Department of Applied Mathematics
Preliminary Examination in Numerical Analysis
January, 2019

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

1. **Root Finding.** Consider the fixed point iteration scheme

$$x_{n+1} = g(x_n).$$

   (a) State the necessary conditions for the convergence of such a scheme to fixed point $x = \alpha$.

   (b) Given

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

   Find an upper bound for the absolute error $|\alpha - x_n|$.

   (c) Derive from first principles the expression that shows the rootfindng method to be pth order convergent.

   (d) Consider the following iteration for calculating $\gamma^{1/3}$:

$$x_{n+1} = ax_n + b\frac{\gamma}{x_n^2} + c\frac{\gamma^2}{x_n^5}$$

   (e) Assuming that this iterative scheme converges for x_0 sufficiently close to $\gamma^{1/3}$ determine a, b, c such that the method has the highest possible convergence rate.

(a) Assume $g(x) \in \boldsymbol{C}^1([a,b])$, $g([a,b]) \subset [a,b]$, and

$$\lambda = \max_{a \le x \le b} |g'| < 1$$

then $x = g(x)$ has a unique fixed point $\alpha \in [a,b]$, $\forall x_0 \in [a,b]$ with $x_{n+1} = g(x_n)$, $n \ge 0$

$$\lim_{n \to \infty} x_n = \alpha,$$

and (although not necessary)

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

(b) By manipulating the Taylor expansion of $(\alpha - x_{n+1}) = \alpha - g(x_n)$ with the above result

$$|\alpha - x_n| \le \frac{\lambda^n}{1-\lambda}|x_1 - x_0|$$

(c) Order of convergence: If $g([a,b]) \in \mathbb{C}^p([a,b])$ for $p \ge 2$ and

$$g'(\alpha) = \cdots = g^{p-1}(\alpha) = 0$$

then for x_0 sufficiently close to α

$$\lim_{n \to \infty} \frac{\alpha - x_{n+1}}{(\alpha - x_n)^p} = (-1)^{p-1} \frac{g^{(p)}(\alpha)}{p!}.$$

(d) First note that there are three unknowns so that we need 3 equations to uniquely determine a, b and c. We also know a-priori that the method should converge with order $p = 3$. We know that $\gamma^{\frac{1}{3}}$ is a fixed point so that $g(\gamma\frac{1}{3}) = \gamma^{\frac{1}{3}}$ implying

$$a + b + c = 1.$$

Also we will require $g'(\gamma\frac{1}{3}) = 0$ giving

$$a - 2b - 5c = 0,$$

and that $g''(\gamma\frac{1}{3}) = 0$ resulting in

$$6b + 30c = 0.$$

Solving this system yields $a = \frac{5}{9}$, $b = \frac{5}{9}$ and $c = \frac{-1}{9}$.

2. **Linear Alegbra.** Let $A \in \mathbb{R}^{n\times n}$ be symmetric and positive definite (spd), and consider the following iteration.

```
Choose A_0 = A
for k = 0, 1, 2, ...
    Compute the Cholesky factor L_k of A_k (so A_k = L_k L_k^T)
    Set A_{k+1} = L_k^T L_k
end
```

Here L_k is lower triangular with positive diagonal elements.

(a) Show that A_k is similar to A, and that A_k is spd (the iteration is therefore well-defined).

(b) Now consider the special case of a 2×2 spd matrix,

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \qquad a \geq c,$$

For this matrix, perform one step of the algorithm and write down A_1.

(c) Use the result from **(b)** to argue that A_k converges to $\operatorname{diag}(\lambda_1, \lambda_2)$, where the eigenvalues of A are ordered as $\lambda_1 \geq \lambda_2 > 0$.

