SUPPLEMENTARY MATERIAL - TECHNICAL PROOFS

Auxiliary Results

As a first go, we recall or prove various auxiliary results that are involved in the proof of Theorem 1, and in that of Theorem 2 as well.

The following inequality follows from the well-known Chernoff bound, see *e.g.* (Boucheron et al., 2013).

Lemma 3 *Let $(Z_i)_{i\geq 1}$ be a sequence of i.i.d. random variables valued in $\{0,1\}$. Set $\mu = n\mathbb{E}[Z_1]$ and $S = \sum_{i=1}^n Z_i$. For any $\delta \in (0,1)$ and all $n \geq 1$, we have with probability $1-\delta$:*

$$S \geq \left(1 - \sqrt{\frac{2\log(1/\delta)}{\mu}}\right)\mu.$$

In addition, for any $\delta \in (0,1)$ and $n \geq 1$, we have with probability $1-\delta$:

$$S \leq \left(1 + \sqrt{\frac{3\log(1/\delta)}{\mu}}\right)\mu.$$

Proof Using the Chernoff lower tail (Boucheron et al., 2013), for any $t > 0$ and $n \geq 1$, it holds that

$$\mathbb{P}\left(S < (1-t)\mu\right) \leq \left(\frac{\exp(-t)}{(1-t)^{1-t}}\right)^{\mu}.$$

Because for any $t \in (0,1)$, $\exp(-t)/(1-t)^{1-t} \leq \exp(-t^2/2)$, we obtain that for any $t > 0$ and $n \geq 1$,

$$\mathbb{P}\left(S < (1-t)\mu\right) \leq \exp\left(-\frac{t^2\mu}{2}\right),$$

the bound being obvious when $t \geq 1$. In the previous bound ,choose $t = \sqrt{2\log(1/\delta)/\mu}$ to get the stated inequality. The second inequality is obtained by inverting the Chernoff upper tail:

$$\mathbb{P}\left(S > (1+t)\mu\right) \leq \left(\frac{\exp(t)}{(1+t)^{1+t}}\right)^{\mu}.$$

■

The following inequality is a well-known concentration inequality for sub-Gaussian random variables, see *e.g.* (Boucheron et al., 2013).

Lemma 4 *Suppose that Z is sub-Gaussian with parameter $s^2 > 0$, i.e. Z is real-valued, centred and for all $\lambda > 0$, $\mathbb{E}[\exp(\lambda Z)] \leq \mathbb{E}[\exp(\lambda^2 s^2/2)]$, then with probability $1-\delta$,*

$$|Z| \leq \sqrt{2s^2\log(2/\delta)}.$$

We shall also need a concentration inequality tailored to Vapnik-Chervonenkis (VC) classes of functions. The result stated in Lemma 5 below is mainly a consequence of the work in Giné and Guillou, 2001. Our formulation is slightly different, the role played by the VC constants (v and A below) being clearly quantified.

Let $\mathcal{F}$ be a bounded class of measurable functions defined on $\mathcal{X}$. Let U be a uniform bound for the class $\mathcal{F}$, i.e. $|f(x)| \leq U$ for all $f \in \mathcal{F}$ and $x \in \mathcal{X}$. The class $\mathcal{F}$ is called VC with parameters (v, A) and uniform bound U if

$$\sup_Q \mathcal{N}\left(\epsilon U, \mathcal{F}, L_2(Q)\right) \leq \left(\frac{A}{\epsilon}\right)^v,$$

where $\mathcal{N}(., \mathcal{F}, L_2(Q))$ denotes the covering numbers of the class $\mathcal{F}$ relative to $L_2(Q)$, see *e.g.* (van der Vaart and Wellner, 1996). For notational simplicity and with no loss of generality, we require in the definition of a VC class that $A \geq 3\sqrt{e}$ and $v \geq 1$. Define $\sigma^2 \geq \sup_{f\in\mathcal{F}} \mathrm{Var}(f(X_1))$. We shall work with the condition

$$\sqrt{n}\sigma \geq c_1\sqrt{U^2 v\log(AU/(\sigma\delta))}, \tag{12}$$

where the constant c_1 and c_2 are specified in the following statement.

Lemma 5 *Let $\mathcal{F}$ be a VC class of functions with parameters (v, A) and uniform bound $U > 0$ such that $\sigma \le U$. Let $n \ge 1$ and $\delta \in (0,1)$. There are two positive universal constants c_1 and c_2 such that, under condition (12), we have with probability $1-\delta$,*

$$\sup_{f\in\mathcal{F}} \left| \sum_{i=1}^n \{f(X_i) - \mathbb{E}f(X_1)\} \right| \le c_2 \sqrt{n\sigma^2 v \log(AU/(\sigma\delta))}.$$

Proof Set $\Lambda = v\log(AU/\sigma)$. Using Giné and Guillou, 2001, equation (2.5) and (2.6), we get

