Math 314

Topics for third exam

Technically, everything covered by the first two exams **plus**

Chapter 4: Eigenvalues

§1: **The beginning**

For A an n×n matrix, v is an *eigenvector* (e-vector, for short) for A if $v \neq 0$ and $Av = \lambda v$ for some (real or complex, depending on the context) number λ. λ is called the associated *eigenvalue* for A.

A matrix which has an eigenvector has *lots* of them; if v is an eigenvector, then so is $2v$, $3v$, etc. On the other hand, a matrix does not have lots of eigenvalues:

If λ is an e-value for A, then $(\lambda I - A)v$=0 for some non-zero vector v. So $\mathcal{N}(\lambda I - A) \neq \{0\}$, so $\det(\lambda I - A) = 0$. But $\det(tI - A) = p_A(t)$, thought of as a function of t, is a polynomial of degree n, so has at most n roots. So A has at most n different eigenvalues.

$p_A(t) = \det(tI - A)$ is called the *characteristic polynomial* of A.

$\mathcal{N}(\lambda I - A) = \mathcal{E}_\lambda(A)$ is (ignoring 0) the collection of all e-vectors for A with e-value λ. it is called the *eigenspace* (or e-space) for A corresponding to λ. An *eigensystem* for a (square) matrix A is a list of all of its e-values, along with their corresponding e-spaces.

One somewhat simple case: if A is (upper or lower) triangular, then the e-values for A are exactly the diagonal entries of A, since $tI - A$ is also triangular, so its determinant is the product of its diaginal entries.

We call $\dim\mathcal{N}(\lambda I - A)$ the *geometric multiplicity* of λ, and the number of times λ is a root of $p_A(t)$ (= number of times $(t - \lambda)$ is a factor) = m(λ) = the algebraic multiplicity of λ .

Some basic facts:

The number of real eigenvalues for an $n \times n$ matrix is $\leq n$.

counting multiplicity and complex root the number of eigenvalues $=n$.

For every e-value λ, 1$\leq$ the geometric multiplicity $\leq$ m(λ)

If the matrix A is symmetric (i.e., $A^T = A$), then every eigenvalue of A is a real number (i.e., every complex root of $p_A(t)$ is actually real).

§2: **Similarity and diagonalization**

TRhe basic idea: to understand a Markov chain $x_n = A^n x_0$, you need to compute large powers of A. This can be hard! There ought to be an easier way. Eigenvalues (or rather, eigenvectors) can help (if you have enough of them).

The matrix $A = \begin{pmatrix} 3 & 2 \\ 3 & 4 \end{pmatrix}$ has e-values 1 and 6 (Check!) with corresponding e-vectors (1,−1) and (2,3) . This then means that

$$\begin{pmatrix} 3 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} \text{, which we write } AP = PD \text{ ,}$$

where P is the matrix whose colummns are our e-vectors, and D is a diagonal matrix. Written slightly differently, this says $A = PDP^{-1}$.

We say two matrices A and B are *similar* if there is an invertible matrix P so that $AP = PB$. (Equivalently, $A = PBP^{-1}$, or $B = P^{-1}AP$.) A matrix A is *diagonalizable* if it is similar to a diagonal matrix.

Why do we care? It is easy to check that if $A = PBP^{-1}$, then $A^n = PB^nP^{-1}$. If B^n is easy to calculate (e.g., if B is diagonal; B^n is then also diagonal, and its diagonal entries are the powers of B's diagonal entries), this means A^n is also fairly easy to calculate!

Also, if A and B are similar, then they have the same characteristic polynomial, so they have the same eigenvalues. They do, however, have different eigenvectors; in fact, if $AP = PB$ and $Bv = \lambda v$, then $A(Pv) = \lambda(Pv)$, i.e., the e-vectors of A are P times the e-vectors of B .

