CS 798 - Algorithmic Spectral Graph Theory, Fall 2015, Waterloo

Lecture 1: Course introduction.

I will start with some basic course information, and then do an overview of the technical content of the course, and finally a review of linear algebra for the course.

Course information

Objective: Learn how to use eigenvalues and eigenvectors to design algorithms and prove theorems.
and more generally use linear algebraic and continuous techniques to solve combinatorial problems

Course page: https://cs.uwaterloo.ca/~lapchi/cs798

Requirement: homework 50%, 2-3 assignments
project 50%

Project: To study one topic in depth
- some suggested topics in the project page (also served as further references for the course).
- encourage to choose a topic close to your research interest
- usually survey type, should be related to spectral graph theory

Prerequisties: linear algebra, probability, algorithms, discrete mathematics (graphs).
All of them are covered in undergraduate computer science curriculum.
Could get a good idea about the background of linear algebra required in today's review.
Should like math, as there will be lots of proofs and calculations.
May be demanding, as definitions and techniques building up, so you should have time to review the material regularly (e.g. missing two weeks will be tough).
Pace may be fast. Material may be too much.

References: Notes will be provided. No textbook.
You can see the lecture notes in my previous offering in 2012 to get a good idea, but there are new topics and some presentations of old topics will be changed.
You can also see the notes of similar courses in other universities as linked in the course page.
I usually provide more details in the notes than in the lectures.

Course Overview

There are some recent breakthroughs in using ideas and techniques from spectral graph theory to design better algorithms for graph problems (both in better approximation and faster running time), and also prove new theorems.

In this course, we aim to study these new techniques and see the new connections.

Spectral graph theory is not a new topic.

About thirty years ago, an important connection was made between the second eigenvalue and graph expansion. This connection made by Cheeger's inequality is useful in different areas:

- construction of expander graphs, which have various applications in theory and practice.
- analysis of mixing time in random walks, with many applications in sampling and counting.
- graph partitioning, and the spectral partitioning algorithm is a widely used heuristic in practice.

We will study ~~these~~ in the beginning of this course.

Most of the early results in spectral graph theory are about the second eigenvalue.

In the last few years, researchers have found new connections between higher eigenvalues and graph properties, including the use of the k-th eigenvalue to find a small non-expanding set, to partition the graph into k non-expanding sets, and to do better analysis of the spectral partitioning algorithm and SDP-based approximation algorithms.

We will discuss some of these, but probably not all in details.

Then, we will study random walks on graphs, and see the classical connection between the second eigenvalue and the mixing time.

After that, we see some recent results in using random walks to find a small non-expanding set. This problem is closed related to the unique games conjecture, which is about the limits of polynomial time approximation algorithms. Combining ideas from random walks and higher eigenvalues gives the best attempt to go beyond the unique games barrier.

This is about half of the course. We will also talk about constructions of expander graphs and small-set expanders if time permits.

Through the study of random walks, we will come across the idea of interpreting the graph as an electrical network, which is used to analyse hitting times of random walks.
This idea has found surprising applications recently.
One line of research is to use ideas from convex optimization to compute these electrical flows quickly, and then somehow these could be used to compute maximum flow faster!
The other direction is to use these concepts for graph sparsification, where the spectral perspective proved to be the right way to look at this combinatorial problem.

Finally, we study an exciting new technique called the method of interlacing polynomials.
This is a new probabilistic method showing the existence of good combinatorial objects, but the proofs involve some mathematical concepts like real-stable polynomials.
This method is used to make some breakthroughs in constructing expander graphs (Ramanujan graphs) and partitioning into expander graphs (the Kadison-Singer problem).
We will also discuss an amazing application of these ideas in designing approximation algorithms for the asymmetric traveling salesman problem.
The ideas developed in graph sparsification (e.g. barrier argument) is also a key component in this last part of the course.

This overview is still very sketchy. In class, we will elaborate with more precise definition and more details. This is an unusual exception that the lecture has more details than the notes.

Some important topis that will not be covered included SDP-based approximation algorithms, eigenvalues of random graphs, and applications of spectral methods in machine learning.

The topics are theoretically oriented. I will try to focus on the underlying techniques that are relevant in broader contexts.

If you are interested in learning more about these topics befor the lectures (e.g. previewing the method of interlacing polynomials), the course project page provides further references.

