Spectral Graph Theory Lecture 4

Bounding Eigenvalues

Daniel A. Spielman September 14, 2015

Disclaimer

These notes are not necessarily an accurate representation of what happened in class. The notes written before class say what I think I should say. I sometimes edit the notes after class to make them way what I wish I had said.

There may be small mistakes, so I recommend that you check any mathematically precise statement before using it in your own work.

These notes were last revised on September 22, 2015.

4.1 Overview

It is unusual when one can actually explicitly determine the eigenvalues of a graph. Usually one is only able to prove loose bounds on some eigenvalues. In this lecture, I will introduce two important techniques for proving such bounds. The first is the Courant-Fischer Theorem, which provides a more powerful characterization of eigenvalues as solutions to optimization problems than the one we derived before. This theorem is useful for doing things like proving upper bounds on the largest eigenvalue of a matrix.

The more powerful technique we will see allows one to compare one graph with another, and prove things like lower bounds on the largest eigenvalue of a matrix. It often goes by the name "Poincaré Inequalities" (see [DS91, SJ89, GLM99]), although I often use the name "Graphic inequlities", as I see them as providing inequalities between graphs.

4.2 The Courant-Fischer Theorem

I gave a hint of the Courant-Fischer Theorem earlier in the semester. I'll do the rest of it now.

Theorem 4.2.1 (Courant-Fischer Theorem)**.** *Let $\boldsymbol{L}$ be a symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then,*

$$\lambda_k = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=k}} \max_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \max_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T)=n-k+1}} \min_{\boldsymbol{x} \in T} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

For example, consier the case $k = 1$. In this case, S is just the span of $\boldsymbol{\psi}_1$ and T is all of $\mathbb{R}^n$. For general k, the optima will be achieved when S is the span of $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$ and when T is the span of $\boldsymbol{\psi}_k, \ldots, \boldsymbol{\psi}_n$.

Proof. We will just verify the first characterization of λ_k. The other is similar.

First, let's verify that λ_k is achievable. Let S_k be the span of $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$. For every $\boldsymbol{x} \in S_k$, we can write

$$\boldsymbol{x} = \sum_{i=1}^{k} c_i \boldsymbol{\psi}_i,$$

so,

$$\frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_{i=1}^{k} \lambda_i c_i^2}{\sum_{i=1}^{k} c_i^2} \leq \frac{\sum_{i=1}^{k} \lambda_k c_i^2}{\sum_{i=1}^{k} c_i^2} = \lambda_k.$$

To verify that this is in fact the maximum, let T_k be the span of $\boldsymbol{\psi}_k, \ldots, \boldsymbol{\psi}_n$. As T_k has dimension $n - k + 1$, for any S of dimension k the intersection of S with T_k has dimension at least 1. So,

$$\max_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \max_{\boldsymbol{x} \in S \cap T_k} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Any such $\boldsymbol{x}$ may be expressed as

$$\boldsymbol{x} = \sum_{i=k}^{n} c_i \boldsymbol{\psi}_i,$$

and so

$$\frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_{i=k}^{n} \lambda_i c_i^2}{\sum_{i=k}^{n} c_i^2} \geq \frac{\sum_{i=k}^{n} \lambda_k c_i^2}{\sum_{i=k}^{n} c_i^2} = \lambda_k.$$

We conclude that for all subspaces S of dimension k,

$$\max_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \lambda_k.$$

□

4.3 Bounds on λ_2

We were essentially using the Courant-Fischer theorem when we observed that the second-smallest eigenvalue of the Laplacian is given by

$$\lambda_2 = \min_{\boldsymbol{v} : \boldsymbol{v}^T \mathbf{1} = 0} \frac{\boldsymbol{v}^T \boldsymbol{L} \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}}.$$

To see this, consider any vector $\boldsymbol{x}$, and let S be the span of $\boldsymbol{x}$ and $\mathbf{1}$. Let $\boldsymbol{v}$ be a vector in S that is orthogonal to $\mathbf{1}$. It is clear that

$$\boldsymbol{v}^T \boldsymbol{L} \boldsymbol{v} = \boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}.$$

You should check for yourself that

$$\boldsymbol{v}^T \boldsymbol{v} \leq \boldsymbol{x}^T \boldsymbol{x}.$$

I'll make this an exercise.

