University of Groningen
Exam Numerical Mathematics 1, June 19, 2017

Use of a simple calculator is allowed. All answers need to be motivated.
In front of each question you find a weight, which gives the number of tenths that can be gained in the final mark. The maximum total score for this exam is 5.4 points.

Exercise 1

(a) Let $n+1$ points $(x_i, y_i),\ i = 0, 1, ..., n$, be given with distinct nodes x_i. A polynomial P is called interpolating if $P(x_i) = y_i,\ i = 0, 1, ..., n$.

(i) 4 Give a complete description of the Lagrange interpolation formula, and explain why this formula provides an interpolating polynomial P of degree $\leq n$.

The Lagrange interpolation formula reads

$$P(x) = \sum_{k=0}^{n} y_k \varphi_k(x),$$

where the functions φ_k are the Lagrange characteristic polynomials defined as

$$\varphi_k(x) = \prod_{\substack{j=0 \\ j\neq k}}^{n} \frac{x - x_j}{x_k - x_j}.$$

One easily verifies that the Lagrange characteristic polynomials φ_k have degree n and satisfy $\varphi_k(x_j) = \delta_{jk}$. It follows that P is a polynomial of degree $\leq n$ satisfying

$$P(x_j) = \sum_{k=0}^{n} y_k \varphi_k(x_j) = \sum_{k=0}^{n} y_k \delta_{jk} = y_j \quad (jk = 0, 1, ..., n).$$

(ii) 2 Show that there cannot exist another interpolating polynomial P of degree $\leq n$.

If P and Q are two interpolating polynomials of degree $\leq n$, their difference $R = P - Q$ satisfies $R(x_j) = 0$ for all $j = 0, 1, ..., n$. But then R is a polynomial of degree $\leq n$ with at least $n+1$ zeros. This is only possible if R is the zero function, proving that $P = Q$.

(iii) 4 Suppose that all the nodes x_i lie in an interval $I = [a, b]$, and that we are interested in evaluating the interpolant P at arbitrary $x \in I$. How is the corresponding Lebesgue constant Λ defined, and what are the implications if its value is large (say, $\Lambda = 10^5$) ?

The Lebesgue constant Λ is defined as

$$\Lambda = \max_{x\in I} \sum_{k=0}^{n} |\varphi_k(x)|.$$

For a given set of nodes and interval I, it is the smallest possible constant in the stability bound that estimates the effect of perturbations of the values y_i on the interpolated value $P(x)$,

$$|\tilde{P}(x) - P(x)| \leq \Lambda \cdot \max_i |\tilde{y}_i - y_i| \quad \text{for all } x \in I.$$

It can therefore be regarded as the condition number of the interpolation problem. If the value of Λ is large, it follows that small perturbations in the values y_i can have a large effect on the interpolated value $P(x)$ for some $x \in I$.

(b) For a smooth function f on the interval $[0, 1]$ we approximate its (one-sided) derivative $f'(0)$ by $P'(0)$, where P is the polynomial (of degree ≤ 2) that interpolates f at the nodes $x_0 = 0$, $x_1 = h$ and $x_2 = 2h$.

(i) [1] Show that P is given by

$$P(x) = \frac{f(0)}{2h^2}(x-h)(x-2h) - \frac{f(h)}{h^2}x(x-2h) + \frac{f(2h)}{2h^2}x(x-h).$$

This follows immediately from the Lagrange interpolation formula since for the given nodes the Lagrange characteristic polynomials are

$$\begin{aligned}
\varphi_0(x) &= \frac{x-x_1}{x_0-x_1}\cdot\frac{x-x_2}{x_0-x_2} = \frac{x-h}{0-h}\cdot\frac{x-2h}{0-2h} = \frac{(x-h)(x-2h)}{2h^2} \\
\varphi_1(x) &= \frac{x-x_0}{x_1-x_0}\cdot\frac{x-x_2}{x_1-x_2} = \frac{x-0}{h-0}\cdot\frac{x-2h}{h-2h} = \frac{x(x-2h)}{-h^2} \\
\varphi_2(x) &= \frac{x-x_0}{x_2-x_0}\cdot\frac{x-x_1}{x_2-x_1} = \frac{x-0}{2h-0}\cdot\frac{x-h}{2h-h} = \frac{x(x-h)}{2h^2}
\end{aligned}$$