<u>Background on linear algebra</u>

Linear transformation, linear independence, nullspace, rank

Let $A \in \mathbb{R}^{n \times n}$ be an $n \times n$ matrix, and $x \in \mathbb{R}^n$ be a column vector (we use x to denote a column vector and x^T to denote a row vector).

One can view A as a matrix or a linear transformation from n-dimensional vectors to n-dimensional vectors.

A set of vectors $v_1, v_2, \ldots, v_k$ are linearly independent if $c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = 0$ implies $c_1 = c_2 = \cdots = c_k = 0$; otherwise they are linearly dependent.

The nullspace of A (or the kernel of A) is defined as $\text{nullspace}(A) := \{ x \in \mathbb{R}^n \mid Ax = 0 \}$, and its dimension is denoted as $\text{null}(A)$.

The range of A (or the image of A) is defined as $\text{range}(A) := \{ Ax \mid x \in \mathbb{R}^n \}$, and its dimension is denoted as the rank of A, denoted by $\text{rank}(A)$.

It is known that $\text{rank}(A) + \text{null}(A) = n$ if A is an $n \times n$ matrix.

It follows from the definition that $\text{rank}(A)$ = maximum number of linearly independent columns of A.

---

determinant

The determinant of A, denoted by $\det(A)$, is defined recursively as

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(\tilde{A}_{ij}),$$

where a_{ij} is the (i,j)-th entry of A and $\hat{A}_{ij}$ is the matrix obtained from A by deleting the i-th row and the j-th column of A.

We can unfold the recursion and get an equivalent (expansion) definition:

$$\det(A) = \sum_{\sigma \in S^n} \text{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)},$$

where $\sigma : [n] \to [n]$ is a permutation (or a bijection) of the indices, and $\text{sgn}(\sigma) = +1$ if σ is even and $\text{sgn}(\sigma) = -1$ if σ is odd. More precisely, $\text{sgn}(\sigma) = (-1)^{\text{inv}(\sigma)}$ where $\text{inv}(\sigma) = \left| \{ (i,j) \mid i < j \text{ and } \sigma(i) > \sigma(j) \} \right|$ is the number of inversions of σ.

A basic property of determinant is the following:

$$\det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u + kv \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + k \det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ v \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}$$

which follows from the (expansion) definition of determinant.

which follows from the (expansion) definition of determinant.

From this it follows that $\det\begin{pmatrix} \vdots \\ ca_i \\ \vdots \end{pmatrix} = c\det\begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix}$, $\det\begin{pmatrix} v \\ v \\ \vdots \\ \vdots \end{pmatrix} = 0$,

$$\det\begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} = -\det\begin{pmatrix} \vdots \\ a_j \\ \vdots \\ a_i \\ \vdots \end{pmatrix}, \quad \det\begin{pmatrix} \vdots \\ a_i + ca_j \\ \vdots \\ a_j \\ \vdots \end{pmatrix} = \det\begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix}$$

Therefore, we can compute the determinant by reducing A into an upper triangular matrix by elementary row operations, and prove that $\det(A) \neq 0$ if and only if $\operatorname{rank}(A) = n$ (because $\operatorname{rank}(A) = n$ if and only if it can be reduced to an upper triangular matrix where all the diagonal entries are nonzero)

One can also prove that $\det(AB) = \det(A)\det(B)$, as this is true for elementary matrices and one can write A and B as products of elementary matrices if $\operatorname{rank}(A) = \operatorname{rank}(B) = n$.

<u>Eigenvalues and Eigenvectors</u>

A nonzero vector x is an <u>eigenvector</u> of A if there exists λ such that $Ax = \lambda x$, and the scalar λ is called an <u>eigenvalue</u> of A

Note that $Ax = \lambda x$ iff $(A - \lambda I)x = 0$ where $I = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$ is the identity matrix.

There is a nonzero vector x satisfying $(A - \lambda I)x = 0$

$\Leftrightarrow$ nullspace $(A - \lambda I) \neq \{0\}$

$\Leftrightarrow$ $\operatorname{rank}(A - \lambda I) < n$

$\Leftrightarrow$ $\det(A - \lambda I) = 0$, i.e. $\det\begin{pmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{pmatrix} = 0$

By the (expansion) definition of determinant, $\det(A - \lambda I)$ is a polynomial of λ of degree n.

