Stat260/CS294: Spectral Graph Methods	**Lecture 2 - 01/27/2015**

Lecture: Basic Matrix Results (1 of 3)

Lecturer: Michael Mahoney *Scribe: Michael Mahoney*

Warning: these notes are still very rough. They provide more details on what we discussed in class, but there may still be some errors, incomplete/imprecise statements, etc. in them.

2 Introduction

Today and next time, we will start with some basic results about matrices, and in particular the eigenvalues and eigenvectors of matrices, that will underlie a lot of what we will do in this class. The context is that eigenvalues and eigenvectors are complex (no pun intended, but true nonetheless) things and—in general—in many ways not so "nice." For example, they can change arbitrarily as the coefficients of the matrix change, they may or may not exist, real matrices may have complex eigenvectors and eigenvalues, a matrix may or may not have a full set of n eigenvectors, etc. Given those and related instabilities, it is an initial challenge is to understand what we can determine from the spectra of a matrix. As it turns out, for many matrices, and in particular many matrices that underlie spectral graph methods, the situation is much nicer; and, in addition, in some cases they can be related to even nicer things like random walks and diffusions.

So, let's start by explaining "why" this is the case. To do so, let's get some context for how/why matrices that are useful for spectral graph methods are nicer and also how these nicer matrices sit in the larger universe of arbitrary matrices. This will involve establishing a few basic linear algebraic results; then we will use them to form a basis for a lot of the rest of what we will discuss. This is good to know in general; but it is also good to know for more practical reasons. For example, it will help clarify when vanilla spectral graph methods can be extended, *e.g.*, to weighted graphs or directed graphs or time-varying graph or other types of normalizations, etc.

2.1 Some basics

To start, recall that we are interested in the Adjacency matrix of a graph $G = (V, E)$ (or $G = (V, E, W)$ if the graph is weighted) and other matrices that are related to the Adjacency matrix. Recall that the $n \times n$ Adjacency matrix is defined to be

$$A_{ij} = \begin{cases} W_{ij} & \text{if } (ij) \in E \\ 0 & \text{otherwise} \end{cases} ,$$

where $W_{ij} = 1$, for all $(i, j) \in E$ if the graph is unweighted. Later, we will talk about directed graphs, in which case the Adjacency matrix is not symmetric, but note here it is symmetric. So, let's talk about symmetric matrices: a symmetric matrix is a matrix A for which $A = A^T$, i.e., for which $A_{ij} = A_{ji}$.

Almost all of what we will talk about will be real-valued matrices. But, for a moment, we will start with complex-valued matrices. To do so, recall that if $x = \alpha + i\beta \in \mathbb{C}$ is a complex number, then

$\bar{x} = \alpha - i\beta \in \mathbb{C}$ is the *complex conjugate* of x. Then, if $M \in \mathbb{C}^{m\times n}$ is a complex-valued matrix, i.e., an $m \times n$ matrix each entry of which is a complex number, then the *conjugate transpose* of M, which is denoted M^*, is the matrix defined as

$$(M^*)_{ij} = \bar{M}_{ji}.$$

Note that if M happens to be a real-valued $m \times n$ matrix, then this is just the transpose.

If $x, y \in \mathbb{C}^n$ are two complex-valued vectors, then we can define their inner product to be

$$\langle x, y \rangle = x^* y = \sum_{i=1}^{n} \bar{x}_i y_i .$$

Note that from this we can also get a norm in the usual way, i.e., $\langle x, x \rangle = \|x\|_2^2 \in \mathbb{R}$. Given all this, we have the following definition.

Definition 1 *If $M \in \mathbb{C}^{n\times n}$ is a square complex matrix, $\lambda \in \mathbb{C}$ is a scalar, and $x \in \mathbb{C}^n \backslash \{0\}$ is a non-zero vector such that*

$$Mx = \lambda x \tag{1}$$

then λ is an eigenvalue *of M and x is the corresponding* eigenvector *of λ.*

Note that when Eqn. (1) is satisfied, then this is equivalent to

$$(M - \lambda I)\, x = 0, \text{ for } x \neq 0, \tag{2}$$

where I is an $n \times n$ Identity matrix. In particular, this means that we have at least one eigenvalue/eigenvector pair. Since (2) means $M - \lambda I$ is rank deficient, this in turn is equivalent to

