Math 128a - Final Exam - Fall 1998

This exam is open book, open notes, open homework, open calculator (you shouldn't need one). You can use any result from the class notes, text of the book (not unassigned problems!) or homework. The total score is 150 pts. The number of points approximately indicates the number of minutes you should spend on the problem (with 30 minutes left at the end for checking your work).

1) (35 points) We will consider how to use polynomial interpolation to implement the function $\exp(x) = e^x$ available on all computers. Most algorithms use the identity $e^x = 2^z$ where $z = (\log_2 e) \cdot x$, so we will just concentrate on the following algorithm for $y = 2^z$:

1) $z1 = \lfloor z \rfloor$ ... i.e. rounded down to the nearest integer
2) $z2 = z - z1$ ... so $0 \le z2 < 1$ and $2^z = 2^{z_1+z_2} = 2^{z_1} \cdot 2^{z_2}$
3) $s1 = 2^{z1}$
4) $s2 = 2^{z2}$
5) $y = s1 * s2$

One can show that the only source of error is step 4), which we will concentrate on.

1. (25 points) Step 4) is typically implemented using polynomial approximation. We will use polynomial interpolation at the points $0 \le z_0 < z_1 < \cdots < z_n \le 1$. Let $p_n(z)$ denote the polynomial interpolant for 2^z at these points. Give explicit expressions for the z_i that guarantee the error bound

$$\max_{0 \le z \le 1} |p_n(z) - 2^z| \le \frac{1}{(n+1)!} \cdot (\ln 2)^{n+1} \cdot 2^{2-2n}$$

Answer: The standard form of the error bound is

$$\max_{0 \le z \le 1} |p_n(z) - 2^z| \le \frac{1}{(n+1)!} \max_{0 \le \xi \le 1} |f^{(n+1)}(\xi)| \left| \max_{0 \le \xi \le 1} \prod_{i=0}^{n} (\xi - z_i) \right|$$

where $f(\xi) = 2^\xi$. Thus $f(\xi) = e^{(\ln 2)\xi}$ so $f^{(n+1)}(\xi) = (\ln 2)^{n+1} 2^\xi$ and $\max_{0 \le \xi \le 1} |f^{(n+1)}(\xi)| \le (\ln 2)^{n+1} 2^1$. To minimize $|\prod(\xi - z_i)|$ we choose Chebyshev points, namely $z_i = \frac{1}{2}(\cos\left(\frac{2i+1}{2n+2}\pi\right) + 1)$, as defined in Homework 8, Problem 5. In that same problem, we showed $\max_{0 \le \xi \le 1} |\prod_{i=0}^{n} (\xi - z_i)| = 2^{1-2n}$.

2. (10 points) Using the inequalities $\ln 2 < 2^{-1/2}$ and $(n+1)! > 2^{2n}$ (true for $n \ge 6$), how big does n have to be to guarantee an approximation good to double precision, that is $|2^z - p_n(z)| \le 2^{-53}$. Consider only the error from interpolation, not roundoff or other possible sources.

Answer: Substituting these inequalites into the bound from the previous part, we seek that smallest $n \ge 6$ that guarantees $2^{-2n} \cdot 2^{-.5(n+1)} \cdot 2^{2-2n} \le 2^{-53}$, or $-2n - .5(n+1) + 2 - 2n \le -53$ or $n \ge 12$.

2) (25 points) Here we consider cubic splines.

1. (15 points) Determine what constraints a_i, b_i, and c_i, for $i = 1, 2, 3$, must satisfy for $f(x)$ to be a cubic spline. Your answer should consist of equations defining or relating these parameters.

$$f(x) = \begin{cases} a_1(x-4)^2 + a_2(x-3)^3 + a_3(x-2)^4 & x \leq 3 \\ b_1(x-4)^2 + b_2(x-4)^3 + b_3(x-4)^4 & 3 \leq x \leq 5 \\ c_1(x-4)^2 + c_2(x-5)^3 + c_3(x-5)^4 & 5 \leq x \end{cases}$$

