Department of Applied Mathematics
Preliminary Examination in Numerical Analysis
August 2022

Instructions. You have three hours to complete this exam. Submit solutions to four (and no more) of the following six problems. All problems have equal value.

Please start each problem on a new page. You MUST prove your conclusions or show a counter-example for all problems unless otherwise noted.
Write your student ID number (not your name!) on your exam.

Problem 1: Root finding

Consider the root finding problem $f(x) = 0$ for $f \in C^2(\mathbb{R})$. For an initial guess x_0 near the single root α, consider the iteration scheme:

$$x_{k+1} = x_k - \gamma_0 f(x_k)$$

Where $\gamma_0 = \frac{1}{f'(x_0)}$. Sometimes this method is called the *chord iteration.*

(a) Rewrite the chord iteration in terms of a fixed point iteration for a function $g(x)$. State necessary conditions (in terms of $f(x)$ and x_0) for the convergence of this scheme to the root $x = \alpha$ of $f(x)$.

(b) Consider the function $f(x) = \sin(3\pi x) + 8x - 4$. One can show it has a unique, simple root at $\alpha \simeq 0.58607509$. Find a set of initial points x_0 for which the chord iteration, if it converges to α, does so *quadratically.*

(c) Recall that the Newton iteration for approximating a root of $f(x)$ is given by

$$x_{k+1} = x_k + p_k$$

where $p_k = -f(x)/f'(x)$.

Consider now an *inexact* Newton iteration where the update to the approximation at step k is q_k where q_k is an approximation of $p_k = -f(x)/f'(x)$. (The chord iteration is an example of an inexact Newton iteration.) This iteration is defined by

$$x_{k+1} = x_k + q_k.$$

Write down an expression for the error of the *inexact Newton* method $e_{k+1} = x_{k+1} - \alpha$ at the $(k+1)^{\text{th}}$ step in terms of the error of the (exact) *Newton* method at the $(k+1)^{\text{th}}$ step.

(d) Assume the difference between the updates in the Newton and inexact Newton iterations at step $k+1$ satisfy

$$|p_k - q_k| \leq \eta_k |f(x_k)|$$

where η_k is small; i.e. your approximate update is close to the Newton update.
With this bound and your solution from part (c), derive an upper bound for the error e_{k+1} of the inexact Newton step. Determine the convergence rate of this method.

Solution:

(a) The chord iteration is a fixed point iteration scheme for the function

$$g(x) = x - \gamma_0 f(x)$$

We know $g \in C^2(\mathbb{R})$, and $g'(x) = 1 - \gamma_0 f'(x)$. By continuity of $g'(x)$, we know that if

$$|g'(\alpha)| = |1 - \gamma_0 f'(\alpha)| < 1$$

Then the chord iteration will converge linearly to α for x_0 in a neighborhood of α and

$$\lim_{n\to\infty} \frac{\alpha - x_{n+1}}{\alpha - x_n} = 1 - \gamma_0 f'(\alpha)$$

The inequality implies that $2 > \gamma_0 f'(\alpha) > 0$, i.e., $2 > f'(\alpha)/f'(x_0)$ and $f'(x_0)$ must be the same sign as $f'(\alpha)$.

(b) We know that for x_0 sufficiently close to α, our fixed point iteration will converge at least quadratically if $g'(\alpha) = 0$. That is,

$$g'(\alpha) = 1 - \gamma_0 f'(\alpha) = 0$$

This means we must choose x_0 such that $f'(x_0) = f'(\alpha)$. We notice that the derivative of $f(x)$ is a periodic function, with period $2/3$. So, assuming the iteration converges, the choice $x_0 = \alpha + (2/3)k$ for $k \in \mathbb{Z}$ will converge at least quadratically.

(c) Given the definition for the inexact Newton iteration:

$$\begin{aligned} x_{k+1} &= x_k + q_k \\ e_{k+1} &= (x_k + p_k) - \alpha + (q_k - p_k) \end{aligned}$$

Which means the error is the sum of two terms: the exact Newton error and the difference between the exact and inexact Newton steps.

