CS 205a Fall 2011 Midterm 1

Please write your name on the top right of this page. The exam is closed book/closed notes and no calculators are allowed. You have 1 hour and 15 minutes to complete the exam.

Multiple Choice (4 x 1 pt each)

For each of the following questions, circle all answers which are correct. You must circle **ALL** of the answers for a given question correctly to receive credit.

1. Which of the following matrices are always diagonalizable (i.e. the matrices that always have a full set of eigenvectors)?

   (a) Symmetric matrices

   (b) Orthogonal matrices

   (c) Householder matrices

   (d) Upper triangular matrices

   Answer: a, b, c

2. Which of the following computations have operation counts of $O(n^3)$? Consider the cases for general $n \times n$ matrices.

   (a) QR factorization with the Gram-Schmidt method

   (b) LU decomposition

   (c) Back substitution step in solving $A\vec{x} = \vec{b}$

   (d) Cholesky factorization

   Answer: a,b,d

3. $\|\vec{x}\|_\infty$ is best for evaluating the numerical solution to a system of equations $A\vec{x} = \vec{b}$ because:

   (a) It is often the only obtainable norm

   (b) It illustrates the worst error for a solution to an individual equation

   (c) It provides information about the total error in the solution

   (d) It is always cheapest to compute

   Answer: b

4. A Householder matrix $H = I - 2\frac{\vec{v}\vec{v}^T}{\vec{v}^T\vec{v}}$

   (a) $\forall \vec{x} \in \mathbb{R}^n, \|\vec{x}\|_2 = \|H\vec{x}\|_2$

   (b) has eigenvalues equal to 1 with multiplicity n

   (c) is a projection matrix onto the hyperplane orthogonal to $\vec{v}$

   (d) has a determinant of -1 (i.e. $det(H) = -1$)

   Answer: a, d

Eigenvalues and Eigenvectors (10 pts)

Given a 2×2 matrix $A = \begin{bmatrix} 7 & 2 \\ 3 & 2 \end{bmatrix}$, answer the following:

1. What are the eigenvalues of A? (3 pts)

$$\begin{aligned} det\left(\begin{bmatrix} 7-\lambda & 2 \\ 3 & 2-\lambda \end{bmatrix}\right) &= 0 \\ (7-\lambda)(2-\lambda) - 6 &= 0 \\ 14 - 9\lambda + \lambda^2 - 6 &= 0 \\ \lambda^2 - 9\lambda + 8 &= 0 \\ (\lambda - 8)(\lambda - 1) &= 0 \\ \Rightarrow \lambda_1 = 1, \lambda_2 &= 8 \end{aligned}$$

2. For each eigenvalue of A, find the corresponding eigenvectors. (3 pts)

   $A\vec{x} = \lambda\vec{x}$

$$\begin{bmatrix} 7 & 2 \\ 3 & 2 \end{bmatrix} \vec{v_1} = \vec{v_1} \qquad \begin{bmatrix} 7 & 2 \\ 3 & 2 \end{bmatrix} \vec{v_2} = 8\vec{v_2}$$

$$\begin{cases} 7\vec{v}_{1,1} + 2\vec{v}_{1,2} = \vec{v}_{1,1} \\ 3\vec{v}_{1,1} + 2\vec{v}_{1,2} = \vec{v}_{1,2} \end{cases} \qquad \begin{cases} 7\vec{v}_{2,1} + 2\vec{v}_{2,2} = 8\vec{v}_{2,1} \\ 3\vec{v}_{2,1} + 2\vec{v}_{2,2} = 8\vec{v}_{2,2} \end{cases}$$

$$\vec{v_1} = \begin{bmatrix} 1 \\ -3 \end{bmatrix} \qquad \vec{v_2} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

3. If you were to perform power method iterations on A with $\vec{x_0} = \begin{pmatrix} 3 & 2 \end{pmatrix}^T$ as the starting vector, to what eigenvector will the power method converge? Why? (1 pt)

   It will converge to the closest eigenvector, $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Closeness can be checked by taking the dot product of $\vec{x_0}$ with the two eigenvectors.