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} \left| \sum_{i=1}^n \{f(X_i) - \mathbb{E}f(X_1)\} \right|\right] \le C\sqrt{\Lambda}\left(\sqrt{n}\sigma + U\sqrt{\Lambda}\right) \le 2C\sqrt{n\sigma^2\Lambda},$$
$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} \left| \sum_{i=1}^n \{f(X_i) - \mathbb{E}f(X_1)\}^2 \right|\right] \le \left(\sqrt{n}\sigma + KU\sqrt{\Lambda}\right)^2 \le 4n\sigma^2 := V,$$

where $C > 0$ and $K > 0$ are two universal constants. Both previous inequalities are obtained by taking c_1 large enough. Let

$$Z = \sup_{f\in\mathcal{F}} \left| \sum_{i=1}^n \{f(X_i) - \mathbb{E}f(X_1)\} \right|$$

We recall Talagrand's inequality (Talagrand, 1996, Theorem 1.4) (or Giné and Guillou, 2001, equation (2.7)), for all $t > 0$,

$$\mathbb{P}\left(|Z - \mathbb{E}Z| > t\right) \le K' \exp\left(-\frac{t}{2K'U}\log(1 + 2tU/V)\right),$$

where $K' > 1$ is a universal constant. Using the fact that for all $t \ge 0$, $t/(2+2t/3) \le \log(1+t)$, we get

$$\mathbb{P}\left(|Z - \mathbb{E}Z| > t\right) \le K' \exp\left(-\frac{t^2}{2K'(V + 2tU/3)}\right).$$

Inverting the bound, we find that for any $\delta \in (0,1)$, with probability $1-\delta$,

$$\begin{aligned} |Z - \mathbb{E}Z| &\le \sqrt{2K'V\log(K'/\delta)} + (4K'U/3)\log(K'/\delta) \\ &\le \sqrt{2K'VK''\log(2/\delta)} + (4K'U/3)K''\log(2/\delta) \end{aligned}$$

for some $K'' > 0$. Taking c_1 large enough and using that $AU/\sigma > 2$, we ensure that $2V = 8n\sigma^2 \ge (4U/3)^2K'K''\log(2/\delta)$. Then using the previous bound on the expectation, it follows that with probability $1-\delta$,

$$\begin{aligned} |Z| &\le 2C\sqrt{n\sigma^2\Lambda} + 2\sqrt{8n\sigma^2K'K''\log(2/\delta)} \\ &= 2C\sqrt{n\sigma^2}\left(\sqrt{\Lambda} + \sqrt{8K'K''\log(2/\delta))}\right). \end{aligned}$$

We then conclude by using the bound $\sqrt{a} + \sqrt{b} \le \sqrt{2}\sqrt{a+b}$. ■

Intermediary Results

We now prove some intermediary results used in the core of the proof of the main results.

Define

$$\overline{\tau}_k = \left(\frac{2k}{nb_fV_D}\right)^{1/D}.$$

Proposition 6 *Suppose that Assumption 1 is fulfilled and that $\overline{\tau}_k \le \tau_0$. Then for any $\delta \in (0,1)$ such that $k \ge 4\log(n/\delta)$, we have with probability $1-\delta$:*

$$\hat{\tau}_k(x) \le \overline{\tau}_k.$$

Proof Using Assumption 1 yields

$$\mathbb{P}(X \in \mathcal{B}(x, \overline{\tau}_k)) = \int_{\mathcal{B}(x,\overline{\tau}_k)} f_X \geq b_f \int_{\mathcal{B}(x,\overline{\tau}_k)} d\lambda = b_f V_D \overline{\tau}_k^D = 2k/n.$$

Consider the set formed by the n balls $\mathcal{B}(x, \overline{\tau}_k)$, $1 \leq k \leq n$. From Lemma 3 with $Z_i = \mathbb{1}_{\mathcal{B}(x,\overline{\tau}_k)}(X_i)$, $\mu \geq 2k$, and the union bound, we obtain that for all $\delta \in (0, 1)$ and any $k = 1, \dots, n$:

$$\sum_{i=1}^{n} \mathbb{1}_{\mathcal{B}(x,\overline{\tau}_k)}(X_i) \geq \left(1 - \sqrt{\frac{2\log(n/\delta)}{2k}}\right) 2k.$$

As $k \geq 4\log(n/\delta)$, it follows that

$$\sum_{i=1}^{n} \mathbb{1}_{\mathcal{B}(x,\overline{\tau}_k)}(X_i) \geq k - (\sqrt{4k\log(n/\delta)} - k) \geq k.$$

Hence $\mathbb{P}_n(\mathcal{B}(x, \overline{\tau}_k)) \geq k/n$, denoting by $\mathbb{P}_n$ the empirical distribution of the X_i's. By definition of $\hat{\tau}_k(x)$ it holds that $\hat{\tau}_k(x) \leq \overline{\tau}_k(x)$. ∎

