Meeting Notes for Week 4/3/20 - 4/9/20

Alden Green

April 9, 2020

Let $v_1, \ldots, v_n$ be eigenvectors of a graph G (normalized so that $\|v_k\|_n^2 = 1$), and $Y_1, \ldots, Y_n$ be observations on the vertices of G, with expectation $\mathbb{E}[Y_i] = f_{0,i}$. The Laplacian eigenmaps estimator is $\widehat{f}_{\mathrm{LE}} = \sum_{k=1}^{\kappa} \langle Y, v_k \rangle_n v_k$. The bias of this estimator is

$$\left\| \mathbb{E}[\widehat{f}_{\mathrm{LE}}] - f_0 \right\|_n^2 = \left\| \sum_{k=1}^{\kappa} (\langle f_0, v_k \rangle_n)^2 - f_0 \right\|_n^2 = \sum_{k=\kappa+1}^{n} (\langle f_0, v_k \rangle_n)^2$$

Let $T_{\mathrm{LE}} = \|\widehat{f}_{\mathrm{LE}}\|_n^2$. The expectation of this statistic is

$$\mathbb{E}[T_{\mathrm{LE}}] = \frac{\kappa}{n} + \sum_{k=1}^{\kappa} (\langle v_k, f_0 \rangle_n)^2 = \frac{\kappa}{n} + \|f_0\|_n^2 - \sum_{k=\kappa+1}^{n} (\langle v_k, f_0 \rangle_n)^2$$

Therefore in either the testing or estimation problems, we arrive at the same question: how to upper bound the tail sum $\sum_{k=\kappa+1}^{n} (\langle v_k, f_0 \rangle_n)^2$?

In our current analysis, we use the following upper bound, which is reminiscent of standard calculations involving the truncated series estimator:

$$
\begin{aligned}
\sum_{k=\kappa+1}^{n} (\langle v_k, f_0 \rangle_n)^2 &\leq \frac{1}{\lambda_{\kappa+1}^s} \sum_{k=\kappa+1}^{n} \lambda_k^s (\langle v_k, f_0 \rangle_n)^2 \\
&\leq \frac{1}{\lambda_{\kappa+1}^s} \sum_{k=1}^{n} \lambda_k^s (\langle v_k, f_0 \rangle_n)^2 \\
&= \frac{f_0^T L^s f_0}{\lambda_{\kappa+1}^s n}
\end{aligned}
\tag{1}
$$

When the inner product $\langle v_k, f_0 \rangle$ is taken with respect to a basis v_k fixed a priori, the above calculation is tight: we can verify this by setting $f_0 \propto v_{\kappa+1}$. However when the basis v_k is random, setting $f_0 \propto v_{\kappa+1}$ is no longer permissible, and we do not know whether or not the above calculation is tight. We investigate the matter empirically.

1   Simulations

In all of our simulations, $d = 1$. Each of the figures below examines the approximation error as function of n, for different values of the neighborhood graph radius r, and the smoothness of the function f_0. The regression function $f_0 = \sum_{k=1}^{100} \theta_k \phi_k$ is constructed out of a basis ϕ_k of $H^s([0,1]^d)$, with θ_k chosen so that $|f_0|_{H^s} = 1$, and $|f_0|_{H^q} \gg 1$ for all $q > s$.

We plot four different quantities in each figure. The red line plots the actual approximation error $\sum_{k=\kappa+1}^{n}(\langle v_k, f_0 \rangle_n)^2$. The blue line plots our bound on approximation error, $(f_0^T L^s f_0)(\lambda_{\kappa+1}^s n)$ The orange line plots the best approximation error we could hope for, $\sum_{k=\kappa}^{100} \theta_k^2$. The green line plots κ^{-2s}, the order of magnitude of the best approximation error (so the green and orange lines should track closely.) We usually choose κ based on balancing bias and variance in the estimation context, but when things get interesting below we will see that the story is the largely the same for different values of κ.