$$\det (M - \lambda I) = 0.$$

Note that this latter expression is a polynomial with λ as the variable. That is, if we fix M, then the function given by $\lambda \to \det (M - \lambda I)$ is a univariate polynomial of degree n in λ. Now, it is a basic fact that every non-zero, single-variable, degree polynomial of degree n with complex coefficients has—counted with multiplicity—exactly n roots. (This counting multiplicity thing might seem pedantic, but it will be important latter, since this will correspond to potentially degenerate eigenvalues, and we will be interested in how the corresponding eigenvectors behave.) In particular, any square complex matrix M has n eigenvectors, counting multiplicities, and there is at least one eigenvalue.

As an aside, someone asked in class if this fact about complex polynomials having n complex roots is obvious or intuitive. It is sufficiently basic/important to be given the name the fundamental theorem of algebra, but its proof isn't immediate or trivial. We can provide some intuition though. Note that related formulations of this state that every non-constant single-variable polynomial with complex coefficients has at least one complex root, etc. (e.g., complex roots come in pairs); and that the field of complex numbers is algebraically closed. In particular, the statements about having complex roots applies to real-valued polynomials, i.e., since real numbers are complex numbers polynomials in them have complex roots; but it is false that real-valued polynomials always have real roots. Equivalently, the real numbers are not algebraically closed. To see this, recall that the equation $x^2 - 1 = 0$, viewed as an equation over the reals has two real roots, $x = \pm 1$; but the equation $x^2 + 1 = 0$ does not have any real roots. Both of these equations have roots over the complex plane: the former having the real roots $x = \pm 1$, and the latter having imaginary roots $x = \pm i$.

2.2 Two results for Hermitian/symmetric matrices

Now, let's define a special class of matrices that we already mentioned.

Definition 2 *A matrix $M \in \mathbb{C}^{n \times n}$ is* Hermitian *if $M = M^*$. In addition, a matrix $M \in \mathbb{R}^{n \times n}$ is* symmetric *if $M = M^* = M^T$.*

For complex-valued Hermitian matrices, we can prove the following two lemmas.

Lemma 1 *Let M be a Hermitian matrix. Then, all of the eigenvalues of M are real.*

Proof: Let M be Hermitian and $\lambda \in \mathcal{C}$ and x non-zero be s.t. $Mx = \lambda x$. Then it suffices to show that $\lambda = \lambda^*$, since that means that $\lambda \in \mathbb{R}$. To see this, observe that

$$\begin{aligned} \langle Mx, x \rangle &= \sum_i \sum_j \bar{M_{ij}} \bar{x_j} x_i \\ &= \sum_i \sum_j M_{ji} x_i \bar{x_j} \qquad (3) \\ &= \langle x, Mx \rangle \end{aligned}$$

where Eqn. (3) follows since M is Hermitian. But we have

$$\langle Mx, x \rangle = \langle \lambda x, x \rangle = \bar{\lambda} \langle x, x \rangle = \bar{\lambda} \|x\|_2^2$$

and also that

$$\langle x, Mx \rangle = \langle x, \lambda x \rangle = \lambda \langle x, x \rangle = \lambda \|x\|_2^2.$$

Thus, $\lambda = \bar{\lambda}$, and the lemma follows.

$\diamond$

Lemma 2 *Let M be a Hermitian matrix; and let x and y be eigenvectors corresponding to different eigenvalues. Then x and y are orthogonal.*

Proof: Let $Mx = \lambda x$ and $My = \lambda' y$. Then,

$$\langle Mx, y \rangle = (Mx)^* y = x^* M^* y = x^* M y = \langle x, My \rangle .$$

But,

$$\langle Mx, y \rangle = \lambda \langle x, y \rangle$$

and

$$\langle x, My \rangle = \lambda' \langle x, y \rangle .$$

Thus

$$\left(\lambda - \lambda' \right) \langle x, y \rangle = 0.$$

Since $\lambda \neq \lambda'$, by assumption, it follows that $\langle x, y \rangle = 0$, from which the lemma follows.