Define

$$\underline{\tau}_k = \left(\frac{k}{2nU_f V_D}\right)^{1/D}.$$

Proposition 7 *Suppose that Assumption 1 is fulfilled and that $\underline{\tau}_k \leq \tau_0$. Then for any $\delta \in (0, 1)$ such that $k \geq 4\log(n/\delta)$, we have with probability $1 - \delta$:*

$$\hat{\tau}_k \geq \underline{\tau}_k.$$

Proof Using Assumption 1 yields

$$\mathbb{P}(X \in \mathcal{B}(x, \underline{\tau}_k)) = \int_{\mathcal{B}(x,\underline{\tau}_k)} f_X \leq U_f \int_{\mathcal{B}(x,\underline{\tau}_k)} d\lambda = U_f V_D \underline{\tau}_k^D = k/(2n).$$

Consider the set formed by the n balls $\mathcal{B}(x, \underline{\tau}_k)$, $1 \leq k \leq n$. From Lemma 3 with $Z_i = \mathbb{1}_{\mathcal{B}(x,\underline{\tau}_k)}(X_i)$, $\mu \leq k/2$, and the union bound, we obtain that for all $\delta \in (0, 1)$ and $k = 1, \dots, n$

$$\sum_{i=1}^{n} \mathbb{1}_{\mathcal{B}(x,\underline{\tau}_k)}(X_i) \leq \left(1 + \sqrt{\frac{6\log(n/\delta)}{k}}\right) k/2.$$

Using that $k \geq 6\log(n/\delta)$, it follows that

$$\sum_{i=1}^{n} \mathbb{1}_{\mathcal{B}(x,\underline{\tau}_k)}(X_i) \leq k + (\sqrt{(6/4)k\log(n/\delta)} - k/2) \leq k.$$

Hence $\mathbb{P}_n(\mathcal{B}(x, \underline{\tau}_k)) \leq k/n$. By definition of $\hat{\tau}_n(k)(x)$ it holds that $\underline{\tau}_k \leq \hat{\tau}_k(x)$. ∎

Proposition 8 *Suppose that Assumption 2 is fulfilled. Then for any $\delta \in (0, 1)$, we have with probability $1 - \delta$:*

$$\left|\sum_{i=1}^{n} \xi_i \mathbb{1}_{\mathcal{B}(x,\hat{\tau}_k(x))}(X_i)\right| \leq \sqrt{2k\sigma^2 \log(2/\delta)}.$$

Proof Set $w_i = \mathbb{1}_{\mathcal{B}(x,\hat{\tau}_k(x))}(X_i)$. Note that $\sum_{i=1}^{n} w_i^2 = k$ almost surely. The result follows from the application of Lemma 4 to the random variable $\sum_{i=1}^{n} \xi_i w_i$, which is sub-Gaussian with parameter $k\sigma^2$. To check this, it is enough to write

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^n \xi_i w_i\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^n \xi_i w_i\right) \mid X_1,\dots X_n\right]\right] \\ &= \mathbb{E}\left[\prod_{i=1}^n \mathbb{E}\left[\exp\left(\lambda \xi_i w_i\right) \mid X_1,\dots X_n\right]\right] \\ &\le \mathbb{E}\left[\prod_{i=1}^n \mathbb{E}\left[\exp\left(\lambda^2\sigma^2 w_i^2/2\right) \mid X_1,\dots X_n\right]\right] \\ &= \mathbb{E}\left[\exp\left(\lambda^2\sigma^2\sum_{i=1}^n w_i^2/2\right)\right] = \exp\left(\lambda^2\sigma^2 k/2\right).\end{aligned}$$

■

Proposition 9 *Suppose that Assumption 1 and 2 are fulfilled and that* $\overline{\tau}_k \le \tau_0$*. Let* $\hat{h}_i := h_i(X_1,\dots,X_n)$ *such that* $a_k = \sup_{i:X_i\in\mathcal{B}(x,\overline{\tau}_k)} |\hat{h}_i|$*. Then for any* $\delta \in (0,1)$ *such that* $k \ge 4\log(2n/\delta)$*, we have with probability* $1-\delta$*:*

$$\left|\sum_{i=1}^n \xi_i \hat{h}_i \mathbb{1}_{\mathcal{B}(x,\hat{\tau}_k(x))}(X_i)\right| \le \sqrt{2k\sigma^2 a_k^2 \log(4/\delta)}.$$

Proof Set $w_i = \mathbb{1}_{\mathcal{B}(x,\hat{\tau}_k(x))}(X_i)$. Note that $\sum_{i=1}^n w_i^2 = k$ almost surely. The result follows from the fact that conditioned upon $X_1,\dots,X_n$, the random variable $\sum_{i=1}^n \xi_i h_i w_i$ is sub-Gaussian with parameter $\sigma^2 k\hat{a}_k^2$ with $\hat{a}_k = \sup_{i:X_i\in\mathcal{B}(x,\overline{\tau}_k)} |\hat{h}_i|$. To check this, it suffices to write