The type of bounds that we get from the Courant-Fischer theorem are those that follow from this. Every vector $\boldsymbol{v}$ orthogonal to $\mathbf{1}$ provides an upper bound on λ_2:

$$\lambda_2 \leq \frac{\boldsymbol{v}^T L \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}}.$$

When we use a vector $\boldsymbol{v}$ in this way, we call it a *test vector.*

Let's see what a test vector can tell us about λ_2 of a path graph on n vertices. I would like to use the vector that assigns i to vertex i as a test vector, but it is not orthogonal to $\mathbf{1}$. So, we will use the next best thing. Let $\boldsymbol{x}$ be the vector such that $\boldsymbol{x}(i) = (n+1) - 2i$, for $1 \leq i \leq n$. This vector satisfies $\boldsymbol{x} \perp \mathbf{1}$, so

$$\begin{aligned} \lambda_2(P_n) &\leq \frac{\sum_{1 \leq i < n} (x(i) - x(i+1))^2}{\sum_i x(i)^2} \\ &= \frac{\sum_{1 \leq i < n} 2^2}{\sum_i (n+1-2i)^2} \\ &= \frac{4(n-1)}{(n+1)n(n-1)/3} && \text{(clearly, the denominator is } n^3/c \text{ for some } c) \\ &= \frac{12}{n(n+1)}. && (4.1) \end{aligned}$$

We will soon see that this bound is of the right order of magnitude.

The Courant-Fischer theorem is not as helpful when we want to prove lower bounds on λ_2. To prove lower bounds, we need the form with a maximum on the outside, which gives

$$\lambda_2 \geq \max_{S : \dim(S) = n-1} \min_{v \in S} \frac{\boldsymbol{v}^T \boldsymbol{L} \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}}.$$

This is not too helpful, as it is difficult to prove lower bounds on

$$\min_{v \in S} \frac{\boldsymbol{v}^T \boldsymbol{L} \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}}$$

over a space S of large dimension. We need another technique.

4.4 Graphic Inequalities

I begin by recalling an extremely useful piece of notation that is used in the Optimization community. For a symmetric matrix $\boldsymbol{A}$, we write

$$\boldsymbol{A} \succcurlyeq 0$$

if $\boldsymbol{A}$ is *positive semidefinite*. That is, if all of the eigenvalues of $\boldsymbol{A}$ are nonnegative, which is equivalent to

$$\boldsymbol{v}^T \boldsymbol{A} \boldsymbol{v} \geq \boldsymbol{0},$$

for all $\boldsymbol{v}$. We similarly write

$$\boldsymbol{A} \succcurlyeq \boldsymbol{B}$$

if

$$\boldsymbol{A} - \boldsymbol{B} \succcurlyeq \boldsymbol{0}$$

which is equivalent to

$$\boldsymbol{v}^T \boldsymbol{A} \boldsymbol{v} \geq \boldsymbol{v}^T \boldsymbol{B} \boldsymbol{v}$$

for all $\boldsymbol{v}$.

The relation $\preccurlyeq$ is an example of a *partial order*. It applies to some pairs of symmetric matrices, while others are incomparable. But, for all pairs to which it does apply, it acts like an order. For example, we have

$$\boldsymbol{A} \succcurlyeq \boldsymbol{B} \text{ and } \boldsymbol{B} \succcurlyeq \boldsymbol{C} \text{ implies } \boldsymbol{A} \succcurlyeq \boldsymbol{C},$$

and

$$\boldsymbol{A} \succcurlyeq \boldsymbol{B} \text{ implies } \boldsymbol{A} + \boldsymbol{C} \succcurlyeq \boldsymbol{B} + \boldsymbol{C},$$

for symmetric matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$.

