May 8, 2023

FINAL EXAM

- Give reasons for your answers. Without valid reasons, you may not get any marks.

1. **(3 marks)** Consider the gradient descent method applied to a convex function $f : \Re^n \to \Re$, where at each step we move along the negative gradient direction to a point which minimizes the function f along this line. Let x_k and x_{k+1} be two consecutive points generated by this method. Prove that $\nabla f(x_k)^T \nabla f(x_{k+1}) = 0$.

   Solution: Gradient descent has the following rule **(0.5 marks)**:

   $$x_{k+1} = x_k - \eta \nabla f(x_k).$$

   Now the value η is found by minimizing $g(\eta) = f(x_k - \eta \nabla f(x_k)^T)$ **(1 mark)**. Using chain rule, $g'(\eta) = \nabla f(x_k - \eta \nabla f(x_k))^T \nabla f(x_k)$, **(1 mark)**and so this must be 0. This implies that $\nabla f(x_{k+1})^T \nabla f(x_k) = 0$ **(0.5 marks)**.

   Common Mistakes: Lack of mathematical rigour and not mentioning differentiation.

2. **(4 marks)** Determine the rate of convergence of Newton's method (with step size $\eta = 1$) for minimizing the function

   $$f(x) = ||x||_2^3,$$

   where $x \in \Re^n$. You may want to use the following formula (you need not prove this). If u is a **unit** vector in $\Re^n$ and I is the identity $n \times n$ matrix, then

   $$(I + uu^T)^{-1} = I - \frac{1}{2}uu^T.$$

   Solution: We need to compute the Hessian and gradient. Standard calculations show that gradient of f is $3||x||x$ **(1 mark)**. Its Hessian is $3||x||(I + uu^T)$, where u is the unit vector $x/||x||$ **(1 mark)**. So using the above **(1 mark)**,

   $$H(x)^{-1} \nabla f(x) = \frac{1}{3||x||}(I - 1/2uu^T) \cdot 3||x||x = x/2.$$

   Thus, $x_{k+1} = x_k/2$, which implies linear convergence **(1 mark)**.

3. (i) **(3 marks)** Let $f_1, \ldots, f_m$ be m convex functions with a common domain. Prove that $f(x) = \max_{i=1}^m f_i(x)$ is also a convex function.

   Solution: We need to use the definition of a convex function. Let x_1, x_2 be two points in the domain of f and $\lambda \in [0, 1]$. We need to show that **(0.5 marks)**

   $$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

   Now, LHS is equal to

   $$\max_{i=1}^{m} f_i(\lambda x_1 + (1 - \lambda)x_2).$$

Since each of the f_i is convex, the above is at most **(1 mark)**

$$\max_{i=1}^{m} \left[\lambda f_i(x_1) + (1-\lambda) f_i(x_2)\right].$$

But the above is at most **(1 mark)**

$$\lambda \max_i f_i(x_1). + (1-\lambda) \max_i f_i(x_2),$$

which is same as $\lambda f(x_1) + (1-\lambda) f(x_2)$ **(0.5 marks)**.

Common Mistakes: Many students assumed. that x is a real number, but this was never assumed.

(ii) **(2 marks)** Given a vector $x \in \Re^n$, define $f(x)$ as the sum of the k largest coordinates in x, where k is a fixed integer between 0 and n. Prove that f is a convex function.

Solution: For every subset S of size k of $\{1, \ldots, n\}$, define a function $f_S(x)$ which adds the coordinates of x corresponding to locations in S, i.e., $f_S(x) = \sum_{i \in S} x_i$ **(1 mark)**. Now observe that $f(x) = \max_S f_S(x)$ **(1 mark)**and so one can use the previous result.

Common Mistakes: Many students used direct method as in part (a). But then one has to be careful: the top k coordinates of sum of two vectors is not same as sum of the top k coordinates of each of the two vectors. Some people assumed this without proving it, and so lost some points here.

4. **(2 marks)** Suppose $\cos(x)$ is to be approximated by an interpolating polynomial of degree $n+1$, using n equally spaced points in the interval $[0,1]$. How accurate is the approximation (give bounds, no exact answer is needed) ?

We know that the error is of the form Mh^n, where $h = 1/n$ here. **(1 mark)**Here M is an upper bound on the n^{th} derivative of $\cos(x)$, which is at most 1. Thus, the error is at most $1/n^n$. **(1 mark)**

Common Mistakes: Not mentioning the bound as above, and using some sort of expansion.

5. **(3 marks)** Determine a formula of the form

$$\int_a^b f(x)dx = Af(c) + Bf'(a) + Cf''(b),$$

where c is the mid-point of $[a,b]$. The formula should be exact for polynomials of as high degree as possible.

There are 3 parameters A, B, C, and so we can hope to satisfy for all polynomials of degree upto 2, i.e., $1, x, x^2$ **(0.5 marks)**. So this gives the following equations **(1.5 marks)**:

$$\begin{aligned} (b-a) &= A \\ (b^2-a^2)/2 &= A(b+a)/2 + B \\ (b^3-a^3)/3 &= A(b+a)^2/4 + 2Ba + 2C \end{aligned}$$

Solving, we get $A = (b-a), B = 0, C = \frac{(a-b)(a^2+b^2+4ab)}{12}$. **(1 mark)**

Common Mistakes: Mistake in solving equations.

