Meeting Notes for Week 1/23/20 - 1/30/20

Alden Green

February 8, 2020

Suppose we observe samples (y_i, x_i) for $i = 1, \ldots, n$, where $x_1, \ldots, x_n$ are sampled independently from a distribution P with density p, and conditional on X the responses $Y = \{y_1, \ldots, y_n\}$ are assumed to follow the model

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1) \tag{1}$$

Our task is to distinguish

$$\mathbf{H}_0 : f = f_0 := 0 \quad \text{vs} \quad \mathbf{H}_{\text{a}} : f \neq f_0$$

and we worst-case risk to assess performance: for a test ϕ and function class $\mathcal{H}$,

$$\mathcal{R}_\epsilon(\phi; \mathcal{H}) = \mathbb{E}_{f=f_0}(\phi) + \sup_{f \in \mathcal{H}, \|f - f_0\|_{\mathcal{L}^2} \geq \epsilon} \mathbb{E}_f(1 - \phi).$$

Theorem 1 presents our formal result, that the graph spectral test

$$\phi_{\text{spec}}(G_{n,r}) := \mathbf{1}\bigl\{T_{\text{spec}}(G_{n,r}) \geq \tau\bigr\}, \quad T_{\text{spec}}(G_{n,r}) := \frac{1}{n}\sum_{k=1}^{\kappa}\biggl(\sum_{i=1}^{n} v_{k,i}(G_{n,r}) y_i\biggr)$$

is a minimax optimal test over the Sobolev balls $W_\sigma^{s,2}(\mathcal{X}; R)$ whenever $4s > d$ and $\mathcal{X} = [0, 1]^d$.

Here, $W_\sigma^{s,2}(\mathcal{X}; L)$ consists of all those functions $f \in W^{s,2}(\mathcal{X})$ such that $\text{supp}(f^{(\alpha)}) \subseteq \mathcal{X}_\sigma$ for each $|\alpha| \leq s$.

Theorem 1. *Suppose we observe samples $(x_i, y_i)_{i=1}^n$ according to the model (1). Let $R, \sigma > 0$, $b, s, d \geq 1$ be fixed constants, with s and d integers. Suppose that P is an absolutely continuous probability measure over $\mathcal{X} = [0, 1]^d$ with density function $p \in C^{s-1}(\mathcal{X}; p_{\max})$ for some $p_{\max} < \infty$, and further that $p(x)$ is bounded away from zero, i.e. there exists $p_{\min} > 0$ such that*

$$p_{\min} < p(x), \quad \text{for all } x \in \mathcal{X}.$$

Then the following statement holds for all n sufficiently large: if the test $\phi_{\text{spec}}(G_{n,r})$ is performed with parameter choices

$$n^{-1/(2(s-1)+d)} \leq r(n) \leq \min\bigl\{n^{-4/((4s+d)(2+d))}, \sigma/s\bigr\}, \ \kappa = n^{2d/(4s+d)}, \ \tau = \frac{\kappa}{n} + b\sqrt{\frac{2\kappa}{n^2}}$$

then there exists constants c_1, c_2 which do not depend on n, b or R such that for every $\epsilon \geq 0$ satisfying

$$\epsilon^2 \geq c_1^2 \cdot b^2 \cdot \max\{R^2, R^{2d/(4s+d)}\} \cdot n^{-4s/(4s+d)} \tag{2}$$

the worst-case risk is upper bounded

$$\mathcal{R}_\epsilon(\phi_{\text{spec}}; W_\sigma^{s,2}(\mathcal{X}; R)) \leq \frac{c_2}{b}. \tag{3}$$

1 Proofs

1.1 Proof of Theorem 1

We want to apply Lemma 3, which analyzes testing over an arbitrary graph G, to the case where $G = G_{n,r}$. In order to do this, we will need to show that when r and κ are appropriately tuned and $\|f\|_{\mathcal{L}^2(\mathcal{X})}$ is sufficiently large, the inequality (18) holds with respect to $G = G_{n,r}$ and $\beta = (f(x_1), \ldots, f(x_n))$. In particular, we will show that the following statements each hold with probability at least $1 - c/b$ for sufficiently large n (Here and in what follows, c denotes a constant which is fixed in n and f and does not depend on b, but may depend on other fixed quantities such as d, s, etc.):

(E1) **Graph Sobolev norm:** For any $n^{-1/(2(s-1)+d)} \leq r \leq \sigma/s$,

$$f^T L^s f \leq c \cdot b \cdot \|f\|_{W^{s,2}(\mathcal{X})}^2 \cdot n^{s+1} r^{s(d+2)} \tag{4}$$

(E2) **Eigenvalue tail bound:** There exists a constant c such that for any $\kappa = 1, \ldots, n$, if $\max\{3d_\infty(X, \overline{X}), n^{-1/d}\} \leq r \leq \kappa^{-2/(d(2+d))}$ then,

$$\lambda_\kappa \geq c \cdot n r^{d+2} \kappa^{2/d} \tag{5}$$

In particular, the inequality (5) is satisfied for sufficiently large n when $\kappa = n^{2d/(4s+d)}$ and $c(\log n/n)^{1/d} \leq r \leq n^{-4/((2+d)(4s+d))}$.

(E3) **Empirical norm of f:** There exists a constant c_1 such that if $\|f\|_{\mathcal{L}^2(\mathcal{X})} \geq c_1 \cdot b \cdot n^{-2s/(4s+d)}$ and $4s > d$,

$$\|f\|_n^2 \geq \frac{1}{b} \cdot \|f\|_{\mathcal{L}^2}^2 \tag{6}$$

1.1.1 Proof of (4)

The probabilistic bound (4) follows from the more general Lemma 1 by Markov's inequality.

Lemma 1. *Let $\mathcal{X}$ be a Lipschitz domain, let $s \geq 1$ be an integer, and let $0 < \sigma < 1$. Suppose that $f \in W_\sigma^{s,2}(\mathcal{X})$, and further that $p \in C^{s-1}(\mathcal{X}; p_{\max})$ for some constant $p_{\max} < \infty$. Then for any 2nd-order kernel K and any $n^{-1/(2(s-1)+d)} \leq r(n) < \sigma/s$, the expected graph Sobolev seminorm is upper bounded*

$$\mathbb{E}\big[f^T L^s f\big] \leq c \cdot \|f\|_{W^{s,2}(\mathcal{X})}^2 \cdot n^{s+1} r^{s(d+2)} \tag{7}$$

We note that the proof of Lemma 1 is where we rely on the fact that f and its derivatives are supported on $\mathcal{X}_\sigma$. The proof of Lemma 1 is lengthy, and we defer it to Section 1.2.

1.1.2 Proof of (5)

We prove (5) by comparing $G_{n,r}$ to the tensor product of a d-dimensional lattice and a complete graph. The latter is a highly structured graph with known eigenvalues, which as we will see are sufficiently lower bounded for our purposes.

Let $\widetilde{r} = r/(3(\sqrt{d}+1)), M = (1/\widetilde{r})^d$ and $N = n\widetilde{r}^d$; assume without loss of generality that M and N are integers. Additionally for $m = M^{1/d}$ define

$$\overline{Z} = \left\{\frac{1}{m}(j_1, \ldots, j_d) : j \in [m]^d\right\}$$

to be the M evenly spaced grid points over $[0,1]^d$. For a given $\overline{z}_j \in \overline{Z}$, we write $Q(z_j) = m^{-1}[j_1 - 1, j_1] \times \cdots \times m^{-1}[j_d - 1, j_d]$ for the cube of side length $1/m$ with z_j at one corner.

Consider the graph $H = (\overline{X}, E_H)$, where $(\overline{x}_k, \overline{x}_\ell) \in E_H$ if

$$\text{there exists } \overline{z}_i, \overline{z}_j \in \overline{Z} \text{ such that } \overline{x}_k \in Q(\overline{z}_i),\ \overline{x}_\ell \in Q(\overline{z}_j),\ \text{and } \|i - j\|_1 \leq 1.$$

On the one hand $H \cong \overline{G}_d^M \otimes K_N$ where $\overline{G}_d^M$ is the d-dimensional lattice on M nodes, and K_N is the complete graph on N nodes. On the other hand, we now show that when $\max\{3d_\infty(X, \overline{X}), n^{-1/d}\} \leq r$ then $G_{n,r} \succeq H$ as a result of the triangle inequality. If $(\overline{x}_k, \overline{x}_\ell) \in E(H)$, by definition there exist $\overline{z}_i, \overline{z}_j$ connected in $\overline{G}_d^M$ such that $\overline{x}_k \in Q(\overline{z}_i)$ and $\overline{x}_\ell \in Q(\overline{z}_j)$. This implies that $\overline{x}_k$ and $\overline{x}_\ell$ must themselves be close together, since

$$\|\overline{x}_k - \overline{x}_\ell\|_2 \leq \|\overline{x}_k - \overline{z}_i\|_2 + \|\overline{z}_i - \overline{z}_j\|_2 + \|\overline{z}_j - \overline{x}_\ell\|_2 \leq \widetilde{r}(1 + 2\sqrt{d}) = r/3.$$

Since we also assume $r/3 \geq d_\infty(X, \overline{X})$, another application of the triangle inequality gives

$$\|\pi(\overline{x}_k) - \pi(\overline{x}_\ell)\|_2 \leq \|\pi(\overline{x}_k) - \overline{x}_\ell\|_2 + \|\overline{x}_k - \overline{x}_\ell\|_2 \leq \|\overline{x}_\ell - \pi(\overline{x}_\ell)\|_2 \leq r,$$

implying that $(\pi(\overline{x}_k), \pi(\overline{x}_\ell)) \in E$ and consequently that $G_{n,r} \succeq \overline{G}_d^M \otimes K_N$.