I find it convenient to overload this notation by defining it for graphs as well. Thus, I'll write

$$G \succcurlyeq H$$

if $\boldsymbol{L}_G \succcurlyeq \boldsymbol{L}_H$. For example, if $G = (V, E)$ is a graph and $H = (V, F)$ is a subgraph of G, then

$$\boldsymbol{L}_G \succcurlyeq \boldsymbol{L}_H.$$

To see this, recall the Laplacian quadratic form:

$$\boldsymbol{x}^T \boldsymbol{L}_G \boldsymbol{x} = \sum_{(u,v) \in E} w_{u,v} (\boldsymbol{x}(u) - \boldsymbol{x}(v))^2.$$

It is clear that dropping edges can only decrease the value of the quadratic form. The same holds for decreasing the weights of edges.

This notation is most powerful when we consider some multiple of a graph. Thus, I could write

$$G \succcurlyeq c \cdot H,$$

for some $c > 0$. What is $c \cdot H$? It is the same graph as H, but the weight of every edge is multiplied by c.

Using the Courant-Fischer Theorem, we can prove

Lemma 4.4.1. *If G and H are graphs such that*

$$G \succcurlyeq c \cdot H,$$

then

$$\lambda_k(G) \geq c\lambda_k(H),$$

for all k.

Proof. The Courant-Fischer Theorem tells us that

$$\lambda_k(G) = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=k}} \max_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T L_G \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq c \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=k}} \max_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T L_H \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = c\lambda_k(H).$$

□

Corollary 4.4.2. *Let G be a graph and let H be obtained by either adding an edge to G or increasing the weight of an edge in G. Then, for all i*

$$\lambda_i(G) \leq \lambda_i(H).$$

4.5 Approximations of Graphs

An idea that we will use in later lectures is that one graph approximates another if their Laplacian quadratic forms are similar. For example, we will say that H is a c-approximation of G if

$$cH \succcurlyeq G \succcurlyeq H/c.$$

Surprising approximations exist. For example, expander graphs are very sparse approximations of the complete graph. For example, the following is known.

Theorem 4.5.1. *For every $\epsilon > 0$, there exists a $d > 0$ such that for all sufficiently large n there is a d-regular graph G_n that is a $(1+\epsilon)$-approximation of K_n.*

These graphs have many fewer edges than the complete graphs!

In a later lecture we will also prove that every graph can be well-approximated by a sparse graph.

4.6 The Path Inequality

By now you should be wondering, "how do we prove that $G \succcurlyeq c \cdot H$ for some graph G and H?" Not too many ways are known. We'll do it by proving some inequalities of this form for some of the simplest graphs, and then extending them to more general graphs. For example, we will prove

$$(n-1) \cdot P_n \succcurlyeq G_{1,n}, \tag{4.2}$$

where P_n is the path from vertex 1 to vertex n, and $G_{1,n}$ is the graph with just the edge $(1, n)$. All of these edges are unweighted.

The following very simple proof of this inequality was discovered by Sam Daitch.

Lemma 4.6.1.

$$(n-1) \cdot P_n \succcurlyeq G_{1,n}.$$

Proof. We need to show that for every $\boldsymbol{x} \in \mathbb{R}^n$,

$$(n-1)\sum_{i=1}^{n-1}(\boldsymbol{x}(i+1)-\boldsymbol{x}(i))^2 \geq (\boldsymbol{x}(n)-\boldsymbol{x}(1))^2.$$

For $1 \leq i \leq n-1$, set

$$\boldsymbol{\Delta}(i) = \boldsymbol{x}(i+1) - \boldsymbol{x}(i).$$

The inequality we need to prove then becomes

$$(n-1)\sum_{i=1}^{n-1}\boldsymbol{\Delta}(i)^2 \geq \left(\sum_{i=1}^{n-1}\boldsymbol{\Delta}(i)\right)^2.$$

But, this is just the Cauchy-Schwartz inequality. I'll remind you that Cauchy-Schwartz just follows from the fact that the inner product of two vectors is at most the product of their norms:

$$(n-1)\sum_{i=1}^{n-1}\boldsymbol{\Delta}(i)^2 = \|\mathbf{1}_{n-1}\|^2 \|\boldsymbol{\Delta}\|^2 = (\|\mathbf{1}_{n-1}\| \, \|\boldsymbol{\Delta}\|)^2 \geq \left(\mathbf{1}_{n-1}^T \boldsymbol{\Delta}\right)^2 = \left(\sum_{i=1}^{n-1}\boldsymbol{\Delta}(i)\right)^2.$$

□

While I won't cover it in lecture, I will also state the version of this inequality for weighted paths.

