{B}_{\tilde{\tau}_1}(0)} \|\mathrm{H}_{f_\tau}(\mu; y_1, \ldots, y_d)v_i - \lambda_i v_i\|_2 \xrightarrow[\tilde{\tau}_1,\tilde{\tau}_2\to 0^+]{} 0,$$

which implies that

$$\sup_{0<\tau\leq\tilde{\tau}_2}\sup_{y_1,\ldots,y_d\in\overline{B}_{\tilde{\tau}_1}(0)}|\langle \mathrm{H}_{f_\tau}(\mu;y_1,\ldots,y_d)v_i,v_i\rangle-\lambda_i|\xrightarrow[\tilde{\tau}_1,\tilde{\tau}_2\to 0^+]{}0. \tag{3.4.10}$$

Moreover, by Bauer–Fike theorem (Theorem 2.1 in Eisenstat and Ipsen (1998)), for all $\tilde{\lambda}_\tau(\mu;y_1,\ldots,y_d)\in\sigma(\mathrm{H}_{f_\tau}(\mu;y_1,\ldots,y_d))$, we have that

$$\min_{i=1,\ldots,d}|\tilde{\lambda}_\tau(\mu;y_1,\ldots,y_d)-\lambda_i|\leq\|\mathrm{H}_{f_\tau}(\mu;y_1,\ldots,y_d)-\mathrm{H}_f(\mu)\|_{\mathcal{M},2}. \tag{3.4.11}$$

By (3.4.4), (3.4.10), (3.4.11) and (3.4.8), it follows that, there exists $0<\tau_2\leq\tau_1$ such that, for all $0<\tau\leq\tau_2$ and $y_1,\ldots,y_d\in\overline{B}_{\tau_2}(0)$,

$$\operatorname{sgn}\left(\langle \mathrm{H}_{f_\tau}(\mu;y_1,\ldots,y_d)v_i,v_i\rangle\right)=\operatorname{sgn}(\lambda_i) \tag{3.4.12}$$

and $\sigma(\mathrm{H}_{f_\tau}(\mu;y_1,\ldots,y_d))=\{\tilde{\lambda}_{\tau,1}(\mu;y_1,\ldots,y_d),\ldots,\tilde{\lambda}_{\tau,d}(\mu;y_1,\ldots,y_d)\}$ with

$$\operatorname{sgn}(\tilde{\lambda}_{\tau,i}(\mu;y_1,\ldots,y_d))=\operatorname{sgn}(\lambda_i). \tag{3.4.13}$$

Now, let $0<\tau\leq\tau_2$, $0<h\leq h^*$, where $h^*=\min(\delta_2,\tau_2)/(2\sqrt{d})$, and $t_i\in[-2h,2h]$. By the mean value theorem, there exist $0\leq c_{i,j}\leq 1$ such that

$$\nabla f(\mu\pm hv_i+\sum_{j=1,j\neq i}t_jv_j)=\mathrm{H}_f(\mu;y_1,\ldots,y_d)\left(\pm hv_i+\sum_{j=1,j\neq i}^{d}t_jv_j\right),$$

where $y_j=c_{i,j}\left(\pm hv_i+\sum_{j=1,j\neq i}^{d}t_jv_j\right)$, implying that

$$\frac{1}{h}\langle\nabla f(\mu\pm hv_i+\sum_{j=1,j\neq i}t_jv_j),v_i\rangle=\pm\langle \mathrm{H}_f(\mu;y_1,\ldots,y_d)\left(v_i\pm\sum_{j=1,j\neq i}^{d}(t_j/h)v_j\right),v_i\rangle.$$

Since $\|y_j\|_2\leq 2\sqrt{d}h^*\leq\delta_2$, by (3.4.9),

$$\operatorname{sgn}\left(\langle\nabla f(\mu\pm hv_i+\sum_{j=1,j\neq i}t_jv_j),v_i\rangle\right)=\operatorname{sgn}(\pm\lambda_i). \tag{3.4.14}$$

Now, let us define the hypercube $F_{h^*}(\mu)$ with center μ, directions v_j, and side length $3/2h^*$ by

$$F_{h^*}(\mu)=\left\{\mu+\sum_{j=1}^{d}t_jv_j:t_j\in[-3/4h^*,3/4h^*]\right\}$$

and its "hyperfaces" by

$$F^{\pm}_{h^*,i}(\mu)=\left\{\mu\pm 3/4h^*v_i+\sum_{j=1,j\neq i}^{d}t_jv_j:t_j\in[-3/4h^*,3/4h^*]\right\}.$$

Since, by (2.2.1), for $0<\tau\leq\tau^*$, where $\tau^*=\min(\tau_2,h^*/(4\rho))$,

$$\overline{S}_{h_{0,\tau}(\cdot,x_2,\ldots,x_k)}\subset\overline{B}_{\rho\tau}(0)\subset\left\{\sum_{j=1}^{d}s_jv_j:s_j\in[-h^*/4,h^*/4]\right\},$$

we have that, for $\mu_i^{\pm} \in F_{h^*,i}^{\pm}(\mu)$ and $x_1 \in \overline{S}_{h_{0,\tau}(\cdot,x_2,\ldots,x_k)}$,

$$\mu_i^{\pm} + x_1 \in \mu + \left\{ \pm h v_i + \sum_{j=1, j\neq i}^{d} s_j v_j : h \in [h^*/2, h^*], s_j \in [-h^*, h^*] \right\}.$$

Now, by (3.4.14),

$$\operatorname{sgn}\left(\langle \nabla f(\mu_i^{\pm} + x_1), v_i \rangle\right) = \operatorname{sgn}(\pm\lambda_i),$$

for all $x_1 \in \overline{S}_{h_{0,\tau}(\cdot,x_2,\ldots,x_k)}$ and $t_j \in [-3/4h^*, 3/4h^*]$. It follows from (3.4.2) that

$$\operatorname{sgn}\left(\langle \nabla f_\tau(\mu_i^{\pm}), v_i \rangle\right) = \operatorname{sgn}(\pm\lambda_i).$$

In particular, for all $\mu_i^+ \in F_{h^*,i}^+(\mu)$ and $\mu_i^- \in F_{h^*,i}^-(\mu)$,

$$\operatorname{sgn}(\langle \nabla f_\tau(\mu_i^+), v_i \rangle) = -\operatorname{sgn}(\langle \nabla f_\tau(\mu_i^-), v_i \rangle) \neq 0. \tag{3.4.15}$$

Notice that $\mu_i^+ \in F_{h^*,i}^+(\mu)$ if and only if $\mu_i^+ - 3/2h^* v_i \in F_{h^*,i}^-(\mu)$ and let $\alpha_i : F_{h^*,i}^+(\mu) \times [0,1] \to F_{h^*}(\mu)$ be given by

$$\alpha_i(y,t) = (1-t)y + t(y - 3/2h^* v_i) = y - 3/2h^* t v_i.$$

Using (3.4.15) and the continuity of ∇f_τ, for all $\mu_i^+ \in F_{h^*,i}^+(\mu)$, there exists $0 < t_1 < 1$ such that $\langle \nabla f_\tau(\alpha_i(\mu_i^+, t_1)), v_i \rangle = 0$. Next, we show that t_1 is unique. To this end, let $0 < t_2 < 1$ be such that $\langle \nabla f_\tau(\alpha_i(\mu_i^+, t_2)), v_i \rangle = 0$. By the mean value theorem, there exist $0 \le c_j \le 1$ such that

$$\nabla f_\tau(\alpha_i(\mu_i^+, t_2)) = \nabla f_\tau(\alpha_i(\mu_i^+, t_1)) + \mathrm{H}_{f_\tau}(\mu; y_1, \ldots, y_d)^\top (\alpha_i(\mu_i^+, t_2) - \alpha_i(\mu_i^+, t_1))),$$

where $y_j = (1-c_j)\alpha_i(\mu_i^+, t_2) + c_j \alpha_i(\mu_i^+, t_1) - \mu$, implying that

$$3/2h^*(t_2 - t_1)\langle \mathrm{H}_{f_\tau}(\mu; y_1, \ldots, y_d) v_i, v_i \rangle = 0.$$

Using (3.4.12) we obtain that $t_2 = t_1$. For $i = 1, \ldots, d$ let

$$Z_{\tau,i}(\mu) = \{\alpha_i(y,t) : \langle \nabla f_\tau(\alpha_i(y,t_1)), v_i \rangle = 0, y \in F_i^+(\mu), t \in [0,1]\}.$$

Notice that $Z_{\tau,i}(\mu)$ are closed subsets of the hypercube $F_{h^*}(\mu)$ of dimension $d-1$ and divide $F_{h^*}(\mu)$ into two parts with only the faces $F_{h^*,i}^+(\mu)$ and $F_{h^*,i}^-(\mu)$ entirely contained in the same part. It follows that $\cap_{i=1}^d Z_{\tau,i}(\mu) = \{\mu_\tau\}$, where μ_τ satisfies $\langle \nabla f_\tau(\mu_\tau), v_i \rangle = 0$ for all $i = 1, \ldots, d$ yielding that $\nabla f_\tau(\mu_\tau) = 0$. Finally, by (3.4.13) and $\|\mu_\tau - \mu\|_2 \le 3/4\sqrt{d}h^* \le \tau_2$, it follows that μ_τ is of type l. Also, by letting $\tau_2 \to 0^+$, we see that $\|\mu_\tau - \mu\|_2 \to 0$.
We now prove (iii). Since $\mathrm{H}_f(\mu)^{-1}$ is symmetric, it holds that

$$\xi = \|\mathrm{H}_f(\mu)^{-1}\|_{\mathcal{M},2} = \max_{i=1,\ldots,d} 1/|\lambda_i| > 0.$$

By (3.4.8) there exists $0 < \tau_3 \le \tau_2$ such that for all $0 < \tau \le \tau_2$ and $y_j \in \overline{B}_{\tau_3}(0)$

$$\|\mathrm{H}_{f_\tau}(\mu; y_1, \ldots, y_d) - \mathrm{H}_f(\mu)\|_{\mathcal{M},2} \le 1/(2\xi). \tag{3.4.16}$$

It follows from (3.4.16) and the triangle inequality that, for all $v \in \mathbb{R}^d$,

$$\begin{aligned}\|v\|_2 &\le 2\xi \left(\|\mathrm{H}_f(\mu)v\|_2 - 1/(2\xi)\|v\|_2\right)\\ &\le 2\xi \left(\|\mathrm{H}_f(\mu)v\|_2 - \|(\mathrm{H}_{f_\tau}(\mu;y_1,\dots,y_d) - \mathrm{H}_f(\mu))v\|_2\right)\\ &\le 2\xi\|\mathrm{H}_{f_\tau}(\mu;y_1,\dots,y_d)v\|_2.\end{aligned}$$

By setting $w = \mathrm{H}_{f_\tau}(\mu;y_1,\dots,y_d)v$ we see that $\|\mathrm{H}_{f_\tau}(\mu;y_1,\dots,y_d)^{-1}w\|_2 \le 2\xi\|w\|_2$ yielding that

$$\|\mathrm{H}_{f_\tau}(\mu;y_1,\dots,y_d)^{-1}\|_{\mathcal{M},2} \le 2\xi. \tag{3.4.17}$$

Moreover, by the mean value theorem, there exist $0 \le \tilde{c}_j \le 1$, $j = 1,\dots,d$, such that,

$$\nabla f_\tau(\mu) = \nabla f_\tau(\mu) - \nabla f_\tau(\mu_\tau) = \mathrm{H}_{f_\tau}(\mu;y_1,\dots,y_d)(\mu - \mu_\tau),$$

where $y_j = \tilde{c}_j\mu + (1-\tilde{c}_j)\mu_\tau - \mu = (1-\tilde{c}_j)(\mu - \mu_\tau)$. Since $\|y_j\|_2 \le \|\mu - \mu_\tau\|_2 \le \tau_2$, $\mathrm{H}_{f_\tau}(\mu;y_1,\dots,y_d)$ is invertible by (3.4.13). We now apply Proposition 2.3.3 with $K = \overline{B}_\delta(\mu)$ and get constants $\tilde{\tau}(K), \tilde{c}_2(K) > 0$ such that, for all $y \in K$ and $0 < \tau \le \min(\tau_2, \tilde{\tau}(K))$,

$$\|\nabla f_\tau(y) - \nabla f(y)\|_2 \le \tilde{c}_2(K)\tau^2. \tag{3.4.18}$$

Using (3.4.17) and (3.4.18), we conclude that, for all $0 < \tau \le \min(\tau_2, \tilde{\tau}(K))$,

$$\|\mu - \mu_\tau\|_2 \le \|\mathrm{H}_{f_\tau}(\mu;y_1,\dots,y_d)^{-1}\|_{\mathcal{M},2}\|\nabla f_\tau(\mu) - \nabla f(\mu)\|_2 \le 2\xi\tilde{c}_2(K)\tau^2.$$

■

In the remaining of this section, we further develop the above results by providing some conditions under which the stationary points (resp. modes, antimodes) of f are *exactly* the stationary points (resp. modes, antimodes) of f_τ for $\tau > 0$. The key criteria for the identification of the modes is the notion of symmetry proposed below.

Definition 3.4.1 *Given $\tau > 0$, a density function f is said to be τ-centrally symmetric about $\mu \in S_f$ if, for all $x \in \mathbb{R}^d$ with $\|x\|_2 \le \tau$, $f(\mu + x) = f(\mu - x)$.*

In particular, for $d = 1$, f is τ-centrally symmetric about $\mu \in \mathbb{R}$ if $f(\mu - x) = f(\mu + x)$ for all $x \in [0, \tau]$. If f has a continuous derivative, a direct computation using Corollary 2.2.1 shows that, for $G = L, S, B, K_\beta$, $f_\tau^{'}(\mu) = 0$. Indeed, by (2.2.25), we see that

$$f_\tau(x) = \frac{1}{\tau}\sqrt{L_G(x,\tau)}$$

where

$$L_G(x,\tau) = 2\int_{T_{++}^\tau} f(x + x_1)f(x - x_2)dx_1dx_2$$

and

$$T_{++}^\tau = \{(x_1, x_2) : x_1 \ge 0, x_2 \ge 0, x_1 + x_2 \le \tau\}.$$

If f has a continuous derivative, we obtain that

$$f_\tau^{'}(x) = \frac{1}{\tau\sqrt{L_G(x,\tau)}}\int_{T_{++}^\tau} f^{'}(x + x_1)f(x - x_2) + f(x + x_1)f^{'}(x - x_2)dx_1dx_2.$$

Therefore, the sign of $f_\tau(x)$ depends on the sign of $f^{'}$ in the interval $(x - \tau, x + \tau)$. In particular, if f is τ-centrally symmetric about μ, then $f^{'}(\mu - x) = -f^{'}(\mu + x)$, yielding that $f_\tau^{'}(\mu) = 0$.

Our next result uses the above idea and gives sufficient conditions for a stationary point μ and a mode m of f to be a stationary point and a mode of f_τ.

Theorem 3.4.2 *Suppose* (2.2.1) *holds true and let* $\tau > 0$. *Then the following hold:*
(i) If f has continuous first order partial derivatives in $\overline{B}_\tau(\mu) \subset S_f$ and f is τ-centrally symmetric about the stationary point μ, then μ is a stationary point for f_τ.
(ii) Suppose that f is τ-centrally symmetric about a mode (resp. an antimode) m and has continuous second order partial derivatives in $\overline{B}_\tau(m)$. If, for all $x_1, \dots, x_k \in \overline{B}_\tau(m)$, the matrix

$$\mathrm{J}_f(x_1, \dots, x_k) = \mathrm{H}_f(x_1) f(x_2) \dots f(x_k) + (k-1) \nabla f(x_1) \nabla f(x_2)^\top f(x_3) \dots f(x_k)$$

is negative (resp. positive) definite, then m is also a mode (resp. an antimode) for f_τ.

Notice that $\mathrm{J}_f(m, \dots, m) = \mathrm{H}_f(m) f^{k-1}(m)$ is negative (resp. positive) definite and therefore the last condition of Theorem 3.4.2 is satisfied by f, for τ small.

Proof of Theorem 3.4.2. For (i) notice that if f is τ-centrally symmetric about μ, then, for all $y \in \mathbb{R}^d$ with $\|y\|_2 \le \tau$, $f(\mu + y) = f(\mu - y)$ and $\partial_j f(\mu - y) = -\partial_j f(\mu + y)$. By the change of variable $-(x_1, \dots, x_k)$ for $(x_1, \dots, x_k)$ on the LHS of (3.4.1) and (2.2.6) it follows that, for all $1 \le j \le d$,

$$\begin{aligned}
&\int h_{0,\tau}(x_1, \dots, x_k) \nabla f(\mu + x_1) f(\mu + x_2) \dots f(\mu + x_k) dx_1 \dots dx_k \\
&= \int h_{0,\tau}(x_1, \dots, x_k) \nabla f(\mu - x_1) f(\mu - x_2) \dots f(\mu - x_k) dx_1 \dots dx_k \\
&= -\int h_{0,\tau}(x_1, \dots, x_k) \nabla f(\mu + x_1) f(\mu + x_2) \dots f(\mu + x_k) dx_1 \dots dx_k,
\end{aligned}$$

and therefore (3.4.1) and $\nabla f_\tau(\mu) = 0$.
We now prove (ii). Since f is τ-centrally symmetric about m, (i) yields that

$$\partial_j f_\tau(m) = 0 \text{ for } j = 1, \dots, d \tag{3.4.19}$$

and m is a stationary point for f_τ. Moreover, (3.4.19) implies that, for $i, j = 1, \dots, d$,

$$\begin{aligned}
\partial_i \partial_j f_\tau(m) &= \frac{1}{k}\left(\frac{1}{k} - 1\right) (f_\tau(m))^{1-2k} (\partial_i f_\tau^k(m))(\partial_j f_\tau^k(m)) + \frac{1}{k}(f_\tau(m))^{1-k}(\partial_i \partial_j f_\tau^k(m)) \\
&= \frac{1}{k}(f_\tau(m))^{1-k}(\partial_i \partial_j f_\tau^k(m)),
\end{aligned}$$

where, by Proposition 2.2.1, (2.2.4) and (2.2.6),

$$\begin{aligned}
\partial_i \partial_j f_\tau^k(m) = k \int \frac{h_{0,\tau}(x_1, \dots, x_k)}{\tau^{kd} \Lambda_1} \Bigg[&\partial_i \partial_j f(m + x_1) \prod_{l=2}^k f(m + x_l) \\
&+ (k-1) \partial_j f(m + x_1) \partial_i f(m + x_2) \prod_{l=3}^k f(m + x_l) \Bigg] dx_1 \dots dx_k.
\end{aligned}$$

Using that the integral of a matrix is the matrix of the integrals, we get that

$$\mathrm{H}_{f_\tau}(m) = \frac{1}{k}(f_\tau(m))^{k-1} \int \frac{h_{0,\tau}(x_1, \dots, x_k)}{\tau^{kd} \Lambda_1} \mathrm{J}_f(m + x_1, \dots, m + x_k) dx_1 \dots dx_k.$$

Since the Hessian is symmetric, there exists an orthogonal matrix $M \in \mathcal{U}$ such that

$$
\begin{aligned}
D &= M^\top \mathrm{H}_{f_\tau}(m) M \\
&= \frac{1}{k}(f_\tau(m))^{k-1} \int \frac{h_{0,\tau}(x_1,\ldots,x_k)}{\tau^{kd}\Lambda_1} M^\top \mathrm{J}_f(m+x_1,\ldots,m+x_k) M dx_1 \ldots dx_k
\end{aligned}
$$

is a diagonal matrix. Now, since $\mathrm{J}_f(m+x_1,\ldots,m+x_k)$ is negative (resp. positive) definite, for all $y \in \mathbb{R}^d \setminus \{0\}$, $y^\top \mathrm{J}_f(m+x_1,\ldots,m+x_k) y < 0$ (resp. > 0), and therefore the diagonal elements of $M^\top \mathrm{J}_f(m+x_1,\ldots,m+x_k)M$ are negative (resp. positive). It follows that the diagonal elements of D (that is, the eigenvalues of $\mathrm{H}_{f_\tau}(m)$) are negative (resp. positive) and m is a mode (resp. an antimode) for f_τ. ■

3.5 Convergence of the gradient system under extreme localization

In this section, we replace f by f_τ in (3.3.2) and consider the gradient system

$$u'(t) = \nabla f_\tau(u(t)). \tag{3.5.1}$$

Then, we study the relationship between the gradient systems (3.5.1) and (3.3.2) under extreme localization. To this aim, notice that the sets $\{S_{f_\tau}\}_{\tau>0}$ contain S_f by Proposition 2.3.1. Under Assumption 3.3.1, Proposition 2.3.2 (iv) shows that the gradient and the Hessian matrix of f_τ converge to those of f. Recall that, by Remark (2.3.1), if f is m-times continuously differentiable, then, f_τ is m-times continuously differentiable in S_{f_τ}. If it exists, we denote by $u_{x,\tau}(t)$ the solution of (3.5.1) with initial point $u_{x,\tau}(0) = x$. Since f_τ is continuous, for $\alpha > 0$, the sets $\mathrm{R}_{\alpha,\tau} = \{x \in \mathbb{R}^d : f_\tau(x) \geq \alpha\}$ are closed. The next lemma along with the boundedness of R_α for all $\alpha > 0$ shows that they are also bounded. As shown in Section 3.3 for the gradient system (3.3.2), we conclude that, for all $x \in S_f$, $u_{x,\tau}$ exists and is unique in a maximal time interval (a, ∞), for some $-\infty \leq a < 0$.

Lemma 3.5.1 *Under assumption* (2.2.1), $(\mathrm{R}_\alpha)^{-\rho\tau} \subset \mathrm{R}_{\alpha,\tau} \subset (\mathrm{R}_\alpha)^{+\rho\tau}$, *for all* $\tau > 0$ *and* $\alpha > 0$. *In particular, if* R_α *is bounded for* $\alpha > 0$, *then* $\mathrm{R}_{\alpha,\tau}$ *is also bounded for any* $\tau > 0$.

Proof of Lemma 3.5.1. Since $x \in (\mathrm{R}_\alpha)^{-\rho\tau}$ satisfies $\inf_{y \in \mathbb{R}^d \setminus \mathrm{R}_\alpha} \|x-y\|_2 > \rho\tau$, we have that $\overline{B}_{\rho\tau}(x) \subset \mathrm{R}_\alpha$. By (2.2.5) and (2.2.1), we also have that $S_{h_{x,\tau}} \subset (\overline{B}_{\rho\tau}(x))^k \subset (\mathrm{R}_\alpha)^k$. It follows that

$$f_\tau(x) = \left(\int \frac{h_{x,\tau}(x_1,\ldots,x_k)}{\tau^{kd}\Lambda_1} f(x_1)\ldots f(x_k) dx_1 \ldots dx_k \right)^{1/k} \geq \alpha, \tag{3.5.2}$$

and therefore $x \in \mathrm{R}_{\alpha,\tau}$. Next, let $x \in \mathrm{R}_{\alpha,\tau}$. Then, there exists $(x_1,\ldots,x_k) \in S_{h_{x,\tau}}$ such that $f(x_1)\ldots f(x_k) \geq \alpha^k$. In particular, since $S_{h_{x,\tau}} \subset (\overline{B}_{\rho\tau}(x))^k$, there exists a point $z \in \overline{B}_{\rho\tau}(x)$ with $f(z) \geq \alpha$. Hence, $z \in \mathrm{R}_\alpha$ and, since $\|x-z\|_2 \leq \rho\tau$, we obtain that $x \in (\mathrm{R}_\alpha)^{+\rho\tau}$. Finally, suppose that R_α is bounded for $\alpha > 0$. Then, there exists $r > 0$ such that $\mathrm{R}_\alpha \subset \overline{B}_r(0)$. It follows that, for $\tau > 0$, $x \in \mathrm{R}_{\alpha,\tau} \subset (\mathrm{R}_\alpha)^{+\rho\tau}$ satisfies $\|x\|_2 \leq \inf_{y \in \mathrm{R}_\alpha}(\|y\|_2 + \|y-x\|_2) \leq r + \inf_{y\in\mathrm{R}_\alpha}\|y-x\|_2 \leq r + \rho\tau$. Hence, $\mathrm{R}_{\alpha,\tau} \subset \overline{B}_{r+\rho\tau}(0)$ is bounded. ■

We have shown that, under Assumption 3.3.1, the gradient system (3.5.1) has a unique solution $u_{x,\tau}$. For a stationary point $\mu_\tau \in S_{f_\tau}$ of f_τ, the stable manifold generated by μ_τ is

$$C_\tau(\mu_\tau) = \{x \in S_{f_\tau} : \lim_{t\to\infty} u_{x,\tau}(t) = \mu_\tau\}.$$

We exploit the differentiability properties of f_τ to show that the solutions of the gradient system (3.5.1) converge for $\tau \to 0^+$ to those of the gradient system (3.3.2). Under Assumption 3.3.1, let $\mathrm{M}_f = \{m_1, \ldots, m_M\}$ and $\mathrm{N}_f = \{m_1, \ldots, m_M, \mu_1, \ldots, \mu_L\}$ denote the set of modes and stationary points of f, respectively.

Theorem 3.5.1 *Suppose that (2.2.1) holds true. (i) If f is continuously differentiable in $\mathbb{R}^d$ and, for all $\alpha > 0$, R_α is compact, then, for all $t \geq 0$ and $x \in S_f$,*

$$\lim_{\tau\to 0^+} \|u_{x,\tau}(t) - u_x(t)\|_2 = 0.$$

(ii) If Assumption 3.3.1 holds true, then, for all $m \in \mathrm{M}_f$ and $x \in C(m)$,

$$\lim_{\tau\to 0^+} \sup_{t\in[0,\infty)} \|u_{x,\tau}(t) - u_x(t)\|_2 = 0.$$

(iii) If Assumption 3.3.1 holds true, $\{\tau_j\}_{j=1}^\infty$ is a sequence of positive scalars converging to zero, and $m \in \mathrm{M}_f$, then there are $\tau^{**} > 0$ and $\{m_{\tau_j}\}_{j=1,\tau_j\leq\tau^{**}}^\infty$ such that m_{τ_j} is a mode of f_{τ_j} and

$$C(m) \subset \liminf_{j\to\infty} C_{\tau_j}(m_{\tau_j}) \subset \limsup_{j\to\infty} C_{\tau_j}(m_{\tau_j}) \subset C(m) \cup \cup_{\mu\in \mathrm{N}_f\setminus \mathrm{M}_f} C(\mu) \cup \partial S_f.$$

In particular, if $\lambda(C(\mu)) = 0$ for all $\mu \in \mathrm{N}_f \setminus \mathrm{M}_f$ and $\lambda(\partial S_f) = 0$, then $\lim_{j\to\infty} C_{\tau_j}(m_{\tau_j}) = C(m)$ λ-a.e.

The proof of Theorem 3.5.1 (i)-(ii) is based on Proposition 2.3.2 (iv) and Grönwall's inequality. We use (ii) to obtain convergence of the stable manifold $C_{\tau_j}(m_{\tau_j})$.

