# Notes for Week of 1/9/20 - 1/16/20

Alden Green

January 19, 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 $K_r(x_i, x_j) = \|x_i - x_j\| \leq r$, 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 $f = f_0$, 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 $k$th 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 Holder spaces $C_d^s(\mathcal{X}; L)$.

**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 $d, R$, and $s$ but are independent of the sample size $n$ such that for every $\epsilon \geq 0$ satisfying*

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

*the worst-case risk is upper bounded*

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

# 1 Proof

Let $G = (V, E)$ be a graph over vertices $V = \{v_1, \dots, v_n\}$, and let $\beta = (\beta_1, \dots, \beta_n) \in \mathbb{R}^n$ be a signal over the vertices $V$. We observe responses $Y = (y_1, \dots, y_n)$ according to the model

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

Letting $L = VSV^T$ be the spectral decomposition of the Laplacian $L$ of $G$, our graph spectral test statistic is

$$T_{\text{spec}} = \frac{1}{n} \sum_{k=1}^{\kappa} \left( \sum_{i=1}^{n} v_i y_i \right)^2$$

where $\kappa$ is a tuning parameter. The resulting test we will use is

$$\phi_{\text{spec}} = \mathbf{1}\{T_{\text{spec}} \geq \frac{\kappa}{n} + t(b)\}, \quad \text{where } t(b) = b\sqrt{\frac{2\kappa}{n^2}} \quad \text{for } b \geq 1.$$

Let $S_s(\beta; G)$ be a measure of smoothness the signal $\beta$ displays over the graph $G$, given by

$$S_s(\beta; G) := \beta^T L^s \beta$$

In Lemma 1, we upper bound the Type I and Type II error of the test $\phi_{\text{spec}}$. Our bound on the Type II error will be stated as a function of $S_2(\beta; G)$ as well as the $\kappa$th eigenvalue $s_\kappa$.

**Lemma 1.** *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 $\phi_{\text{spec}}$ is upper bounded*

$$\mathbb{E}_{\beta_0}(\phi_{\text{spec}}) \leq \frac{1}{b^2}. \tag{5}$$

2. ***Type II error:*** *For any $b$ and $\beta$ such that*

$$\frac{1}{n} \sum_{i=1}^{n} \beta_i^2 \geq 2b\sqrt{\frac{2\kappa}{n^2}} + \frac{S_s(\beta; G)}{n s_\kappa^s} \tag{6}$$

*the Type II error of $\phi_{\text{spec}}$ is upper bounded,*

$$\mathbb{E}_\beta(1 - \phi_{\text{spec}}) \leq \frac{3}{b}. \tag{7}$$

To prove Lemma 1 we will first compute (bounds on) the expectation and variance of the test statistic $T_{\text{spec}}$, and then use Chebyshev's inequality to show (5) and (7).

*Mean of $T_{\text{spec}}$:* Using the notation $\langle v, w \rangle = \sum_{i=1}^{n} v_i w_i$, we have

$$\begin{aligned}
\mathbb{E}(T_{\text{spec}}) &= \frac{1}{n} \left( \sum_{k=1}^{\kappa} \langle \beta, v_k \rangle^2 + \mathbb{E}\big( \langle \varepsilon, v_k \rangle^2 + 2\langle \varepsilon, v_k \rangle \langle \beta, v_k \rangle \big) \right) \\
&= \frac{\kappa}{n} + \frac{1}{n} \sum_{k=1}^{\kappa} \langle \beta, v_k \rangle^2.
\end{aligned}$$

When $\beta = 0$, this equals $\kappa/n$. Otherwise, we have the following lower bound:

$$\begin{aligned}\sum_{k=1}^{\kappa} \langle \beta, v_k \rangle^2 &= \|\beta\|_2^2 - \sum_{k=\kappa+1}^{n} \langle \beta, v_k \rangle^2 \\ &\geq \|\beta\|_2^2 - \frac{1}{s_\kappa^s} \sum_{k=\kappa+1}^{n} \langle \beta, v_k \rangle^2 s_k^s \\ &\geq \|\beta\|_2^2 - \frac{S_s(\beta; G)}{s_\kappa^s},\end{aligned}$$

and therefore $\mathbb{E}(T_{\text{spec}}) \geq \kappa/n + n^{-1}(\|\beta\|_2^2 - S_s(\beta; G)/s_\kappa^s)$.

