Math 314/814

Topics for second exam

Technically, everything covered by the first exam **plus**

Subspaces, bases, dimension, and rank

Basic idea: $W \subseteq \mathbb{R}^n$ is a subspace if whenever $c \in \mathbb{R}$ and $u, v \in W$, we **always** have $cu, u+v \in W$ (W is "closed" under addition and scalar multiplication).

Examples: $\{(x,y,z) \in \mathbb{R}^3 : z = 0\}$ is a subspace of $\mathbb{R}^3$
$\{(x,y,z) \in \mathbb{R}^3 : z = 1\}$ is **not** a subspace of $\mathbb{R}^3$

Basic construction: $v_1, \cdots, v_n \in V$
$W = \{a_1v_1 + \cdots a_nv_n : a_1, \ldots, a_n \in \mathbb{R}$ = all linear combinations of $v_1, \cdots, v_n = \text{span}\{v_1, \cdots, v_n\}$
= the *span* of $v_1, \cdots, v_n$, is a subspace of $\mathbb{R}^k$

Basic fact: if $w_1, \ldots, w_k \in \text{span}\{v_1, \cdots, v_n\}$, then $\text{span}\{w_1, \cdots, w_k\} \subseteq \text{span}\{v_1, \cdots, v_n\}$

Subspaces from matrices
column space of $A = \text{col}(A) = \text{span}\{\text{the columns of } A\}$
row space of $A = \text{row}(A) = \text{span}\{(\text{transposes of the }) \text{ rows of } A\}$
nullspace of $A = \text{null}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$

(Check: null(A) is a subspace!)

Alternative view Ax = lin comb of columns of A, so is in col(A); in fact, $\text{col}(A) = \{Ax : x \in \mathbb{R}^n\}$. So col($A$) is the set of vectors b for which $Ax = b$ has a solution. Any two solutions $Ax = b = Ay$ have $A(x-y) = AX - Ay = b - b = 0$, so $x - y$ is in null(A). So the collection of <u>all</u> solutions to $AX = b$ are (particular solution)+(vector in null(A)). So col(A) tells is which SLEs have solutions, and null(A) tells us how many solutions there are.

Bases:

A **basis** for a subspace V is a set of vectors $v_1, \ldots, v_n$ so that (a) they are linearly independent, and (b) V=span$\{v_1, \ldots, v_n\}$.
The idea: a basis allows you to express every vector in the subspace as a linear combination <u>in exactly one way</u>.

A system of equations $Ax = b$ has a solution iff $b \in$col(A) .

If $Ax_0 = b$, then every other solution to $Ax = b$ is $x = x_0 + z$, where $z \in$null(A) .

The row, column, and nullspaces of a matrix A are therefore useful spaces (they tell us useful things about solutions to the corresponding linear system), so it is useful to have bases for them.

Finding a basis for the row space.
Basic idea: if B is obtained from A by elementary row operations, then row(A) =row(B).
So of R is the reduced row echelon form of A, row(R) =row(A)
But a basis for row(R) is quick to identify; take all of the non-zero rows of R ! (The zero rows are clearly redundant.) These rows are linearly independent, since each has a 'special coordinate' where, among the rows, only it is non-zero. That coordinate is the *pivot* in that row. So in any linear combination of rows, only that vector can contribute something non-zero to that coordinate. *Consequently*, in any linear combination, that coordinate is the **coefficient** of our vector! **So**, if the lin comb is $\vec{0}$, the coefficient of our vector (i.e., of each vector!) is 0.

Put bluntly, to find a basis for row(A), row reduce A, to R; the (transposes of) the non-zero rows of R form a basis for row(A).

This in turn gives a way to find a basis for col(A), since col(A) =row(A^T) !

To find a basis for col(A), take A^T, row reduce it to S; the (transposes of) the non-zero rows of S form a basis for row(A^T) =col(A) .

This is probably in fact the most useful basis for col(A), since each basis vector has that special coordinate. This makes it very quick to decide if, for any given vector b, $Ax = b$ has a solution. You need to decide if b can be written as a linear combination of your basis vectors; but each coefficient will be the coordinate of b lying at the special coordinate of each vector. Then just check to see if **that** linear combination of your basis vectors adds up to b !

There is another, perhaps less useful, but faster way to build a basis for col(A); row reduce A to R, locate the pivots in R, and take the columns of A (Note: A, **not** R !) that correspond to the columns containing the pivots. These form a (different) basis for col(A).

