Spectral Graph Theory Lecture 1

Introduction

Daniel A. Spielman September 2, 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 3, 2015.

1.1 First Things

1. Please call me "Dan". If such informality makes you uncomfortable, you can try "Professor Dan". If that fails, I will also answer to "Prof. Spielman".
2. If you are going to take this course, please sign up for it on Classes V2. This is the only way you will get emails like "Problem 3 was false, so you don't have to solve it".
3. This class meets this coming Friday, September 4, but not on Labor Day, which is Monday, September 7.

1.2 Introduction

I have three objectives in this lecture: to tell you what the course is about, to help you decide if this course is right for you, and to tell you how the course will work.

As the title suggests, this course is about the eigenvalues and eigenvectors of matrices associated with graphs, and their applications. I will never forget my amazement at learning that combinatorial properties of graphs could be revealed by an examination of the eigenvalues and eigenvectors of their associated matrices. I hope to both convey my amazement to you and to make it feel like common sense. I'm now shocked when any important property of a graph is not revealed by its eigenvalues and eigenvectors.

You can get an idea for the topics that I'll cover during the semester from looking at the lecture notes from the 2012 version of this course. This year's version will be pretty similiar. If you look

at the web page for the course, you will see a list of the lectures I plan to give. To help you decide whether or not to take the course, I'll make this lecture fairly technical. If you want to see something closer to a sales pitch, I recommend looking at the notes from my first lecture in 2009.

This class will fundamentally be a math class, but my emphasis is on material that I find useful. I'll present a lot of theorems, a few algorithms, and a bunch of open problems.

1.3 Mechanics

There is no book for this course. I do, however, plan to produce notes for every lecture. **You should read the lecture notes.** They will often contain material that I did not have time to cover in class. They will sometimes contain extra expositions of elementary topics. I will try to make the notes available before lecture. Some students will want to print them out for reference during lecture.

Given that I am providing lecture notes, you might not need to take notes during lecture. I, however, take notes during every lecture that I attend. It helps me pay attention and remember what is going on. But, there are many different learning styles. You may prefer to just listen.

If you would like a book that covers some of the material from the course, I suggest one of

"Algebraic Graph Theory" by Chris Godsil and Gordon Royle,

"Spectral Graph Theory" by Fan Chung, or

"Algebraic Combinatorics" by Chris Godsil.

I expect to produce around 5 or 6 problem sets during the semester. Some of the problems I assign in these will be very hard. You will be allowed to work on them in small groups.

For some lectures, such as today's, I have assigned a number of "exercises" at the end of the lecture notes. You should solve these on your own, as soon after lecture as possible. You should not hand them in. They are just to help you practice the material. Today's exercises are a review of fundamental linear algebra. I will put the solutions to some of them on Classes V2.

There will be no tests or exams.

1.3.1 This is a graduate course

As some undergrads are thinking about taking this course, I thought I should explain the main differences between an undergraduate course and a graduate course, and the differences in outlook between undergrads and graduate students.

Graduate school is essentially pass/fail. Graduate students either write a thesis and graduate, or they do not. Their grades in courses do not matter very much. Most are here because they think they might learn something in this course that they will find useful in their careers. This means that some of them will work very hard.

Graduate students are also occasionally occupied with other responsibilities, like teaching and research. For this reason, I will give students at least two weeks to complete the problems I assign. However, I recommend that you solve the easier problems immediately.

Graduate students routinely take courses for which they do not have all the prerequisite knowledge. I assume that they can learn anything elementary as needed. Wikipedia makes this much easier than it used to be.

Finally, graduate courses are not as "user friendly" as undergraduate courses. I make no guarantees about what will happen in this course. I may assign more or fewer problem sets than I have announced. I may completely change the topics that I decide to cover. You have been warned.

1.4 Background: Graphs

First, we recall that a graph $G = (V, E)$ is specified by its vertex[1] set, V, and edge set E. In an undirected graph, the edge set is a set of unordered pairs of vertices. Unless otherwise specified, all graphs will be undirected, simple (having no loops or multiple edges) and finite. We will sometimes assign weights to edges. These will usually be real numbers. If no weights have been specified, we view all edges as having weight 1. This is an arbitrary choice, and we should remember that it has an impact.

Graphs are typically used to model connections or relations between things, where "things" are vertices. However, I often prefer to think of the edges in a graph as being more important than the vertices. In this case, I may just specify an edge set E, and ignore the ambient vertex set.

Common "natural" examples of graphs are:

- Friendship graphs: people are vertices, edges exist between pairs of people who are friends (assuming the relation is symmetric).
- Network graphs: devices, routers and computers are vertices, edges exist between pairs that are connected.
- Circuit graphs: electronic components, such as transistors, are vertices: edges exist between pairs connected by wires.
- Protein-Protein Interaction graphs: proteins are vertices. Edges exist between pairs that interact. These should really have weights indicating the strength and nature of interaction. Most other graphs should to.

