Notes for Week 3/12/20 - 3/19/20

Alden Green

March 28, 2020

1 Introduction

Suppose we observe data according to the following random design regression sampling model: we observe independent samples $x_1, \ldots, x_n \sim P$ and responses

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0,1) \tag{1}$$

where $f \in L^2(P)$ is a regression function and ε is independent standard Gaussian noise. We wish to test

$$\mathbf{H}_0 : f = 0, \quad \mathbf{H}_a : f \neq 0.$$

Our strategy will be to estimate $\|f\|_{\mathcal{L}^2(P)}$ using a test statistic T to be specified momentarily, and reject the null hypothesis (which we can equivalently state as $\|f\|_{\mathcal{L}^2} = 0$) if T is sufficiently large.

The truncated series estimator of f,

$$\Pi_{\kappa,G}(y) = \sum_{k=1}^{\kappa}\left(\sum_{i=1}^{n} v_{k,i}(G) y_i\right) v_k(G),$$

leads to the somewhat natural choice of test statistic $T = \|\Pi_{\kappa,G}(y)\|_n^2$. (Here, G is a neighborhood graph built on the data X.) A set of standard calculations lead to the following upper bound on approximation error:

$$\left|\|f\|_n^2 - \|\Pi_{\kappa,G}(y)\|_n^2\right| \leq \frac{f^T L^s f}{n \lambda_\kappa(G)^s}$$

When the graph G is formed appropriately, the Laplacian matrix L can be viewed as an estimate of a density-weighted Laplace-Beltrami operator Δ_P. Then $\lambda_\kappa(G)$ can be viewed as an estimate of the (κ)th eigenvalue of Δ_P, and $f^T L^s f$ as an estimate of a Sobolev semi-norm built via the operator Δ_P. When the density p satisfies appropriate regularity conditions, the population level quantities these estimates approximate are roughly

$$\langle \Delta_P^s f, f \rangle_{\mathcal{L}^2(P)} \sim |f|_{H^s(\mathcal{X})}, \quad \lambda_\kappa(G)^2 \sim \kappa^{-2s/d}$$

and the resulting approximation error incurred by projecting f onto κ eigenvectors in $\mathcal{L}^2(P_n)$ is comparable to that incurred by projecting f onto κ elements of the Fourier basis in $\mathcal{L}^2$.

Of course, this not a formal proof, and indeed when $s > 1$ we will see that problems arise. Let us assume for the moment that $f \in C^2(\mathcal{X}; R)$. In this case the second-order graph Sobolev seminorm $f^T L_{n,r}^2 f$ is a biased estimator of the second-order Sobolev seminorm $\int (f^{(2)}(x))^2 \, dP(x)$. This bias will be non-trivial when r is sufficiently small. On the other hand $\lambda_\kappa(G)$ is also a biased estimator of $\lambda_\kappa(\Delta_P)$, and this bias will be non-trivial when r and κ are sufficiently large. In particular, when $d > 4$, taking $\kappa = n^{2d/(8+d)}$ as dictated by the usual bias-variance tradeoff in nonparametric testing problems, there exists no choice of r for which the biases of both the seminorm $f^T L_{n,r}^2 f$ and the eigenvalue $\lambda_\kappa(G)$ are trivial. A test formed using the statistic $\|\Pi_{\kappa,G}(y)\|_n^2$ is therefore, at best, only minimax optimal when $d \leq 4$.

For this reason we introduce a bias correction of our statistic.

1.1 Bias-Corrected Test Statistic

Formally, define the (unweighted) neighborhood graph $G_{n,r} = (X, E)$, where $X = \{x_1, \ldots, x_n\}$, and $(x_i, x_j) \in E \subset X \times X$ if and only if $\|x_i - x_j\|_2 \leq r$. Let $L_{n,r} := L_{G_{n,r}}$ be the Laplacian matrix associated with $G_{n,r}$. Our test statistic will be defined as a simple function of the norm of the projection of Y onto the eigenvectors of $L_{n,r}$,

$$T_{\mathrm{spec}}(G_{n,r}) := \|\Pi_{\kappa, G_{n,r}}(Y)\|_n^2 + \frac{y^T L_{n,r} y}{2n\big(\lambda_\kappa(G_{n,r})\big)^2}$$

The second term on the right hand side is the bias-correction term. Our test is then

$$\phi_{\mathrm{spec}}(G_{n,r}) := \mathbf{1}\{T_{\mathrm{spec}}(G_{n,r}) \geq \tau(b)\}$$

where b is a user-specified model encoding tolerance for error, and τ is a function of b to be specified later.

The bias corrected graph spectral projection test is minimax optimal over the compactly supported Holder ball $C_0^2(\mathcal{X}; L)$ whenever $d < 8$.

Theorem 1. *Suppose we observe samples $(x_i, y_i)_{i=1}^n$ according to the model (1). Let $L > 0$ and $b \geq 1$ be fixed constants, and $d < 8$. Suppose that P is an absolutely continuous probability measure over $\mathcal{X} = [0,1]^d$ with density function $p(x) \in C^1(\mathcal{X}; p_{\max})$ bounded away from zero and infinity,*

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

and the test $\phi_{\mathrm{spec}}(G_{n,r})$ is performed with parameter choices

$$r(n) = n^{-2/(8+d)}, \ \kappa = n^{2d/(4+d)}, \ \tau(b) = \frac{\kappa}{n} + \frac{1}{2n\lambda_\kappa^2}\sum_{k=1}^{n}\lambda_k + \sqrt{6}b\sqrt{\frac{2\kappa}{n^2}}$$

Then the following statements holds for every n sufficiently large: there exists constants c_1, c_2 which do not depend on n, b or R such that for every $\epsilon \geq 0$ satisfying

$$\epsilon^2 \geq c_1^2 \cdot b^2 \cdot L^2 \cdot n^{-8/(8+d)} \tag{2}$$

the worst-case risk is upper bounded

$$\mathcal{R}_\epsilon(\phi_{\mathrm{spec}}(G_{n,r}); C_0^2(\mathcal{X}; R)) \leq \frac{c_2}{b}. \tag{3}$$

2 Analysis

2.1 Fixed graph testing error

To prove Theorem 1, we show that there exists a high-probability set $E \subseteq \mathcal{X}^n$ such that conditional on $X \in E$, the test $\phi_{\mathrm{spec}}(G_{n,r})$ has small risk. Since $G_{n,r}$ is a function only of X and not of Y, this amounts to reasoning about the behavior of the test ϕ_{spec} over a fixed graph $G = (X, E)$, where we observe

$$y_i = \beta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,1) \tag{4}$$

for some fixed $\beta \in \mathbb{R}^n$.

In Lemma 1, we upper bound the Type I and Type II error of the test $\phi_{\mathrm{spec}}(G)$. Our bound on the Type II error will be stated as a function of $\beta^T L^2 \beta$–a measure of the smoothness the signal β displays over the graph G–as well as the κth eigenvalue λ_κ. (For ease of reading, in the statement and proof of this Lemma, we drop the notational dependence of the Laplacian L and eigenvalues λ on the graph G.)

