AMATH 586 Assignment 2

Tyler Chen

Preliminaries

Recall the general expression for a r-step LMM,

$$\sum_{j=0}^{r} \alpha_j U^{n+j} = k \sum_{j=0}^{r} \beta_j f(U^{n+j}, t_{n+j})$$

The local truncation error is,

$$\tau_{n+2} = \frac{1}{k} \left(\sum_{j=0}^{r} \alpha_j \right) u(t_n) + \sum_{q=1}^{\infty} \left(k^{q-1} \left(\sum_{j=0}^{2} \left(\frac{1}{q!} j^q \alpha_j - \frac{1}{(q-1)!} j^{q-1} \beta_j \right) \right) u^{(q)}(t_n) \right)$$

Note that for any integer $q > 0$,

$$k^{q-1} \left(\sum_{j=0}^{r} \left(\frac{1}{q!} j^q \alpha_j - \frac{1}{(q-1)!} j^{q-1} \beta_j \right) \right) u^{(q)}(t_n) = 0 \qquad \iff \qquad \sum_{j=0}^{r} j^q \alpha_j = q \sum_{j=0}^{r} j^{q-1} \beta_j$$

Problem 1

Determine the coefficients β_0, β_1, β_2 for the third order, 2-step Adams-Moulton method:

$$U^{n+2} = U^{n+1} + k[\beta_0 f(U^n, t_n) + \beta_1 f(U^{n+1}, t_{n+1}) + \beta_2 f(U^{n+2}, t_{n+2})].$$

Do this in two different ways:

(a) Using the expression for the local truncation error in Section 5.9.1.

(b) Using the relation

$$u(t_{n+2}) = u(t_{n+1}) + \int_{t_{n+1}}^{t_{n+2}} f(u(s), s)\, ds,$$

and replacing f in the integral by a quadratic polynomial $p(s)$ that takes the values $f(U^n, t_n)$, $f(U^{n+1}, t_{n+1})$, and $f(U^{n+2}, t_{n+2})$ at the points t_n, t_{n+1}, and t_{n+2}.

Solution

(a) This is a 2-step LMM with $\alpha_0 = 0$, $\alpha_1 = -1$, and $\alpha_2 = 1$.

Clearly $\sum_{j=0}^{2} \alpha_j = 0$. We have three unknowns, so we hope to satisfy at least 3 of the equations. The first three equations are,

$$\sum_{j=0}^{2} j\alpha_j = \sum_{j=0}^{2} \beta_j, \qquad \sum_{j=0}^{2} j^2\alpha_j = 2\sum_{j=0}^{2} j\beta_j, \qquad \sum_{j=0}^{2} j^3\alpha_j = 3\sum_{j=0}^{2} j^2\beta_j$$

This gives the linear system,

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2\cdot 1^1 & 2\cdot 2^1 \\ 0 & 3\cdot 1^2 & 3\cdot 2^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1^1\alpha_1 + 2^1\alpha_2 \\ 1^2\alpha_1 + 2^2\alpha_2 \\ 1^3\alpha_1 + 2^3\alpha_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}$$

This has solution,

$$\beta_0 = -1/12, \qquad \beta_1 = 2/3, \qquad \beta_2 = 5/12$$

(b) We approximate $f(u(s), s)$ with a polynomial $F(x)$ passing through the points $f_n := (f(u(t_n), t_n)$, $f_{n+1} := (f(u(t_{n+1}), t_{n+1})$, and $f_{n+2} := (f(u(t_{n+2}), t_{n+2})$. In particular, this is the Lagrange interpolating polynomial, with equation,

$$P(s) = f_n \frac{(s - t_{n+1})(s - t_{n+2})}{(t_n - t_{n+1})(t_n - t_{n+2})} + f_{n+1} \frac{(s - t_n)(s - t_{n+2})}{(t_{n+1} - t_n)(t_{n+1} - t_{n+2})} + f_{n+2} \frac{(s - t_n)(s - t_{n+1})}{(t_{n+2} - t_n)(t_{n+2} - t_{n+1})}$$