This is called the <u>characteristic polynomial</u> of A.

In the last offering of this course, the characteristic polynomials did not play a major role. but they will be very important in the last part of this offering as they will be the key object in the interlacing method. Also, determinants will play an important role in the last part.

One can compute the characteristic polynomial by Gaussian elimination or interpolation.

Any root of this polynomial is an eigenvalue, and any vector in nullspace $(A-\lambda I)$ is an eigenvector.

Geometrically, an eigenvector is a direction that is "fixed" (but can be scaled) by the linear transformation.

So, not all matrices have an eigenvector with a real eigenvalue, e.g. the rotation matrix has no direction fixed, and thus no eigenvector with a real eigenvalue.

But we will see that all real symmetric matrices have real eigenvalues, a corollary of the spectral theorem that we will prove.

Orthogonality

To state the spectral theorem, we need one more concept.

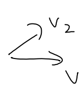
Given two n-dimensional vectors $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ and $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, the inner product of u and v is defined as $\langle u,v \rangle = \sum_{i=1}^{n} u_i v_i$, and we define the norm (or length) of u as $\|u\| = \sqrt{\langle u,u \rangle}$.

Two vectors u and v are orthogonal (or perpendicular) if $\langle u,v \rangle = 0$.

A set S of vectors is orthogonal if u,v are orthogonal for any distinct $u,v \in S$.

A set S of vectors is orthonormal if S is orthogonal and $\|u\|=1$ for every $u \in S$.

A set S of vectors is a basis for a vector space V if S is linearly independent and every vector of V is a linear combination of the vectors in S.

Given any basis, we can construct an orthogonal basis by the Gram-Schmidt orthogonalization process. For example, given 

, we can construct 

by subtracting the projection of v_2 onto v_1.

Given a basis $S = \{w_1, w_2, \dots, w_n\}$, define $S' = \{v_1, v_2, \dots, v_n\}$ where

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j \quad \text{for } 2 \le k \le n.$$

Then S' is an orthogonal basis.

Then $S'' = \left\{ \frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|} \right\}$ is an orthonormal basis.

Having an orthonormal basis is easy for computation, e.g. if A is a matrix whose columns form an orthonormal basis, then $A^T A = I$ and hence $A^{-1} = A^T$ where A^{-1} denotes the inverse of A.

Spectral theorem for real symmetric matrices (follow the presentation in Godsil and Royle)

Theorem Let A be a real symmetric $n\times n$ matrix. Then there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A, and the corresponding eigenvalues are real numbers.

First we prove that there is a real eigenvalue.

Claim 1 Let A be a real symmetric matrix.

If u and v are eigenvectors with different eigenvalues, then u and v are orthogonal.

Proof Let $Au=\alpha u$ and $Av=\beta v$ where $\alpha\neq\beta$.

Then $v^TAu = v^T(Au) = v^T\alpha u = \alpha\langle v,u\rangle$.

On the other hand, $v^TAu = (v^TA)u \underset{\substack{\uparrow \\ \text{because } A \text{ is symmetric}}}{=} \beta v^Tu = \beta\langle v,u\rangle$.

Now $\alpha\langle v,u\rangle = \beta\langle v,u\rangle$ and $\alpha\neq\beta$ implies that $\langle u,v\rangle = 0$, as required. □

Claim 2 The eigenvalues of a real symmetric matrix, if exist, are real numbers.

Proof Let $Au=\lambda u$. By taking the complex conjugate, we get $A\bar{u}=\bar{\lambda}\bar{u}$.

So $\bar{u}$ is also an eigenvector of A.

Since $u^T\bar{u}>0$, by Claim 1, λ and $\bar{\lambda}$ cannot be of distinct values, and so λ is real. □

Since $\det(A-\lambda I)=0$ always has a solution, Claim 2 implies that there is a real eigenvalue.

Now we would like to use induction to finish the proof.

To do this, we consider the vectors orthogonal to the existing eigenvectors, and show that there is an eigenvector in that subset of vectors.

We say a subspace U is <u>A-invariant</u> if $Au\in U$ for every $u\in U$, e.g. if U is the subspace generated by eigenvectors then U is A-invariant.