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^n \xi_i \hat{h}_i w_i\right) \mid X_1,\dots X_n\right] &= \prod_{i=1}^n \mathbb{E}\left[\exp\left(\lambda\xi_i\hat{h}_i w_i\right) \mid X_1,\dots X_n\right] \\ &\le \prod_{i=1}^n \exp\left(\lambda^2\sigma^2\hat{h}_i^2 w_i^2/2\right) \\ &= \exp\left(\lambda^2\sigma^2\sum_{i=1}^n \hat{h}_i^2 w_i/2\right) \le \exp\left(\lambda^2\sigma^2 k\hat{a}_k^2/2\right).\end{aligned}$$

Then, for any $t>0$,

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^n \xi_i h_i w_i\right| > t\right) &\le \mathbb{P}\left(\left|\sum_{i=1}^n \xi_i h_i w_i\right| > t,\, \hat{\tau}_k(x) \le \tau_k(x)\right) + \mathbb{P}(\hat{\tau}_k(x) \le \tau_k(x)) \\ &\le \mathbb{E}\left[\mathbb{P}\left(\left|\sum_{i=1}^n \xi_i h_i w_i\right| > t \mid X_1,\dots,X_n\right)\mathbb{1}_{\{\hat{\tau}_k(x)\le\tau_k(x)\}}\right] + \mathbb{P}(\hat{\tau}_k(x) \le \tau_k(x)) \\ &\le \mathbb{E}\left[2\exp(-t^2/(2k\sigma^2\hat{a}_k^2))\mathbb{1}_{\{\hat{\tau}_k(x)\le\tau_k(x)\}}\right] + \mathbb{P}(\hat{\tau}_k(x) \le \tau_k(x)) \\ &\le 2\exp(-t^2/(2k\sigma^2 a_k^2)) + \mathbb{P}(\hat{\tau}_k(x) \le \tau_k(x))\end{aligned}$$

We obtain the result by choosing $t = \sqrt{2k\sigma^2 a_k^2\log(4/\delta)}$ and applying Proposition 6 (to obtain that $\mathbb{P}(\hat{\tau}_k(x) \le \tau_k(x)) \le \delta/2$). ■

Proposition 10 *Suppose that Assumption 1 and 4 is fulfilled. Let* $\tau > 0$*,* $n \ge 1$*, and* $\delta \in (0,1)$ *such that* $\tau \le \tau_0$ *and* $24nU_f(2\tau)^D \ge \log(2D^2/\delta)$*, then with probability* $1-\delta$*,*

$$\begin{aligned}\max_{1\le j,j'\le D}&\left|\sum_{i=1}^n \left\{(X_{i,j}-x)(X_{i,j'}-x)^T\mathbb{1}_{\mathcal{B}(x,\tau)}(X_i) - \mathbb{E}[(X_{1,j}-x)(X_{1,j'}-x)^T\mathbb{1}_{\mathcal{B}(x,\tau)}(X_1)]\right\}\right| \\ &\le (2\tau)^2\sqrt{\frac{2U_f n(2\tau)^D}{3}\log(2D^2/\delta)}.\end{aligned}$$

Proof We use Bernstein's inequality: for any collection $(Z_1, \ldots, Z_n)$ of independent zero-mean random variables such that for all $i = 1, \ldots, n$, $|Z_i| \leq m$ and $\mathbb{E}Z_i^2 \leq v$, it holds that with probability $1 - \delta$,

$$\left| \sum_{i=1}^{n} Z_i \right| \leq \sqrt{2nv \log(2/\delta)} + (m/3) \log(2/\delta).$$

Applying this with

$$\begin{aligned} W_i &= \frac{(X_{i,j} - x)}{2\tau} \frac{(X_{i,j'} - x)}{2\tau} \mathbb{1}_{\mathcal{B}(0,\tau)}(X_i), \\ Z_i &= W_i - \mathbb{E}[W_i], \end{aligned}$$

we can use

$$|Z_i| \leq 2|W_i| \leq 1/4 = m,$$

and

$$\begin{aligned} \mathbb{E}[(W_i - \mathbb{E}W_i)^2] \leq \mathbb{E}[W_i^2] &= \mathbb{E}\left[\left| \frac{(X_{i,j} - x)}{2\tau} \frac{(X_{i,j'} - x)}{2\tau} \right|^2 \mathbb{1}_{\mathcal{B}(0,\tau)}(X_i) \right] \\ &= \int \left| \frac{(y_j - x)}{2\tau} \frac{(y_{j'} - x)}{2\tau} \right|^2 \mathbb{1}_{\mathcal{B}(0,\tau)}(y) f(y) dy \\ &\leq U_f \int \left| \frac{(y_j - x)}{2\tau} \frac{(y_{j'} - x)}{2\tau} \right|^2 \mathbb{1}_{\mathcal{B}(0,\tau)}(y) dy \\ &= U_f (2\tau)^D \int |u_j u_{j'}|^2 \mathbb{1}_{\mathcal{B}(0,1/2)}(u) du \\ &\leq U_f (2\tau)^D \int (u_j^2 + u_{j'}^2)/2 \mathbb{1}_{\mathcal{B}(0,1/2)}(u) du \\ &= U_f (2\tau)^D \int u_1^2 \mathbb{1}_{\mathcal{B}(0,1/2)}(u) du \\ &= U_f (2\tau)^D \int_{[-1/2,1/2]} u_1^2 du_1 = \frac{U_f (2\tau)^D}{12} = v. \end{aligned}$$