With the assumption that $k = t_{n+2} - t_{n+1} = t_{n+1} - t_n$ we easily compute (using Mathematica),

$$\int_{t_{n+1}}^{t_{n+2}} P(s)\mathrm{d}s = \frac{k}{12}\left(-f(U^n, t_n) + 8f(U^{n+1}, t_{n+1}) + 5f(U^{n+2}, t_{n+2})\right)$$

Using this approximation of the integral to construct a 2-step LMM gives the coefficients,

$$\beta_0 = -1/12, \qquad \beta_1 = 2/3, \qquad \beta_2 = 5/12$$

Problem 2

What is the order of the local truncation error for each of the following linear multistep methods, and which of these methods are *convergent*? Justify your answers.

(a) $U^n - U^{n-2} = k[f(U^n, t_n) - 3f(U^{n-1}, t_{n-1}) + 4f(U^{n-2}, t_{n-2})]$.

(b) $U^n - 2U^{n-1} + U^{n-2} = k[f(U^n, t_n) - f(U^{n-1}, t_{n-1})]$.

(c) $U^n - U^{n-1} - U^{n-2} = k[f(U^n, t_n) - f(U^{n-1}, t_{n-1})]$.

Solution

We expand the first terms of the local truncation error for a 2-step LMM ,

$$\alpha_0 + \alpha_1 + \alpha_2 = \sum_{j=0}^{2} \alpha_j = 0 \tag{c$_1$}$$

$$\alpha_1 + 2\alpha_2 = \sum_{j=0}^{2} j\alpha_j = \sum_{j=0}^{2} \beta_j = \beta_0 + \beta_1 + \beta_2 \tag{c$_2$}$$

$$\alpha_1 + 4\alpha_2 = \sum_{j=0}^{2} j^2\alpha_j = 2\sum_{j=0}^{2} j\beta_j = 2\beta_1 + 4\beta_2 \tag{1}$$

$$\alpha_1 + 8\alpha_2 = \sum_{j=0}^{2} j^3\alpha_j = 3\sum_{j=0}^{2} j\beta_j = 3\beta_1 + 12\beta_2 = 3\sum_{j=0}^{2} j^2\beta_j \tag{2}$$

We explicitly write the first terms of the local truncation error. If (c$_1$) and (c$_2$) hold the method is $\mathcal{O}(k)$. If (1) holds the method is $\mathcal{O}(k^2)$, and if (2) holds the method is $\mathcal{O}(k^3)$

(a) We write this method as,

$$U^{n+2} - U^n = k[f(U^{n+2}, t_{n+2}) - 3f(U^{n+1}, t_{n+1}) + 4f(U^n, t_n)]$$

This is a 2-step LMM with coefficients,

$$\alpha_0 = -1, \qquad \alpha_1 = 0, \qquad \alpha_2 = 1, \qquad \beta_0 = 4, \qquad \beta_1 = -3, \qquad \beta_2 = 1$$

We have (c$_1$) and (c$_2$) but not (1). Therefore the local truncation error is $\mathcal{O}(k)$.

The characteristic polynomial of this LMM (in z) is $z^2 - 1$ which has roots $z = \pm 1$. These are distinct and have modulus less than or equal to one so the method is zero-stable. This, along with consistency implies that the method is convergent.

(b) We write this method as,

$$U^{n+2} - 2U^{n+1} + U^n = k[f(U^{n+2}, t_{n+2}) - f(U^{n+1}, t_{n+1})]$$

This is a 2-step LMM with coefficients,

$$\alpha_0 = 1, \qquad \alpha_1 = -2, \qquad \alpha_2 = 1, \qquad \beta_0 = 0, \qquad \beta_1 = -1, \qquad \beta_2 = 1$$

We have (c$_1$), (c$_2$), and (1) but not (2). Therefore the local truncation error is $\mathcal{O}(k^2)$.

The characteristic polynomial of this LMM (in z) is $z^2 - 2z + 1$ which has repeat root $z = 1$. Since these roots are repeated and do not have modulus less than one the method is not zero-stable and therfore not convergent.

