Math 314/814 Section 5
Topics for second exam

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

More on Bases.

A **basis** for a subspace V of $\mathbb{R}^k$ is a set of vectors $v_1, \dots, v_n$ so that (a) they are linearly independent, and (b) V=span$\{v_1, \dots, v_n\}$.
Example: The vectors $e_1 = (1, 0, \dots, 0), e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$ 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 $v_1, \dots, v_n$ 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 $v_1, \dots, v_n$. We can then <u>think</u> of v as the vector $(a_1, \dots a_n)^T$ = the <u>coordinates</u> of v with respect to the basis $v_1, \dots, v_n$, 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 $T_A^{-1} = T_{A^{-1}}$.

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

$A^{-1} = \frac{1}{\det(A)}A_c^T$ (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

$$\frac{\det(B_i)}{\det(A)}$$

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 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 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 at most 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 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(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 also 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 always 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.

Chapter 5: Orthogonality.

Length and inner product.

"Norm" means length! In $\mathbb{R}^n$ this is computed as $||x|| = ||(x_1, \dots, x_n)|| = (x_1^2 + \dots + x_n^2)^{1/2}$

Basic facts: $||x|| \geq 0$, and $||x|| = 0$ iff $x = \vec{0}$,
$||cu|| = |c| \cdot ||u||$, and $||u + v|| \leq ||u|| + ||v||$ (triangle inequality)

unit vector: the norm of $u/||u||$ is 1; $u/||u||$ is the *unit vector* in the direction of u.

Inner product:
idea: assign a number to a pair of vectors (think: angle between them?)
In $\mathbb{R}^n$, we use the *dot product*: $v = (v_1, \dots, v_n)$, $w = (w_1, \dots, w_n)$
$v \bullet w = \langle v, w \rangle = v_1 w_1 + \dots + v_n w_n = v^T w$

Basic facts:
$\langle v, v \rangle = ||v||^2$ (so $\langle v, v \rangle \geq 0$, and equals 0 iff $v = \vec{0}$)
$\langle v, w \rangle = \langle w, v \rangle$; $\langle cv, w \rangle = \langle v, cw \rangle = c\langle v, w \rangle$ $\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle$

Orthogonality.

Two vectors are *orthogonal* if their angle is $\pi/2$, i.e., $\langle v, w \rangle$=0. Notation: $v \perp w$. (Also say they are *perpendicular*.)

A collection of vectors $\{v_1, \dots, v_k\}$ is an *orthogonal set* if $v_i \perp v_j$ for every $i \neq j$. If all of the vectors in an orthogonal set are non-zero, then they are linearly independent.
An *orthogonal basis* for a subspace W is basis for W that is also an orthogonal set. If we have an orthogonal basis $v_1, \dots, v_k$ for W, then determining the coordinates for a vector $w \in W$ is quick: $w = \Sigma a_i v_i$ for $a_i = \langle w, v_i \rangle / ||v_i||^2$.

A collection of vectors $\{v_1, \dots, v_k\}$ is an *orthonormal set* (we write "o.n. set") if they are an orthogonal set and $||v_i|| = 1$ for every i. An *orthonormal basis* (o.n. basis) is a basis that is also an orthonormal set. For an o.n. basis for W, the coordinates of $w =\in W$ are even shorter: $w = \Sigma a_i v_i$ for $a_i = \langle w, v_i \rangle$.

Orthogonal Matrices.

We've seen that having a basis consisting of orthonormal vectors can simplify some of our previous calculations. Now we'll see where some of them come from.
An $n \times n$ matrix Q is called **orthogonal** if it's columns form an orthonormal basis for R^n. This means $<$(ith column of Q),(jth column of Q¿ $= 1$ if $i = j$, and is 0 otherwise . This in turn means that $Q^T Q = I$, which in turn means $Q^T = Q^{-1}$! So an orthogonal matrix is one whose inverse is equal to its own transpose.

Basic facts about an orthogonal matrix Q :
Q is orthogonal $\Leftrightarrow$ for every $v, w \in R^n$, $< Qv, Qw > = < v, w >$.

Q is orthogonal $\Leftrightarrow$ for every $v \in R^n$, $||Qv|| = ||v||$.
If Q is orthogonal, then Q^T is orthogonal. So the rows of Q form an o.n. basis!
If Q is orthogonal, then Q^{-1} is orthogonal.
If Q is orthogonal, then $\det(Q) = \pm 1$.

Orthogonal Complements.

This notion of orthogonal vectors can even be used to reinterpret some of our dearly-held results about systems of linear equations, where all of this stuff began.

Starting with $Ax = 0$, this can be interpreted as saying that $<$(every row of A),$x >$=0, i.e., x is orthogonal to every row of A. This in turn implies that x is orthogonal to every linear combination of rows of A, i.e., x is orthogonal to every vector in the row space of A.
This leads us to introduce a new concept: the **orthogonal complement** of a subspace W in a vector space V, denoted $W^\perp$, is the collection of vectors v with $v \perp w$ for **every** vector $w \in W$. It is not hard to see that these vectors form a subspace of V; the sum of two vectors orthogonal to w, for example, is orthogonal to w, so the sum of two vectors in $W^\perp$ is also in $W^\perp$. The same is true for scalar multiples.

Some basic facts:
For every subspace W, $W \cap W^\perp = \{0\}$ (since anything in both is orthogonal to *itself*, and only the 0-vector has that property).
Finally, $(W^\perp)^\perp = W$; this is because W is contained in $(W^\perp)^\perp$ (a vector in W is orthogonal to every vector that is orthogonal to things in W), and the dimensions of the two spaces are the same.

