Linear Algebra Background Knowledge.

determinant: $\det(A)$, $\det A$, $|A|$

Geometrically, it can be viewed as the scaling factor of the linear transformation described by the matrix.

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

The determinant gives the signed n-dimensional volume of this parallelotope, $\det(A) = \pm \mathrm{vol}(P)$.
In particular, if the determinant is zero, then this parallelotope has volume zero and is not fully n-dimensional, which indicates that the dimension of the image of A is less than n.
This means that A produces a linear transformation which is neither onto nor one-to-one, and so is not invertible.

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{vmatrix} = a \begin{vmatrix} f & g & h \\ j & k & l \\ n & o & p \end{vmatrix} - b \begin{vmatrix} e & g & h \\ i & k & l \\ m & o & p \end{vmatrix} + c \begin{vmatrix} e & f & h \\ i & j & l \\ m & n & p \end{vmatrix} - d \begin{vmatrix} e & f & g \\ i & j & k \\ m & n & o \end{vmatrix}$$

characteristic polynomial:

The characteristic polynomial of a square matrix is a polynomial which is invariant under matrix similarity and has eigenvalues as roots.

$$p_A(t) = \det(tI - A)$$

invertible = nonsingular = nondegenerate

Singular: A square matrix A that is not invertible is called singular or degenerate.
A square matrix is singular if and only if its determinant is 0.

multiplicity: the multiplicity of a member of a multiset is the number of times it appears in the multiset.

Spectral Graph Theory by Daniel A. Spielman

{Lecture 1}

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

Matrices for Graphs

adjacency matrix, M_G, whose entries $M_G(a,b)$ are given by

$$M_G(a,b) = \begin{cases} 1 & \text{if } (a,b) \in E \\ 0 & \text{otherwise} \end{cases}$$

* index the rows and columns of the matrix by vertices, rather than by number.

view M as 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 and how random walks behave.

use M to define a quadratic form:

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

$$L_G \overset{\text{def}}{=} D_G - M_G$$

D_G is the diagonal matrix in which $D_G(a,a)$ is the degree of vertex a. $d(a)$.

In a weighted graph, we use the weighted degree: the sum of the weights of edges attached to the vertex a.

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

$$x^T L_G x = \sum_{(a,b) \in E} (x(a) - x(b))^2$$

This form measures the smoothness of the function x.
It will be small if the function x does not jump too much over any edge.
$x(a)$ denotes the coordinate of vector x corresponding to vertex a

a vector ψ is an eigenvector of a Matrix M with eigenvalue λ if

$$M\psi = \lambda\psi$$

λ is an eigenvalue if and only if $\lambda I - M$ is a singular matrix ?

Thus, the eigenvalues are the roots of the characteristic polynomial of M:

$$\det(xI - M)$$

Theorem 1.6.1 [The Spectral Theorem] If 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 $\psi_1, \ldots, \psi_n$ and such that ψ_i is an eigenvector of M of eigenvalue λ_i, for each i.

- If the matrix is not symmetric, it might not have n eigenvalues.
- If the eigenvectors are orthogonal, then the matrix is symmetric ?

Fact 1.6.2 The Laplacian matrix of a graph is positive semidefinite. That is, all its eigenvalues are nonnegative.

proof Let ψ be a unit eigenvector of L of eigenvalue λ. Then,

$$\psi^T L \psi = \psi^T \lambda \psi = \lambda = \sum_{(a,b)\in E} (\psi(a) - \psi(b))^2 \geq 0$$

we always number the eigenvalues of the Laplacian from smallest to largest.

Thus, $\lambda_1 = 0$. ? $\longrightarrow$ has a constant eigenvector

We will refer to λ_2, and in general λ_k for small k, as low-frequency eigenvalues.

λ_n is a high-frequency eigenvalue.

$\downarrow$

the curves they traces out resemble the low-frequency modes of vibration of a string

We will relate low-frequency eigenvalues to connectivity.

We will relate high-frequency eigenvalues to problems of graph coloring and finding independent sets.

1.7.2 Spectral Graph Drawing

We can often use the low-frequency eigenvalues to obtain a nice drawing of a graph.

That's a great way to draw a graph if you start out knowing nothing about it.

1.7.3 Graph Isomorphism

If we permute the vertices then the eigenvectors are similarly permuted. That is, if P is a permutation matrix, then

$$L\psi = \lambda\psi \quad \text{if and only if} \quad (PLP^T)(P\psi) = \lambda(P\psi)$$

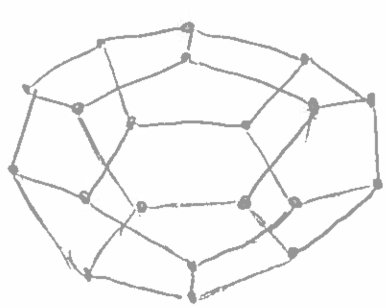
because $P^TP = I$.