Why? Imagine building a matrix B out of just the pivot columns. Then in row reduced form there is a pivot in every column. Solving $Bv = \vec{0}$ in the case that there are no free variables, we get $v = \vec{0}$, so the columns are linearly independent. If we now add a free column to B to get C, we get the same collection of pivots, so our added column represents a free variable. Then there are non-trivial solutions to $Cv = \vec{0}$, so the columns of C are not linearly independent. This means that the added columns can be expressed as a linear combination of the bound columns. This is true for all free columns, so the bound columns span col(A).

Finally, there is the nullspace null(A). To find a basis for null(A):

Row reduce A to R, and use each row of R to solve $Rx = \vec{0}$ by expressing each bound variable in terms of the frees. collect the coefficients together and write $x = x_{i_1}v_1 + \cdots + x_{i_k}v_k$ where the x_{i_j} are the free variables. Then the vectors $v_1, \ldots, v_k$ form a basis for null(A).

Why? By construction they span null(A); and just as with our row space procedure, each has a special coordinate where only it is not 0 (the coordinate corresponding to the free variable!).

Note: since the number of vectors in the bases for row(A) and col(A) is the same as the number of pivots (= number of nonzero rows in the RREF) = rank of A, we have dim(row(A))=dim(col(A))=$r(A)$. And since the number of vectors in the basis for null(A) is the same as the number of free variables for A (= the number of columns without a pivot) = nullity of A (hence the name!), we have dim(null(A)) = $n(A) = n - r(A)$ (where n=number of columns of A).
So, dim(col(A)) + dim(null(A)) = the number of columns of A .

More on Bases.

A **basis** for a subspace V of $\mathbb{R}^k$ is a set of vectors $v_1, \ldots, v_n$ so that (a) they are linearly independent, and (b) V=span$\{v_1, \ldots, v_n\}$.
Example: The vectors $e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ are a basis for $\mathbb{R}^n$, the *standard basis.*

To find a basis: start with a collection of vectors that span, and repeatedly throw out redundant vectors (so you don't change the span) until the ones that are left are linearly independent. Note: each time you throw one out, you need to ask: are the remaining ones lin indep?

Basic fact: If $v_1, \ldots, v_n$ is a basis for V, then every $v \in V$ can be expressed as a linear combination of the v_i's in *exactly one way.* If $v = a_1v_1 + \cdots + a_nv_n$, we call the a_i the **coordinates** of v with respect to the basis $v_1, \ldots, v_n$. We can then <u>think</u> of v as the vector $(a_1, \ldots a_n)^T$ = the <u>coordinates</u> of v with respect to the basis $v_1, \ldots, v_n$, so we can think of V as "really" being $\mathbb{R}^n$.

The Basis Theorem: Any two bases of the same vector space contain the same number of vectors. (This common number is called the *dimension* of V, denoted dim(V) .)

Reason: if $v_1, \dots, v_n$ is a basis for V and $w_1, \dots, w_k \in V$ are linearly independent, then $k \leq n$

As part of that proof, we also learned:
If $v_1, \dots, v_n$ is a basis for V and $w_1, \dots, w_k$ are linearly independent, then the spanning set $v_1, \dots, v_n, w_1, \dots, w_k$ for V can be thinned down to a basis for V by throwing away v_i's .

In reverse: we can take any linearly independent set of vectors in V, and **add** to it from any basis for V, to produce a new basis for V.

Some consequences:
If dim(V)=n, and $W \subseteq V$ is a subspace of V, then dim(W)$\leq n$

If dim(V)=n and $v_1, \dots, v_n \in V$ are linearly independent, then they also span V
If dim(V)=n and $v_1, \dots, v_n \in V$ span V, then they are also linearly independent.

Linear Transformations.

$T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if $T(cu + dv) = cT(u) + dT(v)$ for all $c, d \in \mathbb{R}$, $u, v \in \mathbb{R}^n$. This can be verified in two steps: check $T(cu) = cT(u)$ for all $c \in \mathbb{R}$ and $u \in \mathbb{R}^n$, <u>and</u> $T(u + v) = T(u) + T(v)$ for all $u, v \in \mathbb{R}^n$.

Example: $T_A : \mathbb{R}^n \to \mathbb{R}^m$, $T_A(v) = Av$, is linear
T : {functions defined on $[a, b]$} $\to \mathbb{R}$, $T(f) = f(b)$, is linear

$T : \mathbb{R}^2 \to \mathbb{R}$, $T(x, y) = x - xy + 3y$ is **not** linear!