The eigenvalues of lattices and complete graphs are known to satisfy, respectively

$$\lambda_k(\overline{G}_d^M) \geq \frac{k^{2/d}}{M^{2/d}} \text{ for } k = 0, \ldots, M-1, \quad \text{and } \lambda_j(K_N) \geq N\mathbf{1}\{j > 0\} \text{ for } j = 0, \ldots, N-1.$$

and by standard facts regarding the eigenvalues of tensor product graphs, we have that the spectrum $\Lambda(H)$ satisfies

$$\Lambda(H) = \Big\{N\lambda_k(\overline{G}_d^M) + M\lambda_j(K_N) : \text{for } k = 0, \ldots, M-1 \text{ and } j = 0, \ldots, N-1\Big\}$$

For all $j = 1, \ldots, N-1$, we have that $M\lambda_j(K_N) = MN = n$. Therefore,

$$\begin{aligned}
\lambda_\kappa(H) &\geq \{n \wedge N\lambda_\kappa(\overline{G}_d^M)\} \\
&\geq \{n \wedge n\widetilde{r}^d \frac{\kappa^{2/d}}{M^{2/d}}\} \\
&\geq \{n \wedge (3\sqrt{d} + 3)^{-(2+d)} n r^{d+2} \kappa^{2/d}\} \\
&\geq (3\sqrt{d} + 3)^{-(2+d)} n r^{d+2} \kappa^{2/d},
\end{aligned}$$

where the last inequality is satisfied since $r \leq \kappa^{-2/(d(d+2))}$, completing the proof of (5).

1.1.3 Proof of (6)

Lemma 2. *Let $\mathcal{X}$ be a Lipschitz domain over which the density is upper and lower bounded*

$$0 < p_{\min} \leq p(x) \leq p_{\max} < \infty \quad \textit{for all } x \in \mathcal{X},$$

and let $f \in W^{s,2}(\mathcal{X})$. Then for any $b \geq 1$, there exists c_1 such that if

$$\|f\|_{\mathcal{L}^2(\mathcal{X})} \geq \begin{cases} c_1 \cdot b \cdot \|f\|_{W^{s,2}(\mathcal{X})} \cdot \max\{n^{-1/2}, n^{-s/d}\}, & \textit{if } 2s \neq d \\ c_1 \cdot b \cdot \|f\|_{W^{s,2}(\mathcal{X})} \cdot n^{-a/2}, & \textit{if } 2s = d \textit{ for any } 0 < a < 1 \end{cases} \tag{8}$$

then,

$$\mathbb{P}\left[\|f\|_n^2 \geq \frac{1}{b}\mathbb{E}\big[\|f\|_n^2\big]\right] \geq 1 - \frac{5}{b} \tag{9}$$

where c_1 and c_2 are constants which may depend only on s, $\mathcal{X}$, d, $p_{\min}$ and $p_{\max}$.

The lower bound (6) results from the more general Lemma 2, which can be verified by checking the various orderings of $2s/(4s + d)$, s/d and $1/2$ whenever $4s < d$.

Proof of Lemma 2: To prove (9) we will show

$$\mathbb{E}\big[\|f\|_n^4\big] \le \left(1 + \frac{1}{b^2}\right) \cdot \big(\mathbb{E}\big[\|f\|_n^2\big]\big)^2$$

whence the claim follows from the Paley-Zygmund inequality (Lemma 9). Since $p \le p_{\max}$ is uniformly bounded, we can relate $\mathbb{E}\big[\|f\|_n^4\big]$ to the $\mathcal{L}^4$ norm,

$$\mathbb{E}\big[\|f\|_n^4\big] = \frac{(n-1)}{n}\left(\mathbb{E}\Big[\|f\|_n^2\Big]\right)^2 + \frac{\mathbb{E}\Big[\big(f(x_1)\big)^4\Big]}{n} \le \left(\mathbb{E}\Big[\|f\|_n^2\Big]\right)^2 + p_{\max}^2\frac{\|f\|_{\mathcal{L}^4}^4}{n}.$$

We will use a Sobolev inequality to relate $\|f\|_{\mathcal{L}^4}$ to $\|f\|_{W_d^{s,2}(\mathcal{X})}$. The nature of this inequality depends on the relationship between s and d (see Theorem 6 in Section 5.6.3 of Evans for a formal statement), so from this point on we divide our analysis into three cases: (i) the case where $2s > d$, (ii) the case where $2s < d$, and (iii) the borderline case $2s = d$.

Case 1: $2s > d$. When $2s > d$, since $\mathcal{X}$ is a Lipschitz domain the Sobolev inequality establishes that $f \in C^\gamma(\overline{\mathcal{X}})$ for some $\gamma > 0$ which depends on s and d, with the accompanying estimate

$$\sup_{x\in\mathcal{X}} |f(x)| \le \|f\|_{C^\gamma(\mathcal{X})} \le c\|f\|_{W^{s,2}(\mathcal{X})}.$$

Therefore,

$$\begin{aligned}
\|f\|_{\mathcal{L}^4}^4 &= \int_{\mathcal{X}} [f(x)]^4\,dx \\
&\le \left(\sup_{x\in\mathcal{X}} |f(x)|\right)^2 \cdot \int_{\mathcal{X}} [f(x)]^2\,dx \\
&\le c\|f\|_{W^{s,2}(\mathcal{X})}^2 \cdot \|f\|_{\mathcal{L}^2(\mathcal{X})}^2.
\end{aligned}$$

Since by assumption

$$\|f\|_{\mathcal{L}^2(\mathcal{X})}^2 \ge c_1^2 \cdot b^2 \cdot \|f\|_{W_d^{s,2}(\mathcal{X})}^2 \cdot \frac{1}{n},$$

we have

$$p_{\max}^2\frac{\|f\|_{\mathcal{L}^4(\mathcal{X})}^4}{n} \le c\|f\|_{W^{s,2}(\mathcal{X})}^2 \cdot \frac{\|f\|_{\mathcal{L}^2(\mathcal{X})}^4}{n\|f\|_{\mathcal{L}^2(\mathcal{X})}^2} \le c\frac{\|f\|_{\mathcal{L}^2(\mathcal{X})}^4}{c_1^2 b^2} \le \frac{\mathbb{E}\big[\|f\|_n^2\big]}{b^2},$$

where the last inequality follows by taking c_1 sufficiently large.

Case 2: $2s < d$. When $2s < d$, since $\mathcal{X}$ is a Lipschitz domain the Sobolev inequality establishes that $f \in \mathcal{L}^q(\mathcal{X})$ for $q = 2d/(d-2s)$, and moreover that

$$\|f\|_{\mathcal{L}^q(\mathcal{X})} \le c\|f\|_{W^{s,2}(\mathcal{X})}.$$

Since $4 = 2\theta + (1-\theta)q$ for $\theta = 2 - d/(2s)$, Lyapunov's inequality implies

$$\|f\|_{\mathcal{L}^4(\mathcal{X})}^4 \le \|f\|_{\mathcal{L}^2}^{2\theta} \cdot \|f\|_{\mathcal{L}^q(\mathcal{X})}^{(1-\theta)q} \le c\|f\|_{\mathcal{L}^2(\mathcal{X})}^4 \cdot \left(\frac{\|f\|_{W^{s,2}(\mathcal{X})}}{\|f\|_{\mathcal{L}^2(\mathcal{X})}}\right)^{d/s}.$$

By assumption, $\|f\|_{\mathcal{L}^2(\mathcal{X})} \ge c_1 b\|f\|_{W^{s,2}(\mathcal{X})} n^{-s/d}$, and therefore

$$p_{\max}^2\frac{\|f\|_{\mathcal{L}^4(\mathcal{X})}^4}{n} \le c\|f\|_{\mathcal{L}^2(\mathcal{X})}^4\left(\frac{\|f\|_{W^{s,2}(\mathcal{X})}}{n^{s/d}\|f\|_{\mathcal{L}^2(\mathcal{X})}}\right)^{d/s} \le \frac{c\|f\|_{\mathcal{L}^2(\mathcal{X})}^4}{c_1 b^{d/s}} \le \frac{\|f\|_{\mathcal{L}^2(\mathcal{X})}^4}{b^2}.$$

where the last inequality follows when c_1 is sufficiently large, and keeping in mind that $d/s > 2$ and $b \geq 1$.