Lemma 1. *Let $1 \leq \kappa \leq n$ be an integer. Suppose we observe data according to the model (4), and we perform the test $\phi_{\mathrm{spec}}(G)$ with*

$$\tau(b) = \frac{\kappa}{n} + \frac{1}{2n\lambda_\kappa^2}\sum_{k=1}^{n}\lambda_k + 2\sqrt{6}\cdot b\sqrt{\frac{\kappa}{n^2}}.$$

Assume that

$$\max\Big\{\beta^T L^2 \beta, n\lambda_n^2\Big\} \leq \kappa \cdot \lambda_\kappa^4 \tag{5}$$

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

$$\mathbb{E}_{\beta_0}(\phi_{\mathrm{spec}}) \leq \frac{1}{b^2}. \tag{6}$$

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

$$\frac{\beta^T L \beta}{2n\lambda_\kappa^2} + \|\Pi_{\kappa,G}(\beta)\|_n^2 \geq 2\sqrt{6}\cdot b\sqrt{\frac{\kappa}{n^2}} \tag{7}$$

the Type II error of $\phi_{\mathrm{spec}}(G)$ is upper bounded,

$$\mathbb{E}_{\beta}(1 - \phi_{\mathrm{spec}}) \leq \frac{1}{b^2} + \frac{16}{\sqrt{6}\cdot b\sqrt{\kappa}}. \tag{8}$$

In particular if

$$\frac{1}{n}\sum_{i=1}^{n}\beta_i^2 \geq 2\sqrt{6}\cdot b\sqrt{\frac{\kappa}{n^2}} + \frac{U_2(\beta;G)}{4n\lambda_\kappa^2} \tag{9}$$

where

$$U_2(\beta;G) := \sum_{i\in(n)^3}(\beta_{i_1} - \beta_{i_2})(\beta_{i_1} - \beta_{i_3})A_{i_1i_2}A_{i_1i_3}$$

then (7) and thus (8) follow.

2.2 Variance of bias-correction term

We pay a price in increased variance for introducing the bias-correction term $y^T Ly/(2n\lambda_\kappa^2)$; the condition (5) is sufficient to guarantee this increase in variance is no greater than the variance $\mathrm{Var}(||\Pi_{\kappa,G}(Y)||_n^2)$ of the uncorrected statistic. In this section, we show that under the conditions of Theorem 1, the condition (5) holds with high probability.

Lemma 2. *Fix $b \geq 1$. Suppose $f \in C_0^2(\mathcal{X}; L)$, and let $r = n^{-2/(8+d)}$ and $\kappa = n^{2d/(8+d)}$. Then there exist constants c and C such that for all n sufficiently large and all $d < 8$,*

$$\max\left\{f^T L_{n,r}^2 f,\ n\big(\lambda_n(G_{n,r})\big)^2\right\} \leq \kappa\cdot\big(\lambda_\kappa(G_{n,r})\big)^4 \tag{10}$$

with probability at least $1 - 1/b - 1/(nr^4) - C(n + r^{-d})\exp\{-cnr^d\}$.

2.3 Approximation error

The following Lemma is the main result of this section.

Lemma 3. *Set $r = n^{-2/(8+d)}$ and $\kappa = n^{2d/(8+d)}$. Then if $f \in C_0^2(\mathcal{X}; L)$, $p \in C^1(\mathcal{X}; p_{\max})$, and additionally there exists*

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

then the following statement holds: there exist constants c and C which depend only on $p_{\min}, p_{\max}$ and d such that if

$$\|f\|_{\mathcal{L}^2}^2 \geq C \cdot b^2 \cdot L^2 \cdot n^{-8/(8+d)} \tag{11}$$

then

$$\|f\|_n^2 \geq 2\sqrt{6}b\sqrt{\frac{\kappa}{n^2}} + \frac{U_2(f; G_{n,r})}{4n[\lambda_\kappa(G_{n,r})]^2}$$

with probability at least $1 - 5/b - 1/(nr^4) - C(n + r^{-d})\exp\{-cnr^d\}$.

We note that for $r = n^{-2/(8+d)}$, each of

$$\frac{1}{nr^4} \to 0, \; C(n + r^{-d})\exp\{-cnr^d\} \to 0 \quad \text{as } n \to \infty$$

and these contributions to the overall testing error are therefore asymptotically negligible. Theorem 1 then follows from combining Lemmas 2 and 3 with Lemma 1,

The following three Lemmas will be used in the proof of Lemma 3. Lemma 4 – copied from the graph testing notes – implies a sufficient lower bound on $\|f\|_n^2$ when $s = 2$. It is stronger than we need, in the sense that it applies to all functions bounded in the second-order Sobolev norm rather than merely the second-order Holder norm; in our case we could obtain tighter bounds on the probability, but since we eventually hope to apply our argument to Sobolev functions anyway, we won't bother with this.

Lemma 4. *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 $f \in H^s(\mathcal{X})$. Then for any $b \geq 1$, there exists c_1 such that if

$$\|f\|_{\mathcal{L}^2(\mathcal{X})} \geq \begin{cases} c_1 \cdot b \cdot \|f\|_{H^s(\mathcal{X})} \cdot \max\{n^{-1/2}, n^{-s/d}\}, & \textit{if } 2s \neq d \\ c_1 \cdot b \cdot \|f\|_{H^s(\mathcal{X})} \cdot n^{-a/2}, & \textit{if } 2s = d \textit{ for any } 0 < a < 1 \end{cases} \tag{12}$$

then,

$$\mathbb{P}\left[\|f\|_n^2 \geq \frac{1}{b}\mathbb{E}\big[\|f\|_n^2\big]\right] \geq 1 - \frac{5}{b} \tag{13}$$

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

Lemma 5 establishes that the eigenvalues $\lambda_\kappa(G_{n,r})$ scale at sufficiently fast rate as κ increases.

Lemma 5. *Suppose p is bounded away from 0 and ∞ everywhere on its support, i.e.*

$$0 < p_{\min} \leq p(x) \leq p_{\max} < \infty, \quad \textit{for all } x \in [0,1]^d.$$

Then there exists constants c and C which depend only on $p_{\min}, p_{\max}$ and d, such that for all $1 \leq k \leq n$,

$$\lambda_k(G_{n,r}) \geq c \cdot \min\Big\{r^{d+2}nk^{2/d}, nr^d\Big\} \tag{14}$$

with probability at least $1 - C(n + r^{-d})\exp\big\{-cnr^d\big\}$.

The first term on the right hand side of (14) is the scaling of the eigenvalues we desire. The second term is an artificial truncation on this spectrum. For the choices of r and κ in Lemma 3, these terms are exactly equal, and the truncation therefore does not come into play.

In Lemma 6, we establish that U-statistic $U_2(f; G_{n,r})$ scales appropriately with the Holder norm of f.

Lemma 6. *If $f \in C_0^2(\mathcal{X}; L)$, $p \in C^1(\mathcal{X}; L)$, and $n^{-1/d} \leq r \leq 1$, then there exists a constant c such that*

$$\left|U_2(f; G_{n,r})\right| \leq c n^3 r^{2(d+2)} L^2 p_{\max}^2$$

with probability at least $1 - 1/(nr^4)$.