4. Recall that if $A\vec{x} = \lambda\vec{x}$, one can form the Rayleigh Quotient. Use $\vec{x_0} = \begin{pmatrix} 3 & 2 \end{pmatrix}^T$ as an approximate eigenvector, and compute the corresponding approximate eigenvalue using the Rayleigh Quotient. You may round to an integer in the final calculation. To which exact eigenvector, calculated in (2), does this approximate eigenvalue correspond? In general, how can the Rayleigh Quotient be used to accelerate the power method? (3 pts)

   The Rayleigh Quotient is $\lambda = \frac{\vec{x}^T A \vec{x}}{\vec{x}^T \vec{x}}$.

   Direct substitution gives $\lambda = \dfrac{\begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} 7 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}}{\begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}} = \frac{101}{13} \approx 7.7$

   This approximate eigenvalue corresponds to the actual eigenvalue of 8, and thus to the actual eigenvector $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

   In general, the Rayleigh Quotient can be used to accelerate the power method by using its estimate of the eigenvalue at each step of the iteration, instead of the basic ratio.

Optimization and Nonlinear Equations (7 pts)

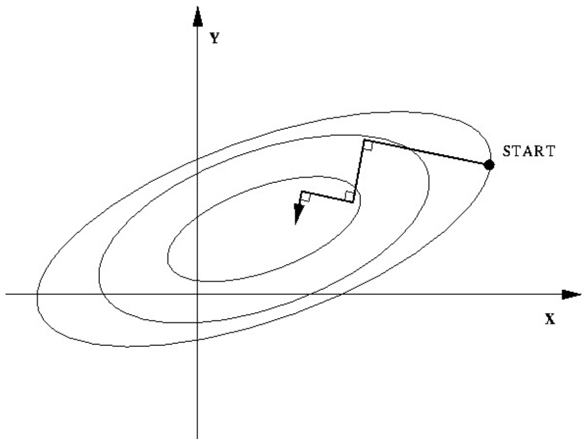
1. Rank the following methods from slowest convergence rate to fastest convergence rate: Newton's Method, Secant Method, Bisection. Which of these is the most robust? Devise a method to maintain the fastest convergence rate while preserving robustness. (3 pts)

   From slowest to fastest: Bisection, Secant Method, Newton's Method.
   Bisection is the most robust.
   Proposed method: combining bisection method (for robustness) with Newton's method (for faster convergence rate, using secant method is also correct). Start with a bracketing interval [a,b], where f(a) and f(b) have opposite signs, and x_0 inside this interval; apply Newton's method to find x_{k+1}; if x_{k+1} is inside the interval, it is accepted and update the interval accordingly, otherwise apply bisection method as usual to update the interval.

2. Suppose you use an iterative method to solve $f(x) = 0$ for a root that is ill-conditioned, and you need to choose a convergence test. Would it be better to terminate the iteration when you find x_k for which $|f(x_k)|$ is small, or when $|x_k - x_{k-1}|$ is small? Why? (3 pts)

   It is better to terminate the iteration when $|x_k - x_{k-1}|$ is small. The reason is that ill-conditioning of the root finding problem of a given function is opposite of evaluating the function, which means in the neighborhood of the root, slight variation of the function value would lead to large variation of x. If we use $|f(x_k)|$ as criteria, the iteration may stop making progress when our solution is still far away from the root.

3. Draw a 2-dimensional example of convergence of the steepest descent method. Clearly indicate the angles between search direction and contours of the objective function. (1 pt)



Least Squares (9 pts)

1. Prove that the method of normal equations minimizes the residual. (3 pts)

   This proof was shown in Review 1.

$$
\begin{aligned}
\phi(x) = \|\vec{r}\|_2^2 &= \|\vec{b} - A\vec{x}\|_2^2 \\
\vec{r}^T\vec{r} &= \left(\vec{b} - A\vec{x}\right)^T \left(\vec{b} - A\vec{x}\right) \\
&= \vec{b}^T\vec{b} - \vec{x}^T A^T\vec{b} - \vec{b}^T A\vec{x} + \vec{x}^T A^T A\vec{x} \\
&= \vec{b}^T\vec{b} - 2\vec{x}^T A^T\vec{b} + \vec{x}^T A^T A\vec{x}
\end{aligned}
$$