(d) We take the estimate obtained in the point above,

$$\begin{aligned} |e_{k+1}| &\le |x_k + p_k - \alpha| + |q_k - p_k| \\ &\le C|e_k|^2 + \eta_k |f(x_k)| \end{aligned}$$

$|f(x_k)| = |f(x_k) - f(\alpha)|$. By the Mean Value Theorem, there exists ζ between x_k and α such that $|f(x_k)| = |f'(\zeta) e_k| \le |f'(\zeta)||e_k|$. Finally, since $f'(x)$ is assumed to be continuous, there exist M such that $|f'(x)| \le M$ on this interval. So,

$$|e_{k+1}| \le C|e_k|^2 + \eta_k |f'(\zeta)| \le C|e_k|^2 + \eta_k M |e_k|$$

This tells us that the inexact Newton will converge quadratically only if η_k is small enough (smaller or comparable to $|e_k|$). Otherwise, it will converge linearly.

Problem 2: Quadrature
For functions defined on a closed interval $[0, 1]$, we want to compute the definite integral

$$I[f] = \int_0^1 f(x) \log(1/x) dx$$

that is, with a logarithmic weight function $\log(1/x) = -\log(x)$. It is possible to find a basis of orthogonal polynomials $P_n(x)$ for the corresponding weighted inner product, and use them to derive n-point Gaussian quadratures with nodes x_k^n and weights ω_k^n. In this problem, you will explore these polynomials and techniques for creating the corresponding Gaussian quadrature.

(a) Let $P_0(x) = 1$. Use Gram-Schmidt or another technique to find $P_1(x)$. Find the corresponding quadrature node x_1^1 and weight ω_1^1 for the 1-point Gaussian quadrature rule.

You may use the formula $\int_0^1 x^m \log(1/x) dx = \frac{1}{(m+1)^2}$.

(b) A general result for orthogonal polynomials is the *interlacing theorem*: considering the partition of $[a, b]$ into $n + 1$ subintervals with endpoints $a < x_n^1 < x_n^2 \cdots < x_n^n < b$, there is a unique node for the $n + 1$ rule on each of these intervals.

Based on this information, state a quadratically convergent algorithm to find the quadrature nodes for the $(n + 1)$ rule, given accurate nodes for the n point rule.

(c) The family P_n satisfies a recursion formula of the form:

$$P_{n+1}(x) = (x - \alpha_n)P_n(x) - \beta_n P_{n-1}(x)$$

Consider the normalized family $Q_n(x) = P_n(x)/\sqrt{h_n}$, with $h_n = \int_0^1 P_n(x)^2 \log(1/x)dx$. Given that $\beta_n = h_n/h_{n-1}$, find a recursive formula for $Q_n(x)$ of the form:

$$\sqrt{\beta_{n+1}} Q_{n+1}(x) = (x - \alpha_n)Q_n(x) - \sqrt{\beta_n} Q_{n-1}(x)$$

(d) Consider the recursive formula above for $n = 0, 1, 2, 3$. Show that $x = \lambda$ is a node of the 4 point Gaussian quadrature if and only if it is an eigenvalue of a symmetric, tridiagonal matrix with diagonal $[\alpha_0, \alpha_1, \alpha_2, \alpha_3]$ and super / sub diagonal $[\sqrt{\beta_1}, \sqrt{\beta_2}, \sqrt{\beta_3}]$.

Indicate why deriving this eigenvalue problem using the normalized polynomials might be preferable to solving the eigenvalue problem with the polynomials from the original recursive formula.

Solution:

(a) We use the Gram-Schmidt process to find:

$$P_1(x) = x - \frac{< x, 1 >_\omega}{< 1, 1 >_\omega} 1 = x - \frac{1}{4}$$

So the single quadrature node is $x_1^1 = 1/4$. We can find the corresponding weight by requiring that this rule integrate constants exactly. That yields:

$$\int_0^1 \log(1/x)dx = 1 = \omega_1^1$$

(b) Each quadrature node x_k^{n+1} is one of the roots of $P_{n+1}(x)$, and by the interlacing theorem, it is between $a_k = x_{k-1}^n$ and $b_k = x_k^n$ (where $a_1 = a, b_{n+1} = b$).

Our algorithm would use bisection until the midpoint lies in the basin of convergence for the Newton iteration:

$$y_k^{m+1} = y_k^m - P_{n+1}(y_k^m)/P'_{n+1}(y_k^m).$$

Newton would then be used until the desired accuracy was achieved.

