CS205 - Class 4

1. As a review, all the matrices *A* we have looked at up to this point in the class have been full rank.
    a. For matrices with full rank, the first thing to consider is whether or not it is square.
        i. If the matrix is square, it is invertible, and Gaussian Elimination can be used to get an LU factorization. Furthermore, if the matrix is symmetric positive definite, a faster Cholesky factorization can be done to get LL^T.
        ii. If the matrix is not square, then it is taller than it is wide, and in this case we do the QR factorization to get the solution. We also considered using the normal equations instead of QR, but said this was bad since it squares the condition number. For QR, there are two ways we consider doing it: Gram-Schmidt and Householder. G.S. has numerical drift for larger matrices, so we prefer to use Householder.
    b. We will next consider matrices without full rank. In considering these types of matrices, we will look at the Singular Value Decomposition and Principal Component Analysis. In order to talk about these methods we first review eigenvalues and eigenvectors.

Eigenvalues and Eigenvectors (Readings Heath pp157-160)

2. For an $n \times n$ matrix A, $Ax = \lambda x$ is the standard eigenvalue problem where λ is an eigenvalue and x is a left eigenvector.
    a. The right eigenvectors y are defined by $y^T A = \lambda y^T$. If y is a right eigenvector of A, then it is a left eigenvector of A^T, since $A^T y = \lambda y$.
    b. Usually we refer to "left" eigenvectors simply as eigenvectors while still referring to "right" eigenvectors as right eigenvectors.
3. Complex matrices
    a. *Hermitian* matrices have $A^H = A$ where the "H" superscript indicates the complex conjugate of the transpose. Thus, for matrices with real values only, $A^H = A^T$, i.e. "H" corresponds to simple transposition
    b. *Unitary* matrices have $A^H A = AA^H = I$.
    c. *Normal* matrices have $A^H A = AA^H$
4. Recall an eigenvalue is a scalar that satisfies $Ax = \lambda x$.
    a. The set of all eigenvalues of A is called the *spectrum* of A, and the *spectral radius* of A is the magnitude of its largest magnitude eigenvalue. $\rho(A) = \max_i |\lambda_i|$.
    b. An eigenvector is a direction along which the action of a matrix is rather simple. The matrix merely expands or contracts vectors in that direction (according to the eigenvalue) leaving the direction unchanged.
    c. For diagonal matrices, the eigenvalues are on the diagonal, and the eigenvectors are the columns of the identity matrix. For example, $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 3\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
    d. For upper triangular and lower triangular matrices, the eigenvalues appear on the diagonal. For example, $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

e. A symmetric matrix is guaranteed to have real eigenvalues, i.e. $\lambda \in R$, while nonsymmetric matrices can have complex eignevalues. For example $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ i \end{bmatrix} = i\begin{bmatrix} 1 \\ i \end{bmatrix}$. It is easy to see why symmetric (or Hermitian, in general) matrices have real eigenvalues. Let Ax= λ x, then λ <x,x>= λ x^Hx=x^H λ x= x^HAx= $x^H A^H$x = $\overline{\lambda}$ x^Hx = $\overline{\lambda}$ <x,x>

f. Eigenvectors can be arbitrarily scaled by a constant, for example $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} s \\ s \end{bmatrix} = 3\begin{bmatrix} s \\ s \end{bmatrix}$ for any s. Thus, we usually require that eigenvectors be normalized (usually in the L_2 norm). Thus, $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = 3\begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}$ gives the eigenvector in standard form.

g. $Ax = \lambda x$ can be written equivalently as $(A - \lambda I)x = 0$ and there is a nontrivial solution x to this problem when $A - \lambda I$ is NOT invertible, or singular, or $\det(A - \lambda I) = 0$.

i. Note that $\det(A - \lambda I)$ is an n-th degree polynomial and we refer to it as the *characteristic polynomial* of A. The roots of the *characteristic equation* $\det(A - \lambda I) = 0$ are the eigenvalues of A.

ii. $\det\left(\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix}\right) = (2-\lambda)(3-\lambda) = 0$. Thus the matrix $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$ has $\lambda = 2,3$ as eignevalues.

iii. An $n \times n$ matrix A has an n-th degree characteristic polynomial with n roots, and thus n eigenvalues. However, there may be multiple roots or complex roots.

1. In the repeated eigenvalue case, if there are fewer linearly independent eigenvectors than repeated roots, the eigenvalue and the matrix is said to be *defective*.
2. $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ so the multiple eigenvalue of 2 has two linearly independent eigenvectors.
3. $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is the only eignevector relationship for the multiple eigenvalue 2 here, and thus this matrix is defective.