Basic fact: <u>every</u> linear transf $T : \mathbb{R}^n \to \mathbb{R}^m$ is $T = T_A$ for some matrix A: A = the matrix with i-th column $T(e_i)$, e_i = the i-th coordinate vector in $\mathbb{R}^n$.

Using the idea of coordinates for a subspace, we can extend these notions to linear transformations $T : V \to W$; thinking of vectors as their coordinates, each T is "really" T_A for some matrix A.

The composition of two linear transformations is linear; in fact $T_A \circ T_B = T_{AB}$. So matrix multiplication is really function composition!
If A is invertible, then T_A is an invertible function, with inverse $T_A^{-1} = T_{A^{-1}}$.

Application: Markov chains.

In many situations we wish to study a characteristic (or characteristics) of a population, which changes over time. Often the rule for how the quantities change may be linear; our goal is to understand what the long term behavior of the situation is.

Initially, the characteristic of the population (think: favorite food, political affiliation, choice of hair color) is distributed among some collection of values; we represent this *initial state* as a vector $\vec{v}_0$ giving the fraction of the whole population which takes each value. (The entries of $\vec{v}_0$ sum to 1.) As time progresses, with each fixed time interval the distribution of the population changes by multiplication by a *transition matrix* A, whose (i, j) entry $a_{i,j}$ records what fraction of the population having the i-th characteristic chooses to switch to the j-th characteristic. Since every person/object in the population ends up with some characteristic, each column of A (which describes how the i-th charactersitc gets redistributed) must sum to 1. After the tick of the clock, the distribution of our initial population, given by $\vec{v}_0$, changes to $\vec{v}_1 = A\vec{v}_0$. After n ticks of the clock, the population districution is given by $\vec{v}_n = A^n\vec{v}_0$.

The main questions to answer are: does the population distribution stabilize over time? And if so, what does it stabilize to? A stable distribution $\vec{v}$ is one which is unchanged as time progresses: $A\vec{v} = \vec{v}$. This can be determined by reinterpreting stability as $(A - I)\vec{v} = \vec{0}$, i.e., $\vec{v}$ lies in the nullspace of the matrix $A - I$. Which we can compute! Our solution should also have all entries non-negative (to reflect that its entries represent parts of a whole) and add up to 1.

It is a basic fact that every transition vector has a stable solution. This is because $A^T\vec{x} = \vec{x}$, where $\vec{x}$ is the all-1's vector. So $A^T - I = (A - I)^T$ is not invertible, so $A - I$ is not invertible! Moreover, under very mild assumptions (e.g., no entry of A is 0) every initial state $\vec{v}_0$, under repeated multiplication by A, will converge to the exact same stable distribution. Which we can compute by the method above!

Application: counting paths in a graph.

A graph is a collection of points (= *vertices*) $\{V_1, \dots, V_n\}$ which are connected (in pairs) by *edges*. We can encode a graph using an $n \times n$ matrix, its *incidence matrix* A, whose $(i,j)^{th}$ entry is the number of edges between V_i and V_j. (Note: this matrix is symmetric). If we work with a *directed* graph, in which edges have a direction, our incidence matrix counts the number of edges going from V_i to V_j (this need not be symmetric).

The incidence matrix $A = (a_{ij})$ can be used to compute the number of distinct paths of any fixed length (length = the number of edges (not necessarilty distinct!) that we cross) running from one vertex to another (or the same vertex!). This is because the collection of all length-two paths from V_i to V_j can be partitioned according to which vertex is in the middle, and the number of such paths with middle vertex V_k is $a_{ik}a_{kj}$. But the sum of these is the $(i,j)^{th}$ entry of A^2 ! By a similar line of reasoning, we can establish that the number of distinct length k paths from V_i to V_j is equal to the $(i,j)^{th}$ entry of A^k. If we are working with the incidence matrix of a directed graph, this number is the number of distinct length k directed paths from V_i to V_j (that is, the directions all match up, always pointing in the forward direction).

Such calculations are routinely used in many applications, for example, in physics, where a 'path integral' averages quantities over all paths (in a crystal lattice, for example) between a pair of points. We can also extract useful theoretical information about a graph, for example, the number of triangles (loops of length 3) is equal to the sum of the diagonal entries of A^3 (which counts the number of paths of length 3 that start and end at the same vertex) times one-third (since we count each triangle 3 times, one for each possible starting point!).

Chapter 4: Eigenvalues, eigenvectors, and determinants

Determinants.

(Square) matrices come in two flavors: invertible (all $Ax = b$ have a solution) and non-invertible ($Ax = \vec{0}$ has a non-trivial solution). It is an amazing fact that one number identifies this difference; the determinant of A.

