Notes for Week 1/15/20 - 1/22/20

Alden Green

January 25, 2020

Let $X = \{x_1, \ldots, x_n\}$ be a sample drawn i.i.d. from a distribution P on $\mathbb{R}^d$, with density p. For a radius $r > 0$, we define $G_{n,r} = (V, E)$ to be the *r-neighborhood graph* of X, an unweighted, undirected graph with vertices $V = X$, and an edge $(x_i, x_j) \in E$ if and only if $K(\|x_i - x_j\|) := \mathbf{1}\{\|x_i - x_j\| \leq r\}$, where $\|\cdot\|$ is the Euclidean norm. We denote by $A \in \mathbb{R}^{n \times n}$ the adjacency matrix, with entries $A_{uv} = 1$ if $(u, v) \in E$ and 0 otherwise. We also denote by D the diagonal degree matrix, with entries $D_{uu} := \sum_{v \in V} A_{uv}$. The graph Laplacian is $L = D - A$, and we write its spectral decomposition as $L = VSV^T$.

Suppose in addition to the random design points $X = \{x_1, \ldots, x_n\} \sim P$, we observe responses

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

To test whether $f = 0$, we propose the following *eigenvector projection* test statistic:

$$T_{\text{spec}} := \frac{1}{n} \sum_{k=0}^{\kappa} \left(\sum_{i=1}^{n} v_{k,i} y_i \right)^2 \tag{2}$$

where v_k is the kth eigenvector of L (ordered according to eigenvalues $s_1 \leq s_2 \leq \ldots \leq s_n$).

The eigenvector projection test is minimax optimal over the balls in higher order Sobolev spaces $W_d^{s,2}(\mathcal{X})$.

Theorem 1. *Let $b \geq 1$ be a fixed constant, and let d and s be positive integers such that $d < 4s$. Suppose that P is an absolutely continuous probability measure over $\mathcal{X} = [0, 1]^d$ with density function $p \in C^{s-1}(\mathcal{X}; R)$ bounded above and below by constants, i.e*

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

Then the following statement holds: if the test $\phi_{\text{spec}} = \mathbf{1}\{T_{\text{spec}} \geq \tau\}$ is performed with parameter choices

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

then there exists constants c_1, c_2 which may depend on d and s but are independent of the sample size n such that for every $\epsilon \geq 0$ satisfying

$$\epsilon^2 \geq c_1 \cdot b^2 \cdot R^2 \cdot n^{-4s/(4s+d)} \tag{3}$$

the worst-case risk is upper bounded

$$\mathcal{R}_\epsilon(\phi_{\text{spec}}; W_d^{s,2}(\mathcal{X}; R)) \leq \frac{c_2}{b}. \tag{4}$$

Proof sketch. Note: I have actually proved Theorem 1, but for brevity I just included the new parts here.

From previous work, we have that for any $X \in \mathcal{X}^n$ such that

$$\|f\|_n^2 \geq 2b\sqrt{\frac{\kappa}{n^2}} + \frac{f^T L^s f}{n(nr^{(d+2)}\kappa^{1/d})^s}$$

the type II error of the test ϕ_{spec} conditional on X is upper bounded by c/b. By some standard calculations, it is therefore sufficient to show that

$$\|f\|_n^2 \geq \frac{c}{b}\|f\|_{\mathcal{L}^2(\mathcal{X})} \tag{5}$$

and

$$f^T L^s f \leq b \cdot n^{s+1} r^{d(s+2)} \|f\|_{W_d^{s,2}(\mathcal{X})} \tag{6}$$

are each satisfied with probability at least $1 - c/b$. □

1 Proof of (5)

We find it more convenient to work with a normalized version of the graph Sobolev seminorm,

$$R_{s,n}(f) = \frac{1}{n^{s+1} r^{d(s+2)}} f^T L^s f.$$

We have that in expectation, the roughness functional $R_{s,n}(f)$ is (up to constants), no greater than the Sobolev norm $\|f\|_{W_d^{1,2}(\mathcal{X})}$.

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

$$\mathbb{E}\big[R_{s,n}(f)\big] \leq c \cdot \|f\|_{W_d^{s,2}(\mathcal{X})} \tag{7}$$

for some constant c which may depend on s, $p_{\max}$, $K_{\max}$, d and $\mathcal{X}$, but not on f, r or n.