(ii) [1] Use the above explicit expression for $P(x)$ to show that

$$P'(0) = \frac{1}{2h}\left[-3f(0) + 4f(h) - f(2h)\right].$$

We have

$$\varphi_0'(x) = \frac{2x-3h}{2h^2}, \quad \varphi_1'(x) = \frac{2x-2h}{-h^2}, \quad \varphi_2'(x) = \frac{2x-h}{2h^2}$$

implying

$$\varphi_0'(0) = \frac{-3}{2h}, \quad \varphi_1'(0) = \frac{2}{h}, \quad \varphi_2'(0) = \frac{-1}{2h}.$$

The expression for $P'(0)$ immediately follows from

$$P'(0) = f(0)\varphi_0'(0) + f(h)\varphi_1'(0) + f(2h)\varphi_2'(0).$$

(iii) [3] Show that $P'(0)$ is a second order approximation of $f'(0)$ (with respect to h).

Taylor expansion of $4f(h)$ and $f(2h)$ yields

$$\begin{aligned}
4f(h) &= 4f(0) + 4hf'(0) + 2h^2 f''(0) + \mathcal{O}(h^3) \\
f(2h) &= f(0) + 2hf'(0) + 2h^2 f''(0) + \mathcal{O}(h^3)
\end{aligned}$$

Substitution of these Taylor expansions into the expression for $P'(0)$ shows that

$$P'(0) = \frac{2hf'(0) + \mathcal{O}(h^3)}{2h} = f'(0) + \mathcal{O}(h^2).$$

Exercise 2

(a) Consider a system of nonlinear equations $f(x) = 0$, where $f : \mathbb{R}^n \to \mathbb{R}^n$ is smooth.

(i) [4] Derive Newton's method for the above system, and explain briefly how this method works.

Newton's method is based on a linearization (first order Taylor expansion) of f about the last iterate $x^{(k)}$,

$$f(x) \approx f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}),$$

where $f'(x^{(k)})$ is the Jacobian matrix of f with entries $\frac{\partial f_i}{\partial x_j}$ evaluated at $x = x^{(k)}$. The next iterate $x^{(k+1)}$ is defined as the vector x for which the linearization is zero, i.e.,

$$f(x^{(k)}) + f'(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0.$$

In each iteration step, the method therefore requires solving the system of linear equations

$$f'(x^{(k)})\delta^{(k)} = -f(x^{(k)}),$$

followed by setting $x^{(k+1)} = x^{(k)} + \delta^{(k)}$.

(ii) [3] Consider the above system with $n = 2$ and

$$f_1(x_1, x_2) = x_1 + x_2^2 + \sin(x_1 x_2) - 3, \quad f_2(x_1, x_2) = x_1 + x_2 + \cos(x_1 x_2) - 4.$$

Starting from the initial guess $x^{(0)} = (\pi, 0)^T$, show that Newton's method converges to the root $\alpha = (3, 0)^T$ in a single step.

The Jacobian matrix is given by

$$f'(x) = \begin{bmatrix} 1 + x_2 \cos(x_1 x_2) & 2x_2 + x_1 \cos(x_1 x_2) \\ 1 - x_2 \sin(x_1 x_2) & 1 - x_1 \sin(x_1 x_2) \end{bmatrix}$$

For the first Newton step, starting from $x^{(0)} = (\pi, 0)^T$, the linear system $f'(x^{(0)})\delta^{(0)} = -f(x^{(0)})$ has the form

$$\begin{bmatrix} 1 & \pi \\ 1 & 1 \end{bmatrix} \cdot \delta^{(0)} = \begin{bmatrix} -(\pi - 3) \\ -(\pi - 3) \end{bmatrix},$$

which has the solution

$$\delta^{(0)} = \begin{bmatrix} 3 - \pi \\ 0 \end{bmatrix},$$

so that

$$x^{(1)} = x^{(0)} + \delta^{(0)} = \begin{bmatrix} \pi \\ 0 \end{bmatrix} + \begin{bmatrix} 3 - \pi \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

Note that $\alpha = (3, 0)^T$ is indeed a root of $f(x) = 0$.

(b) Consider the fixed point iteration $x^{(k+1)} = \phi(x^{(k)})$ with $x^{(0)}$ given and $\phi(x) = \frac{1}{3}x(4+x-2x^2)$.

