Day 7 Tue 2/6/2018

Questions about anything?
Collect formal ps1 from Matthew and Alaeldin.

Continue material from prior lecture. This outline will probably extend over two lectures.

1. Review eigenvalues and eigenvectors definition
   a. def of eigenvalue; eig vector; need to find eigvalues first, then vectors.
   b. Final idea from last time: λ is an eigenvalue iff $\det(A - \lambda I) = 0$ iff $\det(\lambda I - A) = 0$
   c. Solve equation to find all the eigenvalues. Substitute each eigenvalue into the equation and solve resulting homogeneous matrix equation to find all the corresponding eigenvalues
2. Example: $A = \begin{bmatrix} 1 & -15 & -3 \\ -6 & 28 & 6 \\ 30 & -150 & -32 \end{bmatrix}$. Equation $\det(\lambda I - A) = 0$ reduces to $(\lambda - 1)(\lambda + 2)^2 = 0$. This shows the eigenvalues are 1, -2, -2. To find eigenvectors for $\lambda = 1$, we have to solve the homogeneous system $(I - A)\mathbf{v} = \mathbf{0}$. So find the reduced row echelon form for $I - A$. This shows that the eigenvectors are expressed in the form $c\,[1\ \ -2\ \ 10]^T$ where c can be any scalar. For $\lambda = -2$, we have to solve the homogeneous system $(-2I - A)\mathbf{v} = \mathbf{0}$. , or equivalently for $(A + 2I)\mathbf{v} = \mathbf{0}$. So this time find the reduced row echelon form for $A + 2I$. This time there are *two* free variables and the general solution has the form
   $\mathbf{v} = b[5\ \ 1\ \ 0]^T + c[1\ \ 0\ \ 1]^T$, where b and c can be any scalars. Notice that this time we actually have two linearly independent eigenvectors (and all their combinations). It is not a coincidence that this happens for the repeated eigenvalue. A general theorem says that the maximum number of independent eigenvectors for any eigenvalue cannot exceed the multiplicity of the eigenvalue as a root of the equation $\det(\lambda I - A) = 0$. In this example the number of independent eigenvectors equals the multiplicity. But in other cases it can be strictly less than the multiplicity. See next example.

5. characteristic polynomial: $\det(\lambda I - A)$; characteristic equation: $\det(\lambda I - A) = 0$. Eigenvalues are roots of the characteristic polynomial, which has degree n for an $n \times n$ matrix. The multiplicity of the eigenvalue is its multiplicity as a root. There can be at most n distinct eigenvalues for an $n \times n$ matrix.

3. Example: $A = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}$. Here, using what we know about determinants of triangular matrices, we can see by inspection that the characteristic polynomial is $(\lambda - 5)^3$. So there is only one eigenvalue, 5, and it has multiplicity 3. To find eigenvectors, we have to solve the homogeneous system $(5I - A)\mathbf{v} = \mathbf{0}$ (or equivalently, $(A - 5I)\mathbf{v} = \mathbf{0}$,), and this time the matrix $A - 5I$ is already in reduced row echelon form. It tells us that the eigenvectors can all be expressed in the form $c[1\ \ 0\ \ 0]^T$. There is only one independent eigenvector even though the multiplicity of the eigenvalue is 3.

4. Example: $W = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.
   a. This is a permutation matrix for a cyclic permutation, so $W^3 = I$. This *implies* that the characteristic polynomial has to be $\lambda^3 - 1$, though we won't see a proof just yet.
   b. Direct computation: $\det(\lambda I - W) = \det \begin{bmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -1 & 0 & \lambda \end{bmatrix} = \lambda \begin{vmatrix} \lambda & -1 \\ 0 & \lambda \end{vmatrix} + \begin{vmatrix} 0 & -1 \\ -1 & \lambda \end{vmatrix} = \lambda^3 - 1.$
   c. This factors as $(\lambda - 1)(\lambda^2 + \lambda + 1)$ so the eigenvalues are 1 and $\frac{-1 \pm i\sqrt{3}}{2}$. These are called the complex third (or cube) roots of unity.

