Electronic Journal of Statistics
ISSN: 1935-7524

Appendices to: Analytical and statistical properties of local depth functions motivated by clustering applications

Giacomo Francisci[1], Claudio Agostinelli[2], Alicia Nieto-Reyes [3] and Anand N. Vidyashankar [1]

[1] *Department of Statistics, George Mason University, e-mail:* gfranci@gmu.edu; avidyash@gmu.edu

[2] *Department of Mathematics, University of Trento, e-mail:* claudio.agostinelli@unitn.it

[3] *Department of Mathematics, Statistics and Computer Science, University of Cantabria, e-mail:* alicia.nieto@unican.es

A. Discussion and additional proofs

A.1. Useful properties of local depth functions

We begin with a brief discussion concerning the properties of local depth functions that are used in the proofs and discussions in the main paper. We first observe, using (**P2**), that $h_0^{(G)}(x;\cdot) = l_G \mathbf{I}(\cdot \in \{(x,\dots,x)\})$, where $l_G = G(0,\dots,0)$. Also, by Definition 2.1, and by (**P1**), (**P2**), $h_1^{(G)}(0;\cdot) = G(\cdot)$ and

$$0 \le h_{(\cdot)}^{(G)}(\cdot;\cdot) \le l_G. \tag{A.1}$$

Furthermore, (**P4**) ensures that $h_\tau(0;\cdot)$ is non-trivial for all $\tau > 0$, since there is a region including the origin of $(\mathbb{R}^d)^{k_G}$ and having positive Lebesgue measure where $h_\tau(0;\cdot)$ is positive. We note that (**P4**) is satisfied whenever $l_G > 0$ and $G(\cdot)$ is continuous in $(0,\dots,0)$. We will further suppose without loss of generality (w.l.o.g.) that $G(x_1,\dots,x_{k_G}) = G(x_{i_1},\dots,x_{i_{k_G}})$ for every permutation $(i_1,\dots,i_{k_G})$ of $(1,\dots,k_G)$ yielding

$$h_\tau^{(G)}(x;x_1,\dots,x_{k_G}) = h_\tau^{(G)}(x;x_{i_1},\dots,x_{i_{k_G}}); \tag{A.2}$$

since otherwise, one can replace $G(\cdot)$ by $\bar{G}(\cdot)$, where, for $(x_1,\dots,x_{k_G}) \in (\mathbb{R}^p)^{k_G}$,

$$\bar{G}(x_1,\dots,x_{k_G}) = \frac{1}{k_G!}\sum G(x_{i_1},\dots,x_{i_{k_G}}),$$

and the summation is over all $k_G!$ permutations $(i_1,\dots,i_{k_G})$ of $(1,\dots,k_G)$. Also, notice that

$$h_\tau^{(G)}(x+v;x_1+v,\dots,x_{k_G}+v) = h_\tau^{(G)}(x;x_1,\dots,x_{k_G}), \quad v \in \mathbb{R}^p \tag{A.3}$$

$$\text{and} \quad h_\tau^{(G)}(-x;-x_1,\dots,-x_{k_G}) = h_\tau^{(G)}(x;x_1,\dots,x_{k_G}). \tag{A.4}$$

If P is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^p$ with density $f(\cdot)$, then, by (A.3) and (A.4), for all $x \in \mathbb{R}^p$ and $\tau \in [0,\infty]$, it holds that

$$\begin{aligned} LGD(x,\tau,P) &= \int h_\tau(0;x-x_1,\dots,x-x_k)f(x_1)\dots f(x_k)dx_1\dots dx_k \\ &= (h_\tau(0;\cdot) * f^{\otimes k}(\cdot))(x,\dots,x), \end{aligned} \tag{A.5}$$

where $*$ is the convolution operator and $f^{\otimes k}(x_1,\dots x_k) = f(x_1)\dots f(x_k)$. Thus, (2.5) holds.

A.2. Additional proofs

In the following we will use the following notations: $\{e_i\}_{i=1}^p$ is the standard basis of $\mathbb{R}^p$ and the coordinates of a vector $x \in \mathbb{R}^p$ are given by $x^{(i)} \coloneqq \langle x, e_i\rangle$, $i = 1,\dots,p$.

Proof of Proposition 2.1. We start by proving (i). For the monotonicity, observe that, by Definition 2.1 and (**P2**), for all $x \in \mathbb{R}^p$, $(x_1,\dots,x_k) \in (\mathbb{R}^p)^k$ and $0 \le \tau_1 \le \tau_2 \le \infty$, $h_{\tau_1}(x;x_1,\dots,x_k) \le h_{\tau_2}(x;x_1,\dots,x_k)$ and therefore $LGD(x,\tau_1) \le LGD(x,\tau_2)$. Using dominated convergence theorem (DCT) and Definition 2.1, we get that

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

and

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

For (ii) let $0 < \epsilon < 1$. Using tightness and (**P3**), let $r_1, r^* > 0$ such that $P(\overline{B}_{r_1}(0)) \ge 1-\epsilon$ and $h_\tau(x;x_1,\dots,x_k) \le \epsilon$ for all $x \in \mathbb{R}^p$, $x_1 \in \mathbb{R}^p \setminus \overline{B}_{\tau r^*}(x)$, and $(x_2,\dots,x_k) \in (\mathbb{R}^p)^{k-1}$. Since, for $r_2 > \tau r^*$ and $x \in \mathbb{R}^p \setminus \overline{B}_{r_1+r_2}(0)$, it holds that $\overline{B}_{\tau r^*}(x) \subset \mathbb{R}^p \setminus \overline{B}_{r_1}(0)$, using (A.1), we see that for $r \ge r_1 + r_2$

$$\sup_{x\in\mathbb{R}^p\setminus\overline{B}_r(0)} LGD(x,\tau) \le l \sup_{x\in\mathbb{R}^p\setminus\overline{B}_{r_1+r_2}(0)} P(\overline{B}_{\tau r^*}(x)) + \epsilon \le (l+1)\epsilon.$$

We now prove (iii). Let $f(\cdot)$ be the density function of P with respect to λ. By (A.1), we have that $0 \le LGD(x,\tau) \le l$. Furthermore, by (2.5) and (A.3), it holds that

$$|LGD(y,\tau) - LGD(x,\tau)| \le l \int \left| \prod_{j=1}^k f(y-x_j) - \prod_{j=1}^k f(x-x_j) \right| dx_1\dots dx_k.$$

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

Next, notice that, by (iii) and (2.5), (iv) holds when $m = 0$. One can show by induction that for all $0 \le j \le m$, the partial derivatives of $LGD(\cdot,\tau)$ up to order j exist and are given by

$$\partial_{i_j} \dots \partial_{i_1} LGD(x,\tau) = (h_\tau(0;\cdot) * (g_{i_j,\dots,i_1}(\cdot))(x,\dots,x),$$

where, for $(x_1,\dots,x_k) \in (\mathbb{R}^p)^k$, $g_{i_j,\dots,i_1}(x_1,\dots,x_k) \coloneqq \partial_{i_j} \dots \partial_{i_1} f(x_1) \dots f(x_k)$. Finally, one observes that $\partial_{i_m} \dots \partial_{i_1} LGD(\cdot,\tau)$ is continuous. Details are in Francisci et al. (2022). ■

Before we prove Theorem 2.1 we recall the following results on the approximation of the identity, for the function $G(\cdot)$ (see Section 9.2 in Wheeden and Zygmund (2015) and Section XIII.2 in Torchinsky (1995)).

Lemma A.1 *Let $\tilde{G}_\tau(\cdot) \coloneqq \tau^{-kp} h_\tau(0;\cdot)$. Then the following hold:*

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

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

(iii) Additionally, let $\tilde{f} : (\mathbb{R}^p)^k \to \mathbb{R}^p$ and suppose that assumption (2.4) holds true. Then, at every point $(x_1,\dots,x_k) \in (\mathbb{R}^p)^k$ of continuity of $\tilde{f}(\cdot)$

$$\lim_{\tau\to 0^+} (\tilde{G}_\tau * \tilde{f})(x_1,\dots,x_k) = \tilde{f}(x_1,\dots,x_k). \tag{A.6}$$

Furthermore, (A.6) holds uniformly on any set $A \subset (\mathbb{R}^p)^k$ where $\tilde{f}(\cdot)$ is uniformly continuous.

Proof of Theorem 2.1. The proof of (i) follows from Lemma A.1 (iii). Turning to (ii), we first notice that, since $f(\cdot) \in L^\infty(\mathbb{R}^p)$, $f^q(\cdot) \in L^\infty(\mathbb{R}^p)$. Then, we compute

$$|\tau^{-kp} LGD(x,\tau) - f^k(x)| \le \int \left| \prod_{j=1}^k f(x-x_j) - f^k(x) \right| \tilde{G}_\tau(x_1,\dots,x_k)dx_1 \dots dx_k. \tag{A.7}$$

We recursively apply the triangle inequality and obtain

$$\left| \prod_{j=1}^k f(x-x_j) - f^k(x) \right| \le \sum_{i=1}^k \prod_{j=1}^{i-1} f(x-x_j)|f(x-x_i) - f(x)| f^{k-i}(x), \tag{A.8}$$

thus implying that

$$|\tau^{-kp} LGD(x,\tau) - f^k(x)| \le c_\infty^{k-1} \sum_{i=1}^k \int |f(x-x_i) - f(x)| \tilde{G}_\tau(x_1,\dots,x_k)dx_1 \dots dx_k,$$

where $f(\cdot) \le c_\infty < \infty$ a.e. Now, by Lemma A.1 (ii), for all $\delta > 0$ there exists $\tilde{\tau}(\delta) > 0$ such that, for all $0 < \tau \le \tilde{\tau}(\delta)$,

$$\int_{(\mathbb{R}^p)^k \setminus (\overline{B}_\delta(0))^k} \tilde{G}_\tau(x_1, \dots, x_k) dx_1 \dots x_k \le \epsilon. \tag{A.9}$$

If $x \in \mathbb{R}^p$ is a continuity point for $f(\cdot)$, then for all $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x-y) - f(x)| \le \epsilon$ for all $y \in \overline{B}_\delta(0)$. Using Lemma A.1(i), we conclude that, for all $0 < \tau \le \tilde{\tau}(\delta)$,

$$|\tau^{-kp} LGD(x,\tau) - f^k(x)| \le k c_\infty^{k-1}(1 + 2c_\infty)\epsilon. \tag{A.10}$$

Finally, if $f(\cdot)$ is uniformly continuous on $A \subset \mathbb{R}^p$, then (A.10) holds for all $x \in A$.

For (iii), notice that, by (A.4) and a change of variable in (2.5),

$$\tau^{-kp} LGD(x,\tau) - f^k(x) = \int h_1(0; x_1, \dots, x_k) \left[\prod_{j=1}^k f(x + \tau x_j) - f^k(x) \right] dx_1 \dots dx_k. \tag{A.11}$$

Since $f(\cdot)$ 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,

$$\begin{aligned} \prod_{j=1}^k f(x+\tau x_j) &= f^k(x) + \tau f^{k-1}(x) \langle \nabla f(x), \sum_{i=1}^k x_i \rangle \\ &+ \tau^2 f^{k-1}(x) \sum_{i=1}^k \int_0^1 (1-z) x_i^\top H_f(x + z\tau x_i) x_i dz \\ &+ \tau^2 f^{k-2}(x) \sum_{i=1}^k \sum_{j=i+1}^k \langle \nabla f(x), x_i \rangle \langle \nabla f(x), x_j \rangle + O(\tau^2). \end{aligned} \tag{A.12}$$

The continuity of the second order partial derivatives implies that, for small τ, 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 with

$$\lim_{\tau \to 0^+} \int_0^1 (1-z) x_i^\top H_f(x + z\tau x_i) x_i dz = \frac{1}{2} x_i^\top H_f(x) x_i. \tag{A.13}$$

By substituting (A.12) in (A.11), we see that

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

Using (A.4), we see that

$$\int h_1(0;x_1,\dots,x_k)\langle\nabla f(x),\sum_{i=1}^{k} x_i\rangle dx_1\dots dx_k = 0.$$

Now, (A.2) implies that

$$\begin{aligned}
&\int h_1(0;x_1,\dots,x_k)\Big[\sum_{i=1}^{k}\int_0^1(1-z)x_i^\top H_f(x+z\tau x_i)x_i dz\Big]dx_1\dots dx_k\\
=&k\int h_1(0;x_1,\dots,x_k)\Big[\int_0^1(1-z)x_1^\top H_f(x+z\tau x_1)x_1 dz\Big]dx_1\dots dx_k
\end{aligned}$$

and

$$\begin{aligned}
&\int h_1(0;x_1,\dots,x_k)\Big[\sum_{i=1}^{k}\sum_{j=i+1}^{k}\langle\nabla f(x),x_i\rangle\langle\nabla f(x),x_j\rangle\Big]dx_1\dots dx_k\\
=&\frac{k(k-1)}{2}f^{k-2}(x)\int h_1(0;x_1,\dots,x_k)\langle\nabla f(x),x_1\rangle\langle\nabla f(x),x_2\rangle dx_1\dots dx_k.
\end{aligned}$$

By (A.13) and DCT, we conclude that

$$\lim_{\tau\to 0^+}\tau^{-2}(\tau^{-kp}LGD(x,\tau) - f^k(x)) = R(x),$$

where $R(x) = R_1(x) + R_2(x)$ and

$$\begin{aligned}
R_1(x) &\coloneqq \frac{k}{2}f^{k-1}(x)\int h_1(0;x_1,\dots,x_k)x_1^\top H_f(x)x_1 dx_1\dots dx_k,\\
R_2(x) &\coloneqq \frac{k(k-1)}{2}f^{k-2}(x)\int h_1(0;x_1,\dots,x_k)\langle\nabla f(x),x_1\rangle\langle\nabla f(x),x_2\rangle dx_1\dots dx_k.
\end{aligned}$$

We now prove (iv). We first notice that, since $f(\cdot)\in L^1(\mathbb{R}^p)\cap L^{kd}(\mathbb{R}^p)$, then $f(\cdot)\in L^q(\mathbb{R}^p)$, for all $1\le q\le kd$. Next, using Hölder inequality, Lemma A.1

(i), Jensen inequality, and (A.8), we obtain that

$$|\tau^{-kp}LGD(x,\tau)-f^k(x)|^d \leq \int \left|\prod_{j=1}^{k} f(x-x_j)-f^k(x)\right|^d \tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k$$
$$\leq k^{d-1}\sum_{i=1}^{k} I_{\tau,i},$$

where $I_{\tau,i}$ is defined to be

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

By Fubini's theorem, we have that

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

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

Now, we apply again Hölder inequality with exponents $s=k/(k-1)$ and $t=k$, and see that $J_{\tau,i}(x_1,\dots,x_k)\leq c_1 K(x_1,\dots,x_k)$, where

$$c_1 \coloneqq \max_{i=1,\dots,p}\left[\int\prod_{j=1}^{i-1} f^{sd}(x-x_j)f^{(k-i)sd}(x)dx\right]^{1/s} \text{ and}$$
$$K(x_1,\dots,x_k) \coloneqq \max_{i=1,\dots,p}\left[\int|f(x-x_i)-f(x)|^{td}dx\right]^{1/t}.$$

Notice that $K(x_1,\dots,x_k)\leq c_2 \coloneqq 2^d[\int f(x)^{td}dx]^{1/t}<\infty$, and, for all $\epsilon>0$, by Theorem 8.19 in Wheeden and Zygmund (2015), there exists $\delta>0$ such that $K(x_1,\dots,x_k)\leq\epsilon$ for all $(x_1,\dots,x_k)\in(\overline{B}_\delta(0))^k$. Using Lemma A.1 (i) and (A.9), we conclude that for all $0<\tau\leq\tilde{\tau}(\delta)$

$$\int|\tau^{-kp}LGD(x,\tau)-f^k(x)|^d dx \leq c_1 k^d\int K(x_1,\dots,x_k)\tilde{G}_\tau(x_1,\dots,x_k)dx_1\dots dx_k$$
$$\leq c_1 k^d(1+c_2)\epsilon.$$

■

Before proving Proposition 2.2 we establish useful inequalities in the following lemma.

