DEGREE OF MASTER OF SCIENCE

MATHEMATICAL MODELLING AND SCIENTIFIC COMPUTING

B1 Numerical Linear Algebra and Numerical Solution of Differential Equations

---

HILARY TERM 2017

FRIDAY, 13 JANUARY 2017, 9.30am to 11.30am

Candidates should submit answers to a maximum of four questions that include an answer to at least one question in each section.

Please start the answer to each question in a new booklet.
All questions will carry equal marks.

Do not turn this page until you are told that you may do so

Section A: Numerical Solution of Differential Equations

1. Suppose that for $t > 0$, the function $u(t)$ satisfies the differential equation

$$u'(t) = f(t, u(t)), \qquad u(0) = 0. \tag{1}$$

Suppose further that f is Lipschitz continuous.

(a) [2 marks] Consider the general Runge-Kutta method of order R, as given by the Butcher table:

$$\begin{array}{c|ccccc} a_1 & 0 & & & & \\ a_2 & b_{2,1} & 0 & & & \\ \vdots & \vdots & \ddots & & & \\ a_R & b_{R,1} & \cdots & b_{R,R-1} & 0 \\ \hline & c_1 & \cdots & c_{R-1} & c_R \end{array} =: \begin{array}{c|c} \mathbf{a} & B \\ \hline & \mathbf{c}^T \end{array}$$

Write this as a one-step method of the form

$$U_{n+1} = U_n + h\Phi(U_n, t_n; h),$$

where you should explicitly identify the function $\Phi(U_n, t_n; h)$.

(b) [13 marks] If $T = \max_n |T_n|$, where T_n is the truncation error at step n, and L_Φ is the Lipshitz constant for the function Φ, show that the global error of such a method satisfies

$$|e_n| \leqslant \frac{T}{L_\Phi}\left(\mathrm{e}^{L_\Phi t_n} - 1\right).$$

[*You may use without proof the facts that the associated function $\Phi(u, t; h)$ is Lipschitz continuous in the first argument provided that f is Lipschitz continuous, and that* $1 + x \leqslant \mathrm{e}^x$.]

(c) [5 marks] Consider (1) with

$$f(t, u) = u + t^{l-1}, \tag{2}$$

where $l > 0$ is an integer. Let k_i, $i = 1, \ldots, R$, denote the values of the function evaluations in the Runge-Kutta method. Show that, for the first iterate of the Runge-Kutta method above, the k_i can be found by solving the matrix equation

$$(I - hB)\mathbf{k} = h^{l-1}A^{l-1}\mathbf{1},$$

where $A = \operatorname{diag}\{a_1, \ldots, a_R\}$, $\mathbf{k} = [k_1, \ldots, k_R]^T$, B is as in the Butcher table in part (a), and $\mathbf{1} = [1, \ldots, 1]^T$.

(d) [5 marks] Using the result of the previous part show that, for h sufficiently small,

$$U_1 = h^l\mathbf{c}^T(I + hB + \cdots + h^kB^k + \cdots)A^{l-1}\mathbf{1}.$$

Use the fact that the analytic solution of (1) with the function (2) is $u(t) = \int_0^t \mathrm{e}^{t-s}s^{l-1}\,\mathrm{d}s$ to show that a necessary condition for the method to be of order p is that

$$\mathbf{c}^TB^kA^{l-1}\mathbf{1} = \frac{(l-1)!}{(l+k)!},$$

for a range of k which you should identify.

[*You may use without proof the fact that* $u^{(l-j)}(0) = 0$, $1 \leqslant j \leqslant l$, *and* $u^{(l+j)}(0) = (l-1)!$ *for* $j \geqslant 0$.]

2. Let $t_n = t_0 + nh$ for some $h > 0$. For $t \in [t_0, t_m]$, the function $u(t)$ satisfies the ordinary differential equation

$$u'(t) = f(t, u(t)), \qquad u(t_0) = u_0.$$

Recall that a linear multi-step method, which takes the form

$$\sum_{j=0}^{k} \alpha_j U_{n+j} = h \sum_{j=0}^{k} \beta_j f(t_{n+j}, U_{n+j}), \tag{3}$$

computes values $U_n \approx u(t_n)$.

(a) [4 marks] Define the *truncation error* of a linear multi-step method.

What does it mean for the method (3) to be *consistent*?

(b) [6 marks] Define $\rho(z)$ and $\sigma(z)$, the *first* and *second characteristic polynomials* of (3) respectively.

What does it mean for the method (3) to be *zero-stable*?