$\diamond$

So, Hermitian and in particular real symmetric matrices have real eigenvalues and the eigenvectors corresponding to to different eigenvalues are orthogonal. We won't talk about complex numbers and complex matrices for the rest of the term. (Actually, with one exception since we need to establish that the entries of the eigenvectors are not complex-valued.)

2.3 Consequences of these two results

So far, we haven't said anything about a full set of orthogonal eigenvectors, etc., since, e.g., all of the eigenvectors could be the same or something funny like that. In fact, we will give a few counterexamples to show how the niceness results we establish in this class and the next class fail to hold for general matrices. Far from being pathologies, these examples will point to interesting ways that spectral methods and/or variants of spectral method ideas do or do not work more generally (e.g., periodicity, irreducibility etc.)

Now, let's restrict ourselves to real-valued matrices, in which case Hermitian matrices are just symmetric matrices. With the exception of some results next time on positive and non-negative matrices, where we will consider complex-valued things, the rest of the semester will consider real-valued matrices. Today and next time, we are only talking about complex-valued matrices to set the results that underlie spectral methods in a more general context. So, let's specialize to real-values matrices.

First, let's use the above results to show that we can get a full set of (orthogonalizable) eigenvectors. This is a strong "niceness" result, for two reasons: (1) there is a full set of eigenvectors; and (2) that the full set of eigenvectors can be chosen to be orthogonal. Of course, you can always get a full set of orthogonal vectors for $\mathbb{R}^n$—just work with the canonical vectors or some other set of vectors like that. But what these results say is that for symmetric matrices we can also get a full set of orthogonal vectors that in some sense have something to do with the symmetric matrix under consideration. Clearly, this could be of interest if we want to work with vectors/functions that are in some sense adapted to the data.

Let's start with the following result, which says that given several (i.e., at least one) eigenvector, then we can find another eigenvector that is orthogonal to it/them. Note that the existence of at least one eigenvector follows from the existence of at least one eigenvalue, which we already established.

Lemma 3 *Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and let $x_1, \ldots, x_k$, where $1 \leq k < n$, be orthogonal eigenvectors of M. Then, there is an eigenvector x_{k+1} of M that is orthogonal to $x_1, \ldots, x_k$.*

Proof: Let V be the $(n-k)$-dimensional subspace of $\mathbb{R}^n$ that contains all vectors orthogonal to $x_1, \ldots, x_k$. Then, we claim that: for all $x \in V$, we have that $Mx \in V$. To prove the claim, note that for all $i \in [k]$, we have that

$$\langle x_i, Mx \rangle = x_i^T Mx = (Mx_i)^T x = \lambda_i x_i x = \lambda_i \langle x_i, x \rangle = 0,$$

where x_i is one of the eigenvectors assumed to be given.

Next, let

- $B \in \mathbb{R}^{n\times(n-k)}$ be a matrix consisting of the vectors $b_1, \ldots, b_{n-k}$ that form an orthonormal basis for V. (This takes advantage of the fact that $\mathbb{R}^n$ has a full set of exactly n orthogonal vectors that span it—that are, of course, not necessarily eigenvectors.)
- $B' = B^T$. (If B is any matrix, then B' is a matrix such that, for all $y \in V$, we have that $B'y$ is an $(n-k)$-dimensional vector such that $BB'y = y$. I *think* we don't loose any generality by taking B to be orthogonal.)

- λ be a real eigenvalue of the real symmetric matrix

$$M' = B'MB \in \mathbb{R}^{(n-k)\times(n-k)},$$

with y a corresponding real eigenvector of M. I.e., $M'y = \lambda y$.

Then,

$$B'MBy = \lambda y,$$

and so

$$BB'MBy = \lambda By,$$

from which if follows that

$$MBy = \lambda By.$$

The last equation follows from the second-to-last since $By \perp \{x_1, \ldots, x_k\}$, from which it follows that $MBy \perp \{x_1, \ldots, x_k\}$, by the above claim, and thus $BB'MBy = MBy$. I.e., this doesn't change anything since $BB'\xi = \xi$, for ξ in that space.

So, we can now construct that eigenvector. In particular, we can choose $x_{k+1} = By$, and we have that $Mx_{k+1} = \lambda x_{k+1}$, from which the lemma follows.

$\diamond$

Clearly, we can apply the above lemma multiple times. Thus, as an important aside, the following "spectral theorem" is basically a corollary of the above lemma.