"Graph Isomorphism Testing Problem"

First, check if the two graphs have the same sets of eigenvalues.
If they don't, they they are not isomorphic.
If they do, and the eigenvalues have multiplicity? one then draw the pictures. If the pictures are the same, up to horizontal or vertical flips, and no vertex is mapped to the same location as another, then by lining up the pictures we can recover the permutation.

1.7.4 Platonic Solids

dodecahedron



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

1.7.5 The Fiedler Value

The second-smallest eigenvalue of the Laplacian matrix of a graph is zero if and only if the graph is disconnected.

If G is disconnected, then we can partition it into two graphs G_1 and G_2 with no edges between them.

$$L_G = \begin{pmatrix} L_{G_1} & 0 \\ 0 & L_{G_2} \end{pmatrix}$$

Fiedler suggested that think of λ_2 as a measure of how well connected the graph is. He called it "Algebraic Connectivity" of a graph, and we call it the "Fiedler value".

Fiedler proved that the further λ_2 is from 0, the better connected the graph is.

Cheeger's inequality.

If λ_2 is small, then for some t, the set of vertices.

$$S_i \overset{\text{def}}{=} \{ i : v_2(i) < t \}$$

may be removed by cutting much less than $|S_i|$ edges.

The smallest eigenvalue of the diffusion matrix is zero if and only if the graph is bipartite.

1.7.7 Planar Graphs

We will prove that graphs that can be drawn nicely must have small Fiedler value.

1.7.9 Expanders

Roughly speaking, expanders are sparse graphs (say a number of edges linear in the number of vertices), in which λ_2 is bounded away from zero by a constant.

2.1 Eigenvalues and Optimization {Lecture 2}

Eigenvalues arise as the solution to natural optimization problems.

Theorem 2.1.1 Let M be a symmetric matrix and let x be a non-zero vector that maximizes the Rayleigh quotient with respect to M:

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

Then, x is an eigenvector of M with eigenvalue equal to the Rayleigh quotient. Moreover, this eigenvalue is the largest eigenvalue of M.

Proof. We recall that the gradient of a function at its maximum must be the zero vector. Let's compute the gradient.

We have

$$\nabla x^T x = 2x$$

and

$$\nabla x^T M x = 2Mx$$

So,

$$\nabla \frac{x^T M x}{x^T x} = \frac{(x^T x)(2Mx) - (x^T M x)(2x)}{(x^T x)^2}$$

In order for this to be zero, we must have

$$Mx = \frac{x^T M x}{x^T x} x$$

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

Theorem 2.1.2 (Courant-Fischer Theorem). Let L be a symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then

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

For example, consider $k=1$, S is just the span ψ_1 and T is all of $\mathbb{R}^n$.

Lemma 2.1.3 Let M be a symmetric matrix with eigenvalues $\mu_1, \dots, \mu_n$ and a corresponding orthonormal basis of eigenvectors $\psi_1, \dots, \psi_n$. Let x be a vector and expand x in the eigenbasis as

$$x = \sum_{i=1}^{n} c_i \psi_i.$$

Then,

$$x^T M x = \sum_{i=1}^{n} c_i^2 \lambda_i.$$

You should check for yourself (or recall) that $c_i = x^T \psi_i$

Proof. Compute:

$$\begin{aligned} x^T M x &= \Big(\sum_i c_i \psi_i\Big)^T M \Big(\sum_j c_j \psi_j\Big) \\ &= \Big(\sum_i c_i \psi_i\Big)^T \Big(\sum_j c_j \lambda_j \psi_j\Big) \\ &= \sum_{i,j} c_i c_j \lambda_j \psi_i^T \psi_j \\ &= \sum_i c_i^2 \lambda_i, \end{aligned}$$

as $\psi_i^T \psi_j = 0$ for $i \neq j$ □

Proof of 2.1.2 Let $\psi_1, \dots, \psi_n$ be an orthonormal set of eigenvectors of L corresponding to $\lambda_1, \dots, \lambda_n$. 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 $\psi_1, \dots, \psi_k$. We can expand every $x \in S_k$ as

$$x = \sum_{i=1}^{k} c_i \psi_i$$

Applying Lemma 2.1.3 we obtain

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

To show that this is in fact the maximum, we will prove that for all subspaces S of dimension k,

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

Let T_k be the span of $\psi_k, \dots, \psi_n$. As T_k has dimension $n-k+1$, every S of dimension k has an intersection with T_k of dimension at least 1. So,

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

Any such x may be expressed as

$$x = \sum_{i=k}^{n} c_i \psi_i.$$

and so

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

Theorem 2.1.4 Let L be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ with corresponding eigenvectors $\psi_1, \dots, \psi_n$. Then

$$\lambda_i = \min_{x \perp \psi_1, \dots, \psi_{i-1}} \frac{x^T L x}{x^T x},$$

and the eigenvectors satisfy

$$\psi_i = \arg \min_{x \perp \psi_1, \dots, \psi_{i-1}} \frac{x^T L x}{x^T x}.$$