From Lemma 1, the result (5) follows by Markov's inequality.

Note that the bound (7) involves, on the right hand side, the norm $\|f\|_{W_d^{s,2}(\mathcal{X})}$ as opposed to the seminorm $[f]_{W_d^{s,2}(\mathcal{X})}$. To better understand this, consider the following operator

$$L_r f(x) = \int (f(z) - f(x)) K(z,x) \, dP(x)$$

The operator L_r is the expectation of the graph Laplacian, in the sense that $\mathbb{E}(L_n f(x)) = L_r f(x)$, and so it makes sense that the behavior of the associated seminorm $\langle L_r^s f, f \rangle$ is related to the behavior of $\langle L_n^s f, f \rangle$. If p is not uniform, for any s the only functions obviously in the null space of L_r^s are constant functions. Therefore, the magnitude of each derivative $f^{(\alpha)}$ of f is relevant to the overall expected graph Sobolev seminorm.

To be clear, $\mathbb{E}[\langle L_n^s f, f \rangle] \neq \langle L_r^s f, f \rangle$, and bounding the former turns out to be non-trivial. Recall the graph difference operator D_k for $k = (k_1, \ldots, k_n) \in [n]^s$, defined to satisfy the recursive relation

$$D_{k_1} f(x) = (f(x_{k_1}) - f(x)) K_r(x_{k_1}, x), \quad D_k f(x) = \big(D_{k_1}(D_{(k_2,\ldots,k_n)} f)(x)$$

where $K_r(x,z) = \frac{1}{r^d} K(\|x - z\|/r)$. Then when s is even, letting $q = s/2$ we have

$$R_{s,n}(f) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{(nr^2)^q} \sum_{k \in [n]^q} D_k f(x_i) \right)^2 \tag{8}$$

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

$$R_{s,n}(f) = \frac{1}{2n^2r^2} \sum_{i,j=1}^{n} \left(\frac{1}{(nr^2)^q} \sum_{k \in [n]^q} \big(D_k f(x_i) - D_k f(x_j)\big) \right)^2 K_r(x_i, x_j) \tag{9}$$

The proof of Lemma 1 therefore relies heavily on Lemma 2.

Lemma 2. *Let $\mathcal{X}$ be a Lipschitz domain, and suppose $f \in W_d^{s,2}(\mathcal{X})$ for some $s \in \mathbb{N}_+$. Let $q = s/2$ when s is even, and $q = (s-1)/2$ when s is odd. For any indices $k = (k_1, \ldots, k_q)$ and $\ell = (\ell_1, \ldots, \ell_q)$, we have*

$$\mathbb{E}(D_k f(x_i) D_\ell f(x_i)) = \begin{cases} O(r^{2s}) \cdot \|f\|_{W_d^{s,2}(\mathcal{X})}^2, & \text{if all indices are distinct} \\ O(r^2 r^{d(|k \cup \ell \cup i| - (2q+1))}) \cdot [f]_{W_d^{1,2}(\mathcal{X})}^2, & \text{otherwise} \end{cases} \tag{10}$$

Additionally, we have

$$\mathbb{E}(d_i D_k f(x_j) d_i D_\ell f(x_j)) = \begin{cases} O(r^{2s}) \cdot \|f\|_{W_d^{s,2}(\mathcal{X})}^2, & \text{if all indices are distinct} \\ O(r^2 \cdot r^{d(|k \cup \ell \cup i \cup j| - (2q+2))}) \cdot [f]_{W_d^{1,2}(\mathcal{X})}^2, & \text{otherwise} \end{cases} \tag{11}$$

The lower bound on $r(n)$ in Lemma 1 is needed to ensure the leading term in (10) (or (11)), where all indices are distinct, dominates the lower-order terms, where indices are repeated.