3 Proofs

3.1 Proof of Lemma 1

In this proof, we will drop all notational dependence on the graph G for ease of reading.

To prove Lemma 1 we will first compute (bounds on) the expectation and variance of the test statistic T_{spec}, and then use Chebyshev's inequality to show (6) and (8).

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

$$\begin{aligned}
\mathbb{E}\big[T_{\text{spec}}\big] &= \frac{1}{n}\sum_{k=1}^{\kappa}\Big(\langle \beta, v_k\rangle^2 + \mathbb{E}\Big[\langle \varepsilon, v_k\rangle^2 + 2\langle \varepsilon, v_k\rangle\langle \beta, v_k\rangle\Big]\Big) + \frac{1}{2n\lambda_\kappa^2}\mathbb{E}\Big[(\beta+\varepsilon)^T L(\beta+\varepsilon)\Big] \\
&= \frac{\kappa}{n} + \|\Pi_{\kappa,G}(\beta)\|_n^2 + \frac{1}{2n\lambda_\kappa^2}\left(\beta^T L\beta + \sum_{k=1}^n \lambda_k\right)
\end{aligned}$$

Variance of T_{spec}: Note that $T_{\text{spec}} = n^{-1}y^T V_\kappa V_\kappa^T y$ where V_κ is the $n\times\kappa$ matrix with eigenvector v_k as its kth column. Therefore, applying Cauchy-Schwarz,

$$\begin{aligned}
\mathrm{Var}(T_{\text{spec}}) &\leq \frac{2}{n^2}\left(\mathrm{Var}\Big(y^T V_\kappa V_\kappa^T y\Big) + \frac{1}{8\lambda_\kappa^4}\mathrm{Var}\Big(y^T L y\Big)\right) \\
&= \frac{2}{n^2}\left(\mathrm{Var}\Big((\beta+\varepsilon)^T V_\kappa V_\kappa^T(\beta+\varepsilon)\Big) + \frac{1}{8\lambda_\kappa^4}\mathrm{Var}\Big((\beta+\varepsilon)^T L(\beta+\varepsilon)\Big)\right) \\
&= \frac{2}{n^2}\left(4n\|\Pi_{\kappa,G}(\beta)\|_n^2 + 2\kappa + \frac{1}{4\lambda_\kappa^4}\Big(2\beta^T L^2\beta + \sum_{k=1}^n \lambda_k^2\Big)\right) \\
&\leq \frac{2}{n^2}\left(4n\|\Pi_{\kappa,G}(\beta)\|_n^2 + 2\kappa + \frac{1}{4\lambda_\kappa^4}\Big(2\beta^T L^2\beta + n\lambda_n^2\Big)\right) \\
&\leq \frac{2}{n^2}\left(4n\|\Pi_{\kappa,G}(\beta)\|_n^2 + 3\kappa\right)
\end{aligned}$$

where the last equality follows from standard properties of the Gaussian distribution, and the last inequality follows from (5). We now move on to showing the desired inequalities (6) and (8). Let $t(b) := b\sqrt{\frac{\kappa}{n^2}}$.

Proof of (6): By Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}_{\beta=0}\left(T_{\text{spec}} \geq \frac{\kappa}{n} + \frac{1}{2n\lambda_\kappa^2}\sum_{k=1}^n \lambda_k + 4t(b)\right) &\leq \mathbb{P}_{\beta=0}\left(T_{\text{spec}} - \mathbb{E}\big[T_{\text{spec}}\big] \geq \sqrt{6}t(b)\right) \\
&\leq \frac{\mathrm{Var}(T_{\text{spec}})}{\big(\sqrt{6}t(b)\big)^2} = \frac{1}{b^2}.
\end{aligned}$$

Proof of (8): For simplicity, we introduce the notation

$$\Delta = \|\Pi_{\kappa,G}(\beta)\|_n^2 + \frac{1}{2n\lambda_\kappa^2}\beta^T L\beta$$

Assumption (7) implies $\Delta \geq 2\sqrt{6}\cdot t(b)$, and we have already shown that $\mathbb{E}_\beta(T_{\mathrm{spec}}) = \Delta + \kappa/n + \big(\sum_{k=1}^n \lambda_k\big)/(n\lambda_\kappa^2)$. Then another application of Chebyshev's inequality gives us

$$\begin{aligned}
\mathbb{P}_\beta\bigg(T_{\mathrm{spec}} \leq \frac{\kappa}{n} + \frac{1}{n\lambda_\kappa^2}\sum_{k=1}^n \lambda_k^2 + \sqrt{6}\cdot t(b)\bigg) &= \mathbb{P}_\beta\bigg(T_{\mathrm{spec}} - \mathbb{E}_\beta(T_{\mathrm{spec}}) \leq \sqrt{6}\cdot t(b) - \Delta\bigg) \\
&\leq \mathbb{P}_\beta\bigg(|T_{\mathrm{spec}} - \mathbb{E}_\beta(T_{\mathrm{spec}})| \leq \Delta/2\bigg) \qquad (\text{since } \Delta \geq 2\sqrt{6}\cdot t(b)) \\
&\leq \frac{4\mathrm{Var}_\beta(T_{\mathrm{spec}})}{\Delta^2} \\
&\leq \frac{24\kappa/n^2 + 32\Delta/n}{\Delta^2}.
\end{aligned}$$

We handle each summand in the numerator separately. For the first term, since $\Delta \geq 2\sqrt{6}\cdot t(b)$, we have

$$\frac{24\kappa}{n^2\Delta^2} \leq \frac{1}{b^2}. \tag{15}$$

For the second term we have

$$\frac{32}{n\Delta} \leq \frac{16}{\sqrt{6}\cdot nt(b)} = \frac{16}{\sqrt{6}\cdot b\sqrt{\kappa}}, \tag{16}$$

and combining (15) and (16) yields (8).

(9) **implies** (7). Since the eigenvalues λ_k are ordered from smallest to largest,

$$\begin{aligned}
\|\Pi_{\kappa,G}(\beta)\|_n^2 &= \|\beta\|_n^2 - \frac{1}{n}\sum_{k=\kappa+1}^n \langle \beta, v_k\rangle^2 \\
&\geq \|\beta\|_n^2 - \frac{1}{n\lambda_\kappa^2}\sum_{k=\kappa+1}^n \langle \beta, v_k\rangle^2\lambda_k^2 \\
&\geq \|\beta\|_n^2 - \frac{\beta^T L^2\beta}{n\lambda_\kappa^2},
\end{aligned}$$

Therefore since we may rewrite the second-order graph Sobolev seminorm as,

$$\begin{aligned}
\frac{\beta^T L^2\beta}{n\lambda_\kappa^2} &= \frac{1}{n\lambda_\kappa^2}\sum_{i=1}^n (L\beta)_i^2 \\
&= \frac{1}{4n\lambda_\kappa^2}\sum_{i\in[n]^3}\Big(\beta_{i_1} - \beta_{i_2}\Big)\Big(\beta_{i_1} - \beta_{i_3}\Big)A_{i_1,i_2}A_{i_1,i_3} \\
&= \frac{1}{2n\lambda_\kappa^2}\bigg(\beta^T L\beta + \frac{1}{2}\sum_{i\in(n)^3}\Big(\beta_{i_1} - \beta_{i_2}\Big)\Big(\beta_{i_1} - \beta_{i_3}\Big)A_{i_1,i_2}A_{i_1,i_3}\bigg) \\
&= \frac{1}{2n\lambda_\kappa^2}\bigg(\beta^T L\beta + \frac{1}{2}U_2(\beta; G)\bigg)
\end{aligned}$$

we have

$$\|\Pi_{\kappa,G}(\beta)\|_n^2 + \frac{\beta^T L\beta}{2n\lambda_\kappa^2} \geq \|\beta\|_n^2 - \frac{1}{4n\lambda_\kappa^2}U_2(\beta; G)$$

whence we see that (9) implies (7).