Lemma 4.6.2. *Let $w_1, \ldots, w_{n-1}$ be positive. Then*

$$G_{1,n} \preccurlyeq \left(\sum_{i=1}^{n-1}\frac{1}{w_i}\right)\sum_{i=1}^{n-1} w_i G_{i,i+1}.$$

Proof. Let $\boldsymbol{x} \in \mathbb{R}^n$ and set $\boldsymbol{\Delta}(i)$ as in the proof of the previous lemma. Now, set

$$\boldsymbol{\gamma}(i) = \boldsymbol{\Delta}(i)\sqrt{w_i}.$$

Let $\boldsymbol{w}^{-1/2}$ denote the vector for which

$$\boldsymbol{w}^{-1/2}(i) = \frac{1}{\sqrt{w_i}}.$$

Then,

$$\sum_i \boldsymbol{\Delta}(i) = \boldsymbol{\gamma}^T \boldsymbol{w}^{-1/2},$$

$$\left\| \boldsymbol{w}^{-1/2} \right\|^2 = \sum_i \frac{1}{w_i},$$

and

$$\| \boldsymbol{\gamma} \|^2 = \sum_i \boldsymbol{\Delta}(i)^2 w_i.$$

So,

$$\begin{aligned} \boldsymbol{x}^T L_{G_{1,n}} \boldsymbol{x} &= \left(\sum_i \boldsymbol{\Delta}(i) \right)^2 = \left(\boldsymbol{\gamma}^T \boldsymbol{w}^{-1/2} \right)^2 \\ &\leq \left(\| \boldsymbol{\gamma} \| \left\| \boldsymbol{w}^{-1/2} \right\| \right)^2 = \left(\sum_i \frac{1}{w_i} \right) \sum_i \boldsymbol{\Delta}(i)^2 w_i = \left(\sum_i \frac{1}{w_i} \right) \boldsymbol{x}^T \left(\sum_{i=1}^{n-1} w_i L_{G_{i,i+1}} \right) \boldsymbol{x}. \end{aligned}$$

$\square$

4.6.1 Bounding λ_2 of a Path Graph

I'll now demonstrate the power of Lemma 4.6.1 by using it to prove a lower bound on $\lambda_2(P_n)$ that will be very close to the upper bound we obtained from the test vector.

To prove a lower bound on $\lambda_2(P_n)$, we will prove that some multiple of the path is at least the complete graph. To this end, write

$$L_{K_n} = \sum_{i<j} L_{G_{i,j}},$$

and recall that

$$\lambda_2(K_n) = n.$$

For every edge (i,j) in the complete graph, we apply the only inequality available in the path:

$$G_{i,j} \preccurlyeq (j-i) \sum_{k=i}^{j-1} G_{k,k+1} \preccurlyeq (j-i) P_n. \tag{4.3}$$

This inequality says that $G_{i,j}$ is at most $(j-i)$ times the part of the path connecting i to j, and that this part of the path is less than the whole.

Summing inequality (4.3) over all edges $(i,j) \in K_n$ gives

$$K_n = \sum_{i<j} G_{i,j} \preccurlyeq \sum_{i<j} (j-i) P_n.$$

To finish the proof, we compute

$$\sum_{1 \leq i < j \leq n} (j-i) = \sum_{k=1}^{n-1} k(n-k) = n(n+1)(n-1)/6.$$

So,

$$L_{K_n} \preccurlyeq \frac{n(n+1)(n-1)}{6} \cdot L_{P_n}.$$

Applying Lemma 4.4.1, we obtain

$$\frac{6}{(n+1)(n-1)} \leq \lambda_2(P_n).$$

This only differs from the upper bound (4.1) by a factor of 2.