2.2 Drawing with Laplacian Eigenvalues

'drawing a graph on a line, that is, mapping each vertex to a real number. Let $x \in \mathbb{R}^V$ be the vector that describes the assignment of a real number to each vertex. We would like most pairs of vertices that are neighbours to be close to one another. So Hall suggested that we choose an x minimizing

$$x^T L x \tag{2.1}$$

Unless we place restrictions on x, the solution will be degenerate.
To avoid this, and to fix the scale of the embedding overall, we require

$$\sum_{a \in V} x(a)^2 = \|x\|^2 = 1. \qquad (2.2)$$

Even with this restriction, another degenerate solution is possible; every vertex maps to $\frac{1}{\sqrt{n}}$.
To prevent this from happening, we add the additional restriction that

$$\sum_{a} x(a) = 1^T x = 0 \qquad (2.3)$$

As 1 is the eigenvector of the 0 eigenvalue of the Laplacian, the nonzero vectors that minimize (2.1) subject to (2.2) and (2.3) are the unit eigenvectors of the Laplacian of eigenvalue λ_2.

Of course, we really want to draw a graph in two dimensions. So, we will assign two coordinates to each vertex given by x and y. As opposed to minimizing (2.1), we will minimize

$$\sum_{(a,b) \in E} \left\| \begin{pmatrix} x(a) \\ y(a) \end{pmatrix} - \begin{pmatrix} x(b) \\ y(b) \end{pmatrix} \right\|^2$$

This turns out not to be so different from minimizing (2.1), as it equals

$$\sum_{(a,b) \in E} (x(a) - x(b))^2 + (y(a) - y(b))^2 = x^T L x + y^T L x.$$

as before, we impose the scale conditions

$$\|x\|^2 = 1 \quad \text{and} \quad \|y\|^2 = 1,$$

and the centering constraints

$$1^T x = 0 \quad \text{and} \quad 1^T y = 0.$$

However, this still leaves us with the degenerate solution $x = y = \psi_2$.
To ensure that the two coordinates are different, Hall introduced the restriction that x be orthogonal to y. One can use the spectral theorem to prove that the solution is given by setting $x = \psi_2$ and $y = \psi_3$, or by taking a rotation of this solution.

2.3 Isoperimetry and λ_2

Let S be a subset of the vertices of a graph. One way of measuring how well S can be separated from the graph is to count the number of edges connecting S to the rest of the graph. These edges are called the boundary of S, which we formally define by

$$\partial(S) \stackrel{def}{=} \{ (a,b) \in E : a \in S, b \notin S \}$$

We are less interested in the total number of edges on the boundary than in the ratio of this number to the size of S itself.

We will call this ratio the isoperimetric ratio of S, and define it by

$$\theta(S) \stackrel{def}{=} \frac{|\partial(S)|}{|S|}$$

The isoperimetric number of a graph is the minimum isoperimetric number over all sets of at most half the vertices:

$$\theta_G \stackrel{def}{=} \min_{|S| \leq n/2} \theta(S)$$

We will now derive a lower bound on θ_G in terms of λ_2.

Theorem 2.3.1 For every $S \subset V$

$$\theta(S) \geq \lambda_2 (1-s)$$

where $s = \frac{|S|}{|V|}$. In particular

$$\theta_G \geq \frac{\lambda_2}{2}$$

Proof. As

$$\lambda_2 = \min_{x : x^T 1 = 0} \frac{x^T L_G x}{x^T x}$$

for every non-zero x orthogonal to 1 we know that

$$x^T L_G x \geq \lambda_2 x^T x$$

To exploit this inequality, we need a vector related to the set S.
A natural chioce is χ_S, the characteristic vector of S,

$$\chi_S(a) = \begin{cases} 1 & \text{if } a \in S \\ 0 & \text{otherwise} \end{cases}$$

We find

$$\chi_S^T L_G \chi_S = \sum_{(a,b)\in E} (\chi_S(a) - \chi_S(b))^2 = |\partial(S)|$$

However, χ_S is not orthogonal to $\mathbf{1}$. To fix this, use

$$x = \chi_S - s\mathbf{1}.$$

so

$$x(a) = \begin{cases} 1-s & \text{for } a \in S, \text{ and} \\ -s & \text{otherwise} \end{cases}$$

We have $x^T\mathbf{1} = 0$, and

$$x^T L_G x = \sum_{(a,b)\in E} ((\chi_S(a) - s) - (\chi_S(b) - s))^2 = |\partial(S)|.$$

To finish the proof, we compute

$$x^T x = |S|(1-s)^2 + (|V| - |S|)s^2 = |S|(1-2s+s^2) + |S|s - |S|s^2 = |S|(1-s)$$

This gives

$$\lambda_2 \leq \frac{\chi_S^T L_G \chi_S}{\chi_S^T \chi_S} = \frac{|\partial(S)|}{|S|(1-s)}$$

This theorem says if λ_2 is big, then G is very well connected.

<u>Claim 2.3.2</u> Let $S \subseteq V$ have size $s|V|$. Then

$$\| \chi_S - s\mathbf{1} \|^2 = s(1-s)|V|.$$