   Answer: First, $a_3 = b_3 = c_3 = 0$ for $f(x)$ to be piecewise cubic. Continuity of $f(x)$ at 3 and 5 means $a_1 = b_1 - b_2$ and $c_1 = b_1 + b_2$, respectively. Continuity of $f'(x)$ at 3 and 5 means $-2a_1 = -2b_1 + 3b_2$ and $2c_1 = 2b_1 + 3b_2$. Together $a_1 = b_1 - b_2$ and $-2a_1 = -2b_1 + 3b_2$ mean $b_1 - a_1 = b_2 = (2/3)(b_1 - a_1)$. This means $b_2 = 0$, and $a_1 = b_1$. Similarly, $c_1 = b_1 + b_2$ and $2c_1 = 2b_1 + 3b_2$ mean $c_1 - b_1 = b_2 = (2/3)(c_1 - b_1)$ so $c_1 = b_1 = a_1$. Finally, continuity of $f''(x)$ at 3 and 5 mean $2a_1 = 2b_1 - 6b_2$ and $2c_1 = 2b_1 + 6b_2$, which is also satisfied if $b_2 = 0$ and $a_1 = b_1 = c_1$. a_2 and c_2 can be chosen arbitrarily.

2. (10 points) Determine the values of the parameters a_i, b_i, and c_i so that the cubic spline interpolates the points $(x_0, y_0) = (3, 5)$, $(x_1, y_1) = (0, -1)$ and $(x_2, y_2) = (6, 21)$.

   Answer: Plugging (3,5) into $y = b_1(x-4)^2$ yields $a_1 = b_1 = c_1 = 5$. Plugging (0,-1) into $y = 5(x-4)^2 + a_2(x-3)^3$ yields $a_2 = 3$. Plugging (6,21) into $y = 5(x-4)^2 + c_2(x-5)^3$ yields $c_2 = 1$.

3) (20 points) Determine the truncation errors for the following 2 approximations of $f^{(3)}(x)$, and determine which is more accurate. The truncation error should be in the form (something depending on f)$\cdot h^c + O(h^{c+1})$.

$$\begin{aligned} p_1(x) &= \frac{1}{h^3}[f(x+3h) - 3f(x+2h) + 3f(x+h) - f(x)] \\ p_2(x) &= \frac{1}{2h^3}[f(x+2h) - 2f(x+h) + 2f(x-h) + f(x-2h)] \end{aligned}$$

Answer: Substituting Taylor expansions in h for each of the $f(x+ih)$ in the two expressions, we get $p_1(x) = f^{(3)}(x) + 1.5hf^{(4)}(x) + O(h^2)$ and $p_2(x) = f^{(3)}(x) + .5h^2 f^{(5)}(x) + O(h^3)$, so $p_2(x)$ is more accurate.

4) (35 points) Consider the implicit second order integration formula for $x'(t) = f(x(t))$: $x_{n+1} = x_n + \frac{h}{2}(f(x_{n+1}) + f(x_n))$, $h > 0$. Consider applying this formula to the differential equation $x'(t) = \lambda x(t)$ where λ is a constant. λ may be any complex number $\lambda = \lambda_r + i \cdot \lambda_i$, where $i = \sqrt{-1}$ and λ_r and λ_i are real.

1. (5 points) Write down an explicit expression for x_n (the numerical solution from the formula) in terms of $x_0 = x(0)$.
2. (5 points) Write down an explicit expression for $x(t)$ (the true solution) in terms of $x(0)$.
3. (5 points) Under what conditions on λ does $\lim_{n\to\infty} |x_n| = 0$ for any $x_0 \neq 0$?
4. (5 points) Under what conditions on λ does $\lim_{n\to\infty} |x_n| = \infty$ for any $x(0) \neq 0$?
5. (5 points) Under what conditions on λ does $\lim_{t\to\infty} |x(t)| = 0$ for any $x_0 \neq 0$?
6. (5 points) Under what conditions on λ does $\lim_{t\to\infty} |x(t)| = \infty$ for any $x(0) \neq 0$?
7. (5 points) Show that $\lim_{n\to\infty} |x_n| = \lim_{t\to\infty} |x(t)|$ for any $x(0) = x_0$.

Answer:

1. $x_n = \left(\frac{1+h\lambda/2}{1-h\lambda/2}\right)^n x_0$
2. $x(t) = e^{\lambda t} x(0)$
3. $\lim_{n\to\infty} |x_n| = 0$ for any x_0 if and only if

$$\begin{aligned} 1 &> \left|\frac{1+h\lambda/2}{1-h\lambda/2}\right| \\ &= \left(\frac{(1+h\lambda_r/2)^2 + (h\lambda_i)^2}{(1-h\lambda_r/2)^2 + (h\lambda_i)^2}\right)^{1/2} \\ &\equiv \gamma \end{aligned}$$

   which is true if and only if $\lambda_r < 0$.