It is much easier to study abstract, mathematically defined graphs. For example,

- The path on n vertices. The vertices are $\{1, \ldots n\}$. The edges are $(i, i+1)$ for $1 \leq i < n$.
- The ring on n vertices. The vertices are $\{1, \ldots n\}$. The edges are all those in the path, plus the edge $(1, n)$.

[1] I will use the words "vertex" and "node" interchangeably. Sorry about that.

- The hypercube on 2^k vertices. The vertices are elements of $\{0,1\}^k$. Edges exist between vertices that differ in only one coordinate.

1.5 Matrices for Graphs

The naive view of a matrix is that it is essentially a spreadsheet—a table we use to organize numbers. This is like saying that a car is an enclosed metal chair with wheels. It says nothing about what it does!

I will use matrices to do two things. First, I will view a matrix $\boldsymbol{A}$ as providing an function that maps a vector $\boldsymbol{x}$ to the vector $\boldsymbol{A}\boldsymbol{x}$. That is, I view $\boldsymbol{A}$ as an operator. Second, I view a matrix $\boldsymbol{A}$ as providing a function that maps a vector $\boldsymbol{x}$ to a number $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$. That is, I use $\boldsymbol{A}$ to define a quadratic form.

1.5.1 A spreadsheet

We will usually write V for the set of vertices of a graph, and let n denote the number of vertices. There are times that we will need to order the vertices and assign numbers to them. In this case, they will usually be $\{1,\ldots,n\}$. For example, if we wish to draw a matrix as a table, then we need to decide which vertex corresponds to which row and column.

The most natural matrix to associate with a graph G is its adjacency matrix, $\boldsymbol{A}_G$, whose entries $\boldsymbol{A}_G(u,v)$ are given by

$$\boldsymbol{A}_G(u,v) = \begin{cases} 1 & \text{if } (u,v) \in E \\ 0 & \text{otherwise.} \end{cases}$$

It is important to observe that I index the rows and columns of the matrix by vertices, rather than by number. Almost every statement that we make in this class will remain true under renaming of vertices. The first row of a matrix has no special importance. To understand this better see the exercises at the end of the lecture.

While the adjacency matrix is the most natural matrix to associate with a graph, I also find it the least useful. Eigenvalues and eigenvectors are most meaningful when used to understand a natural operator or a natural quadratic form. The adjacency matrix provides neither.

1.5.2 An operator

The most natural operator associated with a graph G is probably its diffusion operator. This operator describes the diffusion of stuff among the vertices of a graph. Imagine a process in which each vertex can contain some amount of stuff (such as a gas). At each time step, the stuff at a vertex will be uniformly distributed to its neighbors. None of the stuff that was at a vertex remains at the vertex, but stuff can enter from other vertices. This is a discrete-time and slightly unnatural notion of diffusion, but it provides a nice matrix.

To construct the diffusion matrix, let $\boldsymbol{D}_G$ be the diagonal matrix in which $\boldsymbol{D}_G(u,u)$ is the degree of vertex u. We will usually write $\boldsymbol{d}(u)$ for the degree of vertex u. In an unweighted graph, the degree of a vertex is the number of edges attached to it. In the case of a weighted graph, we use the *weighted degree*: the sum of the weights of the edges attached to the vertex u.

We then set

$$\boldsymbol{W}_G = \boldsymbol{D}_G^{-1}\boldsymbol{A}_G.$$

Of course, when the graph is *regular*, that is when every vertex has the same degree, $\boldsymbol{W}_G$ is merely a rescaling of $\boldsymbol{A}_G$. I think this is why researchers got away with studying the adjacency matrix for so long.

Formally[2], we use a vector $\boldsymbol{p} \in \mathbb{R}^V$ to indicate how much stuff is at each vertex, with $\boldsymbol{p}(u)$ being the amount of stuff at vertex u. When describing diffusion, I will treat $\boldsymbol{p}$ as a row vector. After one time step, the distribution of stuff at each vertex will be $\boldsymbol{p}\,\boldsymbol{W}_G$. To see this, first consider the case when $\boldsymbol{p}$ is an elementary unit vector, $\boldsymbol{\delta}_u$, where I define $\boldsymbol{\delta}_u$ to be the vector for which $\boldsymbol{\delta}_u(u) = 1$, and for every other vertex v, $\boldsymbol{\delta}_u(v) = 0$. The vector $\boldsymbol{\delta}_u\boldsymbol{D}_G^{-1}$ has the value $1/\boldsymbol{d}(u)$ at vertex u, and is zero everywhere else. So, the vector $\boldsymbol{\delta}_u\boldsymbol{D}_G^{-1}A_G$ has value $1/\boldsymbol{d}(u)$ at every vertex v that is a neighbor of u, and is zero everywhere else. If this is not immediately obvious, think about it until it is.

A few lectures from now we will see that spectral theory provides a good understanding of what happens when one repeatedly applies a linear operator like $\boldsymbol{W}_G$.