Proof of Theorem 3.5.1. We start by proving (i). Fix $x \in S_f$ and let $\alpha > 0$ be such that $x \in \mathrm{R}_\alpha$. Since $\mathrm{R}_\alpha \subset S_f$ is compact, $\mathbb{R}^d \setminus S_f$ is closed, and these two sets are disjoint, we have that $\mathrm{dist}(\mathrm{R}_\alpha, \mathbb{R}^d \setminus S_f) > 0$. Let $\delta = \mathrm{dist}(\mathrm{R}_\alpha, \mathbb{R}^d \setminus S_f)/(3\rho)$ and notice that

$$\mathrm{dist}((\mathrm{R}_\alpha)^{+\delta}, \mathbb{R}^d \setminus S_f) = \inf_{y\in\mathbb{R}^d\setminus S_f, z\in(\mathrm{R}_\alpha)^{+\delta}} \|y - z\|_2 = \inf_{y\in\mathbb{R}^d\setminus S_f} \inf_{z\in\mathbb{R}^d : \inf_{w\in\mathrm{R}_\alpha}\|w-z\|_2\leq\delta} \|y - z\|_2.$$

By the triangle inequality, we have that, for $w \in \mathrm{R}_\alpha$,

$$\|y - z\|_2 \geq \|y - w\|_2 - \|w - z\|_2 \geq \|y - w\|_2 - \delta.$$

It follows that

$$\mathrm{dist}((\mathrm{R}_\alpha)^{+\delta}, \mathbb{R}^d \setminus S_f) \geq \mathrm{dist}(\mathrm{R}_\alpha, \mathbb{R}^d \setminus S_f) - \delta = 2\delta > 0. \tag{3.5.3}$$

Let $0 < \tau \leq \delta/\rho$. Using Lemma 3.5.1 we see that $\mathrm{R}_{\alpha,\tau} \subset (\mathrm{R}_\alpha)^{+\rho\tau} \subset (\mathrm{R}_\alpha)^{+\delta} \subset S_f$. Also, for all $s \geq 0$, $u_x(s) \in \mathrm{R}_\alpha$ and $u_{x,\tau}(s) \in \mathrm{R}_{\alpha,\tau}$; as shown in Section 3.3, the solutions of the gradient system (3.3.2) cannot leave the regions R_α, and the same is true for the gradient system (3.5.1) and $\mathrm{R}_{\alpha,\tau}$. In particular, for all $s \geq 0$, $u_x(s), u_{x,\tau}(s) \in K$, where $K = (\mathrm{R}_\alpha)^{+\delta}$ is a compact subset of S_f.

Now, noticing that the integral of a vector is the vector of the integrals of its components, we obtain, for $t \geq 0$,

$$u_{x,\tau}(t) - u_x(t) = \int_0^t \nabla f_\tau(u_{x,\tau}(s)) - \nabla f(u_x(s)) ds.$$

Next, by adding and subtracting $\nabla f_\tau(u_x(s))$ inside the integral and taking the norm on both sides, we see that

$$\begin{aligned}\|u_{x,\tau}(t) - u_x(t)\|_2 \leq & \int_0^t \|\nabla f_\tau(u_{x,\tau}(s)) - \nabla f_\tau(u_x(s))\|_2 ds \\ & + \int_0^t \|\nabla f_\tau(u_x(s)) - \nabla f(u_x(s))\|_2 ds.\end{aligned}$$

Since ∇f_τ is locally Lipschitz in S_f, it is Lipschitz in the compact subset K; that is, there exists a constant $L_\tau < \infty$ such that for all $y, z \in K$

$$\|\nabla f_\tau(y) - \nabla f_\tau(z)\|_2 \leq L_\tau \|y - z\|_2. \tag{3.5.4}$$

It follows that

$$\|u_{x,\tau}(t) - u_x(t)\|_2 \leq L_\tau \int_0^t \|u_{x,\tau}(s) - u_x(s)\|_2 ds + \int_0^t \|\nabla f_\tau(u_x(s)) - \nabla f(u_x(s))\|_2 ds.$$

We now apply Grönwall's inequality (see Corollary 6.6 in Hale (1980)) with $a = 0$, $\beta(s) = L_\tau$, $0 \leq s \leq t$, $\alpha = \int_0^t \|\nabla f_\tau(u_x(s)) - \nabla f(u_x(s))\|_2 ds$ and $\varphi(t) = \|u_{x,\tau}(t) - u_x(t)\|_2$, and obtain that

$$\|u_{x,\tau}(t) - u_x(t)\|_2 \leq e^{L_\tau} \int_0^t \|\nabla f_\tau(u_x(s)) - \nabla f(u_x(s))\|_2 ds. \tag{3.5.5}$$

To prove (i) we need to show that this converges to 0 as $\tau \to 0^+$. To this end, since ∇f is also locally Lipschitz in S_f, there exists a constant $L < \infty$ such that, for all $y, z \in K$

$$\|\nabla f(y) - \nabla f(z)\|_2 \leq L \|y - z\|_2. \tag{3.5.6}$$

By (3.5.6) and Proposition 2.3.2 (iv), it follows that, for all $y, z \in K$ with $y \neq z$,

$$\lim_{\tau \to 0^+} \frac{\|\nabla f_\tau(y) - \nabla f_\tau(z)\|_2}{\|y - z\|_2} = \frac{\|\nabla f(y) - \nabla f(z)\|_2}{\|y - z\|_2} \leq L. \tag{3.5.7}$$

Hence, $\{L_\tau\}_{0 < \tau \leq \delta/\rho}$ in (3.5.4) can be chosen in such a way that $\lim_{\tau \to 0^+} L_\tau = L$. In particular, there exists $0 < \tau^* \leq \delta/\rho$, such that

$$L_\tau \leq L + 1, \tag{3.5.8}$$

for all $0 < \tau \leq \tau^*$. Therefore, from (3.5.5), it follows that

$$\lim_{\tau \to 0^+} \|u_{x,\tau}(t) - u_x(t)\|_2 = 0,$$

if we can show that

$$\lim_{\tau \to 0^+} \int_0^t \|\nabla f_\tau(u_x(s)) - \nabla f(u_x(s))\|_2 ds = 0. \tag{3.5.9}$$

To show this, we first enlarge the compact set K by δ in such a way that it is still contained in S_f by considering the set $(K)^{+\delta} \subset (\mathrm{R}_\alpha)^{+2\delta}$. As in (3.5.3), we see that

$$\mathrm{dist}((\mathrm{R}_\alpha)^{+2\delta}, \mathbb{R}^d \setminus S_f) \geq \mathrm{dist}(\mathrm{R}_\alpha, \mathbb{R}^d \setminus S_f) - 2\delta = \delta > 0,$$

and $(K)^{+\delta}$ is indeed a compact subset of S_f. Furthermore, for all $y \in K$, $\overline{B}_{\rho\tau}(y) \subset \overline{B}_\delta(y) \subset (K)^{+\delta} \subset S_f$, and, in particular, by (2.2.5) and (2.2.1), $\overline{S}_{h_{y,\tau}} \subset (\overline{B}_\delta(y))^k \subset (S_f)^k$. Now, by (3.4.2), we see that for $y \in K$, the j-th partial derivative of f_τ at y is given by

$$\partial_j f_\tau(y) = (f_\tau(y))^{1-k} \left(\int \frac{h_{y,\tau}(x_1, \ldots, x_k)}{\tau^{kd}\Lambda_1} (\partial_j f(x_1)) f(x_2) \ldots f(x_k)) dx_1 \ldots dx_k \right)$$

and

$$|\partial_j f_\tau(y)| \leq \left(\frac{\alpha_0}{\beta_0} \right)^{k-1} \alpha_{1,j} < \infty,$$

where $\alpha_0 = \max_{z \in (K)^{+\delta}} f(z)$, $\beta_0 = \min_{z \in (K)^{+\delta}} f(z)$, and $\alpha_{1,j} = \max_{z \in (K)^{+\delta}} \partial_j f(z)$ satisfy $0 < \alpha_0, \beta_0, \alpha_{1,j} < \infty$. It follows that for $y \in K$

$$\begin{aligned} \|\nabla f_\tau(y) - \nabla f(y)\|_2 &\leq \|\nabla f_\tau(y)\|_2 + \|\nabla f(y)\|_2 \\ &\leq (1 + (\alpha_0/\beta_0)^{k-1}) \|(\alpha_{1,1}, \ldots, \alpha_{1,d})^\top\|_2 < \infty. \end{aligned}$$

Therefore, for all $0 \leq s \leq t$, $\|\nabla f_\tau(u_x(s)) - \nabla f(u_x(s))\|_2$ is bounded and by Proposition 2.3.2 (iv), for all $0 \leq s \leq t$, $\lim_{\tau \to 0^+} \|\nabla f_\tau(u_x(s)) - \nabla f(u_x(s))\|_2 = 0$. Now, (3.5.9) follows using LDCT completing the proof of (i).

We now prove (ii). We first show that for all $\alpha > 0$ such that $\mathrm{N}_f \subset \mathring{\mathrm{R}}_\alpha$, there exists $\tau^{**} > 0$ such that, for $0 < \tau \leq \tau^{**}$, $\{\mu_\tau\}_{\mu \in \mathrm{N}_f}$ are the only stationary points of f_τ in R_α. To this end, let $0 < \delta < \mathrm{dist}(\mathrm{N}_f, \mathbb{R}^d \setminus \mathring{\mathrm{R}}_\alpha)$ and notice that, by Remark 2.3.1 and Proposition 2.3.1, f_τ is twice continuously differentiable in $\mathrm{R}_\alpha \subset S_f \subset S_{f_\tau}$. By Theorem 3.4.1 (ii), for all $\mu \in \mathrm{N}_f$, there exist $h^*, \tau^* > 0$ and a closed hypercube $F_{h^*}(\mu) \subset \overline{B}_\delta(\mu)$ with side length $3/2h^*$, such that, for $0 < \tau \leq \tau^*$, f_τ has a unique stationary point μ_τ in $\mathring{F}_{h^*}(\mu)$, μ_τ is of the same type as μ, and $\lim_{\tau \to 0^+} \|\mu_\tau - \mu\|_2 = 0$. Let $K = \mathrm{R}_\alpha \setminus \cup_{\mu \in \mathrm{N}_f} \mathring{F}_{h^*}(\mu)$ and $\eta = \min_{z \in K} \|\nabla f(z)\|_2 > 0$. By Proposition 2.3.2 (iv), there exists $0 < \tau^{**} \leq \tau^*$ such that $\|\nabla f_\tau(y) - \nabla f(y)\|_2 < \eta$ for all $y \in \mathrm{R}_\alpha$ and $0 < \tau \leq \tau^{**}$. Hence, for all $z \in K$,

$$\|\nabla f_\tau(z)\|_2 \geq \|\nabla f(z)\|_2 - \|\nabla f_\tau(z) - \nabla f(z)\|_2 > 0.$$

It follows that $\{\mu_\tau\}_{\mu \in \mathrm{N}_f}$ are the only stationary points of f_τ in R_α. Let $0 < \epsilon < \xi \leq \min_{\mu, \nu \in \mathrm{N}_f, \mu \neq \nu} \|\mu - \nu\|_2/3$. Using Lemma 3.5.1 and Theorem 3.4.1 (ii) with $\delta = \epsilon$, let $0 < \tilde{\tau}(\xi) \leq \tau^{**}$ and $\tilde{\alpha}(\xi) > 0$ such that, for all $0 < \tau \leq \tilde{\tau}(\xi)$, $m_\tau \in \overline{B}_\epsilon(m)$ is the only stationary point of f_τ in $\overline{B}_\xi(m)$ and

$$\begin{aligned} \overline{B}_{2\epsilon}(m) &\subset (\mathrm{R}_{\tilde{\alpha}(\xi)})^{-\rho\tau} \cap C(m) \subset \mathrm{R}_{\tilde{\alpha}(\xi),\tau} \cap C(m) \\ &\subset (\mathrm{R}_{\tilde{\alpha}(\xi)})^{+\rho\tau} \cap C(m) \subset \overline{B}_\xi(m) \subset \mathrm{R}_\alpha \cap C(m). \end{aligned}$$

Let $\tilde{t}(\epsilon) = \inf\{t \in [0, \infty) : u_x(t) \in \overline{B}_\epsilon(m)\}$ be the first time that u_x reaches the ball $\overline{B}_\epsilon(m)$. Using (i), there exists $0 < \hat{\tau} \leq \tilde{\tau}(\xi)$ such that

$$\|u_{x,\tau}(\tilde{t}(\epsilon)) - u_x(\tilde{t}(\epsilon))\|_2 \leq \epsilon,$$

for all $0 < \tau \le \hat{\tau}$. Therefore, $u_{x,\tau}(\tilde{t}(\epsilon)) \in \overline{B}_{2\epsilon}(m)$ for all $0 < \tau \le \hat{\tau}$. Since $u_{x,\tau}$ cannot leave the connected components of the level set $\mathrm{R}_{\tilde{\alpha}(\xi),\tau}$, it follows that

$$\{u_{x,\tau}(t) : t \in [\tilde{t}(\epsilon), \infty)\} \subset \mathrm{R}_{\tilde{\alpha}(\xi),\tau} \cap C(m) \subset \overline{B}_{\xi}(m) \tag{3.5.10}$$

and

$$\lim_{t\to\infty} u_{x,\tau}(t) = m_{\tau}. \tag{3.5.11}$$

Using (3.5.5), (3.5.8), and (3.5.9), we obtain that

$$\lim_{\tau\to 0^+} \sup_{t\in[0,\tilde{t}(\epsilon)]} \|u_{x,\tau}(t) - u_x(t)\|_2 \le e^{L+1} \lim_{\tau\to 0^+} \int_0^{\tilde{t}(\epsilon)} \|\nabla f_{\tau}(u_x(s)) - \nabla f(u_x(s))\|_2 ds = 0.$$

It follows from (3.5.10) that

$$\begin{aligned}\lim_{\tau\to 0^+} \sup_{t\in[0,\infty)} \|u_{x,\tau}(t) - u_x(t)\|_2 \le& \lim_{\tau\to 0^+} \sup_{t\in[0,\tilde{t}(\epsilon)]} \|u_{x,\tau}(t) - u_x(t)\|_2 \\ &+ \lim_{\tau\to 0^+} \sup_{t\in[\tilde{t}(\epsilon),\infty)} \|u_{x,\tau}(t) - u_x(t)\|_2 \le 2\xi.\end{aligned}$$

We now prove (iii). Notice that m_{τ_j} is a mode of f_{τ_j} for all $\tau_j \le \tau^{**} \le \tau^*$. (3.5.11) yields that

$$C(m) \subset \liminf_{j\to\infty} C_{\tau_j}(m_{\tau_j}).$$

Next, using the definition of $C_{\tau_j}(m_{\tau_j})$, Proposition 2.3.1, and (3.3.3), we obtain that

$$\limsup_{j\to\infty} C_{\tau_j}(m_{\tau_j}) \subset \lim_{j\to\infty} S_{f_{\tau_j}} \subset \overline{S}_f = \cup_{\nu\in \mathrm{N}_f} C(\nu) \cup \partial S_f.$$

Suppose by contradiction that $\limsup_{j\to\infty} C_{\tau_j}(m_{\tau_j}) \cap C(\nu) \neq \emptyset$ for some $\nu \in \mathrm{M}_f \setminus \{m\}$. Then, there exist $x \in C(\nu)$ and a subsequence $\{\tilde{\tau}_j\}_{j=1}^{\infty}$ of $\{\tau_j\}_{j=1}^{\infty}$ such that $\lim_{t\to\infty} u_{x,\tilde{\tau}_j}(t) = m_{\tilde{\tau}_j}$. In particular, $\lim_{j\to\infty} \lim_{t\to\infty} u_{x,\tilde{\tau}_j}(t) = m$. On the other hand, by (ii), $u_{x,\tilde{\tau}_j}$ converges uniformly on $[0,\infty)$ to u_x. Therefore,

$$\lim_{t\to\infty} \lim_{j\to\infty} u_{x,\tilde{\tau}_j}(t) = \lim_{t\to\infty} u_x(t) = \nu.$$

By Moore-Osgood theorem (see Theorem 7.11 in Rudin (1976)), it follows that

$$m = \lim_{j\to\infty} \lim_{t\to\infty} u_{x,\tilde{\tau}_j}(t) = \lim_{t\to\infty} \lim_{j\to\infty} u_{x,\tilde{\tau}_j}(t) = \nu.$$

We conclude that $\limsup_{j\to\infty} C_{\tau_j}(m_{\tau_j}) \cap C(\nu) = \emptyset$ and

$$\limsup_{j\to\infty} C_{\tau_j}(m_{\tau_j}) \subset C(m) \cup \cup_{\mu\in \mathrm{N}_f\setminus \mathrm{M}_f} C(\mu) \cup \partial S_f.$$

Finally, notice that, if $\lambda(C(\mu)) = 0$ for all $\mu \in \mathrm{N}_f \setminus \mathrm{M}_f$ and $\lambda(\partial S_f) = 0$, then

$$\liminf_{j\to\infty} C_{\tau_j}(m_{\tau_j}) = \limsup_{j\to\infty} C_{\tau_j}(m_{\tau_j}) = C(m)$$

up to a set of Lebesgue measure zero. ∎

As an application of Lemma B.4 in Appendix B, we also have that $\lim_{j\to\infty}(C_{\tau_j}(m_{\tau_j}))^{+\xi} = (C(m))^{+\xi}$ λ-a.e. for all $\xi > 0$ and $\lim_{j\to\infty} \overline{C_{\tau_j}(m_{\tau_j})} = \overline{C(m)}$ λ-a.e.

3.6 Algorithm and consistency of empirical clusters

In this section, we describe the algorithm for the numerical approximation of the clusters induced by the system (3.5.1) and establish its consistency. Proofs are deferred to Sections 3.7 and 3.8.

Since the sample τ-approximation is not, in general, differentiable in x, we use a finite difference approximation that converges to the directional derivative. For $h > 0$, the finite difference approximation of the directional derivative of $g : \mathbb{R}^d \to \mathbb{R}$, in the direction of $v \in S^{d-1}$, is given by

$$\nabla_v^h g(x) = \frac{g(x+hv) - g(x)}{h}.$$

Notice that, if g is differentiable at $x \in \mathbb{R}^d$, then

$$\lim_{h\to 0^+} \nabla_v^h g(x) = \nabla_v g(x),$$

where $\nabla_v g(\cdot) = \langle \nabla g(\cdot), v\rangle$ is the directional derivative of g.

Our first result shows that under the condition $\lim_{n\to\infty} nh_n^{2k}\tau_n^{2kd} = \infty$, the finite difference approximation to the directional derivative of $f_{\tau_n,n}$ converges uniformly on compact sets, in probability.

Theorem 3.6.1 *Suppose (2.2.1) holds true. Let K be a compact subset of S_f, $\{h_n\}_{n=1}^\infty$ and $\{\tau_n\}_{n=1}^\infty$ sequences of positive scalars converging to 0 and $\{v_n\}_{n=1}^\infty$ be a sequence in S^{d-1} converging to $v \in S^{d-1}$. (i) If f is continuously differentiable, then*

$$\lim_{n\to\infty} \sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n}(x) - \nabla_v f(x)| = 0.$$

(ii) If, additionally, $\mathcal{H}_G$ is a VC subgraph class of functions and $\lim_{n\to\infty} nh_n^{2k}\tau_n^{2kd} = \infty$, then, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n,n}(x) - \nabla_v f(x)| \geq \epsilon\right) = 0.$$

The first step towards identifying the modes is finding a local maximum of a function. To this end, we use the steepest ascent or gradient ascent idea; that is, starting from a point in the space, the next point is chosen in the direction given by the gradient of the function at that point. This procedure is repeated until convergence to a local maximum is achieved. When clustering using modes, this procedure is often combined with kernel density estimators to find the modes of the density underlying the given data points, and the clusters associated with them (Fukunaga and Hostetler, 1975; Menardi, 2016). In the following, we propose a similar technique (using instead sample τ-approximation), which does not require existence of gradients, and considers data as potential candidate points for the next move. This yields a computationally efficient procedure (see Theorem 3.6.2 below).

Turning to the consistency result, we need arguments that allows one to approximate uniformly the directional derivative of points over (i) a compact set, (ii) the step-size, and (iii) directions. The next lemma addresses this issue and critically uses the Bernstein-type inequality developed in Theorem 2.4.3. Part (iii) of the lemma below also provides a upper bound on the uniform approximation mentioned above.

Lemma 3.6.1 *Suppose (2.2.1) holds true. Let K be a compact subset of S_f and let $h^* > 0$ be such that $(K)^{+h^*} \subset S_f$. Also, let $\{\tau_n\}_{n=1}^{\infty}$ and $\{h_n\}_{n=1}^{\infty}$ be sequences of positive scalars converging to 0. Assume also that f is three times continuously differentiable. Then*
(i) the finite difference approximation of the directional derivative of f_τ converges uniformly to that of f. That is,

$$\lim_{n\to\infty} \sup_{h\in[h_n,h^*]} \sup_{v\in S^{d-1}} \sup_{x\in K} |\nabla_v^h f_{\tau_n}(x) - \nabla_v^h f(x)| = 0.$$

(ii) If, additionally, $\mathcal{H}_G$ is a VC subgraph class of functions and $\lim_{n\to\infty} nh_n^{2k}\tau_n^{2kd} = \infty$, then, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{h\in[h_n,h^*]} \sup_{v\in S^{d-1}} \sup_{x\in K} |\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f(x)| \geq \epsilon \right) = 0.$$

(iii) Let $\lim_{n\to\infty} \frac{n}{\log(n)} h_n^{2k}\tau_n^{2kd} = \infty$ and $\mathcal{H}_G$ be a VC subgraph class of functions. Then, for all $\epsilon > 0$, there are constants $0 < \tilde{c} < \infty$ and $\tilde{n} \in \mathbb{N}$ such that, for all $n \geq \tilde{n}$,

$$\mathbb{P}\left(\sup_{h\in[h_n,h^*]} \sup_{v\in S^{d-1}} \sup_{x\in K} |\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f(x)| \geq \epsilon \right) \leq \frac{\tilde{c}}{n^2}.$$

We are now ready to state the main result of this section, namely, consistency of the empirically chosen clusters.

Theorem 3.6.2 *Suppose that $\mathcal{H}_G$ is a VC subgraph class of functions, Assumption 3.3.1 and (2.2.1) hold true and f is three times continuously differentiable. Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a sample of i.i.d. random variables from P with density f and $\{h_n\}_{n=1}^{\infty}$ and $\{\tau_n\}_{n=1}^{\infty}$ be sequences of positive scalars converging to zero with $\lim_{n\to\infty} nh_n^{2k}\tau_n^{2kd} = \infty$. For $x \in S_f$ and $r > 0$, define*

$$\mathcal{X}_{n,r}(x) = \{X \in \mathcal{X}_n : h_n \leq \|X - x\|_2 \leq r\},$$

$Y_{n,r,0} = x$ and, recursively, if

$$\max_{X\in\mathcal{X}_{n,r}(Y_{n,r,j})\cup\{Y_{n,r,j}\}} f_{\tau_n,n}(X) - f_{\tau_n,n}(Y_{n,r,j}) > 0, \tag{3.6.1}$$

then

$$Y_{n,r,j+1} = \operatorname*{argmax}_{X\in\mathcal{X}_{n,r}(Y_{n,r,j})} \frac{f_{\tau_n,n}(X) - f_{\tau_n,n}(Y_{n,r,j})}{\|X - Y_{n,r,j}\|_2}; \tag{3.6.2}$$

else stop and let $j^* = j$. It holds that $j^* \leq n$. Furthermore, if $x \in C(m_i)$, and given $0 < \eta \leq 1$, $\alpha^* \leq \alpha < f(m_i)$, and $0 < r \leq r^*$, for some $\alpha^*, r^* > 0$, then there exist $n^* \in \mathbb{N}$ such that, with probability at least $1 - \eta$, $Y_{n,r,j^*} \in R_\alpha \cap C(m_i)$, for all $n \geq n^*$.

Using the above theorem, one can estimate the mode using the last iterate, namely, Y_{n,r,j^*}. The Corollary 3.6.1 below provides strong consistency of this estimate. The proof of Theorem 3.6.2 and Corollary 3.6.1 is given in Section 3.8, whereas Section 3.7 contains the proof of Theorem 3.6.1, Lemma 3.6.1, and other preliminary results. Turning to the proof of Theorem 3.6.2, it is divided into four distinct but connected steps. For the first step, let j^* be a non-negative integer and define $\{y_{r,j}\}$ recursively as follows: let $y_{r,0} = x$ and

$$y_{r,j+1} = y_{r,j} + h_j v_j, \quad 0 \leq j \leq j^* - 1,$$

where $0 < h_j \leq r$ for some small $r > 0$, and where v_j is "close" to the normalized gradient of f at $y_{r,j}$. We show that the sequence $\{y_{r,j}\}$ is close to the solution u_x of (3.3.2). This is achieved, using a version of the discrete Grönwall lemma (Lemma 3.7.1 in Section 3.7). Next, we show that $\{Y_{n,r,j}\}$ in (3.6.2) behaves like the sequence $\{y_{r,j}\}$ described in Step 1, with probability $1 - \eta$. This is achieved in Step 2 using Lemma 3.6.1. The proof of this step requires the existence of sufficient number of data points in a small neighborhood of all points in the direction of the normalized gradient. We establish that this is indeed the case using compactness arguments in Step 3. Finally, we apply the results of Step 1 to $\{Y_{n,r,j}\}_{j=0}^{j^*}$ yielding that this sequence is close to the solution u_x. Since for all points that are not close to a mode, there exists, by Step 3, data points yielding a positive finite difference approximation of the directional derivative, (3.6.1) occurs with the desired probability. This observation allows to conclude, in Step 4, that Y_{n,r,j^*} is close to the mode.

As a consequence of the above theorem, setting $\mathbf{J}_n = \mathbf{I}_{[Y_{n,r,j^*} \notin R_\alpha \cap C(m_i)]}$, $\delta \in (0,1]$, and $\{\eta_n\}_{n=1}^{\infty}$ be a sequence of scalars in $(0,1]$ with $\lim_{n\to\infty} \eta_n = 0$ one can show by Theorem 3.6.2 that

$$\lim_{n\to\infty} \mathbb{P}(\mathbf{J}_n \geq \delta) = \lim_{n\to\infty} \mathbb{P}(\mathbf{J}_n = 1) \leq \lim_{n\to\infty} \eta_n = 0,$$

implying that $\mathbf{J}_n$ converges in probability to zero. Since Y_{n,r,j^*} is the estimate of the mode, we obtain weak consistency of the mode. Furthermore, using (iii) of Lemma 3.6.1, one can strengthen the conclusion to almost sure convergence. We summarize this observation as a corollary.

Corollary 3.6.1 *Suppose that* $\lim_{n\to\infty} \frac{n}{\log(n)} h_n^{2k} \tau_n^{2kd} = \infty$ *and the assumptions of Theorem 3.6.2 hold. Then* $\mathbf{J}_n \xrightarrow{a.s.} 0$.

It is important to note that one can weaken some of the conditions in the above Theorem. Indeed, it follows from the next proposition shows that, for $d \geq 6k+1$, the conditions involving $\{h_n\}_{n=1}^{\infty}$ can be removed provided that the sequence $\{\tau_n\}_{n=1}^{\infty}$ does not converge to zero too fast.

Proposition 3.6.1 *Let* $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ *a sample of i.i.d. random variables from a probability distribution* P *with continuous and bounded density* f, $x \in S_f$, *and* $\tilde{h}_n = \min_{y,z \in \mathcal{X}_n \cup \{x\}, y \neq z} \|y - z\|_2$. *Then, (i)* $\tilde{h}_n > 0$ *a.s., (ii)* $\tilde{h}_n \xrightarrow{a.s.} 0$, *and (iii) for* $d \geq 6k+1$ *and* $0 < \delta < 1 - \frac{6k}{d}$, $n^{1-\delta}\tilde{h}_n^{2k} \xrightarrow{a.s.} \infty$.

Proposition 3.6.1 shows that the distance $\tilde{h}_n$ between all sample points and a point $x \in S_f$ is positive a.s. for all $n \in \mathbb{N}$ and converges to zero a.s. as $n \to \infty$. However, $d \geq 6k+1$ is needed for $n\tilde{h}_n^{2k}\tau_n^{2kd} \xrightarrow{a.s.} \infty$, for some sequence of positive scalars $\{\tau_n\}_{n=1}^{\infty}$ converging to zero. Specifically, by Proposition 3.6.1 (iii) we can take $\tau_n^{2kd} = n^{-\delta}$, for some $0 < \delta < 1 - \frac{6k}{d}$, that is, $\tau_n = n^{-\delta/(2kd)}$. This shows that, for $d \geq 6k+1$, by choosing a suitable sequence $\{\tau_n\}_{n=1}^{\infty}$, we can replace h_n by $\tilde{h}_n$ in Theorem 3.6.2. In turn, this allows replacement of the set $\mathcal{X}_{n,r}(x) = \{X \in \mathcal{X}_n : h_n \leq \|X - x\|_2 \leq r\}$ by $\tilde{\mathcal{X}}_{n,r}(x) = \{X \in \mathcal{X}_n : \|X - x\|_2 \leq r, X \neq x\}$.