Claim 3 Let A be a real symmetric matrix. If U is an A-invariant subspace, then $U^\perp$ is also A-invariant, where $U^\perp=\{v \mid \langle v,u\rangle=0 \ \forall u\in U\}$ is the set of vectors orthogonal to the vectors in U.

Proof Let $v\in U^\perp$ and $u\in U$. Then $v^TAu = v^T(Au) = v^Tu' = 0$ where $u'\in U$.

It also means that $(v^TA)u = 0$ for all $v\in U^\perp$ and $u\in U$

So $v^T A$ is in $U^\perp$ for all $v \in U^\perp$.

Since A is symmetric, it implies that $Av \in U^\perp$ for all $v \in U^\perp$.

Hence $U^\perp$ is also A-invariant, as required. $\square$

<u>Claim 4</u> Let A be a real symmetric matrix. If U is a nonzero A-invariant subspace, then U contains a real eigenvector of A.

<u>Proof</u> Let R be a matrix whose columns form an orthonormal basis of U.

Then $AR = RB$ for some square matrix B, since each column of AR is a linear combination of the columns of R.

Then $R^T A R = B$, which implies that B is symmetric.

Therefore, there exists u such that $Bu = \lambda u$.

This implies that $ARu = RBu = \lambda Ru$.

Note that $Ru \neq 0$ for $u \neq 0$ because the columns of R are linearly independent.

Thus $Ru \in U$ is an eigenvector contained in U. $\square$

Now we can finish the proof easily by induction.

Let $\{u_1, \ldots, u_m\}$ be an orthonormal set of eigenvectors of A, where $1 \leq m < n$.

Let U be the subspace that they generate. Then U is A-invariant.

By Claim 3, $U^\perp$ is also A-invariant.

By claim 4, $U^\perp$ contains an eigenvector, and hence repeating this argument will finish the proof.

Applications of the spectral theorem

Power of matrices

Having an orthonormal set of eigenvectors is very good for computation.

For example, let V be the matrix whose columns form an orthonormal basis of eigenvectors of A.

Then $AV = VD$ where D is the diagonal matrix of eigenvalues, and hence $A = VDV^{-1} = VDV^T$.

To compute A^k, we observe that it is just $\overbrace{(VDV^T)(VDV^T)\ldots(VDV^T)}^{k \text{ times}} = VD^kV^T$ since $V^TV = I$, and D^k can be computed readily since D is a diagonal matrix, i.e. if $D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}$, $D^k = \begin{bmatrix} d_1^k & & \\ & \ddots & \\ & & d_n^k \end{bmatrix}$

This is very useful, for example, in analyzing random walk of a graph.

Eigenspace and multiplicity

If λ is an eigenvalue of A, then we call nullspace$(A-\lambda I)$ the *eigenspace* of λ.

When we write $A = VDV^T$ as above, if the dimension of the eigenspace of λ is k, then λ appears in k diagonal entries in D.

Note that $\det(A-\lambda I) = \det(VDV^T - \lambda I) = \det(VDV^T - \lambda VV^T) = \det(V(D-\lambda I)V^T)$

$= \det(D-\lambda I)$ as $\det(AB) = \det(A)\det(B)$ and $\det(A^{-1}) = 1/\det(A)$

$= \prod_{i=1}^{n} (\lambda_i - \lambda)$ where λ_i is the i-th diagonal entry of D.

Since all the eigenvalues are real, the characteristic polynomial of A is real-rooted.

The multiplicity of a root is equal to the dimension of its eigenspace.

Determinant and trace

We prove two identities for real symmetric matrix A.

Fact $\det(A) = \prod_{i=1}^{n} \lambda_i$.

~~proof~~ $\det(A) = \det(VDV^T) = \det(D) = \prod_{i=1}^{n} \lambda_i$. $\square$

The trace of A, denoted by $\mathrm{Tr}(A)$, is defined as the sum of diagonal entries of A

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

Proof Consider $\det(\lambda I - A)$. Its roots are the eigenvalues of A, and so

$\det(\lambda I - A) = (\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)$.

Note that the coefficients of $\lambda^{n-1} = -\sum_{i=1}^{n} \lambda_i$.

On the other hand, $\det(\lambda I - A) = \det\begin{pmatrix} \lambda-a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda-a_{22} & & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda-a_{nn} \end{pmatrix}$.

