Notes for Week 12/7/19 - 12/29/19

Alden Green

January 22, 2020

We observe random design and responses (x_i, y_i) according to the following regression model: $x_1, \ldots, x_n$ are drawn i.i.d from distribution P with density p supported on $\mathcal{X} \subset \mathbb{R}^d$, and

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

Suppose we form the undirected, weighted graph $G_{n,r} = (X, W)$ with edge weights $W = (W_{ij})$ formed according to the kernel function K as follows:

$$W_{ij} = K_r(x_i, x_j) := \frac{1}{r^d} K(\|x_i - x_j\|^2)$$

Let L_n be the graph Laplacian operator associated with $G_{n,r}$, defined by the action

$$L_n f(x) := \frac{1}{nr^2} \sum_{i=1}^{n} (f(x_i) - f(x)) K_r(x_i, x).$$

The graph Laplacian L_n induces a class of roughness functionals $R_s(f)$, defined by

$$R_{s,n}(f) = R_s(f; G_{n,r}) = \frac{1}{n} f^T L_n^s f$$

We refer to $R_{s,n}(f)$ as the order-s roughness functional. As part of our graph testing work, we have shown that if $f \neq 0$, then for any design points X and resulting neighborhood graph G with Laplacian L such that

$$\|f\|_n^2 \geq 2b\sqrt{\frac{2\kappa}{n^2}} + \frac{R_s(f; G)}{\lambda_\kappa^s} \tag{1}$$

the graph spectral test ϕ_{spec} makes a Type II error with probability at most $3/b$. Suppose

$$\|f\|_n^2 \geq \frac{1}{2}\|f\|_2^2, \ R_s(f; G_{n,r}) \leq ||f||_{W_d^s(\mathcal{X})}^2 \ \text{ and } \ \lambda_\kappa \leq \kappa^{2/d}; \tag{2}$$

then it can be verified that choosing $\kappa = n^{-2d/(4s+d)}$, the equation (1) is satisfied whenever $\|f\|_2^2 \geq n^{-4s/(4s+d)}$. It is therefore sufficient to show that (2) holds with high probability over the random design points X.

In this week's notes, we will focus our attention on upper bounding the roughness functional $R_{n,s}(f)$. Our main results should take the following form: when (a) $f \in C^s(L)$ for a fixed constant L (or more ideally $f \in \mathcal{W}^s(L)$), (b) the neighborhood graph radius $r = r(n)$ is properly chosen as a function of n, and (c) K is an appropriately chosen kernel, the roughness functional satisfies

$$\mathbb{E}\left[R_{s,n}(f)\right] = O\left(1\right) \tag{3}$$

uniformly over all $f \in C^s(1)$. These notes detail our current progress in showing such a result for different values of s. We showed the following useful representation of $R_{s,n}(f)$ in the *8.7.19* Notes. For $k = (k_1, \ldots, k_n) \in [n]^s$, recursively defined the graph difference operator D_k by

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

Then when s is even, letting $q = s/2$ we have

$$R_{s,n}(f) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{(nr^2)^q}\sum_{k\in[n]^q} D_k f(x_i)\right)^2 \tag{4}$$

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

$$R_{s,n}(f) = \frac{1}{2n^2r^2}\sum_{i,j=1}^{n}\left(\frac{1}{(nr^2)^q}\sum_{k\in[n]^q}\big(D_k f(x_i) - D_k f(x_j)\big)\right)^2 K_r(x_i, x_j) \tag{5}$$

1 Bounding the Roughness Functional when f is Holder

For an integer $\ell > 0$, we will say K is a 2nd-order kernel if it is compactly supported on $B(0,1)$, uniformly bounded $|K(x)| \leq K_{\max}$ for all $x \in B(0,1)$, and additionally

$$\int K(z) = 1 \quad \text{and} \quad \int zK(z) = 0$$

That K be a 2nd-order kernel is crucial for L_n to act as a 2nd-order differential operator, and for $R_{s,n}$ to scale at the proper rate.

Lemma 1. *Suppose $f \in C^s(L)$ for some positive integer s and $L > 0$, and further suppose $p \in C^\ell(L)$ for $\ell = \lfloor s \rfloor$ the largest integer strictly less than s. For any 2nd-order kernel K and any $1 \geq r(n) \geq n^{-1/(2(s-1)+d)}$, we have that*

$$\mathbb{E}(R_{s,n}(f)) = O(1),$$

uniformly over all $f \in C^s(L)$.

Note that Lemma 1 implies a high-probability bound on $R_{s,n}(f)$ by Markov's inequality. To prove Lemma 1, we first break the sum (4) (or (5)) into cases according to the number of unique indices, then bound the expectation case by case. The following Lemma will be the workhorse which supplies a sufficient bound in each case.

Lemma 2. *Let $q = s/2$ when s is even, and $q = (s-1)/2$ when s is odd. Under the same conditions as Lemma 1, for any indices $k = (k_1, \ldots, k_q)$ and $\ell = (\ell_1, \ldots, \ell_q)$, we have*

$$\mathbb{E}(D_k f(x_i) D_\ell f(x_i)) = \begin{cases} O(r^{2s}), & \text{if all indices are distinct} \\ O(r^2 r^{d(|k\cup\ell\cup i| - (2q+1))}), & \text{otherwise} \end{cases} \tag{6}$$

Additionally, we have

$$\mathbb{E}(d_i D_k f(x_j) d_i D_\ell f(x_j)) = \begin{cases} O(r^{2s}), & \text{if all indices are distinct} \\ O(r^2 r^{d(|k\cup\ell\cup i\cup j| - (2q+2))}), & \text{otherwise} \end{cases} \tag{7}$$

The minimum scaling condition $r(n) \geq n^{-1/(2(s-1)+d)}$ is needed in order to ensure the diagonal terms of the sums in (4) and (5) (where not all indices are distinct) do not dominate the off-diagonal terms (where all indices are distinct).