Case 3: $2s = d$. Assume f satisfies (25) for a given $0 < a < 1$. When $2s = d$, since $\mathcal{X}$ is a Lipschitz domain we have that $f \in L^q(\mathcal{X})$ for any $q < \infty$, with the accompanying estimate

$$\|f\|_{\mathcal{L}^q(\mathcal{X})} \leq c\|f\|_{W^{s,2}(\mathcal{X})}.$$

In particular the above holds for $q = 2/(1-a)$ when $1/2 < a < 1$, and for any $q > 4$ when $0 < a < 1/2$. Using Lyapunov's inequality as in the previous case then implies the desired result.

1.2 Proof of Lemma 1

To simplify exposition, we introduce the iterated difference operator, defined recursively as

$$D_{jk}f(x) = (D_k f(x_j) - D_k f(x))\frac{K_r(x_j, x)}{r^d}, \quad D_j f(x) = (f(x_j) - f(x))\frac{K_r(x_j, x)}{r^d} \quad \text{for } j \in [n], k \in [n]^q$$

We will also use the notation $d_j f(x) := (f(x_j) - f(x))$. We split our analysis into cases based on whether s is even or odd.

1.2.1 Case 1: s is even.

When s is even, letting $q = s/2$ we have the decomposition

$$f^T L^s f = r^{ds} \cdot \sum_{i=1}^{n} \sum_{k \in [n]^q} \sum_{\ell \in [n]^q} D_k f(x_i) D_\ell f(x_i). \tag{10}$$

For given index vectors $k, \ell \in [n]^q$ and indices i, j, let $I = |k \cup \ell \cup i|$ be the total number of unique indices. We separate our analysis into cases based on the magnitude of I, specifically whether $I = s+1$ (the leading terms where all indices are distinct) $I < s+1$ (the terms where at least one index is repeated) and show that

$$\mathbb{E}(D_k f(x_i) D_\ell f(x_i)) = \begin{cases} O(r^{2s}) \cdot \|f\|_{W^{s,2}(\mathcal{X})}^2, & \text{if } I = s+1 \\ O(r^2 r^{d(I-(s+1))}) \cdot [f]_{W^{1,2}(\mathcal{X})}^2, & \text{if } I < s+1 \end{cases} \tag{11}$$

We will prove (11) in Section 1.2.3. First, we verify that (10) and (11) are together enough to show Lemma 1 when s is even. In the sum on the right hand side of (10), there are $O(n^I)$ terms with exactly I distinct indices. When $I < s+1$, by (11) the total contribution of such terms to the sum is $O(n^I r^{d(I-1)+2}) \cdot [f]_{W^{1,2}(\mathcal{X})}^2$. Since by assumption $r \geq n^{-1/d}$, this increases with I. Taking $I = s$ to be the largest integer less than $s+1$, the contribution of these terms to the sum is therefore $O(n^s r^{d(s-1)+2}) \cdot [f]_{W^{1,2}(\mathcal{X})}^2$ which in light of the restriction $r \geq n^{-1/(2(s-1)+d)}$ is $O(n^{s+1} r^{s(d+2)}) \cdot [f]_{W^{1,2}(\mathcal{X})}^2$. On the other hand when $I = s+1$, by (11) we immediately have that the total contribution of these terms is $O(n^{s+1} r^{2(s+d)}) \cdot \|f\|_{W^{s,2}(\mathcal{X})}$. Therefore,

$$\mathbb{E}(f^T L^s f) = O(n^{s+1} r^{s(d+2)}) \cdot \|f\|_{W^{s,2}(\mathcal{X})}.$$

1.2.2 Case 2: s is odd.

When s is odd, letting $q = (s-1)/2$ we have

$$f^T L^s f = r^{ds} \cdot \sum_{i,j=1}^{n} \sum_{k \in [n]^q} \sum_{\ell \in [n]^q} \big(d_j D_k f(x_i)\big) \cdot \big(d_j D_\ell f(x_i)\big) \cdot K_r(x_i, x_j). \tag{12}$$

For given index vectors $k, \ell \in [n]^q$ and indices $i, j \in [n]$, let $I = |k \cup \ell \cup i \cup j|$ be the total number of unique indices. Similar to the case when s is even, we show that

$$\mathbb{E}(d_i D_k f(x_j) d_i D_\ell f(x_j)) = \begin{cases} O(r^{2s}) \cdot \|f\|^2_{W^{s,2}(\mathcal{X})}, & \text{if } I = s+1 \\ O(r^2 \cdot r^{d(I-(s+1))}) \cdot [f]^2_{W^{1,2}(\mathcal{X})}, & \text{if } I < s+1 \end{cases} \tag{13}$$

Then Lemma 1 follows from similar reasoning to the case where s was even.

1.2.3 Proof of (11) and (13)

Note that if f is constant almost everywhere in $\mathcal{X}$, the claim is immediate as $D_k f(x_i) = 0$ with probability one. Otherwise $[f]_{W^{1,2}(\mathcal{X})} > 0$, which we shall assume in what follows.

Let $\delta = \min\{r^{2s}, 1\} \cdot [f]_{W^{1,2}(\mathcal{X})} > 0$. Our analysis will make heavy use of Taylor expansions, and we therefore would like to show that there exists some $g \in C^s(\mathcal{X}) \cap W^{s,2}_{\sigma/2}(\mathcal{X})$ such that

$$\Big|\mathbb{E}[D_k f(x_i) D_\ell f(x_i)] - \mathbb{E}[D_k g(x_i) D_\ell g(x_i)]\Big| < \delta, \quad \text{and} \quad [g]_{W^{\ell,2}(\mathbb{R}^d)} \le c[f]_{W^{\ell,2}(\mathbb{R}^d)} \text{ for each } \ell = 0, \ldots, s.$$

Then if (11) and (13) hold with respect to g, they hold (up to constants) with respect to f as well. Note that since f is supported on $\mathcal{X}_\sigma$, the extension

$$\widetilde{f} := \begin{cases} f(x), & x \in \mathcal{X}_\sigma \\ 0, & x \in \mathbb{R}^d - \mathcal{X}_\sigma \end{cases}$$

satisfies $\widetilde{f}(x) = f(x)$ for all $x \in \mathcal{X}$, and $\|\widetilde{f}\|_{W^{s,2}(\mathbb{R}^d)} = \|f\|_{W^{s,2}}(\mathcal{X})$.

To construct our function g, we consider as candidates the convolutions

$$g_m := \widetilde{f} * \eta_{1/m}, \quad \text{for each } m \in \mathbb{N}. \tag{14}$$

where η is the standard mollifier (for a definition see Evans.) We make the following observations:

- Since η is a smooth function, $g_m \in C^\infty(\mathbb{R}^d)$ and therefore $g_m \in C^s(\mathcal{X})$.
- For all sufficiently large m and all multiindices α,

  $$g_m^{(\alpha)}(x) = 0 \;\text{ if } x \in \mathbb{R}^d - X_{\sigma/2}.$$

  and along with the previous observation this implies $g_m \in W^{s,2}_{\sigma/2}(\mathcal{X})$.
- By Theorem 1 of Evans 5.3.1, $\|g_m - \widetilde{f}\|_{W^{s,2}_{\mathrm{loc}}(\mathbb{R}^d)} \to 0$ as $m \to \infty$, which implies $\|g_m - f\|_{W^{s,2}(\mathcal{X})} \to 0$.

By the Cauchy-Schwarz inequality

$$\big|\mathbb{E}\big[D_k f(x_i) D_\ell f(x_i)\big] - \mathbb{E}\big[D_k g_m(x_i) D_\ell g_m(x_i)\big]\big| \le \frac{c}{r^{sd}} \cdot \|f - g_m\|_{\mathcal{L}^2(\mathcal{X})}$$

and taking m to be sufficiently large, we can make the right hand side less than δ. On the other hand, since $\|g_m - f\|_{W^{s,2}(\mathcal{X})} \to 0$, there exists m sufficiently large such that $[g_m]_{W^{\ell,2}(\mathcal{X})}$ is at most say $2[f]_{W^{\ell,2}(\mathcal{X})}$ for each $\ell = 0, \ldots, s$. We therefore take $g = g_m$ for m large enough to satisfy both conditions.

Our task is now to prove that (11) and (13) hold with respect to g. We first prove the desired bounds in the case when some indices are repeated, and then the desired bounds in the case when all indices are distinct.

Repeated indices. Since the proofs of (11) and (13) are essentially the same for the case where some index is repeated, we will assume without loss of generality that s is even. Let $k, \ell \in [n]^q$ be index vectors for $q = s/2$.