*Variance of $T_{\text{spec}}$:* We write $T_{\text{spec}} = n^{-1} y^T V_\kappa V_\kappa^T y$ where $V_\kappa$ is the $n \times \kappa$ matrix with eigenvectors $v_1, \ldots, v_\kappa$ as columns. Consequently,

$$\begin{aligned}\operatorname{Var}(T_{\text{spec}}) &= \frac{1}{n^2} \operatorname{Var}(y^T V_\kappa V_\kappa^T y) && (8) \\ &= \frac{1}{n^2} \operatorname{Var}((\beta + \varepsilon)^T V_\kappa V_\kappa^T (\beta + \varepsilon)) && (9) \\ &= \frac{1}{n^2} \operatorname{Var}(2\beta^T V_\kappa V_\kappa^T \varepsilon + \varepsilon^T V_\kappa V_\kappa^T \varepsilon) && (10) \\ &\leq \frac{1}{n^2} (4\beta^T V_\kappa V_\kappa^T \beta + 2\kappa) && (11)\end{aligned}$$

where the last inequality follows from standard properties of the Gaussian distribution. We now move on to showing the desired inequalities (5) and (7).

*Proof of (5):* By Chebyshev's inequality,

$$\begin{aligned}\mathbb{P}_{\beta=0}\big(T_{\text{spec}} \geq \frac{\kappa}{n} + t(b)\big) &\leq \mathbb{P}_{\beta=0}\Big(\Big|T_{\text{spec}} - \frac{\kappa}{n}\Big| \geq t(b)\Big) \\ &\leq \frac{\operatorname{Var}_{\beta=0}(T_{\text{spec}})}{t(b)^2} = \frac{1}{b^2}.\end{aligned}$$

*Proof of (7):* For simplicity, we introduce the notation

$$\Delta = \frac{\|\beta\|_2^2}{n} - \frac{S_s(\beta; G)}{n s_\kappa^s}.$$

Assumption (6) implies $\Delta \geq 2t(b)$. Then another application of Chebyshev's inequality gives us

$$\begin{aligned}\mathbb{P}_\beta\big(T_{\text{spec}} \leq \frac{\kappa}{n} + t(b)\big) &= \mathbb{P}_\beta\big(T_{\text{spec}} - \mathbb{E}_\beta(T_{\text{spec}}) \leq t(b) - \Delta\big) \\ &\leq \mathbb{P}_\beta\big(|T_{\text{spec}} - \mathbb{E}_\beta(T_{\text{spec}})| \leq \Delta - t(b)\big) && \text{(since } \Delta \geq t(b)) \\ &\leq \frac{\operatorname{Var}_\beta(T_{\text{spec}})}{(\Delta - t(b))^2} \\ &\leq 4\frac{\operatorname{Var}_\beta(T_{\text{spec}})}{\Delta^2} && \text{(since } \Delta \geq 2t(b)) \\ &\leq 4\frac{2\kappa/n^2 + 4\beta^T V_\kappa V_\kappa^T \beta/n^2}{\Delta^2}.\end{aligned}$$

We handle each summand in the numerator separately. For the first term, since $\Delta \geq 2t(b)$, we have

$$\frac{2\kappa}{n^2\Delta^2} \leq \frac{1}{2b^2}. \tag{12}$$

For the second term, noting that $\Delta = \beta^T V_\kappa V_\kappa^T \beta/n$, we have

$$\begin{aligned}\frac{\beta^T V_\kappa V_\kappa^T \beta/n^2}{\Delta^2} &= \frac{1}{n\Delta} \\ &\leq \frac{1}{2nt(b)} \\ &= \frac{1}{2b\sqrt{2\kappa}},\end{aligned} \tag{13}$$

and combining (12) and (13) yields (7).