Solution. For **(a)** $A_k = L_{k-1}^T L_{k-1}$ is manifestly symmetric and $\langle \mathbf{x}, A_k \mathbf{x}\rangle = \|L_{k-1}\mathbf{x}\|_2^2$. The right-hand side $\|L_{k-1}\mathbf{x}\|_2^2 \neq 0$, since L_{k-1} is nonsingular (A_{k-1} is spd and so nonsingular). Therefore, A_k is spd. Note also that

$$\begin{aligned} A_k &= L_{k-1}^T L_{k-1} \\ &= L_{k-1}^{-1}(L_{k-1}L_{k-1}^T)L_{k-1} \\ &= L_{k-1}^{-1} A_{k-1} L_{k-1} \\ &= L_{k-1}^{-1} \dots L_0^{-1} A L_0 \dots L_{k-1}. \end{aligned}$$

For **(b)** the Cholesky factorization of $A = A_0$ is

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} = \begin{pmatrix} \sqrt{a} & 0 \\ b/\sqrt{a} & \sqrt{ac-b^2}/\sqrt{a} \end{pmatrix} \begin{pmatrix} \sqrt{a} & b/\sqrt{a} \\ 0 & \sqrt{ac-b^2}/\sqrt{a} \end{pmatrix}.$$

Since A is spd, both $a = e_1^T A e_1$ and $ac - b^2$ have to be positive; therefore, the square roots in the Cholesky factorization are well defined. Then

$$\begin{aligned} A_1 &= \begin{pmatrix} \sqrt{a} & b/\sqrt{a} \\ 0 & \sqrt{ac-b^2}/\sqrt{a} \end{pmatrix} \begin{pmatrix} \sqrt{a} & 0 \\ b/\sqrt{a} & \sqrt{ac-b^2}/\sqrt{a} \end{pmatrix} \\ &= \begin{pmatrix} a + b^2/a & (b/a)\sqrt{ac-b^2} \\ (b/a)\sqrt{ac-b^2} & c - b^2/a \end{pmatrix} \\ &= \begin{pmatrix} a_1 & b_1 \\ b_1 & c_1 \end{pmatrix}. \end{aligned}$$

For **(c)** since A_1 is similar to A, we know it has the same eigenvalues. Also, in the one iteration the off-diagonal terms b have been multiplied by a factor (assume $b \neq 0$ or the problem is trivial)

$$\frac{\sqrt{ac - b^2}}{a} < \frac{\sqrt{ac}}{a} \leq \frac{\sqrt{a^2}}{a} = 1.$$

This implies that the magnitude $|b_1| < |b|$. Note also $a_1 \geq a$ and $0 < c_1 \leq c$ (the inequality $0 < c_1$ holds since A_1 is again spd). These results will then apply to each iteration. A_k has eigenvalues λ_1 and λ_2, with $a_k \geq a_{k-1}$, $0 < c_k \leq c_{k-1}$, and $|b_k| < |b_{k-1}|$. The two-norm $\|A_k\|_2 = \lambda_1$ is then constant, which implies the entries of A_k remain bounded as $k \to \infty$. The sequences then converge: $b_k \to 0$ and $\lim a_k \geq \lim c_k$ (since $a \geq c$, and a_k and c_k are respectively non-decreasing and non-increasing). So $\lim a_k = \lambda_1$ and $\lim c_k = \lambda_2$.

3. **Numerical quadrature.**

   (a) State the simple (one panel) midpoint and trapezoidal rules for approximating the integral $\int_0^h f(x)\,dx$.

   (b) Derive the error formulas for both methods.

   (c) Let $f(x) = x^3$ and show that it is possible to combine the results from the two methods so that the answer is exact.

   (d) Show that you also get the exact answer if you perform Richardson extrapolation (using the error expansion from the composite rule) using the answers obtained by the trapezoidal method with one panel (as above) and the trapezoidal method with two panels of equal size $h/2$.

(a) The Trapezoidal rule is:

$$T = \frac{h}{2}(f(h) + f(0)).$$

The midpoint rule is:

$$M = hf(\frac{h}{2})$$

(b) The error formula for the Trapezoidal method can for example be found by the Peano Theorem. The error is

$$E_T f = \int_0^h f(x)dx - \frac{h}{2}(f(h) + f(0)),$$

and when $f = p$ is a linear polynomial we know that $E_T p = 0$. The Peano kernel for this problem can thus be computed as

$$K(u) = \int_0^h (x-u)_+ dx - \frac{h}{2}((x-u)_+ + 0) = \frac{(h-u)^2}{2} - \frac{h(h-u)}{2} = -\frac{u(h-u)}{2} < 0,$$