2 Bounding the Roughness Functional when f is Sobolev

Suppose f is assumed to belong to the Sobolev space $W_d^{s,2}(\mathcal{X})$, and have small semi-norm in this space, rather than belonging to the Holder space $C_d^s(\mathcal{X})$ and having small Holder norm. Can we still prove–possibly up to constants–the same bound on the expected graph semi-norm of f? Before answering this question, we review the definition of Sobolev spaces and their associated norms.

2.1 Sobolev spaces

For a given $s > 0$, the Sobolev space $W_d^{s,2}(\mathcal{X})$ consists of all functions $f \in \mathcal{L}^2(\mathcal{X})$ such that for each $\alpha = (\alpha_1, \ldots, \alpha_d)$ with $|\alpha| := \sum_{i=1}^d \alpha_i \leq s$, the weak derivative $D^\alpha f$ belongs to $\mathcal{L}^2(\mathcal{X})$. The Sobolev $\{s, 2\}$ semi-norm is then

$$[f]_{W_d^{s,2}(\mathcal{X})} = \left(\sum_{|\alpha|=s} \int_{\mathcal{X}} |D^\alpha f|^2 \, dx\right)^{1/2}$$

and the Sobolev $\{s, 2\}$ norm is

$$\|f\|_{W_d^{s,2}(\mathcal{X})}^2 = \left(\sum_{q=1}^{s} [f]_{W_d^{s,2}(\mathcal{X})}^2\right)^{1/2}.$$

For a $R > 0$, the corresponding ball is $W_d^{s,2}(\mathcal{X}; R) = \big\{f : \|f\|_{W^{s,2}(\mathcal{X})} \leq R\big\}$.

2.2 Bounds on the graph semi-norm

Under largely the same conditions as Lemma 1, we have that 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 3. *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*

$$\mathbb{E}\big[R_{s,n}(f)\big] \leq c \cdot \|f\|_{W_d^{s,2}(\mathcal{X})} \tag{8}$$

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

Note that the bound (8) 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

$$L_r f(x) = \frac{1}{r^2} \int (f(z) - f(x)) K_r(z, x) \, dP(x)$$

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 $\langle L_n^s f, f\rangle$. However, for any s the seminorm $\langle L_r^s f, f\rangle$ assigns zero length only to constant functions f. As a one-dimensional example, if $s = 2$, and $f(x) = c \cdot x$ for c large, the seminorm $\langle L_r^s f, f\rangle$ will also be quite large, despite the fact that the second derivative $f'' = 0$.

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. The proof of Lemma 3 relies on Lemma 4.

Lemma 4. *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*

$$\mathbb{E}(D_k f(x_i) D_\ell f(x_i)) = \begin{cases} O(r^{2s}) \cdot \|f\|^2_{W_d^{s,2}(\mathcal{X})}, & \text{if all indices are distinct} \\ O(r^2 r^{d(|k \cup \ell \cup i| - (2q+1))}) \cdot [f]^2_{W_d^{1,2}(\mathcal{X})}, & \text{otherwise} \end{cases} \tag{9}$$

Additionally, we have

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

The lower bound on $r(n)$ in Lemma 3 is needed to ensure the leading term in (9) (or (10)), where all indices are distinct, dominates the lower-order terms, where indices are repeated.

3 Proofs

3.1 Proof of Lemma 4

We first prove the desired bound in the case when some indices are repeated, and then the desired bound in the case when all indices are distinct.

3.1.1 Repeated indices.

Since the proofs of (9) and (10) 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 $|D_k f(x_i)|$ is upper bounded by

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

Therefore,

$$\begin{aligned} |D_k f(x_i)|\,|D_\ell f(x_i)| &\leq \left(\frac{2K_{\max}}{r^d}\right)^{2(q-1)} \sum_{h,j \in k \cup \ell \cup i} \left|D_{k_q} f(x_h)\right| \cdot \left|D_{\ell_q} f(x_j)\right| \cdot \mathbf{1}\{G_{n,r}[X_{k \cup i}], G_{n,r}[X_{\ell \cup i}] \text{ are connected graphs.}\} \\ &= \left(\frac{2K_{\max}}{r^d}\right)^{2(q-1)} \sum_{h,j \in k \cup \ell \cup i} \left|D_{k_q} f(x_h)\right| \cdot \left|D_{\ell_q} f(x_j)\right| \cdot \mathbf{1}\{G_{n,r}[X_{k \cup \ell \cup i}] \text{ is a connected graph.}\} \end{aligned} \tag{11}$$

We now break our analysis into three cases, based on the number of distinct indices in k_q, ℓ_q, h, j.

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

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

where the last equality follows from Lemma 10.