#### 1.0.1 Step 2: Bounding neighborhood graph functionals

To make use of Lemma 1 we will need to show that when $r$ and $\kappa$ are appropriately tuned and $\|f\|_{\mathcal{L}^2(\mathcal{X})}$ is sufficiently large, the inequality (6) holds with respect to $G = G_{n,r}$ and $\beta = (f(x_1), \ldots, f(x_n))$. In particular, we will show that for some constants $c_1, c_2, c_3, c_4$ which may depend on $L$, $d$ and $s$ but do not depend on $n$, $f$ or $b$, the following statements:

(E1) **Graph Sobolev norm:** For $f \in C_d^s(\mathcal{X}; R)$, $p \in C_d^{s-1}(\mathcal{X}; R)$, and $1 \geq r(n) \geq n^{-1/(2(s-1)+d)}$,

$$S_s(f; G_{n,r}) \leq c_1 \cdot b \cdot n^{s+1} r^{s(d+2)} \tag{14}$$

(E2) **Eigenvalue tail bound:** For any $a > 0$ and $(\log n/n)^{1/d} n^a \leq r \leq n^{-4/((2+d)(4s+d))}$,and for $\kappa = n^{2d/(4s+d)}$,

$$s_\kappa \geq c_2 \cdot nr^{d+2} \kappa^{2/d} \tag{15}$$

(E3) **Empirical norm of $f$:** When $f \in C^s(\mathcal{X}; R)$ and $\|f\|_{\mathcal{L}^2} \geq c_3 \cdot b \cdot n^{-4s/(4s+d)}$,

$$\|f\|_n^2 \geq \frac{1}{b} \cdot \|f\|_{\mathcal{L}^2}^2 \tag{16}$$

each hold with probability at least $1 - c_4/b$ for sufficiently large $n$.

**Proof of (14):** We will take $s$ to be even, as the proof when $s$ is odd follows essentially the same steps. To simplify exposition, we introduce the iterated difference operator, defined recursively as

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

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

$$f^T L^s f = \sum_{i=1}^n \sum_{k \in [n]^q} \sum_{\ell \in [n]^q} r^{ds} D_k f(x_i) D_\ell f(x_i) \tag{17}$$

For given index vectors $k, \ell$ 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 $I$, specifically whether $I < s+1$ or $I = s+1$, and show that

$$\mathbb{E}(D_k f(x_i) D_\ell f(x_i)) = \begin{cases} O(r^{2s}), & \text{if } I = s+1 \\ O(r^2 r^{d(I-(2q+1))}), & \text{otherwise} \end{cases} \tag{18}$$

uniformly over $f \in C^s(L)$. Before proving (18), we verify that (14) is directly implied by (18). In the sum on the right hand side of (17), there are $O(n^I)$ terms with exactly $I$ distinct indices. When $I < s+1$, the total contribution of such terms to the sum is $O(n^I r^{d(I-1)+2})$. Since $r(n) \geq n^{-1/d}$, 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 $O(n^s r^{d(s-1)+2})$ which in light of the restriction $r \geq n^{-1/(2(s-1)+d)}$ is $O(n^{s+1} r^{s(d+2)})$. On the other hand when $I = s+1$, by (18) we immediately have that the total contribution of these terms is $O(n^{s+1} r^{2(s+d)})$. Therefore,

$$\mathbb{E}(f^T L^s f) = O(n^{s+1} r^{s(d+2)})$$

uniformly over $f \in C^s(L)$, and (14) by Markov's inequality.

Now we prove (18). Since $f \in C_d^s(R) \subseteq C_d^1(R)$, using a first-order Taylor expansion of $f(x)$ we can show that 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

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

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

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

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 second upper bound in (18).

To show the first upper bound in (18), we apply the law of iterated expectation,

$$\begin{aligned}\mathbb{E}\left[D_k f(x_i) D_\ell f(x_i)\right] &= \mathbb{E}\left[\mathbb{E}\left(D_k f(x) | x_i = x\right) \mathbb{E}\left(D_\ell f(x) | x_i = x\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(D_k f(x) | x_i = x\right)^2\right]. \end{aligned} \tag{19}$$

By Lemma 2, we have that $\mathbb{E}(D_k f(x)) = O(r^s)$ for all values of $x$, implying the desired result.