3.2 Proof of Lemma 2

By Lemma 5 there exist constants c and C such that

$$\lambda_\kappa(G_{n,r}) \geq c \cdot \min\Big\{nr^{d+2}\kappa^{2/d}, nr^d\Big\} \tag{17}$$

with probability at least $1 - C(n + r^{-d})\exp\{-cnr^d\}$. Note that when $r = n^{-2/(8+d)}$, and $\kappa = n^{2d/(8+d)}$ as specified in the statement of Lemma 2, the two terms inside the minimum are equal. We will now supply upper bounds on both $f^T L_{n,r}^2 f$ and $n\big(\lambda_n(G_{n,r})\big)^2$, which combined with (17) are sufficient to prove the claim of Lemma 2.

Upper bound on $f^T L_{n,r}^2 f$. We begin by rewriting the quadratic form $f^T L_{n,r}^2 f$ as

$$f^T L_{n,r}^2 f = 2 f^T L_{n,r} f + U_2(f; G_{n,r})$$

To upper bound the first term on the right hand side, we compute its expectation

$$\begin{aligned}
\mathbb{E}\Big[f^T L_{n,r} f\Big] &= n(n-1)\mathbb{E}\Big[\big(f(x_i) - f(x_j)\big)^2 K_r(x_i, x_j)\Big] \\
&\leq n(n-1)L^2 r^2 \mathbb{E}\Big[K_r(x_i, x_j)\Big] \\
&\leq n(n-1)L^2 r^{2+d} p_{\max}.
\end{aligned}$$

Since $f^T L_{n,r} f$ is always non-negative, by Markov's inequality

$$f^T L_{n,r} f \leq b n(n-1) L^2 r^{2+d} p_{\max}$$

with probability at least $1 - 1/b$.

In combination with the upper bound on $|U_2(f; G_{n,r})|$ given by Lemma 6, choosing c as in that Lemma we have

$$f^T L_{n,r}^2 f \leq 2bn^2 L^2 r^{2+d} p_{\max} + cn^3 r^{2(d+2)} L^2 p_{\max}^2 \tag{18}$$

with probability at least $1 - 1/b - 1/(nr^4)$.

Assuming the high probability bounds (17) and (18) hold, the routine calculations

$$\frac{\Big(\lambda_\kappa(G_{n,r})\Big)^4 \kappa}{n^3 r^{2(d+2)}} \geq n^3 r^{2(d+2)} \geq n^3 n^{-4(d+2)/(d+8)} = n^{(16-d)/(8+d)} \geq n^{1/2}$$

and

$$\frac{\Big(\lambda_\kappa(G_{n,r})\Big)^4 \kappa}{n^2 r^{(d+2)}} \geq n^4 r^{3(d+2)} \geq n^4 n^{-6(d+2)/(d+8)} = n^{(20-2d)/(8+d)} \geq n^{1/4}$$

imply $f^T L_{n,r}^2 f \leq \kappa\Big(\lambda_\kappa(G_{n,r})\Big)^4$ for all n sufficiently large.

Upper bound on $n\lambda_n^2$ It is well known that for any graph G, it's largest eigenvalue $\lambda_n(G) \leq 2\deg_{\max}(G)$. For the particular case of $G = G_{n,r}$, by Lemma 7

$$\deg_{\max}(G_{n,r}) \leq 2p_{\max}\nu_d n r^d$$

with probability at least $1 - n\exp\big\{-p_{\min}\nu_d n r^d\big\}$. Along with (17) this implies that for an appropriate choice of constant c

$$\frac{\kappa\Big(\lambda_\kappa(G_{n,r})\Big)^4}{n\Big(\lambda_n(G_{n,r})\Big)^2} \geq cnr^{2d}\kappa = cn^{(8-d)/(8+d)}$$

whence the claim (10) follows since $d < 8$.

3.3 Proof of Lemma 3

Let us take the high probability conclusions of Lemmas 4, 5, and 6 as given, namely that for $\kappa = n^{2d/(8+d)}$ and $r = n^{-2/(8+d)}$,

1. $\|f\|_n^2 \geq \frac{1}{b}\|f\|_{\mathcal{L}^2}^2$,
2. for some constant c_1, $\left[\lambda_\kappa(G_{n,r})\right]^2 \geq c_1 n r^{d+2}\kappa^{2/d}$, and
3. for some constant c_2, $|U_2(f; G_{n,r})| \leq c_2 n^3 r^{2(d+2)}\|f\|_{C^2(\mathcal{X})}^2$.

Then, for a sufficiently large choice of constant C in (11),

$$\begin{aligned}
2\sqrt{6}b\sqrt{\frac{\kappa}{n^2}} + \frac{U_2(f; G_{n,r})}{4n[\lambda_\kappa(G_{n,r})]^2} &\leq \Big(2\sqrt{6}b + \frac{c_2}{c_1}\Big)n^{-8/(8+d)} \\
&\leq \frac{1}{b}\|f\|_{\mathcal{L}^2}^2 \\
&\leq \|f\|_n^2.
\end{aligned}$$

The probability with which this holds then comes from accumulating probabilities with which the conclusions of Lemmas 4, 5, and 6 hold.

3.4 Proof of Lemma 5

In the following proof, for graphs G and H defined on a common vertex set V, we write $G \preceq H$ if $f^T L_G f \leq f^T L_H f$ for every $f \in L^2(V)$.