When at least one index is repeated, we obtain a sufficient upper bound by reducing the problem of upper bounding the iterated difference operator to that of upper bounding a single difference operator. Letting $k = (k_1, \ldots, k_q)$, we can show by induction that the absolute value of the iterated difference operator $|D_k g(x_i)|$ is upper bounded by

$$|D_k g(x_i)| \leq \left(\frac{2K_{\max}}{r^d}\right)^{q-1} \sum_{h \in k \cup i} \left|D_{k_q} g(x_h)\right| \cdot \mathbf{1}\{G_{n,r}[X_{k\cup i}] \text{ is a connected graph}\}.$$

Therefore,

$$\begin{aligned}|D_k g(x_i)| \cdot |D_\ell g(x_i)| &\leq \left(\frac{2K_{\max}}{r^d}\right)^{2(q-1)} \sum_{h,j \in k\cup\ell\cup i} \left|D_{k_q} g(x_h)\right| \cdot \left|D_{\ell_q} g(x_j)\right| \cdot \mathbf{1}\{G_{n,r}[X_{k\cup i}], G_{n,r}[X_{\ell\cup i}] \text{ are connected graphs.}\} \\ &= \left(\frac{2K_{\max}}{r^d}\right)^{2(q-1)} \sum_{h,j \in k\cup\ell\cup i} \left|D_{k_q} g(x_h)\right| \cdot \left|D_{\ell_q} g(x_j)\right| \cdot \mathbf{1}\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is a connected graph.}\}\end{aligned} \tag{15}$$

We now break our analysis into three cases, based on the number of distinct indices in k_q, ℓ_q, h, j. In each case we will obtain the same rate

$$\mathbb{E}\Big[\left|D_{k_q} g(x_h)\right| \cdot \left|D_{\ell_q} g(x_j)\right|\Big] = O(r^{(|k\cup\ell\cup i|-3)d+2}) \cdot [g]^2_{W^{1,2}(\mathbb{R}^d)},$$

and plugging this back in to (15) we have that for any $k, \ell \in [n]^q$

$$\mathbb{E}\Big[\left|D_k g(x_i)\right| \cdot \left|D_\ell g(x_i)\right|\Big] = O(r^{(|k\cup\ell\cup i|-(2q+1))d+2}) \cdot [g]^2_{W^{1,2}(\mathbb{R}^d)}.$$

Case 1: Two distinct indices. Let $k_q = \ell_q = i$, and $h = j$. Using the law of iterated expectation, we obtain

$$\begin{aligned}\mathbb{E}\left[\big(D_i g(x_j)\big)^2 \cdot \mathbf{1}\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is connected}\}\right] &= \mathbb{E}\left[\big(D_i g(x_j)\big)^2 \cdot \mathbb{P}\big[\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is connected}\}|x_i, x_j\big]\right] \\ &= O(r^{(|k\cup\ell\cup i|-2)d}) \cdot \mathbb{E}\left[\big(D_i g(x_j)\big)^2\right] \\ &= O(r^{(|k\cup\ell\cup i|-3)d}) \cdot \mathbb{E}\left[\big(d_i g(x_j)\big)^2 K_r(x_i, x_j)\right] \\ &= O(r^{(|k\cup\ell\cup i|-3)d+2}) \cdot [g]^2_{W^{1,2}(\mathcal{X})}\end{aligned}$$

where the last equality follows from Lemma 4.

Case 2: Three distinct indices. Let $k_q = \ell_q = i$, for some $i \neq j \neq h$. Using the law of iterated expectation, we obtain

$$\begin{aligned}&\mathbb{E}\Big[\,|D_i g(x_j)| \cdot |D_i g(x_h)| \cdot \mathbf{1}\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is connected}\}\Big] = \\ &\quad \mathbb{E}\Big[|D_i g(x_j)| \cdot |D_i g(x_h)| \cdot \mathbb{P}\big[\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is connected}\}|x_i, x_j, x_h\big]\Big] \\ &\quad = O(r^{(|k\cup\ell\cup i|-3)d}) \cdot \mathbb{E}\Big[|D_i g(x_j)| \cdot |D_i g(x_h)|\Big] \\ &\quad = O(r^{(|k\cup\ell\cup i|-3)d+2}) \cdot [g]^2_{W^{1,2}(\mathbb{R}^d)}\end{aligned}$$

where the last equality follows from Lemma 5.

Case 3: Four distinct indices. Using the law of iterated expectation, we find that

$$\begin{aligned}
&\mathbb{E}\Big[\big|D_{k_q}g(x_i)\big|\cdot\big|D_{\ell_q}g(x_j)\big|\cdot\mathbf{1}\{G_{n,r}[X_{k\cup\ell\cup i}]\text{ is connected}\}\Big]\\
&\quad=\mathbb{E}\Big[\big|D_{k_q}g(x_i)\big|\cdot\big|D_{\ell_q}g(x_j)\big|\cdot\mathbb{P}\big[G_{n,r}[X_{k\cup\ell\cup i}]\text{ is connected}|x_i,x_j,x_{k_q},x_{\ell_q}\big]\Big]\\
&\quad=O(r^{(|k\cup\ell\cup i|-4)d})\cdot\mathbb{E}\Big[\big|D_{k_q}g(x_i)\big|\cdot\big|D_{\ell_q}g(x_j)\big|\cdot\mathbf{1}\{\|x_i-x_j\|\leq(2q+1)r\}\Big]\\
&\quad=O(r^{(|k\cup\ell\cup i|-3)d+2})\cdot[g]^2_{W^{1,2}(\mathbb{R}^d)}
\end{aligned}$$

where the last equality follows from Lemma 6.

All indices distinct. We first show the desired result when s is even, and then when s is odd.

Case 1: s is even.

By Lemma 8 there exists some $f_s\in\mathcal{L}^2(\mathcal{X})$ which satisfies $\|f_s\|_{\mathcal{L}^2(\mathcal{X})}\leq c\|g\|_{W^{s,2}(\mathcal{X})}$ such that

$$\mathbb{E}\Big[\big(D_kg(x)\big)^2\Big]=r^{2s}\cdot\|f_s\|_{\mathcal{L}^2(\mathcal{X})}.$$

Therefore, by the law of iterated expectation

$$\mathbb{E}\big[D_kg(x_i)D_kg(x_j)\big]=\mathbb{E}\Big[\big(\mathbb{E}[D_kg(x_i)|x_i]\big)^2\Big]=r^{2s}\cdot\mathbb{E}\Big[\big(f_s(x_i)\big)^2\Big]\leq r^{2s}\cdot c\|g\|_{W^{s,2}(\mathcal{X})},$$

proving the claimed result.

Case 2: s is odd.

By the law of iterated expectation, we have

$$\begin{aligned}
\mathbb{E}\Big[\Big(d_iD_kg_m(x_j)\Big)\Big(d_iD_\ell g_m(x_j)\Big)K_r(x_i,x_j)\Big]&=\mathbb{E}\Big[\Big(d_i\big(\mathbb{E}[D_kf|x_i,x_j]\big)(x_j)\Big)^2K_r(x_i,x_j)\Big]\\
&=\mathbb{E}\Big[\Big(d_i\big(r^{s-1}\cdot f_{s-1}+r^sf_s(x_j)\big)(x_j)\Big)^2K_r(x_i,x_j)\Big].
\end{aligned}\tag{16}$$

where the latter equality follows from Lemma 8 and f_{s-1} and f_s satisfy the conclusions of that Lemma, namely that $f_{s-1}\in C^1(\mathcal{X})\cap W^{1,2}(\mathcal{X})$ and $f_s\in C^0(\mathcal{X})\cap\mathcal{L}^2(\mathcal{X})$, and

$$\|f_{s-1}\|_{W^{1,2}(\mathcal{X})},\|f_s\|_{\mathcal{L}^2(\mathcal{X})}\leq c\|g\|_{W^{s,2}(\mathcal{X})}.$$

Applying these estimates inside (16), we obtain

$$\begin{aligned}
\mathbb{E}\Big[\Big(d_i\big(r^{s-1}\cdot f_{s-1}+r^sf_s(x_j)\big)(x_j)\Big)^2K_r(x_i,x_j)\Big]&=\mathbb{E}\Big[\Big(r^{s-1}\cdot d_if_{s-1}(x_j)+r^sd_if_s(x_j)\Big)^2K_r(x_i,x_j)\Big]\\
&\leq 2r^{2(s-1)}\mathbb{E}\Big[\big(d_if_{s-1}(x_j)\big)^2K_r(x_i,x_j)\Big]+2r^{2s}\mathbb{E}\Big[\big(d_if_s(x_j)\big)^2\Big]\\
&\leq 2r^{2(s-1)}\mathbb{E}\Big[\big(d_if_{s-1}(x_j)\big)^2K_r(x_i,x_j)\Big]+4r^{2s}\|g\|^2_{W^{s,2}(\mathbb{R}^d)}\\
&\leq r^{2s}\cdot c\|g\|^2_{W^{s,2}(\mathbb{R}^d)}
\end{aligned}$$

where the last inequality follows from Lemma 4. This concludes the proof of (11) and (13).