3.7 Proof of preliminary results

This section contains the proof of Theorem 3.6.1, Lemma 3.6.1, and other lemmas for the proof of Theorem 3.6.2.

Proof of Theorem 3.6.1. We begin by proving (i). Let $h^*, n^* > 0$ be such that $(K)^{+h^*} \subset S_f$ and $0 < h_n \leq h^*$ for all $n \geq n^*$. Notice that f_{τ_n} is continuously differentiable in $(K)^{+h^*}$ (see Remark 2.3.1). By the mean value theorem, there exist $0 \leq c_{1,n}, c_{2,n} \leq 1$ such that

$$f(x+h_n v_n) - f(x) = h_n \langle \nabla f(x + c_{1,n} h_n v_n), v_n \rangle \tag{3.7.1}$$

and

$$f_{\tau_n}(x+h_n v_n) - f_{\tau_n}(x) = h_n \langle \nabla f_{\tau_n}(x + c_{2,n} h_n v_n), v_n \rangle. \tag{3.7.2}$$

Using the triangle inequality, we have that

$$\sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n}(x) - \nabla_v f(x)| \leq \sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n}(x) - \nabla_{v_n}^{h_n} f(x)| + \sup_{x\in K} |\nabla_{v_n}^{h_n} f(x) - \nabla_v f(x)|.$$

We show that each term converges to 0 as $n \to \infty$. First, by (3.7.1), the uniform continuity of ∇f in $(K)^{+h^*}$ and $\lim_{n\to\infty} \|v_n - v\|_2 = 0$, it follows that

$$\begin{aligned}
\lim_{n\to\infty} \sup_{x\in K} |\nabla_{v_n}^{h_n} f(x) - \nabla_v f(x)| &= \lim_{n\to\infty} \sup_{x\in K} |\langle \nabla f(x + c_{1,n} h_n v_n), v_n\rangle - \langle \nabla f(x), v\rangle| \\
&\leq \lim_{n\to\infty} \sup_{x\in K} |\langle \nabla f(x + c_{1,n} h_n v_n), v_n - v\rangle| \\
&\quad + \lim_{n\to\infty} \sup_{x\in K} |\langle \nabla f(x + c_{1,n} h_n v_n) - \nabla f(x), v\rangle| \\
&\leq \sup_{y\in (K)^{+h^*}} \|\nabla f(y)\|_2 \lim_{n\to\infty} \|v_n - v\|_2 \\
&\quad + \lim_{n\to\infty} \sup_{x\in K} \|\nabla f(x + c_{1,n} h_n v_n) - \nabla f(x)\|_2 = 0.
\end{aligned}$$

Also, by (3.7.1) and (3.7.2), it holds that

$$\begin{aligned}
\sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n}(x) - \nabla_{v_n}^{h_n} f(x)| &= \sup_{x\in K} |\langle \nabla f_{\tau_n}(x + c_{2,n} h_n v_n) - \nabla f(x + c_{1,n} h_n v_n), v_n\rangle| \\
&\leq \sup_{x\in K} \|\nabla f_{\tau_n}(x + c_{2,n} h_n v_n) - \nabla f(x + c_{1,n} h_n v_n)\|_2.
\end{aligned}$$

Finally, Proposition 2.3.2 (iv) and the uniform continuity of ∇f in $(K)^{+h^*}$ imply that

$$\begin{aligned}
\lim_{n\to\infty} \sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n}(x) - \nabla_{v_n}^{h_n} f(x)| &\leq \lim_{n\to\infty} \sup_{x\in K} \|\nabla f_{\tau_n}(x + c_{2,n} h_n v_n) - \nabla f(x + c_{2,n} h_n v_n)\|_2 \\
&\quad + \lim_{n\to\infty} \sup_{x\in K} \|\nabla f(x + c_{2,n} h_n v_n) - \nabla f(x + c_{1,n} h_n v_n)\|_2 \\
&\leq \lim_{n\to\infty} \sup_{y\in (K)^{+h^*}} \|\nabla f_{\tau_n}(y) - \nabla f(y)\|_2 \\
&\quad + \sup_{y\in (K)^{+h^*}} \lim_{n\to\infty} \sup_{z\in \overline{B}_{h_n}(y)\cap (K)^{+h^*}} \|\nabla f(y) - \nabla f(z)\|_2 = 0.
\end{aligned}$$

We now prove (ii). Since

$$\sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n,n}(x) - \nabla_v f(x)| \leq \sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n,n}(x) - \nabla_{v_n}^{h_n} f_{\tau_n}(x)| + \sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n}(x) - \nabla_v f(x)|,$$

using (i), it is enough to show that

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{x\in K}|\nabla^{h_n}_{v_n} f_{\tau_n,n}(x) - \nabla^{h_n}_{v_n} f_{\tau_n}(x)| \geq \frac{\epsilon}{2}\right) = 0.$$

Notice that, by Lemma 2.3.1,

$$\begin{aligned}\sup_{x\in K}|\nabla^{h_n}_{v_n} f_{\tau_n,n}(x) - \nabla^{h_n}_{v_n} f_{\tau_n}(x)| &= \sup_{x\in K}\left|\frac{f_{\tau_n,n}(x+h_n v_n) - f_{\tau_n}(x+h_n v_n)}{h_n} - \frac{f_{\tau_n,n}(x) - f_{\tau_n}(x)}{h_n}\right| \\ &\leq \sup_{x\in K}\left|\frac{L_{G,n}(x+h_n v_n,\tau_n) - L_G(x+h_n v_n,\tau_n)}{h_n^k\tau_n^{kd}\Lambda_1}\right|^{1/k} \\ &+ \sup_{x\in K}\left|\frac{L_{G,n}(x,\tau_n) - L_G(x,\tau_n)}{h_n^k\tau_n^{kd}\Lambda_1}\right|^{1/k}.\end{aligned}$$

We now use that $\lim_{n\to\infty}\sqrt{n}h_n^k\tau_n^{kd} = \infty$ and apply Theorem 2.4.3 with $t = t_n = \sqrt{n}h_n^k\tau_n^{kd}\Lambda_1(\epsilon/4)^k$. It follows that there are constants $\sigma_G \geq 0$ and $1 < c_{G,0}, c_{G,1}, c_{G,2} < \infty$ such that, for all $n \geq n^{**}$, $t_n \geq \max(2^3\sigma_G, 2^4 c_{G,0})$ and

$$\begin{aligned}\mathbb{P}\left(\sup_{x\in K}|\nabla^{h_n}_{v_n} f_{\tau_n,n}(x) - \nabla^{h_n}_{v_n} f_{\tau_n}(x)| > \frac{\epsilon}{2}\right) &\leq \mathbb{P}\left(\sup_{(x,\tau)\in\mathbb{R}^d\times[0,\infty]}\left|\frac{L_{G,n}(x,\tau) - L_G(x,\tau)}{h_n^k\tau_n^{kd}\Lambda_1}\right|^{1/k} \geq \frac{\epsilon}{4}\right) \\ &= \mathbb{P}\left(\sqrt{n}\sup_{(x,\tau)\in\mathbb{R}^d\times[0,\infty]}|L_{G,n}(x,\tau) - L_G(x,\tau)| \geq t_n\right) \\ &\leq M_G(n,t_n),\end{aligned}$$

where M_G is defined in (2.4.5). Now, the result follows from $\lim_{n\to\infty} M_G(n,t_n) = 0$. ■

Proof of Lemma 3.6.1. We begin by proving (i). By Proposition 2.3.3 there are constants $\tilde{\tau}((K)^{+h^*}), \tilde{c}_2((K)^{+h^*}) > 0$ such that, for all $y \in (K)^{+h^*}$ and $0 < \tau \leq \tilde{\tau}((K)^{+h^*})$,

$$f_\tau(y) = f(y) + \tilde{R}_\tau(y)\tau^2$$

and $\|\nabla\tilde{R}_\tau(y)\|_2 \leq \tilde{c}_2((K)^{+h^*})$. Let $n^* \in \mathbb{N}$ such that $\tau_n \leq \tilde{\tau}((K)^{+h^*})$ for all $n \geq n^*$. It holds that, for all $n \geq n^*$,

$$\nabla^h_v f_{\tau_n}(x) - \nabla^h_v f(x) = \frac{\tilde{R}_{\tau_n}(x+hv) - \tilde{R}_{\tau_n}(x)}{h}\tau_n^2.$$

Now, by the mean value theorem, there are constants $0 \leq \tilde{c}_n \leq 1$ such that

$$\frac{\tilde{R}_{\tau_n}(x+hv) - \tilde{R}_{\tau_n}(x)}{h} = \langle\nabla\tilde{R}_{\tau_n}(x+\tilde{c}_n hv), v\rangle,$$

implying that

$$\left|\frac{\tilde{R}_{\tau_n}(x+hv) - \tilde{R}_{\tau_n}(x)}{h}\right| \leq \|\nabla\tilde{R}_{\tau_n}(x+\tilde{c}_n hv)\|_2 \leq \tilde{c}_2((K)^{+h^*}).$$

It follows that

$$\lim_{n\to\infty}\sup_{h\in[h_n,h^*]}\sup_{v\in S^{d-1}}\sup_{x\in K}|\nabla^h_v f_{\tau_n}(x) - \nabla^h_v f(x)| \leq \tilde{c}_2((K)^{+h^*})\lim_{n\to\infty}\tau_n^2 = 0.$$

We now prove (ii). By (i), it is enough to show that

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{h\in[h_n,h^*]} \sup_{v\in S^{d-1}} \sup_{x\in K} |\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f_{\tau_n}(x)| \geq \frac{\epsilon}{2} \right) = 0.$$

Notice that, by Lemma 2.3.1,

$$\begin{aligned} |\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f_{\tau_n}(x)| &= \left| \frac{f_{\tau_n,n}(x+hv) - f_{\tau_n}(x+hv)}{h} - \frac{f_{\tau_n,n}(x) - f_{\tau_n}(x)}{h} \right| \\ &\leq \left| \frac{L_{G,n}(x+hv,\tau_n) - L_G(x+hv,\tau_n)}{h_n^k \tau_n^{kd} \Lambda_1} \right|^{1/k} \\ &+ \left| \frac{L_{G,n}(x,\tau_n) - L_G(x,\tau_n)}{h_n^k \tau_n^{kd} \Lambda_1} \right|^{1/k} \\ &\leq 2 \sup_{(x,\tau)\in\mathbb{R}^d\times[0,\infty]} \left| \frac{L_{G,n}(x,\tau) - L_G(x,\tau)}{h_n^k \tau_n^{kd} \Lambda_1} \right|^{1/k}. \end{aligned}$$

We apply again Theorem 2.4.3 with $t = t_n = \sqrt{n} h_n^k \tau_n^{kd} \Lambda_1 (\epsilon/4)^k$. Then, there are constants $\sigma_G \geq 0$ and $1 < c_{G,0}, c_{G,1}, c_{G,2} < \infty$ such that, for large enough n,

$$\begin{aligned} &\mathbb{P}\left(\sup_{h\in[h_n,h^*]} \sup_{v\in S^{d-1}} \sup_{x\in K} |\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f_{\tau_n}(x)| \geq \frac{\epsilon}{2} \right) \\ \leq &\mathbb{P}\left(\sqrt{n} \sup_{(x,\tau)\in\mathbb{R}^d\times[0,\infty]} |L_{G,n}(x,\tau) - L_G(x,\tau)| \geq t_n \right) \\ \leq &M_G(n,t_n), \end{aligned}$$

where M_G is defined in (2.4.5). Since $\lim_{n\to\infty} t_n = \infty$, $\lim_{n\to\infty} h_n = 0$, and $\lim_{n\to\infty} \tau_n = 0$, we conclude that $\lim_{n\to\infty} M_G(n,t_n) = 0$. Finally, for (iii), we apply Lemma 2.5.1 with $a_n = h_n^k \tau_n^{kd}$ and $b = \Lambda_1(\epsilon/4)^k$ and get constants $0 < \tilde{c} < \infty$ and $\tilde{n} \in \mathbb{N}$ such that $M_G(n,t_n) \leq \frac{\tilde{c}}{n^2}$ for all $n \geq \tilde{n}$. ∎

A version of discrete Grönwall lemma (see e.g. Holte (2009)) is needed in Theorem 3.6.2 to evaluate the difference between the sequence $\{y_{n,r,j}\}_{j=1}^{j^*}$ (defined in the proof) and the solution u_x of (3.3.2). Discrete Grönwall lemma is a suitable tool for this scope. Indeed, it is often used to compare the solution of ordinary differential equations with the approximation given by Euler method (see e.g. Theorem 2.4 in Atkinson et al. (2009)).

Lemma 3.7.1 (Discrete Grönwall lemma) *Let $\{a_n\}_{n=0}^\infty$, $\{b_n\}_{n=0}^\infty$ and $\{c_n\}_{n=0}^\infty$ be non-negative sequences. If $a_0 = 0$ and $a_n \leq (1+c_{n-1})a_{n-1} + b_{n-1}$ for all $n \geq 1$, then, $a_n \leq (\sum_{j=0}^{n-1} b_j) \exp(\sum_{j=1}^{n-1} c_j)$.*

Proof of Lemma 3.7.1. By applying recursively the inequality for $\{a_n\}_{n=0}^\infty$ and using $a_0 = 0$, we see that

$$a_n \leq \sum_{j=0}^{n-1} b_j \prod_{l=j+1}^{n-1} (1+c_l).$$

Now, using $1+s \le e^s$ with $s = c_l$, we get that

$$a_n \le \sum_{j=0}^{n-1} b_j \exp(\sum_{l=j+1}^{n-1} c_l) \le (\sum_{j=0}^{n-1} b_j) \exp(\sum_{j=1}^{n-1} c_j).$$

■

The next lemma is used to control the finite difference approximations of the directional derivative of f over certain compact sets. For a differentiable function $g : \mathbb{R}^d \to \mathbb{R}$, let

$$\mathrm{w}_g(x) = \begin{cases} \nabla g(x)/\|\nabla g(x)\|_2 & \text{if } \nabla g(x) \neq 0 \\ 0 & \text{if } \nabla g(x) = 0. \end{cases} \tag{3.7.3}$$

Lemma 3.7.2 *Suppose that f is continuously differentiable and K is a compact subset of S_f with $K \cap \mathrm{N}_f = \emptyset$. Then, there exist $\tilde{r}(K), \tilde{c}(K) > 0$ such that $(K)^{+\tilde{r}(K)} \subset S_f$ and, for all $x \in K$ and $(h, v) \in (0, \tilde{r}(K)] \times (S^{d-1} \cap \overline{B}_{\tilde{r}(K)}(\mathrm{w}_f(x)))$, $\nabla_v^h f(x) \ge \tilde{c}(K)$.*

Proof of Lemma 3.7.2. Let $g : [0, \infty) \to \mathbb{R}$ be given by

$$g(h) = \min_{y \in K} \left(f(y + h\mathrm{w}_f(y)) - f(y) \right).$$

By the mean value theorem, it holds that $g(h) = h \min_{y \in K} \langle \nabla f(y + ch\mathrm{w}_f(y)), \mathrm{w}_f(y) \rangle$, for some $0 \le c \le 1$. Let $\tilde{h}(K) > 0$ such that $(K)^{+\tilde{h}(K)} \subset S_f$. Since, by Remark 2.3.1, ∇f is uniformly continuous in $(K)^{+\tilde{h}(K)}$, we have that

$$g'(0) = \lim_{h \to 0^+} g(h)/h = \min_{y \in K} \|\nabla f(y)\|_2. \tag{3.7.4}$$

Now, by multivariate Taylor's theorem with integral remainder, we have that, for $v \in S^{d-1}$ and $h > 0$,

$$\begin{aligned} f(x+hv) &= f(x + h\mathrm{w}_f(x)) + h\langle \nabla f(x + h\mathrm{w}_f(x)), v - \mathrm{w}_f(x) \rangle \\ &\quad + h^2 \int_0^1 (1-s)(v - \mathrm{w}_f(x))^\top \mathrm{H}_f(x + hs(v - \mathrm{w}_f(x)))(v - \mathrm{w}_f(x)) ds. \end{aligned}$$

It follows that, for $0 < h \le \tilde{h}(K)/2$,

$$\begin{aligned} f(x+hv) &\ge f(x) + g(h) + h\langle \nabla f(x + h\mathrm{w}_f(x)), v - \mathrm{w}_f(x) \rangle \\ &\quad + h^2 \int_0^1 (1-s)(v - \mathrm{w}_f(x))^\top \mathrm{H}_f(x + hs(v - \mathrm{w}_f(x)))(v - \mathrm{w}_f(x)) ds \\ &\ge f(x) + g(h) - h\|v - \mathrm{w}_f(x)\|_2 \|\nabla f(x + h\mathrm{w}_f(x))\|_2 \\ &\quad - h^2 \|v - \mathrm{w}_f(x)\|_2^2 \int_0^1 (1-s)\|\mathrm{H}_f(x + hs(v - \mathrm{w}_f(x))\|_{\mathcal{M},2} ds \\ &\ge f(x) + g(h) - h\|v - \mathrm{w}_f(x)\|_2 c_1 - h^2 \|v - \mathrm{w}_f(x)\|_2^2 c_2/2, \end{aligned}$$

where

$$c_1 = \max_{y \in (K)^{+\tilde{h}(K)/2}} \|\nabla f(y)\|_2$$

and

$$c_2 = \max_{y \in (K)^{+\tilde{h}(K)}} \|\mathrm{H}_f(y)\|_{\mathcal{M},2}.$$

Therefore, we have that

$$\nabla_v^h f(x) \geq \tilde{g}(h) = g(h)/h - \|v - \mathrm{w}_f(x)\|_2 c_1 - h\|v - \mathrm{w}_f(x)\|_2^2 c_2/2.$$

Since f has no stationary points in K, $\min_{y \in K}\|\nabla f(y)\|_2 > 0$, and the result follows from (3.7.4). ∎

3.8 Proof of consistency of empirical clusters

In this section, we prove Theorem 3.6.2, Corollary 3.6.1, and Proposition 3.6.1.

Proof of Theorem 3.6.2. First, notice that, for all $j = 1, \ldots, j^*$, $Y_{n,r,j} \in \mathcal{X}_n$ and, by (3.6.1), $Y_{n,r,j} \neq Y_{n,r,l}$, for all $l < j$. Hence, $j^* \leq n$. The proof of the remaining part is divided into four steps. In the first step below we introduce few notations and preliminary calculations.

Step 0. Let $\alpha_1 = f(x)$ and $0 < \alpha_2 < \alpha_1$. We recall from Section 3.3 that the solution $u_x(t)$ of (3.3.2) exists for $t \in (a, \infty)$, $a < 0$, and define $\mathrm{G}_x = \{u_x(t) : t \in [0, \infty)\}$. Since $x \in \mathrm{R}_{\alpha_1}$ and $f(u_x(\cdot))$ is monotonically non-decreasing, we have that $\overline{\mathrm{G}}_x = \mathrm{G}_x \cup \{m_i\} \subset C(m_i) \cap \mathrm{R}_{\alpha_1} \subset C(m_i) \cap \mathring{\mathrm{R}}_{\alpha_2}$. $C(m_i)$ is open by Proposition 3.3.1 and therefore there exists $\xi_1 > 0$ such that (i) $\alpha_3 = \sup_{y \in (C(m_i) \setminus (C(m_i))^{-2\xi_1})} f(y) < f(m_i)$ and (ii) $(\mathrm{G}_x)^{+2\xi_1} \subset (C(m_i) \cap \mathrm{R}_{\alpha_2})^{-2\xi_1}$. Let $\max(\alpha_2, \alpha_3) < \alpha^* \leq \alpha < f(m_i)$ and $0 < \xi \leq \xi_1$ such that

$$\overline{B}_{4\xi}(m_i) \subset \mathrm{R}_\alpha \cap C(m_i). \tag{3.8.1}$$

Let $K_\xi = \mathrm{R}_{\alpha_2} \cap \overline{(C(m_i))^{-\xi}}$. It holds that $(\mathrm{G}_x)^{+\xi} \subset K_\xi$, which implies that, for all $\epsilon > 0$,

$$(\mathrm{G}_x)^{+\xi} \setminus B_\epsilon(m_i) \subset K_\xi \setminus B_\epsilon(m_i). \tag{3.8.2}$$

Also, since $\alpha > \alpha_3$, $\mathrm{R}_\alpha \cap C(m_i) = \mathrm{R}_\alpha \cap (C(m_i))^{-2\xi_1}$, implying that

$$\overline{B}_{4\xi}(m_i) \subset K_\xi. \tag{3.8.3}$$

Recall (3.7.3) and for $0 < r \leq \xi$ and $j^* \geq 0$ let

$$\begin{aligned}\tilde{\mathrm{G}}_{x,j^*,r} = \Big\{ &\{y_{r,j}\}_{j=0}^{j^*} : y_{r,0} = x \text{ and, recursively, } y_{r,j+1} = y_{r,j} + h_j v_j \\ &\text{for some } (h_j, v_j) \in (0, r] \times (S^{d-1} \cap \overline{B}_r(\mathrm{w}_f(y_{r,j}))\Big\}.\end{aligned} \tag{3.8.4}$$

Step 1. We show that, for small r, every sequence $\{y_{r,j}\}_{j=0}^{j^*} \in \tilde{\mathrm{G}}_{x,j^*,r}$ either remains in $(\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i)$ or, for some $j \in \{0, \ldots, j^*\}$, $y_{r,j} \in \overline{B}_{4\xi}(m_i)$. To this end, we suppose w.l.o.g. that

$$\|x - m_i\|_2 > 2\xi. \tag{3.8.5}$$

If (3.8.5) does not hold, then $y_{r,0} = x \in \overline{B}_{2\xi}(m_i) \subset \overline{B}_{4\xi}(m_i)$. We now define some quantities that are used in the proof of this fact. Specifically, let $t_0 = 0$ and, recursively, $t_{j+1} = \sum_{l=0}^{j} h_l / \|\nabla f(y_{r,l})\|_2$. Also, let $0 < \tilde{\alpha}_4(\xi) < f(m_i)$ such that $\mathrm{R}_{\tilde{\alpha}_4(\xi)} \cap C(m_i) \subset B_\xi(m_i)$, $t^*(\xi) = \inf\{t \in [0, \infty) : u_x(t) \in \mathrm{R}_{\tilde{\alpha}_4(\xi)}\}$, $\tilde{t}^*(\xi) = \inf\{t \in [0, t^*(\xi)] : u_x(t) \in \overline{B}_{2\xi}(m_i)\}$, $\tilde{K}_\xi = K_\xi \setminus \mathring{\mathrm{R}}_{\tilde{\alpha}_4(\xi)}$, $\tilde{c}_1(\xi) = \inf_{y \in \tilde{K}_\xi}\|\nabla f(y)\|_2 > 0$ and $\tilde{c}_2(\xi) =$

$\sup_{y\in\tilde{K}_\xi}\|\nabla f(y)\|_2 > 0$. Since ∇f is differentiable, it is locally Lipschitz; hence, Lipschitz in $\tilde{K}_\xi$. Denote by L the Lipschitz constant. Let $\tilde{j}^* = \max\{j \in \{0,\dots,j^*\} : t_j \leq \tilde{t}^*(\xi)\}$ and, using the continuity of u_x, let $0 < r_1 \leq \xi$, such that

$$r_1\tilde{t}^*(\xi)\left(\tilde{c}_2(\xi) + \sup_{t\in[0,\tilde{t}^*(\xi)]}\|u_x''(t)\|_2/(2\tilde{c}_1(\xi))\right)\exp(L\tilde{t}^*(\xi)) \leq \xi \tag{3.8.6}$$

and, for all $0 < r \leq r_1$,

$$\|u_x(\tilde{t}(\xi) - r/\tilde{c}_1(\xi)) - u_x(\tilde{t}(\xi))\|_2 \leq \xi. \tag{3.8.7}$$

We show that, for all $j = 0,\dots,\tilde{j}^*$ and $0 < r \leq r_1$, $y_{r,j} \in (\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i)$. We recall that, by (3.8.2), since $\mathrm{R}_{\tilde{\alpha}_4(\xi)} \cap C(m_i) \subset B_\xi(m_i)$, $(\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i) \subset \tilde{K}_\xi$. First, notice that $u_x(t_0) = x$ and, by (3.8.5), it holds that $y_{r,0} = x \in (\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i)$. We now suppose by induction that, for $j \geq 1$, $y_{n,r,j-1} \in (\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i)$ and show that $y_{r,j} \in (\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i)$, thus proving the statement. Since $u_x'(t) = \nabla f(u_x(t))$ and f is three times continuously differentiable, then so is u_x. By Taylor theorem with Lagrange's form of remainder, there exists $t_{j-1} \leq \tilde{t}_{j-1} \leq t_j$ such that

$$u_x(t_j) = u_x(t_{j-1}) + \frac{h_{j-1}}{\|\nabla f(y_{r,j-1})\|_2}\nabla f(u_x(t_{j-1})) + \frac{h_{j-1}^2}{2\|\nabla f(y_{r,j-1})\|_2^2}u_x''(\tilde{t}_{j-1}).$$

It follows that

$$\begin{aligned}(y_{r,j} - u_x(t_j)) = {} & (y_{r,j-1} - u_x(t_{j-1})) + h_{j-1}(v_{j-1} - \mathrm{w}_f(y_{r,j-1})) \\ & + \frac{h_{j-1}}{\|\nabla f(y_{r,j-1})\|_2}\left(\nabla f(y_{r,j-1}) - \nabla f(u_x(t_{j-1}))\right) \\ & - \frac{h_{j-1}^2}{2\|\nabla f(y_{r,j-1})\|_2^2}u_x''(\tilde{t}_{j-1}).\end{aligned}$$

Now, we use the Lipschitz property of ∇f and get

$$\begin{aligned}\|y_{r,j} - u_x(t_j)\|_2 \leq {} & \left(1 + \frac{h_{j-1}L}{\|\nabla f(y_{r,j-1})\|_2}\right)\|y_{r,j-1} - u_x(t_{j-1})\|_2 + r_1 h_{j-1} \\ & + \frac{h_{j-1}^2}{2\|\nabla f(y_{r,j-1})\|_2^2}\sup_{t\in[0,\tilde{t}^*(\xi)]}\|u_x''(t)\|_2.\end{aligned}$$

We now apply Lemma 3.7.1 with $a_j = \|y_{r,j} - u_x(t_j)\|_2$,

$$b_j = r_1 h_j + \frac{h_j^2}{2\|\nabla f(y_{r,j})\|_2^2}\sup_{t\in[0,\tilde{t}^*(\xi)]}\|u_x''(t)\|_2,$$

$c_j = \frac{h_j L}{\|\nabla f(y_{r,j})\|_2}$ and, using (3.8.6) and $t_j \le \tilde{t}^*(\xi)$, we get that $\|y_{r,j} - u_x(t_j)\|_2$ is bounded above by

$$\left(r_1 \sum_{l=0}^{j-1} h_l + \sum_{l=0}^{j-1} \frac{h_j^2}{2\|\nabla f(y_{r,l})\|_2^2} \sup_{t\in[0,\tilde{t}^*(\xi)]} \|u_x''(t)\|_2 \right) \exp\left(L \sum_{l=1}^{j-1} \frac{h_l}{\|\nabla f(y_{r,j}\|_2} \right)$$
$$\le r_1 t_j \left(\tilde{c}_2(\xi) + \sup_{t\in[0,\tilde{t}^*(\xi)]} \|u_x''(t)\|_2 / (2\tilde{c}_1(\xi)) \right) \exp(Lt_j) \le \xi.$$