(i) [1] Determine all fixed points α of ϕ.

We have

$$\begin{aligned}
&\phi(x) = x \iff \\
&\frac{1}{3}x(4+x-2x^2) = x \iff \\
&x = 0 \ \vee \ 4+x-2x^2 = 3 \iff \\
&x = 0 \ \vee \ 1+x-2x^2 = 0 \iff \\
&x = 0 \ \vee \ (1-x)(1+2x) = 0 \iff \\
&x = 0 \ \vee \ x = 1 \ \vee \ x = -1/2.
\end{aligned}$$

The fixed points are therefore $\alpha = 0$, $\alpha = 1$ and $\alpha = -1/2$.

(ii) [4] For each of these fixed points α, check whether $\{x^{(k)}\}$ converges to α if $x^{(0)}$ is chosen sufficiently close to α. If that occurs, also determine the order of convergence.

Note that

$$\phi'(x) = \frac{1}{3}(4+x-2x^2) + \frac{1}{3}x(1-4x) = \frac{1}{3}(4+2x-6x^2).$$

We conclude that

- $\alpha = 0$: no convergence since $|\phi'(0)| = \frac{4}{3} > 1$
- $\alpha = 1$: convergence of order at least 2 since $\phi'(1) = 0$; the order is exactly 2 since $\phi''(1) = -10/3 \neq 0$
- $\alpha = -1/2$: convergence of order 1 since $\phi'(-1/2) = 1/2 \in (0,1)$

Exercise 3

(a) Consider the system of linear equations $Ax = b$, where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 9 \\ 10 \end{bmatrix}.$$

(i) $\boxed{4}$ Determine the Cholesky factorization and LU factorization of A.

The Cholesky factorization $A = R^T R$ and LU factorization $A = LU$ are the same in this case, namely

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

(ii) $\boxed{3}$ Use one of these factorizations to solve $Ax = b$.

We solve $Ax = b$ by first solving $Ly = b$ for y and then solving $Rx = y$ for x.
Solving $Ly = b$ with the forward substitution method we find

$$y = \begin{bmatrix} 6 \\ 3 \\ 1 \end{bmatrix}.$$

Subsequently solving $Rx = y$ with the backward substitution method we find

$$x = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

(b) For solving a general linear system $Ax = b$ we consider iterative methods of the form

$$Px^{(k+1)} = (P - A)x^{(k)} + b,$$

where P is a nonsingular preconditioner of A.

(i) [1] Determine the iteration matrix B and show that the error $e^{(k)} = x^{(k)} - x$ satisfies $e^{(k+1)} = Be^{(k)}$. When does the method converge?

The iteration method can be rewritten as

$$x^{(k+1)} = P^{-1}(P - A)x^{(k)} + P^{-1}b = Bx^{(k)} + P^{-1}b, \tag{1}$$

where the iteration matrix is given by

$$B = P^{-1}(P - A) = I - P^{-1}A.$$

Note that the solution x of $Ax = b$ obviously satisfies $Px = (P - A)x + b$, so we also have

$$x = Bx + P^{-1}b.$$

Subtracting the latter relation from (1) we see that

$$e^{(k+1)} = Be^{(k)}.$$

The method converges if the spectral radius $\rho(B)$ satisfies $\rho(B) < 1$.

(ii) [5] What is the name of the iterative method that corresponds to the preconditioner $P = D = \text{diag}(a_{11}, a_{22}, ..., a_{nn})$? Show that this method converges if A is strictly diagonally dominant by row.

This is Jacobi's method. To show that it converges under the mentioned condition we write $A = D + N$, where N is the non-diagonal part of A. The iteration matrix is given by

$$B = -D^{-1}N = \begin{bmatrix} 0 & -a_{12}/a_{11} & \dots & -a_{1n}/a_{11} \\ -a_{21}/a_{22} & 0 & & -a_{2n}/a_{22} \\ \vdots & & \ddots & \vdots \\ -a_{n1}/a_{nn} & -a_{n2}/a_{nn} & \dots & 0 \end{bmatrix}.$$

In the following we will show that $\rho(B) < 1$. It follows from the strict diagonal dominance of A that the matrix B satisfies

$$\sum_{j=1}^{n} |b_{ij}| < 1 \quad \text{(for all } i = 1, 2, ..., n).$$

Let λ be an arbitrary eigenvalue of B and v a corresponding eigenvector. Then we have

$$\sum_{j=1}^{n} b_{ij}v_j = \lambda v_i \quad \text{(for all } i = 1, 2, ..., n).$$