2 Additional Results

2.1 Fixed Graph Testing

Lemma 3. *Let $1 \leq \kappa \leq n$ be an integer.*

1. ***Type I error:*** *Under the null hypothesis $\beta = \beta_0 = 0$, the Type I error of ϕ_{spec} is upper bounded*

$$\mathbb{E}_{\beta_0}(\phi_{\mathrm{spec}}) \leq \frac{1}{b^2}. \tag{17}$$

2. ***Type II error:*** *For any $b \geq 1$ and β such that*

$$\frac{1}{n}\sum_{i=1}^{n} \beta_i^2 \geq 2b\sqrt{\frac{2\kappa}{n^2}} + \frac{\beta^T L^s \beta}{n\lambda_\kappa^s} \tag{18}$$

the Type II error of ϕ_{spec} is upper bounded,

$$\mathbb{E}_{\beta}(1 - \phi_{\mathrm{spec}}) \leq \frac{3}{b}. \tag{19}$$

2.2 Integrals

For Lemmas 4 - 6, we will assume that K is a kernel function compactly supported on $B(0,1)$ and upper bound $K(x) \leq K_{\max}$.

Lemma 4. *Suppose $g \in C^1(\mathcal{X})$ for $\mathcal{X}$ a Lipschitz domain and that $|p(x)| \leq p_{\max}$ for all $x \in \mathcal{X}$. Then*

$$\mathbb{E}\Big[(g(x_j) - g(x_i))^2 K_r(x_i, x_j)\Big] \leq c K_{\max} p_{\max}^2 r^2 [g]_{W^{1,2}(\mathcal{X})}^2$$

for a constant c which depends only on $\mathcal{X}$ and d.

Proof. Since $\mathcal{X}$ is a Lipschitz domain, we may take $h \in C^1(\mathbb{R}^d)$ to be an extension of g such that $h = g$ a.e. on $\mathcal{X}$, and additionally

$$[h]_{W^{1,2}(\mathbb{R}^d)} \leq c[g]_{W^{1,2}(\mathcal{X})}$$

Since $h = g$ a.e on $\mathcal{X}$, the expectations satisfy

$$\mathbb{E}\Big[(g(x_j) - g(x_i))^2 K_r(x_i, x_j)\Big] = \mathbb{E}\Big[(h(x_j) - h(x_i))^2 K_r(x_i, x_j)\Big].$$

By the fundamental theorem of calculus we have for any $y, x \in \mathbb{R}^d$,

$$h(y) - h(x) = \int_0^1 \frac{d}{dt}\big[h(x + t(y-x))\big]\, dt = \int_0^1 \langle \nabla(h(x + t(y-x))), y - x\rangle\, dt \tag{20}$$

where the integral is well-defined as $\nabla h(z)$ exists almost everywhere since $h \in C^1(\mathbb{R}^d)$. We now perform some standard calculus:

$$\begin{aligned}
\mathbb{E}[(h(x_j) - h(x_i))^2 K_r(x_i, x_j)] &\leq p_{\max}^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (h(y) - h(x))^2 K_r(y, x)\, dy\, dx \\
&= p_{\max}^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\int_0^1 \langle \nabla (h(x + t(y-x))), y - x \rangle\, dt \right)^2 K_r(y, x)\, dy\, dx \\
&\overset{(i)}{\leq} p_{\max}^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\int_0^1 \|\nabla (h(x + t(y-x)))\| \|y - x\|\, dt \right)^2 K_r(y, x)\, dy\, dx \\
&\overset{(ii)}{\leq} p_{\max}^2 r^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\int_0^1 \|\nabla (h(x + t(y-x)))\|\, dt \right)^2 K_r(y, x)\, dy\, dx \\
&\overset{(iii)}{\leq} p_{\max}^2 r^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_0^1 \|\nabla (h(x + t(y-x)))\|^2\, dt K_r(y, x)\, dy\, dx \\
&\overset{(iv)}{\leq} p_{\max}^2 K_{\max} r^{2-d} \int_{\mathbb{R}^d} \int_0^1 \int_{B(x,r)} \|\nabla (h(x + t(y-x)))\|^2\, dy\, dt\, dx \\
&\overset{(v)}{\leq} p_{\max}^2 K_{\max} r^{2-d} \int_{\mathbb{R}^d} \int_0^1 \int_{B(0,r)} \|\nabla (h(x + z))\|^2\, dz\, dt\, dx
\end{aligned}$$

where (i) follows by Cauchy-Schwarz, (ii) follows since either $\|y - x\| \leq r$ or $K_r(y, x) = 0$, (iii) follows by Jensen's, (iv) follows by the assumption $K \leq K_{\max}$ supported on $B(0, 1)$, and (v) follows from the change of variables $z = x + t(y - x)$. Finally, again using Fubini's Theorem, we have

$$\begin{aligned}
K_{\max} r^{2-d} \int_{\mathbb{R}^d} \int_0^1 \int_{B(0,r)} \|\nabla (h(x + z))\|^2\, dz\, dt\, dx &= r^{2-d} \int_{B(0,r)} \int_0^1 \int_{\mathbb{R}^d} \|\nabla (h(x + z))\|^2\, dz\, dt\, dx \\
&= K_{\max} r^2 [h]_{W_d^{1,2}(\mathbb{R}^d)}.
\end{aligned}$$

□

Lemma 5. *Suppose $f \in C^1(\mathcal{X})$ for $\mathcal{X}$ a Lipschitz domain and that $|p(x)| \leq p_{\max}$ for all $x \in \mathcal{X}$. Then*

$$\mathbb{E}\Big[|D_i f(x_h)| \cdot |D_i f(x_j)| \Big] \leq c K_{\max}^3 p_{\max}^3 r^2 [f]_{W^{1,2}(\mathcal{X})}^2$$

for a constant c which depends only on $\mathcal{X}$ and d.

Proof. Since $\mathcal{X}$ is a Lipschitz domain, we may take $g \in C^1(\mathbb{R}^d)$ to be an extension of f such that $g = f$ a.e. on $\mathcal{X}$, and additionally

$$[g]_{W^{1,2}(\mathbb{R}^d)} \leq c[f]_{W^{1,2}(\mathcal{X})}$$

Since $g = f$ a.e on $\mathcal{X}$, the expectations satisfy

$$\mathbb{E}\Big[|D_i f(x_h)| \cdot |D_i f(x_j)| \Big] = \mathbb{E}\Big[|D_i g(x_h)| \cdot |D_i g(x_j)| \Big]$$

We rewrite $\mathbb{E}\big[|D_i g(x_j)| \cdot |D_i g(x_h)| \big]$ as follows,

$$\begin{aligned}
\mathbb{E}\Big[|D_i g(x_j)| \cdot |D_i g(x_h)| \Big] &= \int \int \int |g(z) - g(x)| \cdot |g(z) - g(y)|\, K_r(z, y) K_r(z, x)\, dP(x)\, dP(y)\, dP(x) \\
&= \int \left[\int |g(z) - g(x)|\, K_r(z, x)\, dP(x) \right]^2 dP(z) \\
&\leq p_{\max}^3 \int_{\mathcal{X}} \left[\int_{\mathcal{X}} |g(z) - g(x)|\, K_r(z, x)\, dx \right]^2 dz
\end{aligned}$$

Applying (20) inside the integral gives

$$\begin{aligned}\int_{\mathcal{X}} |g(z)-g(x)|\, K_r(z,x)\, dx &\le \int_{\mathbb{R}^d} |g(z)-g(x)|\, K_r(z,x)\, dx \\ &= \int_{\mathbb{R}^d} \left| \int_0^1 \langle \nabla g(x+t(z-x)), z-x\rangle\, dt \right| K_r(z,x)\, dx \\ &\le \int_{\mathbb{R}^d} \int_0^1 \|\nabla g(x+t(z-x))\| \cdot \|z-x\|\, dt K_r(z,x)\, dx \\ &\le r \int_{\mathbb{R}^d} \int_0^1 \|\nabla g(x+t(z-x))\|\, dt K_r(z,x)\, dx \\ &\le r\frac{K_{\max}}{r^d} \int_{B(z,r)} \int_0^1 \|\nabla g(x+t(z-x))\|\, dt\, dx \\ &\le rK_{\max} \int_{B(0,1)} \int_0^1 \|\nabla g(x-try)\|\, dt\, dy,\end{aligned}$$

and as a result,

$$p_{\max}^3 \int_{\mathcal{X}} \left[\int_{\mathcal{X}} |g(z)-g(x)|\, K_r(z,x)\, dx \right]^2 dz \le c \cdot p_{\max}^3 r^2 K_{\max}^3 [f]^2_{W_d^{1,2}(\mathbb{R}^d)}.$$

□

Lemma 6. *Suppose $f \in C^1(\mathcal{X})$ and that $|p(x)| \le p_{\max}$ for all $x \in \mathcal{X}$. Then for any distinct i,j,k,ℓ each in $[n]$,*

$$\mathbb{E}\Big[|D_k f(x_i)| \cdot |D_\ell f(x_j)| \cdot \mathbf{1}\{\|x_i - x_j\| \le (2q+1)r\} \Big] \le cK_{\max} p_{\max}^2 r^{2+d} [f]^2_{W^{1,2}(\mathbb{R}^d)}$$

for a constant c which depends only on $\mathcal{X}$ and d.