2 Proof of Lemma 2

Note that if f is constant almost everywhere, 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 smooth $g \in C^\infty(\mathbb{R}^d)$ such that

$$\Big| \mathbb{E}[D_k f(x_i) D_\ell f(x_i)] - \mathbb{E}[D_k g(x_i) D_\ell g(x_i)] \Big| < \delta,$$

and additionally $[g]_{W^{\ell,2}(\mathbb{R}^d)} \leq c[f]_{W^{\ell,2}(\mathbb{R}^d)}$ for each $\ell = 0, \ldots, s$. Then if (10) and (11) hold with respect to g, they hold (up to constants) with respect to f as well.

To construct such a g, we first take an extension of f to be defined over $\mathbb{R}^d$, and then mollify. Since $\mathcal{X}$ is a Lipschitz domain, there exists an extension citation $\widetilde{f}$ of f compactly supported on $\mathbb{R}^d$ such that $\widetilde{f} = f$ a.e. on $\mathcal{X}$, and $[\widetilde{f}]_{W^{\ell,2}(\mathbb{R}^d)} \leq c[f]_{W^{\ell,2}(\mathcal{X})}$ for each $\ell = 0, \ldots, s$. Since $\widetilde{f} = f$ a.e. on $\mathcal{X}$, the expected difference operators satisfy $\mathbb{E}[D_k \widetilde{f}(x_i) D_\ell \widetilde{f}(x_i)] = \mathbb{E}[D_k f(x_i) D_\ell f(x_i)]$.

Now, since $\widetilde{f} \in W^{s,2}(\mathbb{R}^d)$, there exists a sequence $(g_m) \in C^\infty(\mathbb{R}^d)$ such that $\left\| g_m - \widetilde{f} \right\|_{W^{s,2}(\mathbb{R}^d)} \to 0$. On the one hand, by the Cauchy-Schwarz inequality

$$\Big| \mathbb{E}[D_k \widetilde{f}(x_i) D_\ell \widetilde{f}(x_i)] - \mathbb{E}[D_k g_m(x_i) D_\ell g_m(x_i)] \Big| \leq \frac{c}{r^{sd}} \cdot \|f - g_m\|_{\mathcal{L}^2}(\mathbb{R}^d)$$

and taking m to be sufficiently large, we can make the right hand side less than δ. On the other hand, since $\left\| g_m - \widetilde{f} \right\|_{W^{s,2}(\mathbb{R}^d)} \to 0$, there exists m sufficiently large such that $[g_m]_{W^{s,2}(\mathbb{R}^d)} \leq 2[\widetilde{f}]_{W^{s,2}(\mathbb{R}^d)}$. We take m sufficiently large such that $g = g_m$ satisfies both conditions.

Our task is now to prove that (10) and (11) 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.

2.0.1 Repeated indices.

Since the proofs of (10) and (11) 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,

$$\begin{aligned}
|D_k g(x_i)| \cdot |D_\ell g(x_i)| &\leq \left(\frac{2K_{\max}}{r^d}\right)^{2(q-1)} \sum_{h,j \in k\cup\ell\cup i} \left|D_{k_q} g(x_h)\right| \cdot \left|D_{\ell_q} g(x_j)\right| \cdot \mathbf{1}\{G_{n,r}[X_{k\cup i}], G_{n,r}[X_{\ell\cup i}] \text{ are connected graphs.}\} \\
&= \left(\frac{2K_{\max}}{r^d}\right)^{2(q-1)} \sum_{h,j \in k\cup\ell\cup i} \left|D_{k_q} g(x_h)\right| \cdot \left|D_{\ell_q} g(x_j)\right| \cdot \mathbf{1}\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is a connected graph.}\}
\end{aligned} \tag{12}$$

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 (12) 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

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

where the last equality follows from Lemma 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

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

where the last inequality follows from Lemma 6.

2.0.2 All indices distinct.

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

Case 1: s even. By Lemma 3 there exists some $f_s \in \mathcal{L}^2(\mathbb{R}^d)$ which satisfies $\|f_s\|_{\mathcal{L}^2(\mathbb{R}^d)} \le c\|g\|_{W^{s,2}(\mathbb{R}^d)}$. Therefore, by the law of iterated expectation along with this Lemma,

$$\mathbb{E}\big[D_k g(x_i) D_k g(x_j)\big] = \mathbb{E}\Big[\big(\mathbb{E}[D_k g(x_i)|x_i]\big)^2\Big] = O(r^{2s}) \cdot \mathbb{E}\Big[\big(f_s(x_i)\big)^2\Big] = O(r^{2s}) \cdot \|g\|_{W^{s,2}(\mathbb{R}^d)},$$

proving the claimed result.