We have shown that, with probability $1 - \delta$,

$$\left| \sum_{i=1}^{n} Z_i \right| \leq \sqrt{\frac{nU_f (2\tau)^D}{6} \log(2/\delta)} + (1/12) \log(2/\delta).$$

Because $24nU_f(2\tau)^D \geq \log(2/\delta)$, we obtain that

$$\left| \sum_{i=1}^{n} Z_i \right| \leq 2\sqrt{\frac{nU_f (2\tau)^D}{6} \log(2/\delta)}.$$

Replacing δ by δ/D^2 and using the union bound, we get the desired result. ■

An important quantity in the framework we develop is

$$\sum_{i: X_i \in \mathcal{B}(x, \hat{\tau}_k(x))} (X_{i,j} - x_j),$$

for which we provide an upper bound in the following theorem. Note that we improve upon the straightforward bound of $k\hat{\tau}_k(x)$ which is unfortunately not enough for the analysis carried out here. We shall work with the following assumption

$$C_1 \log(Dn/\delta) \leq k \leq C_2 n, \tag{13}$$

where the two constants $C_1 > 0$ and $C_2 > 0$ are given in the following proposition.

Proposition 11 *Suppose that Assumption 1 and 4 are fulfilled. Let $n \geq 1$, $k \geq 1$ and $\delta \in (0,1)$. There exist universal positive constants C_1, C_2, and C_3 such that, under (13), we have with probability $1-\delta$,*

$$\max_{j=1,\dots,D}\left|\sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))}(X_{i,j}-x_j)\right| \leq C_3\left(\overline{\tau}_k\sqrt{k\log(nD/\delta)}+\frac{Lk\overline{\tau}_k^2}{b_f}\right).$$

Proof Taking C_1 greater than 4, we ensure that $k \geq 4\log(2n/\delta)$. Taking C_2 small enough, we guarantee that $\overline{\tau}_k \leq \tau_0$. From Proposition 6, we have that $\hat\tau_k(x) \leq \overline{\tau}_k$ is valid with probability $1-\delta/2$.

Let $\mu(\tau) = \mathbb{E}[(X_1 - x)\mathbb{1}_{\mathcal{B}(x,\tau)}(X_1)]$. Consider the following decomposition

$$\begin{aligned}|\sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))}(X_{i,j}-x_j)| &\leq \left|\sum_{i=1}^n\{(X_{i,j}-x_j)\mathbb{1}_{\mathcal{B}(x,\hat\tau_k(x))}(X_{i,j})-\mu_j(\hat\tau_k(x))\}\right| + n\mu_j(\hat\tau_k(x)) \\ &\leq \sup_{0<\tau\leq\overline{\tau}_k}\left|\sum_{i=1}^n\{(X_{i,j}-x_j)\mathbb{1}_{\mathcal{B}(x,\tau)}(X_{i,j})-\mu_j(\tau)\}\right| + n\mu_j(\hat\tau_k(x)).\end{aligned}$$

Notice that

$$\begin{aligned}\mu(\tau) &= \int (y-x)\mathbb{1}_{\mathcal{B}(x,\tau)}(y)f(y)dy = (2\tau)^{1+D}\int_{\mathcal{B}(0,1/2)} vf(x+\tau v)dv \\ &= (2\tau)^{1+D}\int_{\mathcal{B}(0,1/2)} v(f(x+\tau v)-f(x))dv.\end{aligned}$$

Hence

$$|\mu_j(\tau)| \leq \frac{L}{2}(2\tau)^{2+D}\int_{\mathcal{B}(0,1/2)} v_j|v|_\infty dv \leq \frac{L}{8}(2\tau)^{2+D} = \frac{L}{8}(2\tau)^{2+D}.$$

And we find

$$\sup_{j=1,\dots,D}|\mu_j(\hat\tau_k)| \leq \frac{L}{8}(2\overline{\tau}_k)^{2+D} = \frac{Lk}{b_f n}\overline{\tau}_k^2.$$

The class of rectangles $\mathcal{R} = \{y \mapsto \mathbb{1}_{\mathcal{B}(x,\tau)}(y) \,:\, \tau > 0\}$ cannot shatter 2 points x_1 and x_2. Considering the case $\|x_1 - x\|_\infty < \|x_2 - x\|\infty$, it fails to pick out x_2. Hence its VC index is $v = 2$. From Theorem 2.6.4 in van der Vaart and Wellner, 1996, we have

$$\mathcal{N}(\epsilon, \mathcal{R}, L_2(Q)) \leq Kv(4e)^v\left(\frac{1}{\epsilon}\right)^{2(v-1)}$$

for any probability measure Q. This implies that $\mathcal{N}(\epsilon, \mathcal{R}, L_2(Q)) \leq (A/\epsilon)^2$, where A is a universal constant. As a result, the class

$$\mathcal{F}_j = \left\{y \mapsto \frac{(y-x_j)}{\overline{\tau}_k}\mathbb{1}_{\mathcal{B}(x,\tau)}(y) \,:\, \tau \in (0,\overline{\tau}_k]\right\},$$

which is uniformly bounded by 1, satisfies the exact same bound for its covering number, that is

$$\mathcal{N}(\epsilon, \mathcal{F}_j, L_2(Q)) \leq \left(\frac{A}{\epsilon}\right)^2.$$