(c) We substitute $P_n(x) = \sqrt{h_n} Q_n(x)$ in the original recursive formula. We then divide both sides by $\sqrt{h_n}$:

$$\begin{aligned}
\sqrt{h_{n+1}} Q_{n+1}(x) &= (x - \alpha_n)\sqrt{h_n} Q_n(x) - \beta_n \sqrt{h_{n-1}} Q_{n-1}(x) \\
\frac{\sqrt{h_{n+1}}}{\sqrt{h_n}} Q_{n+1}(x) &= (x - \alpha_n) Q_n(x) - \frac{h_n}{h_{n-1}} \frac{\sqrt{h_{n-1}}}{\sqrt{h_n}} Q_{n-1}(x) \\
\sqrt{\beta_{n+1}} Q_{n+1}(x) &= (x - \alpha_n) Q_n(x) - \sqrt{\beta_n} Q_{n-1}(x)
\end{aligned}$$

(d) We write down the formulas for $n = 0, 1, 2, 3$, separating the term $xP_n(x)$:

$$\begin{aligned}
\sqrt{\beta_1} Q_1(x) + \alpha_0 Q_0(x) &= x Q_0(x) \\
\sqrt{\beta_2} Q_2(x) + \alpha_1 Q_1(x) + \sqrt{\beta_1} Q_0(x) &= x Q_1(x) \\
\sqrt{\beta_3} Q_3(x) + \alpha_2 Q_2(x) + \sqrt{\beta_2} Q_1(x) &= x Q_2(x) \\
\sqrt{\beta_4} Q_4(x) + \alpha_3 Q_3(x) + \sqrt{\beta_3} Q_2(x) &= x Q_3(x)
\end{aligned}$$

We plug in $x = \lambda$ and collect these equations in matrix form:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ \sqrt{\beta_4} Q_4(\lambda) \end{bmatrix} + \begin{bmatrix} \alpha_0 & \sqrt{\beta_1} & 0 & 0 \\ \sqrt{\beta_1} & \alpha_1 & \sqrt{\beta_2} & 0 \\ 0 & \sqrt{\beta_2} & \alpha_2 & \sqrt{\beta_3} \\ 0 & 0 & \sqrt{\beta_3} & \alpha_3 \end{bmatrix} \begin{bmatrix} Q_0(\lambda) \\ Q_1(\lambda) \\ Q_2(\lambda) \\ Q_3(\lambda) \end{bmatrix} = \lambda \begin{bmatrix} Q_0(\lambda) \\ Q_1(\lambda) \\ Q_2(\lambda) \\ Q_3(\lambda) \end{bmatrix}$$

If λ is one of the roots of $Q_4(x)$, the first term above vanishes and hence it is an eigenvalue of the symmetric tridiagonal matrix, with eigenvector $[Q_0(\lambda), Q_1(\lambda), Q_2(\lambda), Q_3(\lambda)]$. Since Q_4 has 4 distinct simple roots, it follows that the spectra of this matrix is the set of 4-point Gaussian quadrature nodes.

If we use the original recursive formula, the resulting matrix will be tridiagonal, but non-symmetric which is less preferable from an algorithmic point of view.

Problem 3: Interpolation / Approximation

The Chebyshev polynomials of the third kind, sometimes known as airfoil polynomials, are defined as

$$V_n(x) = \frac{\cos(n + \frac{1}{2})\theta}{\cos \frac{\theta}{2}}$$

where $x = \cos\theta$, n non-negative integer.

There is a table of trigonometric identities provided at the end of this problem for your convenience.

(a) Show that the Chebyshev polynomials of the third kind satisfy the recursion:

$$\begin{aligned} V_0(x) &= 1 \\ V_1(x) &= 2x - 1 \\ V_{n+1}(x) &= 2xV_n(x) - V_{n-1}(x) \end{aligned}$$

(b) Show that these polynomials are an orthogonal basis for polynomials in $[-1, 1]$ with respect to the inner product:

$$< f, g >= \int_{-1}^{1} f(x)g(x)\sqrt{\frac{1+x}{1-x}}dx$$

That is, with weight function $w(x) = \sqrt{\frac{1+x}{1-x}}$.

Hint: Decide which representation of the polynomials will make the integrals the easiest.