The importance that this has to systems of equations stems from the following facts:
$\text{null}(A) = \text{row}(A)^\perp$
$\text{row}(A) = \text{null}(A)^\perp$
$\text{col}(A) = \text{null}(A^T)^\perp$

This provides us with yet another (quicker?) way to decide if a system of equations $A\vec{x} = \vec{b}$ is consistent, or rather, for *which* $\vec{b}$ is it consistent; $\vec{b}$ must lie in col(A), i.e., in $\text{null}(A^T)^\perp$. So it must be $\perp$ to a basis for null(A^T). So we can compute a basis for null(A^T), $v_1 \ldots, v_k$, and use this to check for consistency: $A\vec{x} = \vec{b}$ is consistent $\Leftrightarrow \langle \vec{b}, v_i \rangle = 0$ for every v_i.

And to compute a basis for $W^\perp$: start with a basis for W, writing them as the columns of a matrix A, so $W = \text{col}(A)$, then $W^\perp = \text{col}(A)^\perp = \text{row}(A^T)^\perp = \text{null}(A^T)$, which we know how to compute a basis for!

Orthogonal Projections.

Any vector $v \in V$ can be written, uniquely, as $v = w + w^\perp$, for $w \in W$ and $w^\perp \in W^\perp$; the uniqueness comes from the result above about intersections. That it can be written that way at all comes from orthogonal projections.
We've seen that if $w_1, \ldots, w_k$ is an orthogonal basis for a subspace W of $\mathbb{R}^n$, and $w \in W$, then $w = \dfrac{< w_1, w >}{< w_1, w_1 >} w_1 + \ldots + \dfrac{< w_k, w >}{< w_{k-1}, w_k >} w_k$
On the other hand, if $v \in V$, we can define the orthogonal projection

$$\text{proj}_W(v) = \frac{< w_1, v >}{< w_1, w_1 >} w_1 + \ldots + \frac{< w_k, v >}{< w_k, w_k >} w_k$$

of v into W. This vector is in W, and we can show that $v - \text{proj}_W(v)$ is orthogonal to all of the w_i, so it is orthogonal to every linear combination, i.e., it is orthogonal to every vector in W. So $v - \text{proj}_W(v) = w' \in W^\perp$

In the case that the w_i are not just orthogonal but also *orthnormal*, we can simplify this somewhat:

$\text{proj}_W(v) = < w_1, v > w_1 + \cdots + < w_n, v > w_n = (w_1 w_1^T + \cdots + w_n w_n^T)v = Pv$,
where $P = (w_1 w_1^T + \cdots + w_n w_n^T)$ is the **projection matrix** giving us orthogonal projection.

For any subspace $W \subseteq \mathbb{R}^n$, $\dim(W)+\dim(W^\perp) = n = \dim(\mathbb{R}^n)$. Even more, a basis for W and a basis for $W^\perp$ together form a basis for $\mathbb{R}^n$.

All that we need now is a method for building orthogonal bases for subspaces!

Gram-Schmidt Orthogonalization.

We've seen how a basis consisting of vectors orthogonal to one another can prove useful; this section is about how to *build* such a basis.
The starting point is our old formula for the projection of one vector onto another; If $W = \text{span}\{w\}$, then orthogonal projection onto W is given by
$\dfrac{< w, v >}{< w, w >}w$, so $v - \dfrac{< w, v >}{< w, w >}w$ is perpendicular to w.
Gram-Schmidt orthogonalization consists of repeatedly using this formula to replace a collection of vectors with ones that are orthogonal to one another, **without changing their span**. Starting with a collection $\{v_1, \ldots, v_n\}$ of vectors in V,

let $w_1 = v_1$, then let $w_2 = v_2 - \frac{<w_1,v_2>}{<w_1,w_1>}w_1$.
Then w_1 and w_2 are orthogonal, and since w_2 is a linear combination of $w_1 = v_1$ and v_2, while the above equation can also be rewritten to give v_2 as a linaear combination of w_1 and w_2, the span is unchanged. Continuing,
let $w_3 = v_3 - \dfrac{< w_1, v_3 >}{< w_1, w_1 >}w_1 - \dfrac{< w_2, v_3 >}{< w_2, w_2 >}w_2$; then since w_1 and w_2 are orthogonal, it is not hard to check that w_3 is orthogonal to **both** of them, and using the same argument, the span is unchanged (in this case, $\text{span}\{w_1, w_2, w_3\}$ $=\text{span}\{w_1, w_2, v_3\}$$=\text{span}\{v_1, v_2, v_3\}$).
Continuing this, we let $w_k = v_k - \dfrac{< w_1, v_k >}{< w_1, w_1 >}w_1 - \ldots - \dfrac{< w_{k-1}, v_k >}{< w_{k-1}, w_{k-1} >}w_{k-1}$
Doing this all the way to n will replace $v_1, \ldots, v_n$ with orthogonal vectors $w_1, \ldots, w_n$, without changing the span.

One thing worth noting is that the if two vectors are orthogonal, then any scalar multiples of them are, too. This means that if the coordinates of one of our w_k are not to our satisfaction (having an ugly denomenator, perhaps), we can scale it to change the coordinates to something more pleasant. It is interesting to note that in so doing, the the later vectors w_k are unchanged, since our scalar, can be pulled out of both the top inner product and the bottom one in later calculations, and cancelled.