Numerical Analysis

INSTRUCTIONS FOR QUALIFYING EXAMS

Start each problem on a new sheet of paper. Write your university identification number at the top of each sheet of paper. **DO NOT WRITE YOUR NAME!**

Complete this sheet and staple to your answers. Read the directions of the exam carefully.

STUDENT ID NUMBER: ____

DATE: ____

EXAMINEES: DO NOT WRITE BELOW THIS LINE
****

1. ______________________
2. ______________________
3. ______________________
4. ______________________
5. ______________________
6. ______________________
7. ______________________
8. ______________________

Pass/fail recommend on this form.

Total score: ____________________

DO NOT FORGET TO WRITE YOUR SID NO. ON YOUR EXAM.

There are 8 problems. Problems 1-4 are worth 5 points and problems 5-8 are worth 10 points. All problems will be graded and counted towards the final score.

You have to demonstrate a sufficient amount of work on both groups of problems [1-4] and [5-8] to obtain a passing score.

---

[1] (5 Pts.) Consider numerically evaluating the soft-max function $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ defined as

$$\big(\mu(z)\big)_i = \frac{e^{z_i}}{\sum_{j=1}^{d} e^{z_j}} \qquad \text{for } i = 1, \ldots, d.$$

When the following pseudocode is evaluated

```
d = 5
z = [700, 800, 1000, 900, -40]
out = zeros(d)
S = 0
for i = 1,...,d
   S += exp(z[i])
for i = 1,...,d
   out[i] = exp(z[i])/S
```

an overflow error occurs. Why does this error occur? (Hint. Pay attention to the exponential.) How can we fix this problem?

[2] (5 Pts.) Let $a > 0$. Consider the problem of computing $\sqrt{a}$ with the Newton iteration

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

for $n = 0, 1, \ldots$ with $f(x) = x^2 - a$ and $x_0 > 0$.

(a) Show that if $x_0 < \sqrt{a}$, then $x_1 > \sqrt{a}$.

(b) Show that if $x_n > \sqrt{a}$, then $\sqrt{a} < x_{n+1} < x_n$.

(c) Show that $x_n \to \sqrt{a}$.

[3] (5 Pts.) Assuming that $f \in C^4[a, b]$ is real, derive the error of approximation when the second order derivative is substituted by the finite-difference formula

$$f''(x) \approx \frac{f(x+h) - 2\, f(x) + f(x-h)}{h^2},$$

where the parameter h is called the mesh size (assume that x, $x + h$, $x - h \in (a, b)$).

[4] (5 Pts.)
(a) Consider the linear system $Ax = b$ in the unknown x, with $x, b \in \mathbb{R}^n$ and $A = M - N \in \mathbb{R}^{n\times n}$ is nonsingular. If M is nonsingular and if $(M^{-1}N)^k \to O$ as $k \to \infty$, show that the iterates x_k, defined by

$$Mx_{k+1} = Nx_k + b,$$

converge to $x = A^{-1}b$ for any starting vector x_0. **(b)** Find a splitting $A = M - N$ for the matrix $A = \begin{pmatrix} 10 & -1 \\ -1 & 10 \end{pmatrix}$, so that the iteration in (a) is convergent. Justify your answer.

[5] (10 Pts.) Consider the linear constant-coefficient system of ODEs

$$\frac{du}{dt} = Au, \qquad u(0) = u_0$$

for $t \geq 0$, where $u_0 \neq 0$ and

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

(a) Show that $\|u(t)\| = \|u_0\|$ for all $t \geq 0$.

(b) Consider finding an approximate solution of the ODE using the one-stage Runge–Kutta method with Butcher tableau

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

with a sufficiently small time discretization $h > 0$. Show that $\|u^k\| \to \infty$.

(c) Consider finding an approximate solution of the ODE using the two-stage Runge–Kutta method with Butcher tableau

$$\begin{array}{c|cc} 0 & & \\ 1 & 1 & \\ \hline & 1/2 & 1/2 \end{array}$$

with a sufficiently small time discretization $h > 0$. Show that $\|u^k\| \to \infty$.