1.5.3 A quadratic form

The most natural quadratic form associated with a graph is defined in terms of its Laplacian matrix,

$$\boldsymbol{L}_G \stackrel{\text{def}}{=} \boldsymbol{D}_G - \boldsymbol{A}_G.$$

Given a function on the vertices, $\boldsymbol{x} \in \mathbb{R}^V$, the Laplacian quadratic form is

$$\boldsymbol{x}^T\boldsymbol{L}_G\boldsymbol{x} = \sum_{(u,v)\in E} (\boldsymbol{x}(u) - \boldsymbol{x}(v))^2. \tag{1.1}$$

This form measures the smoothness of the function $\boldsymbol{x}$. It will be small if the function $\boldsymbol{x}$ does not jump too much over any edge.

1.6 Background: Spectral Theory

I now review the highlights of the spectral theory for symmetric matrices. Almost all of the matrices we consider in this course will be symmetric or will be similar[3] to symmetric matrices.

[2] I write $\mathbb{R}^V$ instead of $\mathbb{R}^n$ to emphasize that each coordinate of the vector corresponds to a vertex of the graph.

[3] A matrix $\boldsymbol{A}$ is similar to a matrix $\boldsymbol{B}$ if there is a non-singular matrix $\boldsymbol{M}$ such that $\boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{M} = \boldsymbol{B}$. In this case, $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same eigenvalues. See the exercises at the end of this lecture.

We recall that a vector $\boldsymbol{\psi}$ is an eigenvector of a matrix $\boldsymbol{M}$ with eigenvalue λ if

$$\boldsymbol{M}\boldsymbol{\psi} = \lambda\boldsymbol{\psi}. \tag{1.2}$$

That is, λ is an eigenvalue if and only if $\lambda\boldsymbol{I} - \boldsymbol{M}$ is a singular matrix. Thus, the eigenvalues are the roots of the characteristic polynomial of $\boldsymbol{M}$:

$$\det(x\boldsymbol{I} - \boldsymbol{M}).$$

Theorem 1.6.1. *[The Spectral Theorem] If $\boldsymbol{M}$ is an n-by-n, real, symmetric matrix, then there exist real numbers $\lambda_1, \ldots, \lambda_n$ and n mutually orthogonal unit vectors $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n$ and such that $\boldsymbol{\psi}_i$ is an eigenvector of $\boldsymbol{M}$ of eigenvalue λ_i, for each i.*

This is the great fact about symmetric matrices. If the matrix is not symmetric, it might not have n eigenvalues. And, even if it has n eigenvalues, their eigenvectors will not be orthogonal[4]. In fact, if M is not symmetric, then its eigenvalues and eigenvalues might be the wrong thing to look at.

Before I go on to explain why, I should first remind you that the eigenvectors are not uniquely determined, although the eigenvalues are. If $\boldsymbol{\psi}$ is an eigenvector, then $-\boldsymbol{\psi}$ is as well. Some eigenvalues can be repeated. If $\lambda_i = \lambda_{i+1}$, then $\boldsymbol{\psi}_i + \boldsymbol{\psi}_{i+1}$ will also be an eigenvector of eigenvalue λ_i. Generally, the eigenvectors of a given eigenvalue are only determined up to an orthogonal transformation.

The equation (1.2) defining eigenvalues and eigenvectors does not give much of a hint as to why the eigenvalues and eigenvectors should have combinatorial significance. Fortunately, the eigenvalues and eigenvectors of symmetric matrices have many characterizations. The ones that will be most useful to us will come from optimization problems. In particular, they arise when maximizing or minimizing the Rayleigh quotient with respect to a matrix $\boldsymbol{M}$.

Definition 1.6.2. *The Rayleigh quotient of a vector $\boldsymbol{x}$ with respect to a matrix $\boldsymbol{M}$ is the ratio*

$$\frac{\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}.$$

Observe that if $\boldsymbol{\psi}$ is an eigenvector of $\boldsymbol{M}$ of eigenvalue λ, then

$$\frac{\boldsymbol{\psi}^T\boldsymbol{M}\boldsymbol{\psi}}{\boldsymbol{\psi}^T\boldsymbol{\psi}} = \frac{\boldsymbol{\psi}^T\lambda\boldsymbol{\psi}}{\boldsymbol{\psi}^T\boldsymbol{\psi}} = \frac{\lambda\boldsymbol{\psi}^T\boldsymbol{\psi}}{\boldsymbol{\psi}^T\boldsymbol{\psi}} = \lambda.$$

Theorem 1.6.3. *Let $\boldsymbol{M}$ be a symmetric matrix and let $\boldsymbol{x}$ be a non-zero vector that maximizes the Rayleigh quotient with respect to $\boldsymbol{M}$. Then, $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{M}$ with eigenvalue equal to the Rayleigh quotient. Moreover, this eigenvalue is the largest eigenvalue of $\boldsymbol{M}$.*

Before I continue, I should point out that such a vector is guaranteed to exist. To see this, observe that the Rayleigh quotient does not change if we multiply $\boldsymbol{x}$ by a nonzero constant. So, it suffices to consider vectors $\boldsymbol{x}$ of unit norm, which is a compact set. As the Rayleigh quotient is a continuous function away from the origin, there exists a vector in this set at which it is maximized.

This theorem is so important that I will show you at least two proofs of it. First, we will assume the spectral theorem.