Case 2: s odd. By the law of iterated expectation, we have

$$\begin{aligned}
\mathbb{E}[d_i D_k g_m(x_j) d_i D_\ell g_m(x_j) K_r(x_i,x_j)] &= \mathbb{E}\Big[\Big(d_i\big(\mathbb{E}(D_k f)\big)(x_j)\Big)^2 K_r(x_i,x_j)\Big] \\
&= \mathbb{E}\Big[\Big(d_i\big(I_{s-1} \cdot f_{s-1} + O(r^s) f_s(x_j)\big)(x_j)\Big)^2 K_r(x_i,x_j)\Big].
\end{aligned}$$

where the latter equality follows from Lemma 3. Here I_{s-1}, f_{s-1} and f_s satisfy the conclusions of that Lemma, namely that $|I_{s-1}| \le r^{s-1}$, f_{s-1} and $f_s \in C^\infty(\mathbb{R}^d)$, and

$$\|f_{s-1}\|_{W^{1,2}(\mathbb{R}^d)}, \|f_s\|_{\mathcal{L}^2(\mathbb{R}^d)} \le c\|g\|_{W^{s,2}(\mathbb{R}^d)}.$$

By the linearity and boundedness of the difference operator d_i, we have

$$\begin{aligned}
\mathbb{E}\Big[\Big(d_i\big(I_{s-1} \cdot f_{s-1} + O(r^s) f_s(x_j)\big)(x_j)\Big)^2 K_r(x_i,x_j)\Big] &= \mathbb{E}\Big[\Big(I_{s-1} d_i f_{s-1}(x_j) + O(r^s) f_s(x_j)\Big)^2 K_r(x_i,x_j)\Big] \\
&\le 2I_{s-1}^2 \mathbb{E}\Big[\big(d_i f_{s-1}(x_j)\big)^2 K_r(x_i,x_j)\Big] + O(r^{2s})\mathbb{E}\Big[f_s(x_j)^2\Big] \\
&= O(r^{2(s-1)})\mathbb{E}\Big[\big(d_i f_{s-1}(x_j)\big)^2 K_r(x_i,x_j)\Big] + O(r^{2s})\|g_m\|_{W^{s,2}(\mathbb{R}^d)}^2 \\
&\le O(r^{2s}) \cdot \|g_m\|_{W^{s,2}(\mathbb{R}^d)}^2
\end{aligned}$$

where the last inequality follows from Lemma 4.

2.1 Additional Lemmas

We use multiindex notation to represent higher order partial derivatives and polynomials. For $\alpha \in [\mathbb{N}]^d$, and $x, z \in \mathbb{R}^d$ we write

$$|\alpha| = \alpha_1 + \cdots + \alpha_d, \quad f^{(\alpha)} = \frac{\partial^{|\alpha|} f}{\partial x_{\alpha_1} \ldots \partial x_{\alpha_q}}, \quad (x - z)^\alpha := (x_{\alpha_1} - z_{\alpha_1}) \cdots (x_{\alpha_d} - z_{\alpha_d})$$

Lemma 3. *Let $k \in [n]^q$ for some $q \geq 1$. Suppose that $g \in C^{\infty}(\mathbb{R}^d)$ and $p \in C^{s-1}(\mathbb{R}^d; p_{\max})$, and that K_r is a second order kernel. Then there exist*

- *functions $f_\ell, \ell = 2q, \ldots, s$ satisfying $f_\ell \in W_d^{s-\ell,2}(\mathbb{R}^d) \cap C_d^{\infty}(\mathbb{R}^d)$ and*

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

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

- *constants $I_\ell, \ell = 2q, \ldots, s$ which depend only on $K(\cdot)$ satisfying $|I_\ell| \leq r^\ell$,*

such that

$$\mathbb{E}(D_k f(x)) = \begin{cases} \sum_{\ell=2q}^{s-1} I_\ell \cdot f_\ell(x) + O(r^s) \cdot f_s(x), & \text{if } 2q < s \\ O(r^s) \cdot f_s(x), & \text{if } 2q \geq s \end{cases} \tag{13}$$

The $O(\cdot)$ term may depend on $s, K_{\max}$ and $p_{\max}$, but does not depend on f.