Lemma A.2 *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 of Proposition 2.2. We start by proving (i). By Lemma A.2

$$\sup_{x\in\mathbb{R}^p}|f_\tau^{(G)}(x)-f(x)|\le \sup_{x\in\mathbb{R}^p} F_\tau^{(G)}(x)^{1/k_G}=(\sup_{x\in\mathbb{R}^p} F_\tau^{(G)}(x))^{1/k_G},\ \text{where}$$

$$F_\tau^{(G)}(x):=\left|\int h_1^{(G)}(0;x_1,\dots,x_{k_G})\left[\prod_{j=1}^{k_G} f(x+\tau x_j)-f^{k_G}(x)\right]dx_1\dots dx_{k_G}\right|.$$

We observe that $\sup_{x\in\mathbb{R}^p} F_\tau^{(G)}(x)$ is bounded above by

$$\int h_1^{(G)}(0;x_1,\dots,x_{k_G})\sup_{x\in\mathbb{R}^p}\left|\prod_{j=1}^{k_G} f(x+\tau x_j)-f^{k_G}(x)\right|dx_1\dots dx_{k_G}.$$

Since $f(\cdot)$ is uniformly continuous, for all $(x_1,\dots,x_{k_G})\in(\mathbb{R}^p)^{k_G}$, it holds that

$$\lim_{\tau\to 0^+}\sup_{x\in\mathbb{R}^p}\left|\prod_{j=1}^{k_G} f(x+\tau x_j)-f^{k_G}(x)\right|=0.$$

The result now follows from DCT, since $h_1^{(G)}(0;\cdot)\in L^1((\mathbb{R}^p)^{k_G})$ and the supremum is bounded because $f(\cdot)$ is bounded as it is uniformly continuous.
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 $\mathbb{R}^p$ replaced by K. For the second part of (ii), notice that

$$\sup_{y\in\overline{B}_\epsilon(x)}|f_\tau^{(G)}(y)-f(x)|\le\sup_{y\in\overline{B}_\epsilon(x)}|f_\tau^{(G)}(y)-f(y)|+\sup_{y\in\overline{B}_\epsilon(x)}|f(y)-f(x)|.$$

The result now follows from the first part of (ii) and continuity of $f(\cdot)$. Finally, for (iii), notice that, by Lemma A.2 and Theorem 2.1 (iv),

$$\int|f_\tau^{(G)}(y)-f(y)|^{k_G d}dy\le\int|(f_\tau^{(G)})^{k_G}(y)-f^{k_G}(y)|^d dy\xrightarrow[\tau\to 0^+]{}0.$$

Before we prove (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 A.1 *Let $\varphi:\mathbb{R}^p\to\mathbb{R}$ and $\psi:\mathbb{R}\to\mathbb{R}$ be m-times continuously differentiable in $A\subset\mathbb{R}^p$ and $\varphi(A)\subset\mathbb{R}$, respectively. Then, $\psi(\varphi(\cdot))$ is m-times continuously differentiable in A and, for $x\in A$ and $i_1,\dots,i_m\in\{1,\dots,p\}$, it holds that*

$$\partial_{i_m}\dots\partial_{i_1}\psi(\varphi(x))=\sum_{R\in\mathcal{R}_m}[\partial^{[\#R]}\psi](\varphi(x))\prod_{\{i_{j_l},\dots,i_{j_1}\}\in R}\partial_{i_{j_l}}\dots\partial_{i_{j_1}}\varphi(x),$$

where $\mathcal{R}_m$ is the set of all partitions of $\{1,\dots,m\}$ and $\partial^{[l]}$ denotes the (unidimensional) l^{th} derivative.

The following lemma, which is standard, is required for completing the proof of the proposition.

Lemma A.3 *Let $\varphi_n : \mathbb{R}^p \to \mathbb{R}$, $\phi_n : \mathbb{R}^p \to \mathbb{R}$ and $A \subset \mathbb{R}^p$. Suppose that $\varphi_n(\cdot)$ and $\phi_n(\cdot)$ converge uniformly on A to $\varphi(\cdot)$ and $\phi(\cdot)$, respectively. It holds that (i) $(\varphi_n + \phi_n)(\cdot)$ converges uniformly on A to $(\varphi + \phi)(\cdot)$ and (ii) if $\varphi_n(\cdot)$ and $\phi_n(\cdot)$ are bounded on A, then $(\varphi_n\phi_n)(\cdot)$ converges uniformly on A to $(\varphi\phi)(\cdot)$.*

We now turn to the proof of (iv). We first notice that, by Proposition 2.1 (iv) and Remark 2.1, $LGD(\cdot, \tau)$ and $f_\tau(\cdot)$ 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 2.1 (i), there exists $\tau^* > 0$ such that, for all $0 < \tau \leq \tau^*$, $\sup_{x\in K}|f^k_\tau(x) - f^k(x)| \leq c_1/2$, implying that $f^k_\tau(x) \in [c_3, c_4]$, where $c_3 := c_1/2$ and $c_4 := c_2 + c_1/2$. Next, we apply Lemma A.1 with $\varphi(\cdot) = f^k(\cdot)$ and $\psi(\cdot) = (\cdot)^{1/k}$, and obtain that

$$\partial_{i_m} \dots \partial_{i_1} f(x) = \sum_{R\in\mathcal{R}_m} [\partial^{[\#R]}\psi](\varphi(x)) \prod_{\{i_{j_l},\dots,i_{j_1}\}\in R} \partial_{i_{j_l}} \dots \partial_{i_{j_1}} \varphi(x). \quad \text{(A.14)}$$

Similarly, with $\varphi_\tau(\cdot) := f^k_\tau(\cdot)$, we have that

$$\partial_{i_m} \dots \partial_{i_1} f_\tau(x) = \sum_{R\in\mathcal{R}_m} [\partial^{[\#R]}\psi](\varphi_\tau(x)) \prod_{\{i_{j_l},\dots,i_{j_1}\}\in R} \partial_{i_{j_l}} \dots \partial_{i_{j_1}} \varphi_\tau(x). \quad \text{(A.15)}$$

By Proposition 2.1 (iv), it holds that

$$\partial_{i_{j_l}} \dots \partial_{i_{j_1}} \varphi_\tau(x) = (\tilde{G}_\tau(\cdot) * (\partial_{i_{j_l}} \dots \partial_{i_{j_1}} f^k(\cdot)))(x, \dots, x).$$

We apply Lemma A.1 (iii) with $\tilde{f}(\cdot, \dots, \cdot) = \partial_{i_{j_l}} \dots \partial_{i_{j_1}} (f(\cdot) \dots f(\cdot))$ and $A = (K)^k$, and obtain that $\partial_{i_{j_l}} \dots \partial_{i_{j_1}} \varphi_\tau(\cdot)$ converges uniformly on K to $\partial_{i_{j_l}} \dots \partial_{i_{j_1}} \varphi(\cdot)$. Next, notice that, for all $j \in \{1, \dots, m\}$, $\partial^{[j]}\psi(\cdot)$ 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 2.1 (i), there exists $0 < \tau^{**} \leq \tau^*$, such that, for all $0 < \tau \leq \tau^{**}$, $\sup_{x\in K}|f^k_\tau(x) - f^k(x)| \leq \delta$. Therefore, we have that, for all $0 < \tau \leq \tau^{**}$, $\sup_{x\in K}|[\partial^{[j]}\psi](\varphi_\tau(x)) - [\partial^{[j]}\psi](\varphi(x))| \leq \epsilon$; that is, $[\partial^{[j]}\psi](\varphi_\tau(\cdot))$ converges uniformly on K to $[\partial^{[j]}\psi](\varphi(\cdot))$. Now, the result follows from (A.14), (A.15), and Lemma A.3 with $A = K$. ■

We now return to the proof of Theorem 2.2 over an additional parameter space Θ uniformly over an additional parameter space Θ.

Assumption A.1 for (2.12): *We need the following assumptions on $G_\theta(\cdot)$.*

*(**A1**) $G_\theta(\cdot)$ satisfies (**P1**)-(**P4**), where $k_{G_\theta} = k_{G_\Theta}$ is independent of θ.*
*(**A2**) $\mathcal{H}_{G_\Theta} = \cup_{\theta\in\Theta}\mathcal{H}_{G_\theta}$ is a VC-subgraph class.*
*(**A3**) $\sup_{\theta\in\Theta} G_\theta(\cdot) \leq l_{G,\Theta}$.*
*(**A4**) $G_{(\cdot)}(\cdot)$ is jointly Borel measurable.*

Proof of (2.12). The proof follows essentially as in Section 4 with minor changes. We include an argument for completeness. To this end, first note that

that $\mathcal{H}_{G_\theta} = \{h_\tau^{(G_\theta)}(x;\cdot) : x \in \mathbb{R}^p, \tau \in [0,\infty]\}$ and $\mathcal{H}_{G_\theta,1} \coloneqq \{h_\tau^{(G_\theta,1)}(x;\cdot) : x \in \mathbb{R}^p, \tau \in [0,\infty]\}$. We verify (i) and (ii) in the proof of Theorem 2.2 for the class $\mathcal{H}_{G_\Theta}$. For (i), notice that, in view of **(A3)**, both $\sup_{\theta\in\Theta}\sup_{h^{(G_\theta)}\in\mathcal{H}_{G_\theta}}|h^{(G_\theta)}(\cdot)|$ and $\sup_{\theta\in\Theta}\sup_{h^{(G_\theta,1)}\in\mathcal{H}_{G_\theta,1}}|h^{(G_\theta,1)}(\cdot)|$ are bounded above by $l_{G,\Theta}$. We now turn to (ii). Let $\mathfrak{i}_{G_\Theta} : [0,\infty]\times\Theta\times\mathbb{R}^p\times(\mathbb{R}^p)^{k_{G_\Theta}} \to \mathbb{R}$ be given by

$$\mathfrak{i}_{G_\Theta}(\tau;\theta;x;x_1,\ldots,x_{k_{G_\Theta}}) = h_\tau^{(G_\theta)}(x;x_1,\ldots,x_{k_{G_\Theta}})$$

and $F_{G_\Theta} : (0,\infty)\times\Theta\times\mathbb{R}^p\times(\mathbb{R}^p)^{k_{G_\Theta}} \to (\mathbb{R}^p)^{k_{G_\Theta}}$ be given by

$$F_{G_\Theta}(\tau;\theta;x;x_1,\ldots,x_{k_{G_\Theta}}) = \left(\theta, \frac{x_1-x}{\tau},\ldots,\frac{x_{k_{G_\Theta}}-x}{\tau}\right)^\top.$$

For simplicity, let $H_\tau^{(G_\Theta)}(\theta;\cdot,\cdot) = h_\tau^{(G_\theta)}(\cdot;\cdot)$, $\tau\in\{0,\infty\}$, and $G_*(\theta;\cdot) = G_\theta(\cdot)$, yielding $h_\tau^{(G_\theta)}(\cdot;\cdot) = G_*(F_{G_\Theta}(\tau;\theta;\cdot;\cdot)$, $\theta\in\Theta$, $\tau\in(0,\infty)$. It follows from **(A4)** that $G_*(F_{G_\Theta}(\cdot;\cdot;\cdot;\cdot))$, $H_0^{G_\Theta}(\cdot;\cdot,\cdot)$, and $H_\infty^{G_\Theta}(\cdot;\cdot,\cdot)$ are Borel measurable. Therefore, for all $A\in\mathscr{B}(\mathbb{R})$,

$$\begin{aligned}\mathfrak{i}_{G_\Theta}^{-1}(A) =&(F_{G_\Theta}^{-1}(G_*^{-1}(A))\cup(\{0\}\times(H_0^{(G_\Theta)})^{-1}(A))\cup(\{\infty\}\times(H_\infty^{(G_\Theta)})^{-1}(A))\\ &\in\mathscr{B}([0,\infty]\times\Theta\times\mathbb{R}^p\times(\mathbb{R}^p)^{k_{G_\Theta}}).\end{aligned}$$

We conclude that $\mathfrak{i}_{G_\Theta}(\cdot)$ is Borel measurable and the class $\mathcal{H}_{G_\Theta}$ is image admissible Suslin via $\mathfrak{e}_{G_\Theta} : [0,\infty]\times\Theta\times\mathbb{R}^p \to \mathcal{H}_{G_\Theta}$ given by $\mathfrak{e}_{G_\Theta}(\tau;\theta;x) = h_\tau^{(G_\theta)}(x;\cdot)$. ∎

Before we state Proposition A.1, which is used in the proof of Theorem 2.3, we recall that T is a subset of $\mathbb{R}^p\times[0,\infty]$ such that, for $(x,\tau)\in T$, $E[(\tilde{h}_\tau^{(1)}(x;X_1))^2] > 0$. For $m\geq 1$ and $(x_1,\tau_1),\ldots,(x_m,\tau_m)\in T$, we also use the notations

$$\boldsymbol{LGD_n}(x_l,\tau_l) \coloneqq (LGD_n(x_1,\tau_1),\ldots,LGD_n(x_m,\tau_m))^\top,$$

$$\boldsymbol{LGD}(x_l,\tau_l) \coloneqq (LGD(x_1,\tau_1),\ldots,LGD(x_m,\tau_m))^\top,$$

and, for $j=1,\ldots,k$ and $y_1,\ldots,y_j\in\mathbb{R}^p$,

$$\boldsymbol{h}_{\tau_l}^{(j)}(x_l;y_1,\ldots,y_j) \coloneqq (h_{\tau_1}^{(j)}(x_1;y_1,\ldots,y_j),\ldots,h_{\tau_m}^{(j)}(x_m;y_1,\ldots,y_j))^\top.$$

Proposition A.1 *For $(x_1,\tau_1),\ldots,(x_m,\tau_m)\in T$, $\sqrt{n}(\boldsymbol{LGD_n}(x_l,\tau_l)-\boldsymbol{LGD}(x_l,\tau_l))$ converges in distribution to a m-variate normal distribution with mean 0 and covariance matrix whose $(l_1,l_2)^{th}$ element is given by $k^2\gamma((x_{l_1},\tau_{l_1}),(x_{l_2},\tau_{l_2}))$, where $l_1,l_2=1,\ldots,m$.*

The proof is based on Hoeffding's decomposition of U-statistics and can be found in Francisci et al. (2022). An immediate consequence of Proposition A.1 is the following corollary.

Corollary A.1 *If $x\in\mathbb{R}^p$ and $\tau\in(0,\infty]$ satisfy $E[(\tilde{h}_\tau^{(1)}(x;X_1))^2]>0$, then*

$$\sqrt{n}(LGD_n(x,\tau)-LGD(x,\tau)) \xrightarrow[n\to\infty]{d} N(0,k^2E[(\tilde{h}_\tau^{(1)}(x;X_1))^2]). \tag{A.16}$$

Next, we state and prove a result concerning the quantity $D_G(\cdot,\cdot)$ in 2.13 that will be used for the proof of Proposition 2.3.

Lemma A.4 *Let $D_G(\cdot,\cdot)$, σ_G, and $C_{G,0}$ be as in Theorem 2.4, $\{a_n\}_{n=1}^{\infty}$ be a sequence of positive scalars converging to zero with $\lim_{n\to\infty}\frac{n}{\log(n)}a_n^2=\infty$, $b>0$, and $t_n \coloneqq \sqrt{n}a_n b$. Then, there are constants $0<\tilde{C}_G<\infty$ and $\tilde{n}\in\mathbb{N}$ such that, for all $n\geq\tilde{n}$, $t_n\geq\max(2^3\sigma_G,2^4C_{G,0})$ and*

$$D_G(n,t_n)\leq\frac{\tilde{C}_G}{n^2}.$$

Proof of Lemma A.4. 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 $n\geq n_1$, $t_n\geq\max(2^3\sigma_G,2^4C_{G,0})$ and $t_n/\sqrt{n}=a_nb\leq 1$. Then, for all $n\geq n_1$, it holds that

$$\begin{aligned}D_G(n,t_n)\leq{}& 8\exp\left(-\frac{t_n^2}{2^{15}k_G^2(\sigma_G^2+l_G)}\right)+2\exp\left(-\frac{t_n^2}{2^{6+k_G}k_G^{k_G+1}l_GC_{G,0}(\sigma_G^2+l_G)}\right)\\&+8C_{G,1}^{2C_{G,2}}(\sigma_G^2+2a_nbl_G)^{-C_{G,2}}\exp\left(-\left(\frac{n\sigma_G^2}{2l_G^2}+\frac{\sqrt{n}t_n}{4l_G}\right)\right)\\&\leq 16\exp\left(-\frac{t_n^2}{C_{G,3}}\right)+C_{G,4}a_n^{-C_{G,2}}\exp\left(-\frac{\sqrt{n}t_n}{C_{G,5}}\right),\end{aligned}$$

where

$$\begin{aligned}&C_{G,3}\coloneqq(\sigma_G^2+l_G)\max(2^{15}k_G^2,2^{6+k_G}k_G^{k_G+1}l_GC_{G,0}),\\&C_{G,4}\coloneqq 8C_{G,1}^{2C_{G,2}}(2bl_G)^{-C_{G,2}},\ \text{and}\ C_{G,5}\coloneqq 4l_G.\end{aligned}$$

Next, we use that $\lim_{n\to\infty}\frac{n}{\log(n)}a_n^2=\infty$ to show 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

$$\begin{aligned}n^2a_n^{-C_{G,2}}\exp\left(-\frac{\sqrt{n}t_n}{C_{G,5}}\right)&=\exp\left(2\log(n)-\frac{na_nb}{2C_{G,5}}\right)\exp\left(-C_{G,2}\log(a_n)-\frac{na_nb}{2C_{G,5}}\right)\\&=\exp\left(-\log(n)\left(\frac{b}{2C_{G,5}}\frac{na_n}{\log(n)}-2\right)\right)\\&\quad\exp\left(-\frac{b}{2C_{G,5}}na_n\left(1+\frac{2C_{G,2}C_{G,5}}{b}\frac{\log(a_n)}{na_n}\right)\right).\end{aligned}$$

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

$$\lim_{n\to\infty}n^2a_n^{-C_{G,2}}\exp\left(-\frac{\sqrt{n}t_n}{C_{G,5}}\right)=0.$$

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 $\tilde{n} = \max_{i=1,\dots,3} n_i$. Then, for all $n \geq \tilde{n}$, it holds that

$$D_G(n, t_n) \leq \frac{\tilde{C}_G}{n^2}, \text{ where } \tilde{C}_G \coloneqq 16 + C_{G,4}.$$

∎

Proof of Proposition 2.3. For (i), observe that

$$\sup_{x\in\mathbb{R}^p} |f^{(G)}_{\tau_n,n}(x) - f(x)| \leq \sup_{x\in\mathbb{R}^p} |f^{(G)}_{\tau_n,n}(x) - f^{(G)}_{\tau_n}(x)| + \sup_{x\in\mathbb{R}^p} |f^{(G)}_{\tau_n}(x) - f(x)|$$

and, by Proposition 2.2 (i), it is enough to show that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^p} |f^{(G)}_{\tau_n,n}(x) - f^{(G)}_{\tau_n}(x)| = 0 \text{ a.s.} \tag{A.17}$$