5. Idea is to compute eigenvalues using matrix form because it is in general easier than solving a *n*-degree polynomial.
   a. Must be careful though as we need to preserve least squares solution. i.e. Gaussian-Elimination will not work.
   b. Use a similarity transform that produces a new matrix with "same" eigenvalues/eigenvectors.
6. Formally, a matrix A is said to be **similar** to a matrix B, if $B = T^{-1}AT$ for a nonsingular matrix T.
   a. If A and B are similar, then they have the same eigenvalues. $By = \lambda y$ or $T^{-1}ATy = \lambda y$ or $A(Ty) = \lambda(Ty)$. Note that the eigenvectors of A are Ty where y are the eigenvectors of B (this formalizes the notion of "same-ness" from 1(b)).
   b. If the matrix A has distinct eigenvalues (no repeated eigenvalues), then similarity transforms can be used to put it into diagonal form where the eigenvalues can be read from the diagonal and the

eigenvectors are the columns of the identity matrix. Then the eigenvectors of A are T times the columns of the identity matrix, i.e. the columns of T.

c. If A is real and symmetric, an orthogonal T can be used to put A into diagonal form. Moreover, the eigenvalues are real valued.
d. If A is complex and Hermitian, a unitary T can be used to put A into diagonal form. Moreover, the eigenvalues are real valued.
e. If A is normal, a unitary T can be used to put A into diagonal form.
f. Any matrix can be put into upper triangular, Shur form, with a unitary T. Then the eigenvalues can be read off the diagonal of the matrix.
g. Any matrix can be put into *Jordan form* where the eigenvalues are on the diagonal, and off diagonal elements only occur on the band above the diagonal and only for defective eigenvalues. For example $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ is in Jordan form where both 2 and 3 are repeated eigenvalues, but only 3 is defective.

7. The condition number for an eigenvalue problem is defined by $1/\left|y^H x\right|$ where x and y are the normalized right and left eigenvectors.
   a. If x and y are real valued, then $y^H x = y^T x$. Note that $y^T x = \left|y\right|\left|x\right|\cos\theta = \cos\theta$ where θ is the angle between the eigenvectors.
   b. For symmetric and Hermitian matrices, the left and right eigenvectors are the same so the condition number is 1.
   c. Eigenvalues are well conditioned for normal matrices.
   d. Multiple or "close" eigenvalues can be poorly conditioned, especially if they are defective or "close" to being defective.
   e. Scaling by a diagonal similarity transform – called *balancing* – can improve the condition number of an eigenvalue problem.
8. Numerical methods for finding eigenvalues and eigenvectors.
   a. Would like a technique that solves them one at a time starting with the largest in magnitude. Similarity transform is less desirable because it gives us all of them even we need only a few.
   b. Using the characteristic polynomial is a bad idea since the coefficients are ill-conditioned. Moreover, one should set up a matrix and compute eigenvalues in order to find the roots of a polynomial equation.
   c. QR iteration. Initially set $A_0 = A$ and then iterate A_1, A_2, etc. For each k, compute the QR factorization $A_k = Q_k R_k$, and then define $A_{k+1} = R_k Q_k = Q_k^H Q_k R_k Q_k = Q_k^H A_k Q_k$.
      i. If the eigenvalues are all distinct, then the A_k converge to a triangular matrix. Moreover, if A is symmetric, the A_k converge to a diagonal matrix.
      ii. This convergence can be accelerated by using *shifts* of the form $A_k - \sigma_k I = Q_k R_k$ and $A_{k+1} = R_k Q_k + \sigma_k I$ where σ_k is a rough approximation to an eigenvalue. Initially one can use, for example, the number in the lower right hand corner of the matrix.
9. The *Power Method* allows one to compute the largest eigenvalue and eigenvector. Starting from a nonzero vector x_0, iterate with $x_{k+1} = Ax_k$.

a. To see why this works, assume that x_0 is a linear combination of eigenvectors $x_0 = \sum_i \alpha_i u_i$ where the u_i are the eigenvectors of A. Then $x_k = Ax_{k-1} = A^2 x_{k-2} = \cdots = A^k x_0$ and so $x_k = A^k x_0 = A^k \sum_i \alpha_i u_i = \sum_i \alpha_i A^k u_i = \sum_i \alpha_i \lambda_i^k u_i$. Now assuming that the largest eigenvalue is λ_1, we can write $x_k = \alpha_1 \lambda_1^k u_1 + \sum_{i=2} \alpha_i \lambda_i^k u_i = \lambda_1^k \left(\alpha_1 u_1 + \sum_{i=2} \alpha_i \left(\lambda_i / \lambda_1 \right)^k u_i \right)$ and note that the second term vanishes as $k \to \infty$ since $\left| \lambda_i / \lambda_1 \right| < 1$. Thus as $k \to \infty$, $x_k \to \lambda_1^k \alpha_1 u_1$. Moreover $\left(x_k \right)_j / \left(x_{k-1} \right)_j \to \lambda_1$ for any component j of x.

b. If the starting vector $x_0 = \sum_i \alpha_i u_i$ happens to have $\alpha_i = 0$ for the largest eigenvalue, the method fails.

c. For a real matrix and real x_0, one can never get complex numbers.

d. The largest eigenvalue may be repeated, in which case the final vector may be a linear combination of the true eigenvectors.

e. After every iteration, x_k can be renormalized to stop x_k from growing too large.

f. Shifts can be used to accelerate convergence.

g. Inverse iteration can be used to find the smallest eigenvalue. This relies on the fact that the eigenvalues of A^{-1} are the reciprocals of those of A. Thus, the largest eigenvalue of A^{-1} is the smallest eigenvalue of A.

h. *Deflation* is a method to remove an eigenvalue from a matrix A once it has been computed. Then the resulting matrix can be analyzed to compute the next largest eigenvalue, etc.

10. If $Ax = \lambda x$, then one can form the Rayleigh Quotient $\lambda = \dfrac{x^T A x}{x^T x}$. This is used in a variety of methods for computing eigenvalues.