4. The limit is ∞ if and only if $\gamma > 1$, which is true if and only if $\lambda_r > 0$.
5. Since $|x(t)| = |e^{\lambda_r t + i\cdot\lambda_i t} x(0)| = |e^{\lambda_r t}| \cdot |e^{i\cdot\lambda_i t}| \cdot |x(0)| = e^{\lambda_r t}|x(0)|$, the limit goes to zero for all $x(0)$ if and only if $\lambda_r < 0$.
6. The limit is infinite for all $x(0)$ if and only if $\lambda_r > 0$.
7. The only cases not considered yet are $x_0 = 0$ (trivial) and $\lambda_r = 0$. In the latter case one can see that $|x_n| = |x_0|$ for all n since $x = 1$, and also $|x(t)| = |x(0)|$ for all t.

5) (35 points) This question is about computing null vectors, i.e. nonzero vectors x satisfying $A \cdot x = 0$.

1. (15 points) Suppose $A = \begin{bmatrix} A_1 & A_2 \\ 0 & 0 \end{bmatrix}$ is an n-by-n matrix where A_1 is i-by-i and nonsingular, A_2 is i-by-$n-i$ and the last $n-i$ rows are zero. Then it is a fact that the rank of A is i, so that there are $n-i$ linearly independent null vectors x_1,...,x_{n-i}. Given an algorithm for computing x_1,...,x_{n-i}. Note that they are not uniquely defined; any linearly independent set of $n-i$ null vectors will do. Your algorithm should be expressed at the level of "Factor the matrix B into $B = PLU$ using Gaussian elimination with partial pivoting" or "Solve the lower triangular system $Lx = y$ using forward substitution" rather than writing out loops in detail. Be sure to show that your vectors are independent.

   Answer: Suppose $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ is a null vector of A, where z_1 has i components, and z_2 has $n - i$ components. Then

$$0 = A \cdot z = \begin{bmatrix} A_1 & A_2 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} A_1 \cdot z_1 + A_2 \cdot z_2 \\ 0 \end{bmatrix}$$

   If we choose $z_2 \neq 0$, then we can solve for z_1, getting $z_1 = -A_1^{-1} \cdot (A_2 \cdot z_2)$, which can be solved by a matrix vector multiplication $A_2 \cdot z_2$ and then Gaussian elimination with partial pivoting. It remains to say how to pick $n-i$ different z_2 vectors so that the resulting $n-i$ z vectors are linearly independent. Write the j-th vector as $z^{(j)} = \begin{bmatrix} z_1^{(j)} \\ z_2^{(j)} \end{bmatrix}$. Then they can be linearly dependent, i.e.

$$0 = \sum_{j=1}^{n-i} \alpha_j z^{(j)} = \begin{bmatrix} \sum_{j=1}^{n-i} \alpha_j z_1^{(j)} \\ \sum_{j=1}^{n-i} \alpha_j z_2^{(j)} \end{bmatrix}$$

   only if the $z_2^{(j)}$ are linearly dependent; thus to guarantee independence it suffices to pick $z_2^{(1)}, ..., z_2^{(n-i)}$ independent. There is a lot of choice here; a simple one is to let $z_2^{(j)}$ be the j-th column of the $n-i$-by-$n-i$ identity matrix.

2. (10 points) Let A be an n-by-n matrix. Show that if A has rank i, then at step $i+1$ of Gaussian elimination with complete pivoting, the largest entry found in the submatrix $A(i+1:n, i+1:n)$ is zero. (Ignore roundoff.)

   Answer: Using notation from class, at the end of step i of Gaussian elimination, we have factored the matrix A (possibly with its rows and columns permuted) into $(I + L_i) \cdot A_i$, where L_i is possibly nonzero only below the diagonal in columns $1, ..., i$, A_i is zero in the same locations, and the first i diagonal entries of A_i are nonzero. $I + L_i$ is nonsingular, so a vector is a null vector of A if and only it is a (possibly permuted) null vector of A_i, and in particular A and A_i have the same rank. By the fact stated in part 1, the rank of A is therefore i if GECP finds nonzero pivots for the first i steps, and then the last $n-i$ rows of A_i are all zero.

3. (10 points) Combine the last two parts to give an algorithm for determining the rank i of a matrix, and a linearly independent set of $n - i$ null vectors. (Again ignore roundoff.)

Answer:

1) Do GECP to get $P_r \cdot A \cdot P_c = L \cdot U$, where
   we stop when the last $n - i$ rows of U are zero.
2) Find $n - i$ independent null vectors $x_1, ..., x_{n-i}$ of U using Part 1.
3) Form $P_c x_j$ for $j = 1, ..., n - i$; these are the desired null vectors.