Show that the multi-step method

$$U_{n+k} = U_{n+k-2} + \sum_{j=0}^{k} \beta_j f(t_{n+j}, U_{n+j}) \tag{4}$$

($k \geqslant 2$) is zero-stable for all values of β_j, stating carefully any theorem that you use.

(c) [7 marks] Consider again the general method, (3). Show that, if u is sufficiently smooth, the truncation error T_n for (3) can be expressed as

$$T_n = \frac{1}{\sigma(1)h} \left[C_0 u(t_n) + C_1 u'(t_n) h + \cdots + C_p u^{(p)}(t_n) h^p + \cdots \right],$$

where $C_0 = \sum_{j=0}^{k} \alpha_j$ and $C_q = \sum_{j=0}^{k} \frac{j^q}{q!} \alpha_j - \sum_{j=0}^{k} \frac{j^{q-1}}{(q-1)!} \beta_j$ for $q \geqslant 1$.

(d) [8 marks] Show that in order to have a consistent method, we must have that

$$\rho(1) = 0 \text{ and } \rho'(1) = \sigma(1).$$

Show further that the scheme (3) is of order $p \geqslant 1$ if and only if there exists a non-zero constant K such that

$$\rho(z) - \sigma(z) \log z = K(z-1)^{p+1} + \mathcal{O}(|z-1|^{p+2}).$$

3. (a) [6 marks] Let $\lambda \in \mathbb{C}$ be a constant with $Re(\lambda) < 0$. By considering directly the test problem

$$y' = \lambda y \tag{5}$$

where y is a function of t defined for $t > t_0$, derive regions of absolute stability for

(i) Explicit Euler's method, and

(ii) Implicit Euler's method.

(b) [6 marks] Let $K \in \mathbb{R}^{n \times n}$ be a symmetric matrix with entries that are constant with respect to time. Show that the system of differential equations

$$\mathbf{y}' = K\mathbf{y} \tag{6}$$

can be written as a series of decoupled scalar differential equations of the form

$$z_i' = \lambda_i z_i,$$

for some functions $z_i(t)$ and scalars λ_i, which you should identify.

Using this result, derive a condition for absolute stability of both the explicit and implicit versions of Euler's method when applied to (6).

(c) [5 marks] Let $u(x,t)$ be a function defined for $x \in [-1,1]$, $t \geqslant 0$, which satisfies the partial differential equation

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

with boundary conditions $u(-1,t) = a$, $u(1,t) = b$, and initial condition $u(x,0) = u_0(x)$. Let N be a positive integer, and consider a uniform mesh $x_k = kh$, where $h = 1/N$ and $k = 0, \pm 1, \cdots, \pm N$. Let $U(t)$ be the vector where $U_i(t) \approx u(x_i, t)$ is the approximation obtained by using central differences in the spatial dimension only.

Show that $U(t)$ satisfies the system of differential equations

$$U' = AU + \mathbf{f} \tag{7}$$

where the matrix $A \in \mathbb{R}^{(2N-1)\times(2N-1)}$ and vector $\mathbf{f} \in \mathbb{R}^{2N-1}$ should be identified.

(d) [8 marks] Show that the eigenvectors of the matrix A take the form $\mathbf{w}^p$, the jth component of which is given by $\mathbf{w}_j^p = \sin(jp\pi h)$, and determine the corresponding eigenvalues.

Show that an implicit Euler scheme applied to (7) is unconditionally stable. Derive a condition on the size of the time step required for an explicit Euler scheme applied to (7) to be absolutely stable.

[*You may use without proof the fact that* $\sin(A \pm B) = \sin(A)\cos(B) \pm \cos(A)\sin(B)$]

4. Let $u(x,t)$ be a function that satisfies the partial differential equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - a(x)\frac{\partial u}{\partial x}, \tag{8}$$

with boundary conditions $u(-1,t) = u(1,t) = 0$, initial condition $u(x,0) = u_0(x)$, where u_0 and a are continuous real-valued functions on $[-1,1]$.

Let N be a given integer, and consider a uniform mesh $x_j = j\Delta x$, where $\Delta x = 1/N$ and $j = 0, \pm 1, \ldots, \pm N$, and $t_n = n\Delta t$, $n = 0, 1, 2, \ldots$. Let $U_j^n \approx u(x_j, t_n)$ be an approximation obtained via the scheme

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{(\Delta x)^2} - a_j \frac{U_{j+1}^n - U_{j-1}^n}{2\Delta x},$$

where $a_j = a(x_j)$.