We will prove Lemma 5 by comparing the graph $G_{n,r}$ to a graph $\widetilde{G}_{n,r}$ which we now define. Let $\widetilde{r} = r/(1+\sqrt{d})$, and $M = 1/\widetilde{r}$; without loss of generality we will assume M is an integer. The grid points

$$\bar{Z} := \Big\{\frac{1}{2M}(2m_1 - 1, \ldots, 2m_d - 1) : m \in [M]^d\Big\}$$

induces a tesselation $\mathcal{Q}\big(\bar{Z}\big)$ of $[0,1]^d$, defined as

$$Q(x) = \Big[x_1 - \frac{1}{2M}, x_1 + \frac{1}{2M}\Big] \times \ldots \times \Big[x_d - \frac{1}{2M}, x_d + \frac{1}{2M}\Big], \quad \mathcal{Q}\big(\bar{Z}\big) := \big\{Q(\bar{z}) : \bar{z} \in \bar{Z}\big\}$$

In the graph $\widetilde{G}_{n,r} = (X, \widetilde{E})$, there is an edge between x_i and x_J if they are in the same or adjoining cubes in $\mathcal{Q}(\bar{Z})$; formally speaking $(x_i, x_j) \in \widetilde{E}$ if either (a) there exists $\bar{z} \in \bar{Z}$ such that $x_i \in \mathcal{Q}(\bar{Z})$ and $x_j \in \mathcal{Q}(\bar{z})$ or (b) there exist $\bar{z}, \bar{z}' \in \mathcal{Q}(\bar{z}')$ such that $x_i \in \mathcal{Q}(\bar{z})$, $x_j \in \mathcal{Q}(\bar{z}')$, and $\|\bar{z} - \bar{z}\| = 1/M$.

Let

$$Q_{\min} = \min_{\bar{z} \in \bar{Z}} \Big|Q(\bar{z}) \cap X\Big|, \quad Q_{\max} = \max_{\bar{z} \in \bar{Z}} \Big|Q(\bar{z}) \cap X\Big|$$

We will prove the following three claims. The first two are deterministic statements regarding graph comparisons, while the third involves the concentration of the functionals $Q_{\min}$ and $Q_{\max}$.

1. The graphs $G_{n,r}$ and $\widetilde{G}_{n,r}$ satisfy the partial ordering

$$\widetilde{G}_{n,r} \preceq G_{n,r}$$

2. For each $k = 1, \ldots, n$, the eigenvalues $\lambda_k(\widetilde{G}_{n,r})$ satisfy

$$\lambda_k(\widetilde{G}_{n,r}) \geq \frac{Q_{\min}^2}{2Q_{\max}} \min\Big\{\frac{k^{2/d}}{M^2}, d\Big\}$$

3. When n is sufficiently large, there exist constants c_1 and c_2 such that

$$c_1 p_{\min} n r^d \leq Q_{\min} \leq Q_{\max} \leq c_2 p_{\max} n r^d$$

with probability at least $1 - c_2 r^{-d} \exp\{-c_1 p_{\min} n r^d\}$.

The conclusion of Lemma 5 follows straightforwardly once these statements are proved.

Claim 1: Partial Ordering. By definition if $x_i \sim x_j$ in $\widetilde{G}_{n,r}$, then there exists $\bar{z}_\ell$ and $\bar{z}_m$ in $\bar{Z}$ such that

$$\|x_i - \bar{z}_\ell\|_2 \leq \frac{\sqrt{d}}{2M}, \|x_j - \bar{z}_m\|_2 \leq \frac{\sqrt{d}}{2M}, \|\bar{z}_\ell - \bar{z}_m\|_2 \leq \frac{1}{M}$$

Therefore by the triangle inequality,

$$\|x_i - x_j\|_2 \leq \frac{1 + \sqrt{d}}{M} = r$$

and so $x_i \sim x_j$ in $G_{n,r}$.

Claim 2: Eigenvalues of $\widetilde{G}_{n,r}$. Let $\bar{G}_d^M$ be the lattice graph formed over the grid points $\bar{Z}$, let $V_m = X \cap Q(\bar{z}_m)$ for each $m \in [M]^d$, and let $\widetilde{H} = (V_{\widetilde{H}}, E_{\widetilde{H}})$ where

$$V_{\widetilde{H}} = \bigcup_{m \in [M]^d} \bigcup_{x_i \in V_m} (\bar{z}_m, x_i), \quad \text{and } \Big((\bar{z}_m, x_i), (\bar{z}_o, x_j)\Big) \in E_{\widetilde{H}} \text{ if either } m = o \text{ or } z_m \sim z_o \text{ in } \bar{G}_d^M$$

Lemma 9 implies the following lower bound on the eigenvalues of H

$$\begin{aligned} \lambda_k(\widetilde{H}) &\geq \frac{1}{2} \frac{Q_{\min}^2}{Q_{\max}} \min\Big\{\lambda_k(\bar{G}_d^M), \deg_{\min}(\bar{G}_d^M)\Big\} \\ &\geq \frac{1}{2} \frac{Q_{\min}^2}{Q_{\max}} \min\Big\{\frac{k^{2/d}}{M}, d\Big\} \end{aligned}$$

where the latter inequality follows from known facts about the eigenvalues and degrees of the d-dimensional lattice. Since $\widetilde{G}_{n,r}$ is isomorphic to $\widetilde{H}$, the same lower bound follows for the eigenvalues $\lambda_k(\widetilde{G}_{n,r})$.

Claim 3: Bounding $Q_{\min}$ and $Q_{\max}$. This follows from Hoeffding's inequality, as laid out in Lemma 8.

3.5 Proof of Lemma 6

We will bound the expectation and variance of $U_2(f; G_{n,r})$; then prove the bound on the latter is sufficient to show concentration around the former at a sufficient rate.

For convenience, we introduce the notation

$$D_i f(x) := (f(x_i) - f(x)) K_r(x_i, x), \quad K_r(x, z) := \mathbf{1}\Big\{\|x - z\|_2 \leq r\Big\}$$

Expectation of $U_2(f; G_{n,r})$. By the linearity of expectation, and the law of iterated expectation

$$\begin{aligned} \mathbb{E}\Big[U_2(f; G_{n,r})\Big] &= n(n-1)(n-2) \cdot \mathbb{E}\Big[D_1 f(x_2) D_1 f(x_3)\Big] \\ &\leq n^3 \cdot \mathbb{E}\Big[\mathbb{E}\Big[D_1 f(x_2) \mid x_1\Big]^2\Big]. \end{aligned}$$

It therefore suffices to show

$$\Big(\mathbb{E}\Big[D_1 f(x_2) \mid x_1\Big]\Big)^2 \leq 2 p_{\max}^2 L^2 r^{2(2+d)} \tag{19}$$

we will do using Taylor expansions of f and p. Writing the expectation as an integral, we have

$$-\mathbb{E}\Big[D_1 f(x_2) \mid x_1 = x\Big] = \int \Big(f(y) - f(x)\Big) K_r(y,x)\; p(y)\, dy$$

Then, since $f \in C^2(\mathcal{X}; L)$, we have

$$\Big|f(y) - f(x) - \big(\nabla f(x)\big)^T (y - x)\Big| \leq L\|y - x\|_2$$

for almost every $x \in \mathcal{X}$, and as a result

$$\left|\int \Big(f(y) - f(x)\Big) K_r(y,x)\; p(y)\, dy - \int \Big(\big(\nabla f(x)\big)^T (y-x)\Big) K_r(y,x)\; p(y)\, dy\right| \leq L p_{\max} r^{d+2}$$

If $\mathrm{dist}(x, \partial\mathcal{X}) \leq r$, then since f has first derivative zero at the boundary of $\mathcal{X}$, $\|\nabla f(x)\|_2 \leq r$, which in turn gives

$$\left|\int \Big(\big(\nabla f(x)\big)^T (y-x)\Big) K_r(y,x)\; p(y)\, dy\right| \leq L p_{\max} r^{d+2}$$

proving (19) when $\mathrm{dist}(x_1, \partial X) \leq r$.