Proof. We proceed by induction on q.

Base case. We begin with the base case of $q = 1$. Since f and p are smooth, they both admit Taylor expansions of the following form for all $x, z \in \mathbb{R}^d$:

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

$$p(z) = \sum_{|\beta|<s-1} \frac{p^{(\beta)}(x)}{\beta!}(x-z)^\beta + O((x-z)^{s-1})$$

where $f^{(\alpha)} \in W^{s-|\alpha|,2}(\mathbb{R}^d) \cap C^{\infty}(\mathbb{R}^d)$ additionally satisfies

$$\left\|f^{(\alpha)}\right\|_{W^{s-|\alpha|,2}(\mathbb{R}^d)} \leq \|f\|_{W^{s,2}(\mathbb{R}^d)}$$

Replacing f by its Taylor expansion inside the expected first order difference operator $\mathbb{E}(D_k f(x))$ and letting $E_{\alpha,P} := \mathbb{E}\left[(x-x_k)^\alpha K_r(x_k, x)\right]$, we have

$$\begin{aligned}
\mathbb{E}(D_k f(x)) &= \sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,P} + \frac{|\alpha|}{\alpha!}\sum_{|\alpha|=s}\int_0^1 (1-t)^{s-1}\mathbb{E}\big[f^{(\alpha)}(x+t(x_k-x))(x_k-x)^\alpha K_r(x_k,x)\big]\,dt \\
&= \sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,P} + O(r^s)\sum_{|\alpha|=s}\int_0^1 \mathbb{E}\big[f^{(\alpha)}(x+t(x_k-x))K_r(x_k,x)\big]\,dt \\
&= \sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,P} + O(r^s)\sum_{|\alpha|=s}\int_0^1 \int_{B(x,r)} \frac{f^{(\alpha)}(x+t(z-x))}{r^d}\,dz\,dt \\
&= \sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,P} + O(r^s)\int_0^1 \int_{B(0,1)} \sum_{|\alpha|=s} f^{(\alpha)}(x+try)\,dy\,dt
\end{aligned} \tag{14}$$

Turning our attention now to the expectation $E_{\alpha,P}$, by replacing p with its Taylor expansion we obtain

$$E_{\alpha,P} = \int_{\mathbb{R}^d}(x-z)^\alpha K_r(x,z)p(z)\,dz \sum_{|\beta|=0}^{s-2}\frac{p^{(\beta)}(x)}{\beta!}\underbrace{\int_{\mathbb{R}^d}(x-z)^{\alpha+\beta}K_r(x,z)\,dz}_{:=I_{|\alpha|+|\beta|}} + O(r^s)$$

where the first sum equals zero when $s=1$. Plugging this back in to (14) yields

$$\mathbb{E}(D_k f(x)) = \sum_{\alpha=1}^{s-1} \sum_{|\beta|=0}^{s-2} \frac{f^{(\alpha)}(x) p^{(\beta)}(x)}{\alpha!\beta!} I_{|\alpha|+|\beta|} + \\ O(r^s) \cdot \underbrace{\left(\sum_{|\alpha|=1}^{s-1} f^{(\alpha)}(x) + \int_0^1 \int_{B(0,1)} \sum_{|\alpha|=s} f^{(\alpha)}(x+try)\, dy\, dt \right)}_{:=g_s(x)}$$

where $g_s \in \mathcal{L}^2(\mathbb{R}^d)$ by Lemma 7. We are now in a position to prove the second part of (13) where $q=1$ and $s \in \{1,2\}$. When $s=1$, the sum in the prior expression is over no terms and is equal to zero. Since the integral $I_1 = 0$, the first term is also zero in the case where $s=2$. For $s \in \{1,2\}$, defining $f_s := g_s$, we have $\mathbb{E}(D_k f(x)) = O(r^s) f_s(x)$ for $f_s \in \mathcal{L}_d^2(\mathcal{X}; R)$. This proves the second part of (13) when $q=1$.