It follows that $y_{r,j} \in (\mathrm{G}_x)^{+\xi}$. Moreover, $t_j \le \tilde{t}^*(\xi)$ implies that $\|m_i - u_x(t_j)\|_2 \ge 2\xi$. Hence,

$$\|m_i - y_{r,j}\|_2 \ge \|m_i - u_x(t_j)\|_2 - \|u_x(t_j) - y_{r,j}\|_2 \ge \xi,$$

that is, $y_{r,j} \notin B_\xi(m_i)$. In particular, if $\tilde{j}^* = j^*$, then $y_{r,j} \in (\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i)$ for all $j = 0, \dots, j^*$. Next, we show that, if $\tilde{j}^* < j^*$, then $y_{r,\tilde{j}^*} \in \overline{B}_{4\xi}(m_i)$. Since $\tilde{t}^*(\xi) - r_1/\tilde{c}_1(\xi) < t_{\tilde{j}^*+1} - r_1/\tilde{c}_1(\xi) \le t_{\tilde{j}^*} \le \tilde{t}^*(\xi)$, by (3.8.7) it holds that $\|u_x(t_{\tilde{j}^*}) - u_x(\tilde{t}_\xi^*)\|_2 \le \xi$. Since $u_x(\tilde{t}^*(\xi)) \in \partial B_{2\xi}(m_i)$, we conclude that

$$\|y_{r,\tilde{j}^*} - m_i\|_2 \le \|y_{r,\tilde{j}^*} - u_x(t_{\tilde{j}^*})\|_2 + \|u_x(t_{\tilde{j}^*}) - u_x(\tilde{t}_\xi^*)\|_2 + \|u_x(\tilde{t}_\xi^*) - m_i\|_2 \le 4\xi.$$

Step 2. Notice that $\tilde{K}_\xi \cap \mathrm{N}_f = \emptyset$. We apply Lemma 3.7.2 with $K = \tilde{K}_\xi$ and get constants $r^* = \min(r_1, \tilde{r}(\tilde{K}_\xi)) > 0$ and $c^* = \tilde{c}(\tilde{K}_\xi) > 0$ such that, for all $x \in \tilde{K}_\xi$ and $(h, v) \in (0, r^*] \times (S^{d-1} \cap \overline{B}_{r^*}(\mathrm{w}_f(x)))$,

$$\nabla_v^h f(x) \ge c^*. \tag{3.8.8}$$

For $X \in \mathcal{X}_n$ and $x \in S_f$ let $\tilde{h}_{X,x} = \|X - x\|_2$ and $\tilde{v}_{X,x} = (X - x)/\tilde{h}_{X,x}$. We show the existence of $0 < r_2 \le r^*$ such that, for all $0 < r \le r_2$, there exist $n_1, n_2 \in \mathbb{N}$ such that, with probability at least $1 - \eta$, for $n \ge \max(n_1, n_2)$ and $x \in \tilde{K}_\xi$, we have that $\mathcal{X}_{n,r}(x) \ne \emptyset$,

$$\max_{X\in\mathcal{X}_{n,r}(x)\cup\{x\}} f_{\tau_n,n}(X) - f_{\tau_n,n}(x) > 0 \tag{3.8.9}$$

and

$$X^*(x) = \operatorname*{argmax}_{X\in\mathcal{X}_{n,r}(x)} \frac{f_{\tau_n,n}(X) - f_{\tau_n,n}(x)}{\|X - x\|_2} = \operatorname*{argmax}_{X\in\mathcal{X}_{n,r}(x)} \nabla_{\tilde{v}_{X,x}}^{\tilde{h}_{X,x}} f_{\tau_n,n}(x)$$

satisfies

$$(\tilde{h}_{X^*(x),x}, \tilde{v}_{X^*(x),x}) \in [h_n, r] \times (S^{d-1} \cap \overline{B}_{r^*}(\mathrm{w}_f(x)). \tag{3.8.10}$$

To this end, suppose w.l.o.g. that $r^* \le 1$. Let $\tilde{d}(r^*) = \inf_{y\in\tilde{K}_\xi} \inf_{v\in S^{d-1}\setminus\overline{B}_{r^*}(\mathrm{w}_f(y))} \langle \mathrm{w}_f(y) - v, \nabla f(y)\rangle > 0$ and choose $0 < d^* < \tilde{d}(r^*)/(5\max_{y\in\tilde{K}_\xi}\|\nabla f(y)\|_2)$. Notice that, since $\tilde{d}(r^*) \le \min_{y\in\tilde{K}_\xi}\|\nabla f(y)\|_2 r^*$, we have $d^* < r^*/5 \le 1/5$. By the mean value theorem, there exists $0 \le c \le 1$ such that $\nabla_v^h f(x) = \langle v, \nabla f(x + chv)\rangle$, $x \in \tilde{K}_\xi$. Next, by the uniform continuity of ∇f over compact sets, we have that $\nabla_v^h f(x)$ converges to $\nabla_v f(x)$ uniformly over $v \in S^{d-1}$ and $x \in \tilde{K}_\xi$. Let $r_3 > 0$ be such that, for all $h \in (0, r_3]$, $v \in S^{d-1}$, and $x \in \tilde{K}_\xi$,

$$|\nabla_v^h f(x) - \nabla_v f(x)| \le \min_{y\in\tilde{K}_\xi}\|\nabla f(y)\|_2 d^*.$$

Then, for all $x \in \tilde{K}_\xi$ and $v \in S^{d-1} \cap \overline{B}_{d^*}(\mathrm{w}_f(x))$, it holds that

$$\nabla_v f(x) \geq \|\nabla f(x)\|_2(1 - \|\mathrm{w}_f(x) - v\|_2) \geq \|\nabla f(x)\|_2(1 - d^*),$$

which implies that, for all $x \in \tilde{K}_\xi$, $h \in (0, r_3]$ and $v \in S^{d-1} \cap \overline{B}_{d^*}(\mathrm{w}_f(x))$,

$$\nabla_v^h f(x) \geq \nabla_v f(x) - \min_{y \in \tilde{K}_\xi} \|\nabla f(y)\|_2 d^* \geq \|\nabla f(x)\|_2(1 - 2d^*). \tag{3.8.11}$$

On the other hand, by definition of d^*, we have that, for all $x \in \tilde{K}_\xi$ and $v \in S^{d-1} \setminus \overline{B}_{r^*}(\mathrm{w}_f(x))$,

$$\nabla_v f(x) \leq (1 - 5d^*)\|\nabla f(x)\|_2,$$

which implies that, for all $x \in \tilde{K}_\xi$, $h \in (0, r_3]$ and $v \in S^{d-1} \setminus \overline{B}_{r^*}(\mathrm{w}_f(x))$,

$$\nabla_v^h f(x) \leq \nabla_v f(x) + \min_{y \in \tilde{K}_\xi} \|\nabla f(y)\|_2 d^* \leq \|\nabla f(x)\|_2(1 - 4d^*). \tag{3.8.12}$$

Now, let $r_2 = \min(r_3, d^*) < r^*$ and $0 < r \leq r_2$. Notice that $(K)^{+r} \subset (K)^{+r^*} \subset (K)^{+r(\tilde{K}_\xi)} \subset S_f$. Using Lemma 3.6.1 (ii) with $K = \tilde{K}_\xi$ and $h^* = r$, we choose $n_2 \in \mathbb{N}$ such that, for all $n \geq n_2$, with probability at least $1 - \eta/2$,

$$\sup_{h \in [h_n, r]} \sup_{v \in S^{d-1}} \sup_{x \in \tilde{K}_\xi} |\nabla_v^h f_{\tau_n, n}(x) - \nabla_v^h f(x)| < d^* \min_{y \in \tilde{K}_\xi} \|\nabla f(y)\|_2. \tag{3.8.13}$$

It follows from (3.8.11), (3.8.12) and (3.8.13) that, with probability at least $1 - \eta/2$, for all $x \in \tilde{K}_\xi$, $h \in [h_n, r]$ and $v \in S^{d-1} \cap \overline{B}_r(\mathrm{w}_f(x))$,

$$\nabla_v^h f_{\tau_n, n}(x) > (1 - 2d^*)\|\nabla f(x)\|_2 - d^* \min_{y \in \tilde{K}_\xi} \|\nabla f(y)\|_2 \geq (1 - 3d^*)\|\nabla f(x)\|_2, \tag{3.8.14}$$

and, for all $x \in \tilde{K}_\xi$, $h \in [h_n, r]$ and $v \in S^{d-1} \setminus \overline{B}_{r^*}(\mathrm{w}_f(x))$,

$$\nabla_v^h f_{\tau_n, n}(x) < (1 - 4d^*)\|\nabla f(x)\|_2 + d^* \min_{y \in \tilde{K}_\xi} \|\nabla f(y)\|_2 \leq (1 - 3d^*)\|\nabla f(x)\|_2. \tag{3.8.15}$$

Since $d^* < 1/5$, $(1 - 3d^*)\|\nabla f(x)\|_2 > 0$ for all $x \in \tilde{K}_\xi$. We show in **Step 3** below that there exists a constant n_1 such that, with probability at least $1 - \eta/2$, for all $x \in \tilde{K}_\xi$ and $n \geq n_1$, there exists $X \in \mathcal{X}_{n,r}(x)$ such that

$$(\tilde{h}_{X,x}, \tilde{v}_{X,x}) \in [h_n, r] \times \left(S^{d-1} \cap \overline{B}_r(\mathrm{w}_f(x))\right). \tag{3.8.16}$$

In particular, since $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$ for all $n \geq \max(n_1, n_2)$, (3.8.13) and (3.8.16) hold simultaneously with probability at least $1 - \eta$. It follows from (3.8.14) and (3.8.15) that, with probability at least $1 - \eta$, for all $x \in \tilde{K}_\xi$,

$$\sup_{(h,v) \in [h_n, r] \times S^{d-1} \setminus \overline{B}_{r^*}(\mathrm{w}_f(x))} \nabla_v^h f_{\tau_n, n}(x) < \nabla_{\tilde{v}_{X,x}}^{\tilde{h}_{X,x}} f_{\tau_n, n}(x) \leq \sup_{(h,v) \in [h_n, r] \times S^{d-1} \cap \overline{B}_r(\mathrm{w}_f(x))} \nabla_v^h f_{\tau_n, n}(x).$$

Thus, we have shown that the finite difference approximation of $f_{\tau_n, n}$ with step $\tilde{h}_{X,x}$ and direction $\tilde{v}_{X,x}$ is larger than all finite difference approximations with step $h \in [h_n, r]$ and directions $v \in S^{d-1} \setminus \overline{B}_{r^*}(\mathrm{w}_f(x))$. (3.8.10) follows. Also, (3.8.14) and (3.8.16) imply (3.8.9).

Step 3. We show (3.8.16). To this end, let $0 < s_1 < s_2 < r$ and n_3 be such that $h_n < s_1$ for all $n \geq n_3$. It is enough to show that there exists $n_1 \geq n_3$ such that, for all $n \geq n_3$,

$$\mathbb{P}([\mathcal{X}_n \cap D_{s_1,s_2}(x) \neq \emptyset \; \forall x \in \tilde{K}_\xi]) \geq 1 - \eta/2,$$

where $D_{s_1,s_2}(x) = A_{s_1,s_2}(x) \cap C_{s_2}(x)$, $A_{s_1,s_2}(x) = \overline{B}_{s_2}(x) \setminus B_{s_1}(x)$, and

$$C_{s_2}(x) = \left\{ y \in \mathbb{R}^d \setminus \{x\} : \left\| \frac{y-x}{\|y-x\|_2} - \mathrm{w}_f(x) \right\|_2 \leq s_2 \right\}.$$

Let $0 < \epsilon_1 < \frac{s_2 - s_1}{2}$. We first notice that

$$A_{s_1+\epsilon_1, s_2-\epsilon_1}(x) \subset \cap_{z \in B_{\epsilon_1}(x)} A_{s_1,s_2}(z). \tag{3.8.17}$$

Indeed, $y \in A_{s_1+\epsilon_1,s_2-\epsilon_1}(x)$ satisfies $s_1 + \epsilon_1 \leq \|y-x\|_2 \leq s_2 - \epsilon_1$. Therefore, for all $z \in B_{\epsilon_1}(x)$, it holds that

$$s_1 \leq \|y-x\|_2 - \|x-z\|_2 \leq \|y-z\|_2 \leq \|y-x\|_2 + \|x-z\|_2 \leq s_2,$$

that is, $y \in A_{s_1,s_2}(z)$. Now, let $h^* > 0$ such that $(\tilde{K}_\xi)^{+h^*}$ does not contain stationary points of f. Since w_f is uniformly continuous in $(\tilde{K}_\xi)^{+h^*}$, there exists $\epsilon_2 \in (0, h^*]$ such that, for all $x \in K$,

$$\sup_{y \in B_{\epsilon_2}(x)} \|\mathrm{w}_f(x) - \mathrm{w}_f(y)\|_2 \leq \epsilon_1/2. \tag{3.8.18}$$

Suppose w.l.o.g. that $\epsilon_2 \leq \min(1, \frac{s_1+\epsilon_1}{4})\epsilon_1$. We show that

$$D_{s_1+\epsilon_1,s_2-\epsilon_1}(x) \subset \cap_{z \in B_{\epsilon_2}(x)} D_{s_1,s_2}(z). \tag{3.8.19}$$

To this end, let $y \in D_{s_1+\epsilon_1,s_2-\epsilon_1}(z)$. By (3.8.17), it holds that $y \in \cap_{z \in B_{\epsilon_2}(x)} A_{s_1,s_2}(z)$. We need to show that $y \in \cap_{z \in B_{\epsilon_2}(x)} C_{s_2}(x)$. Since, for all $z \in B_{\epsilon_2}(x)$,

$$\left\| \frac{y-z}{\|y-z\|_2} - \frac{y-x}{\|y-x\|_2} \right\|_2 \leq 2\frac{\|z-x\|_2}{\|y-z\|_2} \leq \frac{2\epsilon_2}{s_1+\epsilon_1} \leq \epsilon_1/2,$$

using the triangle inequality and (3.8.18), we have that

$$\begin{aligned}\left\| \frac{y-z}{\|y-z\|_2} - \mathrm{w}_f(z) \right\|_2 \leq \left\| \frac{y-z}{\|y-z\|_2} - \frac{y-x}{\|y-x\|_2} \right\|_2 + \left\| \frac{y-x}{\|y-x\|_2} - \mathrm{w}_f(x) \right\| \\ + \|\mathrm{w}_f(x) - \mathrm{w}_f(z)\|_2 \leq s_2.\end{aligned}$$

(3.8.19) follows. Notice that, for all $x \in \tilde{K}_\xi$, $\lambda(D_{s_1+\epsilon_1,s_2-\epsilon_1}(x)) = \lambda(D_{s_1+\epsilon_1,s_2-\epsilon_1}(0)) = \bar{\Lambda} > 0$. Now, by the compactness of $\tilde{K}_\xi \subset \cup_{x \in \tilde{K}_\xi} B_{\epsilon_2}(x)$, there exist $x_1, \ldots, x_k \in \tilde{K}_\xi$ such that $\tilde{K}_\xi \subset \cup_{l=1}^k B_{\epsilon_2}(x_l)$. It follows from (3.8.19) that, for all $z \in \tilde{K}_\xi$, there exists x_l such that $z \in B_{\epsilon_2}(x_l)$ and $D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l) \subset D_{s_1,s_2}(z)$. Therefore, it is enough to show that there exists $n_1 \geq n_3$ such that, for all $n \geq n_1$,

$$\mathbb{P}([\mathcal{X}_n \cap D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l) \neq \emptyset \; \forall l \in \{1, \ldots, k\}]) \geq 1 - \eta/2.$$

To this end, notice that $\cup_{l=1}^{k} D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l) \subset (\tilde{K}_\xi)^{+r} \subset S_f$ and let $\alpha_0 = \min_{y\in(\tilde{K}_\xi)^{+r}} f(y)$. Then, $p_l = P(D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l)) \geq \alpha_0\bar{\Lambda} > 0$. Observe that

$$\begin{aligned}\mathbb{P}(\cap_{l=1}^{k}[\mathcal{X}_n \cap D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l) \neq \emptyset]) &= 1 - \mathbb{P}(\cup_{l=1}^{k}[\mathcal{X}_n \cap D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l) = \emptyset]) \\ &\geq 1 - \sum_{l=1}^{k} \mathbb{P}([\mathcal{X}_n \cap D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l) = \emptyset]).\end{aligned}$$

Let G_l have the geometric distribution with parameter p_l. Since $\{X_l\}$ are independent, it holds that

$$\mathbb{P}([\mathcal{X}_n \cap D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l) = \emptyset]) = \mathbb{P}(G_l > n) = \sum_{j=n}^{\infty}(1-p_l)^j p_l = (1-p_l)^n,$$

which implies that

$$\mathbb{P}(\cap_{l=1}^{k}[\mathcal{X}_n \cap D_{s_1+\epsilon_1,s_2-\epsilon_1}(x_l) \neq \emptyset]) \geq 1 - \sum_{l=1}^{k}(1-p_l)^n \geq 1 - k(1-\alpha_0\bar{\Lambda})^n. \quad (3.8.20)$$

The statement follows by taking $n_1 \geq n_3$ such that $\eta/2 \geq k(1-\alpha_0\bar{\Lambda})^{n_1}$.
Step 4. Let $n^* = \max(n_1, n_2)$ and $n \geq n^*$. Notice that, by **Step 2**, $\{Y_{n,r,j}\}_{j=0}^{j^*} \in \tilde{\mathrm{G}}_{x,n,j^*,r^*}$ with probability at least $1-\eta$. Since, $r \leq r^* \leq r_1 \leq \xi$, by **Step 1** either (i) $\{Y_{n,r,j}\}_{j=0}^{j^*}$ remains in $(\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i)$ or (ii) $Y_{n,r^*,j} \in \overline{B}_{4\xi}(m_i)$ for some $j \in \{0,\ldots,j^*\}$. We show that (i) is not possible. Indeed, if $Y_{n,r,j^*} \in (\mathrm{G}_x)^{+\xi} \setminus B_\xi(m_i) \subset \tilde{K}_\xi$, then, by (3.8.9), there exists $X^*(Y_{n,r,j^*}) \in \mathcal{X}_{n,r}(Y_{n,r,j^*})$ such that $f_{\tau_n,n}(X^*(Y_{n,r,j^*})) > f_{\tau_n,n}(Y_{n,r,j^*})$. However, since j^* is the last iterate, by (3.6.1) it holds that $f_{\tau_n,n}(Y_{n,r,j^*}) \geq \max_{X\in\mathcal{X}_{n,r}(Y_{n,r,j^*})} f_{\tau_n,n}(X)$. Let $j_0 = \min\{j \in \{0,\ldots,j^*\} : Y_{n,r,j} \in \overline{B}_{4\xi}(m_i)\}$. By (3.8.1), $Y_{n,r,j_0} \in \mathrm{R}_\alpha \cap C(m_i)$. We show by induction that $Y_{n,r,j} \in \mathrm{R}_\alpha \cap C(m_i)$, for all $j_0 \leq j \leq j^*$. First, notice that, if $Y_{n,r,j} \in \mathrm{R}_{\tilde{\alpha}_4(\xi)} \cap C(m_i)$, then $Y_{n,r,j+1} \in B_{\xi+r}(m_i) \subset B_{2\xi}(m_i) \subset \mathrm{R}_\alpha \cap C(m_i)$. Second, if $Y_{n,r,j} \in B_{4\xi}(m_i) \setminus (\mathrm{R}_{\tilde{\alpha}_4(\xi)} \cap C(m_i))$, then by (3.8.3) $Y_{n,r,j} \in \tilde{K}_\xi$ and by (3.8.8) it holds that $f(Y_{n,r,j+1}) > f(Y_{n,r,j})$. Using the induction hypothesis, we conclude that $Y_{n,r,j+1} \in \mathrm{R}_\alpha \cap C(m_i)$. This completes the proof of consistency of the algorithm. ∎

Proof of Corollary 3.6.1. Let $\delta > 0$. We show that there exists $n^* \in \mathbb{N}$ and $\{\eta_n\}_{n=1}^{\infty}$ such that

$$\sum_{n=1}^{\infty}\mathbb{P}(\mathrm{J}_n \geq \delta) \leq n^* - 1 + \sum_{n=n^*}^{\infty}\mathbb{P}(\mathrm{J}_n = 1) \leq n^* - 1 + \sum_{n=n^*}^{\infty}\eta_n < \infty.$$

Then the result follows from Borel-Cantelli lemma. To this end, we explicitly express the constant η in Theorem 3.6.2 as a function of n and observe the convergence of the series. We first notice that, for $n \geq n_1$, we can choose $\eta_n/2 \geq k(1-\alpha_0\bar{\Lambda})^n$ in (3.8.20). Next, we apply in (3.8.13) Lemma 3.6.1 (iii) with $K = \tilde{K}_\xi$, $h^* = r$, and $\epsilon = d^* \min_{y\in\tilde{K}_\xi}\|\nabla f(y)\|_2$ and get constants $0 < \tilde{c} < \infty$ and $\tilde{n} \in \mathbb{N}$ such that, for all $n \geq \tilde{n}$,

$$\mathbb{P}(\sup_{h\in[h_n,r]}\sup_{v\in S^{d-1}}\sup_{x\in\tilde{K}_\xi}|\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f(x)| < d^*\min_{y\in\tilde{K}_\xi}\|\nabla f(y)\|_2) \leq 1 - \frac{\tilde{c}}{n^2}.$$

Therefore, for all $n \geq n^* = \max(n_1, \tilde{n})$, we can choose $\eta_n/2 = \max(k(1-\alpha_0\bar{\Lambda})^n, \tilde{c}/n^2)$, yielding $\sum_{n=n^*}^{\infty} \eta_n < \infty$. ∎

Proof of Proposition 3.6.1. We first prove (i). Since $P \in \mathcal{P}_{d,\ll\lambda}$ with density f, it holds that

$$\begin{aligned}
\mathbb{P}(\tilde{h}_n = 0) &= \mathbb{P}(\cup_{i=1}^n [\|X_i - x\|_2 = 0] \cup \cup_{i=1}^n \cup_{j=i+1}^n [\|X_i - X_j\|_2 = 0]) \\
&\leq \sum_{i=1}^n \mathbb{P}(\|X_i - x\|_2 = 0) + \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{P}(\|X_i - X_j\|_2 = 0) \\
&= n\mathbb{P}(\|X_1 - x\|_2 = 0) + \binom{n}{2} \int \mathbb{P}(\|X_1 - y\|_2 = 0) f(y) dy = 0.
\end{aligned}$$

For (ii), observe that, for all $\epsilon > 0$,

$$\mathbb{P}(\tilde{h}_n \geq \epsilon) \leq \mathbb{P}(\min_{i=1,\ldots,n} \|X_i - x\|_2 \geq \epsilon) = (\mathbb{P}(\|X_1 - x\|_2 \geq \epsilon))^n.$$

Since $x \in S_f$ and f is continuous, it holds that $\mathbb{P}(\|X_1 - x\|_2 \geq \epsilon) < 1$ and

$$\sum_{n=2}^{\infty} \mathbb{P}(\tilde{h}_n \geq \epsilon) \leq \sum_{n=2}^{\infty} \mathbb{P}(\|X_1 - x\|_2 \geq \epsilon)^n < \infty.$$

By Borel-Cantelli lemma, it follows that $\tilde{h}_n \xrightarrow{a.s.} 0$. We now prove (iii). To this end, let $M > 0$ and notice that $\mathbb{P}(n^{1-\delta}\tilde{h}_n^{2k} \leq M^{2k})$ is equal to

$$\mathbb{P}(\cup_{i=1}^n [\|X_i - x\|_2 \leq Mn^{-(1-\delta)/(2k)}] \cup \cup_{i=1}^n \cup_{j=i+1}^n [\|X_i - X_j\|_2 \leq Mn^{-(1-\delta)/(2k)}]),$$

which is bounded above by

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{P}(\|X_i - x\|_2 \leq Mn^{-(1-\delta)/(2k)}) + \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{P}(\|X_i - X_j\|_2 \leq Mn^{-(1-\delta)/(2k)}) \\
&= nP(B_{Mn^{-(1-\delta)/(2k)}}(x)) + \frac{n(n-1)}{2} \int P(B_{Mn^{-(1-\delta)/(2k)}}(y)) f(y) dy. \qquad (3.8.21)
\end{aligned}$$

Now, since f is bounded, we have that $\alpha = \sup_{y \in \mathbb{R}^d} f(y) < \infty$. For $y \in \mathbb{R}^d$, it holds that

$$P(B_{Mn^{-(1-\delta)/(2k)}}(y)) \leq \alpha\lambda(\overline{B}_{Mn^{-(1-\delta)/(2k)}}(x)) = \alpha c n^{-d(1-\delta)/(2k)}, \qquad (3.8.22)$$

where $c = M^d \pi^{d/2} / \Gamma(d/2 + 1)$. Using (3.8.22) in (3.8.21), we obtain that

$$\mathbb{P}(n^{1-\delta}\tilde{h}_n^{2k} \leq M^{2k}) \leq \alpha c n^{2-d(1-\delta)/(2k)}.$$

Therefore, using $d \geq 6k + 1$ and $0 < \delta < 1 - \frac{6k}{d}$, we have that

$$\sum_{n=1}^{\infty} \mathbb{P}(n^{1-\delta}\tilde{h}_n^{2k} \leq M^{2k}) \leq \alpha c \sum_{n=1}^{\infty} n^{2-d(1-\delta)/(2k)} < \infty.$$

By another application of Borel-Cantelli lemma and Theorem 5.2 in Billingsley (2012), we conclude that $n^{1-\delta}\tilde{h}_n^{2k} \xrightarrow{a.s.} \infty$. ∎

The next result shows that, the normalized gradient of f_τ converges uniformly to the normalized gradient of f in a compact set not containing the stationary points of f.

Proposition 3.8.1 *Suppose that f is continuously differentiable and (2.2.1) holds true. Let $K \subset S_f$ be a compact set with $K \cap \mathrm{N}_f = \emptyset$. Then,*

$$\lim_{\tau \to 0^+} \sup_{x \in K} \|\mathrm{w}_{f_\tau}(x) - \mathrm{w}_f(x)\|_2 = 0. \tag{3.8.23}$$

Proof of Proposition 3.8.1. We use Proposition 2.3.2 (iv), which shows that, as $\tau \to 0^+$, ∇f_τ converges uniformly in K to ∇f. Since $K \cap \mathrm{N}_f = \emptyset$, there exists τ^* such that $\min_{x \in K} \|\nabla f_\tau(x)\|_2 \geq c/2$ for all $0 < \tau \leq \tau^*$, where $c = \min_{x \in K} \|\nabla f(x)\|_2$. Then, using triangle inequality, we see that

$$\sup_{x \in K} \|\mathrm{w}_{f_\tau}(x) - \mathrm{w}_f(x)\|_2 \leq 4/c \sup_{x \in K} \|\nabla f_\tau(x) - \nabla f(x)\|_2,$$

which gives (3.8.23). ∎

3.9 Clustering Algorithm

In this section, we provide a detailed description of the algorithm for clustering. As a first step, starting from a point $x \in \mathbb{R}^d$, we search, in a given neighborhood of x, for the point y that yields the largest directional derivative $\nabla_v^h f_{\tau,n}$ with $h = \|y - x\|_2$ and $v = (y - x)/\|y - x\|_2$. Since

$$(\tau^{kd}\Lambda_1)^{1/k} \nabla_v^h f_\tau(x) = \frac{(L_G(x + hv, \tau))^{1/k} - (L_G(x, \tau))^{1/k}}{h} \quad \text{and}$$
$$(\tau^{kd}\Lambda_1)^{1/k} \nabla_v^h f_{\tau,n}(x) = \frac{(L_{G,n}(x + hv, \tau))^{1/k} - (L_{G,n}(x, \tau))^{1/k}}{h},$$

the constant $(\tau^{kd}\Lambda_1)^{1/k}$ does not influence the choice of the point y which maximizes both finite differences $\nabla_v^h f_\tau(x)$ and $\nabla_v^h f_{\tau,n}(x)$. This allows one to ignore the constant in the specification of the algorithm. That is, the finite difference approximation of the directional derivative of the k^{th} root of the local depth can be computed *avoiding the computation of the constant* Λ_1. We show, in fact, that the constant $(\tau^{kd}\Lambda_1)^{1/k}$ also does not influence the clusters induced by the system (3.5.1). Since $\tau, \Lambda_1 > 0$, if, for $x \in \mathbb{R}^d$, $u_{x,\tau} : \mathbb{R} \to \mathbb{R}^d$ is a solution of the system (3.5.1) with $u_{x,\tau}(0) = x$, then $\tilde{u}_{x,\tau} : \mathbb{R} \to \mathbb{R}^d$ given by $\tilde{u}_{x,\tau}(t) = u_{x,\tau}((\tau^{kd}\Lambda_1)^{1/k} t)$ also satisfies $\tilde{u}_{x,\tau}(0) = x$ and it is a solution of the system

$$\tilde{u}'(t) = \nabla \left((L_G(\tilde{u}(t), \tau))^{1/k} \right). \tag{3.9.1}$$

Moreover, since $\lim_{t \to \infty} u_{x,\tau}(t) = \lim_{t \to \infty} \tilde{u}_{x,\tau}(t)$ for all $x \in \mathbb{R}^d$, the clusters induced by (3.5.1) and (3.9.1) are the same. Hence, for $x, y \in \mathbb{R}^d$ with $y \neq x$ and $h = \|y - x\|_2 \leq r$ small enough, we consider the finite difference approximation of the directional derivatives of $(L_G(x, \tau))^{1/k}$ and $(L_{G,n}(x, \tau))^{1/k}$ along the direction

$v = \frac{y-x}{\|y-x\|_2}$ given by

$$\mathrm{d}_\tau(x;y) = \frac{(L_G(y,\tau))^{1/k} - (L_G(x,\tau))^{1/k}}{\|y-x\|_2} \quad \text{and} \tag{3.9.2}$$

$$\mathrm{d}_{\tau,n}(x;y) = \frac{(L_{G,n}(y,\tau))^{1/k} - (L_{G,n}(x,\tau))^{1/k}}{\|y-x\|_2}. \tag{3.9.3}$$

We now summarize the procedure for computing the clusters in Algorithm 1.