Now, using Lemma A.2, we see that

$$\sup_{x\in\mathbb{R}^p} |f^{(G)}_{\tau_n,n}(x) - f^{(G)}_{\tau_n}(x)| \leq \tau_n^{-p} \sup_{\substack{x\in\mathbb{R}^p \\ \tau\in[0,\infty]}} |LGD_n(x,\tau) - LGD(x,\tau)|^{1/k_G}.$$

Let $\epsilon > 0$, $t_n \coloneqq \sqrt{n}\tau_n^{k_G p}\epsilon^{k_G}$ and notice that, since $\lim_{n\to\infty} n\tau_n^{2k_G p} = \infty$, $\lim_{n\to\infty} t_n = \infty$. It follows from Theorem 2.4 and Lemma A.4 with $a_n = \tau_n^{k_G p}$ and $b = \epsilon^{k_G}$ 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

$$P^{\otimes n}\left(\sup_{x\in\mathbb{R}^p} |f^{(G)}_{\tau_n,n}(x) - f^{(G)}_{\tau_n}(x)| \geq \epsilon\right) \leq D_G(n, t_n) \leq \frac{\tilde{C}_G}{n^2}.$$

Therefore, we obtain that

$$\sum_{n=1}^{\infty} P^{\otimes n}\left(\sup_{x\in\mathbb{R}^p} |f^{(G)}_{\tau_n,n}(x) - f^{(G)}_{\tau_n}(x)| \geq \epsilon\right) \leq \tilde{n} - 1 + \sum_{n=\tilde{n}}^{\infty} \frac{\tilde{C}_G}{n^2} < \infty.$$

Now, (A.17) follows from Borel-Cantelli lemma. The proof of the first part of (ii) follows from the inequality

$$\sup_{x\in K} |f^{(G)}_{\tau_n,n}(x) - f(x)| \leq \sup_{x\in\mathbb{R}^p} |f^{(G)}_{\tau_n,n}(x) - f^{(G)}_{\tau_n}(x)| + \sup_{x\in K} |f^{(G)}_{\tau_n}(x) - f(x)|,$$

(A.17), and Proposition 2.2 (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 $x \in \mathbb{R}^p$,

$$\sup_{y\in\overline{B}_{\epsilon_n}(x)} |f^{(G)}_{\tau_n,n}(y) - f(x)| \leq \sup_{y\in\overline{B}_{\epsilon^*}(x)} |f^{(G)}_{\tau_n,n}(y) - f^{(G)}_{\tau_n}(y)| + \sup_{y\in\overline{B}_{\epsilon_n}(x)} |f^{(G)}_{\tau_n}(y) - f(x)|.$$

Now, using the compactness of $\overline{B}_{\epsilon^*}(x)$ and the first part of (ii), we have that

$$\lim_{n\to\infty} \sup_{y\in\overline{B}_{\epsilon^*}(x)} |f^{(G)}_{\tau_n,n}(y) - f^{(G)}_{\tau_n}(y)| = 0 \text{ a.s.}$$

Finally, Proposition 2.2 (ii) implies that

$$\lim_{n\to\infty} \sup_{y\in \overline{B}_{\epsilon_n}(x)} |f^{(G)}_{\tau_n}(y) - f(x)| = 0.$$

■

We next turn to the proof of Lemma 3.1.

Proof of Lemma 3.1. We first observe that $x \in S_{f_\tau}$ if and only if $f_\tau(x) > 0$ if and only if $LGD(x,\tau) > 0$. Proposition 2.1 (i) implies that for $x \in \mathbb{R}^p$, $LGD(x,\tau_1) \leq LGD(x,\tau_2)$, from which it follows that $S_{f_{\tau_1}} \subset S_{f_{\tau_2}}$. Next, suppose that $f(\cdot)$ is continuous and let $x \in S_f$ and $\tau > 0$. Since $f(\cdot)$ 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 $\lambda((\overline{B}_\delta(x))^k \cap S_{h_\tau(x;\cdot)}) > 0$ and $h_\tau(x;\cdot) \geq c$ in $(\overline{B}_\delta(x))^k \cap S_{h_\tau(x;\cdot)}$. It follows that

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

Thus $x \in S_{f_\tau}$ and $S_f \subset S_{f_\tau}$. Since the sets $\{S_{f_\tau}\}_{\tau>0}$ are monotonically decreasing with τ, we have that $\lim_{\tau\to 0^+} S_{f_\tau} = \cap_{\tau>0} S_{f_\tau} \supset S_f$. For the last part, let $x \in \mathbb{R}^p \setminus \overline{S}_f$. Since $\mathbb{R}^p \setminus \overline{S}_f$ is open, there exists $\epsilon > 0$ such that $\overline{B}_\epsilon(x) \subset \mathbb{R}^p \setminus \overline{S}_f$. Let $0 < \tau \leq \epsilon/\rho$. By (2.4) it follows that $S_{h_\tau(x;\cdot)} \subset (\overline{B}_{\rho\tau}(x))^k \subset (\overline{B}_\epsilon(x))^k$ implying that $LGD(x,\tau) = 0$. Therefore, $x \notin \cap_{\tau>0} S_{f_\tau}$ and $\cap_{\tau>0} S_{f_\tau} \subset \overline{S}_f$. ■

The next lemma is used in the proof of Theorem 3.1 (iii), Proposition A.2 (ii), and Lemma 3.2 (i) and provides a uniform approximation of $f_\tau(\cdot)$ in compact sets. The proof relies on Theorem 2.1 (iii) and is given in Francisci et al. (2022).

Lemma A.5 *Suppose (2.4) holds true and $f(\cdot)$ is three times continuously differentiable. Let K be a compact subset of S_f. Then, there are constants $\tau(K), c_1(K)$, $c_2(K) > 0$ and a continuously differentiable function $\tilde{R}_\tau : K \to \mathbb{R}$ such that, for all $x \in K$ and $0 < \tau \leq \tau(K)$, $|\tilde{R}_\tau(x)| \leq c_1(K)$, $\|\nabla \tilde{R}_\tau(x)\| \leq c_2(K)$, and*

$$f_\tau(x) = f(x) + \tilde{R}_\tau(x)\tau^2.$$

We now turn to the proof of Theorem 3.1. To this end, we introduce few additional notations. The norm of a $p \times p$ matrix A is given by $\|A\|_{\mathcal{M}} := \sup_{y\in\mathbb{R}^p, y\neq 0} \|Ay\|/\|y\|$ and the spectrum of A, that is, the set of all the eigenvalues of A is denoted by $\sigma(A)$. Finally, the sign function $\text{sgn} : \mathbb{R} \to \mathbb{R}$ is given by

$$\text{sgn}(t) = \begin{cases} -1 & \text{if } t < 0 \\ 0 & \text{if } t = 0 \\ 1 & \text{if } t > 0. \end{cases}$$

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 $f_\tau(\cdot)$ when τ-symmetry prevails. As for the proof part (ii), note that for each stationary point μ of $f(\cdot)$, first a closed hypercube centered at μ with directions given by the orthogonal eigenvalues of $H_f(\mu)$ is constructed. The side lengths of the hypercube are such that, for small enough τ and all points in the hypercube (a) the eigenvalues of $H_{f_\tau}(\cdot)$ and $H_f(\cdot)$ corresponding to the same eigenvector have the same sign and (b) 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 $f_\tau(\cdot)$ to those of $f(\cdot)$. Now, (b) 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 $f_\tau(\cdot)$. Next, using (a), 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 Lemma A.5 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 3.1. We start by proving (i). Notice that, if $f(\cdot)$ is continuously differentiable in $\overline{B}_{\rho\tau}(x) \subset S_f$, then, for $j = 1, \ldots, p$,

$$\partial_j f_\tau(x) = \frac{1}{k}(f_\tau(x))^{1-k}\frac{\partial_j LGD(x,\tau)}{\tau^{kp}}, \tag{A.18}$$

where, by Proposition 2.1, (A.2) and (A.4),

$$\partial_j LGD(x,\tau) = k\int h_\tau(0; x_1, \ldots, x_k)\partial_j f(x+x_1)f(x+x_2)\ldots f(x+x_k)dx_1\ldots dx_k.$$

Hence, $\partial_j f_\tau(\mu) = 0$ if and only if

$$\int h_\tau(0; x_1, \ldots, x_k)\partial_j f(\mu + x_1)f(\mu + x_2)\ldots f(\mu + x_k)dx_1\ldots dx_k = 0, \tag{A.19}$$

and hence (3.5) holds. We next turn to the proof of (ii). Since $H_f(\mu)$ is symmetric, it has orthonormal eigenvectors v_i associated with eigenvalues λ_i, $i = 1, \ldots, p$. Notice that, since μ is of type l, l eigenvalues are negative and $p-l$ are positive. In particular,

$$\min_{i=1,\ldots,p}|\lambda_i| > 0. \tag{A.20}$$

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 $f_\tau(\cdot)$ is twice continuously differentiable in $\overline{B}_{\tau_1}(x)$ and its first order partial derivatives are given by (A.18). By uniform continuity of the second order partial derivatives of $f(\cdot)$ in $\overline{B}_\delta(\mu)$ and Proposition 2.2 (iv), it follows that, for $i, j = 1, \ldots, p$,

$$\sup_{y \in \overline{B}_\delta(\mu)} |\partial_i\partial_j f(y) - \partial_i\partial_j f(\mu)| \xrightarrow[\delta \to 0^+]{} 0. \tag{A.21}$$

and, for $0 < \tilde{\tau}_1, \tilde{\tau}_2 \leq \tau_1$,

$$\sup_{0<\tau\leq\tilde{\tau}_2} \sup_{y\in\overline{B}_{\tilde{\tau}_1}(\mu)} |\partial_i\partial_j f_\tau(y) - \partial_i\partial_j f(y)| \leq \sup_{0<\tau\leq\tilde{\tau}_2} \sup_{y\in\overline{B}_{\tau_1}(\mu)} |\partial_i\partial_j f_\tau(y) - \partial_i\partial_j f(y)| + \sup_{y\in\overline{B}_{\tilde{\tau}_1}(\mu)} |\partial_i\partial_j f(y) - \partial_i\partial_j f(y)| \xrightarrow[\tilde{\tau}_1,\tilde{\tau}_2\to 0^+]{} 0. \tag{A.22}$$

For $y_1, \dots, y_p \in \overline{B}_\delta(0)$, let

$$H_f(\mu; y_1, \dots, y_p) \coloneqq \begin{pmatrix} (\nabla\partial_1 f(x+y_1)))^\top \\ \vdots \\ (\nabla\partial_p f(x+y_p))^\top \end{pmatrix}^\top$$

and, similarly with f replaced by f_τ and $y_1, \dots, y_p \in \overline{B}_{\tau_1}(0)$. (A.21) and (A.22) show that,

$$\sup_{y_1,\dots,y_p\in\overline{B}_\delta(0)} \|H_f(\mu; y_1, \dots, y_p) - H_f(\mu)\|_{\mathcal{M}} \xrightarrow[\delta\to 0^+]{} 0, \tag{A.23}$$

$$\sup_{0<\tau\leq\tilde{\tau}_2} \sup_{y_1,\dots,y_p\in\overline{B}_{\tilde{\tau}_1}(0)} \|H_{f_\tau}(\mu; y_1, \dots, y_p) - H_f(\mu)\|_{\mathcal{M}} \xrightarrow[\tilde{\tau}_1,\tilde{\tau}_2\to 0^+]{} 0. \tag{A.24}$$

In particular, (A.23) implies that, for $i = 1, \dots, p$,

$$\sup_{y_1,\dots,y_p\in\overline{B}_\delta(0)} \|H_f(\mu; y_1, \dots, y_p)v_i - \lambda_i v_i\| \xrightarrow[\delta\to 0^+]{} 0.$$

and, for $t_i \in \mathbb{R}$,

$$\sup_{y_1,\dots,y_p\in\overline{B}_\delta(0)} \left| \langle H_f(\mu; y_1, \dots, y_p)\Big(v_i + \sum_{j=1,j\neq i}^{p} t_j v_j\Big), v_i\rangle - \lambda_i \right| \xrightarrow[\delta\to 0^+]{} 0.$$

By (A.20), there exists $0 < \delta_2 \leq \delta$ such that, for $i = 1, \dots, p$,

$$\operatorname{sgn}\Big(\langle H_f(\mu; y_1, \dots, y_p)\Big(v_i + \sum_{j=1,j\neq i}^{p} t_j v_j\Big), v_i\rangle\Big) = \operatorname{sgn}(\lambda_i), \tag{A.25}$$

for all $y_1, \dots, y_p \in \overline{B}_{\delta_2}(0)$. Similarly, using (A.24), one can show that

$$\sup_{0<\tau\leq\tilde{\tau}_2} \sup_{y_1,\dots,y_p\in\overline{B}_{\tilde{\tau}_1}(0)} |\langle H_{f_\tau}(\mu; y_1, \dots, y_p)v_i, v_i\rangle - \lambda_i| \xrightarrow[\tilde{\tau}_1,\tilde{\tau}_2\to 0^+]{} 0. \tag{A.26}$$

Moreover, by Bauer–Fike theorem (Theorem 2.1 in Eisenstat and Ipsen (1998)), for all $\tilde{\lambda}_\tau(\mu; y_1, \dots, y_p) \in \sigma(H_{f_\tau}(\mu; y_1, \dots, y_p))$, we have that

$$\min_{i=1,\dots,p} |\tilde{\lambda}_\tau(\mu; y_1, \dots, y_p) - \lambda_i| \leq \|H_{f_\tau}(\mu; y_1, \dots, y_p) - H_f(\mu)\|_{\mathcal{M}}. \tag{A.27}$$

By (A.20), (A.26), (A.27) and (A.24), it follows that, there exists $0 < \tau_2 \leq \tau_1$ such that, for all $0 < \tau \leq \tau_2$ and $y_1, \ldots, y_p \in \overline{B}_{\tau_2}(0)$,

$$\operatorname{sgn}(\langle H_{f_\tau}(\mu; y_1, \ldots, y_p) v_i, v_i \rangle) = \operatorname{sgn}(\lambda_i) \tag{A.28}$$

and $\sigma(H_{f_\tau}(\mu; y_1, \ldots, y_p)) = \{\tilde{\lambda}_{\tau,1}(\mu; y_1, \ldots, y_p), \ldots, \tilde{\lambda}_{\tau,p}(\mu; y_1, \ldots, y_p)\}$ with

$$\operatorname{sgn}(\tilde{\lambda}_{\tau,i}(\mu; y_1, \ldots, y_p)) = \operatorname{sgn}(\lambda_i). \tag{A.29}$$

Now, let $0 < \tau \leq \tau_2$, $0 < h \leq h^*$, where $h^* \coloneqq \min(\delta_2, \tau_2)/(2\sqrt{p})$, and $t_i \in [-2h, 2h]$. By the mean value theorem, there exist $0 \leq c_{i,j} \leq 1$ such that

$$\nabla f(\mu \pm h v_i + \sum_{j=1, j \neq i} t_j v_j) = H_f(\mu; y_1, \ldots, y_p) \Big(\pm h v_i + \sum_{j=1, j \neq i}^{p} t_j v_j \Big),$$

where $y_j = c_{i,j}(\pm h v_i + \sum_{j=1, j \neq i} t_j v_j)$, implying that

$$\frac{1}{h} \langle \nabla f(\mu \pm h v_i + \sum_{j=1, j \neq i} t_j v_j), v_i \rangle = \pm \langle H_f(\mu; y_1, \ldots, y_p) \Big(v_i \pm \sum_{j=1, j \neq i}^{p} t_j / h v_j \Big), v_i \rangle.$$

Since $\|y_j\| \leq 2\sqrt{p} h^* \leq \delta_2$, by (A.25),

$$\operatorname{sgn}\Big(\langle \nabla f(\mu \pm h v_i + \sum_{j=1, j \neq i} t_j v_j), v_i \rangle \Big) = \operatorname{sgn}(\pm \lambda_i). \tag{A.30}$$

Now, let us define the hypercube $F_{h^*}(\mu)$ with center μ by

$$F_{h^*}(\mu) \coloneqq \Big\{ \mu + \sum_{j=1}^{p} t_j v_j, t_j \in [-3/4 h^*, 3/4 h^*] \Big\}$$

and its "hyperfaces" by

$$F^{\pm}_{h^*, i}(\mu) \coloneqq \Big\{ \mu \pm 3/4 h^* v_i + \sum_{j=1, j \neq i}^{p} t_j v_j, t_j \in [-3/4 h^*, 3/4 h^*] \Big\}.$$

Since, by (2.4), for $0 < \tau \leq \tau^*$, where $\tau^* \coloneqq \min(\tau_2, h^*/(4\rho))$,

$$\overline{S}_{h_\tau(0; \cdot, x_2, \ldots, x_k)} \subset \overline{B}_{\rho\tau}(0) \subset \Big\{ \sum_{j=1}^{p} s_j v_j : s_j \in [-h^*/4, h^*/4] \Big\},$$

we have that, for $\mu_i^{\pm} \in F^{\pm}_{h^*, i}(\mu)$ and $x_1 \in \overline{S}_{h_\tau(0; \cdot, x_2, \ldots, x_k)}$,

$$\mu_i^{\pm} + x_1 \in \mu + \Big\{ \pm h v_i + \sum_{j=1, j \neq i}^{p} s_j v_j : h \in [h^*/2, h^*], s_j \in [-h^*, h^*] \Big\}.$$