Otherwise when $s > 2$ and $q=1$, we must analyze the first term in the prior expression. Since $f^{(\alpha)} \in W^{s-|\alpha|,2}(\mathbb{R}^d)$ and $p^{(\beta)} \in C^{s-1-|\beta|}(\mathbb{R}^d)$, and $\min\{s-|\alpha|, s-1-|\beta|\} \geq s - (|\alpha|+|\beta|)$, the product $f^{(\alpha)} p^{(\beta)}$ belongs to $W^{s-(|\alpha|+|\beta|),2}(\mathbb{R}^d)$, and moreover

$$\left\| f^{(\alpha)} p^{(\beta)} \right\|_{W^{s-(|\alpha|+|\beta|),2}(\mathbb{R}^d)} \leq p_{\max} \left\| f^{(\alpha)} \right\|_{W^{s-(|\alpha|+|\beta|),2}(\mathbb{R}^d)} \leq p_{\max} \|f\|_{W^{s,2}(\mathbb{R}^d)}.$$

Additionally, the integral $I_1 = 0$, and for $\ell > 1$,

$$|I_\ell| \leq r^\ell \int K_r(x,z)\, dz = r^\ell$$

Therefore,

$$\sum_{\alpha=1}^{s-1} \sum_{|\beta|=0}^{s-2} \frac{f^{(\alpha)}(x) p^{(\beta)}(x)}{\alpha!\beta!} I_{|\alpha|+|\beta|} = \sum_{\ell=1}^{s-1} \sum_{\substack{|\alpha|+|\beta|=\ell, \\ |\alpha|>0}} \frac{f^{(\alpha)}(x) p^{(\beta)}(x)}{\alpha!\beta!} I_\ell + O(r^s) \sum_{|\alpha|=2}^{s-1} \sum_{|\beta|=s-|\alpha|}^{s-2} \frac{f^{(\alpha)}(x) p^{(\beta)}(x)}{\alpha!\beta!}$$

$$= \sum_{\ell=2}^{s-1} I_\ell \left(\underbrace{\sum_{\substack{|\alpha|+|\beta|=\ell, \\ |\alpha|>0}} \frac{f^{(\alpha)}(x) p^{(\beta)}(x)}{\alpha!\beta!}}_{:=f_\ell(x)} \right) + O(r^s) \underbrace{\sum_{|\alpha|=2}^{s} \sum_{|\beta|=s-|\alpha|}^{s-1} \frac{f^{(\alpha)}(x) p^{(\beta)}(x)}{\alpha!\beta!}}_{:=h_s(x)}$$

and setting $f_s = g_s + h_s$ implies the first part of (13) when $q=1$.

Induction Step. In this part of the proof, the functions f_ℓ for $\ell = 2, \ldots, s$ and the constants I_ℓ for $\ell = 2, \ldots, s$ may change from line to line, but will always satisfy the conditions in the theorem statement.

Now, we assume (13) holds for all $k \in [n]^q$, and prove the desired estimate on $\mathbb{E}(D_{kj} f(x))$ for each $j \in [n]$. By the law of iterated expectation and our analysis of the base case,

$$\begin{aligned} \mathbb{E}(D_{kj} f(x)) &= \mathbb{E}(D_k(\mathbb{E}(D_j f))(x)) \\ &= \mathbb{E}\left(D_k \left(\sum_{t=2}^{s-1} I_t f_t + O(r^s) f_s \right)(x) \right) \\ &= \sum_{t=2}^{s-1} I_t \cdot \mathbb{E}(D_k f_t(x)) + O(r^s) f_s(x) \end{aligned}$$

where the second equality follows from the linearity and boundedness of $f \mapsto \mathbb{E}(D_k f)$. We now apply the inductive hypothesis to $\mathbb{E}(D_k f_t(x))$ for each $t = 2, \ldots, s-1$, to prove each part of (13).

First we consider the case when $2(q+1) \geq s$. Note that $f_t \in W^{s-t,2}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$, and $t \geq 2$ implies $2q \geq s-t$. Therefore by hypothesis $\mathbb{E}(D_k f_t(x)) = O(r^{s-t}) f_s$ for some $f_s \in \mathcal{L}^2(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$. As a result

$$\sum_{t=2}^{s-1} I_t \cdot \mathbb{E}(D_k f_t(x)) = \sum_{t=2}^{s-1} I_t \cdot O(r^{s-t}) f_s(x) = O(r^s) f_s(x)$$

establishing that the second part of (13) holds for all q.