Otherwise $\mathrm{dist}(x_1, \partial X) > r$. Since $p \in C^1(\mathcal{X}; p_{\max})$, we have

$$|p(y) - p(x)| \leq p_{\max} r$$

and therefore

$$\left|\int \Big(\big(\nabla f(x)\big)^T (y-x)\Big) K_r(y,x)\; p(y)\, dy - p(x)\int \Big(\big(\nabla f(x)\big)^T (y-x)\Big) K_r(y,x)\, dy\right| \leq L p_{\max} r^{d+2}.$$

However, since $K_r(y,x)$ is a radial kernel,

$$\int \big(\nabla f(x)\big)^T (y-x) K_r(y,x)\, dy = 0$$

and (19) follows by the triangle inequality.

Variance of $U_2(f; G_{n,r})$. We have

$$\mathrm{Var}\Big(U_2(f; G_{n,r})\Big) = \sum_{i \in (n)^3} \sum_{j \in (n)^3} \mathrm{Cov}\Big(D_{i_1} f(x_{i_2}) \cdot D_{i_1} f(x_{i_3}), D_{i_4} f(x_{i_5}) \cdot D_{i_4} f(x_{i_5})\Big)$$

Let $I = i \cup j$ be the joint index set over i and j. Our bound on the covariance term in the preceding display will be a function of the cardinality $|I|$. For example, when $|I| = 6$, by the independence properties of X the covariance is equal to 0. Otherwise if $|I| = 3, 4$ or 5, we have

$$\begin{aligned}
&\left|\mathrm{Cov}\Big(D_{i_1} f(x_{i_2}) \cdot D_{i_1} f(x_{i_3}), D_{i_4} f(x_{i_5}) \cdot D_{i_4} f(x_{i_5})\Big)\right| \\
&\quad \leq \left|\mathbb{E}\Big[D_{i_1} f(x_{i_2}) \cdot D_{i_1} f(x_{i_3}) \cdot D_{i_4} f(x_{i_5}) \cdot D_{i_4} f(x_{i_5})\Big]\right| + \left(\mathbb{E}\Big[D_{i_1} f(x_{i_2}) \cdot D_{i_1} f(x_{i_3})\Big]\right)^2 \\
&\quad \leq \left|\mathbb{E}\Big[D_{i_1} f(x_{i_2}) \cdot D_{i_1} f(x_{i_3}) \cdot D_{i_4} f(x_{i_5}) \cdot D_{i_4} f(x_{i_5})\Big]\right| + p_{\max}^4 L^4 r^{4(d+2)} \\
&\quad \leq L^4 r^4\; \mathbb{P}\Big(G[X_I] \text{ is connected}\Big) + 4 p_{\max}^4 L^4 r^{4(d+2)} \\
&\quad \leq L^4 r^4 p_{\max}^{(|I|-1)} (\nu_d r)^{d(|I|-1)} (|I| - 1)! + 4 p_{\max}^4 L^4 r^{4(d+2)} \\
&\quad \leq L^4 p_{\max}^4 \Big(r^{d(|I|-1)} + r^{4(d+2)}\Big)
\end{aligned}$$

where $c_1 := 24\nu_d^{4d} + 4$.

Our bound depends on the indices I only through $|I|$, and we reiterate that the covariance is non-zero only when $|I| = 3, 4$ or 5. Using our previous expression on the variance of the U-statistic, we see

$$\begin{aligned}\mathrm{Var}\Big(U_2(f; G_{n,r})\Big) &\leq c_1 L^4 p_{\max}^4 \sum_{|I|=3}^{5} n^{|I|}\Big(r^{4+d(|I|-1)} + r^{4(d+2)}\Big) \\ &\leq 3c_1 L^4 p_{\max}^4 n^5 r^{4+4d}\end{aligned}$$

where the latter inequality follows since $n^{-1/d} \leq r \leq 1$.

Concentration. Now we simply apply Chebyshev's inequality to deduce

$$\mathbb{P}\Big(|U_2(f; G_{n,r}) - \mathbb{E}(U_2(f; G_{n,r}))| \geq \sqrt{3c_1} L^2 p_{\max}^2 n^3 r^{2(2+d)}\Big) \leq \frac{1}{nr^4}.$$

Therefore by the triangle inequality,

$$|U_2(f; G_{n,r})| \leq (\sqrt{3c_1} + 2) p_{\max}^2 L^2 n^3 r^{2(2+d)}$$

with probability at least $1 - (nr^4)^{-1}$. This completes the proof of Lemma 6, upon proper choice of the constant $c = \sqrt{3c_1} + 2$.

4 Additional Theory

4.1 Concentration

The maximum eigenvalue of a neighborhood graph is no greater than two times its degree, which can be upper bounded with high probability using, for example, Hoeffding's Inequality.

Lemma 7. *Suppose the density p is bounded away from zero and infinity over its support*

$$0 < p_{\min} \leq p(x) \leq p_{\max} < \infty, \quad \textit{for all } x \in [0,1]^d.$$

Then the maximum degree of $G_{n,r}$ can be upper bounded

$$\deg_{\max}(G_{n,r}) \leq 2p_{\max}\nu_d n r^d$$

with probability at least $1 - n\exp\Big\{-p_{\min}\nu_d n r^d\Big\}$.

Similarly, Hoeffding's inequality can be used to upper (and lower) bound the maximum (and minimum) number of points in any cube $Q(\bar{z})$.

Lemma 8. *Suppose the density p is bounded away from zero and infinity over its support*

$$0 < p_{\min} \leq p(x) \leq p_{\max} < \infty, \quad \textit{for all } x \in [0,1]^d.$$

Then the maximum and minimum number of samples in any grid cell can be bounded,

$$Q_{\min} \geq \frac{p_{\min}}{2(1+\sqrt{d})^d} n r^d$$

and

$$Q_{\max} \leq \frac{2p_{\max}}{(1+\sqrt{d})^d} n r^d$$

with probability at least $1 - \frac{2(1+\sqrt{d})^d}{r^d}\exp\Big\{-\big(p_{\min}(1+\sqrt{d})^d/8\big) n r^d\Big\}$.