(c) Let $q(x) = \sum_{k=0}^{2} C_k V_k(x)$ be the polynomial that minimizes

$$\min \int_{-1}^{1} (e^x - p(x))^2 \sqrt{\frac{1+x}{1-x}}dx$$

over the space of all polynomials of degree ≤ 2.

Using the results from part (b) derive explicit formulas for the coefficients C_0, C_1, C_2.

You do not have to evaluate the formulas for the coefficients.

Table of trigonometric identities

$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$
$\sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)$
$\sin^2(\theta/2) = (1 - \cos\theta)/2$
$\cos^2(\theta/2) = (1 + \cos\theta)/2$
$\tan^2(\theta/2) = (1 - \cos\theta)/(1 + \cos\theta)$
$\cos(\alpha)\cos(\beta) = \frac{1}{2}(\cos(\alpha + \beta) + \cos(\alpha - \beta))$
$\sin(\alpha)\sin(\beta) = \frac{1}{2}(\cos(\alpha - \beta) - \cos(\alpha + \beta))$

Solution:

(a) $V_0(x) = \frac{\cos(\theta/2)}{\cos(\theta/2)} = 1$ follows directly from the definition.

To show $V_1(x) = \frac{\cos(3\theta/2)}{\cos(\theta/2)} = 2\cos\theta - 1$, we use angle summation formulas:

$$\begin{aligned}\frac{\cos(3\theta/2)}{\cos(\theta/2)} &= \cos\theta - \frac{\sin(\theta)\sin(\theta/2)}{\cos(\theta/2)} \\ &= \cos\theta - 2\sin^2(\theta/2) \\ &= \cos\theta + \cos\theta - 1 = 2x - 1\end{aligned}$$

In order to demonstrate the recurrence relation, we note that the numerator of $2xV_n(x)$ is $2\cos(\theta)\cos(n+1/2)\theta$. We use the trigonometric identity $2\cos(\alpha)\cos(\beta) = \cos(\alpha+\beta) + \cos(\alpha-\beta)$:

$$2\cos(\theta)\cos(n+1/2)\theta = \cos(n+3/2)\theta + \cos(n-1/2)\theta$$

Dividing both sides of this equation by $\cos(\theta/2)$ yields the recurrence relation.

(b) We write down the inner product of V_n and V_m, and then change integration variables from x to θ.

$$\begin{aligned}< V_n, V_m > &= \int_{-1}^{1} V_n(x)V_m(x)\sqrt{\frac{1+x}{1-x}}dx \\ &= \int_0^{\pi} \frac{\cos(n+1/2)\theta\cos(m+1/2)\theta}{\cos^2(\theta/2)}\sqrt{\frac{1+\cos\theta}{1-\cos\theta}}\sin\theta d\theta\end{aligned}$$

Using half angle formulas, we can see that

$$\begin{aligned}\frac{1+\cos\theta}{1-\cos\theta} &= \frac{\cos^2\theta/2}{\sin^2(\theta/2)} = \cot^2(\theta/2) \\ \frac{\sin\theta}{\cos^2(\theta/2)} &= 2\frac{\sin(\theta/2)}{\cos(\theta/2)} = \tan(\theta/2)\end{aligned}$$

So, we get

$$\begin{aligned}< V_n, V_m > &= 2\int_0^{\pi} \cos(n+1/2)\theta\cos(m+1/2)\theta d\theta \\ &= \int_0^{\pi} \cos(n+m+1)\theta + \cos(n-m)\theta d\theta\end{aligned}$$

When $n \neq m$, this inner product is thus equal to 0.

(c) Because the basis of polynomials we are using is orthogonal, we have formulas for each coefficient in terms of inner products with $f(x) = e^x$. Using the recurrence relation, we find that $V_2(x) = 4x^2 - 2x - 1$. We can then find that:

$$C_0 = \frac{< e^x, V_0 >}{< V_0, V_0 >} = \frac{\int_{-1}^{1} e^x w(x) dx}{\int_{-1}^{1} w(x) dx}$$

$$C_1 = \frac{< e^x, V_1 >}{< V_1, V_1 >} = \frac{\int_{-1}^{1} e^x (2x - 1) w(x) dx}{\int_{-1}^{1} (2x - 1)^2 w(x) dx}$$

$$C_2 = \frac{< e^x, V_2 >}{< V_2, V_2 >} = \frac{\int_{-1}^{1} e^x (4x^2 - 2x - 1) w(x) dx}{\int_{-1}^{1} (4x^2 - 2x - 1)^2 w(x) dx}$$

Problem 4: Linear algebra

(a) State and apply the Gerschgorin theorem to the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \epsilon & \epsilon \\ \epsilon & 2 & \epsilon \\ \epsilon & \epsilon & 2 \end{bmatrix}$$

with $\epsilon << 1$.

