CS 205a Fall 2010 Midterm 1

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

Multiple Choice (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 can be said about a Householder matrix $H = I - 2\frac{vv^T}{v^Tv}$

   (a) The condition number for H is 1

   (b) It is a projection matrix onto the hyperplane orthogonal to v

   (c) It preserves the 2-norm of a vector (ie $\|x\|_2 = \|Hx\|_2$ for all x)

   (d) One of the eigenvalues of H is 1 with multiplicity $n - 1$

   Answer: a, c, d

2. Which of the following classes of matrices are positive semi-definite?

   (a) Permutation matrices

   (b) Projection matrices

   (c) Reflection matrices

   (d) Symmetric matrices

   Answer: b

3. Suppose that a square matrix A is ill-conditioned. Which of the following matrices could potentially have a better condition number?

   (a) cA, where c is a non-zero scalar

   (b) DA, where D is a nonsingular diagonal matrix

   (c) PA, where P is a permutation matrix

   (d) A^{-1}

   Answer: b

4. Which of the following about the least squares solutions are true?

   (a) It satisfies $Ax = b$

   (b) It can be found by solving the normal equation

   (c) Its associated residual lies in the nullspace of A^T

   (d) It lies in the column space of A

   Answer: b, c

Eigenvalues (10 pts)

1. Given that $A = T^{-1}BT$, prove that A and B have the same eigenvalues. How do the eigenvectors of A and B relate? (2 pts)

   Answer: Let λ, q be an eigenvalue, eigenvector pair of A.

   $$\begin{aligned} Aq &= \lambda q \\ \Rightarrow T^{-1}BTq &= \lambda q \\ \Rightarrow BTq &= \lambda Tq \\ \Rightarrow B\tilde{q} &= \lambda \tilde{q}, \end{aligned}$$

   where $\tilde{q} = Tq$. Thus λ is also an eiqenvalue of B. The corresponding eigenvector is $\tilde{q} = Tq$.

2. What are the eigenvalues of a projection matrix? (Just state. No proof required) (1 pts)

   Answer: $1, 0$

3. If $A = A^T$, and $x^T Ax > 0, \forall x \neq 0$, prove that the eigenvalues of A are all positive. (3 pts)

   Answer: Let λ, q be an eigenvalue, eigenvector pair of A.

$$\begin{aligned} q^T Aq &> 0 \\ \Rightarrow q^T \lambda q &> 0 \\ \Rightarrow \lambda q^T q &> 0 \\ \Rightarrow \lambda &> 0. \end{aligned}$$

   The last step follows since $q^T q > 0$ for any non zero vector and since q is an eigenvector, it is non-zero by definition.

4. Given a symmetric matrix, how might you find the second largest eigenvalue. (4 pts)

   Answer:

   (a) Use power method to find the largest eigenvalue.
   (b) Construct the deflation matrix of the matrix.
   (c) Use power method to find the largest eigenvalue of the deflated matrix.

   You may want to repeat the last two steps if the largest eigenvalue has multiplicity greater than 1.

SVD (2 pts)

1. State (without proof) SVD of the following matrices: (1 pt each)

   (a)

$$\begin{pmatrix} -3 \\ -4 \end{pmatrix}$$

   Answer:

$$\begin{pmatrix} -3 \\ -4 \end{pmatrix} = \begin{pmatrix} -3/5 & 4/5 \\ -4/5 & -3/5 \end{pmatrix} \begin{pmatrix} 5 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix}$$

   (b)

$$\begin{pmatrix} 0 & 0.0637 & 0 \\ 97.83 & 0 & 0 \\ 0 & 0 & 2.763 \end{pmatrix}$$

   Answer:

$$\begin{pmatrix} 0 & 0.0637 & 0 \\ 97.83 & 0 & 0 \\ 0 & 0 & 2.763 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 97.83 & 0 & 0 \\ 0 & 2.763 & 0 \\ 0 & 0 & 0.0637 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

   We just want to permute the rows and columns, to get the right ordering in the diagonal matrix.