**Proof of (15):** We prove (15) 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 $\widetilde{r} = r/(3(2\sqrt{d}+1))$, $M = (1/\widetilde{r})^d$, $N = n\widetilde{r}^d$. Assume without loss of generality that $M$ and $N$ are integers. Additionally, for $t = n^{1/d}$ and $m = M^{1/d}$ let

$$\overline{X} = \left\{\frac{1}{t}(k_1, \ldots, k_d) : k \in [t]^d\right\}, \quad \overline{Z} = \left\{\frac{1}{m}(j_1, \ldots, j_d) : j \in [m]^d\right\}.$$

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 with high probability $G_{n,r} \succeq H$. If $(\overline{x}_k, \overline{x}_\ell) \in E_H$, then there exist $\overline{z}_i, \overline{z}_j$ such that

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

Assuming (25) holds, if $(\overline{x}_k, \overline{x}_\ell) \in E_H$, then for sufficiently large $n$

$$\|\pi(\overline{x}_k) - \pi(\overline{x}_\ell)\|_2 \leq 2c\left(\frac{\log n}{n}\right)^{1/d} + \frac{r}{3} \leq r,$$

implying that $(\pi(\overline{x}_k), \pi(\overline{x}_\ell)) \in E$. Therefore, $G_{n,r} \succeq \overline{G}_d^M \otimes K_N$ whenever (25) holds.

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,

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

where the last inequality can be verified by a quick calculation in light of $\kappa = n^{2d/(4s+d)}$ and $r \leq n^{-4/((2+d)(4s+d))}$. Since we've already shown that $\lambda_\kappa(G_{n,r}) \geq \lambda_\kappa(H)$ when (25) is satisfied, which happens probability $1 - o(n^{-1})$, this completes the proof of (15).

**Proof of (16):** Let $Z = \frac{1}{n}\sum_{i=1}^n f^2(x_i)$. We upper bound $\mathbb{E}[Z^2]$,

$$\begin{aligned}
\mathbb{E}[Z^2] &\leq \mathbb{E}(f^2(x_1))^2 + \frac{1}{n}\mathbb{E}(f^4(x_1)) \\
&\leq \mathbb{E}(f^2(x_1))^2 + \frac{R^2}{n}\mathbb{E}(f^2(x_1)) \\
&\overset{(i)}{\leq} \mathbb{E}(f^2(x_1))^2 \left(1 + \frac{R^2}{n/\mathbb{E}(f^2(x_1))}\right) \\
&\overset{(ii)}{\leq} \mathbb{E}(f^2(x_1))^2 \left(1 + \frac{R^2}{c_3^2 b^2 n^{d/(4s+d)}}\right)
\end{aligned}$$

where $(i)$ follows since $f \in C_d^s(\mathcal{X}; R)$ implies $|f(x)| \leq R$, and $(ii)$ follows by assumption. The statement then follows by the Paley-Zygmund inequqality.

### 1.0.2 Step 3: Conclusion

We note that for all possible values of $X \in \mathcal{X}^n$, under the null hypothesis $f = f_0 = 0$ and therefore $\beta = (f(x_1), \ldots, f(x_n)) = 0$ as well. Therefore by (5), we have the following bound on Type I error:

$$\mathbb{E}_{f_0}(\phi_{\text{spec}}) = \mathbb{E}(\mathbb{E}_{\beta=0}(\phi_{\text{spec}})|X) \leq \frac{1}{b^2}. \tag{20}$$

Now, we bound Type II error under the assumption $f \in C_d^s(\mathcal{X}; R), p \in C_d^{s-1}(\mathcal{X}; R)$ uniformly bounded away from 0 and $\infty$ over $\mathcal{X}$, and

$$\|f\|_{\mathcal{L}^2}^2 \geq \epsilon^2 = c_3^2 \cdot b^2 \cdot n^{-4s/(4s+d)}. \tag{21}$$

Choosing $n^{-1/(2(s-1)+d)} \leq r(n) \leq n^{-4/((2+d)(4s+d))}$, we may therefore apply our conclusions in Step 2; namely, that for every possible choice of $f$ there exists a good set $\mathcal{E}_f \subseteq \mathcal{X}^n$ with $\mathbb{P}(\mathcal{E}_f) \geq 1 - c_4/b$ such that each of (14), (15), and (16) hold for all $X \subseteq \mathcal{E}_f$. Choosing $\kappa = n^{2d/(4s+d)}$ to balance the squared bias and variance terms on the right hand side of (6), we have that for all $X \subseteq \mathcal{E}_f$