1.1 s = 1

When $s = 1$ and $d = 1$, we have shown that

$$\frac{f_0^T L f_0}{\lambda_{\kappa+1} n} \lesssim \kappa^{-2}$$

whenever $(\log(n)/n) \lesssim r \lesssim \kappa^{-1}$. Pictorially, this means that the blue line should be decreasing at least as fast as the orange and green lines. In Figures 1 we plot these lines as function of sample size n on the log-log scale, making two different choices of $r \equiv (\log(n)/n)$ and $r \equiv n^{-1/3}$ (the left and right hand plots, respectively). Some algebra can verify that the latter choice satisfies $r \lesssim \kappa^{-1}$. We see that in both instances, all four lines track fairly closely together.
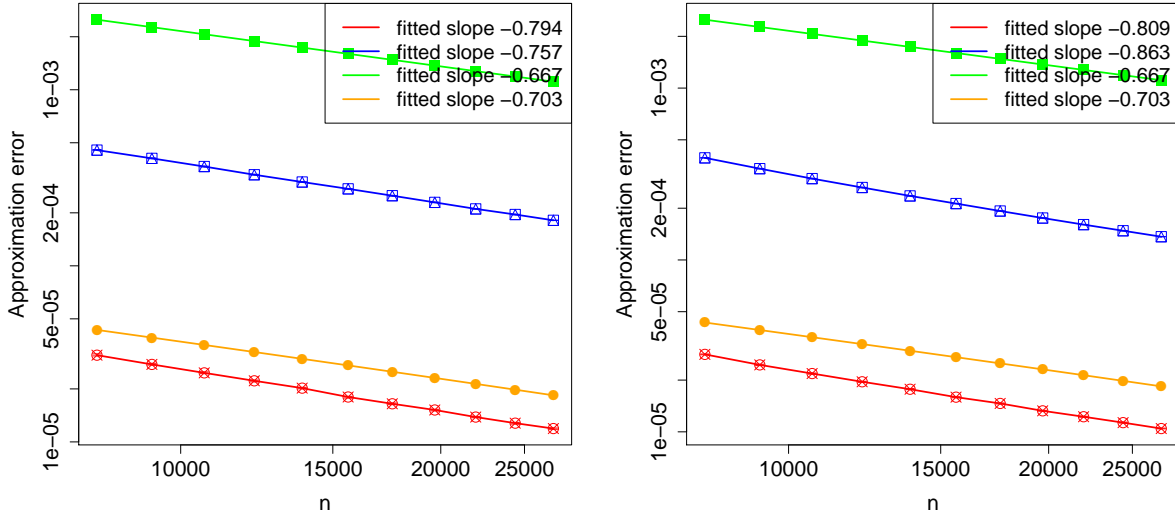

1.2 s = 2

When $s = 2$ and $d = 1$, we have show that

$$\frac{f_0^T L^2 f_0}{\lambda_{\kappa+1}^2 n} \lesssim \kappa^{-4}$$

when $n^{-1/3} \lesssim r \lesssim \kappa^{-1}$. The choice $r \equiv n^{-1/3}$ still satisfies both of these conditions for our choice of κ, however clearly $r \equiv (\log(n)/n)$ violates the lower bound. Pictorially, this means we no longer have theoretical reason to believe that the blue line–and hence, the red line– should decrease at least as fast as the orange and green lines. In Figures 2, we plot each of the four aforementioned lines as a function of sample size on the log-log scale. (The difference between the two rows is that in the second row, we take a larger value