(c) We write this method as,

$$U^{n+2} - U^{n+1} - U^n = k[f(U^{n+2}, t_{n+2}) - f(U^{n+1}, t_{n+1})]$$

This is a 2-step LMM with coefficients,

$$\alpha_0 = -1, \qquad \alpha_1 = -2, \qquad \alpha_2 = 1, \qquad \beta_0 = 0, \qquad \beta_1 = -1, \qquad \beta_2 = 1$$

We do not even have (c_1). The method is not consistent (order $\mathcal{O}(1/k)$).

The characteristic polynomial of this LMM (in z) is $z^2 - 2z - 1$ which has roots $z = 1 + \pm\sqrt{2}$. These are distinct, however one of them does not have modulus less than or equal to one, so the method is not zero-stable and therfore not convergent.

Problem 3

(a) Determine the general solution to the linear difference equation: $2U^{n+3} - 5U^{n+2} + 4U^{n+1} - U^n = 0$. [Hint: One root of the characteristic polynomial $\chi(\lambda)$ is $\lambda = 1$.]

(b) Determine the solution to the difference equation with the starting values $U^0 = 11$, $U^1 = 5$, and $U^2 = 1$.

(c) Consider the LMM

$$2U^{n+3} - 5U^{n+2} + 4U^{n+1} - U^n = k[\beta_0 f(U^n, t_n) + \beta_1 f(U^{n+1}, t_{n+1})].$$

For what values of β_0 and β_1 is the local truncation error $O(k^2)$?

(d) Suppose you use the values of β_0 and β_1 just determined in this LMM. Is this a convergent method? Give a reason.

Solution

(a) This is a linear homogeneous difference equation with characteristic polynomial,

$$\chi(\lambda) = 2\lambda^3 - 5\lambda^2 + 4\lambda - 1 = (\lambda - 1)^2(2\lambda - 1)$$

The general solution to the differece equation is then,

$$U^n = c_1 1^n + c_2 n 1^n + c_3(1/2)^n = c_1 + c_2 n + c_3/2^n$$

(b) If $U^0 = 11$, $U^1 = 5$, and $U^2 = 1$ then,

$$11 = U^0 = c_1 + c_3, \qquad 5 = U^1 = c_1 + c_2 + c_3/2, \qquad 1 = U^2 = c_1 + 2c_2 + c_3/4$$

This is a linear system,

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1/2 \\ 1 & 2 & 1/4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 11 \\ 5 \\ 1 \end{bmatrix}$$

This has solution,

$$c_1 = 3, \qquad c_2 = -2, \qquad c_3 = 8$$

We then have general solution to the differece equation,

$$U^n = 3 - 2n + 8/2^n$$

(c) This is a 3-step LMM with $\alpha_0 = -1$, $\alpha_1 = 4$, $\alpha_2 = -5$, $\alpha_3 = 2$, and $\beta_2 = \beta_3 = 0$. For the method to have local trunction error $\mathcal{O}(k^2)$ we need to satisfy,

$$\sum_{j=0}^{3} \alpha_j = 0, \qquad \sum_{j=0}^{3} j\alpha_j = \sum_{j=0}^{3} \beta_j, \qquad \sum_{j=0}^{3} j^2\alpha_j = 2\sum_{j=0}^{3} j\beta_j,$$

Clearly the leftmost equation is satisfied. The right two equations give the linear system,

$$\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1^1\alpha_1 + 2^1\alpha_2 + 3^1\alpha_3 \\ 1^2\alpha_1 + 2^2\alpha_2 + 3^2\alpha_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

This has solution,

$$\beta_0 = -1, \qquad \beta_1 = 1$$

(d) The method is not zero-stable as the characteristic polynomial has repeated roots of modulus one. This implies the method is not convergent.