[4]You can prove that if the eigenvectors are orthogonal, then the matrix is symmetric.

Proof. First, number the eigenvalues of $\boldsymbol{M}$ so that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Since $\{\boldsymbol{\psi}_i\}_i$ is an orthonormal basis, we can write any vector $\boldsymbol{x}$ in this basis as

$$\boldsymbol{x} = \sum_i (\boldsymbol{\psi}_i^T \boldsymbol{x}) \boldsymbol{\psi}_i.$$

If you don't remember this, the easiest way to verify it is to check that this gives the correct inner products with every vector in the basis:

$$\begin{aligned}
\boldsymbol{\psi}_j^T \left(\sum_i (\boldsymbol{\psi}_i^T \boldsymbol{x}) \boldsymbol{\psi}_i \right) &= \left(\sum_i (\boldsymbol{\psi}_i^T \boldsymbol{x}) \boldsymbol{\psi}_j^T \boldsymbol{\psi}_i \right) && \text{(by linearity)} \\
&= (\boldsymbol{\psi}_j^T \boldsymbol{x}) \boldsymbol{\psi}_j^T \boldsymbol{\psi}_j && (\text{as } \boldsymbol{\psi}_j^T \boldsymbol{\psi}_i = 0 \text{ for } i \neq j) \\
&= \boldsymbol{\psi}_j^T \boldsymbol{x} && (\text{as } \boldsymbol{\psi}_j \text{ is a unit vector , } \boldsymbol{\psi}_j^T \boldsymbol{\psi}_j = 1).
\end{aligned}$$

As the multiplication of $\boldsymbol{x}$ by a constant does not change the Rayleigh quotient, we can assume without loss of generality that $\boldsymbol{x}$ is a unit vector. That is, $\boldsymbol{x}^T \boldsymbol{x} = \|\boldsymbol{x}\|^2 = 1$. We write this in the eigenbasis as

$$\sum_i (\boldsymbol{\psi}_i^T \boldsymbol{x})^2 = 1.$$

We now evaluate the Rayleigh quotient with respect to $\boldsymbol{x}$:

$$\begin{aligned}
\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} &= \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} \\
&= \left(\sum_i (\boldsymbol{\psi}_i^T \boldsymbol{x}) \boldsymbol{\psi}_i \right)^T \boldsymbol{M} \left(\sum_j (\boldsymbol{\psi}_j^T \boldsymbol{x}) \boldsymbol{\psi}_j \right) \\
&= \left(\sum_i (\boldsymbol{\psi}_i^T \boldsymbol{x}) \boldsymbol{\psi}_i \right)^T \left(\sum_j (\boldsymbol{\psi}_j^T \boldsymbol{x}) \lambda_j \boldsymbol{\psi}_j \right) \\
&= \sum_{i,j} (\boldsymbol{\psi}_i^T \boldsymbol{x})(\boldsymbol{\psi}_j^T \boldsymbol{x}) \lambda_j \boldsymbol{\psi}_i^T \boldsymbol{\psi}_j \\
&= \sum_j (\boldsymbol{\psi}_j^T \boldsymbol{x})(\boldsymbol{\psi}_j^T \boldsymbol{x}) \lambda_j, && \text{as } \boldsymbol{\psi}_i^T \boldsymbol{\psi}_j = 0 \text{ for } i \neq j, \\
&= \sum_j (\boldsymbol{\psi}_j^T \boldsymbol{x})^2 \lambda_j \\
&\leq \lambda_n \sum_j (\boldsymbol{\psi}_j^T \boldsymbol{x})^2 \\
&= \lambda_n.
\end{aligned}$$

So, the Rayleigh quotient is never more than λ_n. We have already established that the value of λ_n is achieved by $\boldsymbol{\psi}_n$. It remains to show that only eigenvectors of eigenvalue λ_n can have such a large Rayleigh quotient. This follows from the observation that the inequality above is tight if and only if $\lambda_i = \lambda_n$ for all i such that $\boldsymbol{\psi}_i^T \boldsymbol{x} \neq 0$. That is, if and only if $\boldsymbol{x}$ is in the span of the eigenvectors with eigenvalue λ_n.

□

You can similarly prove that λ_1 is the minimum possible Rayleigh quotient, and that

$$\lambda_i = \min_{\boldsymbol{x} \perp \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i-1}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}},$$

and that one can choose the eigenvectors by

$$\boldsymbol{\psi}_i = \arg \min_{\boldsymbol{x} \perp \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i-1}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

I'll now give another proof of Theorem 1.6.2 that does not assume the spectral theorem. In fact, we can use this proof to prove the spectral theorem.

Proof of Theorem 1.6.2. Let $\boldsymbol{x}$ be a non-zero vector that maximizes the Rayleigh quotient. We recall that the gradient of a function at its maximum must be the zero vector. That is, its derivatives in every direction must be zero. We now compute those derivatives.