Theorem 1 (Spectral Theorem) *Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and let $\lambda_1, \ldots, \lambda_n$ be its real eigenvalues, including multiplicities. Then, there are n orthonormal vectors $x_1, \ldots, x_n$, with $x_i \in \mathbb{R}^n$, such that x_i is an eigenvector corresponding to λ_i, i.e., $Mx_i = \lambda_i x_i$.*

A few comments about this spectral theorem.

- This theorem and theorems like this are very important and many generalizations and variations of it exist.
- Note the wording: there are n vectors "such that x_i is an eigenvector corresponding to λ_i." In particular, there is no claim (yet) about uniqueness, etc. We still have to be careful about that.
- From this we can derive several other things, some of which we will mention below.

Someone asked in class about the connection with the SVD. The equations $Mx_i = \lambda_i x_i$, for all λ_i, can be written as $MX = X\Lambda$, or as $M = X\Lambda X^T$, since X is orthogonal. The SVD writes an arbitrary $m \times n$ matrix A a $A = U\Sigma V^T$, where U and V are orthogonal and Σ is diagonal and non-negative. So, the SVD is a generalization or variant of this spectral theorem for real-valued square matrices to general $m \times n$ matrices. It is *not* true, however, that the SVD of even a symmetric matrix gives the above theorem. It is true by the above theorem that you can write a symmetric matrix as $M = X\Lambda X^T$, where the eigenvectors Λ are real. But they might be negative. For those matrices, you also have the SVD, but there is no immediate connection. On the other hand, some matrices have all Λ positive/nonnegative. They are called SPD/SPSD matrices, and form them the eigenvalue decomposition of the above theorem essentially gives the SVD. (In fact, this is sometimes how the SVD is proven—take a matrix A and write the eigenvalue decomposition of the SPSD matrices AA^T and A^TA.) SPD/SPSD matrices are important, since they are basically covariance or correlation matrices; and several matrices we will encounter, e.g., Laplacian matrices, are SPD/SPSD matrices.

We can use the above lemma to provide the following variational characterization of eigenvalues, which will be very important for us.

Theorem 2 (Variational Characterization of Eigenvalues) *Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix; let $\lambda_l \leq \cdots \leq \lambda_n$ be its real eigenvalues, containing multiplicity and sorted in nondecreasing order; and let $x_1, \ldots, x_k$, for $k < n$ be orthonormal vectors such that $Mx_i = \lambda_i x_i$, for $i \in [k]$. Then*

$$\lambda_{k+1} = \min_{\substack{x\in\mathbb{R}^n\setminus\{\vec{0}\} \\ x\perp x_i \quad \forall i\in[k]}} \frac{x^TMx}{x^Tx},$$

and any minimizer of this is an eigenvector of λ_{k+1}.

Proof: First, by repeatedly applying the above lemma, then we get $n-k$ orthogonal eigenvectors that are also orthogonal to $x_1, \ldots, x_k$. Next, we claim that the eigenvalues of this system of n orthogonal eigenvectors include all eigenvalues of M. The proof is that if there were any other eigenvalues, then its eigenvector would be orthogonal to the other n eigenvectors, which isn't possible, since we already have n orthogonal vectors in $\mathbb{R}^n$.

Call the additional $n-k$ vectors $x_{k+1}, \ldots, x_n$, where x_i is an eigenvector of λ_i. (Note that we are inconsistent on whether subscripts mean elements of a vectors or different vectors themselves; but it should be clear from context.) Then, consider the minimization problem

$$\min_{\substack{x\in\mathbb{R}^n\setminus\{\vec{0}\} \\ x\perp x_i \quad \forall i\in[k]}} \frac{x^TMx}{x^Tx}$$

The solution $x \equiv x_{k+1}$ is feasible, and it has cost λ_{k+1}, and so $\min \leq \lambda_{k+1}$.

Now, consider any arbitrary feasible solution x, and write it as

$$x = \sum_{i=k+1}^{n} \alpha_i x_i.$$

The cost of this solution is

$$\frac{\sum_{i=k+1}^{n} \lambda_i \alpha_i^2}{\sum_{i=k+1}^{n} \alpha_i^2} \geq \lambda_{k+1} \frac{\sum_{i=k+1}^{n} \alpha_i^2}{\sum_{i=k+1}^{n} \alpha_i^2} = \lambda_{k+1},$$

and so $\min \geq \lambda_{k+1}$. By combining the above, we have that $\min = \lambda_{k+1}$.