d. Only one real eigenvalue, but we have three complex eigenvalues, and note that they are distinct. Also note that the nonreal roots are complex conjugates of each other.

e. Geometrically, these three complex numbers are evenly distributed around the unit circle in the complex plane, starting at the point (1,0) representing the real number 1+0*i*.

f. Eigenvector for 1: We have to solve $W\mathbf{v} = \mathbf{v}$, but we know that W performs a cyclic shift on the entries of $\mathbf{v}$. The only way this can leave $\mathbf{v}$ unchanged, is if $\mathbf{v} = [c\ \ c\ \ c]^T$ for some number c. We can also determine this by solving the homogeneous system $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}\mathbf{v} = \mathbf{0}$. That will be left as an exercise.

g. Eigenvector for $\frac{-1+i\sqrt{3}}{2}$ which we will denote as ω. Note that $\omega^2 = \frac{-1-i\sqrt{3}}{2}$ and since $\omega^3 = 1$, $\omega^2 = \frac{1}{\omega}$. This will help us row-reduce the matrix $\begin{bmatrix} \omega & -1 & 0 \\ 0 & \omega & -1 \\ -1 & 0 & \omega \end{bmatrix}$.

   i. Multiply top row and second row by ω^2 : $\begin{bmatrix} 1 & -\omega^2 & 0 \\ 0 & 1 & -\omega^2 \\ -1 & 0 & \omega \end{bmatrix}$

   ii. Add ω^2 times row 2 to row 1: $\begin{bmatrix} 1 & 0 & -\omega^4 \\ 0 & 1 & -\omega^2 \\ -1 & 0 & \omega \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\omega \\ 0 & 1 & -\omega^2 \\ -1 & 0 & \omega \end{bmatrix}$ because $\omega^3 = 1$.

   iii. Add row 1 to row 3: $\begin{bmatrix} 1 & 0 & -\omega \\ 0 & 1 & -\omega^2 \\ 0 & 0 & 0 \end{bmatrix}$. If we think of the solution to the homogeneous equation as $[x\ \ y\ \ z]^T$, then $x = \omega z$ and $y = \omega^2 z$. So $[x\ \ y\ \ z]^T = z[1\ \ \ \omega\ \ \ \omega^2]$ where z is "free".

h. Eigenvector for $\frac{-1-i\sqrt{3}}{2}$ which we can denote as $\overline{\omega}$ or $\frac{1}{\omega}$. Using either properties of complex conjugates, or the same process as we used in the preceding case, we can find the eigenvectors for $\overline{\omega}$ are the complex conjugates of the eigenvectors for ω. That is, they are expressible in the form $[x\ \ y\ \ z]^T = z[1\ \ \ \overline{\omega}\ \ \ \overline{\omega}^2]$ where z is "free".

i. Comment: this example is one instance of a general pattern: When C is a *companion matrix* for the monic polynomial $p(x)$, then each root r of p is an eigenvalue, and multiples of $[1\ \ \ r\ \ \ r^2\ \ \cdots\ \ r^{n-1}]^T$ are corresponding eigenvectors. We will study this in detail a little later.

New Topic: Diagonalizability

1. Overview:
   a. Diagonalizing a matrix means changing it to a diagonal form by multiplying on the left by an invertible matrix and on the right by the inverse of that matrix.
   b. In equation form, the transformation $A \rightarrow P^{-1}AP$ is called a similarity operation (analogous to a *row* operation).

c. We want to apply such a transformation to change *A* into a diagonal matrix, so that $P^{-1}AP = D$ is diagonal.

d. Note that this implies $A = PDP^{-1}$. So $A^k = (PDP^{-1})(PDP^{-1})(PDP^{-1})\cdots(PDP^{-1}) = PD^kP^{-1}$

e. This means we can understand powers of *A* (and also polynomials in *A*) in connection with the much simpler powers (and polynomials) of *D*.

