Preliminary Examination, Numerical Analysis, January 2008

Instructions: This exam is closed books and notes. The time allowed is three hours and you need to work on any three out of questions 1-5 and any two out of questions 6-8. All questions have equal weights and the passing score will be determined after all the exams are graded. Indicate clearly the work that you wish to be graded.

Note: In problems 6-8, the notations $k = \Delta t$ and $h = \Delta x$ are used. Note also that at the end of the exam there is a list of Facts some of which may be useful to you.

1) Singular Value Decomposition Let A be a real $m \times n$ matrix $(m \geq n)$ of full rank. We are interested in the singular value decomposition of $A = U\Sigma V^T$. It is known that algorithms for SVD can be derived by turning the SVD problem into an eigenvalue problem. One such approach is to find the eigenvalue decomposition of the $(m+n)\times(m+n)$ symmetric matrix

$$\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}.$$

Establish the connections between the singular values, left and right singular vectors of A and the eigenvalues, eigenvectors of the above matrix. You should pay particular attention to the case $m > n$. Comment on the advantages and disadvantages of this approach compared with an approach based on the eigendecomposition of $A^T A$.

2) Interpolation and Integration:

a) Consider equally spaced points $x_j = a + jh$, $j = 0, \ldots, J+1$ on the interval $[a, b]$, where $(J+1)h = b - a$. Let $f(x)$ be a function defined on $[a, b]$. Consider the problem of finding a cubic spline approximation $s(x)$ to $f(x)$ that interpolates f at the points x_j, is twice continuously differentiable, and satisfies $s''(a) = s''(b) = 0$. Does this problem always have a solution? If your answer is yes, derive formulas by which to determine the spline. If your answer is no, explain your reasoning.

b) Let $I_n(f)$ denote the result of using the composite Trapezoidal rule to approximate $I(f) \equiv \int_a^b f(x)dx$ using n equally sized subintervals of length $h = (b-a)/n$. It can be shown that the integration error $E_n(f) \equiv I(f) - I_n(f)$ satisfies

$$E_n(f) = d_2h^2 + d_4h^4 + d_6h^6 + \ldots$$

where $d_2, d_4, d_6, \ldots$ are numbers that depend only on the values of f and its derivatives at a and b. Suppose you have a black-box program that, given f, a, b, and n, calculates $I_n(f)$. Show how to use this program to obtain an $O(h^4)$ approximation and an $O(h^6)$ approximation to $I(f)$.

3) Iterative Methods:
Consider the fixed-point iteration

$$\mathbf{u}^{(k+1)} = T\mathbf{u}^{(k)} + \mathbf{c}$$

for finding a solution of the problem

$$\mathbf{u} = T\mathbf{u} + \mathbf{c},$$

where T is an $m \times m$ real matrix and $\mathbf{c}$ is a real m-vector.

a) Show that the fixed point iteration will converge for an arbitrary initial guess $\mathbf{u}^{(0)}$ if and only if the spectral radius of T, $\rho(T)$, is less than 1.

b) Consider the boundary value problem

$$-u''(x) = f(x), \quad \text{for } 0 \le x \le 1$$

with $u(0) = u(1) = 0$, and the following discretization of it:

$$-U_{j-1} + 2\,U_j - U_{j+1} = F_j,$$

for $j = 1, 2, \ldots, N-1$ where $Nh = 1$ and $F_j \equiv h^2\, f(jh)$.

Show that the Jacobi iterative method will converge for this problem for any choice of initial guess. Express the speed of convergence as a function of the discretization stepsize h. How does the number of iterations required to reduce the initial error by a factor δ depend on h? In practice, would you use this method to solve the given problem? If so, explain why this is a good idea? If not, how would you solve it in practice?

4) Sensitivity:

Consider a 6×6 symmetric positive definite matrix A with singular values $\sigma_1 = 1000$, $\sigma_2 = 500$, $\sigma_3 = 300$, $\sigma_4 = 20$, $\sigma_5 = 1$, $\sigma_6 = 0.01$.

a) Suppose you use a Cholesky factorization package on a computer with a machine epsilon 10^{-14} to solve the system $Ax = b$ for some nonzero vector b. How many digits of accuracy do you expect in the computed solution? Justify your answer in terms of condition and stability. You may assume that the entries of A and b are exactly represented in the computer's floating-point system.

b) Suppose that instead you use an iterative method to find an approximate solution to $Ax = b$ and you stop iterating and accept iterate $x^{(k)}$ when the residual $r^{(k)} = Ax^{(k)} - b$ has 2-norm less than 10^{-9}. Give an estimate of the maximum size of the relative *error* in the final iterate? Justify your answer.

