Midterm Exam, Oct 20, 2022

This exam is scheduled for 70 minutes. You can use a one page (front and back) cheat sheet, but no other sources including calculators, phones, lecture notes, or textbooks. There are empty scratch pages on the back, which you can tear off. Good luck!

Name: netID:

1. **Short answers and true/false [40 points]**

Please answer the following questions either with True/False or provide a short answer. You do not need to give details/explanations for these answers.

1	True or False? The fixed point iteration applied to $g(x) = \frac{x+4}{x+1}$ converges to 2, given a sufficiently close starting point.	T
2	When using the bisection method to find a root of $f(x) = 2\sqrt{x} - 1$ with starting interval bounds $(a_0, b_0) = (0, 2)$, what are the interval bounds (a_2, b_2) after *two* bisection steps?	$(0, \frac{1}{2})$
3	Give an example for a sequence x_k that converges linearly to $x^* = 2$ for $k \to \infty$.	$2+(\frac{1}{2})^k$
4	At which convergence order (sublinear, linear, superlinear, quadratic) does the sequence $x_k := 2^{-2^k}$ for $k = 1, 2, \ldots$, converge to 0?	quadratic
5	Assume your computer takes 0.1 seconds to multiply matrices $A, B \in \mathbb{R}^{1000\times 1000}$ with each other. Based on the flop count, how long should it approximately take to multiply two matrices of size 4000×4000?	6.4 sec
6	True or False? For $\boldsymbol{v} \in \mathbb{R}^n$, the definition $N(\boldsymbol{v}) := \Vert\boldsymbol{v}\Vert_2 - \Vert\boldsymbol{v}\Vert_1$ is a norm.	F
7	Give an example for a 2×2 matrix whose induced 1-norm is larger than the induced ∞-norm.	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$
8	Give the 3×3 permutation matrix P that permutes the second and the third row of a matrix when multiplied from the left.	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$
9	What is the induced 1-norm of the 9×9 identity matrix?	1
10	What is the 2-norm based condition number $\kappa_2(A)$ of $A := \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$?	6

2. **Newton's method for optimization [20 points]** Consider a function $f \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$. Recall from single-variable calculus that a necessary condition for $\xi \in \mathbb{R}$ to be the minimum of f is that $f'(\xi) = 0$. Below, we develop a Newton method to solve the nonlinear equation $f'(\xi) = 0$.

   (a) Assuming that $f''(x_k) \neq 0$ for any x_k, write down the Newton iteration for the function $f' \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$.

   Start with $x_0 \in \mathbb{R}$,

   $$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \qquad k = 0, 1, 2, \ldots$$

   (b) Viewing your iteration from part (a) as a fixed-point iteration, write down the corresponding $g : \mathbb{R} \to \mathbb{R}$. Show that the fixed point ξ of g satisfies $f'(\xi) = 0$.

   $$g(x) = x - \frac{f'(x)}{f''(x)}$$

   Fixed point $\xi \Rightarrow \xi = g(\xi) = \xi - \frac{f'(\xi)}{f''(\xi)}$

   $$\Rightarrow -\frac{f'(\xi)}{f''(\xi)} = 0 \Rightarrow f'(\xi) = 0$$

(c) Assuming that $f \in \mathcal{C}^3(\mathbb{R}, \mathbb{R})$, show that if $f''(\xi) \neq 0$, then your iteration from part (a) will have (at least) super-linear convergence. (Hint: recall the condition for local convergence of a fixed-point iteration).

Linear convergence for fixed point problem with

$$\frac{|x_{k+1} - \xi|}{|x_k - \xi|} \xrightarrow{k \to \infty} |g'(\xi)| ; \quad g'(x) = 1 - \frac{f''(x)^2 - f'(x) f'''(x)}{f''(x)^2}$$

$$\Longrightarrow g'(\xi) = 1 - \frac{f''(\xi)^2 - \overset{=0}{f'(\xi)} f'''(\xi)}{f''(\xi)} = 1 - \frac{f''(\xi)^2}{f''(\xi)^2} = 0$$

$\Longrightarrow$ (at least) superlinear convergence

(d) Consider the function $f(x) = \frac{1}{2}x^2$. This function has a minimum as well as a root at $\xi = 0$. Write down the explicit form for the Newton iteration for minimizing f *and* the Newton iteration for root-finding. Which converges faster and why?