We can therefore apply Lemma 5 with $v = 2$, A a universal constant, $U = 1$ and σ^2 defined as

$$\mathrm{Var}\left(\frac{(X_1-x)_j}{\overline{\tau}_k}\mathbb{1}_{\mathcal{B}(x,\tau)}(X_1)\right) \leq \mathbb{E}[\mathbb{1}_{\mathcal{B}(x,\tau)}(X_1)] \leq \mathbb{E}[\mathbb{1}_{\mathcal{B}(x,\overline{\tau}_k)}(X_1)] \leq \frac{2U_f}{b_f}\frac{k}{n} \leq \frac{4k}{n} := \sigma^2.$$

Condition 12 is valid under (13) when C_1 (resp. C_2) is a large (resp. small) enough constant. The fact that $\sigma^2 \leq 1$ is provided by (13) as well. We obtain that

$$\sup_{0<\tau\leq\overline{\tau}_k}\left|\sum_{i=1}^{n}\{(X_{i,j}-x_j)\mathbb{1}_{\mathcal{B}(x,\tau)}(X_{i,j})-\mu_j(\tau)\}\right| \leq \overline{\tau}_k C\sqrt{kD\log(n/\delta)},$$

where C is a universal constant (C should be large enough to absorb the other constants involved until now). Using the union bound, this bound is extended to a uniform bound over $j \in \{1,\dots,D\}$. We then obtain the statement of the proposition. ■

Proof of Theorem 2

We rely on the bias-variance decomposition expressed in (7). On the first hand, we have

$$\begin{aligned}|m_k(x)-m(x)| &= \left|\frac{\sum_{i=1}^n (m(X_i)-m(x))\mathbb{1}_{\{\mathcal{B}(x,\hat{\tau}_k(x))\}}(X_i)}{\sum_{i=1}^n \mathbb{1}_{\{\mathcal{B}(x,\hat{\tau}_k(x))\}}(X_i)}\right| \\ &\leq \sup_{y\in\mathcal{B}(x,\hat{\tau}_k(x))}|m(y)-m(x)| \\ &\leq L_1\hat{\tau}_k(x).\end{aligned}$$

Applying Lemma 6 we obtain that, with probability $1-\delta/2$,

$$|m_k(x)-m(x)| \leq L_1\overline{\tau}_k.$$

On the other hand, we apply Proposition 8 to get that, with probability $1-\delta/2$,

$$|\hat{m}_k(x)-m_k(x)| \leq \sqrt{\frac{2\sigma^2\log(4/\delta)}{k}}.$$

Proof of Theorem 1

Denote by $\mathbb{X}$ the design matrix of the (local) regression problem

$$\mathbb{X} = (X_{i_1}^c,\dots,X_{i_k}^c)^T$$
$$\mathbb{Y} = (y_{i_1}^c,\dots,y_{i_k}^c)^T.$$

where for any $j = 1,\dots,k$, i_j is such that $X_{i_j} \in \mathcal{B}(x;\hat{\tau}_k(x))$. Define $w = \mathbb{Y}-\mathbb{X}\beta^*$, $\hat{\nu} = \hat{\beta}_k(x)-\beta^*$ Following Hastie et al., 2015, define

$$\mathcal{C}(S,\alpha) = \{u\in\mathbb{R}^D \,:\, \|u_{\overline{S}}\|_1 \leq \alpha\|u_S\|_1\}.$$

and let $\hat{\gamma}_n$ be defined as

$$\hat{\gamma}_n = \inf_{u\in\mathcal{C}(S,3)}\frac{\|\mathbb{X}u\|_2^2}{k\|u\|_2^2}.$$

Hence, $\hat{\gamma}_n$ is the smallest eigenvalue (restricted to the cone) of the design matrix $\mathbb{X}$. From Lemma 11.1 in Hastie et al., 2015, we have the following: whenever

$$\lambda \geq (2/k)\|\mathbb{X}^T w\|_\infty,$$

it holds that

$$\hat{\nu} \in \mathcal{C}(S,3),$$
$$\|\hat{\nu}\|_2 \leq \frac{3\sqrt{\#\mathcal{S}_x}}{\hat{\gamma}_n}\lambda.$$