Algorithm 1: Clustering with local depth

Input: $\{x_1,\dots,x_n\}$, $\{y_1,\dots,y_o\}$ (optional), τ, s, r
Output: Local maxima for input points: $\{z_1,\dots,z_{n+o}\}$
1 Compute the local depth of $\{x_1,\dots,x_n\}$ with localization parameter τ
2 Store $\{x_1,\dots,x_n\}$, $\{y_1,\dots,y_o\}$ in new variables
for $i = 1$ *to* n **do**
  $z_i^* = x_i$
end
for $i = 1$ *to* o **do**
  $z_{i+n}^* = y_i$
end
3 For all points, compute the corresponding local maxima
for $i = 1$ *to* $n + o$ **do**
  repeat
4   $z_i = z_i^*$
5   Store the data points (different from z_i) at distance from z_i smaller than r or the s closest data points if they are less than s in new variables $w_1,\dots,w_l$ $(l \geq s)$
6   $z_i^* = \mathrm{argmax}_{j=1,\dots,l}\, \mathrm{d}_{\tau,n}(z_i;w_j)$
  until $L_{G,n}(z_i^*,\tau) < L_{G,n}(z_i,\tau)$
end

The algorithm requires as input, data points $\{x_1,\dots,x_n\}$, the localizing parameter τ, and two additional parameters, r and s. Additional points $\{y_1,\dots,y_o\}$ may also be provided as input. Starting from any point $x \in \{x_1,\dots,x_n\} \cup \{y_1,\dots,y_o\}$, based on the finite difference (3.9.3), the algorithm moves to another data point $y \in \{x_1,\dots,x_n\}$ (hence, except for the initial step, only data points are involved in (3.9.3)). The parameter r gives a bound on the norm $\|y-x\|_2$ in (3.9.3) in order to choose only those points that are close to each other. The parameter s, representing the minimal number of directions at each step of the algorithm, is exploited to ensure that the number of directional derivatives taken into account is not too small. Based on these choices, the steps 5, 6 and 7 of Algorithm 1 are repeated until the local maximum is achieved. The resulting data points are returned as output.

We now turn to the choice of the parameters r, s, and τ. We notice that for a good approximation to the directional derivative, the parameter r cannot be too large. Several exploratory analyses show that, under this condition, the parameter r does not significantly influence the output of Algorithm 1. Hence, we fix $r = 0.05$ in all our numerical work.

Turning to s, it is a good idea to consider a large number of various directions. The parameter s ensures that a sufficient number of directions are evaluated to get close to the maximum (over $v \in S^{d-1}$) of the directional derivative. This is particularly important in regions where data are sparse. The quantity s can also play the role of a smoothing parameter. If τ is small with a small sample size n, then the sample local depth can be noisy and have local peaks with a small basin of attraction that were not present in the original distribution. In this case, the choice of a larger s helps to avoid these local maxima.

We now describe a general method for the choice of s. Let $w(x) = \nabla f(x)/\|\nabla f(x)\|_2$ and V_i, $i = 1, \dots, s$ be independent according to the uniform distribution P_V on the unit sphere S^{d-1}. We take uniform distribution on S^{d-1} because directions of the s data points close to x are, in general, unknown. Then, for a given precision $\epsilon \in (0,1)$, with probability at least $1-\eta$ ($\eta \in (0,1)$), we require that

$$P_V^s\big(\min_{i=1,\dots,s} \|V_i - w(x)\|_2 \le \epsilon\big) \ge 1 - \eta.$$

Using the independence of V_is and due to the uniformity on S^{d-1}, we see that this is equivalent to

$$(1 - P_V(\|V_1 - e_d\|_2 \le \epsilon))^s \le \eta,$$

where $e_d = (0, \dots, 0, 1)^\top \in \mathbb{R}^d$. Therefore, s can be taken to be the smallest integer greater than or equal to

$$g_d(\eta, \epsilon) = \log_{1 - t_d(\epsilon)}(\eta),$$

where $t_d(\epsilon) = P_V(\|V_1 - e_d\|_2 \le \epsilon)$. Next, we compute the quantity $t_d(\epsilon)$. For $d = 1$, P_V is the Rademacher distribution yielding $t_d(\epsilon) = 1/2$. For $d \ge 2$, $t_d(\epsilon)$ is the probability (i.e. the area) of the hyperspherical cap $C_{1,\epsilon} = S^{d-1} \cap \overline{B}_\epsilon(e_d)$. Li (2011) shows that this is given by

$$t_d(\epsilon) = \frac{1}{2} \mathrm{I}_{r^2(\epsilon)}\left(\frac{d-1}{2}, \frac{1}{2}\right),$$

where $\mathrm{I}_z(\alpha, \beta)$ is the cumulative distribution function of a beta probability distribution with parameters $\alpha, \beta > 0$ and $r(\epsilon)$ is the radius of the hyperspherical cap. By Pythagoras theorem,

$$r^2(\epsilon) = 1^2 - (1 - h(\epsilon))^2 = 2h(\epsilon) - h^2(\epsilon),$$

where $h(\epsilon)$ is the height of the hyperspherical cap. To compute $r^2(\epsilon)$, we first compute $h(\epsilon)$. Since every point $x \in C_{1,\epsilon}$ satisfies $\langle x, e_d \rangle = 1 - \epsilon^2/2$, we conclude that $h(\epsilon) = 1 - \langle x, e_d \rangle = \epsilon^2/2$ and $r^2(\epsilon) = \epsilon^2 - \epsilon^4/2$. For $d = 1$, by choosing $\eta = 0.05$ and any $\epsilon \in (0,1)$, the above procedure yields $s = 5$. For $d = 2$, $\eta = 0.05$, and $\epsilon = 0.3$ (thus $h(\epsilon) = 0.045$), one obtains $s = 30$. Similarly, if $d = 5$, $\eta = 0.05$, and $\epsilon = 0.7$ (thus $h(\epsilon) = 0.245$), then $s = 71$. We notice that, for fixed η and ϵ, $g_d(\eta, \epsilon)$ is increasing in d as $\mathrm{I}_{r^2(\epsilon)}(\frac{d-1}{2}, \frac{1}{2})$ is decreasing in d. This implies that a larger sample size is required to obtain the same precision in higher dimensions.

We now turn to the parameter τ. Convergence of the clustering algorithm (cf. Theorem 3.6.2) requires that $\lim_{n\to\infty} n\tau_n^{2kd} = \infty$. Thus, we can take $\tau_n = n^{(-1+\delta)/(2kd)}$, for some $\delta > 0$. More specific choices are possible for some depth functions such as β-skeleton depth ($G = K_\beta$), lens depth ($G = K_2 = L$), spherical depth ($G = K_1 = B$), and simplicial depth ($G = S$). For these depth functions, τ is chosen as a quantile of order $q \in [0,1]$ based on distances between data points $x_1, \dots, x_n$. Specifically, for β-skeleton depth, q is chosen as a quantile of the empirical distribution of the $\binom{n}{2}$ distances $\|x_i - x_j\|_2$, $i, j \in \{1, 2, \dots, n\}$. For simplicial depth, q is a quantile of the $\binom{n}{d+1}$ maxima of the form $\max_{i,j=1,\ldots,d+1}\|x_{i_i} - x_{i_j}\|_2$ for all $\binom{n}{d+1}$ combinations of indices $i_1, \ldots, i_{d+1}$ from $\{1, 2, \ldots, n\}$.

We now turn to the computational complexity of β-skeleton and simplicial depth. To this end, we recall that $L_{K_\beta,n}$ is a U-statistics of order 2, while $L_{S,n}$ is a U-statistics of order $d+1$. This means that the computational complexity of $L_{K_\beta,n}$ is of order $O(\binom{n}{2})$, while the computational complexity of the $L_{S,n}$ is of order $O(\binom{n}{d+1})$, which makes a significant difference, especially in high dimensions. For large d and n, an approximation to L_S can be made by considering a large number of simplices sampled with replacement amongst all the $\binom{n}{d+1}$ simplices that define $L_{S,n}$; in our simulations (see Sections 3.11 and 3.12) we sample 10^8 simplices to reduce the computational cost.

3.10 Illustrative examples

In this section, we illustrate with some examples the role of the localizing parameter in the τ-approximation and the clustering methodology proposed in this chapter. We begin with a one-dimensional example showing the flexibility of the τ-approximation for different values of τ. As described in Chapter 2, for small values of τ, f_τ "resembles" the underlying density, while for larger τ it becomes unimodal, as depth functions are decreasing from the median of the distribution. We take this univariate distribution to be a mixture of four normal distributions with means -2, $0, 3, 4$, standard deviations $0.5, 0.8, 0.5, 0.2$ and weights $0.25, 0.5, 0.15$ and 0.1, respectively. The resulting density is quadrimodal and is depicted in Figure 3.1 along with its sample τ-approximation for $\tau = 0.5, 1, 2, 4$. For reproducibility, we use the seed `1234` for all figures that are based on one-sample and appearing in this section and Sections 3.11 and 3.12 below. As can be seen from Figure 3.1 for $\tau = 0.5$ the approximation has a similar shape to the density with approximately the same number of modes. For $\tau = 1$, the clusters corresponding to the modes at $x = 3$ and $x = 4$ merge yielding only three clusters. As we increase τ from 1 to 2, we notice that one can still identify two clusters, while, for $\tau = 4$, the τ-approximation has a unimodal shape.

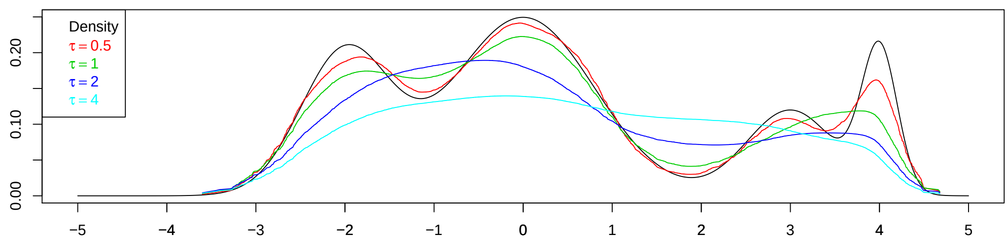

FIGURE 3.1: In black the quadrimodal mixture density and in red, green, blue and cyan its sample τ-approximation $f_{\tau,n}$ for $\tau = 0.5, 1, 2, 4$, respectively, and $n = 6000$.

Turning to two-dimensional examples studied in the literature (see Chacón (2015)), we consider mixtures of bivariate normal distributions with the following characteristics: a two-mixture with equal weights (Bimodal) and identity covariance matrix and the mixtures investigated in Wand and Jones (1993) and Chacón (2009) referred to as (H) Bimodal IV, (K) Trimodal III and #10 Fountain. Their analytical expression

is given below.
(i) The Bimodal density is a two-mixture of normal distributions with equal weights, identity covariance matrix and means $(-2,0)^\top$ and $(2,0)^\top$.
(ii) The (H) Bimodal IV density is a mixture of two normal distributions with equal weights, means $\mu_1 = (1,-1)^\top$, $\mu_2 = (-1,1)^\top$ and covariances

$$\Sigma_1 = \frac{4}{9}\begin{pmatrix} 1 & \frac{7}{10} \\ \frac{7}{10} & 1 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \frac{4}{9}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(iii) The (K) Trimodal III density is a mixture of three normal distributions with weights $w_1 = w_2 = \frac{3}{7}$ and $w_3 = \frac{1}{7}$; means $\mu_1 = (-1,0)^\top$, $\mu_2 = (1, 2\cdot\frac{\sqrt{3}}{3})^\top$ and $\mu_3 = (1,-2\cdot\frac{\sqrt{3}}{3})^\top$; and covariances

$$\Sigma_1 = \begin{pmatrix} \frac{9}{25} & \frac{7}{10}\cdot\frac{9}{25} \\ \frac{7}{10}\cdot\frac{9}{25} & \frac{49}{100} \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \Sigma_3 = \begin{pmatrix} \frac{9}{25} & 0 \\ 0 & \frac{49}{100} \end{pmatrix}.$$

(iv) The #10 Fountain density is a mixture of six normal distributions with weights $w_1 = \frac{1}{2}$ and $w_2 = w_3 = w_4 = w_5 = w_6 = \frac{1}{10}$; means $\mu_1 = \mu_2 = (0,0)^\top$, $\mu_3 = (-1,1)^\top$, $\mu_4 = (-1,-1)^\top$, $\mu_5 = (1,-1)^\top$ and $\mu_6 = (1,1)^\top$; and covariances

$$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \begin{pmatrix} \frac{1}{16} & 0 \\ 0 & \frac{1}{16} \end{pmatrix}.$$

The true clusters corresponding to these densities are in Figure 3.2 (first row). They are constructed as follows. First, we compute the gradient of the densities. Next, we use the R package `deSolve` to solve the (negative) gradient system

$$u'(t) = -\nabla f(u(t))$$

with initial value very close to the saddle points of f (see Chacón (2015)). In this way, we build the "borders" of the clusters (i.e. the curves in black). Finally, we plot modes in red and draw each cluster with a different color.

We apply our algorithm to analyze these models and identify clusters using local lens depth (LLD); these results are displayed in Figure 3.2 (second row). A comparison of our results with the clusters obtained using the kernel density estimator (KDE) are provided in Figure 3.2 (third row). Specifically, clusters are obtained via the kernel mean shift algorithm as implemented by the function `kms` in the R package `ks` (Duong, 2018). We set maximum number of iterations to 5000 and tolerance to 10^{-8}. The plug-in estimator of the bandwidth matrix is given by the function `Hpi` with pilot option `"dunconstr"` and derivatives of order one. The bandwidth matrix is obtained via minimization of the asymptotic mean integrated squared error (AMISE) of the gradient of the estimated density. For more details on the bandwidth matrix selection procedure see Sections 3.6 and 5.6.4 in Chacón and Duong (2018). For more details on the mean shift clustering algorithm see Section 6.2.2 of Chacón and Duong (2018). By a visual inspection of Figure 3.2, LLD performs a better clustering estimation than KDE. A more detailed analysis of the performance of LLD and KDE is provided in the next section.

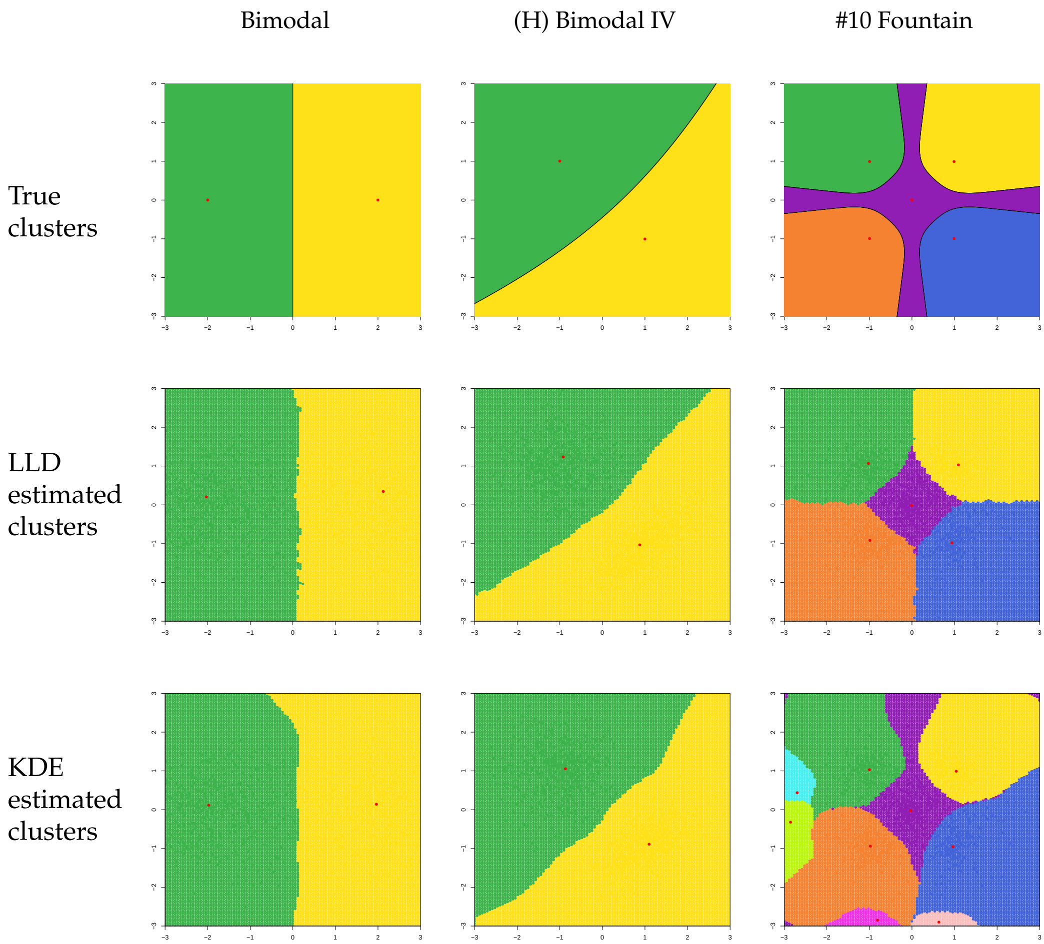

FIGURE 3.2: Clusters associated with the Bimodal (left), (H) Bimocal IV (middle) and #10 Fountain (right) densities. True clusters (first row). Local depth clustering based on $n = 1000$ samples from these densities and parameters $q = 0.05$, $s = 50$ and $r = 0.05$ (second row). Kernel density estimator clustering (third row). The true modes (first row) and the predicted modes (second and third rows) are plotted in red.

3.11 Numerical experiments

In this section, we compare the performance of clustering Algorithm 1 based on LDFs with kernel density estimator. We evaluate the performance in three different ways: (i) true number of clusters identified by the algorithm, (ii) empirical Hausdorff distance between the "true" cluster and the estimated cluster, and (iii) empirical probability distance (see Chacón (2015), for instance). We recall that the symmetric difference between two subsets A and B of $\mathbb{R}^d$ is $A\Delta B = ((\mathbb{R}^d \setminus A) \cap B) \cup (A \cap (\mathbb{R}^d \setminus B))$. Let $P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ be the empirical probability distribution of n i.i.d. samples, $X_1, \ldots, X_n$, from $P \in \mathcal{P}_d$. The empirical probability distance between the clusterings $\mathcal{C} = \{C_1, \ldots, C_l\}$ and $\mathcal{D} = \{D_1, \ldots, D_s\}$ with $l < s$ is given by

$$\hat{\mathrm{d}}_{P_n,\eta}(\mathcal{C}, \mathcal{D}) = \frac{1}{2} \min_{(i_1,\ldots,i_s)\in \mathrm{P}_s} \left(\sum_{j=1}^{l} P_n(C_j \Delta D_{i_j}]) + \eta \sum_{j=l+1}^{s} P_n(D_{i_j}) \right),$$

where P_s is the set of all permutations of $(1, \ldots, s)$ and $\eta \geq 0$ is a penalization coefficient for clusters that do not match with any other. If $l = s$ the second term in the above expression is zero. In our numerical experiments, we choose $\eta = 1$. Next, the empirical Hausdorff distance is given by

$$\tilde{\mathrm{d}}_{H,P_n}(\mathcal{C}, \mathcal{D}) = \max\left(\max_{i\in\{1,\ldots,t\}} \min_{j\in\{1,\ldots,s\}} P_n(C_i \Delta D_j), \max_{j\in\{1,\ldots,s\}} \min_{i\in\{1,\ldots,t\}} P_n(C_i \Delta D_j) \right).$$

In numerical experiments and data analysis, $\mathcal{C}$ is taken to be the set of true clusters while $\mathcal{D}$ is the set of estimated clusters, produced by the algorithm. If the estimated clusters coincide with the true clusters, then both these distances, *viz.* the *clustering errors*, are zero. Thus, small values of these distances suggest a good performance. As explained before, we consider the following distributions commonly used in the literature: Bimodal, (H) Bimodal IV, (K) Trimodal III and #10 Fountain. To test the performance of our methodology in higher dimensions, we also consider a bimodal and a quadrimodal density in dimension five. We refer to these distributions as Mult. Bimodal and Mult. Quadrimodal. Specifically, the Mult. Bimodal and Mult. Quadrimodal densities are obtained as mixtures of normal densities with identity covariance matrix and equal weights. In particular, the Mult. Bimodal density is a mixture of two normal distributions with means $(-2, 0, 0, 0, 0)^\top$ and $(2, 0, 0, 0, 0)^\top$ and the Mult. Quadrimodal density is a mixture of four normal distributions with means $(-2, 2, 0, 0, 0)^\top$, $(-2, -2, 0, 0, 0)^\top$, $(2, -2, 0, 0, 0)^\top$ and $(2, 2, 0, 0, 0)^\top$. The true clusters for the Mult. Bimodal density can be deduced from those of the Bimodal density. Similarly, the true clusters for the Mult. Quadrimodal density are deduced from those of the two-dimensional density given by a mixture of four normal distributions with means $(-2, 2)^\top$, $(-2, -2)^\top$, $(2, -2)^\top$ and $(2, 2)^\top$, and again equal weights and identity covariance matrix.

Our simulation results are based on a sample size of 1000 and 100 numerical experiments and we choose τ so that the corresponding quantiles q are given by 0.01, 0.05 and 0.1 (see Section 3.9). We compare the results of Algorithm 1 based on local lens depth (LLD) and local simplicial depth (LSD), with hierarchical clustering (Hclust) and Kernel density estimator (KDE). The hierarchical clustering requires a pre-specification of the number of clusters while the other methods do not, and it is reported here since it is one of the widely used methods for clustering. Thus, we compute it making use of the true number of clusters, which implies that the obtained results are not comparable with those of the other methodologies. Specifically, we use the R function `hclust` based on the L^2-distance between the observations and the default complete linkage method, i.e. the clusters distance is the maximum distance between the points in each cluster. Next, we apply the function `cutree`, based on the true number of clusters, to the output of `hclust`, yielding the final clusters. For more details about the numerical implementation and the quantiles for LLD and LSD we refer to Section 3.9.

Tables 3.1 and 3.2 provides clustering errors based on the Hausdorff distance and the probability distance. The best results are highlighted in bold. From these results we see that clustering errors based based on the KDE, LLD, LSD, and Hclust are similar for the distributions (H) Bimodal IV, (K) Trimodal III, #10 Fountain, and Bimodal. However, LLD outperforms all the competitors for the distribution in dimension five as can be seen from the columns Mult. Bimodal and Mult. Quadrimodal. Table 3.3 provides a comparison of the number of times the correct number of clusters is detected. The number of times the procedure identifies a lower number of clusters (on the left) and a higher number of clusters (on the right) is also provided. Again we notice that the proposed methods perform as well as the competitors. It is possible to improve the performance of LSD for distributions in dimension 5, by choosing smaller values of q, as described in Section 3.12.

Clustering errors (Hausdorff distance)			
	(H) Bimodal IV	(K) Trimodal III	#10 Fountain
KDE [a]	**0.00 (0.03)**	0.10 (0.15)	0.08 (0.05)
LLD [1]	0.05 (0.10)	**0.01 (0.15)**	**0.06 (0.01)**
LSD [2]	0.05 (0.11)	0.10 (0.15)	**0.06 (0.01)**
Hclust [*]	0.05 (0.09)	0.15 (0.09)	0.29 (0.05)
	Bimodal	Mult. Bimodal	Mult. Quadrimodal
KDE [a]	0.01 (0.05)	0.38 (0.17)	0.16 (0.08)
LLD [3]	0.01 (0.03)	**0.01 (0.04)**	**0.02 (0.01)**
LSD [4]	**0.00 (0.00)**	0.23 (0.18)	0.38 (0.18)
Hclust [*]	0.06 (0.05)	0.05 (0.03)	0.07 (0.03)

[a] `pilot="dunconstr"` [1] $q = 0.1$, $s = 30$. [2] $q = 0.01$, $s = 30$. [3] $q = 0.1$, $s = 50$. [4] $q = 0.05$, $s = 30$. [*] The true number of clusters is given in input.

TABLE 3.1: Mean of the clustering errors based on the Hausdorff distance for the densities (H) Bimodal IV, (K) Trimodal III, #10 Fountain, Bimodal, Mult. Bimodal and Mult. Quadrimodal. In parentheses the standard deviation. The true number of clusters is specified as input for the hierarchical clustering algorithm.

Clustering errors (distance in probability)			
	(H) Bimodal IV	(K) Trimodal III	#10 Fountain
KDE [a]	**0.01 (0.07)**	**0.06 (0.08)**	0.21 (0.31)
LLD [1]	0.13 (0.28)	**0.06 (0.07)**	**0.06 (0.01)**
LSD [2]	0.12 (0.27)	0.07 (0.09)	**0.06 (0.01)**
Hclust [*]	0.05 (0.09)	0.16 (0.09)	0.35 (0.07)
	Bimodal	Mult. Bimodal	Mult. Quadrimodal
KDE [a]	0.01 (0.04)	0.12 (0.13)	0.57 (0.33)
LLD [3]	0.01 (0.02)	**0.01 (0.01)**	**0.03 (0.01)**

LSD [4]	**0.00 (0.00)**	0.20 (0.17)	0.45 (0.17)
Hclust [*]	0.06 (0.05)	0.05 (0.03)	0.10 (0.04)

TABLE 3.2: Mean of the clustering errors based on the distance in probability for the densities (H) Bimodal IV, (K) Trimodal III, #10 Fountain, Bimodal, Mult. Bimodal and Mult. Quadrimodal. In parentheses the standard deviation. The true number of clusters is specified as input for the hierarchical clustering algorithm.