(d) Consider finding an approximate solution of the ODE using the two-stage Runge–Kutta method with Butcher tableau

$$\begin{array}{c|cc} 0 & & \\ 1 & 1 & \\ \hline & 0 & 1 \end{array}$$

with a sufficiently small time discretization $h > 0$. Show that $\|u^k\| \to 0$.

Clarification. For an ODE $y' = f(t, y)$, an explicit s-stage Runge–Kutta method takes the form

$$\begin{aligned} y_{n+1} &= y_n + h\sum_{i=1}^{s} b_i k_i \\ k_1 &= f(t_n, y_n), \\ k_2 &= f(t_n + c_2 h, y_n + h(a_{21}k_1)), \\ k_3 &= f(t_n + c_3 h, y_n + h(a_{31}k_1 + a_{32}k_2)), \\ &\vdots \\ k_i &= f\left(t_n + c_i h, y_n + h\sum_{j=1}^{i-1} a_{ij}k_j\right). \end{aligned}$$

The *Butcher tableau* puts the coefficients of the method in a table as:

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \dots & a_{1s} \\ c_2 & a_{21} & a_{22} & \dots & a_{2s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \dots & a_{ss} \\ \hline & b_1 & b_2 & \dots & b_s \end{array}$$

[6] **(10 Pts.)** Consider the scalar initial/boundary value problem:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(a(x)\frac{\partial u}{\partial x}\right) + b\cos(\pi x)\frac{\partial u}{\partial x}$$

for $0 < x < 1$, $t > 0$, $u(x, 0) = u_0(x)$, b is a nonzero constant, and $a(x)$ smooth.

(a) If $a(x)$ vanishes identically, what boundary conditions, if any, do we need at $x = 0$ and $x = 1$ to make the problem well posed?

(b) What conditions on $a(x)$ guarantee well posedness in general if we assume periodic boundary conditions?

(c) Write a convergent finite difference for the vanishing $a(x)$ case.

Justify your answers.

[7] (10 Pts.) Consider the equation

$$\frac{\partial^2 u}{\partial t^2} + b\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

for b constant to be solved for $t > 0$, $0 < x < 1$, with initial data $u(x, 0)$ given and $\frac{\partial u}{\partial y}(x, 0)$ given and with periodic boundary conditions in x.

(a) Write a convergent finite difference approximation to this problem.

(b) How do you expect the solution to behave as t gets very large?

Justify your answers.

[8] (10 Pts.) Develop and describe the piecewise-linear Galerkin finite element approximation of

$$\begin{aligned} -\triangle u + u &= f(x,y), \quad (x,y) \in T, \\ u &= g_1(x), \quad (x,y) \in T_1, \\ u &= g_2(y), \quad (x,y) \in T_2, \\ \frac{\partial u}{\partial \vec{n}} &= h(x,y), \quad (x,y) \in T_3, \end{aligned}$$

where

$$\begin{aligned} T &= \{(x,y)|\ x > 0,\ y > 0,\ x + y < 1\} \\ T_1 &= \{(x,y)|\ y = 0,\ 0 < x < 1\} \\ T_2 &= \{(x,y)|\ x = 0,\ 0 < y < 1\} \\ T_3 &= \{(x,y)|\ x > 0,\ y > 0,\ x + y = 1\}, \end{aligned}$$

and $\vec{n}$ denotes the exterior unit normal to the boundary, ∂T.

(a) Derive the weak variational formulation of the problem.

(b) Give the necessary assumptions on the functions f, g_1, g_2, and h. and verify the assumptions of the Lax-Milgram Lemma by analyzing the appropriate linear and bilinear forms.

(c) Develop and describe the piecewise linear Galerkin finite element approximation of the problem and a set of basis functions such that the corresponding linear system is sparse. Show that this linear system has a unique solution. Give a convergence estimate and quote the appropriate theorems for convergence.