Let u be any coordinate. We have

$$\frac{\partial}{\partial \boldsymbol{x}(u)} \boldsymbol{x}^T \boldsymbol{x} = \frac{\partial}{\partial \boldsymbol{x}(u)} \sum_v (\boldsymbol{x}(v))^2 = 2\boldsymbol{x}(u).$$

Similarly, we may compute

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}(u)} \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} &= \frac{\partial}{\partial \boldsymbol{x}(u)} \sum_{v,w} \boldsymbol{M}(v,w) \boldsymbol{x}(v) \boldsymbol{x}(w) \\
&= \frac{\partial}{\partial \boldsymbol{x}(u)} \left(\boldsymbol{M}(u,u) \boldsymbol{x}(u) \boldsymbol{x}(u) + \sum_{v \neq u} \boldsymbol{M}(v,u) \boldsymbol{x}(v) \boldsymbol{x}(u) + \sum_{v \neq u} \boldsymbol{M}(u,v) \boldsymbol{x}(u) \boldsymbol{x}(v) \right) \\
&= 2 \left(\boldsymbol{M}(u,u) \boldsymbol{x}(u) + \sum_{v \neq u} \boldsymbol{M}(v,u) \boldsymbol{x}(v) \right) \\
&= 2(\boldsymbol{M}\boldsymbol{x})(u).
\end{aligned}$$

To simplify notation and prevent a mess in the next computation, I recall that $\boldsymbol{\delta}_u$ is the elementary unit vector with a 1 in position u. So,

$$\boldsymbol{x}(u) = \boldsymbol{\delta}_u^T \boldsymbol{x} \quad \text{and} \quad (\boldsymbol{M}\boldsymbol{x})(u) = \boldsymbol{\delta}_u^T \boldsymbol{M} \boldsymbol{x}.$$

We now compute

$$\frac{\partial}{\partial \boldsymbol{x}(u)} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{(\boldsymbol{x}^T \boldsymbol{x})(2\boldsymbol{\delta}_u^T \boldsymbol{M} \boldsymbol{x}) - (\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x})(2\boldsymbol{\delta}_u^T \boldsymbol{x})}{(\boldsymbol{x}^T \boldsymbol{x})^2}.$$

In order for this to be zero, we must have

$$(\boldsymbol{x}^T \boldsymbol{x})(2\boldsymbol{\delta}_u^T \boldsymbol{M} \boldsymbol{x}) - (\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x})(2\boldsymbol{\delta}_u^T \boldsymbol{x}) = 0,$$

which implies

$$\boldsymbol{\delta}_u^T \boldsymbol{M}\boldsymbol{x} = \frac{\boldsymbol{x}^T \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} \boldsymbol{\delta}_u^T \boldsymbol{x}.$$

As this holds for every coordinate u, it implies that

$$\boldsymbol{M}\boldsymbol{x} = \frac{\boldsymbol{x}^T \boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} \boldsymbol{x}.$$

That is, if and only if $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{M}$ with eigenvalue equal to its Rayleigh quotient. □

1.7 Eigenvectors of the Laplacian

We will now use this characterization of eigenvalues to reason about the smallest eigenvalues of the Laplacian. We will do this by thinking about the Laplacian quadratic form (1.1). As all of the terms in the sum defining the form are squares, we can see that it is never negative. If we let $\boldsymbol{x}$ be a constant vector, then this form is zero. So, the smallest eigenvalue of the Laplacian is zero.

One can also show that $\lambda_2 > 0$ if and only if the graph is connected. If the graph is disconnected, one can construct at least two orthogonal vectors with eigenvalue zero: consider vectors that are constant on one component of the graph and zero everywhere else. On the other hand, if the graph is connected then we can show that $\lambda_2 > 0$: Let $\boldsymbol{x}$ be any vector orthogonal to the constant vectors. So, there must be two vertices u and v for which $\boldsymbol{x}(u) \neq \boldsymbol{x}(v)$. As there is a path between these vertices, there must be some edge across which $\boldsymbol{x}$ changes value. So, the quadratic form will be positive on $\boldsymbol{x}$. (For a more formal proof, look at the notes from my second lecture on 2009)

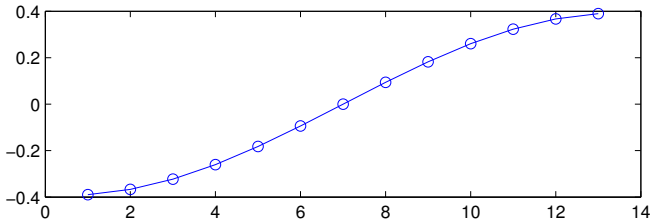
The attempt to make this qualitative statement quantitative led Fiedler [Fie73] to call λ_2 the "algebraic connectivity of a graph". We will relate λ_2 to how well connected a graph is. In general, it will be interesting to turn qualitative statements like this into quantitative ones. For example, we will later see that the smallest eigenvalue of the diffusion matrix is zero if and only if the graph is bipartite. One can relate the magnitude of this eigenvalue to how far a graph is from being bipartite.

1.8 Examples

Let's start by examining the eigenvectors of the Laplacian of the path graph. I'll do examples in Matlab, so you can follow along if you like.