We scale this eigenvector such that $\max_j |v_j| = 1$. Hence there exists at least one index i with $|v_i| = 1$. For this index i we have

$$|\lambda| = |\lambda v_i| = |\sum_{j=1}^{n} b_{ij}v_j| \leq \sum_{j=1}^{n} |b_{ij}| < 1.$$

Since λ was an arbitrary eigenvalue of B, we have shown that $\rho(B) < 1$. For those familiar with the maximum norm, we note that the proof can be shortened to

$$\rho(B) \leq \|B\|_\infty = \max_i \sum_{j=1}^{n} |b_{ij}| < 1.$$

Exercise 4

(a) For the numerical solution of the initial value problem

$$y'(t) = f(t, y(t)), \quad y(t_0) = y_0$$

we use the so-called *implicit midpoint rule*, which is defined as

$$u_{n+1} = u_n + hf(\frac{1}{2}t_n + \frac{1}{2}t_{n+1}, \frac{1}{2}u_n + \frac{1}{2}u_{n+1}),$$

where $u_0 = y_0$ and $t_n = t_0 + nh$.

(i) 3 Show that application of this method to the test problem $y'(t) = \lambda(t)y(t)$ leads to the recurrence relation

$$u_{n+1} = \frac{1 + \frac{1}{2}h\lambda(\frac{1}{2}t_n + \frac{1}{2}t_{n+1})}{1 - \frac{1}{2}h\lambda(\frac{1}{2}t_n + \frac{1}{2}t_{n+1})}u_n.$$

For the function $f(t, y) = \lambda(t)y$ corresponding to the test problem, the implicit midpoint rule reads

$$u_{n+1} = u_n + h\lambda(\frac{1}{2}t_n + \frac{1}{2}t_{n+1})(\frac{1}{2}u_n + \frac{1}{2}u_{n+1}).$$

Collecting terms involving u_{n+1} on the left we obtain

$$\left[1 - \frac{1}{2}h\lambda(\frac{1}{2}t_n + \frac{1}{2}t_{n+1})\right] u_{n+1} = \left[1 + \frac{1}{2}h\lambda(\frac{1}{2}t_n + \frac{1}{2}t_{n+1})\right] u_n,$$

from which the above recurrence relation immediately follows.

(ii) 4 Give the definition of 'A-stability' (unconditional absolute stability) and verify whether the implicit midpoint rule is A-stable.

A method is 'A-stable' if its approximations satisfy

$$\lim_{n\to\infty} u_n = 0$$

whenever it is applied (with arbitrary step size $h > 0$) to the test equation

$$y'(t) = \lambda y(t) \quad (t \geq 0),$$

where λ is a complex number with negative real part.
For the implicit midpoint rule, the latter test problem is a special case of the more general test problem $y'(t) = \lambda(t)y(t)$, and it follows from part (i) that its approximations satisfy the recurrence relation

$$u_{n+1} = \frac{1 + \frac{1}{2}h\lambda}{1 - \frac{1}{2}h\lambda}u_n.$$

We see that the implicit midpoint rule is A-stable iff

$$\left|\frac{1 + \frac{1}{2}z}{1 - \frac{1}{2}z}\right| < 1 \quad \text{(for all } z \in \mathbb{C} \text{ with negative real part)},$$

which is equivalent to

$$|z + 2| < |z - 2| \quad \text{(for all } z \in \mathbb{C} \text{ with negative real part)}.$$

The latter condition is indeed fulfilled since for complex numbers in the (open) left half plane, the distance to the number -2 is smaller than to the number 2.

(b) Consider the Poisson equation on the (open) unit square $\Omega = (0,1) \times (0,1)$,

$$-\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = f(x,y), \tag{1}$$

where $u(x,y) = g(x,y)$ is given on the boundary of Ω (Dirichlet boundary conditions).

(i) $\boxed{2}$ First show that for any smooth function $v : [0,1] \to \mathbb{R}$ and $x \in (0,1)$ the quantity

$$\frac{v(x+h) - 2v(x) + v(x-h)}{h^2} \tag{2}$$

provides an approximation to $v''(x)$ of order 2 with respect to h.