Number of times the true clusters are detected correctly			
	(H) Bimodal IV	(K) Trimodal III	#10 Fountain
KDE [a]	**(0) 99 (1)**	(15) 77 (8)	(0) 79 (21)
LLD [1]	(0) 83 (17)	**(14) 79 (7)**	**(0) 100 (0)**
LSD [2]	(0) 85 (15)	(13) 75 (12)	**(0) 100 (0)**
	Bimodal	Mult. Bimodal	Mult. Quadrimodal
KDE [a]	(0) 97 (3)	(0) 18 (82)	(0) 25 (75)
LLD [3]	(0) 99 (1)	**(0) 99 (1)**	**(0) 100 (0)**
LSD [4]	**(0) 100 (0)**	(12) 63 (25)	(77) 18 (5)

TABLE 3.3: Number of times that the procedure identifies the true number of clusters for the densities (H) Bimodal IV, (K) Trimodal III , #10 Fountain, Bimodal, Mult. Bimodal and Mult. Quadrimodal. In parentheses the number of times the procedure identifies a lower number of clusters (on the left) and a higher number of clusters (on the right).

3.12 Data analysis

In this section, we evaluate the performance of our methodology on two datasets taken from the UCI machine learning repository (http://archive.ics.uci.edu/ml/), namely, the Iris dataset and the Seeds dataset. For the sake of completeness we provide more details concerning the datasets. The Iris dataset consists of $n = 150$ observations from three classes (Iris Setosa, Iris Versicolour, and Iris Virginica) with four measurements each (sepal length, sepal width, petal length, and petal width). We compare our results to those based on KDE (with built-in bandwidth) and Hclust. Our algorithm, based on both lens and simplicial depth, correctly identifies all three clusters (see Table 3.4); furthermore, the Hausdorff distance and probability distance from our algorithm are smaller than those of the competitors.

Seeds dataset consists of $n = 210$ observations concerning three varieties of wheat; namely, Kama, Rosa, and Canadian. High quality visualization of the internal kernel structure was detected using a soft X-ray technique and seven geometric parameters of wheat kernels were recorded. They are area, perimeter, compactness, length of kernel, width of kernel, asymmetry coefficient, and length of kernel groove. All of these geometric parameters were continuous and real-valued. Table 3.5 contains the results of our analysis. The best results are highlighted in bold and correspond to LLD. We notice that both of our methods, LLD and LSD, correctly identify the true number of clusters.

It is worth mentioning here that Hclust was given as input the true number of clusters, three, as required by this methodology. However, the Hausdorff distance and probability distance of our proposed methods are smaller than those of Hclust. KDE, in both the examples, overestimates the true number of clusters.

Clustering errors for Iris data			
	Number of clusters	Distance in prob.	Hausdorff distance
KDE [a]	7	0.37	0.31
LLD [4]	**3**	**0.10**	**0.10**
LSD [5]	**3**	**0.10**	**0.10**
Hclust [*]		0.16	0.16

[5] $q = 10^{-4}$, $s = 20$.

TABLE 3.4: Mean of the clustering errors based on the Hausdorff distance and distance in probability for the Iris data. The true number of clusters is specified as input for the hierarchical clustering algorithm.

Clustering errors for Seeds data			
	Number of clusters	Distance in prob.	Hausdorff distance
KDE [a]	25	0.75	0.33
LLD [4]	**3**	**0.10**	**0.10**
LSD [6]	**3**	0.17	0.17
Hclust [*]		0.20	0.20

[6] $q = 10^{-5}$, $s = 20$.

TABLE 3.5: Mean of the clustering errors based on the Hausdorff distance and distance in probability for the Seeds data. The true number of clusters is specified as input for the hierarchical clustering algorithm.

Glossary of notation

A	a set
∂A, $\overline{A}$, and $\mathring{A}$	boundary, closure, and interior of A
$(A)^{+\delta}$ and $(A)^{-\delta}$	enlargement and negative enlargement of A by δ
$\mathcal{A}$	class of affine transformations
(m-)a.e. and (m-)a.s.	m-almost everywhere and m-almost surely
B	class of balls on $\mathbb{R}^d$
$B(P)$	(P-)Brownian bridge
$B_n(P)$	abbreviation for $\sqrt{n}(U_{1,n}(\cdot,P) - U_1(\cdot,P))$
$\mathcal{B}(E)$	Borel σ-algebra on E
$\mathcal{B}_d$	Borel σ-algebra on $\mathbb{R}^d$
$\overline{B}_r(x)$	closed ball in $\mathbb{R}^d$ with radius r and center x
$B_r(x)$	open ball in $\mathbb{R}^d$ with radius r and center x
$\gamma_{B(P)}$ and $\gamma_{\hat{W}(P)}$	covariance function of $B(P)$ and $\hat{W}(P)$
$\mathcal{C}, \mathcal{D}$	collection of sets or clustering
$\mathcal{C}_{\mathcal{H}}$	collection of subgraphs of functions in $\mathcal{H}$
C	closed convex cone in $\mathbb{R}^d$
$\#A$	cardinality of A
$g \circ h$	composition of functions g and h
$g * h$	convolution of functions g and h
∂_j	partial derivative w.r.t. j^{th} component
$\partial^{[n]}$	unidimensional n^{th} derivative
$d_{\Sigma(P)}$	Mahalanobis distance
$\tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,Q)}$	L^q-pseudodistance on $\mathscr{L}^q(\mathscr{H}_k, Q)$
$\tilde{d}^{(k,0)}_{\mathcal{H},q,\hat{P}^k_n}$	restriction of $\tilde{d}_{\mathscr{L}^q(\mathscr{H}_k,\hat{P}^k_n)}$ to $\mathcal{H}$
$\tilde{d}^{(1,k-1)}_{\mathcal{H},q,\hat{P}^k_n}$	pseudodistance on $\mathcal{H}$
$\tilde{d}^{(1,k-1)}_{\mathcal{H},q,P^k}$ and $\tilde{\tilde{d}}^{(1,k-1)}_{\mathcal{H},q,P^k}$	pseudodistances on $\mathcal{H}$
$\hat{\mathrm{d}}_{P,\eta}$ and $\tilde{\mathrm{d}}_{H,P}$	probability and Hausdorff distance between clusterings
$\mathscr{D}_{E^k}$	domain of $\mathcal{J}_k$
D	depth function
D_G	*Type A* depth function with index/kernel function G
D'_G	*Type A* depth function defined via $h'_{G,x,\infty}$
$\hat{D}_G$	*Type B* depth function with index G
$\tilde{D}_G$	*Type C* depth function with index G
$\tilde{D}_H$	halfspace depth
$\mathcal{D}(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$	class of *statistical* depth functions w.r.t. $(\mathcal{S}, \mathcal{P}_{d,1}, \mathcal{P}_{d,2})$
δ_x	Dirac measure at x
$\Delta[x_1, \ldots, x_{d+1}]$	closed simplex with vertices $x_1, \ldots, x_{d+1}$
det	determinant
$A \Delta B$	symmetric difference between A and B
$\mathrm{dist}(A, B)$	distance between A and B

E	(Hausdorff) topological space
$\mathbb{E}$	expectation
$\mathbb{E}^*$	outer expectation
$\{e_j\}_{j=1}^d$	standard basis of $\mathbb{R}^d$
$\mathfrak{f}(A)$	set of all real-valued functions from A
(F, d_F) and $(F, \tilde{d}_F)$	metric space and pseudometric space
F or F_P	cumulative distribution function for the probability measure P
f or f_P	density function for the probability measure P
f_τ or $f_{G,\tau}$	τ-approximation of f
$f_{\tau,n}$ or $f_{G,\tau,n}$	sample τ-approximation of f
∇ and ∇_v	gradient and directional derivative
∇_v^h	finite difference approximation of ∇_v
$\mathcal{G}$	class of kernel functions
G	(general) index or kernel function for *Type A* (local) depth functions
$\tilde{G}_\tau$	scaled kernel function G
$\mathscr{H}_k$	class of Borel measurable, symmetric functions $h : E^k \to \mathbb{R}$
$\mathcal{H}$	class of functions in $\mathscr{H}_k$
$\mathcal{H}_G$	class of functions $h_{G,x,\tau}$
h	kernel of a U-statistics
$h_{\mathcal{H}}$	envelope function of $\mathcal{H}$
$h_{G,x,\tau}$	function depending on kernel function G
$h'_{G,x,\infty}$	function depending on index G that is unbounded below
H_g	Hessian matrix of g
$H_{x,u}$	closed halfspace with outer normal u and boundary point x
$\langle \cdot, \cdot \rangle$	inner product on $\mathbb{R}^d$
$\mathbf{I}_A$	indicator function of A
$i_{G,x,\infty}$	function depending on index G
$\mathcal{J}_k(h, Q)$	integral of h w.r.t. $Q \in \mathcal{M}_\pm(E^k)$
$\mathcal{J}_k(h, Q)$	integral of h w.r.t. $Q \in \mathcal{M}_\pm(E^k)$
k (of k_G)	the order of a U-statistics (possibly depending on G)
λ	Lebesgue measure on $\mathbb{R}^d$
Λ_1 or $\Lambda_{G,1}$	Lebesgue integral of G over $(\mathbb{R}^d)^{k_G}$
$\ell^\infty(A) \subset \mathfrak{f}(A)$	set of all bounded functions from A
L_G and $L_{G,n}$	*Type A* and sample *Type A* local depth function
$L_{x,u}$	line through x with direction u
L_α and $\mathrm{L}_{\alpha,\tau}$	level sets of f and f_τ
$L^q((\mathbb{R}^d)^k, \lambda^k)$	space of functions $g : (\mathbb{R}^d)^k \to \mathbb{R}$ for which g^q is absolutely integrable
$L^q((\mathbb{R}^d)^k)$	abbreviation for $L^q((\mathbb{R}^d)^k, \lambda^k)$
$\mathscr{L}^q(\mathscr{H}_k, Q)$	class of symmetric functions with finite q^{th}-moment w.r.t. $Q \in \mathcal{P}(E^k)$
$l_{\mathcal{H}}$	constants bounding functions in $\mathcal{H}$
l or l_G	abbreviation for $l_{\mathcal{H}_G}$
M	matrix
$\mathcal{M}(E)$	set of all Borel measures on E
$\mathcal{M}_\pm(E)$	set of all finite signed Borel measures on E
m	measure
M_f	set of modes of f
$\mu(P)$	mean of P
$\mu_m^{(j)}(P)$	j^{th}-moment of P
$\mu_c^{(j)}(P)$	j^{th} central moment of P
$\\|\cdot\\|_q$	L^q-norm on $\mathbb{R}^d$

$\|\cdot\|_{\mathcal{M},q}$	L^q-matrix-norm
$\|\cdot\|_{\mathscr{L}^q(\mathscr{H}_k,Q)}$	L^q-(semi)norm on $\mathscr{L}^q(\mathscr{H}_k, Q)$
$\mathbb{N}$	set of all natural numbers
$N(\mu, \sigma^2)$	normal random variable with mean μ and variance σ^2
N_f	set of stationary points of f
$N(F, \tilde{d}_F, \epsilon)$	ϵ-covering number of $(F, \tilde{d}_F)$
O, o	big O, little o
O_G	function measuring outlyingness and depending on index G
$(\Omega, \Sigma, \mathbb{P})$	probability space
$(\Omega, \overline{\Sigma}, \overline{\mathbb{P}})$	completion of $(\Omega, \Sigma, \mathbb{P})$
P	probability measure
$P_n = \hat{P}_n^1$	empirical probability distribution
$\hat{P}_n^k$	empirical probability distribution for U-statistics of order k
P_T	push-forward measure of P w.r.t. T
P_u	abbreviation for P_{π_u}
$\mathbb{P}_X$	probability distribution of X
$\mathbb{P}^*(A)$	outer probability of A
$\mathcal{P}(E)$	set of all Borel probability measures on E
$\mathcal{P}_{fd}(E)$	set of all finitely discrete Borel probability measures on E
$\mathcal{P}_d$	set of all Borel probability measures on $\mathbb{R}^d$
$\mathcal{P}_{d,1}$, $\mathcal{P}_{d,2}$	(general) subclasses of $\mathcal{P}_d$
$\mathcal{P}_{d,\ll \mathrm{m}}$	subclass of absolutely continuous probabilities w.r.t. m
$\mathcal{P}_{d,A}$ and $\mathcal{P}_{d,H}$	subclasses of angularly and halfspace symmetric probabilities
$\mathcal{P}_{d,C}$ and $\mathcal{P}_{d,S}$	subclass of centrally and spherically symmetric probabilities
$\mathcal{P}_{d,c}$	subclass of continuous probabilities
$\mathcal{P}_{d,d}$ and $\mathcal{P}_{d,fd}$	subclass of discrete and finitely discrete probabilities
$\mathcal{P}_{d,hp}$	subclass of probabilities that assign probability zero to all hyperplanes
$\mathcal{P}_d^{(j)}$	subclass of probabilities with finite j^{th}-moment
$\mathcal{P}_d^{(2,i)}$ (resp. $\mathcal{P}_d^{(2,s)}$)	subclass of probabilities with invertible (resp. singular) covariance matrix
P_n	set of all permutations of $\{1, \dots, n\}$
ϕ_d	d-variate standard normal density
$\prod_{i=1}^n a_i$	n-fold product of sets or measures
a^n	power or n-fold product of identical sets or measures
$g^{\times n}$	n-fold product of identical functions g
$\otimes_{i=1}^n \Sigma_i$	n-fold product of σ-algebras
$\Sigma^{\otimes n}$	n-fold product of identical σ-algebras
π	transcendental number pi
$\pi_u(x)$	projection of x onto the direction u
$\pi_{e_j}(x)$	j^{th}-coordinate of x
$Q_p(P)$	quantile set of order p for P
$Q_p^-(P)$ and $Q_p^+(P)$	lower and upper quantile set of order p for P
$Q_{p,C}^-(P)$ and $Q_{p,C}^+(P)$	lower and upper C-quantile set of order p for P
$\mathcal{R}_m$	set of all partitions of $\{1, \dots, m\}$
$R_{D,\alpha}(P)$	(D-)depth quantile sets of P
R_α, $\mathrm{R}_{\alpha,\tau}$, and $\mathrm{R}_{\alpha,\tau,n}$	upper level sets of f, f_τ, and $f_{\tau,n}$
$\mathbb{R}$	set of all real numbers
r	radius or rank of a U-statistics
$\hat{\sigma}(M)$	spectrum of M
$\tilde{\sigma}(h)$	symmetrization of h
$\Sigma(P)$	covariance of P

S_g	support of a function or measure g
S^{d-1}	unit sphere in $\mathbb{R}^d$
$S_r(x)$ or $S_r^{d-1}(x)$	sphere in $\mathbb{R}^d$ with radius r and center x
$\mathcal{S}$	a class of invertible transformations in $\mathcal{T}$
sgn	sign function
$\mathcal{T}$	class of Borel measurable functions $T : \mathbb{R}^d \to \mathbb{R}^d$
$T_{A,\mu}$	function used for defining angular symmetry
$T_{C,\mu}$	function used for defining central symmetry
$T_{R,U,\mu}$ and $T_{S,U,\mu}$	functions used for defining spherical symmetry
τ	localizing parameter
$\mathcal{U}$	class of orthogonal transformations
$\mathrm{U}(x)$	neighborhood of x
$U_k(h,P)$	abbreviation for abbreviation for $\mathcal{J}_k(h,P^k)$
$U_{k,n}(h,P)$	U-statistics of order k and kernel h
$\mathbb{U}$	spherical uniform measure on $\overline{B}_1(0)$
$\mathbb{V}ar$	variance
$V(\mathcal{C})$	VC index of $\mathcal{C}$
$W_k(P)$	abbreviation for $kB(P)(\overline{p}_{k,1}\cdot)$
$W_{k,n}(P)$	abbreviation for $\sqrt{n}(U_{k,n}(\cdot,P) - U_k(\cdot,P))$
$\hat{W}(P)$	limit process for sample local depth
w_g	normalized gradient of g
X	random variable
$\xrightarrow{d}$ (resp. $\xrightarrow{d^*}$)	convergence in distribution (resp. distribution*)
$\xrightarrow{p}$ (resp. $\xrightarrow{p^*}$)	convergence in probability (resp. probability*)
$\xrightarrow{a.s.}$ (resp. $\xrightarrow{a.s.^*}$)	almost sure (resp. almost sure*) convergence

Appendix

A Measurability in the sense of Arcones and Giné (1993)

In this section, we show that the assumptions in Definition 1.11.2 yield the sufficient measurability conditions required by Arcones and Giné (1993) at page 1497, allowing for randomization by Rademacher and normal random variables and for the use of Fubini's theorem for several classes of functions constructed from $\mathcal{H} \subset L^1(\mathscr{H}_k, P^k)$, where $P \in \mathcal{P}(E)$. These classes include $p_{k,0}\mathcal{H}$, $(\overline{p}_{k,k} - p_{k,0})\mathcal{H}$, $p_{k,j}\mathcal{H}$, $\overline{p}_{k,j}\mathcal{H}$, and $(\overline{p}_{k,j} - p_{k,0})\mathcal{H}$, which are used in the proof of the uniform law of large numbers, as well as the corresponding "δ-classes" (see Proposition A.3 below), which play an important role in the uniform central limit theorem. Specifically, we show that, if $\mathcal{H}$ is image admissible Suslin, then all these classes and their composition via a measurable function are image admissible Suslin. Next, we show that this property yields measurability w.r.t. the completion of $(E^\infty, (\mathcal{B}(E))^{\otimes\infty}, P^\infty)$ of suprema over these classes. For this purpose, we make use of results from Dudley (2014). We begin by showing that if $\mathcal{H}$ is image admissible Suslin, then the classes $p_{k,0}\mathcal{H}$, $(\overline{p}_{k,k} - p_{k,0})\mathcal{H}$, $p_{k,j}\mathcal{H}$, $\overline{p}_{k,j}\mathcal{H}$, and $(\overline{p}_{k,j} - p_{k,0})\mathcal{H}$ are image admissible Suslin.

Proposition A.1 *Let $P \in \mathcal{P}(E)$. If $\mathcal{H} \subset L^1(\mathscr{H}_k, P^k)$ is image admissible Suslin, then the classes $p_{k,0}\mathcal{H}$, $(\overline{p}_{k,k} - p_{k,0})\mathcal{H}$, $p_{k,j}\mathcal{H}$, $\overline{p}_{k,j}\mathcal{H}$, and $(\overline{p}_{k,j} - p_{k,0})\mathcal{H}$ are image admissible Suslin for all $j = 1, \dots, k$.*

Proof of Proposition A.1. Since $\mathcal{H}$ is image admissible Suslin, there exists a Suslin measurable space $(Y, \mathcal{Y})$ and a surjective function $T : Y \to \mathcal{H}$ such that the function $\psi : E^k \times Y \to \mathbb{R}$ given by $\psi(x_1, \dots, x_k, y) = (T(y))(x_1, \dots, x_k)$ is measurable on $(E^k \times Y, (\mathcal{B}(E))^{\otimes k} \times \mathcal{Y})$. It is enough to show that the maps $(x_1, \dots, x_j, y) \mapsto \int (T(y))(x_1, \dots, x_k) dP(x_{j+1}) \dots dP(x_k)$ are measurable on $(E^j \times Y, (\mathcal{B}(E))^{\otimes j} \times \mathcal{Y})$ for all $j = 0, \dots, k$ (when $j = 0$ the corresponding term is missing). We suppose w.l.o.g. that ψ is non-negative (if not apply the same argument to ψ^+ and ψ^-). Let $\{\varphi_n\}_{n=1}^\infty$ be a sequence of simple functions[3] with $\varphi_n \uparrow \psi$ (see Theorem 2.10 of Folland (1999)). For $x_1, \dots, x_j \in E$ and $y \in Y$, let $\varphi_{n,x_1,\dots,x_j,y} : E^{k-j} \to \mathbb{R}$ and $\psi_{x_1,\dots,x_j,y} : E^{k-j} \to \mathbb{R}$ be given by $\varphi_{n,x_1,\dots,x_j,y}(x_{j+1}, \dots, x_k) = \varphi_n(x_1, \dots, x_k)$ and $\psi_{x_1,\dots,x_j,y}(x_{j+1}, \dots, x_k) = \psi(x_1, \dots, x_k)$. By Proposition 2.34 of Folland (1999), $\varphi_{n,x_1,\dots,x_j,y}$ and $\psi_{x_1,\dots,x_j,y}$ are measurable on $(E^{k-j}, (\mathcal{B}(E))^{\otimes(k-j)})$. Next, by using a monotone class argument (see for instance the proof of Theorem 2.36 of Folland (1999)), we see that the functions $(x_1, \dots, x_j, y) \mapsto \int \varphi_{n,x_1,\dots,x_j,y}(x_{j+1}, \dots, x_k, y) dP(x_{j+1}) \dots dP(x_k)$ are measurable on

[3] A function φ_n is called simple if $\varphi_n = \sum_{i=1}^{m_n} c_{n,i} \mathbf{I}_{A_{n,i}}$ for some $m_n \in \mathbb{N}$, $c_{n,i} \geq 0$, and $A_{n,i} \in (\mathcal{B}(E))^{\otimes k} \times \mathcal{Y}$

$(E^j \times Y, (\mathcal{B}(E))^{\otimes j} \times \mathcal{Y})$.[4]. It follows from the monotone convergence theorem (Theorem 2.14 of Folland (1999)) that the functions

$$\begin{aligned}(x_1,\ldots,x_j,y) \mapsto \int \psi_{x_1,\ldots,x_j,y}(x_{j+1},\ldots,x_k)dP(x_{j+1})\ldots dP(x_k) \\ = \int \psi_{x_1,\ldots,x_j,y}(x_{j+1},\ldots,x_k)dP(x_{j+1})\ldots dP(x_k) \\ = \lim_{n\to\infty} \int \varphi_{n,x_1,\ldots,x_j,y}(x_{j+1},\ldots,x_k)dP(x_1)\ldots dP(x_j) \\ = \lim_{n\to\infty} \int \varphi_{n,x_1,\ldots,x_j,y}(x_{j+1},\ldots,x_k)dP(x_1)\ldots dP(x_j)\end{aligned}$$

are measurable on $(E^j \times Y, (\mathcal{B}(E))^{\otimes j} \times \mathcal{Y})$. ■

The next proposition shows that the composition of image admissible Suslin classes of functions is image admissible Suslin (see Theorem 5.3.6 of Dudley (2014)).

Proposition A.2 *Let $\mathcal{H}_1,\ldots,\mathcal{H}_m$ be classes of functions from E^k to $\mathbb{R}$ and $\varphi : \mathbb{R}^m \to \mathbb{R}$ be Borel measurable. If $\mathcal{H}_i$, $i = 1,\ldots,m$, are image admissible Suslin, then $\varphi(\mathcal{H}_1,\ldots,\mathcal{H}_m) = \{\varphi(h_1,\ldots,h_m) : h_i \in \mathcal{H}\}$ is image admissible Suslin.*

Proof of Proposition A.2. Let $(Y_i, \mathcal{Y}_i)$ be Suslin measurable spaces and $T_i : Y_i \to \mathcal{H}_i$ be surjective functions such that $(x_1,\ldots,x_k,y_i) \mapsto (T_i(y_i))(x_1,\ldots,x_k)$ are measurable on $(E^k \times Y_i, (\mathcal{B}(E))^{\otimes k} \times \mathcal{Y}_i)$. Since $(Y_i, \mathcal{Y}_i)$ are Suslin measurable spaces, we have that (i) $\mathcal{Y}_i$ are generated by countable subclasses $\mathcal{Z}_i \subset \mathcal{Y}_i$ and $\{y_i\} \in \mathcal{Y}_i$ for all $y_i \in Y_i$ and (ii) there are Polish spaces X_i and measurable, surjective maps $S_i : X_i \to Y_i$. It follows that $\otimes_{i=1}^m \mathcal{Y}_i$ is generated by $\prod_{i=1}^m \mathcal{Z}_i$ and $\{(y_1,\ldots,y_m)\} \in \otimes_{i=1}^m \mathcal{Y}_i$ for all $(y_1,\ldots,y_m) \in \prod_{i=1}^m Y_i$. Since the product of separable spaces is separable and the product of completely metrizable spaces is completely metrizable via the product metric,[5] we obtain that $\prod_{i=1}^m X_i$ is a Polish space. Next, let $S : \prod_{i=1}^m X_i \to \prod_{i=1}^m Y_i$ be given by $S(z_1,\ldots,z_m) = (S_1(z_1),\ldots,S_m(z_m))$. Notice that S is surjective because S_i are surjective. Also, since S_i are measurable, we have that $\pi_i \circ S : \prod_{i=1}^m X_i \to Y_i$ are measurable, where $\pi_i : \prod_{i=1}^m X_i \to X_i$ is the projection into the i^{th}-component. Measurability of the components $\pi_i \circ S$ implies measurability of S (see Proposition 2.4 of Folland (1999)). We conclude that $(\prod_{i=1}^m Y_i, \otimes_{i=1}^m \mathcal{Y}_i)$ is a Suslin measurable space. Next, notice that $T : \prod_{i=1}^m Y_i \to \varphi(\mathcal{H}_1,\ldots,\mathcal{H}_m)$ given by $T(y_1,\ldots,y_m) = \varphi(T_1(y_1),\ldots,T_m(y_m))$ is surjective because T_i are surjective. To conclude the proof we need to show that the map $(x_1,\ldots,x_k,y_1,\ldots,y_m) \mapsto (T(y_1,\ldots,y_m))(x_1,\ldots,x_k)$ is measurable on $(E^k \times \prod_{i=1}^m Y_i, (\mathcal{B}(E))^{\otimes k} \times \otimes_{i=1}^m \mathcal{Y}_i)$. This follows from the measurability of φ, the measurability of the components $(x_1,\ldots,x_k,y_1,\ldots,y_m) \mapsto (T_i(y_i))(x_1,\ldots,x_k)$ on $(E^k \times \prod_{i=1}^m Y_i, (\mathcal{B}(E))^{\otimes k} \times \otimes_{i=1}^m \mathcal{Y}_i)$, and Proposition 2.4. ■

We show next that, if E is second countable (see [43]), then "δ-classes" are image admissible Suslin. We begin with two lemmas. The first lemma is about separability of classes of functions $\mathcal{F} \subset L^q(\mathscr{H}_j, P^j)$ endowed with a pseudodistance $\tilde{d} : L^q(\mathscr{H}_j, P^j) \times L^q(\mathscr{H}_j, P^j) \to [0,\infty)$ with $\tilde{d} \leq c d_{L^q(\mathscr{H}_j,P^j)}$ for some $c > 0$. In particular,

[4]To see this, choose in Theorem 2.36 $(X, \mathcal{M}, \mu)$ to be $(E^{k-j}, (\mathcal{B}(E))^{\otimes(k-j)}, P^{k-j})$, $(Y, \mathcal{N})$ to be $(E^j \times Y, (\mathcal{B}(E))^{\otimes j} \otimes \mathcal{Y})$, and E to be $A_{n,i} \in (\mathcal{B}(E))^{\otimes k} \times \mathcal{Y}$. Notice that we do not use the measure ν of Theorem 2.36 as we are interested in measurability w.r.t. one variable only.

[5]If $(X_1, d_{X_1}),\ldots,(X_m, d_{X_m})$ are complete metric spaces, then $\tilde{X}^m = \prod_{i=1}^m X_i$ endowed with the product metric $d_{\tilde{X}^m,q} : \tilde{X}^m \times \tilde{X}^m \to [0,\infty)$ given by $d_{\tilde{X}^m,q}((x_1,\ldots,x_m),(y_1,\ldots,y_m)) = \|(d_{X_1}(x_1,y_1),\ldots,d_{X_m}(x_m,y_m))^\top\|_q$ is a complete metric space.

this holds for the pseudodistance $\tilde{d}^{(1,j-1)}_{\mathcal{F},2,P^j}$ and $\tilde{\tilde{d}}^{(1,j-1)}_{\mathcal{F},2,P^j}$ used for continuity of the limiting Gaussian process in the uniform central limit theorem (see Definition 1.11) and for the asymptotic equicontinuity conditions in Theorems 1.11.5, 1.11.6, and Corollary 1.11.3. Indeed, by Jensen's inequality, it holds that $\tilde{d}^{(1,j-1)}_{\mathscr{H}_j,q,P^j} \leq 2\tilde{d}_{L^q(\mathscr{H}_j,P^j)}$[6] and $\tilde{\tilde{d}}^{(1,j-1)}_{\mathscr{H}_j,q,P^j} \leq \tilde{\tilde{d}}_{L^q(\mathscr{H}_j,P^j)}$.

