Notes for Week 11/16/19 - 11/23/19

Alden Green

November 23, 2019

In this week's notes, I pick up the thread of the "11.8.19.NotesForWeek" document. There, we obtained that the critical radius in the two-sample density testing problem over a graph G was a function of

$$\Pi_{\max} = \max_{i=1,\ldots,n} \left\{ \sum_{k=1}^{\kappa} v_{k,i}^2 \right\}$$

where $L = VSV^T$ and $V = (v_1, \ldots, v_n)$ is an orthonormal matrix, S is a diagonal matrix with increasing entries $0 = s_1 \leq s_2 \leq \ldots \leq s_n$. In particular, we showed that if $\Pi_{\max}$ could be upper bounded by a constant less than 1, then the critical radius in the two-sample density testing problem over a graph G was the same as in the one-sample regression testing problem. Our goal was therefore to upper bound $\Pi_{\max} \leq \frac{1}{2}$ for an appropriate choice of κ. We showed that such an upper bound holds when $G = G_{dn}$ the d-dimensional grid, and $\kappa \leq \frac{n}{4}$.

1 Bounding $\Pi_{\max}$ for neighborhood graphs in 1-dimension.

Now, suppose we additionally have a graph $\widetilde{G}$ with Laplacian matrix $\widetilde{L}$ such that

$$(1-\delta)x^T L x \leq x^T \widetilde{L} x \leq (1+\delta) x^T L x, \quad \text{for some } 0 \leq \delta \leq 1 \text{ and all } x \in \mathbb{R}^n. \tag{1}$$

Write the spectral decomposition $\widetilde{L} = U \Lambda U^T$, and suppose that the eigenvectors U satisfy

$$u_{l,i}^2 \leq \frac{2}{n} \quad \text{for all } l, i = 1, \ldots, n. \tag{2}$$

Lemma 1. *Let $\widetilde{G} = G_{1n}$ be the 1-dimensional grid graph (i.e. the chain), and assume G satisfies (1). Then we have the following bound on incoherence of the eigenvectors $v_{k,i}$: for any $R \gtrsim n\delta^{1/2}(s_{\max}^{1/2} \vee 1)$,*

$$v_{k,i}^2 \lesssim \frac{1}{n}\left(R + \frac{n^4 \delta^2 s_{\max}^2}{R^3}\right) \quad \textit{for any } i = 1, \ldots, n$$

for any $k \leq n/8$.

Before proving Lemma 1, we demonstrate its consequence. To state things concisely, we use tilde-order notation $\widetilde{O}$ and $\widetilde{\Theta}$ to suppress $\log(n)$ terms. Assume that

$$G = G_{n,r},\ \widetilde{G} = G_{1n} \text{ satisfy (1) with } \delta = \widetilde{O}(n^{-1}), \quad \text{the maximum degree } d_{\max}(G) = \widetilde{O}(1) \tag{3}$$

Using the upper bound on maximum eigenvalue of a Laplacian $s_{\max} \leq 2d_{\max}$), by Lemma 1 the eigenvectors v of G satisfy

$$v_{k,i}^2 = \widetilde{O}\left(\frac{1}{n}\left(R + \frac{n}{R^3}\right)\right)$$

for any $k \leq n/8$ and $R = \widetilde{\Theta}(n^{1/2} s_{\max}^{1/2}) = \widetilde{\Theta}(n^{1/2})$. Taking $R = \widetilde{\Theta}(n^{1/2})$, we have that

$$\max_{i=1,\ldots,n} v_{k,i}^2 = \widetilde{O}\left(n^{-1/2}\right) \tag{4}$$

for all $k \leq n/8$. This is quite a bit weaker than the incoherence bound (2) that the eigenvectors of the chain satisfy. Nevertheless, it will suffice for our purposes. When $\kappa \asymp n^{2/5}$, equation (4) implies that

$$\Pi_{\max} = \widetilde{O}(\kappa n^{-1/2}) = o(1).$$

We have previously shown that when $r = \widetilde{O}(n^{-1})$, the graph $G_{n,r}$ satisfies (3) with high probability; therefore, $\Pi_{\max}$ is less than constant order. Looking back at the "11.8.19.NotesForWeek" (specifically equation (8) which gives the critical radius) we see that this is sufficient to ensure that the critical radius in the two-sample density testing and one-sample regression testing problems will be the same when $d = 1$.

2 Technical Results.

2.1 Proof of Lemma 1

The proof of Lemma 1 will proceed according to the following steps.

1. We show that the incoherence $\max_i v_{k,i}^2$ can be upper bounded

$$\max_i v_{k,i}^2 \leq \frac{1}{n}\left(\sum_{j=1}^{n} |\langle v_k, u_j\rangle|\right)^2 \tag{5}$$

   replacing the norm $\|v_k\|_\infty$ by the inner products $\langle v_k, u_j\rangle$.

2. Let

$$I(k,r) = \{j \geq 0 : |j-k| \leq r\}.$$

   Using Davis-Kahan, we establish that the following upper bounds

$$\sum_{j \notin I(k,r)}^{n} (\langle v_k, u_j\rangle)^2 \leq 400 \frac{n^4 \delta^2 s_{\max}^2}{R^4} \tag{6}$$

   hold for each $k \leq n/8$ and any $r \geq 8k\sqrt{\delta}$.