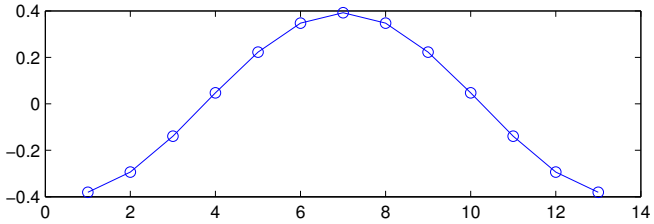
I'll first construct the path graph on 20 vertices, and plot its second eigenvector. The x-axis in this plot gives the number of the vertex. The y-axis is the value of the eigenvector at a vertex.

```
a = diag(ones(1,19),1);
a = a + a';
la = diag(sum(a)) - a;
[v,d] = eig(full(la));
diag(d)
plot(v(:,2));
hold on; plot(v(:,2),'o')
```



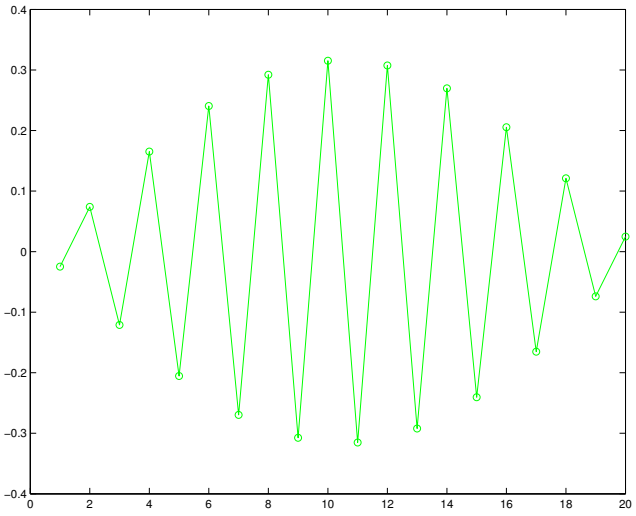
I'll now plot the third and fourth eigenvectors on top of this. You can see that the number of sign changes in the eigenvectors equals its index. The higher the eigenvalue, the more times it oscillates.

```
plot(v(:,3),'k');
plot(v(:,3),'ko');
plot(v(:,4),'r');
plot(v(:,4),'ro');
```



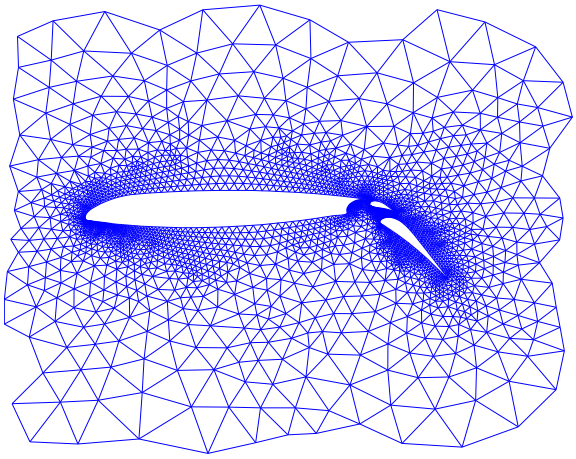
And, here is the eigenvector of highest eigenvalue.

```
clf;
plot(v(:,20),'g');
hold on
plot(v(:,20),'go');
```



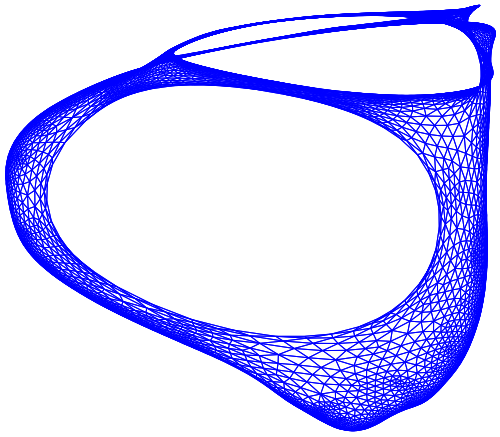
Usually graphs don't come with coordinates. In this case, a good way to try to draw them is to use the eigenvectors to supply coordinates. For example, here is the airfoil graph that comes with Matlab. I'll first draw it using the suggested coordinates. These are given by the variables x and y that come with the file. They indicate where to draw the vertices. The edges are then drawn as straight lines.

```
load airfoil
A = sparse(i,j,1,4253,4253);
A = A + A';
gplot(A,[x,y])
```