where $u \in [0, h]$. By the Peano theorem we also get

$$\frac{E_T x^2}{2!} = \int_0^h \frac{x^2}{2} dx - \frac{h}{2}\frac{h^2}{2} = -\frac{h^3}{12},$$

thus the error is

$$E_T f = -\frac{h^3}{12} f''(\xi), \quad \xi \in (0, h).$$

Alternatively the error can be obtained by the exact remainder formula in Newton interpolation. For the midpoint rule a direct application of Taylor's theorem yields

$$f(x) - (f(\frac{h}{2}) + (x - \frac{h}{2})f'(\frac{h}{2}) = \frac{1}{2}(x - \frac{h}{2})^2 f''(\xi), \quad \xi \in [0, h].$$

Integrating the above expression give the error for the midpoint method

$$E_M f = \int_0^h \frac{1}{2}(x - \frac{h}{2})^2 f''(\xi)dx = \frac{h^3}{24} f''(\xi), \quad \xi \in (0, h).$$

(c) For $f = x^3$ the exact integral is $h^4/4$. A better approximation to the integral is obtained by $(2M + T)/3$, this cancels the leading order errors. Here the result is

$$\frac{2M+T}{3} = \frac{1}{3}\left(2h\left(\frac{h}{2}\right)^3 + \frac{h}{2}\left(h^3+0\right)\right) = \frac{h^4}{3}\left(\frac{1}{4}+\frac{2}{4}\right) = \frac{h^4}{4}.$$

(d) The composite rule has error expansion $\mathcal{O}(h^2)$ so the Richardson extrapolated approximation is

$$RE = \frac{4T(h/2) - T(h)}{3}.$$

Here

$$T(h) = \frac{h^4}{2}, \quad T(h/2) = \frac{h}{2}\left(\frac{h^3}{2} + \left(\frac{h}{2}\right)^3 + 0\right) = h^4\frac{5}{16}.$$

Thus,

$$RE = \frac{4T(h/2) - T(h)}{3} = \frac{h^4}{3}\left(\frac{5}{4} - \frac{2}{4}\right) = \frac{h^4}{4}.$$

4. **Interpolation/Approximation.** We are given three data points as follows:

x	-1	0	1
y	1	-1	-1

Determine the interpolating polynomial of lowest degree possible using

(a) Lagrange's interpolation formula,

(b) Newton's interpolation formula,

(c) Stirling's interpolation formula,

Verify that the answers you get agree.

(d) Quote the formula for the error, applied to this special case.

Hint: In standard operator notation on a grid with spacing h, Stirling's interpolation formula can be written

$$f(x_0 + th) = f_0 + t\mu\delta f_0 + \frac{t^2}{2!}\delta^2 f_0 + \frac{t(t^2-1)}{3!}\mu\delta^3 f_0 + \frac{t^2(t^2-1)}{4!}\delta^4 f_0 + \cdots$$

(a) Lagrange:

$$\frac{(x-0)(x-1)}{(-1-0)(-1-1)}(1) + \frac{(x+1)(x-1)}{(0+1)(0-1)}(-1) + \frac{(x+1)(x-0)}{(1+1)(1-0)}(-1)$$
$$= \frac{x^2-x}{2} + \frac{x^2-1}{1} - \frac{x^2+x}{2} = x^2 - x - 1$$

(b) Newton's Divided differences:

x	y		
-1	1		
		-2	
0	-1		1
		0	
1	-1		

It follows

$$p(x) = 1 - 2(x+1) + 1(x+1)x = 1 - 2x - 2 + x^2 + x = x^2 - x - 1$$

(c) Stirlings: Use $x_0 = 0, h = 1, t = x$. Regular (non-divided) differences.

x	y	δ	δ^2
-1	1		
		-2	
0	-1		2
		0	
1	-1		

It follows

$$p(x) = -1 + x(-1) + \frac{x^2}{2}2 = x^2 - x - 1$$

(d) Error

$$\frac{(x+1)(x)(x-1)}{3!} f^{(3)}(\zeta) \quad \text{with} \quad \zeta \in [-1, 1].$$

5. **ODEs** Consider the initial value problem $y' = f(t, y)$, $y(t_0) = y_0$. The Milne method is a linear multistep method defined by

$$y_n = y_{n-2} + \int_{t_{n-2}}^{t_n} P(t)dt$$

where $P(t)$ is the unique quadratic polynomial that interpolates f at the points t_{n-2}, t_{n-1}, t_n, and $t_n = hn$.