3. We carefully upper bound the L_1 norm present in (5) given the various bounds on L_2 norm we've established in (6).

Step 1: Proof of (5). We can re-express v_k in the basis (u_j) as

$$v_k = \sum_{j=1}^{n} \langle v_k, u_j\rangle u_j,$$

whereupon the bound (5) follows from (2).

Step 2: Proof of (6). Let $\widetilde{L} = L + H$. By Davis Kahan, we have that

$$\sum_{j \notin I(k,r)} (\langle v_k, u_j\rangle)^2 \leq \left(\frac{\|H\|_{\text{op}}}{\min |\lambda_j - s_k| : j \in I(k,r)}\right)^2 \tag{7}$$

To upper bound the numerator, we use (1) to get

$$\|H\|_{op} \leq \delta s_{\max}.$$

To lower bound the denominator, we note

$$\begin{aligned}|\lambda_j - s_k| &\geq |\lambda_j - \lambda_k| - |\lambda_k - s_k| \\ &\geq |\lambda_j - \lambda_k| - \delta\lambda_k. \end{aligned} \tag{8}$$

The eigenvalues of the chain graph are well known to be

$$\lambda_j = 2\left(1 - \cos\left(\frac{j\pi}{n}\right)\right) = 4\sin^2\left(\frac{k\pi}{n}\right) \quad \text{for } k = 0, \ldots, n-1.$$

By Taylor expansion, for $k \leq j \leq n/(2\pi)$ we have

$$\lambda_j - \lambda_k = 2\left(\cos\left(\frac{k\pi}{n}\right) - \cos\left(\frac{j\pi}{n}\right)\right) \geq \frac{(k-j)^2\pi^2}{n^2}. \tag{9}$$

and when $j \geq n/(2\pi)$ we have $\lambda_j - \lambda_k \geq 2(\cos\left(\frac{\pi}{8}\right) - \cos\left(\frac{1}{2}\right) > .09$. Additionally since $\sin(x) \leq x$ for all $x \geq 0$, we have

$$\lambda_k \leq 4\frac{k^2\pi^2}{n^2}. \tag{10}$$

Combining (9),(10), and the lower bound $|k - j| \geq R \geq 8k\sqrt{\delta}$, we have that

$$|\lambda_j - \lambda_k| - \delta\lambda_k \geq .04\frac{(k-j)^2\pi^2}{n^2} - 4\delta\frac{k^2\pi^2}{n^2} \geq .02\frac{(k-j)^2\pi^2}{n^2}. \tag{11}$$

The result follows from (11), (8) and (7).

Step 3: L_1 norm to L_2 norm. Let $a \in \mathbb{R}^n$, and suppose we know that the L_2 norm of a is bounded,

$$\|a\|_2^2 \leq 1. \tag{12}$$

Under no other conditions on a, the upper bound $\|a\|_1 \leq \sqrt{n}$ is the best that can be hoped for (achieved when $a = (n^{-1/2}, \ldots, n^{-1/2})$). However, suppose we also know that for some $B_2^2 \geq B_3^2 \geq \ldots \geq B_n^2$, we have that there exists some $R \in [n]$ such that

$$\sum_{j=r}^{n} a_j^2 \leq B_r^2 \quad \text{for each } r = R, \ldots, n. \tag{13}$$

Clearly, if $B_n^2 \leq \frac{1}{n}$ then a cannot be equal to $(n^{-1/2}, \ldots, n^{-1/2})$. We might hope for more general improvements on the bound $\|a\|_1 \leq \sqrt{n}$ if B_R^2 is quite small once R gets sufficiently large. The following Lemma gives such a result.

Lemma 2. *Suppose $a \in \mathbb{R}^n$ satisfies (12) and (13) for some $1 = B_1^2 = B_2^2 \geq \ldots \geq B_n^2 \geq B_{n+1}^2 = 0$. Assume additionally that there exists an $R \in [n]$ such that*

$$B_r^2 - B_{r+1}^2 \leq \frac{1}{r} \textit{ for each } r = R, R+1, \ldots, n-1. \tag{14}$$

Then,

$$\|a\|_1 \leq \sqrt{R}\sqrt{1 - B_R^2} + \sum_{j=R}^{n} \sqrt{B_j^2 - B_{j+1}^2}. \tag{15}$$

Proof. Set

$$(a_\star)_j^2 = \begin{cases} \frac{1}{R}(1 - B_R^2), \; j \leq R \\ B_j^2 - B_{j+1}^2, j > R \end{cases}$$

so that $\|a_\star\|_1$ is equal to the right hand side of (15). We now prove by contradiction that $a_\star$ maximizes $\|a\|_1$ under the constraints (12) and (13). Suppose this were not true, and let a be a vector such that a satisfies (12) and (13) and $\|a\|_1 > \|a_\star\|$. Then,

- since $\|a\|_1 > \|a_\star\|$ then there exists some index j such that

$$a_j > (a_\star)_j \tag{16}$$

- by (12) $\|a\|_2 \leq 1$. Since $\|a_\star\|_2 = 1$, by (16) there must exist some index k such that

$$(a_\star)_k > a_k. \tag{17}$$

  Choose k to be the largest of all such indices.