Now, by (A.30),

$$\mathrm{sgn}(\langle \nabla f(\mu_i^{\pm} + x_1), v_i\rangle) = \mathrm{sgn}(\pm\lambda_i),$$

for all $x_1 \in \overline{S}_{h_\tau(0;\cdot,x_2,\ldots,x_k)}$ and $t_j \in [-3/4h^*, 3/4h^*]$. It follows from (A.18) that $\mathrm{sgn}(\langle \nabla f_\tau(\mu_i^{\pm}), v_i\rangle) = \mathrm{sgn}(\pm\lambda_i)$. In particular, for all $\mu_i^+ \in F_{h^*,i}^+(\mu)$ and $\mu_i^- \in F_{h^*,i}^-(\mu)$,

$$\mathrm{sgn}(\langle \nabla f_\tau(\mu_i^+), v_i\rangle) = -\,\mathrm{sgn}(\langle \nabla f_\tau(\mu_i^-), v_i\rangle) \neq 0. \tag{A.31}$$

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 $\alpha_i(y,t) = y - 3/2h^* t v_i$. Since $\nabla f_\tau(\cdot)$ is continuous, by (A.31), 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

$$\nabla f_\tau(\alpha_i(\mu_i^+, t_2)) = \nabla f_\tau(\alpha_i(\mu_i^+, t_1)) + H_{f_\tau}(\mu; y_1, \ldots, y_p)^\top(\alpha_i(\mu_i^+, t_2) - \alpha_i(\mu_i^+, t_1))),$$

where $y_j = (1 - c_j)\alpha_i(\mu_i^+, t_2) + c_j\alpha_i(\mu_i^+, t_1) - \mu$, implying that

$$3/2h^*(t_2 - t_1)\langle H_{f_\tau}(\mu; y_1, \ldots, y_p)v_i, v_i\rangle = 0.$$

By (A.28), it follows that $t_2 = t_1$. Let, for $i = 1, \ldots, p$,

$$Z_{\tau,i}(\mu) \coloneqq \{\alpha_i(y,t) : \langle \nabla f_\tau(\alpha_i(y,t)), v_i\rangle = 0, y \in F_i^+(\mu), t \in [0,1]\}.$$

Notice that $Z_{\tau,i}(\mu)$ are closed subsets of the hypercube $F_{h^*}(\mu)$ with dimension $p-1$ that 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 $\cap_{i=1}^p Z_{\tau,i}(\mu) = \{\mu_\tau\}$, where μ_τ satisfies $\langle \nabla f_\tau(\mu_\tau), v_i\rangle = 0$, for all $i = 1, \ldots, p$, implying that $\nabla f_\tau(\mu_\tau) = 0$. Finally, by (A.29) and $\|\mu_\tau - \mu\| \le 3/4\sqrt{p}h^* \le \tau_2$, it follows that μ_τ is of type l. Also, by letting $\tau_2 \to 0^+$, we see that $\|\mu_\tau - \mu\| \to 0$.
Finally, we prove (iii). Since $H_f(\mu)^{-1}$ is symmetric, it holds that

$$\xi \coloneqq \|H_f(\mu)^{-1}\|_{\mathcal{M}} = \max_{i=1,\ldots,p} 1/|\lambda_i| > 0.$$

By (A.24), there exists $0 < \tau_3 \le \tau_1$, such that, for all $0 < \tau \le \tau_3$ and $y_j \in \overline{B}_{\tau_3}(0)$, $j = 1, \ldots, p$,

$$\|H_{f_\tau}(\mu; y_1, \ldots, y_p) - H_f(\mu)\|_{\mathcal{M}} \le 1/(2\xi). \tag{A.32}$$

It follows from (A.32) and the triangle inequality that, for all $v \in \mathbb{R}^p$,

$$\|v\| \le 2\xi(\|H_f(\mu)v\| - 1/(2\xi)\|v\|) \le 2\xi\|H_{f_\tau}(\mu; y_1, \ldots, y_p)v\|.$$

By setting $w = H_{f_\tau}(\mu; y_1, \ldots, y_p)v$, we see that $\|w\|$ is bounded below by $1/(2\xi)\|H_{f_\tau}(\mu; y_1, \ldots, y_p)w\|$ implying that

$$\|H_{f_\tau}(\mu; y_1, \ldots, y_p)^{-1}\|_{\mathcal{M}} \le 2\xi. \tag{A.33}$$

Moreover, by the mean value theorem, there exist $0 \leq \tilde{c}_j \leq 1$, $j = 1, \ldots, p$, such that,

$$\nabla f_\tau(\mu) = \nabla f_\tau(\mu) - \nabla f_\tau(\mu_\tau) = H_{f_\tau}(\mu; y_1, \ldots, y_p)(\mu - \mu_\tau),$$

where $y_j = (1 - \tilde{c}_j)(\mu - \mu_\tau)$. Since $\|y_j\| \leq \|\mu - \mu_\tau\| \leq \tau_2$, using (A.29) we see that $H_{f_\tau}(\mu; y_1, \ldots, y_p)$ is invertible. We now apply Lemma A.5 with $K = \overline{B}_\delta(\mu)$ and get constants $\tau(K), c_2(K) > 0$ such that, for all $y \in K$ and $0 < \tau \leq \min(\tau_2, \tau(K))$,

$$\|\nabla f_\tau(y) - \nabla f(y)\| \leq c_2(K)\tau^2. \tag{A.34}$$

Using (A.33) and (A.34), we conclude that, for all $0 < \tau \leq \min(\tau_2, \tau(K))$,

$$\|\mu - \mu_\tau\| \leq \|H_{f_\tau}(\mu; y_1, \ldots, y_p)^{-1}\|_{\mathcal{M}} \|\nabla f_\tau(\mu) - \nabla f(\mu)\| \leq 2\xi c_2(K)\tau^2.$$

■

We study next the relationship between the gradient systems (3.4) and (3.2) under extreme localization. To this aim, notice that the sets $\{S_{f_\tau}\}_{\tau>0}$ contain S_f by Lemma 3.1. Furthermore, because of Remark 2.1 and Proposition 2.2 (iv), $f_\tau(\cdot)$ is twice continuously differentiable in S_{f_τ} and its gradient and Hessian matrix converge to those of $f(\cdot)$ in S_f. The next lemma is used to show the existence of the solution $u_{x,\tau}(t)$ of (3.4) with initial condition $u_{x,\tau}(0) = x$. The proof involves standard analysis arguments and hence it is omitted. For more details see also Francisci et al. (2022).

Lemma A.6 *Under assumption (2.4), $(R^\alpha)^{-\rho\tau} \subset R^\alpha_\tau \subset (R^\alpha)^{+\rho\tau}$, for all $\tau > 0$ and $\alpha > 0$. In particular, if R^α is bounded for $\alpha > 0$, then R^α_τ is also bounded for any $\tau > 0$.*

The next proposition is required in the proof of Theorem 3.2 and its proof is based on Proposition 2.1 (iv), Grönwall's inequality, and Lemma A.5. The interested reader can refer to Francisci et al. (2022).

Proposition A.2 *Suppose that (2.4) holds true. (i) If $f(\cdot)$ is continuously differentiable in $\mathbb{R}^p$ and, for all $\alpha > 0$, R^α is compact, then, for all $t \geq 0$ and $x \in S_f$,*

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

(ii) If, additionally, $f(\cdot)$ is three times continuously differentiable, then, for $x \in S_f$,

$$\lim_{\tau \to 0^+} \sup_{t \in [0,\infty)} \|u_{x,\tau}(t) - u_x(t)\| = 0.$$

We now prove the convergence of clusters based on τ-approximation to that based on $f(\cdot)$.

Proof of Theorem 3.2. Let $\alpha \coloneqq \min_{\nu \in N_f} f(\nu)/2$, $\delta \coloneqq \operatorname{dist}(R^{2\alpha}, \mathbb{R}^p \backslash R^\alpha)/(1+\rho)$, $\{\alpha_n\}_{n=1}^\infty$ be a sequences of positive scalars converging monotonically to 0 with $\alpha_1 < \alpha$ and $\delta_n \coloneqq \min(\operatorname{dist}(R^{2\alpha}, \mathbb{R}^p \setminus R^\alpha), \operatorname{dist}(R^{\alpha_n}, \mathbb{R}^p \setminus S_f))/(1+\rho)$. We see that

$$N_f \subset R^{2\alpha} \subset (R^\alpha)^{-\delta} \subset (R^\alpha)^{-\delta_n} \subset (R^{\alpha_n})^{-\delta_n}, \tag{A.35}$$

with

$$\text{dist}(N_f, \mathbb{R}^p \setminus (R^{\alpha_n})^{-\delta_n}) \geq \text{dist}(R^{2\alpha}, (\mathbb{R}^p \setminus R^{\alpha})^{+\delta}) \geq \delta \geq \delta_n. \tag{A.36}$$

Furthermore, by Lemma A.6, for $0 < \tau \leq \delta_n/\rho$,

$$(R^{\alpha_n})^{-\delta_n} \subset (R^{\alpha_n})^{-\rho\tau} \subset R^{\alpha_n}_{\tau} \subset (R^{\alpha_n})^{+\rho\tau} \subset (R^{\alpha_n})^{+\delta_n} \subset S_f. \tag{A.37}$$

We notice that, by Assumption 3.1, $(R^{\alpha_n})^{-\delta_n}$ is bounded. Moreover, by Lemma 3.1 and Remark 2.1, $f_\tau(\cdot)$ is twice continuously differentiable in $S_f \subset S_{f_\tau}$. Now, by Theorem 3.1 (ii), there exist $h^*, \tau^* > 0$ and closed hypercubes $F_{h^*}(\mu)$, $\mu \in N_f$, with side length $3/2h^*$, such that, for $0 < \tau_j \leq \tau^*$, $f_{\tau_j}(\cdot)$ has a unique stationary point μ_{τ_j} in $\mathring{F}_{h^*}(\mu)$ and μ_{τ_j} is, for $\tau_j \leq \tau^*$, of the same type as μ, and $\lim_{j\to\infty}\|\mu_{\tau_j} - \mu\| = 0$. We can suppose without loss of generality that $3/2h^* \leq \delta/\sqrt{p}$, that is $F_{h^*}(\mu) \subset \overline{B}_\delta(\mu)$. By (A.35) and (A.36), it follows that $F_{h^*}(\mu) \subset (R^{\alpha_n})^{-\delta_n}$ and $K_n \coloneqq (R^{\alpha_n})^{+\delta_n} \setminus \cup_{\nu\in N_f}\mathring{F}_{h^*}(\nu)$ is compact. Let $\eta_n \coloneqq \min_{y\in K_n}\|\nabla f(y)\| > 0$. By Proposition 2.2 (iv), there exists $0 < \tau_n^* \leq \min(\tau^*, \delta_n/\rho)$ such that $\|\nabla f_\tau(y) - \nabla f(y)\| < \eta_n$, for all $y \in (R^{\alpha_n})^{+\delta_n}$ and $0 < \tau \leq \tau_n^*$. Hence,

$$\|\nabla f_\tau(y)\| \geq \|\nabla f(y)\| - \|\nabla f_\tau(y) - \nabla f(y)\| > 0.$$

It follows that $\{\nu_{\tau_j}\}_{\nu\in N_f}$ are the only stationary points of $f_{\tau_j}(\cdot)$ in $(R^{\alpha_n})^{+\delta_n}$. Now, by (A.37), $(R^{\alpha_n})^{-\delta_n} \subset R^{\alpha_n}_{\tau_j} \subset (R^{\alpha_n})^{+\delta_n}$, which implies that the solutions of (3.4) starting in $(R^{\alpha_n})^{-\delta_n}$ cannot leave the set to reach another possible stationary point of $f_{\tau_j}(\cdot)$ outside $R^{\alpha_n}_{\tau_j}$. Therefore, for $0 < \tau_j \leq \tau_n^*$, we can partition $(R^{\alpha_n})^{-\delta_n}$ as

$$\cup_{\nu\in N_f}(C(\nu) \cap (R^{\alpha_n})^{-\delta_n}) = (R^{\alpha_n})^{-\delta_n} = \cup_{\nu\in N_f}(C_{\tau_j}(\nu_{\tau_j}) \cap (R^{\alpha_n})^{-\delta_n}). \tag{A.38}$$

Next, we show that $(R^{\alpha_n})^{-\delta_n} \uparrow_{n\to\infty} S_f$. To this end, let $x \in S_f$. Clearly, $x \in R^{f(x)} \subset \mathring{R}^{f(x)/2}$. Since $\alpha_n, \delta_n \xrightarrow[n\to\infty]{} 0$, there exists n^* such that, for all $n \geq n^*$, $\alpha_n < f(x)/2$ and $\delta_n < \text{dist}(R^{f(x)}, \mathbb{R}^p \setminus \mathring{R}^{f(x)/2})/2$. Then, $x \in (R^{f(x)/2})^{-\delta_n} \subset (R^{\alpha_n})^{-\delta_n}$. We recall that the symmetric difference between two subsets A and B of $\mathbb{R}^p$ is $A\Delta B = ((\mathbb{R}^p \setminus A) \cap B) \cup (A \cap (\mathbb{R}^p \setminus B))$. For $\mu \in N_f$, it holds that

$$\begin{aligned}\limsup_{j\to\infty} C_{\tau_j}(\mu_{\tau_j})\Delta C(\mu) &= (\lim_{n\to\infty}(R^{\alpha_n})^{-\delta_n}) \cap (\limsup_{j\to\infty} C_{\tau_j}(\mu_{\tau_j})\Delta C(\mu)) \\ &= \lim_{n\to\infty}((R^{\alpha_n})^{-\delta_n} \cap (\limsup_{j\to\infty} C_{\tau_j}(\mu_{\tau_j})\Delta C(\mu))).\end{aligned}$$

Using (A.38), we have that $(R^{\alpha_n})^{-\delta_n} \cap (\limsup_{j\to\infty} C_{\tau_j}(\mu_{\tau_j})\Delta C(\mu))$ is a subset of

$$(R^{\alpha_n})^{-\delta_n} \cap (\cap_{j=1,\tau_j\leq\tau_n^*}^{\infty} \cup_{l=j}^{\infty} C_{\tau_l}(\mu_{\tau_j})\Delta C(\mu)),$$

which is equal to

$$\begin{aligned}&((R^{\alpha_n})^{-\delta_n} \cap (\cap_{j=1,\tau_j\leq\tau_n^*}^{\infty} \cup_{l=j}^{\infty} C_{\tau_l}(\mu)) \cap (\cup_{\nu\in N_f,\nu\neq\mu} C(\nu))) \\ &\cup((R^{\alpha_n})^{-\delta_n} \cap C(\mu) \cap (\cup_{\nu\in N_f,\nu\neq\mu}(\cap_{j=1,\tau_j\leq\tau_n^*} \cup_{l=j}^{\infty} C_{\tau_l}(\nu_{\tau_l})))).\end{aligned}$$

The above union is contained in

$$(R^{\alpha_n})^{-\delta_n} \cap (\cup_{\mu \in N_f} \cup_{\nu \in N_f, \nu \neq \mu} C(\mu) \cap (\cap_{j=1, \tau_j \leq \tau_n^*} \cup_{l=j}^{\infty} C_{\tau_l}(\nu_{\tau_l}))).$$

It follows that $\limsup_{j\to\infty} C_{\tau_j}(\mu_{\tau_j}) \Delta C(\mu)$ is contained in

$$\lim_{n\to\infty} (R^{\alpha_n})^{-\delta_n} \cap (\cup_{\mu \in N_f} \cup_{\nu \in N_f, \nu \neq \mu} C(\mu) \cap (\cap_{j=1, \tau_j \leq \tau_n^*} \cup_{l=j}^{\infty} C_{\tau_l}(\nu_{\tau_l}))).$$

Now, this is equal to

$$\cup_{\mu \in N_f} \cup_{\nu \in N_f, \nu \neq \mu} \lim_{n\to\infty} ((R^{\alpha_n})^{-\delta_n} \cap C(\mu) \cap (\cap_{j=1, \tau_j \leq \tau_n^*} \cup_{l=j}^{\infty} C_{\tau_l}(\nu_{\tau_l}))).$$

Next, let

$$x \in (R^{\alpha_n})^{-\delta_n} \cap C(\mu) \cap (\cap_{j=1, \tau_j \leq \tau_n^*} \cup_{l=j}^{\infty} C_{\tau_l}(\nu_{\tau_l})).$$

Then, there exists a subsequence $\{\tilde{\tau}_j\}_{j=1}^{\infty}$ of $\{\tau_j\}_{j=1}^{\infty}$ such that $\lim_{t\to\infty} u_{x,\tilde{\tau}_j}(t) = \nu_{\tilde{\tau}_j}$. In particular, $\lim_{j\to\infty} \lim_{t\to\infty} u_{x,\tilde{\tau}_j}(t) = \nu$. On the other hand, by Proposition A.2 (ii), $u_{x,\tilde{\tau}_j}(\cdot)$ converges uniformly on $[0,\infty)$ to $u_x(\cdot)$, as $j \to \infty$. Therefore,

$$\lim_{t\to\infty} \lim_{j\to\infty} u_{x,\tilde{\tau}_j}(t) = \lim_{t\to\infty} u_x(t) = \mu.$$