It is often possible to improve the Gerschgorin bounds on the eigenvalue estimate by first applying a similarity transformation to the matrix $\mathbf{A}$ involving a diagonal matrix $\mathbf{D}_n = \mathrm{diag}(d_1, \ldots, d_n)$.

(b) Prove that a reduction in the error bound of one eigenvalue for the matrix above must occur at the expense of relaxing the bounds for the remaining eigenvalues.

(c) Let $\mathbf{D}_3 = \mathrm{diag}(1, k\epsilon, k\epsilon)$, show that the bounding radius for $\lambda_1 \sim 1$ can be reduced from $\rho_1 = 2\epsilon$ to $2\epsilon^2$.

(d) How would you use the Gerschgorin circle for $\lambda \sim 1$ to compute a numerical approximation for the eigenvalue in that ball?

Solution:

(a) The Gerschgorin circle theorem says that all eigenvalues live in the union of the balls in the complex plane defined by

$$|\lambda - a_{ii}| \leq \sum_{j=1//j\neq i}^{n} |a_{ij}|.$$

Applying this theorem to the problem we find that the eigenvalues live in the balls

$$|\lambda - 1| \leq 2\epsilon \quad \text{and} \quad |\lambda - 2| \leq 2\epsilon$$

in the complex plane.

(b)

$$\mathbf{D}^{-1}\mathbf{A}\mathbf{D} = \begin{bmatrix} \frac{1}{d_1} & 0 & 0 \\ 0 & \frac{1}{d_2} & 0 \\ 0 & 0 & \frac{1}{d_3} \end{bmatrix} \begin{bmatrix} 1 & \epsilon & \epsilon \\ \epsilon & 2 & \epsilon \\ \epsilon & \epsilon & 2 \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \epsilon\frac{d_2}{d_1} & \epsilon\frac{d_3}{d_1} \\ \epsilon\frac{d_1}{d_2} & 2 & \epsilon\frac{d_3}{d_2} \\ \epsilon\frac{d_1}{d_3} & \epsilon\frac{d_2}{d_3} & 2 \end{bmatrix}$$

The Gerschgorin circle theorem eigenvalue balls are

$$
\begin{aligned}
|\lambda_1 - 1| &\le \epsilon \left(\left|\frac{d_2}{d_1}\right| + \left|\frac{d_3}{d_1}\right| \right) \\
|\lambda_2 - 2| &\le \epsilon \left(\left|\frac{d_1}{d_2}\right| + \left|\frac{d_3}{d_2}\right| \right) \\
|\lambda_3 - 2| &\le \epsilon \left(\left|\frac{d_2}{d_3}\right| + \left|\frac{d_2}{d_3}\right| \right))
\end{aligned}
$$

Suppose our goal is to reduce the size of the circle for λ_1, then $|d_2| << |d_1|$ and $|d_3| << |d_1|$. This means that

$$\left|\frac{d_1}{d_2}\right| > 1 \text{ and } \left|\frac{d_1}{d_3}\right| > 1.$$

Thus the circles for λ_2 and λ_3 are larger than the original circles.

(c)

$$\mathbf{D}^{-1}\mathbf{A}\mathbf{D} = \begin{pmatrix} 1 & k\epsilon^2 & k\epsilon^2 \\ \frac{1}{k} & 2 & \epsilon \\ \frac{1}{k} & \epsilon & 2 \end{pmatrix}$$

Thus the circle centered at 1 now has radius $2k\epsilon^2$.

(d) One way to use the circle theorem to find an eigenvalue is to use a shifted inverse power iteration where the shift lies in the circle of the eigenvalue you are looking for.

Note: The similarity transformation may not help with the eigenvalues in concentric circles.