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

$$\begin{aligned}
&\mathbb{E}\left[|D_i f(x_j)| \cdot |D_i f(x_h)| \cdot \mathbf{1}\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is connected}\}\right] \qquad (12) \\
&\quad = \mathbb{E}\left[|D_i f(x_j)| \cdot |D_i f(x_h)| \cdot \mathbb{P}\left[\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is connected}\}|x_i, x_j, x_h\right]\right] \\
&\quad = O(r^{(|k\cup\ell\cup i|-3)d}) \cdot \mathbb{E}\left[|D_i f(x_j)| \cdot |D_i f(x_h)|\right]. \qquad (13)
\end{aligned}$$

It remains to upper bound $\mathbb{E}\left[|D_i f(x_j)| \cdot |D_i f(x_h)|\right]$, which we rewrite as

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

By the argument set out in Section 6.6, we may assume without loss of generality that there exists an extension $g \in C_d^\infty(\mathcal{X}) \cap W_d^{1,2}(\mathbb{R}^d)$ such that $[g]_{W_d^{1,2}(\mathbb{R}^d)} \leq c[f]_{W_d^{1,2}(\mathcal{X})}$. Then, by a similar series of steps as in Section 6.6, we obtain

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

and as a result,

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

Plugging back into (13), we have that $\mathbb{E}\left[|D_i f(x_j)| \cdot |D_i f(x_h)| \cdot \mathbf{1}\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is connected}\}\right] = O(r^{d(|k\cup\ell\cup i|-3)+2}) \cdot [f]^2_{W_d^{1,2}(\mathcal{X})}$.

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

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

Replacing f by a smooth extension $g \in C^\infty(\mathcal{X}) \cap W_d^{1,2}(\mathbb{R}^d)$, we rewrite the expectation as an integral,

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

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,

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

Therefore $\mathbb{E}\big[\big|D_{k_q} f(x_i)\big|\,\big|D_{\ell_q} f(x_j)\big|\mathbf{1}\{G_{n,r}[X_{k\cup\ell\cup i}] \text{ is connected}\}\big] = O(r^{d(|k\cup\ell\cup i|-3)+2})\cdot [f]_{W_d^{1,2}(\mathcal{X})}^2$.

Since we obtain the same scaling in each case, plugging this back in to (11) we have that for any $k, \ell \in [n]^q$

$$\mathbb{E}\big[|D_k f(x_i)|\,|D_\ell f(x_i)|\big] = O(r^{d(|k\cup\ell\cup i|-(2q+1))+2})\cdot [f]_{W_d^{1,2}(\mathcal{X})}^2.$$

3.1.2 All indices distinct.

We first show the desired result when s is even, and then when s is odd. In either case, since $\mathcal{X}$ is a Lipschitz domain, there exists an extension $g \in W^{s,2}(\mathbb{R}^d)$ such that $\|g\|_{W^{s,2}(\mathbb{R}^d)} \le c\|f\|_{W^{s,2}(\mathcal{X})}$, and $g = f$ a.e in $\mathcal{X}$. Further, there exists a sequence of smooth approximations $(g_m) \subset C^\infty(\mathbb{R}^d)$, such that

$$\|g_m - g\|_{W^{s,2}(\mathbb{R}^d)} \to 0.$$

s even. By the law of iterated expectation, we have

$$\begin{aligned}
\mathbb{E}\big[D_k f(x_i) D_k f(x_j)\big] = \mathbb{E}\big[D_k g(x_i) D_k g(x_j)\big] &= \mathbb{E}\Big[\big(\mathbb{E}[D_k g(x_i)|x_i]\big)^2\Big] \\
&\le 2\mathbb{E}\Big[\big(\mathbb{E}[D_k g_m(x_i)|x_i]\big)^2\Big] + \mathbb{E}\Big[\big(\mathbb{E}[D_k (g-g_m)(x_i)|x_i]\big)^2\Big] \\
&\le 2\mathbb{E}\Big[\big(\mathbb{E}[D_k g_m(x_i)|x_i]\big)^2\Big] + \left(\frac{2K_{\max}}{r^d}\right)^q \|g - g_m\|_{\mathcal{L}^2(\mathbb{R}^d)}^2.
\end{aligned}$$

By Lemma 8, we have that $\mathbb{E}[D_k g_m(x_i)|x_i]\big)^2 = O(r^{2s})\cdot [f_s(x)]^2$ for some $f_s \in \mathcal{L}^2(\mathbb{R}^d)$ which satisfies $\|f_s\|_{\mathcal{L}^2(\mathbb{R}^d)} \le c\|g_m\|_{W^{s,2}(\mathbb{R}^d)}$ for some constant c which does not depend on g_m. Along with the previous display, this implies

$$\mathbb{E}\big[D_k f(x_i) D_k f(x_j)\big] \le O(r^{2s})\|g_m\|_{W^{s,2}(\mathbb{R}^d)}^2 + \left(\frac{2K_{\max}}{r^d}\right)^q \|g - g_m\|_{\mathcal{L}^2(\mathbb{R}^d)}^2,$$

and taking the limit of the right hand side as $m \to \infty$ we have $\mathbb{E}\big[D_k f(x_i) D_k f(x_j)\big] \le O(r^{2s})\cdot\|f\|_{W^{s,2}(\mathbb{R}^d)}^2$.

s **odd.** By the law of iterated expectation, we have

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

where the latter equality follows from Lemma 8. 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

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

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

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

where the last inequality follows from Lemma 10. By similar arguments to the case where s is even, we can show that this estimate holds (up to constants) when g_m is replaced by f, establishing the claim.