2

of κ; precisely, we multiply the previous value of κ by a constant). In both Figures the conclusion is the same. When $r \equiv n^{-1/3}$ (right hand plot) all four lines still track reasonably close together. However, when $r \equiv (\log(n)/n)$ (left hand plot) the blue line no longer tracks the orange and green lines—in fact, it doesn't even decrease—but the red line still does.
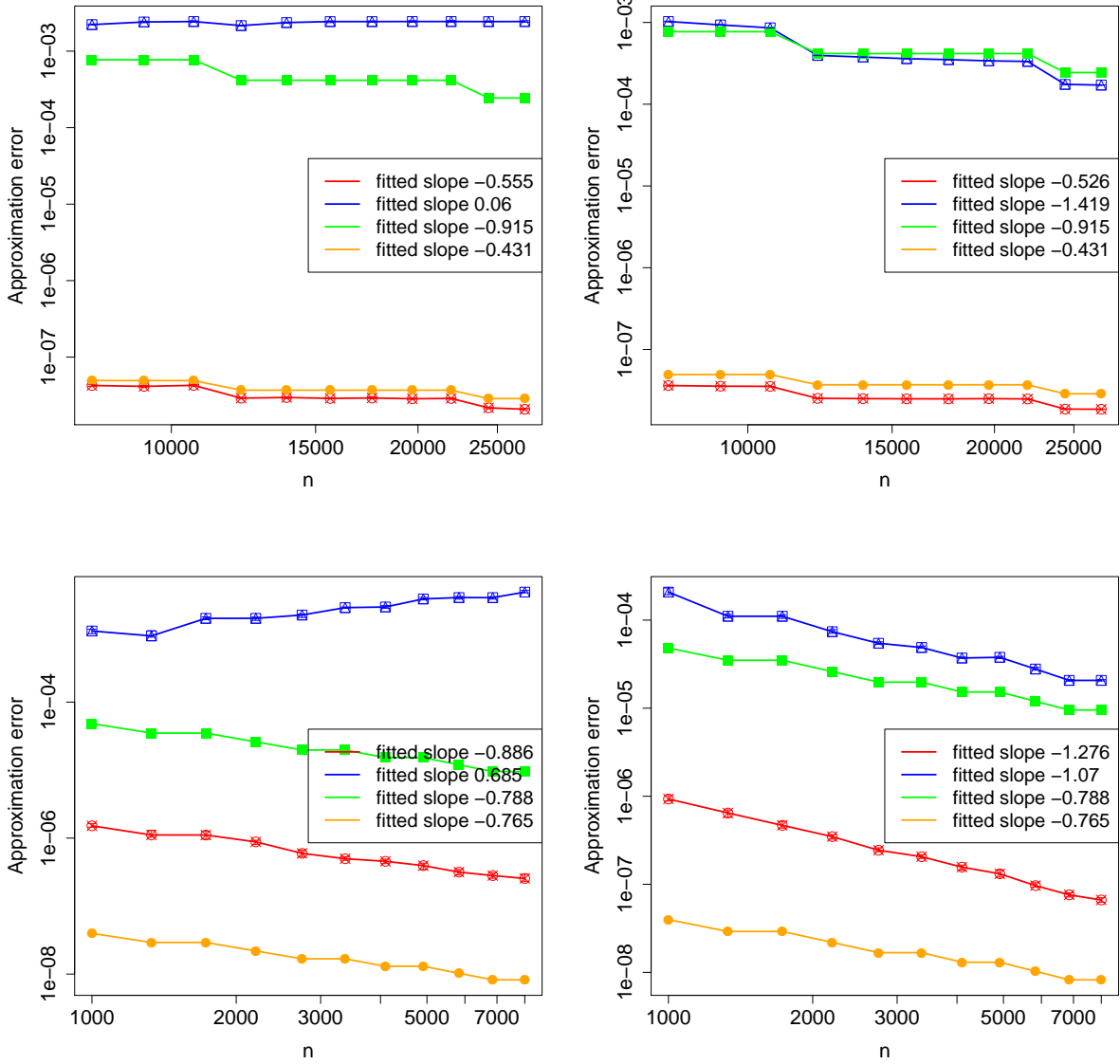
My conclusions are twofold: first, that our bound on $f_0^T L^s f_0 / \lambda_{\kappa+1}^s n$ appears tight, at least in these experiments, and second that the upper bound $f_0^T L^s f_0 / \lambda_{\kappa+1}^s n \geq \sum_{k=\kappa+1}^n (\langle v_k, f_0 \rangle_n)^2$ is not always tight.

2   Simple example showing looseness of our bound.

I will show that for a graph $\bar{G}_{n,r}$ which bears some resemblance to a neighborhood graph, f_0 have non-trivial loading onto many of the "top" eigenvectors, that is, eigenvectors which correspond to the largest eigenvalues in the graph Laplacian.

3

Let $X_1, \ldots, X_n$ be random samples which belong to $[0,1]^d$. Let Z be the grid points spaced at interval width r; for $M = 1/r$,

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

and let $\bar{G}$ be the d-dimensional lattice graph over the grid points X.

Let $\Pi : \{X_1, \ldots, X_n\} \to Z$ assign each sample X_i to its nearest neighbor $z \in Z$. Then $\bar{G}_{n,r} = ([n], E)$ will be the graph with edge set $E \subseteq [n] \times [n]$, with (i,j) connected in $\bar{G}_{n,r}$ only if $(\Pi(X_i), \Pi(X_n))$ belongs to the edge set of $\bar{G}$.

It is an interesting fact—and not hard to show—that the eigenvectors of $\bar{G}_{n,r}$ can be divided into two categories. The first category consists of eigenvectors v_k which are piecewise constant over cubes; that is, if $\Pi(X_i) = \Pi(X_j)$ then $v_{k,i} = v_{k,j}$. The second category consists of delta functions at points in the same cube: precisely, for each i and j such that $\Pi(X_i) = \Pi(X_j)$, the vector $u_k = e_i - e_j$ is an eigenvector of $\bar{G}_{n,r}$. Even if f_0 is smooth, for instance $f_0 \in C^s([0,1]^d)$ or $f \equiv \phi_{\kappa+1}$, the loadings of f onto the latter category of eigenvectors satisfy

$$\langle f_0, u_k \rangle_n^2 = (f_{0,i} - f_{0,j})^2 \approx (\nabla f_0(X_j))^2 \cdot r^2$$

which, while small, is not zero *unless* f_0 itself happens to piecewise constant over cubes.

I don't really know how to use this information. If we were in fact interested in the graph $\bar{G}_{n,r}$, we might try to approximate f_0 by a piecewise constant function $\bar{f}_0$, then proceed to analyze the approximation error of $\bar{f}_0$ using (1). But we aren't interested in $\bar{G}_{n,r}$. Nevertheless, I think the above example does give some useful insight into the empirical results we see in the previous section.