Problem 4

Show that the characteristic polynomial of the linear multistep method

$$\sum_{\ell=0}^{r} a_\ell U^{n+\ell} = k \sum_{\ell=0}^{r} b_\ell f(U^{n+\ell}, t_{n+\ell}), \quad a_r = 1,$$

namely, $\chi(z) = \sum_{\ell=0}^{r} a_\ell z^\ell$, is the characteristic polynomial $\det(zI - A)$ of the r by r companion matrix

$$A = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & \cdots & -a_{r-2} & -a_{r-1} \end{bmatrix}.$$

[Hint: Expand $\det(zI - A)$ along the first column and use induction on r.]

Solution

Define,

$$X_k = \underbrace{\begin{bmatrix} z & -1 & & \\ & \ddots & \ddots & \\ & & z & -1 \\ & & & z \end{bmatrix}}_{k\times k}, \qquad Y_k = \underbrace{\begin{bmatrix} -1 & & & \\ z & -1 & & \\ & \ddots & \ddots & \\ & & z & -1 \end{bmatrix}}_{k\times k}$$

Both X_k and Y_k article triangular, with determinants z^k and $(-1)^k$ respectively. Thus, for $0 \leq k \leq r-1$, where, for notational convencience we take $Y_0 = X_0 = [\ ]$ and $\det(Y_0) = \det(X_0) = 1$,

$$\det \begin{bmatrix} X_k & \\ & Y_{r-k-1} \end{bmatrix} = \det(X_k)\det(Y_{r-k-1}) = (-1)^{r-k-1} z^k$$

Write,

$$zI - A = \begin{bmatrix} z & -1 & & \\ & \ddots & \ddots & \\ & & z & -1 \\ a_0 & \cdots & a_{r-2} & z + a_{r-1} \end{bmatrix}$$

Expanding across the bottom row,

$$\det(zI - A) = \sum_{k=0}^{r} (-1)^{r+k+1} a_k \det \begin{bmatrix} X_k & \\ & Y_{r-k-1} \end{bmatrix} = \sum_{k=0}^{r} (-1)^{2r} a_k z^k = \sum_{k=0}^{r} a_k z^k \qquad \square$$

Problem 5

Consider the system of equations

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}' = \begin{bmatrix} -1000 & 1 \\ 0 & -1/10 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

$$u_1(0) = 1, \quad u_2(0) = 2,$$

whose exact solution is

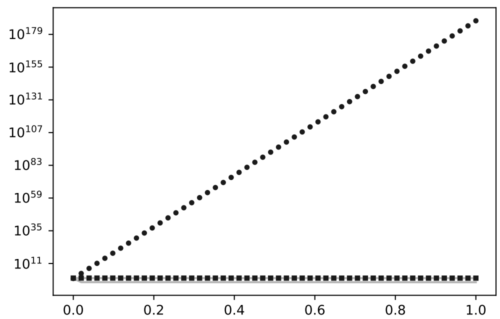
$$u_1(t) = \frac{9979}{9999} e^{-1000t} + \frac{20}{9999} e^{-t/10}, \quad u_2(t) = 2e^{-t/10}.$$

(a) Use the classical fourth-order Runge-Kutta method to solve this system of equations, integrating out to $T = 1$. What size time step is necessary to achieve a reasonably accurate approximate solution? Turn in a plot of $U_1(t)$ and $U_2(t)$ that shows what happens if you choose the time step too large, and also turn in a plot of $U_1(t)$ and $U_2(t)$ once you have found a good size time step.

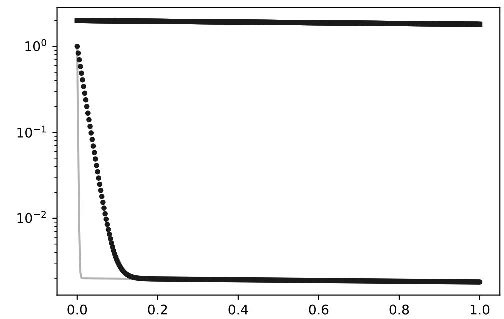
(b) Now solve this system of ODEs using MATLAB's `ode23s` routine (which uses a second-order implicit method). How many time steps does it require? Can you explain why a second-order method can solve this problem accurately using fewer time steps than the fourth-order Runge-Kutta method?