Least Squares (6 pts)

The complete orthogonal factorization of matrix A is $A = QRZ^T$, where $R = \begin{pmatrix} \hat{R} & 0 \\ 0 & 0 \end{pmatrix}$, $\hat{R}$ is upper triangular and non-singular, and Q, Z are orthogonal. Given the complete orthogonal factorization of A, construct an algorithm for computing the minimum norm least squares solution of the problem $Ax = b$. (Note that the QR factorization is not sufficient, and the complete orthogonal factorization is cheaper than the SVD.)

Answer: The least square solution should satisfy the normal equations $A^T A = A^T b$. Using this transformation

$$\begin{aligned} A^T A x &= A^T b \\ \Rightarrow Z R^T Q^T Q R Z^T x &= Z R^T Q^T b \\ \Rightarrow Z R^T R Z^T x &= Z R^T Q^T b \;\; (Q^T Q = I) \\ \Rightarrow R^T R Z^T x &= R^T Q^T b \;\; \text{(Z is orthogonal and hence invertible)} \\ \Rightarrow R^T R \tilde{x} &= R^T \tilde{b}, \end{aligned} \tag{1}$$

where $\tilde{x} = Z^T x, \tilde{b} = Q^T b$. First, let's look at solving $R\tilde{x} = \tilde{b}$. Now

$$\begin{aligned} R\tilde{x} &= \tilde{b} \\ \Rightarrow \begin{pmatrix} \hat{R} & 0 \\ 0 & 0 \end{pmatrix} \tilde{x} &= \tilde{b} \\ \Rightarrow \begin{pmatrix} \hat{R} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \bar{x} \end{pmatrix} &= \begin{pmatrix} \hat{b} \\ \bar{b} \end{pmatrix} \end{aligned}$$

Like for QR decomposition, we cannot eliminate the error in $\bar{b}$, which will correspond to the least square error. Since $\hat{R}$ is non-singular, we can eliminate the error in $\hat{b}$ completely using $\hat{x} = \hat{R}^{-1}\hat{b}$. We now claim that if $\hat{x}$ is chosen like this, then $\tilde{x} = \begin{pmatrix} \hat{x} \\ \bar{x} \end{pmatrix}$ for any choice of $\bar{x}$ will give the least square solution. We can prove this by proving that this choice satisfies equation (1). Restating equation (1)

$$\begin{aligned} R^T R \tilde{x} &= R^T \tilde{b} \\ \Rightarrow R^T \begin{pmatrix} \hat{R} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \bar{x} \end{pmatrix} &= \begin{pmatrix} \hat{R}^T & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{b} \\ \bar{b} \end{pmatrix} \\ \Rightarrow R^T \begin{pmatrix} \hat{R}\hat{x} \\ 0 \end{pmatrix} &= \begin{pmatrix} \hat{R}^T \hat{b} \\ 0 \end{pmatrix} \\ \Rightarrow R^T \begin{pmatrix} \hat{b} \\ 0 \end{pmatrix} &= \begin{pmatrix} \hat{R}^T \hat{b} \\ 0 \end{pmatrix} \;\; \text{(We chose } \hat{x} = \hat{R}^{-1}\hat{b}) \\ \Rightarrow \begin{pmatrix} \hat{R}^T & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{b} \\ 0 \end{pmatrix} &= \begin{pmatrix} \hat{R}^T \hat{b} \\ 0 \end{pmatrix} \\ \Rightarrow \begin{pmatrix} \hat{R}^T \hat{b} \\ 0 \end{pmatrix} &= \begin{pmatrix} \hat{R}^T \hat{b} \\ 0 \end{pmatrix} \end{aligned}$$

which is always true. Thus we proved that this choice of $\tilde{x}$ satisfies equation (1) and thus the normal equations, and thus is a least squares solution. Note that since $\tilde{x} = Z^T x$, our final value of x is given by $x = Z\tilde{x}$. Since Z is an orthonormal matrix, it does not change the norm. Thus for the least norm x, we wan't to choose the least norm $\tilde{x}$, which is nothing but the $\tilde{x}$ created by choosing $\bar{x} = 0$. Thus the minimum norm least squares solution is given by $x = Z \begin{pmatrix} \hat{R}^{-1}\hat{b} \\ 0 \end{pmatrix}$.