- suppose $k < j$ and $j > R$. Since k was chosen to be the largest such index which satisfied (17), clearly $a_i \geq (a_\star)_i$ for all $i > j$, and by (16) $a_j > (a_\star)_j$. But then

$$B_j^2 = \sum_{i=j}^N (a_\star)_i^2 < \sum_{i=j}^N a_i^2$$

  and so a does not satisfy (13). Therefore either $k > j$ or $j \leq R$.

- In either case, we have

$$a_j > (a_\star)_j \geq (a_\star)_k > a_k.$$

  Let $j \leq l < k$ be the largest index for which $a_j > (a_\star)_j$.

  We now construct a vector $\|b\|$ which satisfies the constraints (12) and (13) such that $\|b\|_1 > \|a\|_1$. Once we have shown this, we will have established a contradiction, and the proof of Lemma 2 will be complete. Let $b = (b_i)$ be given as follows:

$$b_i = \begin{cases} a_i, \quad \text{if } i \neq k, l, \\ a_{\star,k}, \quad \text{if } i = k, \\ \sqrt{a_l^2 - (a_{\star,k}^2 - a_k^2)}, \quad \text{if } i = l. \end{cases}$$

  Clearly $\|b\|_2 = \|a\|_2 = 1$ and so b satisfies (12). To see that b satisfies (13) we separate the analysis into cases. When $r \geq l + 1$, we have that

$$\sum_{i=r}^n b_i^2 = \sum_{i=r}^n a_i^2 \leq B_r^2.$$

  When $l < r \leq k$, we have that

$$\sum_{i=r}^n b_i^2 \leq \sum_{i=r}^k a_{\star,i}^2 + \sum_{i=k+1}^n a_i^2 \leq B_r^2 - B_{k+1}^2 + B_{k+1}^2 = B_r^2$$

  Finally, when $R \leq r \leq l$, we have

$$\sum_{i=r}^n b_i^2 \leq a_l^2 - (a_{\star,k}^2 - a_k^2) + (a_{\star,k}^2) + \sum_{i \neq k,l} a_i^2 = \sum_{i=r}^n a_i^2 \leq B_r^2.$$

Therefore b satisfies (13). Finally, we have that

$$\|b\|_1 = \|a\|_1 + (a_{\star,k} - a_k) - (\sqrt{a_l^2 - (a_{\star,k}^2 - a_k^2)} - a_l) > \|a\|_1 + (a_{\star,k} - a_k) + (a_{\star,k} - a_k) - (a_{\star,k} - a_k) = \|a\|_1,$$

so we have established the desired contradiction.

$\square$

Putting the pieces together. We apply Lemma 2 to the vector $a = (a_i)_{i=1}^n$, where

$$a_i = \frac{1}{\sqrt{2}} \left(|\langle v_k, u_{k-i} \rangle| + |\langle v_k, u_{k+i} \rangle| \right).$$

Note the following:

- $\|a\|_2 \leq 1$.
- By (6), the vector a satisfies the constraint (13) with

$$B_r^2 = 400 \frac{n^4 \delta^2 s_{\max}^2}{r^4} \quad \text{when } r \geq 8k\sqrt{\delta}.$$

- Taylor expanding $f(r+1) = (r+1)^{-4}$ around r, we have that

$$B_r^2 - B_{r+1}^2 \leq 3200 \frac{n^4 \delta^2 s_{\max}^2}{r^5} \quad \text{when } r \geq 2.$$

  Therefore, the sequence (B_r) satisfies (14) for all r large enough such that

$$r \geq 2^{5/4} 100^{1/4} n \sqrt{\delta s_{\max}}$$

- Since a and (B_r) satisfy the constraints (12), (13) and (9), we may apply Lemma 2 to a, obtaining that

$$\sum_{j=1}^n |\langle u_j, v_k \rangle| = \sqrt{2} \|a\|_1 \leq \sqrt{2R} + 60\sqrt{2} n^2 \delta s_{\max} \sum_{r=R}^n r^{-5/2}$$

  for any $R \geq 2 \vee k\sqrt{\delta} \vee 2^{5/4} 100^{1/4} n \sqrt{\delta s_{\max}}$. Bounding sum by integral, we have that if $R \geq 2$ then

$$\sum_{r=R}^n r^{-5/2} \leq \int_{R-1}^{n-1} x^{-5/2}\, dx \leq \frac{2^{3/2} 2}{3} R^{-3/2}.$$

  Plugging this into the previous expression, we arrive at

$$\sum_{j=1}^n |\langle u_j, v_k \rangle| \leq \sqrt{R} + 60\sqrt{2} n^2 \delta s_{\max} \frac{2^{3/2} 2}{3R^{3/2}}$$

  for any $R \geq 2 \vee k\sqrt{\delta} \vee 2^{5/4} 100^{1/4} n \sqrt{\delta s_{\max}}$. The desired result then follows from (5).