Otherwise $2(q+1) < s-1$. For each $t = 2, \ldots, s-1$, if additionally $2q \leq s-t-1$, then by hypothesis $\mathbb{E}(D_k f_t(x)) = \sum_{\ell=2q}^{s-t-1} I_\ell \cdot f_{\ell+t}(x) + O(r^{s-t}) f_s$, and otherwise $\mathbb{E}(D_k f_t(x)) = O(r^{s-t}) f_s(x)$. Therefore,

$$\begin{aligned}
\sum_{t=2}^{s-1} I_t \cdot \mathbb{E}(D_k f_t(x)) &= \sum_{t=2}^{s-1-2q} I_t \cdot \left\{ \sum_{\ell=2q}^{s-t-1} I_\ell \cdot f_{\ell+t}(x) + O(r^{s-t}) \cdot f_s(x) \right\} + \sum_{t=s-1-2q}^{s-1} I_t \cdot O(r^{s-t}) \cdot f_s(x) \\
&= \sum_{t=2}^{s-1-2q} I_t \cdot \left\{ \sum_{\ell=2q}^{s-t-1} I_\ell \cdot f_{\ell+t}(x) \right\} + O(r^s) \cdot f_s(x) \\
&= \sum_{\ell=2q}^{s-3} \sum_{t=2}^{s-\ell-1} I_{\ell+t} \cdot f_{\ell+t}(x) + O(r^s) \cdot f_s(x).
\end{aligned}$$

Rewriting the final equation as a sum over $\ell + t = 2(q+1), \ldots, s-1$ establishes (13). □

Lemma 4. *Suppose $g \in C^\infty(\mathbb{R}^d)$, that $|p(x)| \leq p_{\max}$ for all $x \in \mathbb{R}^d$, and that K is a 2nd order kernel. Then*

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

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

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

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

By the upper bound on p, we obtain

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

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

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

□

Lemma 5. *Suppose $g \in C^\infty(\mathbb{R}^d)$, that $|p(x)| \leq p_{\max}$ for all $x \in \mathbb{R}^d$, and that K is a 2nd order kernel. Then*

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

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

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

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

Then we obtain

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

and as a result,

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

□

Lemma 6. *Suppose $g \in C^\infty(\mathbb{R}^d)$, that $|p(x)| \leq p_{\max}$ for all $x \in \mathbb{R}^d$, and that K is a 2nd order kernel. Then*

$$\mathbb{E}\Big[\big|D_{k_q} g(x_i)\big| \cdot \big|D_{\ell_q} g(x_j)\big| \cdot \mathbf{1}\{\|x_i - x_j\| \leq (2q+1)r\}\Big] \leq c K_{\max} p_{\max}^2 r^{2+d} [g]_{W^{1,2}(\mathbb{R}^d)}^2$$

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

Proof. We rewrite the expectation as an integral,

$$\mathbb{E}\Big[\big|D_{k_q}g(x_i)\big|\cdot\big|D_{\ell_q}g(x_j)\big|\cdot\mathbf{1}\{\|x_i-x_j\|\leq(2q+1)r\}\Big]$$
$$\leq p_{\max}^4\int_{\mathcal{X}^4}|g(x)-g(y)|\cdot|g(u)-g(v)|\cdot K_r(x,y)K_r(u,v)\mathbf{1}\{\|y-v\|\leq(2q+1)r\}\,dy\,dx\,du\,dv$$