Consequently, the proof will be completed if, with probability $1-\delta$,

$$\frac{2}{k}\|\mathbb{X}_j^T w\|_\infty \leq \overline{\tau}_k\left(\sqrt{\frac{2\sigma^2\log(16D/\delta)}{k}}+L_2\overline{\tau}_k^2\right), \tag{14}$$

$$\hat{\gamma}_n \geq \frac{\overline{\tau}_k^2}{24\times 8}. \tag{15}$$

Proof of (14). In the next few lines, we show that (14) holds with probability $1-\delta/2$. By definition

$$\mathbb{X}^T w = \sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} w_i^c X_i^c = \sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} w_i X_i^c.$$

Using that $w_i = \xi_i + m(X_i) - \beta^{*T}X_i$,

$$\begin{aligned}\mathbb{X}^T w &= \sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} X_i^c\xi_i + \sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} X_i^c(m(X_i)-\beta^{*T}X_i)\\ &= \sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} X_i^c\xi_i + \sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} X_i^c(m(X_i)-m(x)-\beta^{*T}(X_i-x))\end{aligned}$$

where we have used the covariance structure (with empirically centred terms) to derive the last line. Note that for any $\tau>0$, $\max_{i:X_i\in\mathcal{B}(x,\tau)}|X_{i,j}^c|\le\tau$. Hence, from Proposition 9, because $\overline{\tau}_k\le\tau_0$ and $k\ge 4\log(8Dn/\delta)$ (taking C_1 large enough), we have with probability $1-\delta/(4D)$,

$$\left|\sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} X_{i,j}^c\xi_i\right| \le \sqrt{2k\sigma^2\overline{\tau}_k^2\log(16D/\delta)}$$

Moreover,

$$\sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} |X_{i,j}^c||m(X_i)-m(x)-g(x)^T(X_i-x)| \le kL_2\hat\tau_k(x)^2 \max_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))}|X_{i,j}^c| \le kL_2\hat\tau_k(x)^3$$

Using Proposition 6, because $k\ge 4\log(4Dn/\delta)$, it holds, with probability $1-\delta/(4D)$,

$$\sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} |X_{i,j}^c||m(X_i)-m(x)-\beta^{*T}(X_i-x)| \le kL_2\overline{\tau}_k^3$$

We finally obtain that for any $j=1,\ldots,D$, it holds, with probability $1-\delta/(2D)$,

$$|\mathbb{X}_j^T w| \le \sqrt{2k\sigma^2\overline{\tau}_k^2\log(16/\delta)} + kL_2\overline{\tau}_k^3,$$

and from the union bound, we deduce that, with probability $1-\delta/2$,

$$\max_{j=1,\ldots,D}|\mathbb{X}_j^T w| \le \overline{\tau}_k\left(\sqrt{2k\sigma^2\log(16D/\delta)} + kL_2\overline{\tau}_k^2\right).$$

Proof of (15). We show that (15) holds with probability $1-\delta/2$. Define

$$\hat\Sigma_k = \sum_{i:X_i\in\mathcal{B}(x,\underline{\tau}_k)} (X_i-x)(X_i-x)^T.$$
$$\hat\mu(\tau) = \sum_{i:X_i\in\mathcal{B}(x,\tau)} (X_i-x).$$

First, note that

$$\mathbb{X}^T\mathbb{X} = \sum_{i:X_i\in\mathcal{B}(x,\hat\tau_k(x))} (X_i-x)(X_i-x)^T - k^{-1}\hat\mu(\hat\tau_k)\hat\mu(\hat\tau_k)^T.$$

Then, using Proposition 7, because $k\ge 4\log(4n/\delta)$, with probability $1-\delta/4$, $\hat\tau_k(x)\ge\underline{\tau}_k$, implying that

$$\mathbb{X}^T\mathbb{X} \ge \hat\Sigma_k - k^{-1}\hat\mu(\hat\tau_k)\hat\mu(\hat\tau_k)^T = \mathbb{E}[\hat\Sigma_k] + (\hat\Sigma_k - \mathbb{E}[\hat\Sigma_k]) - k^{-1}\hat\mu(\hat\tau_k)\hat\mu(\hat\tau_k)^T$$

Let $u\in\mathbb{R}^D$. We have that

$$|u^T\hat\mu(\hat\tau_k)|^2 \le \|u\|_1^2 \max_{j=1,\ldots,D}|(\hat\mu(\hat\tau_k))_j|^2 \le \#\mathcal{S}_x\|u\|_2^2 \max_{j=1,\ldots,D}|(\hat\mu(\hat\tau_k))_j|^2.$$

Similarly, we have:

$$|u^T(\hat{\Sigma}_k - \mathbb{E}\hat{\Sigma}_k)u| \leq \|u\|_1^2 \|\hat{\Sigma}_k - \mathbb{E}\hat{\Sigma}_k\|_\infty \leq \#\mathcal{S}_x \|u\|_2^2 \|\hat{\Sigma}_k - \mathbb{E}\hat{\Sigma}_k\|_\infty.$$