$f'(x) = x, \quad f''(x) = 1$

Newton for optimization: $\quad x_{k+1} = x_k - \frac{x_k}{1}$

Newton for root finding: $\quad x_{k+1} = x_k - \frac{\frac{1}{2}x_k^2}{x_k}$

Newton for optimization converges in one iteration to 0, and is thus faster.

3. **LU factorization [15 points]** Let $a > 0$ and consider the matrix $A \in \mathbb{R}^{3\times 3}$,

$$A = \begin{pmatrix} -2 & 1-a & 0 \\ 1+a & -2 & 1-a \\ 0 & 1+a & -2 \end{pmatrix}.$$

(a) For what values of a does an LU factorization without pivoting exist? Justify your answer.

Leading principal matrices:

$A^{(1)} = [-2]$ always invertible

$A^{(2)} = \begin{bmatrix} -2 & 1-a \\ 1+a & -2 \end{bmatrix}$ has determinant of $4-(1+a)(1-a) = 3+a^2$

Not invertible if $a = \pm i\sqrt{3}$, which is not in $\mathbb{R}$

Thus: LU exists for all $a > 0$.

(b) Since A is tridiagonal, its LU factorization (assuming the conditions of part (a) are met) can be written as

$$\begin{pmatrix} -2 & 1-a & 0 \\ 1+a & -2 & 1-a \\ 0 & 1+a & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_2 & 1 & 0 \\ 0 & l_3 & 1 \end{pmatrix} \begin{pmatrix} u_1 & v_1 & 0 \\ 0 & u_2 & v_2 \\ 0 & 0 & u_3 \end{pmatrix}$$

Compute the values of l_2, l_3, v_1, v_2, u_1, u_2, and u_3 for $a = \frac{1}{2}$.

$a = \frac{1}{2} \Rightarrow A = \begin{bmatrix} -2 & \frac{1}{2} & 0 \\ \frac{3}{2} & -2 & \frac{1}{2} \\ 0 & \frac{3}{2} & -2 \end{bmatrix} \Rightarrow u_1 = -2, \ v_1 = \frac{1}{2}$

$l_2 u_1 = \frac{3}{2} \Rightarrow l_2 = -\frac{3}{4}$

$l_2 v_1 + u_2 = -2 \Rightarrow u_2 = -2 + \frac{3}{4} \cdot \frac{1}{2} = -\frac{13}{8}$

$v_2 = \frac{1}{2}$

$l_3 u_2 = \frac{3}{2} \Rightarrow l_3 = \frac{3}{2} \cdot \left(-\frac{8}{13}\right) = -\frac{12}{13}$

$l_3 v_2 + u_3 = -2 \Rightarrow u_3 = -2 + \frac{12}{13} \cdot \frac{1}{2} = -\frac{20}{13}$

4. **Cholesky factorization [15 points]** We say that a matrix $A \in \mathbb{R}^{n\times n}$ is *positive semi-definite* if $A^\mathsf{T} = A$ and if, for every $\boldsymbol{x} \in \mathbb{R}^n$, the quadratic form $\boldsymbol{x}^\mathsf{T} A \boldsymbol{x} \geq 0$.

   (a) Assume that $A = LL^\mathsf{T}$ for $L \in \mathbb{R}^{n\times n}$ a lower-triangular matrix. Show that A is positive semi-definite. (Hint: use that $\boldsymbol{x}^\mathsf{T}\boldsymbol{x} = \|\boldsymbol{x}\|_2^2 \geq 0$).

   $x^T A x = x^T L L^T x = \| L^T x \|_2^2 \geqslant 0$, so A is positive semidefinite.