By Moore-Osgood theorem (see Theorem 7.11 in Rudin (1976)), it follows that $\nu = \lim_{j\to\infty} \lim_{t\to\infty} u_{x,\tilde{\tau}_j}(t) = \lim_{t\to\infty} \lim_{j\to\infty} u_{x,\tilde{\tau}_j}(t) = \mu$. We conclude that

$$\cup_{\mu \in N_f} \cup_{\nu \in N_f, \nu \neq \mu} \lim_{n\to\infty} ((R^{\alpha_n})^{-\delta_n} \cap C(\mu) \cap (\cap_{j=1, \tau_j \leq \tau_n^*} \cup_{l=j}^{\infty} C_{\tau_l}(\nu_{\tau_l}))) = \emptyset,$$

implying that $\lim_{j\to\infty} C_{\tau_j}(\mu_{\tau_j}) \Delta C(\mu) = \limsup_{j\to\infty} C_{\tau_j}(\mu_{\tau_j}) \Delta C(\mu) = \emptyset$. Finally, the equivalence

$$\lim_{j\to\infty} A_j = A \text{ if and only if } \lim_{j\to\infty} (A_j \Delta A) = \emptyset,$$

with $A_j = C_{\tau_j}(\mu_{\tau_j})$ and $A = C(\mu)$ implies that $\lim_{j\to\infty} C_{\tau_j}(\mu_{\tau_j}) = C(\mu)$. ∎

Proof of Theorem 3.3. 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 $f_{\tau_n}(\cdot)$ is continuously differentiable in $(K)^{+h^*}$ (see Remark 2.1). By the mean value theorem, there exist $0 \leq c_{1,n}, c_{2,n} \leq 1$ such that

$$f(x + h_n v_n) - f(x) = h_n \langle \nabla f(x + c_{1,n} h_n v_n), v_n \rangle \tag{A.39}$$

and

$$f_{\tau_n}(x + h_n v_n) - f_{\tau_n}(x) = h_n \langle \nabla f_{\tau_n}(x + c_{2,n} h_n v_n), v_n \rangle. \tag{A.40}$$

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 (A.39), the uniform continuity of $\nabla f(\cdot)$ in $(K)^{+h^*}$ and $\lim_{n\to\infty}\|v_n - v\| = 0$, it can be shown that

$$\lim_{n\to\infty} \sup_{x\in K} |\nabla_{v_n}^{h_n} f(x) - \nabla_v f(x)| \le \sup_{y\in(K)^{+h^*}} \|\nabla f(y)\| \lim_{n\to\infty} \|v_n - v\| + \lim_{n\to\infty} \sup_{x\in K} \|\nabla f(x + c_{1,n} h_n v_n) - \nabla f(x)\| = 0.$$

Also, by (A.39) and (A.40), it holds that

$$\sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n}(x) - \nabla_{v_n}^{h_n} f(x)| \le \sup_{x\in K} \|\nabla f_{\tau_n}(x + c_{2,n} h_n v_n) - \nabla f(x + c_{1,n} h_n v_n)\|.$$

Finally, Proposition 2.2 (iv) and the uniform continuity of $\nabla f(\cdot)$ in $(K)^{+h^*}$ imply that

$$\lim_{n\to\infty} \sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n}(x) - \nabla_{v_n}^{h_n} f(x)| \le \lim_{n\to\infty} \sup_{y\in(K)^{+h^*}} \|\nabla f_{\tau_n}(y) - \nabla f(y)\| + \sup_{y\in(K)^{+h^*}} \lim_{n\to\infty} \sup_{z\in \overline{B}_{h_n}(y)\cap(K)^{+h^*}} \|\nabla f(y) - \nabla f(z)\| = 0.$$

We now prove (ii). By (i) and triangle inequality, it is enough to show that

$$\lim_{n\to\infty} P^{\otimes n}\left(\sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n,n}(x) - \nabla_{v_n}^{h_n} f_{\tau_n}(x)| \ge \frac{\epsilon}{2} \right) = 0.$$

Notice that, by Lemma A.2, $\sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n,n}(x) - \nabla_{v_n}^{h_n} f_{\tau_n}(x)|$ is bounded above by

$$\sup_{x\in K} \left| \frac{LGD_n(x + h_n v_n, \tau_n) - LGD(x + h_n v_n, \tau_n)}{h_n^k \tau_n^{kp}} \right|^{1/k} + \sup_{x\in K} \left| \frac{LGD_n(x, \tau_n) - LGD(x, \tau_n)}{h_n^k \tau_n^{kp}} \right|^{1/k}.$$

We now use that $\lim_{n\to\infty} \sqrt{n} h_n^k \tau_n^{kp} = \infty$ and apply Theorem 2.4 with $t = t_n := \sqrt{n} h_n^k \tau_n^{kp} (\epsilon/4)^k$. It follows that there are constants $\sigma_G \ge 0$, $1 < C_{G,0}, C_{G,1}, C_{G,2} < \infty$, and $n^{**} \in \mathbb{N}$ such that, for all $n \ge n^{**}$, $t_n \ge \max(2^3\sigma_G, 2^4 C_{G,0})$ and

$$\begin{aligned} &P^{\otimes n}\left(\sup_{x\in K} |\nabla_{v_n}^{h_n} f_{\tau_n,n}(x) - \nabla_{v_n}^{h_n} f_{\tau_n}(x)| > \frac{\epsilon}{2} \right) \\ \le &P^{\otimes n}\left(\sqrt{n} \sup_{\substack{x\in\mathbb{R}^p \\ \tau\in[0,\infty]}} |LGD_n(x,\tau) - LGD(x,\tau)| \ge t_n \right) \le D_G(n, t_n), \end{aligned}$$

where $D_G(\cdot,\cdot)$ is defined in 2.13. Now, the result follows from $\lim_{n\to\infty} D_G(n, t_n) = 0$. ∎

Proof of Lemma 3.2. We begin by proving (i). By Lemma A.5 there are constants $\tau((K)^{+h^*}), c_2((K)^{+h^*}) > 0$ such that, for all $y \in (K)^{+h^*}$ and $0 < \tau \le$

$\tau((K)^{+h^*})$, $f_\tau(y) = f(y) + \tilde{R}_\tau(y)\tau^2$ and $\|\nabla \tilde{R}_\tau(y)\| \leq c_2((K)^{+h^*})$. Let $n^* \in \mathbb{N}$ such that $\tau_n \leq \tau((K)^{+h^*})$ for all $n \geq n^*$. It holds that, for all $n \geq n^*$,

$$\nabla_v^h f_{\tau_n}(x) - \nabla_v^h f(x) = \frac{\tilde{R}_{\tau_n}(x+hv) - \tilde{R}_{\tau_n}(x)}{h}\tau_n^2.$$

Now, by the mean value theorem, there are constants $0 \leq \tilde{c}_n \leq 1$ such that

$$\left|\frac{\tilde{R}_{\tau_n}(x+hv) - \tilde{R}_{\tau_n}(x)}{h}\right| \leq \|\nabla \tilde{R}_{\tau_n}(x + \tilde{c}_n hv)\| \leq c_2((K)^{+h^*}).$$

It follows that

$$\lim_{n\to\infty} \sup_{h\in[h_n,h^*]} \sup_{v\in S^{p-1}} \sup_{x\in K} |\nabla_v^h f_{\tau_n}(x) - \nabla_v^h f(x)| \leq c_2((K)^{+h^*}) \lim_{n\to\infty} \tau_n^2 = 0.$$

We now prove (ii). By (i), it is enough to show that

$$\lim_{n\to\infty} P^{\otimes n}\left(\sup_{h\in[h_n,h^*]} \sup_{v\in S^{p-1}} \sup_{x\in K} |\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f_{\tau_n}(x)| \geq \frac{\epsilon}{2}\right) = 0.$$

Notice that, by Lemma A.2,

$$|\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f_{\tau_n}(x)| \leq 2 \sup_{\substack{x\in\mathbb{R}^p \\ \tau\in[0,\infty]}} \left|\frac{LGD_n(x,\tau) - LGD(x,\tau)}{h_n^k \tau_n^{kp}}\right|^{1/k}.$$

We apply again Theorem 2.4 with $t = t_n \coloneqq \sqrt{n} h_n^k \tau_n^{kp} (\epsilon/4)^k$. Then, there are constants $1 < C_{G,0}, C_{G,1}, C_{G,2} < \infty$ such that, for large enough n,

$$\begin{aligned} &P^{\otimes n}\left(\sup_{h\in[h_n,h^*]} \sup_{v\in S^{p-1}} \sup_{x\in K} |\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f_{\tau_n}(x)| \geq \frac{\epsilon}{2}\right) \\ \leq &P^{\otimes n}\left(\sqrt{n} \sup_{\substack{x\in\mathbb{R}^p \\ \tau\in[0,\infty]}} |LGD_n(x,\tau) - LGD(x,\tau)| \geq t_n\right) \leq D_G(n, t_n), \end{aligned}$$

where $D_G(\cdot,\cdot)$ is defined in 2.13. 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 $\lim_{n\to\infty} D_G(n, t_n) = 0$. Finally, for (iii), we apply Lemma A.4 with $a_n = h_n^k \tau_n^{kp}$ and $b = (\epsilon/4)^k$ and get constants $0 < \tilde{C} < \infty$ and $\tilde{n} \in \mathbb{N}$ such that, for all $n \geq \tilde{n}$,

$$D_G(n, t_n) \leq \frac{\tilde{C}}{n^2}.$$

■

A version of discrete Grönwall lemma (see e.g. Holte (2009)) is needed in Theorem 3.2 to evaluate the difference between the sequence $\{y_{n,r,j}\}_{j=1}^{j^*}$ (defined in the proof) and the solution $u_x(\cdot)$ of (3.2).

Lemma A.7 (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, $a_n \leq (\sum_{j=0}^{n-1} b_j) \exp(\sum_{j=1}^{n-1} c_j)$.*

The next lemma is also used in the proof of Theorem 3.2.

Lemma A.8 *Suppose that $f(\cdot)$ is continuously differentiable and K is a compact subset of S_f with $K \cap N_f = \emptyset$. Then, there exist $r(K), c(K) > 0$ such that $(K)^{+r(K)} \subset S_f$ and, for all $x \in K$ and $(h, v) \in (0, r(K)] \times (S^{p-1} \cap \overline{B}_{r(K)}(w(x)))$, $\nabla_v^h f(x) \geq c(K)$.*

Proof of Lemma A.8. Recall (4.6) and let $g : [0, \infty) \to \mathbb{R}$ be given by

$$g(h) = \min_{y \in K}(f(y + hw(y)) - f(y)).$$

By the mean value theorem, it holds that $g(h) = h \min_{y \in K} \langle \nabla f(y + chw(y)), w(y) \rangle$, for some $0 \leq c \leq 1$. Let $h(K) > 0$ such that $(K)^{+h(K)} \subset S_f$. Since, by Remark 2.1, $\nabla f(\cdot)$ is uniformly continuous in $(K)^{+h(K)}$, we have that

$$g'(0) = \lim_{h \to 0^+} g(h)/h = \min_{y \in K} \|\nabla f(y)\|. \tag{A.41}$$

Now, by multivariate Taylor's theorem with integral remainder, we have that, for $v \in S^{p-1}$ and $h > 0$,

$$\begin{aligned} f(x + hv) &= f(x + hw(x)) + h\langle \nabla f(x + hw(x)), v - w(x) \rangle \\ &\quad + h^2 \int_0^1 (1 - s)(v - w(x))^\top H_f(x + hs(v - w(x)))(v - w(x))ds. \end{aligned}$$

It follows that, for $0 < h \leq h(K)/2$,

$$\begin{aligned} f(x + hv) &\geq f(x) + g(h) - h\|v - w(x)\|\|\nabla f(x + hw(x))\| \\ &\quad - h^2\|v - w(x)\|^2 \int_0^1 (1 - s)\|H_f(x + hs(v - w(x))\|_{\mathcal{M}} ds \\ &\geq f(x) + g(h) - h\|v - w(x)\|c_1 - h^2\|v - w(x)\|^2 c_2/2, \end{aligned}$$

where

$$c_1 \coloneqq \max_{y \in (K)^{+h(K)/2}} \|\nabla f(y)\| \text{ and } c_2 \coloneqq \max_{y \in (K)^{+h(K)}} \|H_f(y)\|_{\mathcal{M}}.$$

Therefore, we have that

$$\nabla_v^h f(x) \geq \tilde{g}(h) \coloneqq g(h)/h - \|v - w(x)\|c_1 - h\|v - w(x)\|^2 c_2/2.$$

Since $f(\cdot)$ has no stationary points in K, $\min_{y \in K} \|\nabla f(y)\| > 0$, and the result follows from (A.41). ■

Proof of Corollary 3.1. We show that there exists $n^* \in \mathbb{N}$ and $\{\eta_n\}_{n=1}^{\infty}$ such that $P^{\otimes n}(J_n = 1) \leq \eta_n$ for all $n \geq n^*$ and $\sum_{n=n^*}^{\infty} \eta_n < \infty$. Then the result follows from Borel-Cantelli lemma. To this end, we explicitly express the constant η in Theorem 3.2 as a function of n and observe the convergence of the series. We first notice that, for $n \geq n_1$, we can choose $\eta_n/2 \geq k(1-\alpha_0\Lambda^*)^n$ in (4.23). Next, we apply in (4.16) Lemma 3.2 (iii) with $K = \tilde{K}$, $h^* = r$, and $\epsilon = d^*\underline{\nu}$ and get constants $0 < \tilde{C} < \infty$ and $\tilde{n} \in \mathbb{N}$ such that, for all $n \geq \tilde{n}$,

$$P^{\otimes n}\left(\sup_{h\in[h_n,r]}\sup_{v\in S^{p-1}}\sup_{x\in\tilde{K}}|\nabla_v^h f_{\tau_n,n}(x) - \nabla_v^h f(x)| < d^*\underline{\nu}\right) \leq 1 - \frac{\tilde{C}}{n^2}.$$

Therefore, for all $n \geq n^* := \max(n_1, \tilde{n})$, we can choose

$$\eta_n/2 = \max(k(1-\alpha_0\Lambda^*)^n, \tilde{C}/n^2),$$

yielding $\sum_{n=n^*}^{\infty}\eta_n < \infty$. ■

B. 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 $f_{\tau,n}(\cdot)$ are unknown. Our main result in this section characterizes the asymptotic variance and establishes a related limit distribution. To this end, we let

$$\Lambda_1^{*2} := \int \Lambda_1^2(x_1)dx_1,$$

where, for $x_1 \in \mathbb{R}^p$,

$$\Lambda_1(x_1) := \int h_1(0; x_1, \ldots, x_p)dx_2 \ldots dx_p.$$

Theorem B.1 *Let P be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^p$ with continuous density $f(\cdot)$. Let $x \in S_f$ and $\{\tau_n\}_{n=1}^{\infty}$ be a sequence of positive scalars converging to zero. Suppose (2.4) and $E[(\tilde{h}_\tau^{(1)}(x; X_1))^2] > 0$ hold true. If $\sqrt{n}\tau_n^{((2k-1)p)/2} \xrightarrow[n\to\infty]{} \infty$, then*

$$\sqrt{n}\tau_n^{p/2}(f_{\tau_n,n}(x) - f_{\tau_n}(x)) \xrightarrow[n\to\infty]{d} N(0, \Lambda_1^{*2}f(x)).$$

Remark B.1 *We notice that, for $k > 1$, the limit distribution in Theorem B.1 with $f_{\tau_n}(\cdot)$ replaced by $f(\cdot)$ cannot hold. In fact, the deterministic term $f_{\tau_n}(x) - f(x)$ is, by Lemma A.5, 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 $1/(\sqrt{n}\tau_n^{p/2})$. Since, necessarily, $\sqrt{n}\tau_n^{((2k-1)p)/2} \to \infty$, $f_{\tau_n}(x) - f(x)$ is the dominant term. On the other hand, if $k = 1$, $\sqrt{n}\tau_n^{p/2} \to \infty$ and $\sqrt{n}\tau_n^{p/2+2} \to 0$, then, by Lemma A.5, $\sqrt{n}\tau_n^{p/2}(f_{\tau_n}(x) - f(x)) \to 0$. Hence,*

$$\sqrt{n}\tau_n^{p/2}(f_{\tau_n,n}(x) - f(x)) \xrightarrow[n\to\infty]{d} N(0, \Lambda_1^{*2}f(x)).$$

An alternative form for Theorem B.1 without the factor $f(x)$ in the variance term is given in the following corollary. Before proving Theorem B.1, we provide a lemma concerning the order of convergence of $E[(\tilde{h}_\tau^{(1)}(x; X_1))^2]$ to 0, as $\tau \to 0^+$.

Lemma B.1 *Suppose (2.4) holds true. If $f(\cdot)$ is continuous, then*

$$\lim_{\tau \to 0^+} \frac{E[(\tilde{h}_\tau^{(1)}(x; X_1))^2]}{\tau^{(2k-1)p}} = \Lambda_1^{*2} f^{2k-1}(x),$$

where

$$\Lambda_1^{*2} = \int (\int h_1(0; x_1, \dots, x_k) dx_2 \dots dx_k)^2 dx_1.$$

The proof of Lemma (B.1) follows using Theorem 2.1 (i), continuity of $f(\cdot)$, boundedness of $h_1(\cdot;\cdot)$, (2.4), and DCT. Details are in Francisci et al. (2022).