Note that is x is a minimizer of this expression, i.e., if the cost of x equals λ_{k+1}, then $a_i = 0$ for all i such that $\lambda_i > \lambda_{k+1}$, and so x is a linear combination of eigenvectors of λ_{k+1}, and so it itself is an eigenvector of λ_{k+1}.

$\diamond$

Two special cases of the above theorem are worth mentioning.

- The leading eigenvector.

$$\lambda_1 = \min_{x \in \mathbb{R}^n \setminus \{\vec{0}\}} \frac{x^T M x}{x^T x}$$

- The next eigenvector.

$$\lambda_2 = \min_{x \in \mathbb{R}^n \setminus \{\vec{0}\}, x \perp x_1} \frac{x^T M x}{x^T x},$$

  where x_1 is a minimizer of the previous expression.

2.4 Some things that were skipped

Luca and Dan give two slightly different versions of the variational characterization and Courant-Fischer theorem, i.e., a min-max result, which might be of interest to present.

From wikipedia, there is the following discussion of the min-max theorem which is nice.

- Let $A \in \mathbb{R}^{n \times n}$ be a Hermitian/symmetric matrix, then the Rayleigh quotient $R_A : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$ is $R_A(x) = \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$, or equivalently $f_A(x) = \langle Ax, x \rangle : \|x\|_2 = 1$.
- Fact: for Hermitian matrices, the range of the continuous function $R_A(x)$ or $f_A(x)$ is a compact subset $[a, b]$ of $\mathbb{R}$. The max b and min a are also the largest and smallest eigenvalue of A, respectively. The max-min theorem can be viewed as a refinement of this fact.
- **Theorem 3** *If $A \in \mathbb{R}^{n \times n}$ is Hermitian with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_k \geq \cdots$, then*

$$\lambda_k = \max\{\min\{R_A(x) : x \in U, x \neq 0\}, dim(U) = k\},$$

  and also

$$\lambda_k = \min\{\max\{R_A(x) : x \in U, x \neq 0\}, dim(U) = n - k + 1\}.$$

- In particular,

$$\lambda_n \leq R_A(x) \leq \lambda_1,$$

  for all $x \in \mathbb{R}^n \setminus \{0\}$.
- A simpler formulation for the max and min is

$$\begin{aligned} \lambda_1 &= \max\{R_A(x) : x \neq 0\} \\ \lambda_n &= \min\{R_A(x) : x \neq 0\} \end{aligned}$$

Another thing that follows from the min-max theorem is the Cauchy Interlacing Theorem. See Spielman's 9/16/09 notes and Wikipedia for two different forms of this. This can be used to control eigenvalues as you make changes to the matrix. It is useful, and we may revisit this later.

And, finally, here is counterexample to these results in general. Lest one thinks that these niceness results always hold, here is a simple non-symmetric matrix.

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

(This is an example of a nilpotent matrix.

Definition 3 *A* nilpotent matrix *is a square matrix A such that $A^k = 0$ for some $k \in \mathbb{Z}^+$.*

More generally, any triangular matrix with all zeros on the diagonal; but it could also be a dense matrix.)

For this matrix A, we can define $R_A(x)$ as with the Rayleigh quotient. Then,

- The only eigenvalue of A equals 0.
- The maximum value of $R_A(x)$ is equal to $\frac{1}{2}$, which is larger that 0.

So, in particular, the Rayleigh quotient doesn't say much about the spectrum.

2.5 Summary

Today we showed that any symmetric matrix (e.g., adjacency matrix A of an undirected graph, Laplacian matrix, but more generally) is nice in that it has a full set of n real eigenvalues and a full set of n orthonormal eigenvectors.

Next time, we will ask what those eigenvectors look like, since spectral methods make crucial use of that. To do so, we will consider a different class of matrices, namely positive or nonnegative (not PSD or SPSD, but element-wise positive or nonnegative) and we will look at the extremal, i.e., top or bottom, eigenvectors.