4.7 The Complete Binary Tree

Let's do the same analysis with the complete binary tree.

One way of understanding the complete binary tree of depth $d+1$ is to identify the vertices of the tree with strings over $\{0,1\}$ of length at most d. The root of the tree is the empty string. Every other node has one ancestor, which is obtained by removing the last character of its string, and two children, which are obtained by appending one character to its label.

Alternatively, you can describe it as the graph on $n = 2^{d+1} - 1$ nodes with edges of the form $(i, 2i)$ and $(i, 2i+1)$ for $i < n$. We will name this graph T_d. Pictures of this graph appear below.

Pictorially, these graphs look like this:

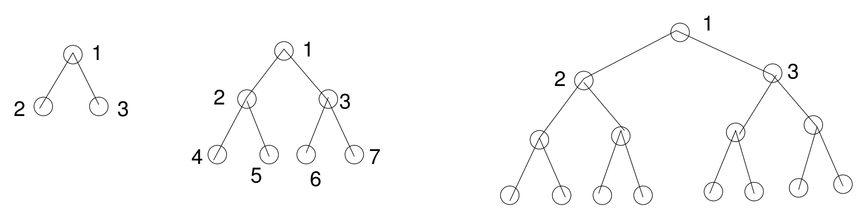

Figure 4.1: T_1, T_2 and T_3. Node 1 is at the top, 2 and 3 are its children. Some other nodes have been labeled as well.

Let's first upper bound $\lambda_2(T_d)$ by constructing a test vector x. Set $x(1) = 0$, $x(2) = 1$, and $x(3) = -1$. Then, for every vertex u that we can reach from node 2 without going through node 1, we set $x(u) = 1$. For all the other nodes, we set $x(u) = -1$.

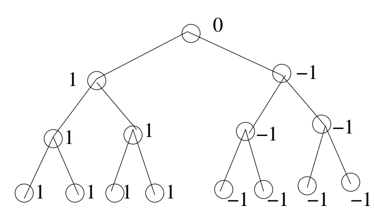

Figure 4.2: The test vector we use to upper bound $\lambda_2(T_3)$.

We then have

$$\lambda_2 \leq \frac{\sum_{(i,j)\in T_d}(x_i - x_j)^2}{\sum_i x_i^2} = \frac{(x_1 - x_2)^2 + (x_1 - x_3)^2}{n-1} = 2/(n-1).$$

We will again prove a lower bound by comparing T_d to the complete graph. For each edge $(i,j) \in K_n$, let $T_d^{i,j}$ denote the unique path in T from i to j. This path will have length at most $2d$. So, we have

$$K_n = \sum_{i<j} G_{i,j} \preccurlyeq \sum_{i<j} (2d) T_d^{i,j} \preccurlyeq \sum_{i<j} (2\log_2 n) T_d = \binom{n}{2}(2\log_2 n) T_d.$$

So, we obtain the bound

$$\binom{n}{2}(2\log_2 n)\lambda_2(T_d) \geq n,$$

which implies

$$\lambda_2(T_d) \geq \frac{1}{(n-1)\log_2 n}.$$

In the next problem set, I will ask you to improve this lower bound to $1/cn$ for some constant c.

4.8 Exercises

1. Let $\boldsymbol{v}$ be a vector so that $\boldsymbol{v}^T\mathbf{1} = 0$. Prove that

$$\|\boldsymbol{v}\|^2 \leq \|\boldsymbol{v} + t\mathbf{1}\|^2,$$

for every real number t.

References

[DS91] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.

[GLM99] S. Guattery, T. Leighton, and G. L. Miller. The path resistance method for bounding the smallest nontrivial eigenvalue of a Laplacian. *Combinatorics, Probability and Computing*, 8:441–460, 1999.

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, July 1989.