3.2 Proof of Lemma 1

We will take s to be even, as the proof when s is odd follows essentially the same steps. Now when s is even, we have the decomposition

$$R_{s,n}(f) = \frac{1}{n^{s+1} r^{2s}} \sum_{i=1}^{n} \sum_{k\in[n]^q} \sum_{\ell\in[n]^q} D_k f(x_i) D_\ell f(x_i)$$

For given index vectors k, ℓ and index i, 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.

When $I < s+1$, there is at least one repeated index, and by Lemma 2 we have that

$$\mathbb{E}(D_k f(x_i) D_\ell f(x_i)) = O(r^2 r^{d(I-(s+1))}).$$

There are $O(n^I)$ terms in $R_{s,n}(f)$ with exactly I distinct indices. Therefore, the total contribution of such terms to the sum is $O(r^{-2(s-1)} r^{d(I-(s+1))} n^{I-(s+1)} r)$. Since $r(n) \ge n^{-1/d}$, this increases with I. Taking $I = s$ to be as large as possible such that there is at least one repeated index, the contribution of these terms to the sum is $O(r^{-(2(s-1)+d)} n^{-1}) = O(1)$, with the equality following by the restriction $r \ge n^{-1/(2(s-1)+d)}$.

When $I = s+1$, every index in ℓ, k and i is unique. Therefore by Lemma 1,

$$\mathbb{E}(D_k f(x_i) D_\ell f(x_i)) = O(r^{2s}).$$

Since there are $O(n^{s+1})$ such terms, we have that the total contribution of these terms is $O(n^{s+1} r^{2s} n^{-(s+1)} r^{-2s}) = O(1)$.

4 Proof of Lemma 2

4.1 Proof of (6)

Let $k, \ell \in [n]^q$ be index vectors. We first prove the desired bound in the case when some indices are repeated, and then the desired bound in the case when all indices are distinct.

Repeated indices. Since $f \in C^1(L)$, for all index vectors $k, \ell \in [n]^q$ and indices $i \in [n]$, the product of iterated difference operators $D_k f(x_i) D_\ell f(x_i)$ satisfies

$$|D_k f(x_i) D_\ell f(x_i)| \leq 4^q L^2 r^{2-ds}$$

Moreover $D_k f(x_i) D_\ell f(x_i)$ will equal zero if there exists $x_j, j \in k \cup \ell \cup i$ such that

$$\|x_j - x_h\| > r, \quad \text{for all } h \in k \cup \ell \cup i$$

Therefore $D_k f(x_i) D_\ell f(x_i)$ is nonzero with probability $O(r^{d(|k \cup \ell \cup i| - 1)})$, which along with the boundedness of $D_k f(x_i) D_\ell f(x_i)$ implies the claimed result.

All indices distinct. By the law of iterated expectation, we have

$$\begin{aligned} \mathbb{E}\left[D_k f(x_i) D_\ell f(x_i)\right] &= \mathbb{E}\left[\mathbb{E}\left(D_k f(x) | x_i = x\right) \mathbb{E}\left(D_\ell f(x) | x_i = x\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(D_k f(x) | x_i = x\right)^2\right] \end{aligned} \tag{14}$$

The result then follows from Lemma 7.

4.2 Proof of (7)

The proof of (7) when some index is repeated is essentially the same, and we do not reproduce it.

When all indices ℓ, k, i and j are distinct, by Lemma 7 there exists some $f_{s-1} \in C^1(L)$ such that

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

where the second equality follows from the linearity of the difference operator. Then since $f_{s-1} \in C^1(L)$, we have $\mathbb{E}(D_i f_{s-1}(x_j)) = O(r)$, and

$$\mathbb{E}\Big(\big(d_i f_{s-1}(x_j))\big)^2 K_r(x_i, x_j)\Big) = \mathbb{E}\Big(\big(O(r)\big)^2 K_r(x_i, x_j)\Big) = O(r^2),$$

establishing $\mathbb{E}(d_i D_k f(x_j) d_i D_\ell f(x_j) K_r(x_i, x_j)) = O(r^{2s})$ as claimed.

5 Auxiliary Results.

Lemma 5. *When $f \in C^s(L)$, for any $k \in [n]$*

$$D_k f(x) = \left(\sum_{r=1}^{s-1} (x_k - x)^r f^{(r)}(x) + O((x_k - x)^s)\right) \cdot K_r(x_k, x)$$

Lemma 6. *When K is a bounded and compactly supported kernel, then for any s and for all $x \in \mathbb{R}^d$*

$$\mathbb{E}(|K(x_i, x)|^s) \leq C L r^{(1-s)d}$$

Lemma 7. *Suppose $f \in C^s(L)$, $p \in C^{s-1}(L)$, $k \in [n]^q$ for some $q \geq 1$, and that K_r is a 2nd-order kernel. Then if $2q \leq s - 1$,*

$$\mathbb{E}(D_k f(x)) = \sum_{\ell = 2q}^{s-1} O(r^\ell) \cdot f_\ell(x) + O(r^s), \tag{15}$$

for some $f_\ell \in C^{s-\ell}(L)$. If $2q \geq s$, $\mathbb{E}(D_k f(x)) = O(r^s)$. All $O(\cdot)$ terms may depend L and s, but do not depend on f or x.

Proof. We will prove Lemma 7 in the case where $d = 1$. When $d \geq 2$, using multivariate Taylor expansions we find the same result, but with more notational overhead.