Problem 5: Numerical ODE

(a) Derive the coefficients c_0, c_1 and c_2 for the backward differentiation formula

$$c_0 u_{n+1} + c_1 u_n + c_2 u_{n-1}$$

which has a truncation error that is second order for approximating $u'(t_{n+1})$.
You must derive the truncation error for this problem.
Be mindful of where you are approximating the derivative.

(b) Show that applying this scheme to approximate the solution of the ODE

$$u' = f(t, u)$$

is convergent.

For the remainder of this problem consider a stiff linear system of ODEs of the form

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}, \quad \mathbf{u}(0) = \mathbf{u}_0 \tag{1}$$

where $\mathbf{A}$ is a negative definite matrix.

(c) Explain what is meant by a stiff system.

(d) Determine if the scheme derived in part (a) is suitable for solving the stiff system of ODEs (1).

Hint: You don't have to derive the entire region stability domain. It suffices to establish stability for the part of the complex plane that is relevant for the system in (1).

Solution:

(a) We are approximating the derivative at t_{n+1} so we will use Taylor expansions around that point for the other two terms in the expression. Also since we want a second order method, we should go out to the third derivative.

$$u(t_n) = u(t_{n+1}) - hu'(t_{n+1}) + \frac{h^2}{2}u''(t_{n+1}) - \frac{h^3}{3!}u^{(3)}(t_{n+1}) + \cdots$$
$$u(t_{n-1}) = u(t_{n+1}) - 2hu'(t_{n+1}) + \frac{4h^2}{2}u''(t_{n+1}) - \frac{8h^3}{3!}u^{(3)}(t_{n+1}) + \cdots$$

To get a second order approximation, we need to add these expressions together in a way that the u'' term is destroyed. To do this, I'll take 4 times the first equation and subtract the second.

The result is

$$4u(t_n) - u(t_{n-1}) = 3u(t_{n+1}) - 2hu'(t_{n+1}) + \frac{4h^3}{3!}u^{(3)}(t_{n+1}) + \cdots$$

Now all we have to do is solve for $u'(t_{n+1})$ to get our approximation.

$$u'(t_{n+1}) = \frac{-4u(t_n) + u(t_{n-1}) + 3u(t_{n+1})}{2h} + \frac{2h^2}{3!}u^{(3)}(t_{n+1}) + \cdots$$

So our constants are $c_0 = \frac{3}{2h}$, $c_1 = -\frac{2}{h}$, and $c_2 = \frac{1}{2h}$.

(b) To show that method is convergent, we need to show that it is consistent and stable. Since the truncation error goes to zero as $h \to 0$, the method is consistent.

To determine stability, we need to look at the roots of the corresponding characteristic polynomial:

$$p(\lambda) = -\frac{3}{2}\lambda^2 + 2\lambda - \frac{1}{2}.$$

If all the roots lie inside the unit ball in the complex plane, then the method is stable.

The roots are $\lambda = -1$ and $\lambda = 1/3$. Both lie in the unit ball so the method is stable.

Since the method is both consistent and stable, it is convergent.

(c) A differential equation is called stiff if there is requirement on the time step size for the time stepping method to be stable when applied to it.

(d) A negative definite matrix is a Hermitian (symmetric in the case of real matrices) matrix that has negative eigenvalues.

To determine the region of stability for the method we found in part (a), we must apply it to test problem:

$$u'(t) = \alpha u$$

for $\alpha < 0$ and determine for what values of h the method is convergent.

Applying the method to this problem, we find the characteristic polynomial is

$$(-h\alpha - \frac{3}{2})r^2 + 2r - \frac{1}{2}$$

We need to look at the roots of this polynomial. I will simplify that equation

$$(2h\alpha + 3)r^2 - 4r + 1 = 0$$

They are

$$r = \frac{2 \pm \sqrt{1 - 2h\alpha}}{2h\alpha + 3}$$

Note that $r = \frac{2+\sqrt{1-2h\alpha}}{2h\alpha+3} > 1$ for all h. Thus the method is not stable for this problem.

Problem 6: Numerical PDE
Consider the advection equation

$$u_t + u_x = 0 \tag{2}$$

with periodic boundary conditions.

(a) Consider solving (2) by first discretizing u_x in space via the centered difference approximation

$$u_x \approx \frac{u(x+h,t) - u(x-h,t)}{2h}.$$

Using this discretization, write down the matrix $\mathbf{A}$ such that (2) can be written as

$$\mathbf{u}_t = \mathbf{A}\mathbf{u}. \tag{3}$$

(b) Consider solving the ODE system (3) using explicit Euler in time. Using properties of $\mathbf{A}$, show that this scheme is unsuitable for solving (2). Explain what the failure mechanism is. (You may state the domain for explicit Euler without proof.)