$$\begin{aligned} 2b\sqrt{\frac{2\kappa}{n^2}} + \frac{S_s(\beta; G_{n,r})}{ns_\kappa} &\le 2bn^{-4s/(4s+d)} + c \cdot b \cdot \frac{n^s r^{s(d+2)}}{s_\kappa^s} && \text{(by (14))} \\ &\le 2bn^{-4s/(4s+d)} + c \cdot b \cdot \frac{1}{n^{4s/(4s+d)}} && \text{(by (15))} \\ &\le \frac{1}{b}\|f\|_{\mathcal{L}^2} \\ &\le \frac{1}{n}\sum_{i=1}\beta_i^2. && \text{(by (16))} \end{aligned}$$

where the last two inequalities follow for a suitably large choice of $c_3$ in (21). We conclude that for all $X \subseteq \mathcal{E}_f$, the inequality (6) is satisfied with respect to $\beta = (f(x_1), \ldots, f(x_n))$ and $G = G_{n,r}$. As a result the worst-case Type II error is bounded

$$\sup_{\substack{f \in C_d^s(\mathcal{X};\mathcal{R}), \\ \|f\|_{\mathcal{L}^2} \ge \epsilon}} \mathbb{E}_f(1 - \phi_{\mathrm{spec}}) \le \sup_{\substack{f \in C_d^s(\mathcal{X};\mathcal{R}), \\ \|f\|_{\mathcal{L}^2} \ge \epsilon}} \mathbb{E}\big[\mathbb{E}_\beta(1 - \phi_{\mathrm{spec}} | X \in \mathcal{E}_f)\big] + \frac{c_4}{b} \le \frac{3 + c_4}{b},$$

completing the proof of Theorem 1.

## 2 Supporting Results

**Lemma 2.** *Let $s$ be a positive integer, and $R \ge 0$. Suppose $f \in C^s(R)$, $p \in C^{s-1}(R)$, $k \in [n]^q$ for some $q \ge 1$, and that $K_r$ is a 2nd-order kernel. Then if $2q \le s - 1$,*

$$\mathbb{E}(D_k f(x)) = \sum_{\ell=2q}^{s-1} O(r^\ell) \cdot f_\ell(x) + O(r^s), \tag{22}$$

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

*Proof.* We will prove Lemma 2 in the case where $d = 1$. When $d \ge 2$, 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 $d = 1$. The When $q = 1$ and $k \in [n]$, by taking Taylor expansions of $f$ and $p$, we obtain

$$\begin{aligned} \mathbb{E}(D_k f(x)) &= \sum_{\ell=1}^{s-1} f^{(\ell)}(x) \int (z - x)^\ell K_r(z, x) p(z) dz + O(r^s) \\ &= \sum_{\ell=1}^{s-1}\sum_{a=0}^{s-1} f^{(\ell)}(x) p^{(a)}(x) \underbrace{\int (z - x)^{\ell+a} K_r(z, x) dz}_{:= I_{t+a}} + O(r^s) \end{aligned} \tag{23}$$

Since $K$ is a 2nd-order kernel, $I_1 = 0$ and $I_t = O(r^t)$ for $t \ge 2$. Additionally, when $\ell + a \le s - 1$, we have that $f^{(\ell)} p^{(a)} \in C^{\min\{s-\ell, s-1-a\}} \subseteq C^{s-(\ell+a)}$, and $\left|f^{(\ell)} p^{(a)}\right| \le L^2$ for any $\ell$ and $a$. We can therefore simplify (23) by combining all terms where $\ell + a = t$, obtaining

$$\mathbb{E}(D_k f(x)) = \sum_{t=1}^{s-1} f_t(x) I_t + O(r^s) = \sum_{t=2}^{s-1} f_t(x) I_t + O(r^s), \tag{24}$$

which establishes (22) in the base case.

Now, we assume (22) 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 (24),

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

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 (22). $\square$

**Theorem 2** (Theorem 1 of "On the rate of Convergence of Empirical Measures in $\infty$-Transportation Distance")**.** *Let $X = \{x_1, \ldots, x_n\} \sim P$, where $P$ is a distribution on $\mathcal{X} = [0,1]^d$ with density $p$ satisfying*

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

*With probability at least $1 - o(n^{-1})$, there exists a bijection between grid points and data points $\pi : \overline{X} \to X$ such that*

$$\max_{k \in [t]^d} |\overline{x}_k - \pi(\overline{x}_k)| \leq c\left(\frac{\log n}{n}\right)^{1/d} \tag{25}$$