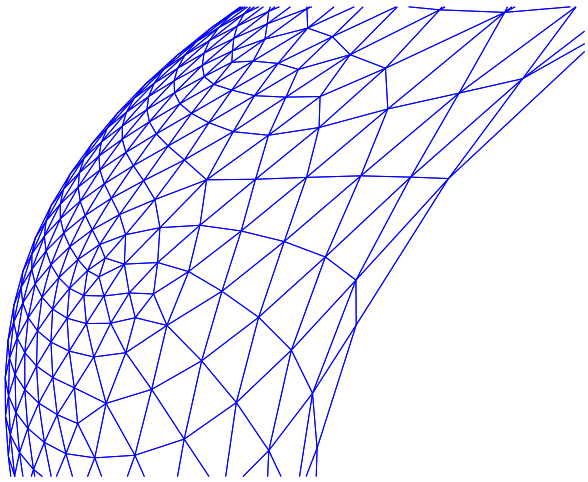
We will now draw it using coordinates from the eigenvectors. I will plot vertex u at point $(\boldsymbol{\psi}_2(u), \boldsymbol{\psi}_3(u))$ I will draw the edges as straight lines between the vertices they connect.

```
L = lap(A);
[v,d] = eigs(L,3,'sa');
gplot(A,v(:,[2 3]))
```



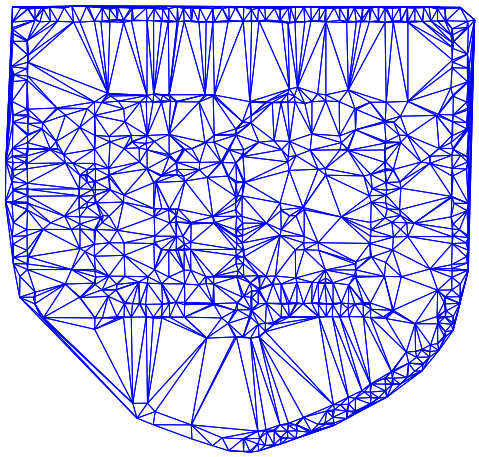
Let's zoom in on a small portion, so that we can see how nice the embedding is.

```
set(gca,'XLim',[-.032, -.024])
set(gca,'YLim',[-.0023, .0052])
```



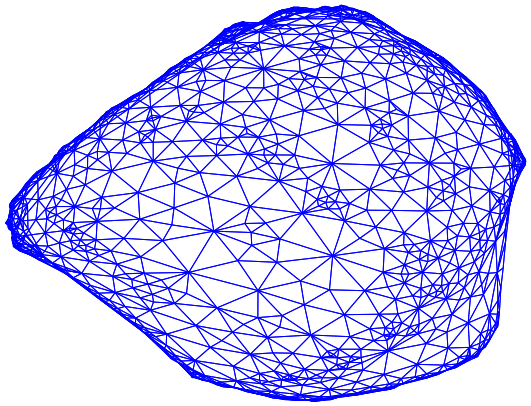
Let's do a more interesting example. I generated the following graph by sampling pixels in the Yale Coat of Arms with probability proportional to their darkness. Those pixels became vertices. I then generated the Delaunay triangulation of these points to form the following graph:

```
load yaleShieldBig
gplot(A,xy)
```



Here is the drawing that I get by using the second and third eigenvectors of the Laplacian.

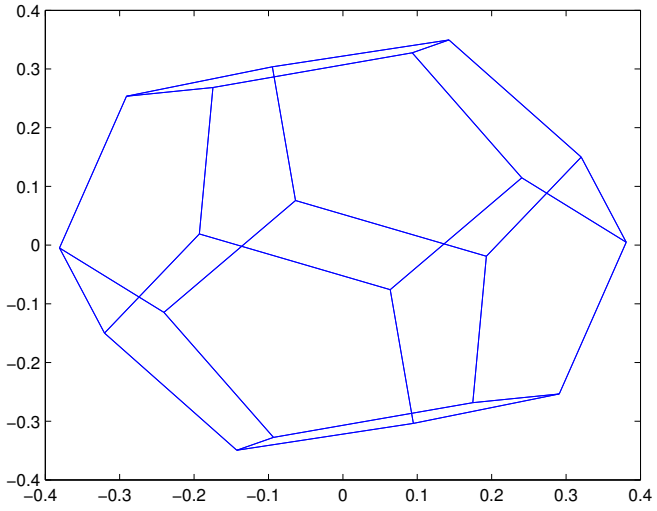
```
L = lap(A);
[v,d] = eigs(L, 3, 'sa');
gplot(A,v(:,[2 3]))
```



That's a great way to draw a graph if you start out knowing nothing about it. It's the first thing I do whenever I meet a strange graph. Note that the middle of the picture is almost planar, although edges do cross near the boundaries.

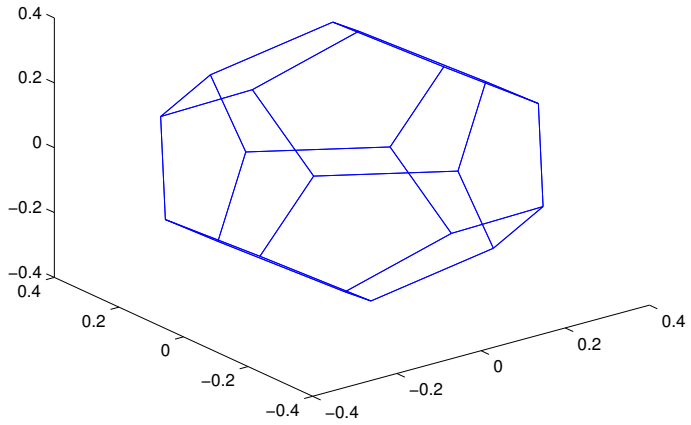
Finally, let's look at a spectral embedding of the edges of the dodecahedron.

```
load dodec
L = lap(A);
[v,d] = eigs(L,3,'sa');
gplot(A,v(:,[2 3]))
```