These facts in turn tell us when a matrix can be diagonalized. Since for a diagonal matrix D, each of the standard basis vectors e_i is an e-vector, R^n has a basis consisting of e-vectors for D. If A is similar to D, via P, then each of $Pe_i = i$th column of P is an e-vector. But since P is invertible, its columns form a basis for R^n, as well. SO there is a basis consisting of e-vectors of A. On the other hand, such a basis guarantees that A is diagonalizable (just run the above argument in reverse...), so we find that:

(The Diagonalization Theorem) An $n \times n$ matrix A is diagonalizable if and only if there is basis of R^n consisting of eigenvectors of A.

And one way to guarantee that such a basis exists: If A is $n \times n$ and has n distinct eigenvalues, then choosing an e-vector for each will always yield a linear independent coillection of vectors (so, since there are n od them, you get a basis for R^n). So:

If A is $n \times n$ and has n distinct (real) eigenvalues, A is diagonalizable. In fact, the dimensions of all of the eigenspaces for A (for real eigenvalues λ) add up to n if and only if A is diagonalizable.

§2: **Discrete dynamical systems**

A discrete dynamical system (DDS) (= a system that moves in discrete steps) is a generalization of the Markov processes we studied before. It consists of an *initial state* x_0 and a transition matrix A . Starting at x_0, at every tick of the clock, we take the vector we are standing on and mutliply by A, so after n ticks, we are standing on $x_n = A^n x_0$.

The main question we wish to study is: what happens to x_n as n gets larger and larger? It turns out that this question has a fairly straightforward answer when A is diagonalizable. The answer depends upon the value of the *spectral radius* of A, $\rho(A)$, which is defined to be $\max\{|\lambda_i|\}$, where λ_i ranges over all of the e-values of A. In essence, it is the size of the 'largest' eigenvalue of A. Then we have:

If A is diagonalizable, and x_0 is an initial state, then

If $\rho(A) < 1$, then $||A^n x_0||$ goes to 0 as n goes to ∞ .

If $\rho(A) = 1$, then for some N, $||A^n x_0|| \leq N$ for all n .

If $\rho(A) = 1$, A has e-value 1, and every other e-value has absolute value less than 1, then $A^n x_0$ has a limit x_∞ as $n \to \infty$, and either $Ax_\infty = 0$ or $Ax_\infty = x_\infty$. (Usually, it equals x_∞ .)

If $\rho(A) > 1$, then for nearly every x_0, $||A^n x_0||$ goes to ∞ as n goes to ∞

A matrix A is called *defective* if for some e-value λ, $\dim\mathcal{N}(\lambda I - A) < \mathrm{m}(\lambda)$. It is fairly easy to show that a matrix is defective if and only if it is not diagonalizable (since the sum of dimensions of e-spaces will then be less than n).

What do we do if A isn't diagonalizable? Some of the statements (when $\rho(A) = 1$) fail to be true. But it turns out that the other two statements are true. This can be shown using *Jordan normal forms*

The idea is that being diagonalizable says that A is similar to a very simple matrix. It turns out that every matrix is similar to a 'kind of' simple matrix. A *Jordan block* $J_\lambda(k)$ is a $k \times k$ matrix most of whose entries are 0, except along the diagonal the entires are equal to λ, and just above the diagonal they are 1.

Every matrix is similar to a *block diagonal matrix*, i.e., a matrix whose entries are all 0 outside of a collection of square blocks whose diagonals sit on the main diagonal of A. Each block is a Jordan block, with possibly different λ's. This matrix is the *Jordan normal form* for A. It is unique, up to reordering the blocks on the diagonal.

We can still talk about the spectral radius $\rho(A)$ of a matrix, even if it isn't diagonalizable. With Jordan normal forms, it is possible to show that the first and last assertions of our theorem hold true, for every matrix A.

Chapter 5: Norms and inner products (again)

§1: **Norms**

We have found the notion of the length of a vector in R^n useful in several circumstances so far, now it is time to extend this concept to more of our favorite vector spaces!

The idea of this section is that our familiar notion of length satisfies some fairly natural properties. What we will now do is assert that any function satisfying those properties is something that we can reasonably called a notion of length, or a *norm*.