5) Linear Least Squares: The Linear Least Squares problem for an $m \times n$ real matrix A and $b \in \mathbb{R}^m$ is the problem:

$$\text{Find} \quad \mathrm{x} \in \mathbb{R}^{\mathrm{n}} \text{ such that } \|\mathrm{Ax} - \mathrm{b}\|_2 \text{ is minimized.}$$

a) Suppose that you have data $\{(t_j, y_j)\}$, $j = 1, 2, \ldots, m$ that you wish to approximate by an expansion

$$p(t) = \sum_{k=1}^{n} x_k \phi_k(t).$$

Here, the functions $\phi_k(t)$ are given functions. Which norm on the difference between the approximation function p and the data gives rise to a linear least squares problem for the unknown expansion coefficients x_k? What is the matrix A in this case, and what is the vector b?

b) Suppose that A is a real $m \times n$ matrix of full rank and let $b \in \mathbb{R}^m$. What are the 'normal equations' for the Least Squares problem? How can they be used to solve the Least Squares problem? What is the QR factorization of A and how can it be used to solve the Least Squares problem? Compare and contrast these approaches for numerically solving the Least Square problem.

6) Elliptic Problems:

Consider the standard five-point difference approximation (centered difference for both the gradient and divergence operators) for the variable coefficient Poisson equation

$$-\nabla \cdot (a\nabla v) = f$$

with Dirichlet boundary conditions, in a two-dimensional rectangular region. We assume that $a(x, y) \geq a_0 > 0$. The approximate solution $\{u_{i,j}\}$ satisfies a linear system $Au = b$.

1. State and prove the maximum principle for the numerical solution $u_{i,j}$.
2. Derive the matrix A in the one-dimensional case and show that it is symmetric and positive definite.
3. For the one-dimensional and *constant-coefficient* case, show that the global error $e_j = v(x_j) - u_j$ satisfies $\|e\|_2 = O(h^2)$ as the space step $h \to 0$.
4. Discuss the advantages and disadvantages of trying to solve the system for the two-dimensional problem using (i) the SOR (Successive Over Relaxation) method and (ii) the (preconditioned) Conjugate Gradient method.

7) Heat Equation Stability:

a) Consider the initial value problem for the constant-coefficient diffusion equation

$$v_t = \beta v_{xx}, \ t > 0$$

with initial data $v(x, 0) = f(x)$. The Crank-Nicolson scheme for this problem is:

$$\frac{u_j^{n+1} - u_j^n}{k} = \frac{\beta}{2h^2}\left\{u_{j-1}^{n+1} - 2u_j^{n+1} + u_{j+1}^{n+1} + u_{j-1}^n - 2u_j^n + u_{j+1}^n\right\}.$$

Analyze the 2-norm stability of this scheme and show that the scheme is stable for any choice of $k > 0$ and $h > 0$.

b) Consider the variable coefficient diffusion equation

$$v_t = (\beta v_x)_x, \qquad 0 < x < 1, \ t > 0$$

with Dirichlet boundary conditions

$$v(0, t) = 0, \qquad v(1, t) = 0$$

and initial data $v(x, 0) = f(x)$. Assume that $\beta(x) \geq \beta_0 > 0$, and that $\beta(x)$ is smooth. Let $\beta_{j+1/2} = \beta(x_{j+1/2})$. The Crank-Nicolson scheme for this problem is:

$$\begin{aligned}\frac{u_j^{n+1} - u_j^n}{k} = \frac{1}{2h^2} \ &\left\{\beta_{j-1/2}u_{j-1}^{n+1} - (\beta_{j-1/2} + \beta_{j+1/2})u_j^{n+1} + \beta_{j+1/2}u_{j+1}^{n+1}\right.\\ &\left. + \beta_{j-1/2}u_{j-1}^n - (\beta_{j-1/2} + \beta_{j+1/2})u_j^n + \beta_{j+1/2}u_{j+1}^n\right\}.\end{aligned}$$

Analyze the 2-norm stability of this scheme for solving this initial boundary value problem. Do not neglect the fact that there are boundary conditions!

8) Numerical Methods for ODEs: Consider the Linear Multistep Method

$$y_{n+2} - \frac{4}{3}y_{n+1} + \frac{1}{3}y_n = \frac{2}{3}hf_{n+2}$$

for solving an initial value problem $y' = f(y, x)$, $y(0) = \eta$. You may assume that f is Lipschitz continuous with respect to y uniformly for all x.

a) Analyze the consistency, stability, accuracy, and convergence properties of this method.

b) Would it be more reasonable to use this method or Euler's method for the initial value problem:

$$y' = -10^8[y - \cos(x)] - \sin(x), \quad y(0) = 1?$$

Justify your answer.