Nonlinear equations (12 pts)

The Taylor series for a sufficiently smooth function $f(x)$ is given by

$$f(x+h) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!} h^n,$$

where $f^{(n)}(x)$ is the n^{th} derivative of f evaluation at x. Truncating the series to the first k terms gives

$$f(x+h) = \sum_{n=0}^{k-1} \frac{f^{(n)}(x)}{n!} h^n + R_k(x),$$

where $R_k(x)$ is the error due to truncation. The Taylor theorem says that there exists a $\xi \in [x, x+h]$ s.t.

$$R_k(x) = \frac{f^{(k)}(\xi)}{k!} h^k$$

1. Derive Newton's method for finding the root of $f(x)$ by truncating the Taylor series. (2 pts)

   Answer: Starting at a guess x_k, using the first two terms of the series, and ignoring the truncation error

   $$f(x_k + h) = f(x_k) + f^{(1)}(x_k)h$$

   We wan't to find h, such that $f(x_k + h) = 0$. Thus

   $$\begin{aligned} f(x_k) + f^{(1)}(x_k)h &= 0 \\ \Rightarrow h &= -\frac{f(x_k)}{f^{(1)}(x_k)} \end{aligned}$$

   Thus the new guess for $x = x_{k+1} = x_k + h = x_k - \frac{f(x_k)}{f^{(1)}(x_k)}$, which is the Newton's method update rule.

2. Derive an update rule for finding the root of $f(x)$ by truncating the series to the first 3 terms. Note that the update rule should provide a unique guess for the next iteration. (4 pts)

   Answer: Starting at a guess x_k, using the first three terms of the series, and ignoring the truncation error

   $$f(x_k + h) = f(x_k) + f^{(1)}(x_k)h + \frac{f^{(2)}(x_k)}{2} h^2$$

   We wan't to find h, such that $f(x_k + h) = 0$. Thus

   $$f(x_k) + f^{(1)}(x_k)h + \frac{f^{(2)}(x_k)}{2} h^2 = 0$$

   This is a quadratic in h and can be solved for h using the quadratic formula, i.e.

   $$h = \frac{-f^{(1)}(x_k) \pm \sqrt{(f^{(1)}(x_k))^2 - 2f^{(2)}(x_k)f(x_k)}}{f^{(2)}(x_k)}$$

   This gives us two options. To choose between $+$ or $-$, we note that when we have reached the solution, i.e. when $f(x_k) = 0$, we want that $h = 0$. Substituting $f(x_k) = 0$ in the above formula we get $h = \frac{-f^{(1)}(x_k) \pm f^{(1)}(x_k)}{f^{(2)}(x_k)}$. Thus for h to go to zero, we need to choose the $+$ sign. (Credit has also been given for choosing the sign which gives the least value of $f(x_{k+1})$.)