(a) Derive the formula for this method.

(b) Find the leading order term in the local truncation error of this method. What is the order of this method?

(c) Is this method 0-stable, strongly stable? Why?

For (a) note that $P(t)$ interpolates f_{n-2}, f_{n-1} and f_n and can for example be found by Lagrange interpolation

$$p(t) = f_{n-2}L_1(t) + f_{n-1}L_2(t) + f_n L_3(t),$$

where

$$L_1(t) = \left(\frac{t - t_{n-1}}{t_{n-2} - t_{n-1}}\right)\left(\frac{t - t_n}{t_{n-2} - t_n}\right) = \frac{(t - t_{n-1})(-t_n)}{2h^2},$$

$$L_2(t) = \left(\frac{t - t_{n-2}}{t_{n-1} - t_{n-1}}\right)\left(\frac{t - t_n}{t_{n-1} - t_n}\right) = -\frac{(t - t_{n-2})(-t_n)}{h^2},$$

$$L_3(t) = \left(\frac{t - t_{n-1}}{t_n - t_{n-1}}\right)\left(\frac{t - t_{n-2}}{t_n - t_{n-2}}\right) = \frac{(t - t_{n-1})(-t_{n-2})}{2h^2}.$$

The method follows from

$$\int_{t_{n-2}}^{t_n} P(t)dt = f_{n-2}\int_{t_{n-2}}^{t_n} L_1(t)dt + f_{n-1}\int_{t_{n-2}}^{t_n} L_2(t)dt + f_n\int_{t_{n-2}}^{t_n} L_3(t)dt.$$

Here

$$\int_{t_{n-2}}^{t_n} L_1(t)dt = \frac{h}{3}, \quad \int_{t_{n-2}}^{t_n} L_2(t)dt = \frac{4h}{3}, \quad \int_{t_{n-2}}^{t_n} L_3(t)dt = \frac{h}{3},$$

so the method is

$$y_n = y_{n-2} + \frac{h}{3}(f_n + 4f_{n-1} + f_{n-2}).$$

For (b), to find the order p, we may check that (see e.g. THM 2.4 in Hairer, Norsett, Wanner, Solving ordinary differential equations I: Nonstiff problems (1993), HNW)

$$\sum_{i=0}^{k} \alpha_i = 0, \quad \sum_{i=0}^{k} \alpha_i i^q = q\sum_{i=0}^{k} \beta_i i^{q-1}, \quad \text{for} \quad q = 1, \ldots, p.$$

Here (in the notation of eq. (2.1) in HNW)

$$\alpha_0 = -1,\ \alpha_1 = 0, \alpha_2 = 1, \quad \beta_0 = \frac{1}{3},\ \beta_1 = \frac{4}{3}, \beta_2 = \frac{1}{3},$$

So

$$\alpha_0 + \alpha_1 + \alpha_2 = 0,$$

and

$$\alpha_1 + 2\alpha_2 = 2 = \beta_0 + \beta_1 + \beta_2,$$

and

$$\alpha_1 + 2^2\alpha_2 = 4 = 2(\frac{4}{3} + 2\frac{1}{3}) = 2(\beta_1 + 2\beta_2),$$

and

$$\alpha_1 + 2^3\alpha_2 = 8 = 3(\frac{4}{3} + 4\frac{1}{3})) = 3(\beta_1 + 2^2\beta_2),$$

and

$$\alpha_1 + 2^4\alpha_2 = 16 = 4(\frac{4}{3} + 8\frac{1}{3})) = 4(\beta_1 + 2^3\beta_2),$$

but

$$\alpha_1 + 2^5\alpha_2 = 32 \neq \frac{100}{3} = 5(\frac{4}{3} + 16\frac{1}{3})) = 5(\beta_1 + 2^4\beta_2).$$