(c) Consider solving (2) using the discretization

$$\frac{u_m^{n+1} - u_m^{n-1}}{2k} + \frac{u_{m+1}^n - u_{m-1}^n}{2h} = 0$$

where $u_m^n = u(mh, nk)$, k denotes the time step size and h denotes the space step size. Use von Neumann analysis to determine the amplification factor and the region of stability.

Solution:

(a) The matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} 0 & -\frac{1}{2h} & 0 & \cdots & \cdots & 0 & \frac{1}{2h} \\ \frac{1}{2h} & 0 & -\frac{1}{2h} & 0 & \cdots & \cdots & 0 \\ 0 & \frac{1}{2h} & 0 & -\frac{1}{2h} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & \frac{1}{2h} & 0 & -\frac{1}{2h} \\ -\frac{1}{2h} & 0 & \cdots & \cdots & 0 & \frac{1}{2h} & 0 \end{bmatrix}$$

(b) The region of absolute stability for explicit Euler is the circle centered at -1 with radius 1 in the complex plane. $\mathbf{A}$ is skew-symmetric thus all of it's eigenvalues are purely imaginary. Thus the condition of the eigenvalues lying in the region of stability for explicit Euler is not met and the method is not guaranteed to converge.

(c) We begin this problem by plugging $u_m^n = e^{at_n}e^{ibx_m}$ into the difference equation. This yields the following equation

$$\frac{e^{a(t_n+k)}e^{ibx_m} - e^{a(t_n-k)}e^{ibx_m}}{2k} + \frac{e^{a(t_n)}e^{ib(x_m+h)} - e^{a(t_n)}e^{ib(x_m-h)}}{2h} = 0$$

We now remove the common factor of $e^{at_n}e^{ibx_m}$ to get

$$\frac{1}{2k}\left(\eta - \frac{1}{\eta}\right) - \frac{i\sin(bh)}{h} = 0$$

where $\eta = e^{ak}$ is the amplification factor. This is a quadratic in η and we know that the region of stability is the where $|\eta| \leq 1$ where

$$\eta = i\frac{k}{h}\sin(bh) \pm \sqrt{1 - \frac{k^2}{h^2}\sin^2(bh)}.$$

We need to investigate conditions on k and h that will force $|\eta| \leq 1$.

If $1 - \frac{k^2}{h^2}\sin^2(bh) \geq 0$. Then

$$\eta_1 = i\frac{k}{h}\sin(bh) + \sqrt{1 - \frac{k^2}{h^2}\sin^2(bh)}$$

and

$$\eta_2 = i\frac{k}{h}\sin(bh) - \sqrt{1 - \frac{k^2}{h^2}\sin^2(bh)}$$

and $|\eta_j|^2 = 1$ for $j = 1, 2$.

Then we find that $0 \leq \frac{k^2}{h^2}\sin^2(bh) \leq 1$. Since $|\sin^2(bh)| \leq 1$, we find that the condition is that $0 \leq \frac{k}{h} \leq 1$.

If $1 - \frac{k^2}{h^2}\sin^2(bh) < 0$, then $\frac{k^2}{h^2}\sin^2(bh) > 1$ and we can write the amplification factors as

$$\eta_1 = i\frac{k}{h}\sin(bh) + i\sqrt{\frac{k^2}{h^2}\sin^2(bh) - 1}$$

and

$$\eta_2 = i\frac{k}{h}\sin(bh) - i\sqrt{\frac{k^2}{h^2}\sin^2(bh) - 1}$$

.

Then $|\eta_j|^2 = 2\frac{k^2}{h^2}\sin^2(bh) - 1$ for $j = 1, 2$. Thus we need

$$2\frac{k^2}{h^2}\sin^2(bh) - 1 < 1$$

which means that $\frac{k^2}{h^2}\sin^2(bh) < 1$ which is a contraction to our assumption about $1 - \frac{k^2}{h^2}\sin^2(bh) < 0$.

Thus the only way that method can be stable is if $\frac{k}{h} < 1$.