Solution

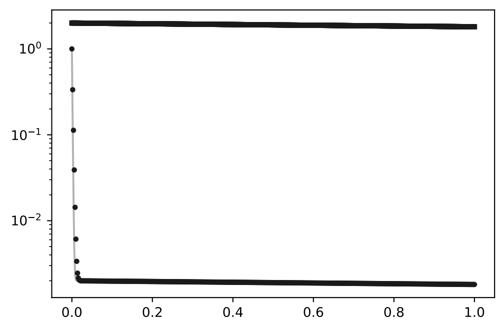
(a) We use the Runge-Kutta 4th order solver from last assignment to solve the given system. Figure 1 shows plots of the found solutions vs time are show for a variety of mesh sizes. Note that a log scale was used on the vertical axis so that the solutions apear piecewise linear. We test multiple mesh sizes and find that around $N = 1200$ the solution has an error (measured as infinity norm of actual solution and compute solution) of roughly the same as the output of `solve_ivp` using the Runge-Kutta 2-3 option and default flags.



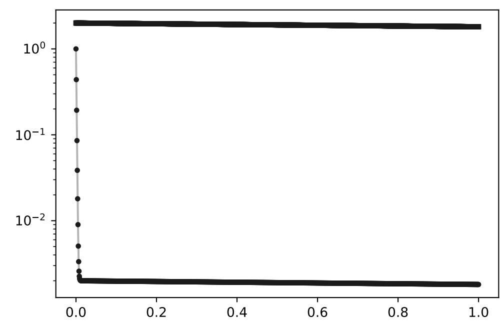
(a) $N = 51$, error = 9.584e188



(b) $N = 375$, error = 4.990



(c) $N = 500$, error = 0.34228

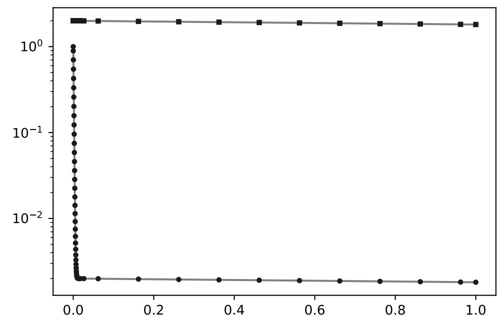


(d) $N = 1200$, error = 0.00921

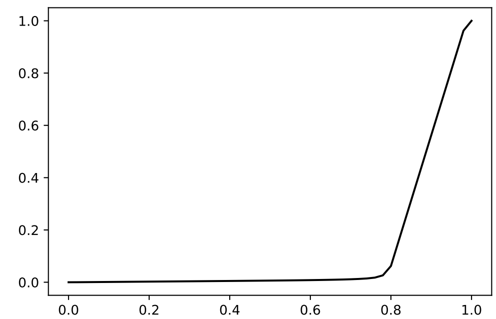
Figure 1: Runge-Kutta 4th order solutions. circle: u_1, square: u_2, actual solution: grey line

(b) We use the `ode23s` to solve the same system. Figure 2a shows the results. In particular note that the algoritm used just 51 mesh points. When 51 mesh points are used with the Runge-Kutta 4th order solver without any sort of error control the results are wildly unstable. In particular, for low values of t the solution is not matched. Figure 2b shows the mesh poins used byt the `ode23s` algorithm vs. linear spaced meshpoints used by the RK4 algorithm above. It is clear that the mesh is far more dense near low values of t where the solution changes most rapidly.

By choosing where to put the mesh points, the algorithm is able to pick a good spacing so that the solution does not require many points. This is illustrated in the differences between Figures 1d and 2a. On the steep portion of the graph, `ode23s` places more mesh points, even though fewer points are used total. This is because not as many points are needed in the later part of the graph.



(a) `ode23s` solution, $N = 51$, error $= 0.00923$



(b) linear time steps vs time steps from RK23

Figure 2