Lemma A.1 *Suppose that E is second countable and let $\mathcal{F} \subset L^q(\mathscr{H}_j, P^j)$, where $j \in \mathbb{N}$, $q \geq 1$ and $P \in \mathcal{P}(E^j)$. Then, if for a pseudodistance $\tilde{d} : L^q(\mathscr{H}_j, P^j) \times L^q(\mathscr{H}_j, P^j) \to [0, \infty)$ and some $c > 0$, $\tilde{d} \leq c\tilde{d}_{L^q(\mathscr{H}_j,P^j)}$, then $(\mathcal{F}, \tilde{d}_\mathcal{F})$ is separable, where $\tilde{d}_\mathcal{F}$ is the restriction of $\tilde{d}$ on $\mathcal{F}$.*

Proof of Lemma A.1. Since E is second countable, E^j is second countable, and, there exists a collection of open sets $\{U_i\}_{i=1}^\infty$ in E^j such that every open set $V \subset E^j$ can be written as union of elements from $\{U_i\}_{i=1}^\infty$. Thus, $\{U_i\}_{i=1}^\infty$ generates all open sets and, therefore, it generates $\mathcal{B}(E^j)$. Next, we apply Theorem 4.13 of Brezis (2011) and obtain that the standard L^q-space $\{h : (E^j, \mathcal{B}(E^j)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) : J(|h|^q, P^j) < \infty\}$ endowed with the usual L^q-metric is separable. Using Proposition 3.25 of the same book, we have that the subspace $(L^q(\mathscr{H}_j, P^j), \tilde{d}_{L^q(\mathscr{H}_j,P^j)})$ of symmetric functions is also separable. Since $\tilde{d} \leq c\tilde{d}_{L^q(\mathscr{H}_j,P^j)}$, the topology induced by $\tilde{d}$ on $L^q(\mathscr{H}_j, P^j)$ is coarsest than the topology induced by $\tilde{d}_{L^q(\mathscr{H}_j,P^j)}$ on $L^q(\mathscr{H}_j, P^j)$.[7] Thus, by definition of separable space (see [43]), we obtain that $(L^q(\mathscr{H}_j, P^j), \tilde{d})$ is separable. Finally, we apply again Proposition 3.25 of Brezis (2011) and obtain that $(\mathcal{F}, \tilde{d}_\mathcal{F})$ is separable. ■

The second lemma is a version of Theorem 5.2.5 of Dudley (2014) and it is about measurability of closed subset of an image admissible class of functions.

Lemma A.2 *Suppose that E is second countable and let $\mathcal{F} \subset L^q(\mathscr{H}_j, P^j)$, where $j \in \mathbb{N}$, $q \geq 1$ and $P \in \mathcal{P}(E^j)$, be image admissible Suslin via $(Y, \mathcal{Y})$ and surjective map $T : Y \to \mathcal{F}$. If $C \subset \mathcal{F}$ is relatively closed in $\mathcal{F}$ w.r.t. $\tilde{d}_\mathcal{F}$, where $\tilde{d}_\mathcal{F}$ is either $\tilde{d}^{(1,j-1)}_{\mathcal{F},q,P^j}$ or $\tilde{\tilde{d}}^{(1,j-1)}_{\mathcal{F},q,P^j}$, then $T^{-1}(C) \in \mathcal{Y}$.*

Proof of Lemma A.2. Let $U = \mathcal{F} \setminus C$. Since $T^{-1}(C) = T^{-1}(\mathcal{F} \setminus U) = Y \setminus T^{-1}(U)$, it is enough to show that $T^{-1}(U) \in \mathcal{Y}$. By Lemma A.1 we have that $(\mathcal{F}, \tilde{d}_\mathcal{F})$ is separable. Using the proof of Proposition 2.1.4 of Dudley (2018) there are sequences $\{f_i\}_{i=1}^\infty$ and $\{r_i\}_{i=1}^\infty$, where $f_i \in \mathcal{F}$ and $0 < r_i < \infty$, such that

$$U = \cup_{i=1}^\infty \{f \in \mathcal{F} : \tilde{d}_\mathcal{F}(f, f_i) < r_i\}.$$

Therefore, it is enough to show that, for all $i \in \mathbb{N}$,

$$\{y \in Y : \tilde{d}_\mathcal{F}(T(y), f_i) < r_i\} = T^{-1}(\{f \in \mathcal{F} : \tilde{d}_\mathcal{F}(f, f_i) < r_i\}) \in \mathcal{Y}.$$

[6] Recall that $\tilde{d}^{(1,j-1)}_{\mathscr{H}_j,q,P^j}(h_1, h_2) = \|p_{j,1}h_1 - p_{j,1}h_2\|_{L^q(\mathscr{H}_j,P^j)}$ and $\tilde{d}_{L^q(\mathscr{H}_j,P^j)}(h_1, h_2) = \|h_1 - h_2\|_{L^q(\mathscr{H}_j,P^j)}$. Since $p_{j,1} = \overline{p}_{j,1} - \overline{p}_{j,0}$, we have that for all $h_1, h_2 \in \mathcal{F}$

$$\|p_{j,1}h_1 - p_{j,1}h_2\|_{L^q(\mathscr{H}_j,P^j)} \leq \|\overline{p}_{j,1}h_1 - \overline{p}_{j,1}h_2\|_{L^q(\mathscr{H}_j,P^j)} + \|\overline{p}_{j,0}h_1 - \overline{p}_{j,0}h_2\|_{L^q(\mathscr{H}_j,P^j)} \leq 2\|h_1 - h_2\|_{L^q(\mathscr{H}_j,P^j)}.$$

[7] U is open in $(L^q(\mathscr{H}_j, P^j), \tilde{d})$ if and only if for all $h \in U$ there exists $r > 0$ such that $\{f \in L^q(\mathscr{H}_j, P^j) : \tilde{d}(f, h) \leq r\} \subset U$. Using that $\tilde{d} \leq c\tilde{d}_{L^q(\mathscr{H}_j,P^j)}$, we have that $\{f \in L^q(\mathscr{H}_j, P^j) : \tilde{d}_{L^q(\mathscr{H}_j,P^j)}(f, h) \leq \frac{r}{c}\} \subset \{f \in L^q(\mathscr{H}_j, P^j) : \tilde{d}(h, f) \leq r\} \subset U$. Thus, U is open in $(L^q(\mathscr{H}_j, P^j), \tilde{d}_{L^q(\mathscr{H}_j,P^j)})$.

To this end, we show that the maps

$$y \mapsto \tilde{d}^{(1,j-1)}_{\mathcal{F},q,P^j}(T(y), f_i) = \|p_{j,1}(T(y)) - p_{j,1}f_i\|_{L^q(\mathscr{H}_1,P)} \text{ and}$$
$$y \mapsto \tilde{\overline{d}}^{(1,j-1)}_{\mathcal{F},q,P^j}(T(y), f_i) = \|\overline{p}_{j,1}(T(y)) - \overline{p}_{j,1}f_i\|_{L^q(\mathscr{H}_1,P)}$$

are measurable. We first notice that, by the proof of Proposition A.1 and Proposition 2.34 of Folland (1999), the map

$$(x_1, y) \mapsto (p_{j,1}(T(y)))(x_1) - (p_{j,1}f_i)(x_1) \text{ and}$$
$$(x_1, y) \mapsto (\overline{p}_{j,1}(T(y)))(x_1) - (\overline{p}_{j,1}f_i)(x_1)$$

are measurable on $(E \times Y, \mathcal{B}(E) \otimes \mathcal{Y})$. The result now follows from Fubini-Tonelli theorem (see Theorem 2.37 of Folland (1999)). ■

We are now ready to show that "δ-classes" are image admissible Suslin.

Proposition A.3 *Suppose that E is second countable and let $\mathcal{F} \subset L^q(\mathscr{H}_j, P^j)$, where $j \in \mathbb{N}$, $q \geq 1$ and $P \in \mathcal{P}(E^j)$, be image admissible Suslin. Then, for $0 < \delta \leq \infty$ and $\tilde{d}_{\mathcal{F}}$ equal to either $\tilde{d}^{(1,j-1)}_{\mathcal{F},q,P^j}$ or $\tilde{\overline{d}}^{(1,j-1)}_{\mathcal{F},q,P^j}$,*

$$\mathcal{F}_{\tilde{d}_{\mathcal{F}},\delta}(P) = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F} \text{ and } \tilde{d}_{\mathcal{F}}(f_1, f_2) \leq \delta\}.$$

is image admissible Suslin.

Proof of Proposition A.3. Using Proposition A.2 with $k = j$, $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{F}$, $m = 2$, and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ given by $\varphi(t_1, t_2) = t_1 - t_2$, we obtain that $\mathcal{F} - \mathcal{F} = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}\} \subset L^q(\mathscr{H}_j, P^j)$ is image admissible Suslin. Let $(Y, \mathcal{Y})$ be the corresponding Suslin measurable space and $T : Y \to \mathcal{F} - \mathcal{F}$ be surjective and measurable. Notice that

$$\mathcal{F}_{\tilde{d}_{\mathcal{F}},\delta}(P) = \{f \in \mathcal{F} - \mathcal{F} : \mathcal{F}_{\tilde{d}_{\mathcal{F}-\mathcal{F}},\delta}(P)(f, 0) \leq \delta\}$$

is the closed ball of radius δ in $(\mathcal{F} - \mathcal{F}, \tilde{d}_{\mathcal{F}-\mathcal{F}})$. Using Lemma A.2, we obtain that $Y_0 = T^{-1}(\mathcal{F}_{\tilde{d}_{\mathcal{F}},\delta}(P)) \in \mathcal{Y}$. It follows that the restriction of T to Y_0, namely, $T_0 : Y_0 \to \mathcal{F}_{\tilde{d}_{\mathcal{F}},\delta}(P)$ given by $T_0(y) = T(y)$, is surjective and measurable on $(Y_0, \mathcal{Y}_0)$, where $\mathcal{Y}_0 = \mathcal{Y} \cap Y_0$. The result follows if we show that $(Y_0, \mathcal{Y}_0)$ is a Suslin measurable space. To see this, we use that $(Y, \mathcal{Y})$ is a Suslin measurable space. Specifically, (i) $\mathcal{Y}$ is generated by a countable subclass $\mathcal{Z} \subset \mathcal{Y}$ and $\{y\} \in \mathcal{Y}$ for all $y \in Y$ and (ii) there is a Polish space X and a measurable, surjective map $S : (X, \mathcal{B}(X)) \to (Y, \mathcal{Y})$. It follows that $\{y\} \in \mathcal{Y}_0$ for all $y \in Y_0$ and $\mathcal{Y}_0$ is generated by $\mathcal{Z}_0 = \mathcal{Z} \cap Y_0$.[8] Next, let $X_0 = S^{-1}(Y_0) \in \mathcal{B}(X)$. Then, the restriction of S to X_0, namely, $S_0 : (X_0, \mathcal{B}(X_0)) \to (Y_0, \mathcal{Y}_0)$ given by $S_0(x) = S(x)$, is surjective and measurable. Finally, we notice that $X_0 \in \mathcal{B}(X)$ is itself a Polish space (see page 388 of Dudley (2018)). ■

The next proposition shows that the envelope of an image admissible class of functions is measurable in $(E^\infty, \overline{(\mathcal{B}(E))^{\otimes\infty}}, \overline{P^\infty})$, where, for $n \in \mathbb{N} \cup \{\infty\}$, $(E^n, \overline{(\mathcal{B}(E))^{\otimes n}}, \overline{P^n})$ is the completion of $(E^n, (\mathcal{B}(E))^{\otimes n}, P^n)$. In particular, it applies to the classes $\mathcal{H}$,

[8] To see this, it is enough to show that the smallest σ-algebra generated by $\mathcal{Z}_0$ contains $\mathcal{Y}_0$. Let $B \in \mathcal{Y}_0$. Then $B \in \mathcal{Y}$ can be obtained starting from elements of $\mathcal{Z}$ via countable union and complementing. Using that $B = B \cap Y_0$ and for arbitrary set $A_i \subset Y$ a) $(\cup_{i=1}^\infty A_i) \cap Y_0 = \cup_{i=1}^\infty (A_i \cap Y_0)$ and b) $(Y \setminus A_1) \cap Y_0 = Y_0 \setminus (A_1 \cap Y_0)$ we see that B can be obtained starting from elements of $\mathcal{Z}_0$.

$p_{k,0}\mathcal{H}$, $(\overline{p}_{k,k} - p_{k,0})\mathcal{H}$, $p_{k,j}\mathcal{H}$, $\overline{p}_{k,j}\mathcal{H}$, and $(\overline{p}_{k,j} - p_{k,0})\mathcal{H}$ of Proposition A.1 and to the classes obtained from these classes via composition as in Proposition A.2. If E is second countable, then it also applies to the corresponding "δ-classes".

Proposition A.4 *Let $P \in \mathcal{P}(E)$ and $\mathcal{F} \subset L^1(\mathscr{H}_j, P^j)$ be image admissible Suslin. Then, $\sup_{f\in\mathcal{F}} f$ and the envelope function $f_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|f|$ are measurable in $(E^j, \overline{(\mathcal{B}(E))^{\otimes j}}, \overline{P^j})$.*

Proof of Proposition A.4. There exists a Suslin measurable space $(Y, \mathcal{Y})$ and a surjective function $T : Y \to \mathcal{F}$ such that $(x_1, \dots, x_j, y) \mapsto (T(y))(x_1, \dots, x_j)$ is measurable on $(E^j \times Y, (\mathcal{B}(E))^{\otimes j} \otimes \mathcal{Y})$. The result is given by Corollary 5.3.5 of Dudley (2014) and the definition of universally measurable (u.m.) at page 186 (see also Example 1.7.5 of van der Vaart and Wellner (1996)). ∎

Remark A.1 *Notice that measurability in $(E^n, \overline{(\mathcal{B}(E))^{\otimes n}}, \overline{P^n})$ canonically yields measurability in $(E^\infty, \overline{(\mathcal{B}(E))^{\otimes\infty}}, \overline{P^\infty})$. That is, if $\psi : E^n \to \mathbb{R}$ is measurable in $(E^n, \overline{(\mathcal{B}(E))^{\otimes n}}, \overline{P^n})$, then $\tilde{\psi} : E^\infty \to \mathbb{R}$ given by $\tilde{\psi}(x_1, x_2, \dots) = \psi(x_1, \dots, x_n)$ is measurable in $(E^\infty, \overline{(\mathcal{B}(E))^{\otimes\infty}}, \overline{P^\infty})$. Indeed, using the notation $\infty - n$ to denote countable product from $n+1$ to ∞, we have that, for all $B \in \mathcal{B}(R)$,*

$$\tilde{\psi}^{-1}(B) = \psi^{-1}(B) \times E^{\infty-n} \in \overline{(\mathcal{B}(E))^{\otimes n}} \times E^{\infty-n} \subset {}^{9}\overline{(\mathcal{B}(E))^{\otimes\infty}}.$$

Remark A.2 *Let $P \in \mathcal{P}(E)$. By choosing $m = 1$ and $\mathcal{H}_1 = \mathcal{F} \subset L^1(\mathscr{H}_j, P^j)$ in Proposition A.2, where $\mathcal{F}$ is image admissible Suslin, we see that the class of functions*

$$\{(x_1, \dots, x_n) \mapsto \sum_{1\le i_1<\dots<i_j\le n} \varphi(f(x_{i_1}, \dots, x_{i_j})) : f \in \mathcal{F}\}$$

is image admissible Suslin. By Proposition A.4, the envelope function

$$(x_1, \dots, x_n) \mapsto \sup_{f\in\mathcal{F}} \left| \sum_{1\le i_1<\dots<i_j\le n} \varphi(f(x_{i_1}, \dots, x_{i_j})) : f \in \mathcal{F} \right|$$

is measurable in $(E^n, \overline{(\mathcal{B}(E))^{\otimes n}}, \overline{P^n})$. In particular, we can take $\varphi : \mathbb{R} \to \mathbb{R}$ of the form $\varphi(t) = a_1 \dots a_j t$, where $a_1, \dots, a_j \in \mathbb{R}$. Then, measurability in $(E^n, \overline{(\mathcal{B}(E))^{\otimes n}}, \overline{P^n})$ and Proposition A.5 below allow for randomization by Rademacher and normal random variables and the use of Fubini's theorem (cf. Arcones and Giné (1993) page 1497 and van der Vaart and Wellner (1996) pages 85 and 109-111).

When computing expectations and probabilities, measurability w.r.t. the completion is, in some sense, equivalent to measurability w.r.t. the original probability space. Specifically, Proposition A.5 shows that the outer expectation of a function that is completion measurable is equal to the expectation with respect to completion of the probability space. We begin by showing that the expectation of a random variable w.r.t. the completion is equal to the original expectation.

Lemma A.3 *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, $(\Omega, \overline{\Sigma}, \overline{\mathbb{P}})$ be its completion w.r.t. $\mathbb{P}$, and $Y : \Omega \to \mathbb{R}$ be a random variable on $(\Omega, \Sigma, \mathbb{P})$. Then, $\overline{\mathbb{E}}[Y] = \mathbb{E}[Y]$, where $\overline{\mathbb{E}}$ denotes expectation w.r.t. the completed space $(\Omega, \overline{\Sigma}, \overline{\mathbb{P}})$.*

[9] Let $A \cup N \in \overline{(\mathcal{B}(E))^{\otimes n}}$, where $A \in (\mathcal{B}(E))^{\otimes n}$, $N \subset N^* \in (\mathcal{B}(E))^{\otimes n}$, and $(P^n)^*(N) = P^n(N^*) = 0$. Then $(A \cup N) \times E^{\infty-n} \in \overline{(\mathcal{B}(E))^{\otimes\infty}}$. Indeed, $(A \cup N) \times E^{\infty-n} = A \times E^{\infty-n} \cup N \times E^{\infty-n}$, where $A \times E^{\infty-n} \in (\mathcal{B}(E))^{\otimes\infty}$ and $(P^\infty)^*(N \times E^{\infty-n}) \le P^\infty(N^* \times E^{\infty-n}) = 0$.

Proof of Lemma A.3. By definition of integral on a general probability space, $\overline{\mathbb{E}}[Y]$ is the supremum of integrals of the form $\overline{\mathbb{E}}[\overline{\varphi}]$, where $\overline{\varphi} \le Y$ is a simple function, that is, $\overline{\varphi} = \sum_{j=1}^{k} c_j \mathbf{I}_{\overline{A}_j}$ for some $k \ge 1$, $c_j \ge 0$, and $\overline{A}_j \in \overline{\Sigma}$. Notice that $\overline{A}_j = A_j \cup N_j$, for some $N_j \subset \Omega$ with $\overline{\mathbb{P}}(N_j) = 0$. This yields that $\overline{\mathbb{P}}(\overline{A}_j) = \mathbb{P}(A_j)$, which, in turn, implies that $\overline{\mathbb{E}}[\overline{\varphi}] = \mathbb{E}[\varphi]$, where $\varphi = \sum_{j=1}^{k} c_j \mathbf{I}_{A_j}$. By taking the supremum over all such $\overline{\varphi}$ and φ, we obtain that $\overline{\mathbb{E}}[Y] = \mathbb{E}[Y]$. ∎

Proposition A.5 *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, $(\Omega, \overline{\Sigma}, \overline{\mathbb{P}})$ be its completion w.r.t. $\mathbb{P}$, and $X : \Omega \to \mathbb{R}$ be a random variable on $(\Omega, \overline{\Sigma}, \overline{\mathbb{P}})$. Then, $\mathbb{E}^*[X] = \mathbb{E}[X^*] = \overline{\mathbb{E}}[X]$ and $X = X^*$ $\overline{\mathbb{P}}$-a.s.*

Proof of Proposition A.5. We apply Proposition 2.12 of Folland (1999) with g replaced by $g + \infty I_N$ (g and N are defined there) and obtain a random variable $Y : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ such that (i) Y is measurable on $(\Omega, \Sigma, \mathbb{P})$, (ii) $Y \ge X$, and (iii) $Y = X$ $\overline{\mathbb{P}}$-a.s. It follows from the definition of X^* that $Y \ge X^*$ and $Y = X^*$ $\overline{\mathbb{P}}$-a.s. Finally, we apply Lemma A.3 to the random variables Y and X^* and conclude that

$$\mathbb{E}[X^*] \le \mathbb{E}[Y] = \overline{\mathbb{E}}[Y] = \overline{\mathbb{E}}[X] \le \overline{\mathbb{E}}[X^*] = \mathbb{E}[X^*]$$

yielding that $\mathbb{E}[X^*] = \overline{\mathbb{E}}[X]$ and $X = X^*$ $\overline{\mathbb{P}}$-a.s. ∎

B Convergence of sets

In this section, we summarize with proofs various properties concerning the limits of sets.

Lemma B.1 *Let $\{A_n\}_{n=1}^{\infty}$, $\{B_n\}_{n=1}^{\infty}$ be sequences of sets in $\mathbb{R}^p$. Then, it holds that*

(i) $\liminf_{n\to\infty}(A_n \cap B_n) = (\liminf_{n\to\infty} A_n) \cap (\liminf_{n\to\infty} B_n)$,

(ii) $\limsup_{n\to\infty}(A_n \cap B_n) \subset (\limsup_{n\to\infty} A_n) \cap (\limsup_{n\to\infty} B_n)$,

(iii) $\liminf_{n\to\infty}(A_n \cup B_n) \supset (\liminf_{n\to\infty} A_n) \cup (\liminf_{n\to\infty} B_n)$, *and*

(iv) $\limsup_{n\to\infty}(A_n \cup B_n) = (\limsup_{n\to\infty} A_n) \cup (\limsup_{n\to\infty} B_n)$.

In particular, if $A = \lim_{n\to\infty} A_n$ and $B = \lim_{n\to\infty} B_n$ exist, then

(v) $\lim_{n\to\infty}(A_n \cap B_n) = A \cap B$ *and (vi)* $\lim_{n\to\infty}(A_n \cup B_n) = A \cup B$.

Proof of Lemma B.1. We begin by proving (i). It holds that

$$\begin{aligned}
x \in \liminf_{n\to\infty}(A_n \cap B_n) &\Leftrightarrow \exists n^* \in \mathbb{N} : x \in \cap_{n=n^*}^{\infty}(A_n \cap B_n) \\
&\Leftrightarrow \exists n^* \in \mathbb{N} : x \in (A_n \cap B_n)\ \forall n \ge n^* \\
&\Leftrightarrow \exists n_A, n_B \in \mathbb{N} : x \in A_n\ \forall n \ge n_A \text{ and } x \in B_n\ \forall n \ge n_B \\
&\Leftrightarrow \exists n_A, n_B \in \mathbb{N} : x \in \cap_{n=n_A}^{\infty} A_n \text{ and } x \in \cap_{n=n_B}^{\infty} B_n \\
&\Leftrightarrow x \in \liminf_{n\to\infty} A_n \text{ and } x \in \liminf_{n\to\infty} B_n \\
&\Leftrightarrow x \in (\liminf_{n\to\infty} A_n) \cap (\liminf_{n\to\infty} B_n).
\end{aligned}$$

For (ii), we have that

$$
\begin{aligned}
x \in \limsup_{n\to\infty}(A_n \cap B_n) &\Leftrightarrow \forall n^* \in \mathbb{N}\ x \in \cup_{n=n^*}^{\infty}(A_n \cap B_n) \\
&\Leftrightarrow \forall n^* \in \mathbb{N}\ \exists \tilde{n}^* \geq n^* : x \in (A_{\tilde{n}^*} \cap B_{\tilde{n}^*}) \\
&\Leftrightarrow \forall n^* \in \mathbb{N}\ \exists \tilde{n}^* \geq n^* : x \in A_{\tilde{n}^*} \text{ and } x \in B_{\tilde{n}^*} \\
&\Rightarrow \forall n_A, n_B \in \mathbb{N}\ \exists \tilde{n}_A \geq n_A \text{ and } \tilde{n}_B \geq n_B : x \in A_{\tilde{n}_A} \text{ and } x \in B_{\tilde{n}_B} \\
&\Leftrightarrow \forall n_A, n_B \in \mathbb{N}\ x \in \cup_{n=n_A}^{\infty} A_n \text{ and } x \in \cup_{n=n_B}^{\infty} B_n \\
&\Leftrightarrow x \in \limsup_{n\to\infty} A_n \text{ and } x \in \limsup_{n\to\infty} B_n \\
&\Leftrightarrow x \in (\limsup_{n\to\infty} A_n) \cap (\limsup_{n\to\infty} B_n).
\end{aligned}
$$

We now prove (iii). We have that

$$
\begin{aligned}
x \in \liminf_{n\to\infty}(A_n \cup B_n) &\Leftrightarrow \exists n^* \in \mathbb{N} : x \in \cap_{n=n^*}^{\infty}(A_n \cup B_n) \\
&\Leftrightarrow \exists n^* \in \mathbb{N} : x \in (A_n \cup B_n)\ \forall n \geq n^* \\
&\Leftarrow \exists n_A \in \mathbb{N} : x \in A_n\ \forall n \geq n_A \text{ or } \exists n_B \in \mathbb{N} : x \in B_n\ \forall n \geq n_B \\
&\Leftrightarrow \exists n_A \in \mathbb{N} : x \in \cap_{n=n_A}^{\infty} A_n \text{ or } \exists n_B \in \mathbb{N} : x \in \cap_{n=n_B}^{\infty} B_n \\
&\Leftrightarrow x \in \liminf_{n\to\infty} A_n \text{ or } x \in \liminf_{n\to\infty} B_n \\
&\Leftrightarrow x \in (\liminf_{n\to\infty} A_n) \cup (\liminf_{n\to\infty} B_n).
\end{aligned}
$$

For (iv), we notice that

$$
\begin{aligned}
x \in \limsup_{n\to\infty}(A_n \cup B_n) &\Leftrightarrow \forall n^* \in \mathbb{N}\ x \in \cup_{n=n^*}^{\infty}(A_n \cup B_n) \\
&\Leftrightarrow \forall n^* \in \mathbb{N}\ \exists \tilde{n}^* \geq n^* : x \in (A_{\tilde{n}^*} \cup B_{\tilde{n}^*}) \\
&\Leftrightarrow \forall n^* \in \mathbb{N}\ \exists \tilde{n}^* \geq n^* : x \in A_{\tilde{n}^*} \text{ or } x \in B_{\tilde{n}^*} \\
&\Leftrightarrow \forall n_A \in \mathbb{N}\ \exists \tilde{n}_A \geq n_A : x \in A_{\tilde{n}_A} \text{ or } \forall n_B \in \mathbb{N}\ \exists \tilde{n}_B \geq n_B : x \in B_{\tilde{n}_B} \\
&\Leftrightarrow \forall n_A \in \mathbb{N}\ x \in \cup_{n=n_A}^{\infty} A_n \text{ or } \forall n_B \in \mathbb{N}\ x \in \cup_{n=n_B}^{\infty} B_n \\
&\Leftrightarrow x \in \limsup_{n\to\infty} A_n \text{ or } x \in \limsup_{n\to\infty} B_n \\
&\Leftrightarrow x \in (\limsup_{n\to\infty} A_n) \cup (x \in \limsup_{n\to\infty} B_n).
\end{aligned}
$$

Finally, using that $A = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$ and $B = \liminf_{n\to\infty} B_n = \limsup_{n\to\infty} B_n$, (v) follows from (i) and (ii) and (vi) follows from (iii) and (iv). ■

Corollary B.1 *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets in $\mathbb{R}^p$ and $B \subset \mathbb{R}^p$. Then, it holds that*

(i) $\liminf_{n\to\infty}(A_n \cap B) = (\liminf_{n\to\infty} A_n) \cap B$

(ii) $\limsup_{n\to\infty}(A_n \cap B) = (\limsup_{n\to\infty} A_n) \cap B$,

(iii) $\liminf_{n\to\infty}(A_n \cup B) = (\liminf_{n\to\infty} A_n) \cup B$, *and*

(iv) $\limsup_{n\to\infty}(A_n \cup B) = (\limsup_{n\to\infty} A_n) \cup B$.