4.2 Graph Comparison

For an integer $\mathcal{M} \geq 1$, we introduce the (Alden product) graph. Let $G = (Z, E)$ with $|Z| = \{z_1, \ldots, z_{\mathcal{M}}\}$, and let $V_1, \ldots, V_{\mathcal{M}}$ be sets of size $N_1, \ldots, N_{\mathcal{M}}$. The (Alden product) graph H is defined as

$$H = \Big(V_H, E_H\Big)$$

where $V_H = \bigcup_{m=1}^{\mathcal{M}} \bigcup_{j=1}^{N_m} (z_m, x_{m,j})$ and $(z_\ell, x_{\ell,j}) \sim (z_m, x_{m,j})$ in H if (a) $z_\ell = z_m$ or (b) $z_\ell \sim z_m$ in G. Let

$$N_{\min} := \min_{m \in [\mathcal{M}]} N_m, \quad N_{\max} := \max_{m \in [\mathcal{M}]} N_m$$

and

$$\deg_{\min}(G) := \min_{m \in [\mathcal{M}]} \deg(z_m; G), \quad \deg_{\max}(G) := \max_{m \in [\mathcal{M}]} \deg(z_m; G)$$

Lemma 9. *Assume that G is a non-empty graph, and that for each $m \in [\mathcal{M}]$ the set V_m is non-empty set.*

1. *If $k \in [\mathcal{M}]$,*

$$\frac{1}{2} \frac{N_{\min}^2}{N_{\max}} \lambda_k(G) \leq \lambda_k(H) \leq \frac{N_{\max}^2}{N_{\min}} \lambda_k(G)$$

2. *Otherwise if $k = \mathcal{M} + 1, \ldots, \sum_{m=1}^{\mathcal{M}} N_m$,*

$$\frac{N_{\min}^2}{N_{\max}} \deg_{\min}(G) \leq \lambda_k(H) \leq 2 \frac{N_{\max}^2}{N_{\min}} \deg_{\max}(G)$$

Proof. To prove Lemma 9, we will

1. Exhibit an orthonormal basis of $L^2(V_H)$, then
2. Compute upper and lower bounds on the Rayleigh quotient, then
3. Apply the Courant-Fischer Theorem, and then
4. Put together the pieces.

Step 1: Orthonormal basis of $L^2(V_H)$. Let $v_1, \ldots, v_{\mathcal{M}}$ be eigenvectors of L_G, with corresponding eigenvalues $\lambda_1(G) \leq \cdots \leq \lambda_{\mathcal{M}}(G)$, Then for $k = 1, \ldots, \mathcal{M}$, let

$$v_k^{(H)}\big((z_m, x_{m,j})\big) = \frac{1}{\sqrt{N_m}} v_k(z_m)$$

Now, for each $m = 1, \ldots, \mathcal{M}$, there is an $(N_m - 1)$-dimensional subspace of $L^2(V_m)$ orthogonal to the constant function. Let $u_{1,m}, \ldots, u_{N_m - 1,m}$ be an orthonormal sequence spanning this subspace, and further let

$$u_{\ell,m}^{(H)}\big((z_{m'}, x_{m',j})\big) = \begin{cases} u_{\ell,m}(x_{m,j}), & \text{if } m = m' \\ 0, & \text{otherwise.} \end{cases}$$

for $m \in [\mathcal{M}]$ and $\ell \in [N_m - 1]$.

We now verify that $\{v_k^{(H)} : k \in [\mathcal{M}]\} \cup \{u_{\ell,m}^{(H)} : m \in [\mathcal{M}], \ell \in [N_m - 1]\}$ is an orthonormal basis of $L^2(V_H)$. First,

$$\begin{aligned}
\langle v_k^{(H)}, v_\ell^{(H)} \rangle_{L^2(V_H)} &= \sum_{m=1}^{\mathcal{M}} \sum_{j=1}^{N_m} v_k^{(H)}\big((z_m, x_{m,j})\big) v_\ell^{(H)}\big((z_m, x_{m,j})\big) \\
&= \sum_{m=1}^{\mathcal{M}} \sum_{j=1}^{N_m} \frac{v_k(z_m) v_\ell(z_m)}{N_m} \\
&= \sum_{m=1}^{\mathcal{M}} v_k(z_m) v_\ell(z_m) \\
&= \mathbf{1}\{k = \ell\}
\end{aligned}$$

Next,

$$\begin{aligned}
\langle u_{\ell,m}^{(H)}, u_{\ell',m'}^{(H)} \rangle_{L^2(V_H)} &= \sum_{o=1}^{\mathcal{M}} \sum_{j=1}^{N_o} u_{\ell,m}(x_{o,j}) u_{\ell',m'}(x_{o,j}) \mathbf{1}\{m = o\} \mathbf{1}\{m' = o\} \\
&= \sum_{j=1}^{N_m} u_{\ell,m}(x_{m,j}) u_{\ell',m}(x_{m,j}) \mathbf{1}\{m = m'\} \\
&= \mathbf{1}\{m = m'\} \mathbf{1}\{\ell = \ell'\}
\end{aligned}$$

Finally,

$$\begin{aligned}
\langle v_k^{(H)}, u_{\ell,m}^{(H)} \rangle_{L^2(V_H)} &= \sum_{o=1}^{\mathcal{M}} \sum_{j=1}^{N_o} \frac{v_k(z_o)}{\sqrt{N_o}} u_{\ell,m}(x_{o,j}) \mathbf{1}\{m = o\} \\
&= \frac{v_k(z_m)}{\sqrt{N_m}} \sum_{j=1}^{N_o} u_{\ell,m}(x_{m,j}) \\
&= 0,
\end{aligned}$$

where the last equality follows since $u_{\ell,m}$ is orthogonal to the constant function in $L^2(V_m)$. Since $\dim L^2(V_H) = \sum_{m=1}^{\mathcal{M}} N_m$ and we have exhibited a total of $\sum_{m=1}^{\mathcal{M}} N_m$ orthonormal functions in $L^2(V_H)$, they form a basis.

Step 2: Bounding the Rayleigh quotient. Let $N : L^2(V_H) \to L^2(V_H)$ be the following diagonal operator,

$$\Big(Nv\Big)\Big((z_m, x_{m,j})\Big) = N_m v\big((z_m, x_{m,j})\big)$$

and let $\widetilde{L}_H = N^{1/2} L_H N^{1/2}$. We will analyze the Rayleigh quotient of the operator $\widetilde{L}_H$ with respect to the basis $\{v_k^{(H)}, u_{\ell,m}^{(H)}\}$.

We begin by establishing the following upper and lower bounds,

$$N_{\min}^2 \lambda_k(G) \leq \langle \widetilde{L}_H v_k^{(H)}, v_k^{(H)} \rangle_{L^2(V_H)} \leq N_{\max}^2 \lambda_k(G). \tag{20}$$

Seeing as $N^{1/2}$ is self-adjoint,

$$\begin{aligned}
\left\langle \widetilde{L}_H v_k^{(H)}, v_k^{(H)} \right\rangle_{L^2(V_H)} &= \left\langle L_H N^{1/2} v_k^{(H)}, N^{1/2} v_k^{(H)} \right\rangle_{L^2(V_H)} \\
&= \frac{1}{2} \sum_{m=1}^{\mathcal{M}} \sum_{o=1}^{\mathcal{M}} \sum_{i=1}^{N_m} \sum_{j=1}^{N_o} \Big(v_k(z_m) - v_k(z_o) \Big)^2 \mathbf{1}\{z_m \sim z_o \text{ in } G\} \\
&= \frac{1}{2} \sum_{m=1}^{\mathcal{M}} \sum_{o=1}^{\mathcal{M}} N_m N_o \Big(v_k(z_m) - v_k(z_o) \Big)^2 \mathbf{1}\{z_m \sim z_o \text{ in } G\} \\
&\leq N_{\max}^2 \lambda_k(G),
\end{aligned}$$

and by reversing the last inequality we have

$$\langle \widetilde{L}_H v_k^{(H)}, v_k^{(H)} \rangle_{L^2(V_H)} \geq N_{\min}^2 \lambda_k(G).$$