6. **(2 marks)** Compute the LU factorization without pivoting of

$$\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}.$$

We need to subtract twice the first row from the second row, i.e.,

$$\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

Thus, L is the inverse of the first matrix on the LHS above, i.e., $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ **(1 mark)**and U is the matrix on the right. **(1 mark)**

Common Mistakes: Not mentioning proper reasoning and taking the inverse of the matrix on the left.

7. **(5 marks)** Consider the fixed point iteration:

$$x_{k+1} = (\alpha+1)x_k - x_k^2, \quad k = 0, 1, ...$$

where α is a real number satisfying $1/2 \leq \alpha \leq 1$. Show that the iteration converges for any initial guess x_0 if $\alpha - 1/5 \leq x_0 \leq \alpha + 1/5$. Assume that the iteration converges, find the values of α for which it converges quadratically. Argue why it will converge quadratically for this range of the starting point x_0.

Solution: The iteration will converge to α. Therefore, define $e_k = x_k - \alpha$ **(0.5 marks)**. Subtracting α from the above recurrence, we get **(1 marks)**:

$$e_{k+1} = e_k + \alpha x_k - x_k^2 = e_k - e_k x_k = e_k(1 - x_k) = e_k(1 - x_k).$$

Now we claim that $|1 - x_k| \leq 1/2$ always **(1 mark)**. To see this, it is true initially by direct calculation. Now suppose it is true at some time, then $|e_{k+1}| \leq |e_k|/2$ **(1 mark)**. Since x_{k+1} gets even closer to α than x_k, $|1 - x_{k+1}| \leq 1/2$ even now. **(0.5 marks)**

For quadratic convergence, we choose $\alpha = 1$ **(1 mark)**. In this case the above equation becomes $e_{k+1} = -e_k^2$ **(0.5 marks)**.

Common Mistakes: Many students just argued that this is a fixed point iteration of the form $g(x) = x$ for some g and then $g'(\alpha) < 1$. This is not enough. Many people argued that derivative of g is less than 1 in the entire interval of interest. This is ok, but it requires a proof about why this implies convergence (doesn't follow from class notes directly). So if the last part is not shown, I have taken off -1.

8. **(3 marks)** Determine the Householder transformation that annihilates all but the first entry of the vector $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, i.e., find the householder matrix H such that $H \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \\ 0 \\ 0 \end{bmatrix}$, where α is a **negative** number. Clearly write down the vector corresponding to the Householder matrix.

Solution: Clearly, $\alpha = -2$ since lengths are preserved. Let v denote the unit vector joining the above two vectors, i.e., $\frac{1}{\sqrt{12}} \cdot (3,1,1,1)^T$ (or negative of this) **(1 mark)**. This is the normal vector to the hyperplane. We now know that the matrix is **(1 mark)**

$$H = I - 2vv^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 9 & 3 & 3 & 3 \\ 3 & 1 & 1 & 1 \\ 3 & 1 & 1 & 1 \\ 3 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -8 & -3 & -3 & -3 \\ -3 & 0 & -1 & -1 \\ -3 & -1 & 0 & -1 \\ -3 & -1 & -1 & 0 \end{pmatrix}$$

Common Mistakes: Incorrect values or α.

9. **(3 marks)** Consider the matrix $A = \begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix}$. To which eigenvector will the power iteration starting with $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ converge?

   The larger magnitude eigenvalue is 3 **(1.5 marks)**. The unit eigenvector corresponding to this is $\frac{1}{\sqrt{5}}(2\ 1)^T$ **(1.5 marks)**.

   Common Mistakes Not normalizing the eigenvalue, power method always normalizes.

10. **(5 marks)** Let Q be an $n \times n$ unitary matrix (i.e., $UU^* = I$). Consider the following computational problem: given a $n \times n$ matrix A, output QA. Assume that you are given stable algorithms for computing addition and multiplication. Prove that the usual algorithm for computing QA is stable. Note that the input is A, Q is fixed and is not part of the input.

    Solution: Let B denote QA and B' be the computed matrix. We need to show that B' can be expressed as QA' for some matrix A' such that $||A' - A||/||A|| = O(\epsilon)$ **(1 mark)**. We know that each $B'_{ij} = B_{ij}(1 + O(\epsilon))$ because addition and multiplication are stable **(1 mark)**. Now, Since Q is unitary, $A' = Q * B'$ and $A = Q * B$ **(1 mark)**. Therefore

    $$\frac{||A' - A|||}{||A||} = \frac{||B' - B||}{||B||},$$

    because multiplying by unitary matrix does not change norm **(1 mark)**. Now each entry of $B' - B$ is $O(\epsilon)$ times the corresponding entry of B and so the second ratio is $O(\epsilon)$ **(1 mark)**.

    Common Mistakes: Almost no one realized what needed to be proved. They directly started doing some sort of error analysis.