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,

$$\begin{aligned}
&\int_{\mathcal{X}^4}|g(x)-g(y)|\cdot|g(u)-g(v)|\cdot K_r(x,y)K_r(u,v)\mathbf{1}\{\|y-v\|\leq(2q+1)r\}\,dy\,dx\,du\,dv\\
&\quad\leq K_{\max}^2r^d\int_{\mathcal{X}}\int_{[B(0,1)]^3}\big|g\big((z_3+z_1)r+v\big)-g(z_1r+v)\big|\cdot\big|g(z_2r+v)-g(v)\big|\,dz_1\,dz_2\,dz_3\,dv\\
&\quad\leq K_{\max}^2r^{d+2}\int_{[B(0,1)]^3}\int_{[0,1]^2}\int_{\mathcal{X}}\|\nabla g(tz_3r+z_1r+v)\|\cdot\|\nabla g(tz_2r+v)\|\,dv\,dt_1\,dt_2\,dz_1\,dz_2\,dz_3\\
&\quad\leq c\nu_d^3K_{\max}^2r^{d+2}[f]_{W_d^{1,2}(\mathcal{X})}^2.
\end{aligned}$$

□

Lemma 7. *Suppose $f\in\mathcal{L}^2(\mathbb{R}^d)$. Then, the function $g(x)=\int_0^1\int_{B(0,1)}f(x+aty)\,dy\,dt$ also belongs to $\mathcal{L}^2(\mathbb{R}^d)$, with norm*

$$\|g\|_{\mathcal{L}^2(\mathbb{R}^d)}\leq\nu_d\cdot\|f\|_{\mathcal{L}^2(\mathbb{R}^d)}$$

Proof. We compute the squared norm of g,

$$\begin{aligned}
\|g\|_{\mathcal{L}^2(\mathbb{R}^d)}^2&=\int_{\mathbb{R}^d}\left(\int_0^1\int_{B(0,1)}f(x+aty)\,dt\,dy\right)^2dx\\
&\leq\nu_d^2\int_{\mathbb{R}^d}\int_0^1\frac{1}{\nu_d}\int_{B(0,1)}f^2(x+aty)\,dt\,dy\,dx && \text{(Jensen's inequality)}\\
&=\nu_d^2\int_0^1\int_{B(0,1)}\frac{1}{\nu_d}\int_{\mathbb{R}^d}f^2(x+aty)\,dt\,dy\,dx && \text{(Fubini's theorem)}\\
&=\nu_d^2\|f\|_{\mathcal{L}^2(\mathbb{R}^d)}^2.
\end{aligned}$$

□

3 Proof of (6)

Lemma 8. *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 $f\in W_d^{s,2}(\mathcal{X})$. Then for any $b\geq1$, there exists c_1 such that if

$$\|f\|_{\mathcal{L}^2(\mathcal{X})}\geq\begin{cases}c_1\cdot b\cdot\|f\|_{W_d^{s,2}(\mathcal{X})}\cdot\max\{n^{-1/2},n^{-s/d}\}, & \textit{if } 2s\neq d\\ c_1\cdot b\cdot\|f\|_{W_d^{s,2}(\mathcal{X})}\cdot n^{-a/2}, & \textit{if } 2s=d \textit{ for any } 0<a<1\end{cases}\tag{15}$$

then,

$$\mathbb{P}\left[\|f\|_n^2\geq\frac{1}{b}\mathbb{E}\big[\|f\|_n^2\big]\right]\geq1-\frac{5}{b}\tag{16}$$

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

Proof. To prove (16) we will show

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

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

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

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

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

Therefore,

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

Since by assumption

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

we have

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

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

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

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

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

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

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

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 (17) 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. $\square$

The proof of Lemma 8 relies on (a variant of) the Paley-Zygmund Inequality.

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

$$\mathbb{E}\big[\|f\|_n^4\big] \leq \left(1 + \frac{1}{b^2}\right) \cdot \Big(\mathbb{E}\big[\|f\|_n^2\big]\Big)^2. \tag{17}$$

Then,

$$\mathbb{P}\left[\|f\|_n^2 \geq \frac{1}{b}\mathbb{E}\big[\|f\|_n^2\big]\right] \geq 1 - \frac{5}{b}. \tag{18}$$

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

$$\mathbb{P}(Z > \lambda \mathbb{E}(Z^p)) \geq \left[(1 - \lambda^p)\frac{\mathbb{E}(Z^p)}{(\mathbb{E}(Z^q))^{p/q}}\right]^{\frac{q}{q-p}} \tag{19}$$

Applying (19) with $Z = \|f\|_n^2$, $p = 1$, $q = 2$ and $\lambda = \frac{1}{b}$, by assumption (17) we have

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

$\square$