Next we show that for all $m \in [\mathcal{M}], \ell \in [N_m - 1]$,

$$N_{\min}^2 \Big(\deg_{\min}(G) + 1 \Big) \leq \langle \widetilde{L}_H u_{\ell,m}^{(H)}, u_{\ell,m}^{(H)} \rangle_{L^2(V_H)} \leq N_{\max}^2 \Big(\deg_{\max}(G) + 1 \Big) \tag{21}$$

Expanding the inner product as double sum, we have

$$\begin{aligned}
&\left\langle \widetilde{L}_H u_{\ell,m}^{(H)}, u_{\ell,m}^{(H)} \right\rangle_{L^2(V_H)} \qquad (22) \\
&= \frac{1}{2} \sum_{m'=1}^{\mathcal{M}} \sum_{o=1}^{\mathcal{M}} \sum_{i=1}^{N_{m'}} \sum_{j=1}^{N_o} \Big(\sqrt{N_{m'}} u_{\ell,m}(x_{m,i}) \mathbf{1}\{m' = m\} - \sqrt{N_o} u_{\ell,m}(x_{m,i}) \mathbf{1}\{o = m\} \Big)^2 \mathbf{1}\{m' = o \text{ or } z_{m'} \sim z_o \text{ in } G\} \\
&= \frac{N_m}{2} \Bigg(\underbrace{\sum_{i,j=1}^{N_m} \Big(u_{\ell,m}(x_{m,i}) - u_{\ell,m}(x_{m,j}) \Big)^2}_{:= S_1} + \underbrace{2 \sum_{o=1}^{\mathcal{M}} \sum_{i=1}^{N_m} \sum_{j=1}^{N_o} \Big(u_{\ell,m}(x_{m,i}) \Big)^2 \mathbf{1}\{z_m \sim z_o \text{ in } G\}}_{:= S_2} \Bigg) \qquad (23)
\end{aligned}$$

S_1 is the sum over all term where $m' = m = o$, and S_2 is the sum over all terms where either $m' = m \neq o$ or $m' \neq m = o$. Then, since by definition $u_{\ell,m}$ is orthonormal to the constant function and has unit norm in $L^2(V_m)$, we have

$$2N_{\min} \leq S_1 = \sum_{i,j=1}^{N_m} \Big(\big(u_{\ell,m}(x_{m,i}) \big)^2 + \big(u_{\ell,m}(x_{m,j}) \big)^2 - u_{\ell,m}(x_{m,i}) u_{\ell,m}(x_{m,j}) \Big) = 2N_m \leq 2N_{\max}$$

and

$$N_{\min} \deg_{\min}(G) \leq S_2 = \left(\sum_{i=1}^{N_m} \Big(u_{\ell,m}(x_{m,i}) \Big)^2 \right) \left(\sum_{o=1}^{\mathcal{M}} N_o \right) \leq N_{\max} \deg_{\max}(G)$$

These bounds along with (23) imply (21),

Step 3: Courant-Fischer Theorem. Now we translate the upper and lower bounds we established in Step 2 into upper and lower bounds on $\widetilde{\lambda}_k(H) = \lambda_k(\widetilde{L}_H)$. Naturally, we do this by means of the Courant-Fischer Theorem, which implies

$$\begin{aligned}
\widetilde{\lambda}_k(H) &= \min_{V} \left\{ \max_{v \in V} \langle \widetilde{L}_H v, v \rangle \right\} \\
\widetilde{\lambda}_k(H) &= \max_{V} \left\{ \min_{v \in V^\perp} \langle \widetilde{L}_H v, v \rangle \right\}
\end{aligned}$$

where the outer minimum (maximum) are taken over all subspaces V of dimension k. We now construct subspaces V_k for $k = 1, \ldots, \sum_{m=1}^{\mathcal{M}} N_m$. If $k \in [\mathcal{M}]$, let $V_k = \mathrm{span}\{v_1^{(H)}, \ldots, v_k^{(H)}\}$; otherwise if $k > \mathcal{M}$, let $V_k = V_{\mathcal{M}} \cup U_{k-\mathcal{M}}$, where $U_{k-\mathcal{M}}$ is the span of $(k-\mathcal{M})$-functions $u_{\ell,m}$ chosen arbitrarily. The following facts are immediate consequences of Courant-Fischer and our derivations in Step 2:

(a) For each $k \in [\mathcal{M}]$,
$$\widetilde{\lambda}_k(H) \le \max_{v \in V_k} \left\langle \widetilde{L}_H v, v \right\rangle \le \lambda_k(G) N_{\max}^2$$

(b) For each $k > \mathcal{M}$,
$$\widetilde{\lambda}_k(H) \le \max_{v \in V_k} \left\langle \widetilde{L}_H v, v \right\rangle \le \max\left\{ \lambda_{\mathcal{M}}(G) N_{\max}^2, N_{\max}^2 \left(\deg_{\max}(G) + 1 \right) \right\} \le 2 \deg_{\max}(G) N_{\max}^2$$

(c) For each $k \in [\mathcal{M}]$,
$$\widetilde{\lambda}_k(H) \ge \min_{v \in V_k^{\perp}} \left\langle \widetilde{L}_H v, v \right\rangle \ge \min\left\{ \lambda_k(G) N_{\min}^2, N_{\min}^2 \cdot \left(\deg_{\min}(G) + 1 \right) \right\} \ge \frac{1}{2} \lambda_k(G) N_{\min}^2$$

(d) For each $k > \mathcal{M}$,
$$\widetilde{\lambda}_k(H) \ge \min_{v \in V_k^{\perp}} \left\langle \widetilde{L}_H v, v \right\rangle \ge N_{\min}^2 \left(\deg_{\min}(G) + 1 \right)$$

Step 4: Completing the Proof. Let v be an eigenvector of $\widetilde{L}_H$, so that
$$\widetilde{L}_H v = \lambda v.$$

Then,
$$w = \frac{N^{1/2} v}{\|N^{1/2} v\|_{L^2(V_H)}}$$
is a solution to the generalized eigenvalue equation
$$L_H w = \lambda N^{-1} w.$$

(To see this, simply left multiple both sides of the generalized eigenvalue equation by $N^{1/2}$. Also note that N^{-1} is positive semi-definite since by assumption $N_m > 0$ for all m, and thus invertible.) Therefore,
$$\frac{\lambda}{N_{\max}} \le \langle L_H w, w \rangle_{L^2(V_H)} \le \frac{\lambda}{N_{\min}}$$
and using Courant-Fischer as in Step 3, we have that for each $k = 1, \ldots, \sum_{m=1}^{\mathcal{M}} N_m$,
$$\frac{\widetilde{\lambda}_k(H)}{N_{\max}} \le \lambda_k(H) \le \frac{\widetilde{\lambda}_k(H)}{N_{\min}}$$

Along with the bounds of Step 3, this establishes the claim of Lemma 9.

$\square$