2. Eigenvector matrices

a. Let *P* be any matrix whose columns are eigenvectors of *A*. (For now, we do not assume that the columns are independent, or even distinct.)

b. To be specific, let the columns of *P* be $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ which are eigenvectors of *A*. Also, for each j let λ_j be an eigenvalue corresponding to the eigenvector $\mathbf{v}_j$. We do not assume that these eigenvalues are distinct, nor that the number of repetitions is governed by multiplicities of the eigenvalues as roots of the characteristic equation, nor that every eigenvalue of *A* is listed among our specific λ_j's. In particular, it might well be that $P = [\mathbf{v}\ \ \mathbf{v}\ \ \mathbf{v}\ \ \mathbf{v}]$ so that all of the columns of *P* are the same eigenvectors, and all of the λ_j's are the same eigenvalue. All we are assuming is that where $A\mathbf{v}_j = \lambda_j\mathbf{v}_j$ for $j = 1, 2, \cdots, k$.

c. Now observe that $AP = A[\mathbf{v}_1\ \ \mathbf{v}_2\ \ \cdots\ \ \mathbf{v}_k] = [A\mathbf{v}_1\ \ A\mathbf{v}_2\ \ \cdots\ \ A\mathbf{v}_k] = [\lambda_1\mathbf{v}_1\ \ \lambda_2\mathbf{v}_2\ \ \cdots\ \ \lambda_k\mathbf{v}_k]$.

d. On the other hand, from our knowledge of row and column operations, we know $[\lambda_1\mathbf{v}_1\ \ \lambda_2\mathbf{v}_2\ \ \cdots\ \ \lambda_k\mathbf{v}_k] = [\mathbf{v}_1\ \ \mathbf{v}_2\ \ \cdots\ \ \mathbf{v}_k]D$ where $D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k \end{bmatrix}$, and $[\mathbf{v}_1\ \ \mathbf{v}_2\ \ \cdots\ \ \mathbf{v}_k] = P$.

e. We have shown that *AP* = *PD* where *D* is a diagonal matrix whose nonzero entries are the eigenvalues corresponding to the columns of *P*

f. As a special case, if the matrix *P* happens to be invertible, then $D = P^{-1}AP$ and we have diagonalized *A*.

g. Conversely, if *A* is diagonalizable, and $D = P^{-1}AP$ is a diagonal matrix, then it follows that *AP* = *PD*. This implies that the columns of *P* are eigenvectors of *A* and the diagonal entries of *D* are the corresponding eigenvalues.

h. This establishes the following theorem: a square matrix *A* is diagonalizable iff there exists and invertible matrix whose columns are all eigenvectors of *A*.

i. From the invertible matrix theorem, this can be restated: an $n \times n$ matrix *A* is diagonalizable iff there exist *n* linearly independent eigenvectors of *A*.

j. To diagonalize *A*, we find its eigenvectors and eigenvalues. If there are *n* independent eigenvectors, we put them as columns into a matrix *P*, and put their corresponding eigenvalues (in order) into the diagonal matrix *D*. That gives us a diagonalization of *A*.

3. Example: Diagonalize the matrix W from our earlier example.

4. Example: $A = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}$ is not diagonalizable.

5. Lots of matrices are not diagonalizable. In fact, we have this
Theorem: a triangular (nondiagonal) matrix with all diagonal entries equal cannot be diagonalizable.

Proof: Let A be an $n \times n$ triangular matrix with all diagonal entries equal to one value λ, and with at least one nonzero entry off the main diagonal. We know that the diagonal entries of A must be eigenvalues of A, and since these are all equal, we see that A has a single eigenvalue λ, with multiplicity n.

Now suppose that A is diagonalizable. Then for some invertible matrix P we have $P^{-1}AP = D$, where D is the diagonal matrix all of whose entries equal λ. That is, $D = \lambda I$. But then we have $P^{-1}AP = \lambda I$ so $A = P\lambda I P^{-1} = \lambda P I P^{-1} = \lambda I$. This shows that A is a diagonal matrix, which is a contradiction. Therefore we conclude that A is not diagonalizable. ■

6. Special case: An $n \times n$ matrix with n distinct eigenvalues is diagonalizable. This follows from the following
Theorem: If $\lambda_1, \lambda_2, \cdots, \lambda_k$ are *distinct* eigenvalues of A, with corresponding *nonzero* eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$, then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is a linearly independent set. [Note: the word *distinct* in this statement means that no two of the eigenvalues are equal.]