In particular, if $A = \lim_{n\to\infty} A_n$ exists, then

$$\text{(v)} \lim_{n\to\infty}(A_n \cap B) = A \cap B \text{ and (vi) } \lim_{n\to\infty}(A_n \cup B) = A \cup B.$$

Proof of Corollary B.1. (i),(iv),(v) and (vi) follow directly from Lemma B.1 (i),(iv),(v) and (vi) with $B_n = B$ for all $n \in \mathbb{N}$. We now prove (ii). It holds that

$$\begin{aligned}
x \in \limsup_{n\to\infty}(A_n \cap B) \Leftrightarrow & \forall n^* \in \mathbb{N}\ x \in \cup_{n=n^*}^{\infty}(A_n \cap B) \\
\Leftrightarrow & \forall n^* \in \mathbb{N}\ \exists \tilde{n}^* \geq n^* : x \in (A_{\tilde{n}^*} \cap B) \\
\Leftrightarrow & \forall n^* \in \mathbb{N}\ \exists \tilde{n}^* \geq n^* : x \in A_{\tilde{n}^*} \text{ and } x \in B \\
\Leftrightarrow & \forall n^* \in \mathbb{N}\ x \in \cup_{n=n^*}^{\infty} A_n \text{ and } x \in B \\
\Leftrightarrow & x \in \limsup_{n\to\infty} A_n \text{ and } x \in B \\
\Leftrightarrow & x \in (\limsup_{n\to\infty} A_n) \cap B.
\end{aligned}$$

For (iii), we have that

$$\begin{aligned}
x \in \liminf_{n\to\infty}(A_n \cup B_n) \Leftrightarrow & \exists n^* \in \mathbb{N} : x \in \cap_{n=n^*}^{\infty}(A_n \cup B) \\
\Leftrightarrow & \exists n^* \in \mathbb{N} : x \in (A_n \cup B)\ \forall n \geq n^* \\
\Leftrightarrow & \exists n^* \in \mathbb{N} : \forall n \geq n^*\ x \in A_n \text{ or } x \in B \\
\Leftrightarrow & \exists n^* \in \mathbb{N} : x \in \cap_{n=n^*}^{\infty} A_n \text{ or } x \in B \\
\Leftrightarrow & x \in \liminf_{n\to\infty} A_n \text{ or } x \in B \\
\Leftrightarrow & x \in (\liminf_{n\to\infty} A_n) \cup B.
\end{aligned}$$

■

Lemma B.2 *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets in $\mathbb{R}^p$ and $B \subset \mathbb{R}^p$. Then*

$$\liminf_{n\to\infty}(B \setminus A_n) = B \setminus (\limsup_{n\to\infty} A_n) \text{ and } \limsup_{n\to\infty}(B \setminus A_n) = B \setminus (\liminf_{n\to\infty} A_n).$$

In particular, if $A = \lim_{n\to\infty} A_n$ exists, then

$$\lim_{n\to\infty}(\mathbb{R}^p \setminus A_n) = \mathbb{R}^p \setminus A.$$

Proof of Lemma B.2. We use that, for a sequence of sets $\{C_n\}_{n=1}^{\infty}$ in $\mathbb{R}^p$ and $D \subset \mathbb{R}^p$, it holds that $D \setminus (\cup_{n=1}^{\infty} C_n) = \cap_{n=1}^{\infty}(D \setminus C_n)$ and $D \setminus (\cap_{n=1}^{\infty} C_n) = \cup_{n=1}^{\infty}(D \setminus C_n)$. Then, we have that

$$\begin{aligned}
\liminf_{n\to\infty}(B \setminus A_n) &= \cup_{n=1}^{\infty}(B \setminus (\cup_{l=n}^{\infty} A_l)) = B \setminus (\limsup_{n\to\infty} A_n), \text{ and} \\
\limsup_{n\to\infty}(B \setminus A_n) &= \cap_{n=1}^{\infty}(B \setminus (\cap_{l=n}^{\infty} A_l)) = B \setminus (\liminf_{n\to\infty} A_n).
\end{aligned}$$

Finally, the last part follows from $A = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$. ■

Lemma B.3 *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets in $\mathbb{R}^p$ and $A \subset \mathbb{R}^p$. Then, $\lim_{n\to\infty} A_n = A$ if and only if $\lim_{n\to\infty}(A_n \Delta A) = \emptyset$.*

Proof of Lemma B.3. First, suppose that $\lim_{n\to\infty} A_n = A$. Using Lemma B.2 and Corollary B.1 (v)-(vi), we have that

$$\begin{aligned}\emptyset &= (A \cap (\mathbb{R}^p \setminus A)) \cup ((\mathbb{R}^p \setminus A) \cap A)\\ &= ((\lim_{n\to\infty} A_n) \cap (\mathbb{R}^p \setminus A)) \cup ((\lim_{n\to\infty} (\mathbb{R}^p \setminus A_n)) \cap A)\\ &= (\lim_{n\to\infty} (A_n \cap (\mathbb{R}^p \setminus A))) \cup (\lim_{n\to\infty} ((\mathbb{R}^p \setminus A_n) \cap A))\\ &= \lim_{n\to\infty} ((A_n \cap (\mathbb{R}^p \setminus A)) \cup ((\mathbb{R}^p \setminus A_n) \cap A)) = \lim_{n\to\infty} (A_n \Delta A).\end{aligned}$$

Second, suppose that $\lim_{n\to\infty}(A_n \Delta A) = \emptyset$. Then, by Lemma B.1 (iv), Corollary B.1 (ii), and Lemma B.2, it holds that

$$\begin{aligned}\emptyset = \limsup_{n\to\infty}(A_n \Delta A) &= \limsup_{n\to\infty}((A_n \cap (\mathbb{R}^p \setminus A)) \cup ((\mathbb{R}^p \setminus A_n) \cap A))\\ &= (\limsup_{n\to\infty}(A_n \cap (\mathbb{R}^p \setminus A))) \cup (\limsup_{n\to\infty}((\mathbb{R}^p \setminus A_n) \cap A))\\ &= ((\limsup_{n\to\infty} A_n) \cap (\mathbb{R}^p \setminus A)) \cup ((\limsup_{n\to\infty}(\mathbb{R}^p \setminus A_n)) \cap A)\\ &= ((\limsup_{n\to\infty} A_n) \cap (\mathbb{R}^p \setminus A)) \cup ((\mathbb{R}^p \setminus (\liminf_{n\to\infty} A_n)) \cap A).\end{aligned}$$

Therefore, $(\limsup_{n\to\infty} A_n) \cap (\mathbb{R}^p \setminus A) = \emptyset$ and $(\mathbb{R}^p \setminus (\liminf_{n\to\infty} A_n)) \cap A = \emptyset$, which imply that $\limsup_{n\to\infty} A_n \subset A$ and $A \subset \liminf_{n\to\infty} A_n$. Hence, $\lim_{n\to\infty} A_n = A$. ■

Lemma B.4 *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets in $\mathbb{R}^p$ and $\xi \geq 0$. Then,*

$$(\liminf_{n\to\infty} A_n)^{+\xi} \subset \liminf_{n\to\infty}(A_n)^{+\xi} \subset \limsup_{n\to\infty}(A_n)^{+\xi} \subset (\limsup_{n\to\infty} A_n)^{+\xi}.$$

In particular, if $A := \lim_{n\to\infty} A_n$ exists, then $\lim_{n\to\infty}(A_n)^{+\xi} = (A)^{+\xi}$ and $\lim_{n\to\infty} \overline{A_n} = \overline{A}$. Finally, if A_n and A are open, then $\lim_{n\to\infty} \partial A_n = \partial A$.

Proof of Lemma B.4. For the first part, it is enough to show the first and third inclusion. With this aim, let $x \in (\liminf_{n\to\infty} A_n)^{+\xi}$. Hence, $\mathrm{dist}(\{x\}, \cup_{j=1}^{\infty} \cap_{n=j}^{\infty} A_n) \leq \xi$. Therefore, there exists a sequence $\{y_l\}_{l=1}^{\infty}$ in $\cup_{j=1}^{\infty} \cap_{n=j}^{\infty} A_n$ such that $\lim_{l\to\infty}\|x - y_l\| \leq \xi$. Then, for some $n^* \in \mathbb{N}$, the sequence $\{y_l\}_{l=1}^{\infty}$ is in $\cap_{n=n^*}^{\infty} A_n$, that is, $\{y_l\}_{l=1}^{\infty}$ is in A_n for all $n \geq n^*$. It follows that, for all $n \geq n^*$, $\mathrm{dist}(\{x\}, A_n) \leq \lim_{l\to\infty}\|x - y_l\| \leq \xi$. Hence, for all $n \geq n^*$, $x \in (A_n)^{+\xi}$, that is, $x \in \cap_{n=n^*}^{\infty}(A_n)^{+\xi} \subset \liminf_{n\to\infty}(A_n)^{+\xi}$. We now prove the third inclusion. To this end, let $x \in \limsup_{n\to\infty}(A_n)^{+\xi}$. Then, for all $j \in \mathbb{N}$, there exists a constant $n \geq j$ such that $x \in (A_n)^{+\xi}$, that is, $\mathrm{dist}(\{x\}, A_n) \leq \xi$. It follows that $\mathrm{dist}(\{x\}, \cup_{k=j}^{\infty} A_k) \leq \mathrm{dist}(\{x\}, A_n) \leq \xi$. Hence, $x \in (\cup_{k=j}^{\infty} A_k)^{+\xi} = (\cap_{l=1}^{j} \cup_{k=l}^{\infty} A_k)^{+\xi}$, for all $j \in \mathbb{N}$, which implies that $x \in (\limsup_{k\to\infty} A_k)^{+\xi}$. For the second part, notice that, by definition of limit of sets, $A = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$. It follows from the first part that $\liminf_{n\to\infty}(A_n)^{+\xi} = \limsup_{n\to\infty}(A_n)^{+\xi} = (A)^{+\xi}$. Next, notice that, for all $\emptyset \neq B \subset \mathbb{R}^p$, $x \in \overline{B}$ if and only if $\mathrm{dist}(\{x\}, B) = 0$. In particular, $\overline{B} = (B)^{+0}$. Hence, $\lim_{n\to\infty} \overline{A_n} = \overline{A}$. Finally, if A_n and A are open, then, using Lemma B.1 (v) and Lemma B.2, we have that

$$\lim_{n\to\infty} \partial A_n = \lim_{n\to\infty} (\overline{A}_n \cap (\mathbb{R}^p \setminus A_n)) = (\lim_{n\to\infty} \overline{A}_n) \cap (\lim_{n\to\infty} (\mathbb{R}^p \setminus A_n)) = \overline{A} \cap (\mathbb{R}^p \setminus A) = \partial A.$$

■

Lemma B.5 *Let* $\{A_n\}_{n=1}^{\infty}$ *be a sequence of sets in* $\mathbb{R}^p$ *and* $\xi > 0$. *If* $\lim_{n\to\infty} A_n = A$, *then there exists* $n^*(\xi) \in \mathbb{N}$ *such that, for all* $n \geq n^*(\xi)$, $A_n^{+\xi} \subset A$ *and* $A_n \subset (A)^{+\xi}$.

Proof of Lemma B.5. Since $\lim_{n\to\infty} \cup_{j=n}^{\infty} A_j = A \subset (A)^{+\xi}$, there exists $n_1^*(\epsilon) \in \mathbb{N}$ such that $\cup_{n=n^*}^{\infty} A_n \subset (A)^{+\xi}$. Hence $A_n \subset (A)^{+\xi}$, for all $n \geq n_1^*(\xi)$. On the other hand, by Lemma B.4, we have that $\lim_{n\to\infty} \cap_{j=n}^{\infty} (A_j)^{+\xi} = (A)^{+\xi} \supset A$. Hence, there is $n_2^*(\xi)$ such that $\cap_{n=n^*}^{\infty} (A_n)^{+\xi} \supset A$, which implies that $(A_n)^{+\xi} \supset A$, for all $n \geq n_2^*(\xi)$. ■

Bibliography

Abbondandolo, Alberto and Majer Pietro (2006). "On the global stable manifold". In: *Studia Mathematica* 177, pp. 113–131.

Agarwal, Ratan Prakash and V. Lakshmikantham (1993). *Uniqueness and nonuniqueness criteria for ordinary differential equations*. Vol. 6. World Scientific.

Agostinelli, Claudio and Mario Romanazzi (2008). *Local depth of multidimensional data*. Working Paper 3. Ca' Foscari University of Venice.

— (2011). "Local depth". In: *Journal of Statistical Planning and Inference* 141.2, pp. 817–830.

Arcones, Miguel A. (1995). "A Bernstein-type inequality for U-statistics and U-processes". In: *Statistics & probability letters* 22.3, pp. 239–247.

Arcones, Miguel A., Hengjian Cui, and Yijun Zuo (2006). "Empirical depth processes". In: *Test* 15.1, pp. 151–177.

Arcones, Miguel A. and Evarist Giné (1993). "Limit theorems for U-processes". In: *The Annals of Probability* 21.3, pp. 1494–1542.

Astola, Jaakko, Petri Haavisto, and Yrjö Neuvo (1990). "Vector median filters". In: *Proceedings of the IEEE* 78.4, pp. 678–689.

Atkinson, Kendall, Weimin Han, and David Stewart (2009). *Numerical solution of ordinary differential equations*. John Wiley & Sons.

Billingsley, Patrick (2012). *Probability and Measure*. Vol. 939. John Wiley & Sons.

Bonnotte, Nicolas (2013). "Unidimensional and evolution methods for optimal transportation". PhD thesis. Université Paris-Sud.

Bremner, David and Rasoul Shahsavarifar (2018). "Approximate Data Depth Revisited." In: *CCCG*, pp. 272–281.

Brenier, Yann (1991). "Polar factorization and monotone rearrangement of vector-valued functions". In: *Communications on Pure and Applied Mathematics* 44.4, pp. 375–417.

Brezis, Haim (2011). *Functional analysis, Sobolev spaces and partial differential equations*. Springer.

Chacón, José E. (2009). "Data-driven choice of the smoothing parametrization for kernel density estimators". In: *Canadian Journal of Statistics* 37.2, pp. 249–265.

— (2015). "A population background for nonparametric density-based clustering". In: *Statistical Science* 30.4, pp. 518–532.

Chacón, José E. and Tarn Duong (2018). *Multivariate kernel smoothing and its applications*. Chapman and Hall/CRC.

Chazal, Frédéric, Leonidas J. Guibas, Steve Y. Oudot, and Primoz Skraba (2013). "Persistence-Based Clustering in Riemannian Manifolds". In: *J. ACM* 60.6.

Chen, Yen-Chi, Christopher R. Genovese, Ryan J. Tibshirani, and Larry Wasserman (2016). "Nonparametric modal regression". In: *Annals of Statistics* 44.2, pp. 489–514.

Chen, Zhiqiang (1995). "Robustness of the half-space median". In: *Journal of statistical planning and inference* 46.2, pp. 175–181.

Chernozhukov, Victor, Alfred Galichon, Marc Hallin, and Marc Henry (2017). "Monge–Kantorovich depth, quantiles, ranks and signs". In: *The Annals of Statistics* 45.1, pp. 223–256.

Chow, Yuan Shih and Henry Teicher (1997). "Probability Theory: Independence, Interchangeability, Martingales". In:

Cuesta, Juan Antonio and Carlos Matran (1989). "Notes on the Wasserstein Metric in Hilbert Spaces". In: *The Annals of Probability* 17.3, pp. 1264–1276.

Cuevas, Antonio, Manuel Febrero, and Ricardo Fraiman (2007). "Robust estimation and classification for functional data via projection-based depth notions". In: *Computational Statistics* 22, pp. 481–496.

Despres, Christian J. J. (2017). "The Vapnik-Chervonenkis dimension of cubes in $\mathbb{R}^d$". In: *arXiv preprint arXiv:1412.6612v3*.

Devroye, Luc and László Györfi (1985). *Nonparametric Density Estimation: The L1 View*. Wiley & Sons.

Donoho, David L. and Miriam Gasko (1992). "Breakdown properties of location estimates based on halfspace depth and projected outlyingness". In: *The Annals of Statistics* 20.4, pp. 1803–1827.

Dudley, Richard M. (1979). "Balls in $\mathbb{R}^k$ do not cut all subsets of $k+2$ points". In: *Advances in Mathematics* 31.3, pp. 306 –308.

— (2014). *Uniform central limit theorems*. Vol. 142. Cambridge university press.

— (2018). *Real analysis and probability*. Chapman and Hall/CRC.

Duong, Tarn (2018). *ks: kernel smoothing*. R package version 1.11.3.

Dutta, Subhajit, Anil K. Ghosh, and Probal Chaudhuri (2011). "Some intriguing properties of Tukey's half-space depth". In: *Bernoulli* 17.4, pp. 1420–1434.

Dyckerhoff, Rainer (2004). "Data depths satisfying the projection property". In: *Allgemeines Statistisches Archiv* 88.2, pp. 163–190.

Edgar, Gerald (2007). *Measure, topology, and fractal geometry*. Springer Science & Business Media.

Eisenstat, Stanley C. and Ilse C. F. Ipsen (1998). "Three absolute perturbation bounds for matrix eigenvalues imply relative bounds". In: *SIAM Journal on Matrix Analysis and Applications* 20.1, pp. 149–158.

Elmore, Ryan T., Thomas P. Hettmansperger, and Fengjuan Xuan (2006). "Spherical data depth and a multivariate median". In: *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 72, p. 87.

Folland, Gerald B. (1999). *Real analysis: modern techniques and their applications*. Vol. 40. John Wiley & Sons.

Francisci, Giacomo, Claudio Agostinelli, Alicia Nieto-Reyes, and Anand N. Vidyashankar (2020). "Analytical and statistical properties of local depth functions motivated by clustering applications". In: *arXiv preprint arXiv:2008.11957*.

Francisci, Giacomo, Alicia Nieto-Reyes, and Claudio Agostinelli (2019). "Generalization of the simplicial depth: no vanishment outside the convex hull of the distribution support". In: *arXiv preprint arXiv:1909.02739*.

Fukunaga, Keinosuke and Larry Hostetler (1975). "The estimation of the gradient of a density function, with applications in pattern recognition". In: *IEEE Transactions on information theory* 21.1, pp. 32–40.

Galpin, Jacqueline S. and Douglas M. Hawkins (1987). "Methods of L1 estimation of a covariance matrix". In: *Computational Statistics & Data Analysis* 5.4, pp. 305–319.

Garling, David J. H. (2007). *Inequalities: a journey into linear analysis*. Cambridge University Press.

Geenens, Gery, Alicia Nieto-Reyes, and Giacomo Francisci (2021). "Statistical depth in abstract metric spaces". In: *arXiv preprint arXiv:2107.13779*.

Genovese, Christopher R., Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman (2016). "Non-parametric inference for density modes". In: *Journal of the Royal Statistical Society Series B* 78.1, pp. 99–126.

Giné, Evarist and Richard Nickl (2016). *Mathematical foundations of infinite-dimensional statistical models*. Vol. 40. Cambridge University Press.

Hale, Jack K. (1980). *Ordinary differential equations*. Vol. 21. Dover Publications Inc.

Hallin, Marc, Eustasio Del Barrio, Juan Cuesta-Albertos, and Carlos Matrán (2021). "Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach". In: *The Annals of Statistics* 49.2, pp. 1139–1165.

Hamel, Andreas H. and Daniel Kostner (2018). "Cone distribution functions and quantiles for multivariate random variables". In: *Journal of Multivariate Analysis* 167, pp. 97–113.

Hardy, Michael (2006). "Combinatorics of Partial Derivatives". In: *The electronic journal of combinatorics* 13.R1.

Harley, Peter and George McNulty (1979). "When is a point Borel?" In: *Pacific Journal of Mathematics* 80.1, pp. 151–157.

Hirsch, Morris W., Robert L. Devaney, and Stephen Smale (1974). *Differential equations, dynamical systems, and linear algebra*. Vol. 60. Academic press.

Hoeffding, Wassily (1948). "A Class of Statistics with Asymptotically Normal Distribution". In: *The Annals of Mathematical Statistics* 19.3, pp. 293–325.

— (1961). "The strong law of large numbers for U-statistics". In: *University of North Carolina Institute of statistics. Mimeo Series* 302.

— (1963). "Probability Inequalities for Sums of Bounded Random Variables". In: *Journal of the American Statistical Association* 58.301, pp. 13–30.

Holte, John M. (2009). "Discrete Gronwall lemma and applications". In: *MAA-NCS meeting at the University of North Dakota*.

Jost, Jürgen (2005). *Riemannian geometry and geometric analysis*. Springer.

Kallenberg, Olav (1997). *Foundations of modern probability*. Vol. 2. Springer.

Kenne Pagui, Euloge Clovis, Alessandra Salvan, and Nicola Sartori (2017). "Median bias reduction of maximum likelihood estimates". In: *Biometrika* 104.4, pp. 923–938.

Kleindessner, Matthäus and Ulrike Von Luxburg (2017). "Lens depth function and k-relative neighborhood graph: versatile tools for ordinal data analysis". In: *The Journal of Machine Learning Research* 18.1, pp. 1889–1940.

Knott, Martin and Cyril S. Smith (1984). "On the optimal mapping of distributions". In: *Journal of Optimization Theory and Applications* 43.1, pp. 39–49.

Korolyuk, Vladimir S. and Yu V. Borovskich (2013). *Theory of U-statistics*. Vol. 273. Springer Science & Business Media.

Kuelbs, James and Joel Zinn (2016). "Convergence of quantile and depth regions". In: *Stochastic Processes and their Applications* 126.12, pp. 3681–3700.

Lee, A. J. (1990). "U-statistics: Theory and Practice". In:

Li, Shengqiao (2011). "Concise formulas for the area and volume of a hyperspherical cap". In: *Asian Journal of Mathematics and Statistics* 4.1, pp. 66–70.

Liu, Regina Y. (1990). "On a notion of data depth based on random simplices". In: *The Annals of Statistics* 18.1, pp. 405–414.

— (1992). "Data depth and multivariate rank tests". In: *L1-statistical analysis and related methods*. Ed. by Dodge Y. North-Holland Amsterdam, pp. 279–294.

Liu, Regina Y. and Kesar Singh (1993). "A quality index based on data depth and multivariate rank tests". In: *Journal of the American Statistical Association* 88.421, pp. 252–260.

Liu, Zhenyu and Reza Modarres (2011). "Lens data depth and median". In: *Journal of Nonparametric Statistics* 23.4, pp. 1063–1074.

Mahalanobis, Prasanta Chandra (1936). "On the generalized distance in statistics". In: *Proceedings of the National Institute of Science of India*. Vol. 12, pp. 49–55.

Major, Péter (1978). "On the invariance principle for sums of independent identically distributed random variables". In: *Journal of Multivariate Analysis* 8.4, pp. 487–517.

Makinde, Olusola Samuel (2019). "Gene expression data classification: some distance-based methods". In: *Kuwait Journal of science* 46.3, pp. 31–39.

Massé, Jean-Claude (2002). "Asymptotics for the Tukey median". In: *Journal of Multivariate Analysis* 81.2, pp. 286–300.

— (2004). "Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean". In: *Bernoulli* 10.3, pp. 397–419.

McCann, Robert J. (1995). "Existence and uniqueness of monotone measure-preserving maps". In: *Duke Mathematical Journal* 80.2, pp. 309–323.

Menardi, Giovanna (2016). "A review on modal clustering". In: *International Statistical Review* 84.3, pp. 413–433.

Mosler, Karl and Pavlo Mozharovskyi (2020). "Choosing among notions of multivariate depth statistics". In: *arXiv preprint arXiv:2004.01927*.

Ohm, Jens-Rainer et al. (2012). "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)". In: *IEEE Transactions on circuits and systems for video technology* 22.12, pp. 1669–1684.

Oja, Hannu (1983). "Descriptive statistics for multivariate distributions". In: *Statistics & Probability Letters* 1.6, pp. 327–332.

Ojo, Oluwasegun Taiwo, Antonio Fernandez Anta, Rosa E. Lillo, and Carlo Sguera (2019). "Detecting and Classifying Outliers in Big Functional Data". In: *arXiv preprint arXiv:1912.07287*.

Oksendal, Bernt (2003). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.

Perko, Lawrence (2013). *Differential equations and dynamical systems*. Vol. 7. Springer Science & Business Media.

Reich, Alfred John (1980). "Robust and adaptive estimation of location and scale". PhD thesis. Texas Tech University.

Rudin, Walter (1976). *Principles of mathematical analysis*. Vol. 3. McGraw-hill New York.

Rüschendorf, L. and S. T. Rachev (1990). "A characterization of random variables with minimum L^2-distance". In: *Journal of Multivariate Analysis* 32.1, pp. 48–54.

Sagan, Hans (1994). *Space-filling curves*. Springer.

Schneemeier, Wilhelm (1989). "Weak convergence and Glivenko-Cantelli results for empirical processes of U-statistic structure". In: *Stochastic processes and their applications* 33.2, pp. 325–334.

Serfling, Robert (2002). "Quantile functions for multivariate analysis: approaches and applications". In: *Statistica Neerlandica* 56.2, pp. 214–232.

Serfling, Robert J. (2009). *Approximation theorems of mathematical statistics*. Vol. 162. John Wiley & Sons.

Stahel, W. A. (1981). *Breakdown of covariance estimators*. Tech. rep. ETH Zürich.

Stein, P. (1966). "A note on the volume of a simplex". In: *The American Mathematical Monthly* 73.3, pp. 299–301.

Tao, Terence (2011). *An introduction to measure theory*. Vol. 126. American Mathematical Society Providence, RI.

Teschl, Gerald (2012). *Ordinary differential equations and dynamical systems*. Vol. 140. American Mathematical Society.

Torchinsky, Alberto (1995). *Real variables*. Chapman and Hall/CRC.

Tu, Loring W. (2011). *An introduction to manifolds.*

Tukey, John W. (1975). “Mathematics and the picturing of data”. In: *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*. Vol. 2, pp. 523–531.

Tyler, David E. (1987). “A distribution-free M-estimator of multivariate scatter”. In: *The annals of Statistics*, pp. 234–251.

van der Vaart, Aad W. and Jon A. Wellner (1996). *Weak convergence*. Springer.

Villani, Cédric (2009). *Optimal transport: old and new*. Vol. 338. Springer.

Wand, Matt P. and M. Chris Jones (1993). “Comparison of smoothing parameterizations in bivariate kernel density estimation”. In: *Journal of the American Statistical Association* 88.422, pp. 520–528.

Wheeden, Richard L. and Antoni Zygmund (2015). *Measure and integral: an introduction to real analysis*. Chapman and Hall/CRC.

Yang, Mengta and Reza Modarres (2018). “β-Skeleton depth functions and medians”. In: *Communications in Statistics-Theory and Methods* 47.20, pp. 5127–5143.

Zuo, Y. (1998). “Contributions to the theory and applications of statistical depth”. PhD thesis. The University of Texas at Dallas. URL: http://www.stt.msu.edu/~zuo/papers_html/main.pdf.

Zuo, Yijun (2003). “Projection-based depth functions and associated medians”. In: *The Annals of Statistics* 31.5, pp. 1460–1490.

Zuo, Yijun and Robert Serfling (2000a). “General notions of statistical depth function”. In: *Annals of statistics* 28.2, pp. 461–482.

— (2000b). “On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry”. In: *Journal of Statistical Planning and Inference* 84, pp. 55–79.

— (2000c). “Structural properties and convergence results for contours of sample statistical depth functions”. In: *The Annals of Statistics* 28.2, pp. 483–499.