By the (expansion) definition of the determinant, the coefficient of λ^{n-1} only appears in the term $(\lambda-a_{11})(\lambda-a_{22})\cdots(\lambda-a_{nn})$, which is $-\sum_{i=1}^{n} a_{ii}$.

Therefore, $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} = \mathrm{trace}(A)$. $\square$

Eigen-decomposition

Also, let $\{v_1, v_2, \ldots, v_n\}$ be a basis of orthonormal eigenvectors.

Then any vector x can be written as $c_1 v_1 + c_2 v_2 + \ldots + c_n v_n$.

By orthonormality, $\langle x, v_i \rangle = \langle c_1 v_1 + \cdots + c_n v_n, v_i \rangle = c_1 \langle v_1, v_i \rangle + \ldots + c_i \langle v_i, v_i \rangle + \ldots + c_n \langle v_n, v_i \rangle = c_i$.

Therefore, $x = \langle x, v_1 \rangle v_1 + \langle x, v_2 \rangle v_2 + \ldots + \langle x, v_n \rangle v_n$

$$= v_1 v_1^T x + v_2 v_2^T x + \ldots + v_n v_n^T x$$

$$= (v_1 v_1^T + v_2 v_2^T + \ldots + v_n v_n^T) x.$$

This is true for all x, and hence $v_1 v_1^T + v_2 v_2^T + \ldots + v_n v_n^T = I$.

Multiplying both sides by A, we get

$$Ax = A(v_1 v_1^T + v_2 v_2^T + \ldots + v_n v_n^T) x$$

$$= (\lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T + \ldots + \lambda_n v_n v_n^T) x.$$

Thus, $A = \lambda_1 v_1 v_1^T + \ldots + \lambda_n v_n v_n^T$.

Finally, we claim that $A^{-1} = \frac{1}{\lambda_1} v_1 v_1^T + \frac{1}{\lambda_2} v_2 v_2^T + \ldots + \frac{1}{\lambda_n} v_n v_n^T$ if $\lambda_i \neq 0$ for all i.

because $(\lambda_1 v_1 v_1^T + \ldots + \lambda_n v_n v_n^T)(\frac{1}{\lambda_1} v_1 v_1^T + \ldots + \frac{1}{\lambda_n} v_n v_n^T) = v_1 v_1^T + \ldots + v_n v_n^T = I$.

Later on, we will use this idea to define the "pseudo-inverse" of a matrix A, when A is not of full rank.

Positive semidefinite matrices

A real symmetric matrix M is said to be <u>positive semidefinite</u> if $x^T M x \geq 0$ for every vector x.

<u>Claim</u> A real symmetric matrix M is positive semidefinite if and only if all eigenvalues are non-negative if and only if $M = BB^T$ for some B.

<u>Proof</u> If $M = BB^T$, then $x^T M x = x^T B B^T x = \|B^T x\|^2 \geq 0$, and is thus positive semidefinite.

Since M is symmetric, $M = RDR^T$ where D is a diagonal matrix, and R is an orthonormal basis.

Suppose M is positive semidefinite. Let $\vec{e_i} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i$ be the i-th unit vector.

Let $x = R\vec{e_i}$.

since $R^T R = R R^T = I$.

Then $0 \leq x^T M x = (R\vec{e_i})^T M (R\vec{e_i}) = (\vec{e_i})^T R^T R D R^T R \vec{e_i} \overset{\downarrow}{=} (\vec{e_i})^T D \vec{e_i} = d_i$ where d_i is the i-th eigenvalue.

So $d_i \geq 0$ for all i, and hence all eigenvalues of M are non-negative.

If all eigenvalues are non-negative, then $M = RDR^T = RD^{\frac{1}{2}} D^{\frac{1}{2}} R^T = (D^{\frac{1}{2}} R^T)^T (D^{\frac{1}{2}} R^T)$.

Let $B = D^{\frac{1}{2}} R^T$, where $D^{\frac{1}{2}} = \begin{pmatrix} \sqrt{d_1} & & & \\ & \sqrt{d_2} & & \\ & & \ddots & \\ & & & \sqrt{d_n} \end{pmatrix}$ which is well-defined since $d_i \geq 0$ for all i.

Therefore, $M = B^T B$, as required. $\square$