This is Theorem 2 in section 5.1 of the file eigenvs_Lay4E.pdf posted on blackboard, where you will find a complete proof. I will say a little about the main idea of the proof momentarily. But for now, let us see how this theorem leads to the preceding statement: *no repeated eigenvalues implies diagonalizability*. If A is $n \times n$, then the characteristic polynomial is degree n, so there will be n roots, aka eigenvalues. Assume they are all distinct. For each eigenvalue there must be a nonzero eigenvector, and now our theorem says that these n eigenvectors are linear independent. Putting them as columns into a matrix P, we know that P will be invertible, and hence A is diagonalizable.

7. Main idea of the proof of the independence result. Suppose A is matrix with 4 distinct eigenvalues, say 2, 3, -5, and 8. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ be nonzero eigenvectors for these eigenvalues. And suppose that the eigenvectors are NOT linearly independent. That means one of the vectors is a combination of the others. Since the numbering purely a matter of notational convenience, we may as well assume that it is $\mathbf{v}_4$ that is dependent on the other vectors, so that we can write $\mathbf{v}_4 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$ for some scalar c's. And the c's cannot all equal 0 because $\mathbf{v}_4 \neq \mathbf{0}$.

Now apply A to both sides of this equation, and we obtain
$A\mathbf{v}_4 = c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 + c_3 A\mathbf{v}_3$, and because the **v**'s are eigenvectors, that gives us
$8\mathbf{v}_4 = c_1 2\mathbf{v}_1 + c_2 3\mathbf{v}_2 - c_3 5\mathbf{v}_3$. But from our first equation we also know
$8\mathbf{v}_4 = c_1 8\mathbf{v}_1 + c_2 8\mathbf{v}_2 + c_3 8\mathbf{v}_3$. Therefore we can subtract to find
$6c_1\mathbf{v}_1 + 5c_2\mathbf{v}_2 + 13c_3\mathbf{v}_3 = \mathbf{0}$, and as already mentioned these c's cannot all equal 0. (and the coefficients in front of the c's are all nonzero, because they are differences of distinct eigenvalues). That means we can solve for one of the **v**'s as a linear combination of the other two.

To summarize, assuming there are four dependent nonzero eigenvectors with distinct eigenvalues leads to the conclusion that there are actually *three* dependent nonzero eigenvectors with distinct eigenvalues. But the argument that we used can actually work with any number of nonzero eigenvectors having distinct eigenvalues. So we can build a kind of reverse induction argument, showing that we can keep reducing the number of vectors until we have only one, and that gives us a contradiction.

8. For matrices with some repeated eigenvalues, in some cases diagonalization is possible and in others it is not. Our theorem above about triangular matrices with all equal diagonal elements provides many examples where diagonalization is impossible. On the other hand, for any diagonal matrix D and any invertible matrix P, we can define the matrix $A = PDP^{-1}$. This matrix is evidently diagonalizable, and by putting some repeated elements on the diagonal of D, we assure that A has some repeated eigenvalues.

9. Here is the most general result about diagonalizability. Let A be an $n \times n$ matrix, and let the distinct eigenvalues be $\lambda_1, \lambda_2, \cdots, \lambda_k$, with $k < n$. For each eigenvalue λ_j, the set of solutions to the equation $(A - \lambda_j I)\mathbf{v} = \mathbf{0}$ is a subspace of $\mathbb{C}^n$ or $\mathbb{R}^n$ (and is actually the null space of the matrix $A - \lambda_j I$). This is referred to as the eigenspace corresponding to λ_j. The dimension of the eigenspace is the maximum number of independent eigenvectors possible for λ_j. With this terminology, A is diagonalizable iff the dimensions of its eigenspaces sum to n. In this case, we can make a set of n independent eigenvectors by choosing as many independent eigenvectors as possible for each eigenvalue. Putting all of these eigenvectors as columns into a single matrix gives us an invertible matrix P that diagonalizes A.

End of Day