Proof of Theorem B.1. Using Hoeffding's decomposition of U-statistics (see (1.1.22) in Korolyuk and Borovskich (2013)), it follows that

$$LGD_n(x, \tau_n) - LGD(x, \tau_n) = \frac{k}{n} \sum_{i=1}^{n} (h_{\tau_n}^{(1)}(x; X_i) - LGD(x, \tau_n)) + \sum_{j=2}^{k} \binom{k}{j} \binom{n}{j}^{-1} \sum_{1 \le i_1 < \cdots < i_j \le n} g_{\tau_n}^{(j)}(x; X_{i_1}, \dots, X_{i_j}). \tag{B.1}$$

Now, applying Lindeberg-Levy Theorem for triangular arrays (Billingsley, 2012, Theorem 27.2) with

$$r_n = n,\ s_n = \sqrt{n}(E[(\tilde{h}_{\tau_n}^{(1)}(x; X_1))^2])^{1/2}, \text{ and } S_n = \sum_{i=1}^{n} (h_{\tau_n}^{(1)}(x; X_i) - LGD(x, \tau_n)),$$

it follows that

$$\sqrt{n} \frac{1}{n} \sum_{i=1}^{n} (h_{\tau_n}^{(1)}(x; X_i) - LGD(x, \tau_n)) / (E[(\tilde{h}_{\tau_n}^{(1)}(x; X_1))^2])^{1/2} \xrightarrow[n \to \infty]{d} N(0, 1), \tag{B.2}$$

provided the Lindeberg condition (Billingsley, 2012, Equation (27.8))

$$\lim_{n \to \infty} \frac{1}{E[(\tilde{h}_{\tau_n}^{(1)}(x; X_1))^2]} \int_{A_{n,\epsilon}} (h_{\tau_n}(x; x_1) - LGD(x, \tau_n))^2 f(x_1) dx_1 = 0 \tag{B.3}$$

holds for all $\epsilon > 0$, where

$$A_{n,\epsilon} := \{x_1 \in \mathbb{R}^p : (h_{\tau_n}(x; x_1) - LGD(x, \tau_n))^2 \ge \epsilon^2 n E[(\tilde{h}_{\tau_n}^{(1)}(x; X_1))^2]\}.$$

Using (A.1), it holds that $(h_{\tau_n}(x; x_1) - LGD(x, \tau_n)^2 \le l^2$, for all $x, x_1 \in \mathbb{R}^p$. Also, due to $x \in S_f$ and $n\tau_n^{(2k-1)p} \xrightarrow[n \to \infty]{} \infty$, Lemma B.1 implies that

$nE[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2] \xrightarrow[n\to\infty]{} \infty$. Let $n^* \in \mathbb{N}$ be such that for all $n \geq n^*$, $l^2 < \epsilon^2 nE[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2]$. It follows that $A_{n,\epsilon} = \emptyset$ for all $n \geq n^*$. Thus, (B.3) holds true and we obtain (B.2). Finally, for $j = 2, \ldots, k$, let

$$R_n^{(j)} = R_n^{(j)}(X_1, \ldots, X_n) \coloneqq \binom{k}{j}\binom{n}{j}^{-1} \sum_{1\leq i_1 < \cdots < i_j \leq n} g^{(j)}_{\tau_n}(x; X_{i_1}, \ldots, X_{i_j}).$$

Using Markov inequality and $E[g^{(j)}_{\tau_n}(x; x_1, \ldots, x_{j-1}, X_j)] = 0$ (see (1.1.22) in Korolyuk and Borovskich (2013)), we obtain that

$$P^{\otimes n}(\sqrt{n}|R_n^{(j)}| > \epsilon) \leq \frac{n}{\epsilon^2}\binom{k}{j}^2\binom{n}{j}^{-1} E\left[\left(g^{(j)}_{\tau_n}(x; X_1, \ldots, X_j)\right)^2\right],$$

which implies that

$$P^{\otimes n}\left(\frac{\sqrt{n}|R_n^{(j)}|}{(E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2])^{1/2}} > \epsilon\right) \leq \frac{n\binom{k}{j}^2\binom{n}{j}^{-1}}{\epsilon^2 E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2]} E\left[\left(g^{(j)}_{\tau_n}(x; X_1, \ldots, X_j)\right)^2\right].$$

Since $j \geq 2$ and $nE[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2] \xrightarrow[n\to\infty]{} \infty$,

$$P^{\otimes n}(\sqrt{n}|R_n^{(j)}|/(E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2])^{1/2} > \epsilon) \xrightarrow[n\to\infty]{} 0. \tag{B.4}$$

From (B.1), (B.2), and (B.4), it follows that

$$\sqrt{n}\frac{LGD_n(x,\tau_n) - LGD(x,\tau_n)}{k(E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2])^{1/2}} \xrightarrow[n\to\infty]{d} N(0,1). \tag{B.5}$$

Now, using the delta method we obtain

$$\sqrt{n}\frac{(LGD(x,\tau_n))^{1-1/k}}{(E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2])^{1/2}}((LGD_n(x,\tau_n))^{1/k} - (LGD(x,\tau_n))^{1/k}) \xrightarrow[n\to\infty]{d} N(0,1);$$

equivalently,

$$Z_n \coloneqq \sqrt{n}\frac{\tau_n^{kp} f^{k-1}_{\tau_n}(x)}{(E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2])^{1/2}}(f_{\tau_n,n}(x) - f_{\tau_n}(x)) \xrightarrow[n\to\infty]{d} N(0,1). \tag{B.6}$$

To complete the proof, since $x \in S_f$ and $\tau_n > 0$, it holds, by Theorem 2.1 (i), that

$$\frac{f^k_{\tau_n}(x)}{f^k(x)} = \frac{LGD(x,\tau_n)}{\tau_n^{kp} f^k(x)} \xrightarrow[n\to\infty]{} 1 \tag{B.7}$$

and, by Lemma B.1,

$$\frac{(E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2])^{1/2}}{\tau_n^{(k-1/2)p}} \xrightarrow[n\to\infty]{} \Lambda_1^* f^{k-1/2}(x) > 0. \tag{B.8}$$

(B.7) and (B.8) imply that

$$
\begin{aligned}
Y_n &:= \frac{(E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2])^{1/2}}{\tau_n^{(k-1/2)p} f^{k-1}_{\tau_n}(x)} \cdot \frac{1}{\Lambda_1^* f^{\frac{1}{2}}(x)} \\
&= \frac{(E[(\tilde{h}^{(1)}_{\tau_n}(x;X_1))^2])^{1/2}}{\tau_n^{(k-1/2)p}} \cdot \frac{1}{\Lambda_1^* f^{k-1/2}(x)} \cdot \frac{f^{k-1}(x)}{f^{k-1}_{\tau_n}(x)} \xrightarrow[n\to\infty]{} 1.
\end{aligned}
$$

From (B.6) and Slutsky's Theorem it follows that

$$Y_n Z_n \xrightarrow[n\to\infty]{d} N(0,1),$$

completing the proof. ∎

Corollary B.1 *Under the hypothesis of Theorem B.1,*

$$\sqrt{n}\tau_n^{\frac{1}{2}p}\left(\sqrt{f_{\tau_n,n}(x)} - \sqrt{f_{\tau_n}(x)}\right) \xrightarrow[n\to\infty]{d} N(0, \Lambda_1^{*2}/4).$$

The proof of Corollary B.1 follows from Theorem B.1, Proposition 2.2, and Slutsky's Theorem. For the details see Francisci et al. (2022). An extension of Theorem B.1 uniformly over S_f, namely,

$$\sqrt{n}\tau_n^{\frac{1}{2}p}(f_{\tau_n,n}(\cdot) - f_{\tau_n}(\cdot)) \xrightarrow[n\to\infty]{d} \Lambda_1^* W(\cdot) \text{ in } \ell^\infty(S_f),$$

where $\{W(x)\}_{x\in S_f}$ is a centered Gaussian process with the covariance function $\gamma : S_f \times S_f \to \mathbb{R}$ given by $\gamma(x,y) = \sqrt{f(x)f(y)}$, requires an extension of the results of Arcones and Giné (1993) to triangular arrays and it is beyond the scope of the present paper. 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 $\{Z^G_{\tau_n}(x)\}_{n=1}^\infty$ depend on n and x.

C. Examples

In this section of the appendix, we provide additional examples of LDFs and verify that they satisfy the VC-subgraph property.

Example C.1 *As in the introduction, let $G(\cdot) = \mathbf{I}(\cdot \in Z^G_1(0))$, for some $k \geq 1$. Then, as before, for $G = L, B, S, K_\beta$, we obtain local lens (Kleindessner and Von Luxburg, 2017), spherical, simplicial (Agostinelli and Romanazzi, 2008), and β-skeleton depth. In particular, $K_1 = B$ and $K_2 = L$. We will now verify that these class of depth functions satisfy the VC-subgraph property. Let $\mathcal{B} := \{\overline{B}_r(x) : x \in \mathbb{R}^p, r > 0\}$ be the class of balls in $\mathbb{R}^p$ and, for $\beta \geq 1$, $\mathcal{K}_\beta := \{\overline{B}_{\frac{\beta}{2}\|x_1-x_2\|}(\frac{\beta}{2}x_1 + (1-\frac{\beta}{2})x_2) \cap \overline{B}_{\frac{\beta}{2}\|x_1-x_2\|}((1-\frac{\beta}{2})x_1 + \frac{\beta}{2}x_2) : x_1, x_2 \in \mathbb{R}^p\}$ be the class of all β-skeleton sets. By Theorem 1 in Dudley (1979), $\mathcal{B}$ is a VC-class of sets. Applying Proposition 3.6.7 (ii) of Giné and Nickl (2016), it follows that* *also the intersection $\mathcal{B} \cap \mathcal{B}$ is a VC-class of sets. Since a subset of a VC-class of sets is still a VC-class (see Proposition 3.6.7 (iv) in Giné and Nickl (2016)), it holds that, for all $\beta \geq 1$, $\mathcal{K}_\beta \subset \mathcal{B} \cap \mathcal{B}$ is a VC-class. We finally notice that the function $\mathbf{I}(\cdot \in Z_1^{K_{(\cdot)}}(0))$ is jointly Borel measurable. Similarly, the class of simplices in $\mathbb{R}^p$, which are given by the intersections of $p+1$ half-spaces, is a VC-class (see Corollary 6.7 of Arcones and Giné (1993)).*

Example C.2 *We turn to the uniform kernel (Devroye and Györfi, 1985) in this example. Again, in this case, $k = 1$ and*

$$G(\cdot) \coloneqq \mathbf{I}(\cdot \in \overline{B}_1(0)).$$

Since closed balls in $\mathbb{R}^p$ form a VC-class of sets by Theorem 1 in Dudley (1979), it follows that $G(\cdot)$ belongs to the VC-subgraph class.

Example C.3 *Local depth functions can also be developed using kernel density techniques. In this case, let $k = 1$ and $G(\cdot)$ be a radially symmetric integrable function with unit integral (Chacón and Duong, 2018).*

D. Supplementary results

In this section we use local depth functions for density level set estimation and identification of stationary points and derive supplementary results complementing Theorem 3.2. Additional details are provided in Francisci et al. (2022).

D.1. Density level set estimation

In this subsection, we provide an application of the theory and methods of the paper to estimate the upper level sets. We briefly describe another application to divergence based inference. We begin with the definition of level sets and upper level sets.

Definition D.1 *For $\alpha > 0$, the level sets of $f(\cdot)$ and $f_\tau(\cdot)$ are $L^\alpha = \{x \in \mathbb{R}^p : f(x) = \alpha\}$ and $L^\alpha_\tau = \{x \in \mathbb{R}^p : f_\tau(x) = \alpha\}$, respectively. The upper level sets of $f(\cdot)$, $f_\tau(\cdot)$ and $f_{\tau,n}(\cdot)$ are $R^\alpha \coloneqq \{x \in \mathbb{R}^p : f(x) \geq \alpha\}$, $R^\alpha_\tau \coloneqq \{x \in \mathbb{R}^p : f_\tau(x) \geq \alpha\}$ and $R^\alpha_{\tau,n} \coloneqq \{x \in \mathbb{R}^p : f_{\tau,n}(x) \geq \alpha\}$, respectively.*

The next proposition shows that in the limit the upper level sets induced by $f_\tau(\cdot)$ and $f_{\tau,n}(\cdot)$ coincide with those induced by $f(\cdot)$. We use the notation $\mathring{A}$ for the interior of a set A.

Proposition D.1 *Suppose that $f(\cdot)$ 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*

$$\mathring{R}^\alpha \subset \liminf_{n\to\infty} R^{\alpha_n}_{\tau_n} \subset \limsup_{n\to\infty} R^{\alpha_n}_{\tau_n} \subset R^\alpha, \tag{D.1}$$

and, if $\lambda(L^\alpha) = 0$, then

$$\lim_{n\to\infty} R^{\alpha_n}_{\tau_n} = R^\alpha \text{ a.e.} \tag{D.2}$$

Suppose additionally that $\mathcal{H}_G$ is a VC-subgraph class of functions and $\lim_{n\to\infty} \frac{n}{\log(n)}\tau_n^{2kp} = \infty$. It holds that

$$\mathring{R}^{\alpha} \subset \liminf_{n\to\infty} R^{\alpha_n}_{\tau_n,n} \subset \limsup_{n\to\infty} R^{\alpha_n}_{\tau_n,n} \subset R^{\alpha}\, a.s., \tag{D.3}$$

and, if $\lambda(L^{\alpha}) = 0$, then

$$\lim_{n\to\infty} R^{\alpha_n}_{\tau_n,n} = R^{\alpha} \;\; a.s. \tag{D.4}$$

The proof of Proposition D.1 involves standard arguments on set convergence and can be found in Francisci et al. (2022). A common approach in modal clustering is to define clusters as the connected components of the upper level sets R^{α} for some $\alpha > 0$ (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.

D.2. Exact identification of stationary points and modes

In this subsection, we further develop the results of Section 3.1 by providing some conditions under which the stationary points (resp. modes, antimodes) of $f(\cdot)$ are *exactly* the stationary points (resp. modes, antimodes) of $f_\tau(\cdot)$ for $\tau > 0$. The key criteria for the identification of the modes is the notion of symmetry proposed below.

Definition D.2 *Given $\tau > 0$, a density function $f(\cdot)$ is said to be τ-centrally symmetric about $\mu \in S_f$ if, for all $x \in \mathbb{R}^p$ with $\|x\| \le \tau$, $f(\mu + x) = f(\mu - x)$.*

In particular, for $p = 1$, $f(\cdot)$ is τ-centrally symmetric about $\mu \in \mathbb{R}$ if $f(\mu - x) = f(\mu + x)$ for all $x \in [0, \tau]$. If $f(\cdot)$ has a continuous derivative, a direct computation shows that, for $G = L, S, B, K_\beta$, $f^{'}_\tau(\mu) = 0$. Indeed, we see that

$$f_\tau(x) = \frac{1}{\tau}\sqrt{LGD(x,\tau)}$$

where

$$LGD(x,\tau) = 2\int_{T^{\tau}_{++}} f(x + x_1)f(x - x_2)dx_1dx_2$$

and

$$T^{\tau}_{++} = \{(x_1, x_2) : x_1 \ge 0, x_2 \ge 0, x_1 + x_2 \le \tau\}.$$

In particular, if $f(\cdot)$ has a continuous derivative, it follows that

$$f^{'}_\tau(x) = \frac{1}{\tau\sqrt{LGD(x,\tau)}}\int_{T^{\tau}_{++}} f^{'}(x + x_1)f(x - x_2) + f(x + x_1)f^{'}(x - x_2)dx_1dx_2.$$

Therefore, the sign of $f_\tau(x)$ depends on the sign of $f^{'}(\cdot)$ in the interval $(x - \tau, x + \tau)$. In particular, if $\mu \in \mathbb{R}$ satisfies $f(\mu - x) = f(\mu + x)$ for all $x \in (0, \tau)$, it follows that $f^{'}(\mu - x) = -f^{'}(\mu + x)$, yielding $f^{'}_\tau(\mu) = 0$.

Our next result, which is about the Hessian matrix, gives sufficient conditions for a stationary point μ and a mode m of $f(\cdot)$ to be a stationary point and a mode of $f_\tau(\cdot)$.

Theorem D.1 *Suppose (2.4) holds true and let $\tau > 0$. Then the following hold:*
(i) If $f(\cdot)$ has continuous first order partial derivatives in $\overline{B}_\tau(\mu) \subset S_f$ and $f(\cdot)$ is τ-centrally symmetric about the stationary point μ, then μ is a stationary point for $f_\tau(\cdot)$.
(ii) Suppose that $f(\cdot)$ 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

$$J_f(x_1, \dots, x_k) \coloneqq H_f(x_1) f(x_2) \dots f(x_k) + (k-1) \nabla f(x_1) \nabla f(x_2)^\top f(x_3) \dots f(x_k)$$

is negative (resp. positive) definite, then m is also a mode (resp. an antimode) for $f_\tau(\cdot)$.

Notice that $J_f(m, \dots, m) = H_f(m) f^{k-1}(m)$ is negative (resp. positive) definite and therefore the last condition of Theorem D.1 is satisfied by $f(\cdot)$, for τ small.