For 2×2 matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, this number is det$(A)$=$ad - bc$; if $\neq 0$, A is invertible, if =0, A is non-invertible (=singular).
For larger matrices, there is a similar (but more complicated formula):
A= $n \times n$ matrix, $M_{ij}(A)$ = matrix obtained by removing ith row and jth column of A.
$\det(A) = \Sigma_{i=1}^{n}(-1)^{i+1}a_{i1}\det(M_{i1}(A))$
(this is called expanding along the first column)

Amazing properties:
If A is upper triangular, then $\det(A)$ = product of the entries on the diagonal
If you multiply a row of A by c to get B, then $\det(B) = c\det(A)$
If you add a mult of one row of A to another to get B, then $\det(B) = \det(A)$
If you switch a pair of rows of A to get B, then $\det(B) = -\det(A)$

In other words, we can understand exactly how each elementary row operation affects the determinant. In part, A is invertible iff $\det(A) \neq 0$.

In fact, we can **use** row operations to calculate $\det(A)$ (since the RREF of a matrix is upper triangular). We just need to *keep track* of the row operations we perform, and compensate for the changes in the determinant;
$\det(A) = (1/c)\det(E_i(c)A)$, $\det(A) = (-1)\det(E_{ij}A)$

More interesting facts:
$\det(AB) = \det(A)\det(B)$; $\det(A^T) = \det(A)$; $\det(A^{-1}) = [\det(A)]^{-1}$

We can expand along other columns than the first: for any fixed value of j (= column),
$\det(A) = \Sigma_{i=1}^{n}(-1)^{i+j}a_{ij}\det(M_{ij}(A))$
(expanding along jth column)
And since $\det(A^T) = \det(A)$, we could expand along **rows**, as well.... for any fixed i (= row),
$\det(A) = \Sigma_{j=1}^{n}(-1)^{i+j}a_{ij}\det(M_{ij}(A))$

A formula for the inverse of a matrix:
If we define A_c to be the matrix whose (i, j)th entry is $(-1)^{i+j}\det(M_{ij}(A))$, then $A_c^T A = (\det A)I$ (A_c^T is called the *adjoint* of A). So if $\det(A) \neq 0$, then we can write the inverse of A as

$A^{-1} = \dfrac{1}{\det(A)}A_c^T$ (This is very handy for 2×2 matrices...)

The same approach allows us to write an explicit formula for the solution to $Ax = b$, when A is invertible:
If we write $B_i = A$ with its ith column replaced by b, then the (unique) solution to $Ax = b$ has ith coordinate equal to
$\dfrac{\det(B_i)}{\det(A)}$

Eigenvectors and Eigenvalues.

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 <u>not</u> 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 null$(\lambda I - A) \neq \{0\}$, so $\det(\lambda I - A) = 0$. But $\det(tI - A) = \chi_A(t)$, thought of as a function of t, is a polynomial of degree n, so has <u>at</u> <u>most</u> n roots. So A has at most n different eigenvalues.
$\chi_A(t) = \det(tI - A)$ is called the *characteristic polynomial* of A.
null$(\lambda I - A) = E_\lambda(A)$ 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 <u>exactly</u> the diagonal entries of A, since $tI - A$ is also triangular, so its determinant is the product of its diaginal entries.
We call dim(null$(\lambda I - A)$) the *geometric multiplicity* of λ, and the number of times λ is a root of $\chi_A(t)$ (= 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 roots the number of eigenvalues $=n$.
For every e-value λ, 1$\leq$ the geometric multiplicity $\leq$ m(λ).
(non-zero) e-vectors having all different e-values are linearly independent.

Similarity and diagonalization

The 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

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

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, $P^{-1}AP = B$, or $A = PBP^{-1}$.) A matrix A is *diagonalizable* if it is similar to a diagonal matrix.

We write $A \sim B$ is A is similar to B, i.e., $P^{-1}AP = B$. We can check:
$A \sim A$; if $A \sim B$ then $B \sim A$; if $A \sim B$ and $B \sim C$, then $A \sim C$. (We sat that "$\sim$" is an *equivalence relation*.)

Why do we care about similarity? We can check that if $A = PBP^{-1}$, then $A^n = PB^nP^{-1}$. If B^n is quick 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 <u>also</u> fairly quick 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 . Similar matrices also have the same determinant, rank, and nullity.

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 <u>always</u> yield a linear independent coillection of vectors (so, since there are n of 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.