We will prove by induction on q in the case where $d = 1$. When $q = 1$ and $k \in [n]$, by taking Taylor expansions of f and p, we obtain

$$\begin{aligned}\mathbb{E}(D_k f(x)) &= \sum_{\ell=1}^{s-1} f^{(\ell)}(x) \int (z-x)^\ell K_r(z,x)p(z)dz + O(r^s) \\ &= \sum_{\ell=1}^{s-1}\sum_{a=0}^{s-1} f^{(\ell)}(x)p^{(a)}(x) \underbrace{\int (z-x)^{\ell+a} K_r(z,x)dz}_{:=I_{t+a}} + O(r^s) \end{aligned} \tag{16}$$

Since K is a 2nd-order kernel, $I_1 = 0$ and $I_t = O(r^t)$ for $t \geq 2$. Additionally, when $\ell + a \leq s-1$, we have that $f^{(\ell)}p^{(a)} \in C^{\min\{s-\ell, s-1-a\}} \subseteq C^{s-(\ell+a)}$, and $\left|f^{(\ell)}p^{(a)}\right| \leq L^2$ for any ℓ and a. We can therefore simplify (16) by combining all terms where $\ell + a = t$, obtaining

$$\mathbb{E}(D_k f(x)) = \sum_{t=1}^{s-1} f_t(x) I_t + O(r^s) = \sum_{t=2}^{s-1} f_t(x) I_t + O(r^s), \tag{17}$$

which establishes (15) in the base case.

Now, we assume (15) 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 (17),

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

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 $\mathbb{E}(D_k f_t(x))$. If $2(q+1) \geq s$, note that since $f_t \in C^{s-t}(L)$ for $t \geq 2$, we have by hypothesis $\mathbb{E}(D_k f_t(x)) = O(r^{s-t})$. As a result

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

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

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

Noting that $g_\ell \in C^{s-(t+\ell)}(L)$ for some $\ell + t = 2(q+1), \ldots, s-1$, and $I_t \cdot O(r^\ell) = O(r^{t+\ell})$, we can rewrite the final equation as a sum over $\ell + t = 2(q+1), \ldots, s-1$, which proves (15). $\square$

The following Lemma establishes a result analogous to Lemma 7 when the function f and density p are assumed to be bounded in Sobolev, rather than Holder, norm. 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 8. *Let $k \in [n]^q$ for some $q \geq 1$. Suppose that $f \in W_d^{s,2}(\mathbb{R}^d) \cap C_d^\infty(\mathbb{R}^d)$ and $p \in C_d^{s-1}(\mathbb{R}^d; p_{\max})$, 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

$$\mathbb{E}(D_k f(x)) = \begin{cases} \sum_{\ell=2q}^{s-1} I_\ell \cdot f_\ell(x) + O(r^s) \cdot f_s(x), & \text{if } 2q < s \\ O(r^s) \cdot f_s(x), & \text{if } 2q \geq s \end{cases} \tag{18}$$

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

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

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

$$\begin{aligned}
\mathbb{E}(D_k f(x)) &= \sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,P} + \frac{|\alpha|}{\alpha!} \sum_{|\alpha|=s} \int_0^1 (1 - t)^{s-1} \mathbb{E}\big[f^{(\alpha)}(x + t(x_k - x))(x_k - x)^\alpha K_r(x_k, x)\big]\, dt \\
&= \sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,P} + O(r^s) \sum_{|\alpha|=s} \int_0^1 \mathbb{E}\big[f^{(\alpha)}(x + t(x_k - x)) K_r(x_k, x)\big]\, dt \\
&= \sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,P} + O(r^s) \sum_{|\alpha|=s} \int_0^1 \int_{B(x,r)} \frac{f^{(\alpha)}(x + t(z - x))}{r^d}\, dz\, dt \\
&= \sum_{|\alpha|=1}^{s-1} \frac{f^{(\alpha)}(x)}{\alpha!} E_{\alpha,P} + O(r^s) \int_0^1 \int_{B(0,1)} \sum_{|\alpha|=s} f^{(\alpha)}(x + try)\, dy\, dt
\end{aligned} \tag{19}$$

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 (19) 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 9. We are now in a position to prove the second part of (18) when $q=1$. Since $q=1$, $s\in\{1,2\}$. 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 (18) 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

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

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 (18) 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 (18) 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 (18).

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 (18) 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 (15). $\square$

Lemma 9. *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,

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

$\square$

Lemma 10. *Suppose $g \in C^\infty(\mathbb{R}^d)$ and $|p(x)| \le p_{\max}$ for all $x \in \mathbb{R}^d$. Then*

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

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

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

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

Plugging this into (22), we obtain

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

where (i) follows by Cauchy-Schwarz, (ii) follows since either $\|y - x\| \le r$ or $K_r(y,x) = 0$, (iii) follows by Jensen's, (iv) follows by the assumption $K \le K_{\max}$ 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

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

□

6 Old Work

6.1 Bounding the 2nd-Order Roughness Functional

The (first and) second order roughness functionals are somewhat special. These roughness functionals do not involve taking compositions of the difference operator D_k, and as a result, we get the desired bounds in expectation even when $r(n)$ is small.