Proof. Since $\mathcal{X}$ is a Lipschitz domain, we may take $g \in C^1(\mathbb{R}^d)$ to be an extension of f such that $g = f$ a.e. on $\mathcal{X}$, and additionally

$$[g]_{W^{1,2}(\mathbb{R}^d)} \le c[f]_{W^{1,2}(\mathcal{X})}$$

Since $g = f$ a.e on $\mathcal{X}$, the expectations satisfy

$$\mathbb{E}\Big[|D_k f(x_i)| \cdot |D_\ell f(x_j)| \cdot \mathbf{1}\{\|x_i - x_j\| \le (2q+1)r\} \Big] = \mathbb{E}\Big[|D_k g(x_i)| \cdot |D_\ell g(x_j)| \cdot \mathbf{1}\{\|x_i - x_j\| \le (2q+1)r\} \Big].$$

We rewrite the expectation as an integral,

$$\begin{aligned}&\mathbb{E}\Big[|D_k g(x_i)| \cdot |D_\ell g(x_j)| \cdot \mathbf{1}\{\|x_i - x_j\| \le (2q+1)r\} \Big] \\ &\quad \le p_{\max}^4 \int_{\mathcal{X}^4} |g(x)-g(y)| \cdot |g(u)-g(v)| \cdot K_r(x,y) K_r(u,v) \mathbf{1}\{\|y-v\| \le (2q+1)r\}\, dy\, dx\, du\, dv\end{aligned}$$

By substituting $z_1 = (y-v)/r$, $z_2 = (u-v)/r$, and $z_3 = (x-y)/r = (x-v)/r + z_1$, we can simplify the integral in the previous display,

$$\begin{aligned}&\int_{\mathcal{X}^4} |g(x)-g(y)| \cdot |g(u)-g(v)| \cdot K_r(x,y) K_r(u,v) \mathbf{1}\{\|y-v\| \le (2q+1)r\}\, dy\, dx\, du\, dv \\ &\quad \le K_{\max}^2 r^d \int_{\mathcal{X}} \int_{[B(0,1)]^3} \Big| g\big((z_3+z_1)r+v\big) - g(z_1 r + v)\Big| \cdot \Big| g(z_2 r + v) - g(v)\Big|\, dz_1\, dz_2\, dz_3\, dv \\ &\quad \le K_{\max}^2 r^{d+2} \int_{[B(0,1)]^3} \int_{[0,1]^2} \int_{\mathcal{X}} \|\nabla g(tz_3 r + z_1 r + v)\| \cdot \|\nabla g(tz_2 r + v)\|\, dv\, dt_1\, dt_2\, dz_1\, dz_2\, dz_3 \\ &\quad \le c\nu_d^3 K_{\max}^2 r^{d+2} [f]^2_{W_d^{1,2}(\mathcal{X})}.\end{aligned}$$

□

Lemma 7. *Suppose that $f \in \mathcal{L}^2(U)$ for some open domain U, and that $h : U \times U \times [0,1] \to \mathbb{R}$ is uniformly bounded. Then, the function $g(x) = \int_0^1 \int_{B(0,1)} f(x + aty)h(y,x,t)\,dy\,dt$ also belongs to $\mathcal{L}^2(U)$, with norm*

$$\|g\|_{\mathcal{L}^2(U)} \le \nu_d \cdot \|f\|_{\mathcal{L}^2(U)} \cdot \|h\|_\infty$$

Proof. We compute the squared norm of g,

$$\begin{aligned}
\|g\|^2_{\mathcal{L}^2(\mathbb{R}^d)} &= \int_U \left(\int_0^1 \int_{B(0,1)} f(x + aty)h(y,x,t)\,dt\,dy \right)^2 dx \\
&\le \|h\|^2_\infty \int_U \left(\int_0^1 \int_{B(0,1)} f(x + aty)\,dt\,dy \right)^2 dx \\
&\le \nu_d^2 \|h\|^2_\infty \int_U \int_0^1 \frac{1}{\nu_d} \int_{B(0,1)} f^2(x + aty)\,dt\,dy\,dx && \text{(Jensen's inequality)} \\
&= \nu_d^2 \|h\|^2_\infty \int_0^1 \int_{B(0,1)} \frac{1}{\nu_d} \int_U f^2(x + aty)\,dt\,dy\,dx && \text{(Fubini's theorem)} \\
&= \nu_d^2 \|h\|^2_\infty \|f\|^2_{\mathcal{L}^2(U)}.
\end{aligned}$$

□

Lemma 8. *Let $k \in [n]^q$ for some $q \ge 1$, let K_r be a second order kernel, and suppose $f \in C^s(\mathcal{X}) \cap W^{s,2}_\sigma(\mathcal{X})$ for some $s \in \mathbb{N}$ and $\sigma > 0$. Let p be a density satisfying $p \in C^0(\mathcal{X}; p_{\max})$ if $s = 0$, and otherwise $p \in C^{s-1}(\mathcal{X}; p_{\max})$ for some $p_{\max} > 0$. For any $qr < \sigma$, there exist functions f_ℓ for $\ell = 2q, \ldots, s-1$ and f_s satisfying $f_\ell \in C^{s-\ell}(\mathcal{X}) \cap W^{s-\ell,2}_{\sigma-qr}(\mathcal{X})$ and*

$$\|f_\ell\|_{W^{s-\ell,2}(\mathcal{X})} \le c\|f\|_{W^{s,2}(\mathcal{X})}$$

for some constant c which depends only on σ, d, $\mathcal{X}$ and $p_{\max}$ such that

$$\mathbb{E}(D_k f(x)) = \begin{cases} \sum_{\ell=2q}^s f_\ell(x) r^\ell, & \text{if } 2q < s \\ r^s \cdot f_s(x), & \text{if } 2q \ge s \end{cases} \tag{21}$$

for any $x \in \mathcal{X}$.

Proof. We proceed by induction on q.

Base case. We begin with the base case of $q = 1$. When $s = 0$, we need to show that $\mathbb{E}[D_k f(x)] = f_s(x)$ for some $f_s \in C^0(\mathcal{X})$ satisfying $\mathrm{supp}(f_s) \subset X_{\sigma-r}$ and $\|f_s\|_{\mathcal{L}^2(\mathcal{X})} \le c\|f\|_{W^{s,2}(\mathcal{X})}$. But

$$\begin{aligned}
\mathbb{E}\Big[D_k f(x)\Big] &= \int f(z) K_r(x,z) p(z)\,dz - f(x)\mathbb{E}\big[K_r(x_k, x)\big] \\
&= \int f(yr + x) K\big(\|y\|\big) p(yr + z)\,dy - f(x)\mathbb{E}\big[K_r(x_k, x)\big],
\end{aligned}$$

whence the boundedness $f_s \in C^0(\mathcal{X})$ follows from the boundedness of f, K and p. We now analyze each term in the above difference. First, we have that $f(x)\mathbb{E}[K_r(x_k, x)] = 0$ unless $x \in \mathcal{X}_\sigma$, and

$$\|f\mathbb{E}[K_r(x_k, \cdot)]\|_{\mathcal{L}^2(\mathcal{X})} \le p_{\max}\|f\|_{\mathcal{L}^2(\mathcal{X})}.$$

Moreover, since K is compactly supported on $B(0,1)$, the integral $\int f(yr + x)K(\|y\|)p(yr + x)\,dy = 0$ unless $x \in \mathcal{X}_{\sigma-r}$. Since

$$\left\| \int f(yr + x) K(\|y\|) p(yr + x)\,dy \right\|_{\mathcal{L}^2(\mathcal{X})} \le p_{\max} K_{\max} \|f\|_{\mathcal{L}^2(\mathcal{X})},$$

the claim follows.