Proof of Theorem D.1. For (i) notice that if $f(\cdot)$ is τ-centrally symmetric about μ, then, for all $y \in \mathbb{R}^p$ with $\|y\| \le \tau$, $f(\mu+y) = f(\mu-y)$ 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 (3.5) and (A.4) it follows that, for all $1 \le j \le p$,

$$\begin{aligned} &\int h_\tau(0; x_1, \dots, x_k) \nabla f(\mu + x_1) f(\mu + x_2) \dots f(\mu + x_k) dx_1 \dots dx_k \\ &= -\int h_\tau(0; x_1, \dots, x_k) \nabla f(\mu + x_1) f(\mu + x_2) \dots f(\mu + x_k) dx_1 \dots dx_k, \end{aligned}$$

and therefore (3.5) and $\nabla f_\tau(\mu) = 0$.
We now prove (ii). Since $f(\cdot)$ is τ-centrally symmetric about m, by (i),

$$\partial_j f_\tau(m) = 0 \text{ for } j = 1, \dots, p \tag{D.5}$$

and hence m is a stationary point for $f_\tau(\cdot)$. Moreover, (D.5) implies that, for $i, j = 1, \dots, p$,

$$\partial_i \partial_j f_\tau(m) = \frac{1}{k} (f_\tau(m))^{1-k} (\partial_i \partial_j f_\tau^k(m)),$$

where, by Proposition 2.1, (A.2) and (A.4),

$$\begin{aligned} \partial_i \partial_j f_\tau^k(m) = k \int \frac{h_\tau(0; x_1, \dots, x_k)}{\tau^{kp}} \Bigg[&\partial_i \partial_j f(m + x_1) \prod_{l=2}^{k} f(m + x_l) \\ &+ (k-1) \partial_j f(m + x_1) \partial_i f(m + x_2) \prod_{l=3}^{k} f(m + x_l) \Bigg] dx_1 \dots dx_k. \end{aligned}$$

Noticing that the integral of a matrix is the matrix of the integrals, we get that

$$H_{f_\tau}(m) = \frac{1}{k}(f_\tau(m))^{k-1} \int \frac{h_\tau(0; x_1, \ldots, x_k)}{\tau^{kp}} J_f(m + x_1, \ldots, m + x_k) dx_1 \ldots dx_k.$$

Since the Hessian is symmetric, there exists an orthogonal matrix Q such that

$$\begin{aligned} D &= Q^\top H_{f_\tau}(m) Q \\ &= \frac{1}{k}(f_\tau(m))^{k-1} \int \frac{h_\tau(0; x_1, \ldots, x_k)}{\tau^{kp}} Q^\top J_f(m + x_1, \ldots, m + x_k) Q dx_1 \ldots dx_k \end{aligned}$$

is a diagonal matrix. Now, since $J_f(m + x_1, \ldots, m + x_k)$ is negative (resp. positive) definite, for all $y \in \mathbb{R}^p \setminus \{0\}$, $y^\top J_f(m + x_1, \ldots, m + x_k) y < 0$ (resp. > 0), and therefore the diagonal elements of $Q^\top J_f(m + x_1, \ldots, m + x_k) Q$ are negative (resp. positive). It follows that the diagonal elements of D (that is, the eigenvalues of $H_{f_\tau}(m)$) are negative (resp. positive) and m is a mode (resp. an antimode) for $f_\tau(\cdot)$. ■

D.3. Supplementary results related to Theorem 3.2

As the next Lemma shows, in Theorem 3.2, the minimum distance between all data points and a point $x \in S_f$ (denoted as $\tilde{h}_n$ below) is positive a.s. for all $n \in \mathbb{N}$ and converges to zero a.s. as $n \to \infty$. However, $p \geq 6k + 1$ is needed for $n\tilde{h}_n^{2k} \tau_n^{2kp} \xrightarrow[n\to\infty]{} \infty$ a.s., for some sequence of positive scalars $\{\tau_n\}_{n=1}^\infty$ converging to zero (by Lemma D.1 (iii) we can take $\tau_n^{2kp} = n^{-\delta}$, for some $0 < \delta < 1 - \frac{6k}{d}$, that is $\tau_n = n^{-\delta/(2kp)}$). This shows that, for $p \geq 6k + 1$, by choosing a suitable sequence $\{\tau_n\}_{n=1}^\infty$, we can replace h_n by $\tilde{h}_n$ in Theorem 3.2. In turn, this allows replacement of the set $\mathcal{X}_{n,r}(x) = \{X \in \mathcal{X}_n : h_n \leq \|X - x\| \leq r\}$ by $\tilde{\mathcal{X}}_{n,r}(x) = \{X \in \mathcal{X}_n : \|X - x\| \leq r,\ X \neq x\}$.

Lemma D.1 *Let $\mathcal{X}_n \coloneqq \{X_1, \ldots, X_n\}$ a sample of i.i.d. random variables from a probability distribution P with continuous and bounded density $f(\cdot)$, $x \in S_f$, and $\tilde{h}_n = \min_{y,z \in \mathcal{X}_n \cup \{x\}, y \neq z} \|y - z\|$. Then, (i) $\tilde{h}_n > 0$ a.s., (ii) $\tilde{h}_n \xrightarrow[n\to\infty]{} 0$ a.s. and (iii) for $p \geq 6k + 1$ and $0 < \delta < 1 - \frac{6k}{\delta}$, $n^{1-\delta}\tilde{h}_n^{2k} \xrightarrow[n\to\infty]{} \infty$ a.s.*

Proof of Lemma D.1. We first prove (i). Since P is absolutely continuous w.r.t. the Lebesgue measure, it holds that

$$\begin{aligned} P^{\otimes n}(\tilde{h}_n = 0) &= P^{\otimes n}(\cup_{i=1}^n [\|X_i - x\| = 0] \cup \cup_{i=1}^n \cup_{j=i+1}^n [\|X_i - X_j\| = 0]) \\ &\leq nP(\|X_1 - x\| = 0) + \frac{n(n-1)}{2} \int P(\|X_1 - y\| = 0) f(y) dy = 0. \end{aligned}$$

For (ii), observe that, for all $\epsilon > 0$,

$$P^{\otimes n}(\tilde{h}_n \geq \epsilon) \leq P^{\otimes n}(\min_{i=1,\ldots,n} \|X_i - x\| \geq \epsilon) = P(\|X_1 - x\| \geq \epsilon)^n.$$

Since $x \in S_f$ and $f(\cdot)$ is continuous, it holds that $P(\|X_1 - x\| \geq \epsilon) < 1$ and

$$\sum_{n=1}^{\infty} P^{\otimes n}(\tilde{h}_n \geq \epsilon) \leq \sum_{n=1}^{\infty} P(\|X_1 - x\| \geq \epsilon)^n < \infty.$$

By Borel-Cantelli lemma, it follows that $\tilde{h}_n \xrightarrow[n\to\infty]{} 0$ a.s. We now prove (iii). To this end, let $M > 0$ and notice that $P^{\otimes n}(n^{1-\delta}\tilde{h}_n^{2k} \leq M^{2k})$ is equal to

$$P^{\otimes n}(\cup_{i=1}^{n}[\|X_i - x\| \leq Mn^{-(1-\delta)/(2k)}] \cup \cup_{i=1}^{n}\cup_{j=i+1}^{n}[\|X_i - X_j\| \leq Mn^{-(1-\delta)/(2k)}]),$$

which is bounded above by

$$nP(B_{Mn^{-(1-\delta)/(2k)}}(x)) + \frac{n(n-1)}{2}\int P(B_{Mn^{-(1-\delta)/(2k)}}(y))f(y)dy. \tag{D.6}$$

Now, since $f(\cdot)$ is bounded, we have that $\alpha \coloneqq \sup_{y\in\mathbb{R}^p} f(y) < \infty$. For $y \in \mathbb{R}^p$, it holds that

$$P(B_{Mn^{-(1-\delta)/(2k)}}(y)) \leq \alpha\lambda(\overline{B}_{Mn^{-(1-\delta)/(2k)}}(x)) = \alpha C n^{-p(1-\delta)/(2k)}, \tag{D.7}$$

where $C = M^p \pi^{p/2}/\Gamma(p/2+1)$. Using (D.7) in (D.6), we obtain that

$$P^{\otimes n}(n^{1-\delta}\tilde{h}_n^{2k} \leq M^{2k}) \leq \alpha C n^{2-p(1-\delta)/(2k)}.$$

Therefore, using $p \geq 6k+1$ and $0 < \delta < 1 - \frac{6k}{p}$, we have that

$$\sum_{n=1}^{\infty} P^{\otimes n}(n^{1-\delta}\tilde{h}_n^{2k} \leq M^{2k}) \leq \alpha C \sum_{n=1}^{\infty} n^{2-p(1-\delta)/(2k)} < \infty.$$

By another application of Borel-Cantelli lemma, we conclude that $n^{1-\delta}\tilde{h}_n^{2k} \xrightarrow[n\to\infty]{} \infty$ a.s. ■

E. Clustering Algorithm

In this section, we provide a detailed description of the algorithm for clustering. As a first step, starting from a point $x \in \mathbb{R}^p$, we search, in a given neighborhood of x, for the point y that yields the largest directional derivative $\nabla_v^h(LGD_n(\cdot,\tau))^{1/k}$ with $h = \|y - x\|$ and $v = (y-x)/\|y-x\|$, that is,

$$d_{\tau,n}(x;y) \coloneqq \frac{(LGD_n(y,\tau))^{1/k} - (LGD_n(x,\tau))^{1/k}}{\|y-x\|}. \tag{E.1}$$

Next, given n data points $x_1, \ldots, x_n$, the localization parameter τ used for the clustering procedure is chosen as the quantile of order q, $0 \leq q \leq 1$, of the empirical distribution of the $\binom{n}{2}$ distances $\|x_i - x_j\|$, $i > j$, $i, j \in \{1, 2, \ldots, n\}$ for lens depth, spherical depth, and β-skeleton depth. Detailed methodology

Algorithm 1: Clustering with local general depth

Input: $\{x_1, \dots, x_n\}$, $\{y_1, \dots, y_o\}$ (optional), q, s, r
Output: Local maxima for input points: $\{z_1, \dots, z_{n+o}\}$
1 Compute the quantile τ of order q of all pairwise distances: $\|x_i - x_j\|$, $i > j$, $i, j \in \{1, 2, \dots, n\}$
2 Store $\{x_1, \dots, x_n\}$, $\{y_1, \dots, y_o\}$ in new variables
for $i = 1$ *to* n **do**
  $z_i^* \coloneqq x_i$
end
for $i = 1$ *to* o **do**
  $z_{i+n}^* \coloneqq y_i$
end
3 Compute the local general depth with localization parameter τ of $\{z_1^*, \dots, z_{n+o}^*\}$ w.r.t. $\{x_1, \dots, x_n\}$
4 For all points, compute the corresponding local maxima
for $i = 1$ *to* $n + o$ **do**
  repeat
5    $z_i \coloneqq z_i^*$
6   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 $w_1, \dots, w_l$ $(l \geq s)$
7    $z_i^* \coloneqq \text{argmax}_{v \in \{w_1, \dots, w_l\}} d_{\tau,n}(z_i; v)$
  until $LGD_n(z_i^*, \tau) > LGD_n(z_i, \tau)$
end

for simplicial depth is provided below. We now summarize the procedure for computing the clusters in Algorithm 1.

The algorithm requires as input, data points $\{x_1, \dots, x_n\}$, quantile q, 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 (E.1), 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 (E.1)). The parameter r gives a bound on the norm $\|y - x\|$ in (E.1) 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 q. 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 $v \in S^{p-1}$) 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 q 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. A general method for the choice of s is described in Francisci et al. (2022). We now turn to the parameter q. We notice that choosing q is equivalent to choosing τ. Thus typical values of τ correspond to typical values of q. Now, convergence of the clustering algorithm (cf. Theorem 3.4) requires that $\lim_{n\to\infty} n\tau_n^{2kp} = \infty$. Thus, we can take $\tau_n = n^{(-1+\delta)/(2kp)}$, for some $\delta > 0$. While for the class of β-skeleton depths q can be taken as the quantile of pairwise distances $\|x_i - x_j\|$, $i > j$, $i, j \in \{1, 2, \dots, n\}$, for the simplicial depth, q can be chosen as a quantile of the $\binom{n}{p+1}$ maxima of the form $\max_{j,l=1,\dots,p+1, j>l} \|x_{i_j} - x_{i_l}\|$ for all $\binom{n}{p+1}$ combinations of indices $i_1, \dots, i_{p+1}$ from $\{1, 2, \dots, n\}$. Alternatively, we could choose τ as described in Theorem 3.2 for all depths, that is, τ_n such that $\lim_{n\to\infty} nh_n^{2k}\tau_n^{2kp} = \infty$.

We now turn to the computational complexity of β-skeleton and simplicial depth. To this end, we recall that $LK_\beta D_n$ is a U-statistics of order 2, while LSD_n is a U-statistics of order $(p+1)$. This means that the computational complexity of $LK_\beta D_n$ is of order $O(\binom{n}{2})$, while the computational complexity of the LSD_n is of order $O(\binom{n}{p+1})$, which makes a significant difference, especially in high dimensions. In dimension 2 an optimal algorithm for $LK_\beta D_n$ requires $O(n^{\frac{3}{2}+\epsilon})$ operations for $\beta > 1$ and $O(n\log(n))$ operations for $\beta = 1$ (Bremner and Shahsavarifar, 2017; Shahsavarifar and Bremner, 2018; Shahsavarifar, 2019). In general, the computational complexity of both $LK_\beta D_n$ and LSD_n can be reduced by random subsampling Yuan (2018); in our simulations with $p \geq 5$ and LSD_n (see Appendix F) we sample uniformly 10^8 simplices to reduce the computational cost.

F. Simulations and data analysis

F.1. Numerical experiments

In this subsection, we describe additional simulation results of our method for identification of clusters. For the sake of completeness and ease of comparisons, we retain the results described in the main paper. Specifically, we evaluate the performance using empirical Hausdorff distance and empirical probability distance between the "true" cluster and the estimated cluster (see Chacón (2015), for instance). 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. We consider the mixtures of bivariate normal distributions investigated in Wand and Jones (1993) and Chacón (2009) referred to as (H) Bimodal IV and #10 Fountain. Their analytical expression and the associated *true clusters* are given in Appendix F.3. To test the performance of our methodology in multivariate models, we consider a quadrimodal density in dimension five. We refer to this distribution as Mult. Quadrimodal. Additionally, we also study a circular distribution, which we refer to Circular Bimodal II. These is also described in Appendix F.3. Additional analyses for several other distributions are provided in Francisci et al. (2022). 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 Algorithm 1). For more details about the numerical implementation and the quantiles for LSD we refer to Appendix E. We compare our results based on LLD and LSD, with hierarchical clustering (Hclust) and Kernel density estimator (KDE) using both Algorithm 1 and mean shift algorithm (Fukunaga and Hostetler, 1975). For mean shift algorithm we use 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 options `"dunconstr"` or `"dscalar"`, and derivatives of order one. In the first case, the starting matrix is obtained via minimization of the asymptotic mean integrated squared error (AMISE) of the gradient of the estimated density, while, in the second case, a diagonal pilot bandwidth matrix is used to estimate the final (full) bandwidth matrix. 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). 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 Euclidean 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. We also apply two other recent clustering algorithms (Chacón, 2019), which are a combination of mixture model clustering (Fraley and Raftery, 2002) and modal clustering (Chacón, 2015) procedures. Both algorithms start by fitting a normal mixture density $\hat{f}(x) = \sum_{t=1}^{\hat{T}} \hat{\pi}_t \phi(x|\hat{\mu}_t, \hat{\Sigma}_t)$, where $\hat{\pi}_t \geq 0$, $\sum_{t=1}^{\hat{T}} \hat{\pi}_t = 1$, and $\phi(\cdot|\mu, \Sigma)$ is the density of a p-variate normal distribution with mean μ and covariance matrix Σ. This is done using the expectation maximization (EM) algorithm implemented in the function `Mclust` from the R package `mclust` (Scrucca et al., 2016). In the above, $\hat{T}$ is the value of $T \in \{1, \ldots, 9\}$ that minimizes the Bayesian information criterion (BIC) for fitting $\hat{f}_T(x) = \sum_{t=1}^{T} \hat{\pi}_t \phi(x|\hat{\mu}_t, \hat{\Sigma}_t)$. Mixture model clusters are then the sets given by

$$\hat{C}_t \coloneqq \{x \in \mathbb{R}^p : \hat{\pi}_t \phi(x|\hat{\mu}_t, \hat{\Sigma}_t) > \max_{\substack{j=1,\ldots,\hat{T}, \\ j \neq t}} \hat{\pi}_j \phi(x|\hat{\mu}_j, \hat{\Sigma}_j)\}$$

for $t = 1, \ldots, \hat{T}$. The first clustering algorithm is called mixture model modal merging (MMMM) and relies on the idea that mixture components are likely to be more than the number modes. Thus, one would like to merge mixture components whose points converge to the same mode into a single cluster. Following this idea, one applies mean shift algorithm starting from the estimated mixture means $\hat{\mu}_1, \ldots, \hat{\mu}_{\hat{T}}$ and merges those clusters whose estimated means converge to the same mode, yielding clusters $\tilde{C}_1, \ldots, \tilde{C}_{\hat{U}}$ where $\hat{U} \leq \hat{T}$. The second clustering algorithm is called mixture model modal clustering (MMMC) and is based on a direct computation of the stable manifolds generated by $\hat{f}(\cdot)$. An algorithm for this is provided in Section 3.1 of Chacón (2019). Turning to LLD, exploratory analysis suggests that some circular distributions require values of q higher than 0.1. To see this, let P be the Circular Bimodal II distribution and draw 100 samples from P. Figure 1 shows the median adjusted Rand index and interquartile range over $C = 100$ subsamples (left). The central plot shows the boxplot of optimal value of q and the right plot displays the number of clusters detected when q is the optimal quantile order. Thus, for the Circular Bimodal II distribution we let q run up to 0.2.

Tables 1 and 2 contain numerical results for several choices of q and s. Since the parameter r does not affect the output of Algorithm 1, we leave it fixed at $r = 0.05$. The expressions LLD-q-s and LSD-q-s refer to LLD and LSD with parameters q and s. The expressions KDE-`"dun"`-s and KDE-`"dsc"`-s refer to KDE with Algorithm 1 and parameter s; and pilot options `"dunconstr"` and `"dscalar"`, respectively. Similarly, the expressions KDE-`"dun"-ms` and KDE-`"dsc"-ms` refer to KDE with mean shift algorithm and pilot options `"dunconstr"` and `"dscalar"`. In Table 1 the first row refers to the case $c = 0$ and the second row to the case $c = 1$. From these two values, it is possible to compute the distance in probability for all values of c. In all the tables the best results are in bold face. For the probability distance, the best results are bolded only for the case $c = 1$. We observe that for mixture of normal distributions, the best results are always obtained by MMMM and MMMC. However, these methods perform poorly for the Circular Bimodal II distribution, where LDS yields the lowest errors. Among other methods, LLD yields the best results for the distributions Mult. Quadrimodal. For the distributions (H) Bimodal IV and #10 Fountain, KDE, LLD, and LSD all yield similar results. Finally, we notice that the merging algorithm in Chazal et al. (2013) may be used to improve the results of KDE, LLD, and LSD for the Circular Bimodal II distribution. Also, it is possible to improve the performance of LSD for Mult. Quadrimodal distribution by choosing smaller values of q, as described in Subsection F.2 below.