3. Prove that Newton's method converges at a quadratic rate. What additional assumptions are required? Note that an iterative method is said to converge with rate r if $\lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|^r} = C$ for some finite constanct $C > 0$, whre e_k is the error at iteration k. (Hint: Use taylor series expansion on $g(x) = \frac{f(x)}{f'(x)}$ around the root x^* of f.) (6 pts)

Answer:

$$
\begin{aligned}
g(x) &= \frac{f(x)}{f'(x)} \\
\Rightarrow g^{'}(x) &= \frac{(f^{'}(x))^2 - f(x)f^{''}(x)}{(f^{'}(x))^2}
\end{aligned}
$$

At $x = x^*$, $f(x^*) = 0$, thus $g^{'}(x^*) = 1$, assuming $f^{'}(x^*) \neq 0$. Using the taylor series explansion of $g(x)$ around x^*, we get

$$
\begin{aligned}
g(x^* + h) &= g(x^*) + g^{'}(x^*)h + g^{''}(\xi)h^2 \quad \text{(for some } \xi \in [x, x+h]) \\
\Rightarrow g(x^* + h) &= 0 + h + g^{''}(\xi)h^2 \quad (g(x^*) = \frac{f(x^*)}{f'(x^*)} = 0, g^{'}(x^*) = 1) \\
\Rightarrow g(x^* + h) &= h + Ch^2 \quad \text{(Let } C = g^{''}(\xi)) \qquad (2)
\end{aligned}
$$

Now the error at any iteration is given by $e_k = x_k - x^*$. Thus $x_k = x^* + e_k$. Using this in the update rule:

$$
\begin{aligned}
x_{k+1} &= x_k - g(x_k) \\
\Rightarrow x_{k+1} &= x_k - g(x^* + e_k) \\
\Rightarrow x_{k+1} &= x_k - (e_k + Ce_k^2) \quad \text{(from (2))} \\
\Rightarrow x_{k+1} &= x_k - (x_k - x^* + Ce_k^2) \\
\Rightarrow x_{k+1} &= x^* - Ce_k^2 \\
\Rightarrow x_{k+1} - x^* &= -Ce_k^2 \\
\Rightarrow e_{k+1} &= -Ce_k^2 \\
\Rightarrow \frac{e_{k+1}}{e_k^2} &= -C
\end{aligned}
$$

Assuming we start close enough, so that e_k is small enough; and $g^{''}(x)$ is bounded, so that $C = g^{''}(\xi)$ is bounded, we can see that $e_k \to 0$ as $C \to \infty$. Since $\xi \in [x^*, x^* + e_k]$, as $e_k \to 0$, $\xi \to x^*$, thus $C \to g^{''}(x^*)$, a constant. Thus

$$
\lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|^r} = C,
$$

where $C = g^{''}(x^*)$, and $r = 2$. Thus with the assumptions stated, Newton's method converges at rate 2.

Optimization (6 pts)

1. The steepest descent method involves choosing a search direction $\vec{s_k} = -\nabla f(\vec{x_k})$, choosing α_k to minimize $f(\vec{x_k} + \alpha_k \vec{s_k})$ and setting $\vec{x_{k+1}} = \vec{x_k} + \alpha_k \vec{s_k}$. Prove that $\nabla f(\vec{x_{k+1}}) \cdot \nabla f(\vec{x_k}) = 0$.

   Answer: To minimize $f(\vec{x_k} + \alpha \vec{s_k})$, we set

$$
\begin{aligned}
\frac{d}{d\alpha} f(\vec{x_k} + \alpha \vec{s_k}) &= 0 \\
\Rightarrow \nabla f(\vec{x_k} + \alpha \vec{s_k}) \cdot \frac{d}{d\alpha}(\vec{x_k} + \alpha \vec{s_k}) &= 0 \quad \text{(Chain rule)} \\
\Rightarrow \nabla f(\vec{x_k} + \alpha \vec{s_k}) \cdot \vec{s_k} &= 0 \\
\Rightarrow \nabla f(\vec{x_k} + \alpha \vec{s_k}) \cdot \nabla f(\vec{x_k}) &= 0 \quad (\vec{s_k} = -\nabla f(\vec{x_k})) \\
\Rightarrow \nabla f(\vec{x_{k+1}}) \cdot \nabla f(\vec{x_k}) &= 0 \quad (\vec{x_{k+1}} = \vec{x_k} + \alpha \vec{s_k})
\end{aligned}
$$