Using the variable change $y = x + 2\underline{\tau}_k v$ and that $\underline{\tau}_k \leq \tau_0$, we have that

$$\begin{aligned}
\mathbb{E}\hat{\Sigma}_k &= n\mathbb{E}[(X_1 - x)(X_1 - x)^T 1_{\mathcal{B}(x,\underline{\tau}_k)}(X_1)] = n \int (y-x)(y-x)^T 1_{\{y \in \mathcal{B}(x,\underline{\tau}_k)\}} f(y) dy \\
&\geq n b_f \int (y-x)(y-x)^T 1_{\{y \in \mathcal{B}(x,\underline{\tau}_k)\}} dy = n(2\underline{\tau}_k)^{2+D} b_f \int_{v \in \mathcal{B}(0,1/2)} v v^T dv \\
&= n(2\underline{\tau}_k)^{2+D} b_f \left(\int_{[-1/2,1/2]} v_1^2 dv_1 \right) I_D \\
&= \frac{b_f}{12} n (2\underline{\tau}_k)^{2+D} I_D = \frac{b_f}{6U_f} \underline{\tau}_k^2 k I_D \geq \frac{\underline{\tau}_k^2 k}{12} I_D,
\end{aligned}$$

using that $U_f / b_f \leq 2$. Consequently,

$$\frac{\|\mathbb{X}u\|_2^2}{\|u\|_2^2} \geq \frac{\underline{\tau}_k^2 k}{12} - \#\mathcal{S}_x \left(\|\hat{\Sigma}_k - \mathbb{E}\hat{\Sigma}_k\|_\infty + k^{-1} \max_{j=1,\dots,D} |(\hat{\mu}(\hat{\tau}_k))_j|^2 \right).$$

Proposition 10 can be applied because $24nU_f(2\underline{\tau}_k)^D = 12k \geq \log(16D^2/\delta)$ which is satisfied whenever C_1 is large. Combined with Proposition 11 (our conditions ensure that (13) is satisfied), we obtain that, with probability $1 - \delta/4$,

$$\begin{aligned}
\frac{\|\mathbb{X}u\|_2^2}{\|u\|_2^2} &\geq \frac{\underline{\tau}_k^2 k}{12} - \#\mathcal{S}_x \left(4\underline{\tau}_k^2 \sqrt{\frac{k}{3} \log(16D^2/\delta)} + 2C^2 \left(\overline{\tau}_k^2 \log(8nD/\delta) + \frac{L^2 k \overline{\tau}_k^4}{b_f^2} \right). \right) \\
&\geq \frac{\overline{\tau}_k^2 k}{24 \times 8} \left(2 - \#\mathcal{S}_x C_3 \left(\sqrt{\frac{\log(nD/\delta)}{k}} + \frac{\log(nD/\delta)}{k} + \frac{\overline{\tau}_k^2 L^2}{b_f^2} \right) \right),
\end{aligned}$$

where $C > 0$ is a universal constant. To obtain the last inequality we use $\overline{\tau}_k = C_f^{1/D} \underline{\tau}_k$ with $C_f \leq 8$, we choose $C_3 > 0$ large enough and $C_2 > 0$ small enough. Choose C_1 large enough to get that $C_3 \#\mathcal{S}_x \sqrt{\log(nD/\delta)/k} \leq 1/3$ and $C_3 \#\mathcal{S}_x \log(nD/\delta)/k \leq 1/3$. Finally, noting that $C_3 \#\mathcal{S}_x \overline{\tau}_k^2 L^2 / b_f^2 \leq 1/3$ we obtain the result.

Proof of Theorem 2

We rely on the bias-variance decomposition expressed in (7). On the first hand, we have

$$\begin{aligned}
|m_k(x) - m(x)| &= \left| \frac{\sum_{i=1}^n (m(X_i) - m(x)) \mathbb{1}_{\{\mathcal{B}(x,\hat{\tau}_k(x))\}}(X_i)}{\sum_{i=1}^n \mathbb{1}_{\{\mathcal{B}(x,\hat{\tau}_k(x))\}}(X_i)} \right| \\
&\leq \sup_{y \in \mathcal{B}(x,\hat{\tau}_k(x))} |m(y) - m(x)| \\
&\leq L_1 \hat{\tau}_k(x).
\end{aligned}$$

Applying Lemma 6 we obtain that, with probability $1 - \delta/2$,

$$|m_k(x) - m(x)| \leq L_1 \overline{\tau}_k.$$

On the other hand, we apply Proposition 8 to get that, with probability $1 - \delta/2$,

$$|\hat{m}_k(x) - m_k(x)| \leq \sqrt{\frac{2\sigma^2 \log(4/\delta)}{k}}.$$

Choose C_1 large enough to get that $C|\mathcal{S}_x| \sqrt{\log(2D/\delta)} \leq \sqrt{k}/3$ and $C|\mathcal{S}_x| D \log(2nD/\delta) \leq k/3$. Finally, noting that $C|\mathcal{S}_x| \overline{\tau}_k^2 L^2 \leq b_f^2/3$ we obtain the result.