Take the gradient and set equal to 0:

$$
\begin{aligned}
\nabla\phi(x) &= -2A^T\vec{b} + 2A^T A\vec{x} \\
0 &= -2A^T\vec{b} + 2A^T A\vec{x}
\end{aligned}
$$

$$\Rightarrow A^T A\vec{x} = A^T\vec{b}$$

2. On the computer, why do we solve the least squares problem $\min_{\vec{x}} \|\vec{b} - A\vec{x}\|$, even when A is invertible? (2 pts)

   1. Small perturbations in A and $\vec{b}$ due to handling by the computer may result in $\vec{b}$ not being in the column space of A or A no longer being invertible.
   2. We generally don't like taking the inverse of matrices, as inverting a matrix can be a very expensive operation.

3. In what case(s) do we want to avoid solving a linear system on the computer using the normal equations, and why? (1 pt)

   If the condition number of A is large, the condition number of $A^T A$ will square it and be very large. Thus the normal equations should be avoided for ill-conditioned matrices.

4. Use the modified Gram-Schmidt method to find the QR decomposition of $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. (3 pts)

$$\vec{q_1} = \frac{\vec{a_1}}{r_{11}} = \frac{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}{\left\|\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$r_{12} = \vec{q_1} \cdot \vec{a_2} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{3}{\sqrt{2}}$$

$$\vec{a_{2*}} = \vec{a_2} - r_{12}\vec{q_1} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{3}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$\vec{q_2} = \frac{\vec{a_{2*}}}{r_{22}} = \frac{\begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}}{\left\|\begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}\right\|} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$$

$$A = QR = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{\sqrt{2}}{2} \\ \frac{1}{\sqrt{2}} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Singular Value Decomposition (10 pts)

1. Write down the SVD for the following matrices (no proof is necessary): (3 pts)

$$\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}:$$

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}:$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}:$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

2. Prove that if A is a real matrix and $A = A^T$, then the singular values of A are the absolute values of the eigenvalues of A. (1 pt)

   Because A is a symmetric matrix, then $A = QDQ^T$, where Q is an orthogonal matrix and D is a real diagonal matrix. Consequently, $A = Q|D|sign(D)Q^T$, where $sign(D)Q^T$ is also an orthogonal matrix. Therefore, the diagonal entries of $|D|$, which are absolute values of eigenvalues of A, are singular values of A.

3. Given the SVD of A as $A = U\Sigma V^T$, what space do the columns of V corresponding to zero singular values span? Prove your statement. (3 pts)

   The columns of V corresponding to zero singular values span the null space of A.
   Proof:
   A=$U\Sigma V^T$=$\Sigma_{j=1}^{r}\sigma_j u_j v_j^T$,
   where σ_j are nonzero singular values, u_j and v_j are column vectors of U and V that correspond to nonzero singular values.
   Take arbitray column of V that corresponds to zero singular values, such as v_k, we have
   Av_k=$\Sigma_{j=1}^{r}\sigma_j u_j v_j^T v_k = 0$
   Take arbitray column of V that corresponds to nonzero singular values, such as v_p, we have
   Av_p=$\sigma_p(v_p^T v_p)u_p \neq 0$
   Since the columns of V are orthogonal and span the whole space, the columns of V corresponding to zero singular values span the null space of A.

4. Let A be an $m \times n$ real matrix. Consider a symmetric matrix eigenvalue problem:

$$\begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} \vec{x} = \lambda \vec{x}$$

Show that if λ satisfies this relation, then $|\lambda|$ is a singular value of A. (3 pts)

Write $\vec{x}$ as $\begin{pmatrix} \vec{y} \\ \vec{z} \end{pmatrix}$,
Because $\begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} \vec{y} \\ \vec{z} \end{pmatrix} = \lambda \begin{pmatrix} \vec{y} \\ \vec{z} \end{pmatrix}$,
$A\vec{z} = \lambda\vec{y}$ and $A^T\vec{y} = \lambda\vec{z}$
$A^T A\vec{z} = \lambda A^T\vec{y} = \lambda^2\vec{z}$
Then, λ^2 is an eigenvalue of $A^T A$,
Therefore, $|\lambda|$ is a singular value of A.