Now, when $s \geq 1$ since $f \in C^s(\mathcal{X})$ it admits a Taylor expansion of the following form for all $x, z \in \mathcal{X}$:

$$f(z) = \sum_{|\alpha|<s} \frac{f^{(\alpha)}(x)}{\alpha!}(x-z)^\alpha + \frac{1}{(s-1)!}\sum_{|\alpha|=s}(x-z)^\alpha \int_0^1 (1-t)^{s-1} f^{(\alpha)}(x+t(z-x))\,dt$$

$f^{(\alpha)} \in C^{s-|\alpha|}(\mathcal{X})$ additionally satisfies

$$\|f^{(\alpha)}\|_{W^{s-|\alpha|,2}(\mathcal{X})} \leq \|f\|_{W^{s,2}(\mathcal{X})} \quad \text{and} \quad \operatorname{supp}\big(f^{(\alpha)}\big) \subset \mathcal{X}_\sigma.$$

Replacing f by its Taylor expansion inside the expected first order difference operator $\mathbb{E}(D_k f(x))$ we have

$$\mathbb{E}(D_k f(x)) = \sum_{1 \leq |\alpha| < s} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,x} + \frac{1}{(s-1)!}\sum_{|\alpha|=s}\int_0^1 (1-t)^{s-1} E_{\alpha,x,t}(f)\,dt \tag{22}$$

where we use the notation $E_{\alpha,x} := \mathbb{E}\left[(x-x_k)^\alpha K_r(x_k,x)\right]$ and $E_{\alpha,x,t}(f) := \mathbb{E}\big[f^{(\alpha)}(x+t(x_k-x))(x_k-x)^\alpha K_r(x_k,x)\big]$.

By a change of variables, we have

$$\begin{aligned} E_{\alpha,x,t}(f) &= \frac{1}{r^d}\int_{\mathbb{R}^d} f^{(\alpha)}\big(x+t(z-x)\big)(z-x)^\alpha K\left(\frac{\|z-x\|}{r}\right)p(z)\,dz \\ &= r^s \int_{\mathbb{R}^d} y^\alpha f^{(\alpha)}\big(x+tyr\big)K\big(\|y\|\big)p(yr+x)\,dy \end{aligned}$$

and as a result the remainder term in (22) reduces to

$$\begin{aligned} \frac{1}{(s-1)!}\sum_{|\alpha|=s}\int_0^1 (1-t)^{s-1}E_{\alpha,x,t}(f)\,dt &= \frac{r^s}{(s-1)!}\sum_{|\alpha|=s}\int_0^1\int_{\mathbb{R}^d}(1-t)^{s-1}y^\alpha f^{(\alpha)}\big(x+tyr\big)K\big(\|y\|\big)p(yr+x)\,dy\,dt \\ &=: r^s g_s(x) \quad \text{for } g_s \in C^0(\mathcal{X}) \cap \mathcal{L}^2_{\sigma-r}(\mathcal{X}), \end{aligned}$$

where additionally $\|g_s\|_{\mathcal{L}^2(\mathcal{X})} \leq c\|f\|_{W^{s,2}(\mathcal{X})}$ by Lemma 7. When $s=1$, there are no multiindices $1 < |\alpha| < s$, and so only the remainder term in (22) is non-zero; thus we have shown (21) when $q=1$ and $s=1$.

When $s \geq 2$, we can analyze $E_{\alpha,x}$ using a Taylor expansion of p. For any $x, z \in \mathcal{X}$, we have

$$p(z) = \sum_{|\beta|<s-1}\frac{p^{(\beta)}(x)}{\beta!}(x-z)^\beta + \frac{1}{(s-2)!}\sum_{|\beta|=s-1}(x-z)^\beta\int_0^1 (1-t)^{s-2}p^{(\beta)}(x+t(z-x))\,dt$$

and $p^{(\beta)} \in C^{s-|\beta|-1}(\mathcal{X})$ additionally satisfies

$$\|p^{(\beta)}\|_{C^{s-|\beta|-1}(\mathcal{X})} \leq p_{\max}.$$

Replacing p by its Taylor expansion, we analyze the term $E_{\alpha,x}$. Note that by assumption $\sigma > r$, and therefore for any $x \in \mathcal{X}_\sigma$,

$$\int_{\mathcal{X}}(x-z)^{\alpha+\beta}K_r(x,z)\,dz = \int_{\mathbb{R}^d}(x-z)^{\alpha+\beta}K_r(x,z)\,dz = r^{|\alpha|+|\beta|}\int_{\mathbb{R}^d}y^{\alpha+\beta}K\big(\|y\|\big)\,dz$$

and similarly

$$\int_{\mathcal{X}}\int_0^1(1-t)^{s-2}(x-z)^{\alpha+\beta}p^{(\beta)}(x+t(z-x))K_r(x,z)\,dz\,dt = r^{|\alpha|+|\beta|}\int_{\mathbb{R}^d}\int_0^1(1-t)^{s-2}y^{\alpha+\beta}p^{(\beta)}(x+try)K\big(\|y\|\big)\,dy\,dt$$

Therefore for any such $x \in \mathcal{X}_\sigma$,

$$\begin{aligned}
E_{\alpha,x} &= \int_{\mathcal{X}} (x-z)^\alpha K_r(x,z) p(z)\, dz \\
&= \sum_{|\beta|=0}^{s-2} \frac{p^{(\beta)}(x)}{\beta!} \int_{\mathcal{X}} (x-z)^{\alpha+\beta} K_r(x,z)\, dz \ \ + \\
&\frac{1}{(s-2)!} \sum_{|\beta|=s-1} \int_{\mathcal{X}} \int_0^1 (1-t)^{s-2} (x-z)^{\alpha+\beta} p^{(\beta)}(x+t(z-x)) K_r(x,z)\, dz\, dt \\
&= \sum_{|\beta|=0}^{s-2} r^{|\alpha|+|\beta|} \frac{p^{(\beta)}(x)}{\beta!} \int_{\mathbb{R}^d} y^{\alpha+\beta} K(y)\, dy + \frac{r^{|\alpha|+s-1}}{(s-2)!} \sum_{|\beta|=s-1} \int_{\mathbb{R}^d} \int_0^1 (1-t)^{s-2} y^{\alpha+\beta} p^{(\beta)}(x+ytr) K\big(\|y\|\big)\, dy\, dt \\
&=: \sum_{|\beta|=0}^{s-2} \mathbf{1}\{|\alpha|+|\beta|>1\} r^{|\alpha|+|\beta|} p_{|\beta|}(x) + r^s p_{s-1}(x) \ \text{ for } p_{|\beta|} \in C^\infty(\mathcal{X}), \|p_{|\beta|}\|_{C^{s-|\beta|-1}(\mathcal{X})} \le c\|p\|_{C^{s-1}(\mathcal{X})},
\end{aligned}$$

where the last line follows since K is a 2nd order kernel. On the other hand when $x \notin \mathcal{X}_\sigma$, the derivatives $f^{(\alpha)}(x) = 0$ for each $\alpha = 1, 2, \ldots, s-1$. As a result, by substituting our expressions for $E_{\alpha,x}$ and $E_{\alpha,x,t}(f)$ back into (22), we obtain that for all $x \in \mathcal{X}$,

$$\begin{aligned}
\mathbb{E}\big[D_k f(x)\big] &= \sum_{|\alpha|=1}^{s-1} \sum_{|\beta|=0}^{s-2} \mathbf{1}\{|\alpha|+|\beta|>1\} r^{|\alpha|+|\beta|} \frac{f^{(\alpha)}(x) p_{|\beta|}(x)}{\alpha!} + r^s \Bigg(\sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x) p_{s-1}(x)}{\alpha!} + g_s(x) \Bigg) \\
&= \sum_{\ell=2}^{s-1} r^\ell \sum_{\substack{|\alpha|+|\beta|=\ell, \\ |\beta|>0}} \frac{f^{(\alpha)}(x) p_{|\beta|}(x)}{\alpha!} + r^s \Bigg(\sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x) p_{s-1}(x)}{\alpha!} + g_s(x) \Bigg) \\
&=: \sum_{\ell=2}^{s-1} r^\ell f_\ell(x) + r^s f_s(x).
\end{aligned}$$

We have already shown $g_s \in C^0(\mathcal{X}) \cap \mathcal{L}^2_{\sigma-r}(\mathcal{X})$. Since additionally $f^{(\alpha)} \in C^{s-|\alpha|}(\mathcal{X}) \cap W^{s-|\alpha|,2}_\sigma(\mathcal{X})$ and $p_{|\beta|} \in C^{s-1-|\beta|}(\mathcal{X})$, the products

$$f^{(\alpha)} \cdot p_{|\beta|} \in W^{m,2}_\sigma(\mathcal{X}) \cap C^m(\mathcal{X}) \ \text{ for } m = \min\{s-|\alpha|, s-1-|\beta|\}.$$

Finally, note that for $|\alpha|+|\beta|=\ell$ the inequality $s-\ell \le m$ holds. Therefore, $f_\ell \in W^{s-\ell,2}_\sigma(\mathcal{X}) \cap C^{s-\ell}(\mathcal{X})$ for each $\ell = 2, 3, \ldots, s$, and

$$\|f^{(\alpha)} p_{|\beta|}\|_{W^{s-\ell,2}(\mathbb{R}^d)} \le \|f^{(\alpha)} p_{|\beta|}\|_{W^{m,2}(\mathbb{R}^d)} \le p_{\max} \|f^{(\alpha)}\|_{W^{s-|\alpha|,2}(\mathcal{X})} \le p_{\max} \|f\|_{W^{s,2}(\mathbb{R}^d)}$$

which finishes the proof of Lemma 8 when $q = 1$.