   (b) It turns out that every positive semi-definite matrix A can be written as $A = LL^\mathsf{T}$ with L lower-triangular. This is called the Cholesky factorization, and the entries of L can be computed according to the formula

$$l_{jj} := \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2}, \quad 1 \leq j \leq n,$$

   $j-1$ mult, $j-1$ additions, 1 squareroot
   overall: $\underline{\underline{2j-1}}$

$$l_{ij} := \frac{1}{l_{jj}}\left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk}\right), \quad 1 \leq j < i \leq n.$$

   $j-1$ mult, $j-1$ addition,
   1 division
   overall: $2j-1$

   Count the floating point operations to compute the Cholesky factorization, only keeping track of leading order terms in n (but do track the constant for the leading-order term). You will need *both* of the below formulas:

$$\sum_{j=1}^{k} j = \frac{k(k+1)}{2} \sim \frac{1}{2}k^2, \qquad \sum_{j=1}^{k} j^2 = \frac{k(k+1)(2k+1)}{6} \sim \frac{1}{3}k^3.$$

   For l_{jj}: $\displaystyle\sum_{j=1}^{n} 2j-1 = 2\sum_{j=1}^{n} j - n \sim n^2 - n$

   For l_{ij}: $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{i-1} 2j-1 = \sum_{i=1}^{n}\left(2\sum_{j=1}^{i-1} j - (i-1)\right)$

$$\sum_{i=1}^{n} i(i-1) - (i-1) = \sum_{i=1}^{n} (i-1)^2 = \sum_{i=1}^{n-1} i^2 = \sim \frac{1}{3}n^3$$

   Overall: $\underline{\underline{\sim \frac{1}{3}n^3 + n^2}}$

(c) How much faster is computing the Cholesky factorization than the standard LU factorization? Give a simple justification for this speedup.

Flops for LU are $\frac{2}{3}n^3$, so Cholesky is by a factor of 2 faster.

5. **Least Squares Problem [20 points]** Let $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ be a matrix, and let $\boldsymbol{b} \in \mathbb{R}^m$ denote a vector.

   (a) Write down the least squares problem for the over-determined system $A\boldsymbol{x} = \boldsymbol{b}$ with $\boldsymbol{x} \in \mathbb{R}^n$ (do not forget to specify the norm used in this formulation!)

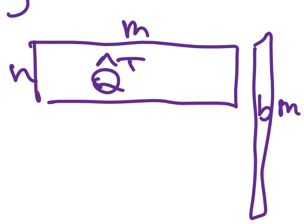
   $$\min_{x \in \mathbb{R}^n} \|Ax-b\|_2$$

   (b) Given a (reduced) QR-factorization $A = \hat{Q}\hat{R}$ with $\hat{Q} \in \mathbb{R}^{m\times n}$ and an upper triangular matrix $\hat{R} \in \mathbb{R}^{n\times n}$, how can you compute the solution of the least squares problem from (a)?

   Solve $\hat{R}x = \hat{Q}^T b$ for $x \in \mathbb{R}^n$

   (c) How many floating point operations (leading-order term and constant) are required for solving the least squares problem in (b), i.e., when you are given the QR-factorization? Besides the flops for the solve with $\hat{R}$, do not forget to include the flops to compute the right hand side of the equation.

   Computing $\hat{Q}^T b$ requires $2m\cdot n$ operations



   Solving with $\hat{R}$ requires n^2 operations, so overall: $n^2 + 2mn$

(d) Given are the three datapoints

	$i=1$	$i=2$	$i=3$
a_i	-1	1	2
y_i	1	2	-1

You would like to fit a quadratic polynomial to these points, i.e., a curve

$$y(a) = x_0 + x_1 a + x_2 a^2, \tag{1}$$

with appropriate coefficients $\boldsymbol{x} = [x_0, x_1, x_2]^T$. Write down the associated least squares system, that is, define the matrix A and the vector $\boldsymbol{b}$ (you don't have to solve the system).

$$\left.\begin{aligned} x_0 + (-1)x_1 + (-1)^2 x_2 &= 1 \\ x_0 + 1x_1 + 1^2 x_2 &= 2 \\ x_0 + 2x_1 + 4x_2 &= -1 \end{aligned}\right\} \quad A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

(e) In (d), $m = n$, i.e., the matrix A is square. What is the value of the norm $\|A\boldsymbol{x}^* - \boldsymbol{b}\|_2$ at the least square solution $\boldsymbol{x}^*$? What does that mean for the best fitting curve (1) and the corresponding data points from the table?

$\|Ax^* - b\|_2 = 0$ as this system is not over-determined.

That means that the curve (1) goes through the 3 data points.