Thus, the order of the method is four. The leading order error term in the local error (with $p = 4$) is (again see e.g. HNW eq. (2.5))

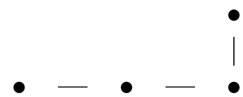
$$\frac{d^5y}{dt^5}\frac{h^5}{120}\left(\sum_{i=0}^{2}\alpha_i i^5 - 5\sum_{i=0}^{2}\beta_i i^4\right) = -\frac{d^5y}{dt^5}\frac{h^5}{90}.$$

To answer (c) note that $\rho(z) = z^2 - 1$ with roots $z = \pm 1$, the method is thus zero stable but not strongly stable as there are two roots with $|z| = 1$.

6. **PDEs** Consider the following finite difference scheme:

$$\frac{u(x,t+k)-u(x,t)}{k}+\frac{\frac{3}{2}u(x,t)-2u(x-h,t)+\frac{1}{2}u(x-2h,t))}{h}=0$$

Graphically, we illustrate its stencil as shown below

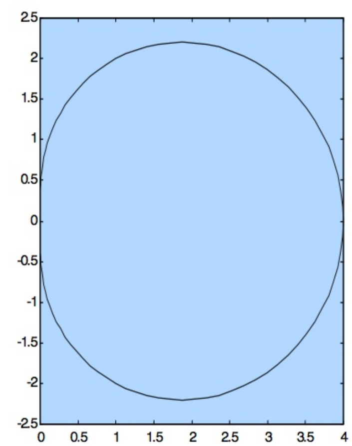


(a) Determine which PDE the scheme is consistent with,

(b) Determine its order of accuracy in time and in space (when applied to the the PDE from (a)),

(c) Use von Neumann analysis to determine the scheme's stability.

Hint to part (c): The figure below shows the curve traced out by $f(s)=\frac{3}{2}-2e^{is}+\frac{1}{2}e^{-2is}$ for $-\pi \le s \le \pi$.



(a) Taylor expansion of the two terms in the difference scheme gives:

$$\frac{u(x,t+k)-u(x,t)}{k}=u_t+\frac{k}{2}u_{tt}+\mathcal{O}(k^2)$$

$$\frac{\frac{3}{2}u(x,t)-2u(x-h,t)+\frac{1}{2}u(x-2h,t))}{h}=u_x-\frac{h^2}{3}u_{xxx}+\mathcal{O}(h^3)$$

Hence, the difference scheme is consistent with $u_t+u_x=0$.

(b) From the error terms above, we see immediately that the scheme is first order in time and second order in space.

(c) Substitute $u(x,t) = \zeta^{t/k} e^{i\omega x}$ in

$$\frac{u(x,t+k) - u(x,t)}{k} + \frac{\frac{3}{2}u(x,t) - 2u(x-h,t) + \frac{1}{2}u(x-2h,t))}{h} = 0$$

and simplifying gives

$$\zeta = 1 - \lambda(\frac{3}{2} - 2e^{-i\omega h} + \frac{1}{2}e^{-2i\omega h})$$

We set $\omega h = s$ and let $f(s)$ be the function defined and displayed in the problem text. The task becomes to decide whether $\zeta(z) = 1 - \lambda f(s)$ will fit entirely inside the unit circle for some choice of $\lambda > 0$. Had the image in the problem text depicted a perfect circle centered at $+2$and with radius 2, $\zeta(s)$ would have fitted inside the unit circle for $\lambda \leq 1/2$. However, the image shows the curve for $f(s)$ to be a lot 'flatter' than a circle in the vicinity of $s = 0$. To see if this flatness causes a failure of ζ to fall inside the unit circle, we Taylor expand $f(s) = is + \mathcal{O}(s^3)$ (note that the s^2 term is missing). Hence, for small s, $\zeta = 1 - \lambda is + \mathcal{O}(s^3)$, and

$$|\zeta|^2 = 1 + (\lambda s)^2 + \mathcal{O}(s^3).$$

No matter how small λ is, ζ must for small s be outside the unit circle. From this, we conclude that the proposed scheme is unconditionally unstable.