Taylor expansion of $v(x+h)$ and $v(x-h)$ yields

$$\begin{aligned} v(x+h) &= v(x) + hv'(x) + \frac{1}{2}h^2 v''(x) + \frac{1}{6}h^3 v'''(x) + \mathcal{O}(h^4) \\ v(x-h) &= v(x) - hv'(x) + \frac{1}{2}h^2 v''(x) - \frac{1}{6}h^3 v'''(x) + \mathcal{O}(h^4) \end{aligned}$$

Substitution of these Taylor expansions into the above difference quotient gives

$$\frac{h^2 v''(x) + \mathcal{O}(h^4)}{h^2} = v''(x) + \mathcal{O}(h^2).$$

(ii) $\boxed{5}$ We choose an integer $N \geq 1$, set $h = 1/(N+1)$ and define grid nodes $(x_i, y_j) = (ih, jh)$, $i,j = 0,1,...,N+1$. We construct approximations $u_{i,j}$ to $u(x_i, y_j)$ by requiring that differential equation (1) is satisfied at all internal grid nodes while replacing both second derivatives by the second order difference quotient of type (2). This leads to a linear system $A\tilde{u} = b$, where the vector $\tilde{u}$ consists of all values $u_{i,j}$ at the internal nodes. Find the matrix A and right-hand-side vector b in case $N = 3$.

For each internal grid node (x_i, y_j) the discretized differential equation reads

$$\frac{1}{h^2}\left[-u_{i-1,j} + 2u_{i,j} - u_{i+1,j}\right] + \frac{1}{h^2}\left[-u_{i,j-1} + 2u_{i,j} - u_{i,j+1}\right] = f_{i,j},$$

where $f_{i,j} = f(x_i, y_j)$. We can rewrite this into

$$\frac{1}{h^2}\left[-u_{i-1,j} - u_{i,j-1} + 4u_{i,j} - u_{i,j+1} - u_{i+1,j}\right] = f_{i,j}.$$

Taking the boundary conditions into account this leads to the following linear system if $N = 3$ (and therefore $h = 1/4$),

$$\frac{1}{h^2}\left[\begin{array}{ccc|ccc|ccc} 4 & -1 & & -1 & & & & & \\ -1 & 4 & -1 & & -1 & & & & \\ & -1 & 4 & & & -1 & & & \\ \hline -1 & & & 4 & -1 & & -1 & & \\ & -1 & & -1 & 4 & -1 & & -1 & \\ & & -1 & & -1 & 4 & & & -1 \\ \hline & & & -1 & & & 4 & -1 & \\ & & & & -1 & & -1 & 4 & -1 \\ & & & & & -1 & & -1 & 4 \end{array}\right] \left[\begin{array}{c} u_{1,1} \\ u_{1,2} \\ u_{1,3} \\ \hline u_{2,1} \\ u_{2,2} \\ u_{2,3} \\ \hline u_{3,1} \\ u_{3,2} \\ u_{3,3} \end{array}\right] = \left[\begin{array}{l} f_{1,1} + h^{-2}(g_{0,1} + g_{1,0}) \\ f_{1,2} + h^{-2} g_{0,2} \\ f_{1,3} + h^{-2}(g_{0,3} + g_{1,4}) \\ \hline f_{2,1} + h^{-2} g_{2,0} \\ f_{2,2} \\ f_{2,3} + h^{-2} g_{2,4} \\ \hline f_{3,1} + h^{-2}(g_{3,0} + g_{4,1}) \\ f_{3,2} + h^{-2} g_{4,2} \\ f_{3,3} + h^{-2}(g_{4,3} + g_{3,4}) \end{array}\right].$$

Note that the same matrix A is obtained if we make the following alternative logical choice for the solution vector $\tilde{u}$ and right-hand-side vector b,

$$\begin{aligned} \tilde{u} &= (u_{1,1}, u_{2,1}, u_{3,1}, u_{1,2}, u_{2,2}, u_{3,2}, u_{1,3}, u_{2,3}, u_{3,3})^T, \\ b &= (f_{1,1} + ..., f_{2,1} + ..., f_{3,1} + ..., f_{1,2} + ..., f_{2,2}, f_{3,2} + ..., f_{1,3} + ..., f_{2,3} + ..., f_{3,3} + ...)^T. \end{aligned}$$