Lemma 11. *Fix $L > 0$, and form the neighborhood graph $G_{n,r}$ using kernel $K(z) = \mathbf{1}\{z^2 \leq r^2\}$. Suppose $f \in C^2(L)$ and $p \in C^1(L)$. Then for any $r(n) \geq n^{-1/(2+d)}$ and , we have*

$$\mathbb{E}(R_{2,n}(f)) = O(L^2)$$

6.2 Bounding the 3rd-Order Roughness Functional

For the order-s roughness functionals when $s \geq 3$, the graph Laplacian is forced to approximate higher order derivatives by composing difference operators– that is, taking differences of differences, etc. In for the roughness functional to continue scaling at the desired rate, we will need $r(n)$ to be larger, and we cannot use the uniform kernel K any longer. In the particular case $s = 3$, we will require a kernel K which is bounded and compactly supported on $B(0, 1)$, and additionally satisfies

$$\int z^j K(z) = 0 \text{ for } j = 1, 2.$$

We then have the following result.

Lemma 12. *Suppose $f \in C^3(L)$ for some $L > 0$, and further suppose $p \in C^2(L)$. For any kernel K which satisfies the previous conditions, and any $r(n) \geq n^{-1/(4+d)}$, we have that*

$$\mathbb{E}(R_{s,n}(f)) = O(1)$$

6.3 Proof of Lemma 11

We prove Lemma 11 in the case when $d = 1$.

TODO: Extend the proof to all values of d. It's not hard, just requires you to replace derivatives by partial derivatives.

Separating the squared sum in (4) into diagonal and off-diagonal terms, we obtain

$$\begin{aligned} R_{2,n}(f) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{nr^2}\sum_{j=1}^{n} D_j f(x_i)\right)^2 \\ &= \frac{1}{n^3 r^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\underbrace{(D_j f(x_i))^2}_{V_1} + \frac{1}{n^3 r^4}\sum_{i=1}^{n}\sum_{j=1}\sum_{k\neq j}\underbrace{(D_j f(x_i))(D_k f(x_i))}_{U_1} \end{aligned}$$

The expectation of U_1 will be of the right order of magnitude thanks to a cancellation of the first-order term in a Taylor expansion. Noting that the kernel K satisfies

$$\int zK(z)\,dz = 0$$

we therefore derive that for any $x \in \mathcal{X}$,

$$\begin{aligned}
\mathbb{E}(D_j f(x)) &= \frac{1}{r}\int ((y-x)f'(x) + O(L(y-x)^2)K(y-x)p(y)\,dy && \text{(by } f \in C^2(L))\\
&= \frac{1}{r}\int (y-x)K(y-x)f'(x)(p(x) + O(L(y-x)))\,dy + O(Lr^2) && \text{(by } p \in C^1(L))\\
&= \underbrace{\frac{f'(x)p(x)}{r}\int zK(z)\,dz}_{=0} + O(Lr^2) = O(Lr^2).
\end{aligned}$$

As a result, we have

$$\frac{1}{n^3r^4}\sum_{i=1}^{n}\sum_{j=1}\sum_{k\neq j}\mathbb{E}\left[(D_j f(x_i))(D_k f(x_i))\right] = \frac{1}{n^3r^4}\sum_{i=1}^{n}\sum_{j=1}\sum_{k\neq j}\mathbb{E}\left[\mathbb{E}(D_k f(x) \mid x_i = x)^2\right] = O(L^2).$$

The expectation of the diagonal terms may be much larger, but on the other hand there are many fewer of them. Specifically, we compute

$$\begin{aligned}
\mathbb{E}\big((D_k f(x_i))^2\big) &= \mathbb{E}\Big(O\left(L^2(x_k - x_i)^2\right)\cdot K_r(x_k, x)^2\Big)\\
&= O(L^2r^2)\cdot\mathbb{E}(K_r(x_k, x_i)^2) = O(L^2 r)
\end{aligned}$$

and therefore

$$\frac{1}{n^3r^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\big((D_j f(x_i))^2\big) = \frac{1}{n}O(L^2r^{-3}) = O(L^2)$$

where the second equality follows since by assumption when $d = 1$, $n^{-1} = O(r^3)$.

6.4 Proof of Lemma 12

We prove Lemma 12 when $d = 1$ and $r(n) \geq n^{-1/5}$.

TODO: Prove for all dimensions d, by replacing derivatives with partial derivatives.

From (5), we have

$$\begin{aligned}
R_{3,n}(f) &= \frac{1}{2n^2r^2}\sum_{i,j=1}^{n}\left(\frac{1}{(nr^2)}\sum_{k\in n}\big(D_k f(x_i) - D_k f(x_j)\big)\right)^2 K_r(x_i, x_j)\\
&= \frac{1}{2n^4r^6}\sum_{i,j=1}^{n}\sum_{k,\ell=1}^{n} D_{i\ell}f(x_j)\cdot d_i(D_k f(x_j)).
\end{aligned}$$

We separate our analysis into cases based on the distinct indices in the previous summation.