(a) [8 marks] Consider the case where $a(x) = 0$. By using a semi-discrete Fourier transform, show that this scheme is stable provided that $\Delta t/(\Delta x)^2 \leqslant 1/2$.

[You may use Parseval's identity:

$$\|U^n\|_{\ell_2}^2 = \frac{1}{2\pi}\|\hat{U}^n\|_{L_2}^2.$$

(without proof).]

(b) [9 marks] Now consider the case where $a(x)$ is a continuous function. Define the truncation error of the scheme. If $\partial^2 u/\partial t^2, \partial^3 u/\partial x^3$ and $\partial^4 u/\partial x^4$ exist and are bounded in the domain, show that $|T_n| \leqslant C(\Delta t + \Delta x^2)$, where C is a constant independent of Δt and Δx.

(c) [8 marks] Show that the global error vanishes as $\Delta x \to 0$, $\Delta t \to 0$ provided that $\Delta t \leqslant \frac{1}{2}\Delta x^2$, and $\Delta x \leqslant \lambda$, where λ is a constant you should identify.

Section B: Numerical Linear Algebra

5. (a) [10 marks] What is an orthogonal matrix? Show that for any vector, x, $\|Qx\|_2 = \|x\|_2$ when Q is an orthogonal matrix. Prove also that $\|QAP\|_2 = \|A\|_2$ when Q and P are orthogonal.

   Explain what a QR factorisation is and, briefly, how it might be computed for an $n \times n$ upper Hessenberg matrix with just $n-1$ Givens rotations.

   (b) [4 marks] Explain how a QR factorisation can be employed in the solution of a linear least squares problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$$

   where $A \in \mathbb{R}^{m \times n}$ has rank n and $b \in \mathbb{R}^m$ with $m > n$.

   (c) [5 marks] Define the Singular Value Decomposition (SVD) of a matrix $A \in \mathbb{R}^{m \times n}$. [*You do not need to prove that the SVD exists, but you may assume that it does for the rest of this question.*] How might an SVD be employed in the solution of a linear least squares problem?

   (d) [6 marks] Prove that for any $A \in \mathbb{R}^{n \times n}$, there exists an orthogonal matrix, Q, and a real symmetric and positive semi-definite matrix, H, such that $A = QH$. Further prove that if P is any orthogonal matrix and $\|A\|_2 \geqslant 2$, then

$$\|A - P\|_2 \geqslant \|A - Q\|_2$$

   .

6. (a) [6 marks] What is a simple iteration for the solution of a (square) linear system of equations, $Ax = b$ with starting vector x_0? What is the associated iteration matrix? If the iteration matrix is diagonalisable, state and prove a necessary and sufficient condition for the simple iteration to generate a sequence of vectors which converge to x. [*For the remainder of this question, you may assume that the assumption of diagonalisability is not necessary for this result to hold.*]

(b) [10 marks] What is (i) the Jacobi iteration, (ii) the relaxed Jacobi iteration with relaxation parameter $\theta \in \mathbb{R}, 0 < \theta \leqslant 1$?

For any square matrix, B, let

$$\rho = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } B\}.$$

Show that for the relaxed Jacobi iteration matrix, the value of ρ cannot be less than the value of ρ for the Jacobi iteration matrix when Jacobi iteration converges. Are there any circumstances in which the relaxed Jacobi iteration converges, but the Jacobi iteration does not?

State briefly in words a context in which the relaxed Jacobi iteration might be preferable to the Jacobi iteration.

(c) [9 marks] Let

$$A = \begin{bmatrix} 2 & -2 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \ddots & 2 & -1 \\ 0 & 0 & \cdots & 0 & -2 & 2 \end{bmatrix} \in \mathbb{R}^{(n+2)\times(n+2)}.$$

Show that

$$v_r = \begin{bmatrix} \cos\frac{0r\pi}{n+1} \\ \cos\frac{1r\pi}{n+1} \\ \cos\frac{2r\pi}{n+1} \\ \vdots \\ \cos\frac{(n+1)r\pi}{n+1} \end{bmatrix}$$

is a right eigenvector of A for $r = 0, 1, 2, \ldots, n+1$. Further show that the relaxed Jacobi iteration matrix for A has eigenvalues all lying in the real interval $[1 - 2\theta, 1]$. Deduce that for $0 < \theta < 1$, the relaxed Jacobi iteration for A will generate a sequence of vectors which will converge to a solution, x, whenever $x - x_0$ has no component in the direction of v_0.