Clustering errors (distance in probability)				
	(H) Bimodal IV	#10 Fountain	Mult. Quadrimodal	Circular Bimodal II
MMMM	0.00 (0.00)	0.23 (0.06)	0.01 (0.00)	0.28 (0.02)
	0.00 (0.00)	0.23 (0.07)	**0.01 (0.00)**	0.55 (0.05)
MMMC	0.00 (0.00)	0.10 (0.06)	0.01 (0.00)	0.27 (0.02)
	0.00 (0.00)	0.10 (0.07)	**0.01 (0.00)**	0.53 (0.05)

KDE-"dun"-30	0.00 (0.00) **0.00 (0.00)**	0.06 (0.01) **0.06 (0.01)**	0.29 (0.31) 0.34 (0.37)	0.07 (0.06) 0.36 (0.06)
KDE-"dun"-50	0.00 (0.00) **0.00 (0.00)**	0.06 (0.01) **0.06 (0.01)**	0.12 (0.20) 0.14 (0.24)	0.10 (0.07) 0.34 (0.11)
KDE-"dsc"-30	0.07 (0.17) 0.11 (0.26)	0.06 (0.01) **0.06 (0.01)**	0.24 (0.30) 0.28 (0.35)	0.04 (0.03) 0.37 (0.05)
KDE-"dsc"-50	0.06 (0.16) 0.09 (0.24)	0.06 (0.01) **0.06 (0.01)**	0.10 (0.18) 0.11 (0.21)	0.07 (0.05) 0.37 (0.05)
KDE-"dun"-ms	0.01 (0.05) 0.01 (0.07)	0.19 (0.26) 0.21 (0.31)	0.43 (0.27) 0.57 (0.33)	0.19 (0.06) 0.38 (0.12)
KDE-"dsc"-ms	0.09 (0.19) 0.13 (0.28)	0.34 (0.30) 0.41 (0.36)	0.24 (0.30) 0.29 (0.36)	0.24 (0.05) 0.48 (0.10)
LLD-0.05-30	0.33 (0.21) 0.53 (0.33)	0.07 (0.05) 0.07 (0.06)	0.08 (0.19) 0.09 (0.22)	0.29 (0.03) 0.58 (0.06)
LLD-0.05-50	0.31 (0.25) 0.47 (0.37)	0.06 (0.01) **0.06 (0.01)**	0.03 (0.01) 0.03 (0.01)	0.27 (0.04) 0.53 (0.07)
LLD-0.1-30	0.09 (0.19) 0.13 (0.28)	0.06 (0.01) **0.06 (0.01)**	0.08 (0.19) 0.09 (0.22)	0.22 (0.05) 0.43 (0.11)
LLD-0.1-50	0.04 (0.12) 0.05 (0.18)	0.06 (0.01) **0.06 (0.01)**	0.03 (0.01) 0.03 (0.01)	0.20 (0.06) 0.39 (0.11)
LLD-0.2-30	/ /	/ /	/ /	0.15 (0.08) 0.27 (0.16)
LLD-0.2-50	/ /	/ /	/ /	0.12 (0.08) 0.21 (0.17)
LSD-0.01-30	0.08 (0.18) 0.12 (0.27)	0.06 (0.01) **0.06 (0.01)**	0.32 (0.07) 0.57 (0.15)	0.21 (0.05) 0.41 (0.10)
LSD-0.01-50	0.06 (0.15) 0.08 (0.23)	0.06 (0.01) **0.06 (0.00)**	0.33 (0.05) 0.64 (0.13)	0.20 (0.06) 0.37 (0.11)
LSD-0.05-30	0.00 (0.00) **0.00 (0.00)**	0.19 (0.07) 0.32 (0.14)	0.29 (0.11) 0.45 (0.17)	0.13 (0.08) 0.23 (0.17)
LSD-0.05-50	0.00 (0.00) **0.00 (0.00)**	0.20 (0.06) 0.35 (0.14)	0.28 (0.07) 0.52 (0.16)	0.07 (0.07) **0.11 (0.14)**
Hclust *	0.05 (0.09)	0.35 (0.07)	0.10 (0.04)	0.17 (0.05)

* The true number of clusters is given in input.

Table 1: Mean of the clustering errors based on distance in probability distance over 100 replications with $n = 1000$ samples for the distributions (H) Bimodal IV, #10 Fountain, Mult. Quadrimodal, and Circular Bimodal II. In parentheses the standard deviation.

Clustering errors (Hausdorff distance)				
	(H) Bimodal IV	#10 Fountain	Mult. Quadrimodal	Circular Bimodal II
MMMM	**0.00 (0.00)**	0.22 (0.03)	**0.01 (0.00)**	0.55 (0.04)
MMMC	**0.00 (0.00)**	0.09 (0.02)	**0.01 (0.00)**	0.53 (0.05)
KDE-"dun"-30	**0.00 (0.00)**	**0.06 (0.01)**	0.10 (0.08)	0.44 (0.07)
KDE-"dun"-50	**0.00 (0.00)**	**0.06 (0.01)**	0.07 (0.07)	0.40 (0.13)
KDE-"dsc"-30	0.04 (0.10)	**0.06 (0.01)**	0.09 (0.08)	0.48 (0.06)
KDE-"dsc"-50	0.04 (0.09)	**0.06 (0.01)**	0.06 (0.07)	0.46 (0.07)
KDE-"dun"-ms	**0.00 (0.03)**	0.08 (0.05)	0.16 (0.08)	0.44 (0.07)
KDE-"dsc"-ms	0.05 (0.11)	0.11 (0.06)	0.08 (0.08)	0.48 (0.07)
LLD-0.05-30	0.27 (0.17)	**0.06 (0.02)**	0.03 (0.04)	0.55 (0.05)
LLD-0.05-50	0.22 (0.17)	**0.06 (0.01)**	0.02 (0.01)	0.52 (0.06)
LLD-0.1-30	0.05 (0.11)	**0.06 (0.01)**	0.03 (0.05)	0.47 (0.07)
LLD-0.1-50	0.02 (0.08)	**0.06 (0.01)**	0.02 (0.01)	0.45 (0.09)
LLD-0.2-30	/	/	/	0.34 (0.19)
LLD-0.2-50	/	/	/	0.27 (0.21)
LSD-0.01-30	0.05 (0.11)	**0.06 (0.01)**	0.55 (0.19)	0.46 (0.06)
LSD-0.01-50	0.04 (0.09)	**0.06 (0.01)**	0.64 (0.17)	0.45 (0.08)
LSD-0.05-30	**0.00 (0.00)**	0.35 (0.15)	0.38 (0.18)	0.28 (0.20)
LSD-0.05-50	**0.00 (0.00)**	0.38 (0.15)	0.48 (0.18)	**0.13 (0.17)**
Hclust *	0.05 (0.09)	0.29 (0.05)	0.07 (0.03)	0.34 (0.10)

* The true number of clusters is given in input.

Table 2: Mean of the clustering errors based on distance Hausdorff distance over 100 replications with $n = 1000$ samples for the distributions (H) Bimodal IV, #10 Fountain, Mult. Quadrimodal, and Circular Bimodal II. In parentheses the standard deviation.

F.2. Data analysis

In this section, we revisit the data analysis with more details. As explained in the main paper, 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 data sets. 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); furthermore, the Hausdorff distance and probability distance from our algorithm are smaller than those of the competitors.

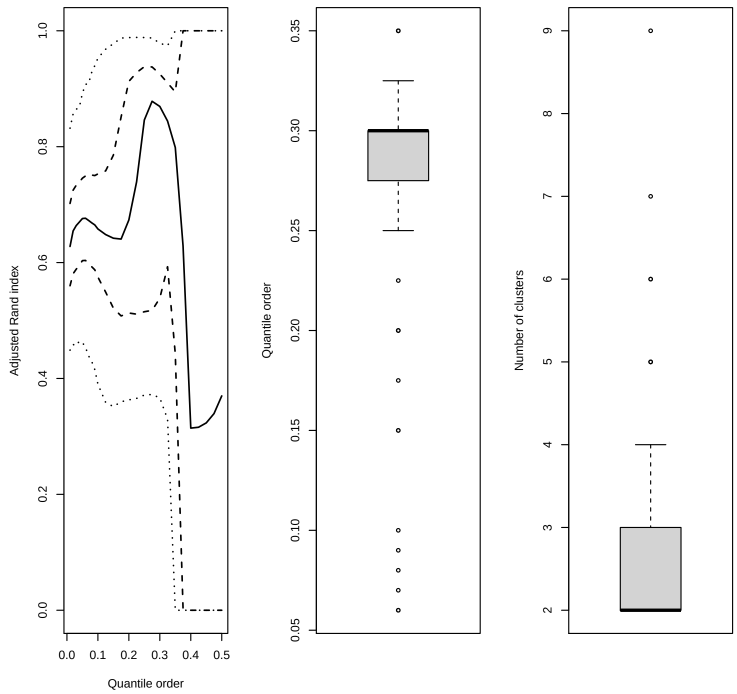


Fig 1: For 100 replications with $n = 1000$ samples for the Circular Bimodal II distribution and LLD with $s = 30$ (i) median adjusted Rand index and interquantile range as a function of the quantile order q, (ii) boxplot of the optimal quantile q (center), and (iii) boxplot of the number of clusters for the optimal quantile q (right).

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 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, as well as KDE, 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-`ms`, 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]	**3**	**0.03**	**0.03**
KDE-`ms` [b]	7	0.37	0.31
LLD [1]	**3**	0.10	0.10
LSD [2]	**3**	0.10	0.10
Hclust [*]		0.16	0.16
Clustering errors for Seeds data			
	Number of clusters	Distance in prob.	Hausdorff distance
KDE [a]	**3**	0.16	0.16
KDE-`ms` [b]	25	0.75	0.33
LLD [1]	**3**	**0.10**	**0.10**
LSD [3]	**3**	0.17	0.17
Hclust [*]		0.20	0.20

[a] `pilot="dunconstr",` $s = 30$
[b] `pilot="dunconstr", mean shift algorithm` [1] $q = 0.05$, $s = 30$.
[2] $q = 10^{-4}$, $s = 20$. [3] $q = 10^{-5}$, $s = 20$.

Table 3: Mean of the clustering errors based on the Hausdorff distance and distance in probability for the Iris and Seeds data. The true number of clusters is specified as input for the hierarchical clustering algorithm.

F.3. True clusters

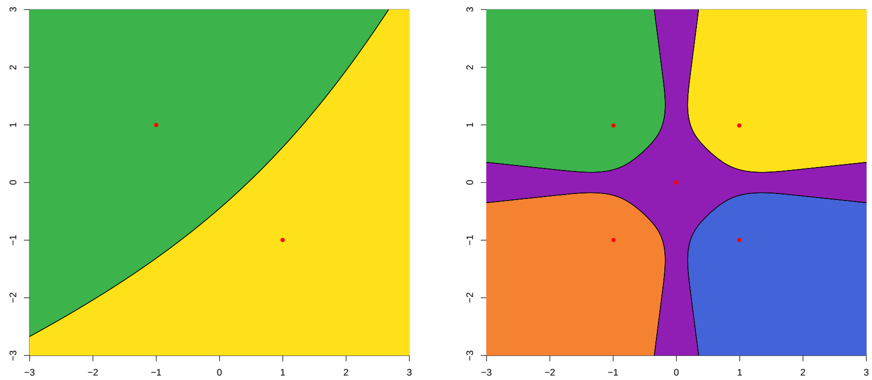
In this subsection we provide the analytical expression for the distributions (H) Bimodal IV, #10 Fountain, Mult. Quadrimodal, and Circular Bimodal II considered in Section F.1 and the corresponding true clusters. We now describe these distributions.
(i) The (H) Bimodal IV density is a mixture of two normal distributions with equal weights, means $\mu_1 = (1, -1)^\top$, $\mu_2 = (-1, 1)^\top$ and covariances

$$\Sigma_1 = \frac{4}{9}\begin{pmatrix} 1 & \frac{7}{10} \\ \frac{7}{10} & 1 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \frac{4}{9}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(ii) 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 Fig 2.
(iii) The Mult. Quadrimodal density is a mixture of four normal distributions



(H) Bimodal IV

#10 Fountain

Fig 2: True clusters associated with the (H) Bimodal IV and #10 Fountain densities.

with means $(-2, 2, 0, 0, 0)$, $(-2, -2, 0, 0, 0)$, $(2, -2, 0, 0, 0)$ and $(2, 2, 0, 0, 0)$. The true clusters for this distribution can be deduced from those of the projection onto the first two components.
(iv) The Circular Bimodal II distribution is a mixture with weights $w_1 = 3/4$ and $w_2 = 1/4$ of the distribution $(X^{(1)} \sin(X^{(2)}), X^{(1)} \cos(X^{(2)}))$, where X is normal with mean $\mu_1 = (2\pi, 0)^T$ and covariance matrix

$$\Sigma_1 = \begin{pmatrix} 0.2 & 0 \\ 0 & 1 \end{pmatrix},$$

and a centered normal distribution with covariance matrix $\Sigma_2 = 2I$. For this distribution, we evaluate, for simplicity, the clustering algorithm on the basis of the membership to the correct mixture components and not on the basis of the membership to the true underlying clusters, that is, the stable manifolds associated to the gradient system (3.2).

References

Agostinelli, C. and Romanazzi, M. (2008). Local depth of multidimensional data. Working Paper No. 3, Ca' Foscari University of Venice.

Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *The Annals of Probability* **21** 1494-1542.

Billingsley, P. (2012). *Probability and Measure* **939**. John Wiley & Sons.

Bremner, D. and Shahsavarifar, R. (2017). An optimal algorithm for computing the spherical depth of points in the plane. *arXiv preprint arXiv:1702.07399*.

Chacón, J. E. (2009). Data-driven choice of the smoothing parametrization for kernel density estimators. *Canadian Journal of Statistics* **37** 249–265.

Chacón, J. E. (2015). A population background for nonparametric density-based clustering. *Statistical Science* **30** 518–532.

Chacón, J. E. (2019). Mixture Model Modal Clustering. *Adv. Data Anal. Classif.* **13** 379–404.

Chacón, J. E. and Duong, T. (2018). *Multivariate kernel smoothing and its applications*. Chapman and Hall/CRC.

Chazal, F., Guibas, L. J., Oudot, S. Y. and Skraba, P. (2013). Persistence-Based Clustering in Riemannian Manifolds. *Journal of the ACM* **60**.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L1 View*. Wiley & Sons.

Dudley, R. M. (1979). Balls in $\mathbb{R}^k$ do not cut all subsets of $k+2$ points. *Advances in Mathematics* **31** 306 - 308.

Duong, T. (2018). ks: kernel smoothing R package version 1.11.3.

Eisenstat, S. C. and Ipsen, I. C. F. (1998). Three absolute perturbation bounds for matrix eigenvalues imply relative bounds. *SIAM Journal on Matrix Analysis and Applications* **20** 149–158.

Fraley, C. and Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* **97** 611-631.

Francisci, G., Agostinelli, C., Nieto-Reyes, A. and Vidyashankar, A. N. (2022). Analytical and statistical properties of local depth functions motivated by clustering applications. *arXiv preprint arXiv:2008.11957*.

Fukunaga, K. and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory* **21** 32–40.

Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models* **40**. Cambridge University Press.

Hardy, M. (2006). Combinatorics of Partial Derivatives. *The electronic journal of combinatorics* **13**.

Holte, J. M. (2009). Discrete Gronwall lemma and applications. In *MAA-NCS meeting at the University of North Dakota*.

Kleindessner, M. and Von Luxburg, U. (2017). Lens depth function and k-relative neighborhood graph: versatile tools for ordinal data analysis. *The Journal of Machine Learning Research* **18** 1889–1940.

Korolyuk, V. S. and Borovskich, Y. V. (2013). *Theory of U-statistics* **273**. Springer Science & Business Media.

Menardi, G. (2016). A review on modal clustering. *International Statistical Review* **84** 413–433.

Rudin, W. (1976). *Principles of mathematical analysis* **3**. McGraw-hill New

York.

Schneemeier, W. (1989). Weak convergence and Glivenko-Cantelli results for empirical processes of U-statistic structure. *Stochastic processes and their applications* **33** 325–334.

Scrucca, L., Fop, M., Murphy, T. B. and Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal* **8** 289–317.

Shahsavarifar, R. (2019). Algorithmic and geometric aspects of data depth with focus on β-skeleton depth. *arXiv preprint arXiv:1905.10716*.

Shahsavarifar, R. and Bremner, D. (2018). Computing the Planar β-skeleton Depth. *arXiv preprint arXiv:1803.05970*.

Torchinsky, A. (1995). *Real variables*. Chapman and Hall/CRC.

Wand, M. P. and Jones, M. C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association* **88** 520–528.

Wheeden, R. L. and Zygmund, A. (2015). *Measure and integral: an introduction to real analysis*. Chapman and Hall/CRC.

Yuan, M. (2018). Efficient fast computing for U-statistics via A-Optimal Subsampling, PhD thesis, Purdue University.