Case 1: $i = k = \ell$. In this case, we can rewrite the summand as

$$\begin{aligned}
D_i i f(x_j) d_i D_i f(x_j) &= (d_i f(x_j))^2 K_r(x_i, x_j)^3\\
&\leq 4L^2r^2\left|K_r(x_i, x_j)^3\right|
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}(D_i i f(x_j) d_i D_i f(x_j)) &\leq 4L^2r^2\mathbb{E}\big(\left|K_r(x_i, x_j)^3\right|\big)\\
&\leq 4CL^3, && \text{(Lemma 6)}
\end{aligned}$$

and as a result

$$\frac{1}{2n^4r^6}\sum_{i,j=1}^{n}\mathbb{E}(D_{ii}f(x_j)\cdot d_i(D_i f(x_j))) \leq \frac{2CL^3}{n^2r^6} \leq \frac{2CL^3}{n^{4/5}}.$$

Case 2: $i = \ell \neq k$**.** In this case, we can rewrite the summand as $-(D_i f(x_j))(d_i D_k f(x_j))K_r(x_i, x_j)$. We have that

$$|D_i f(x_j)| \leq 2LrK(x_i, x_j), \quad \text{and } |d_i D_k f(x_j)| \leq 2Lr\big(|K_r(x_k, x_i)| + |K_r(x_\ell, x_i)|\big)$$

Therefore, we can upper bound

$$|D_i f(x_j) d_i D_k f(x_j) K_r(x_i, x_j)| \leq 4Lr^2 K(x_i, x_j)^2 \big(|K_r(x_k, x_i)| + |K_r(x_\ell, x_i)|\big)$$

and the resultant expectation satisfies

$$\mathbb{E}\big(|D_i f(x_j) d_i D_k f(x_j) K_r(x_i, x_j)|\big) \leq 8CL^4 r.$$

Since there are order n^3 terms in the sum for which $i = k \neq \ell$, we have

$$\frac{1}{2n^4 r^4} \sum_{\ell \neq k = i \neq j \neq}^{n} \mathbb{E}(D_{ii} f(x_j) \cdot d_i(D_\ell f(x_j))) \leq \frac{8CL^3}{r^5 n} \leq 8CL^3.$$

Note that by symmetry, the same analysis applies to the case $i = \ell \neq k$.

Case 3: $i \neq k = \ell$**.** In this case, we can rewrite the summand as $(d_i D_k f(x_j))^2 K_r(x_i, x_j)$. We bound the order of magnitude of the composed difference operators using Taylor expansion:

$$\begin{aligned}
|d_i D_k f(x_j)| &= |(f(x_k) - f(x_i))K_r(x_k, x_i) - (f(x_k) - f(x_j))K_r(x_k, x_j)| \\
&\leq |(f(x_k) - f(x_i))|\,|K_r(x_k, x_i)| + |(f(x_k) - f(x_j))|\,|K_r(x_k, x_ij)| \\
&\leq 2L\left(|x_k - x_j|\,|K_r(x_k, x_i)| + |x_k - x_i|\,|K_r(x_k, x_i)|\right) \\
&\leq 2Lr\big(|K_r(x_k, x_i)| + |K_r(x_\ell, x_i)|\big)
\end{aligned}$$

Taking expectation with respect to x_k, we have

$$\begin{aligned}
\mathbb{E}\big(\big(d_i D_k f(x_j)\big)^2 | x_i, x_j\big) &\leq 8L^2 r^2 \left(\int K_r(x, x_i)^2 p(x)\, dx + \int K_r(x, x_j)^2 p(x)\, dx\right) \\
&\leq 16L^3 r^2 \int \left(\frac{1}{r} K(z/r)\right)^2 dz && (p \in C^2(L)) \\
&\leq 16L^3 r \int \big(K(t)\big)^2 dt && \text{(change of variables)} \\
&\leq 16CL^3 r. && \text{(K compactly supported and bounded)}
\end{aligned}$$

As a result,

$$\begin{aligned}
\mathbb{E}((d_i D_k f(x_j))^2 K_r(x_i, x_j)) &= \mathbb{E}(\mathbb{E}((d_i D_k f(x_j))^2 | x_i, x_j) K_r(x_i, x_j)) \\
&\leq 16CL^3 r \mathbb{E}(|K_r(x_i, x_j)|) \\
&\leq 16CL^3 r.
\end{aligned}$$

Since there are not quite n^3 terms in the sum with $i \neq k = \ell$, and since $r \geq n^{-1/5}$,

$$\frac{1}{2n^4 r^6} \sum_{i \neq j \neq k = \ell \neq i}^{n} \mathbb{E}(D_{ik} f(x_j) \cdot d_i(D_k f(x_j))) \leq \frac{8CL^3}{r^5 n} \leq 8CL^3.$$

Case 4: i, k, ℓ all distinct. We rewrite the summand as $d_i(D_\ell f(x_j))d_i(D_k f(x_j))K_r(x_i, x_j)$. Noting that

$$d_i(D_k f(x_j)) = [f(x_k) - f(x_i)]K_r(x_k, x_i) - [f(x_k) - f(x_j)]K_r(x_k, x_j)$$

by Lemma 7 we have that

$$|\mathbb{E}(d_i(D_k f(x_j))|x_i, x_j)| \leq 2Lr^3.$$

Therefore,

$$\begin{aligned}
|\mathbb{E}(d_i(D_\ell f(x_j))d_i(D_k f(x_j))K_r(x_i, x_j))| &\leq \mathbb{E}\Big(|\mathbb{E}(d_k(D_\ell f(x_j))|x_i, x_j) \cdot \mathbb{E}(d_k(D_\ell f(x_j))|x_i, x_j) \cdot K_r(x_k, x_j)|\Big) \\
&\leq 4L^2 r^6 \mathbb{E}(|K_r(x_k, x_j)|) \leq 4CL^2 r^6.
\end{aligned}$$

Since there are order n^4 terms in the summation for which i, j, k and ℓ are all distinct, we have that

$$\frac{1}{2n^4 r^6} \sum_{i \neq j \neq k \neq \ell}^{n} \mathbb{E}(D_{ik} f(x_j) d_i D_\ell f(x_j) K_r(x_i, x_j)) \leq 4C.$$

The above four cases cover all terms in the summation, and so we have proved Lemma 12.