Induction Step. Now, we assume (21) holds for all $k \in [n]^q$, and prove the desired estimate on $\mathbb{E}[D_k D_j f(x)]$ for each $j \in [n]$. We will build a proof piece-by-piece, depending on the relative size of s and q.

If $s \le 2$, by hypothesis there exists $f_s \in C^0(\mathcal{X}) \cap \mathcal{L}^2_{\sigma-r}(\mathcal{X})$ satisfying

$$\|f_s\|_{\mathcal{L}^2(\mathcal{X})} \le c\|f\|_{W^{s,2}(\mathcal{X})}$$

such that for any $z \in \mathcal{X}$,

$$\mathbb{E}\Big[D_j f(z)\Big] = r^s f_s(z).$$

Therefore by the law of iterated expectation,

$$\mathbb{E}\Big[D_k D_j f(x)\Big] = \mathbb{E}\Big[D_k(\mathbb{E}[D_j f])(x)\Big] = \mathbb{E}\Big[D_k\big(r^s f_s\big)(x)\Big] = r^s \mathbb{E}\Big[D_k f_s(x)\Big]$$

Seeing as $f_s \in C^0(\mathcal{X}) \cap \mathcal{L}^2_{\sigma-r}(\mathcal{X})$, we may apply the inductive hypothesis to obtain that $\mathbb{E}[D_k f_s(x)] \in C^0(\mathcal{X}) \cap \mathcal{L}^2_{\sigma-(q+1)r}(\mathcal{X})$, and additionally

$$\Big\|\mathbb{E}\Big[D_k f_s(x)\Big]\Big\|_{\mathcal{L}^2(\mathcal{X})} \leq c\|f_s\|_{\mathcal{L}^2(\mathcal{X})}. \tag{23}$$

We have therefore established (21) for all q in the case when $s \leq 2$.

Otherwise, when $s \geq 3$ by hypothesis there exist functions $f_\ell \in C^{s-\ell}(\mathcal{X}) \cap W^{s-\ell,2}_{\sigma-r}(\mathcal{X})$ for each $\ell = 2, \ldots, s$ satisfying

$$\|f_\ell\|_{W^{s-\ell,2}(\mathcal{X})} \leq c\|f\|_{W^{s,2}(\mathcal{X})}$$

such that for any $z \in \mathcal{X}$,

$$\mathbb{E}(D_j f(z)) = \sum_{\ell=2}^{s} r^\ell f_\ell(z).$$

By the law of iterated expectation

$$\begin{aligned}
\mathbb{E}\Big[D_k D_j f(x)\Big] &= \mathbb{E}\Big[D_k\big(\mathbb{E}[D_j f]\big)(x)\Big] \\
&= \mathbb{E}\Big[D_k\Big(\sum_{\ell=2}^{s-1} r^\ell f_\ell + r^s f_s\Big)(x)\Big] \\
&= \sum_{\ell=2}^{s-1} r^\ell \cdot \mathbb{E}\Big[D_k f_\ell(x)\Big] + r^s \mathbb{E}\Big[D_k f_s(x)\Big].
\end{aligned}$$

Recalling that $\mathbb{E}\Big[D_k f_s(x)\Big]$ satisfies (23), we now apply the inductive hypothesis to $\mathbb{E}(D_k f_\ell(x))$ for each $\ell = 2, \ldots, s-1$, to prove (21).

First we consider the case when $2(q+1) \geq s$. Note that $2q \geq s - \ell$ for each $\ell \geq 2$. Therefore by hypothesis, for each $\ell = 2, \ldots, s-1$ the expectation $\mathbb{E}[D_k f_\ell(x)] = r^{s-\ell} f_{\ell,s}(x)$ for some $f_{\ell,s} \in C^{s-\ell}(\mathcal{X}) \cap \mathcal{L}^2_{\sigma-(q+1)r}(\mathbb{R}^d)$ satisfying

$$\|f_{\ell,s}\|_{\mathcal{L}^2(\mathcal{X})} \leq c\|f_\ell\|_{W^{s-\ell,2}(\mathcal{X})} \leq c\|f\|_{W^{s,2}(\mathcal{X})} \tag{24}$$

and as a result

$$\sum_{\ell=2}^{s-1} r^\ell \cdot \mathbb{E}(D_k f_\ell(x)) = r^s \sum_{\ell=2}^{s-1} f_{\ell,s}(x)$$

establishing that the second part of (21) holds for all q and all $s \leq 2(q+1)$.

Otherwise $2(q+1) < s-1$. For each $\ell = 2, \ldots, s-1$, if additionally $2q \leq s - \ell - 1$, then by hypothesis

$$\mathbb{E}(D_k f_\ell(x)) = \sum_{m=2q}^{s-\ell-1} r^m \cdot f_{\ell,\ell+m}(x) + r^{s-\ell} f_{\ell,s}$$

where $f_{\ell,\ell+m} \in C^{s-(\ell+m)}(\mathcal{X}) \cap W^{s-(\ell+m),2}_{\sigma-(q+1)r}(\mathcal{X})$ satisfies

$$\|f_{\ell,\ell+m}\|_{W^{s-(\ell+m),2}(\mathbb{R}^d)} \leq c\|f_\ell\|_{W^{s-\ell}} \leq c\|f\|_{W^{s,2}(\mathbb{R}^d)}.$$

On the other hand if $2s > s-\ell-1$, then

$$\mathbb{E}\big[D_k f_\ell(x)\big] = r^s f_{\ell,s}$$

for some $f_{\ell,s} \in C^0(\mathcal{X}) \cap \mathcal{L}^2_{\sigma-(q+1)r}(\mathcal{X})$ which additionally satisfies (24). Therefore,

$$\begin{aligned}
\sum_{\ell=2}^{s-1} r^\ell \cdot \mathbb{E}(D_k f_\ell(x)) &= \sum_{\ell=2}^{s-1-2q} r^\ell \cdot \left\{ \sum_{m=2q}^{s-\ell-1} r^m \cdot f_{\ell,\ell+m}(x) + r^{s-\ell} \cdot f_{\ell,s}(x) \right\} + \sum_{\ell=s-1-2q}^{s-1} r^s \cdot f_{\ell,s}(x) \\
&= \sum_{\ell=2}^{s-1-2q} r^\ell \cdot \left\{ \sum_{m=2q}^{s-\ell-1} r^m \cdot f_{\ell,\ell+m}(x) \right\} + r^s \sum_{\ell=2}^{s-1} f_{\ell,s}(x) \\
&= \sum_{m=2q}^{s-3} \sum_{\ell=2}^{s-m-1} r^{m+\ell} \cdot f_{\ell,\ell+m}(x) + r^s \sum_{\ell=2}^{s-1} f_{\ell,s}(x).
\end{aligned}$$

Rewriting the first sum in the final equation as a sum over $\ell + m = 2(q+1), \ldots, s-1$ establishes (21). ☐

2.3 One-Sided Concentration

The proof of Lemma 2 relies on (a variant of) the Paley-Zygmund Inequality.

Lemma 9. *Let f satisfy the following moment inequality for some $b \geq 1$:*

$$\mathbb{E}\big[\|f\|_n^4\big] \leq \left(1 + \frac{1}{b^2}\right) \cdot \left(\mathbb{E}\big[\|f\|_n^2\big]\right)^2. \tag{25}$$

Then,

$$\mathbb{P}\left[\|f\|_n^2 \geq \frac{1}{b}\mathbb{E}\big[\|f\|_n^2\big]\right] \geq 1 - \frac{5}{b}. \tag{26}$$

Proof. Let Z be a non-negative random variable such that $\mathbb{E}(Z^q) < \infty$. The Paley-Zygmund inequality says that for all $0 \leq \lambda \leq 1$,

$$\mathbb{P}(Z > \lambda \mathbb{E}(Z^p)) \geq \left[(1-\lambda^p)\frac{\mathbb{E}(Z^p)}{(\mathbb{E}(Z^q))^{p/q}}\right]^{\frac{q}{q-p}} \tag{27}$$

Applying (27) with $Z = \|f\|_n^2$, $p = 1$, $q = 2$ and $\lambda = \frac{1}{b}$, by assumption (25) we have

$$\mathbb{P}\Big(\|f\|_n^2 > \frac{1}{b}\mathbb{E}[\|f\|_n^2]\Big) \geq \Big(1 - \frac{1}{b}\Big)^2 \cdot \frac{\big(\mathbb{E}[\|f\|_n^2]\big)^2}{\mathbb{E}[\|f\|_n^4]} \geq \frac{\Big(1-\frac{2}{b}\Big)}{\Big(1+\frac{1}{b^2}\Big)} \geq 1 - \frac{5}{b}.$$

☐