You will notice that this looks like what you would get if you squashed the dodecahedron down to the plane. The reason is that we really shouldn't be drawing this picture in two dimensions: the smallest non-zero eigenvalue of the Laplacian has multiplicity three. So, we can't reasonably choose just two eigenvectors. We should be choosing three that span the eigenspace. If we do, we would get the canonical representation of the dodecahedron in three dimensions.

```
[v,d] = eigs(L,4,'sa');
gplot3(A,v(:,2:4))
```



As you would guess, this happens for all Platonic solids. In fact, if you properly re-weight the edges, it happens for every graph that is the one-skeleton of a convex polytope.

1.9 Exercises

The following exercises are for your own practice. They are intended as a review of fundamental linear algebra. I have put the solutions in a separate file that you can find on Classes V2. I recommend that you try to solve all of these before you look at the solutions, so that you can get back in practice at doing linear algebra.

1. Orthogonal eigenvectors. Let $\boldsymbol{A}$ be a symmetric matrix, and let $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ be vectors so that

$$\boldsymbol{A}\boldsymbol{\psi} = \mu\boldsymbol{\psi} \quad \text{and} \quad \boldsymbol{A}\boldsymbol{\phi} = \nu\boldsymbol{\phi}.$$

Prove that if $\mu \neq \nu$ then $\boldsymbol{\psi}$ must be orthogonal to $\boldsymbol{\phi}$. Note that your proof should exploit the symmetry of $\boldsymbol{A}$, as this statement is false otherwise.

2. Invariance under permutations.

Let $\boldsymbol{\Pi}$ be a permutation matrix. That is, there is a permutation $\pi : V \to V$ so that

$$\boldsymbol{\Pi}(u, v) = \begin{cases} 1 & \text{if } u = \pi(v), \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Prove that if

$$\boldsymbol{A}\boldsymbol{\psi} = \lambda\boldsymbol{\psi},$$

then

$$\left(\boldsymbol{\Pi}\boldsymbol{A}\boldsymbol{\Pi}^T\right)(\boldsymbol{\Pi}\boldsymbol{\psi}) = \lambda(\boldsymbol{\Pi}\boldsymbol{\psi}).$$

That is, permuting the coordinates of the matrix merely permutes the coordinates of the eigenvectors, and does not change the eigenvalues.

3. Invariance under rotations.

Let $\boldsymbol{Q}$ be an orthonormal matrix. That is, a matrix such that $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$. Prove that if

$$\boldsymbol{A}\boldsymbol{\psi} = \lambda\boldsymbol{\psi},$$

then

$$\left(\boldsymbol{Q}\,\boldsymbol{A}\,\boldsymbol{Q}^T\right)(\boldsymbol{Q}\boldsymbol{\psi}) = \lambda(\boldsymbol{Q}\boldsymbol{\psi}).$$

4. Similar Matrices.

A matrix $\boldsymbol{A}$ is similar to a matrix $\boldsymbol{B}$ if there is a non-singular matrix $\boldsymbol{M}$ such that $\boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{M} = \boldsymbol{B}$. Prove that similar matrices have the same eigenvalues.

5. Spectral decomposition.

Let $\boldsymbol{A}$ be a symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and let $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n$ be a corresponding set of orthonormal column eigenvectors. Let $\boldsymbol{\Psi}$ be the orthonormal matrix whose ith column is $\boldsymbol{\psi}_i$. Prove that

$$\boldsymbol{\Psi}^T\boldsymbol{A}\,\boldsymbol{\Psi} = \boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix with $\lambda_1, \ldots, \lambda_n$ on its diagonal. Conclude that

$$\boldsymbol{A} = \boldsymbol{\Psi}\,\boldsymbol{\Lambda}\,\boldsymbol{\Psi}^T = \sum_{i\in V}\lambda_i\boldsymbol{\psi}_i\boldsymbol{\psi}_i^T.$$

6. Traces.

Recall that the trace of a matrix $\boldsymbol{A}$, written $\mathrm{Tr}\,(\boldsymbol{A})$, is the sum of the diagonal entries of $\boldsymbol{A}$. Prove that for two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\mathrm{Tr}\,(\boldsymbol{A}\boldsymbol{B}) = \mathrm{Tr}\,(\boldsymbol{B}\,\boldsymbol{A})\,.$$

Note that the matrices **do not** need to be square for this to be true. They can be rectangular matrices of dimensions $n \times m$ and $m \times n$.

Use this fact and the previous exercise to prove that

$$\mathrm{Tr}\,(\boldsymbol{A}) = \sum_{i=1}^{n}\lambda_i,$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$.

References

[Fie73] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(98):298–305, 1973.