A *norm* on a vector space V is a function $||\cdot||:V \to R$ which satisfies:

(1) for every v in V, $||v|| \geq 0$, and $||v|| = 0$ if and only if $v = 0$

(2) for every v in V and c in R, $||c \cdot v|| = |c| \cdot ||v||$

(3) for every v and w in V, $||v + w|| \leq ||v|| + ||w||$ (Triangle Inequality)

The pair (V,$||\cdot||$) is called a *normed linear space*.

For example, on R^n there are lots of different norms: for every $p \geq 1$, the function

$$||v||_p = (|v_1|^p + \cdots + |v_n|^p)^{1/p}$$

is a norm, called the p-norm . There is a similar norm for '$p = \infty$':

$$||v||_\infty = \max\{|v_1|, \ldots, |v_n|\}$$

Also, for $C[a, b]$ = the cts fcns from $[a, b]$ to R,

$$||f|| = \int_a^b |f(x)| \, dx$$

is a norm. For many of these, especially the p-norms, proving the triangle inequality takes some work!

With a norm we can talk about convergence: $v_n \to v$ as $n \to \infty$ means (as with the usual norm) that $||v_n - v|| \to 0$ as $n \to \infty$.

We can also talk about the ball of radius r around a vector v; it is all of the vectors w with $||w - v|| < r$.

§2: **Inner products**

Just as with norms, we can adapt our notion of an innner product $< \cdot, \cdot >$ to more general vector spaces, by taking some of its familiar properties and making these a definition of an inner product!

An *inner product* on a vector space V is a function $< \cdot, \cdot >$ which takes pairs of vectors and hands you a number, which satisfies:

(1) for every v in V, $< v, v > \geq 0$, and $< v, v >= 0$ if and only if $v = 0$

(2) for every v and w in V, $< v, w > = < w, v >$

(3) for every v and w in V, and c in R, $< cv, w > = c < v, w >$

(4) for every u, v, and w in V, $< u + v, w > = < u, w > + < v, w >$

The pair (V,$< \cdot, \cdot >$) is called an *inner product space*.

Again, it turns out that there are lots of inner products on R^n, besides the usual one. For example, on R^2, $< v, w >= 2v_1w_1 + 5v_2w_2$ is an inner product; you can check that the four properties hold. More generally, for any invertible $n \times n$ matrix A, the function

$$< v, w >_A = < Av, Aw > = v^T(A^TA)w$$

is an inner product on R^n. On $C[a, b]$,

$$< f, g > = \int_a^b f(x)g(x) \, \mathrm{d}x$$

is an inner product.

It turns out that every inner product on V can be used to define a norm on V, by doing what we know is true for the usual norm and inner product:

Define $||v|| = (< v, v >)^{1/2}$. Property (1) for an inner product implies that property (1) for a norm holds; property (3) for an inner product implies property (2) for a norm holds; and finally, property (3) for this norm hold because

$$(< v, w >)^2 \leq < v, v >< w, w >$$

This is our (old) Schwartz inequality; but a look at the reasons why this was true for the ordinary inner product will convince you that all we need to know was the properties (1)-(4) for the inner product. So our argument there carries over to this more general setting without any change!

So every inner product can be used to define a norm. But not every norm comes from an inner product! There are several properties (for example, $||u + v||^2 + ||u - v||^2 = 2||u||^2 + 2||v||^2$) which one can show always hold, if your norm comes from an inner product! By evaluating both sides suing specific vectors, however, one can show that such equalities don't hold, showing that the norms in question do not come from inner products!

Just as with the ordinary inner product, we say that two vectors v and w are *orthogonal* if $< v, w >$=0.

If the vectors $v_1, \ldots, v_n$ are all non-zero and all orthogonal to one another, and v is in the span of the v_i's then it is easy to show that

$$v = \frac{< v_1, v >}{< v_1, v_1 >}v_1 + \cdots + \frac{< v_n, v >}{< v_n, v_n >}v_n$$

In fact,m this is the only way to write v as a linear combination of the v_i's, implying that the v_i's are linearly independent!