6.5 First-order graph Sobolev seminorm

We begin with a bound on the first-order graph semi-norm, under the assumption f has small semi-norm in the first order Sobolev space $W_d^{1,2}(\mathcal{X})$.

Lemma 13. *Let $b \geq 1$ be a fixed number. Suppose K is a uniformly bounded kernel function, compactly supported on $B(0,1)$, $f \in W_d^{1,2}(\mathcal{X})$, and additionally that $p(x) \leq p_{\max}$ for all $x \in \mathcal{X}$. Then there exists a constant c which may depend only on p such that*

$$R_{s,n}(f) \leq c \cdot b \cdot [f]^2_{W_d^{s,2}(\mathcal{X})} \tag{20}$$

with probability at least $1 - \frac{1}{b}$.

6.6 Proof of Lemma 13

We will bound $\mathbb{E}(R_{s,n}(f)) \leq c \cdot [f]^2_{W_d^{s,2}(\mathcal{X})}$, whence (20) holds with probability at least $1 - \frac{1}{b}$ by Markov's inequality. We rewrite

$$\mathbb{E}(R_{s,n}(f)) = \frac{1}{r^2} \int_{\mathcal{X}} \int_{\mathcal{X}} (f(y) - f(x))^2 K_r(y, x) p(y) p(x) dy dx \leq \frac{p_{\max}^2}{r^2} \int_{\mathcal{X}} \int_{\mathcal{X}} (f(y) - f(x))^2 K_r(y, x)\, dy\, dx \tag{21}$$

Since $\mathcal{X} \subset \mathbb{R}^d$ is a Lipschitz domain, there exists an extension operator $E : W_d^{1,2}(\mathcal{X}) \to W_d^{1,2}(\mathbb{R}^d)$ such that $Ef = f$ a.e. on $\mathcal{X}$, and $[Ef]_{W_d^{s,2}(\mathbb{R}^d)} \leq c[f]_{W_d^{s,2}(\mathcal{X})}$ for some c which depends only on $\mathcal{X}$ and d. For notational simplicity, we denote this extension $Ef =: g$, and observe that

$$\int_{\mathcal{X}} \int_{\mathcal{X}} (f(y) - f(x))^2 K_r(y, x)\, dy\, dx \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (g(y) - g(x))^2 K_r(y, x)\, dy\, dx. \tag{22}$$

Assume for the moment that $g \in C_d^\infty(\mathcal{X}_{2r})$, where $\mathcal{X}_{2r} = \{x \in \mathbb{R}^d : \text{dist}(x, \mathcal{X}) \leq 2r\}$. If g is not smooth, we can approximate it by smooth functions (g_m) in such a way that the double integral will not be substantially changed, a fact that we will show at the end of the proof.

If $g \in C_d^\infty(\mathcal{X})$, by the fundamental theorem of calculus we have for any $y, x \in \mathbb{R}^d$,

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

Plugging this into (22), we obtain

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

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

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

and plugging back in to (21), we have $\mathbb{E}(R_{s,n}(f)) \leq c \cdot [f]^2_{W_d^{s,2}(\mathcal{X})}$.

To complete the proof, it remains to treat the case where $g \notin C_d^\infty(\mathcal{X}_{2r})$. Since $g \in W_d^{1,2}(\mathcal{X}_{2r})$, by Theorem 2 of (Evans), there exists a sequence of functions (g_m) such that $g_m \in C_d^\infty(\mathcal{X}_{2r}) \cap W_d^{1,2}(\mathcal{X}_{2r})$ and $\|g_m - g\|_{W_d^{1,2}(\mathcal{X}_{2r})} \to 0$. By the triangle inequality,

$$\int_{\mathcal{X}}\int_{\mathcal{X}} (g(y)-g(x))^2 K_r(y,x)\,dy\,dx \leq 6\int_{\mathcal{X}}\int_{\mathcal{X}} (g_m(x)-g(x))^2 K_r(y,x)\,dx + 3\int_{\mathcal{X}}\int_{\mathcal{X}} (g_m(y)-g_m(x))^2 K_r(y,x)\,dy\,dx.$$

The first term shrinks to zero as $m \to \infty$ since

$$\int_{\mathcal{X}}\int_{\mathcal{X}} (g_m(x)-g(x))^2 K_r(y,x)\,dy\,dx = \int_{\mathcal{X}} (g_m(x)-g(x))^2\,dx \leq \|g_m - g\|_{W_d^{1,2}(\mathcal{X}_{2r})},$$

and by our previous arguments,

$$\limsup_{m\to\infty}\int_{\mathcal{X}}\int_{\mathcal{X}} (g_m(y)-g_m(x))^2 K_r(y,x)\,dy\,dx \leq \limsup_{m\to\infty} K_{\max} r^2 [g_m]_{W_d^{1,2}(\mathcal{X}_{2r})} = K_{\max} r^2 [g]_{W_d^{1,2}(\mathcal{X}_{2r})} \leq c K_{\max} r^2 [g]_{W_d^{1,2}(\mathcal{X})}.$$

This concludes the proof of Lemma 13.