NTNU

TMA4215 Numerical Mathematics — Fall 2016

Problem Set 3

1 Kutta's method from 1901 is definitely the most famous of all explicit Runge–Kutta pairs. Its Butcher tableau is given by

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$$

a) Verify that the method is of order 4 by confirming that it obeys all the eight order conditions.

Solution. The eight order conditions for a 4th-order Runge-Kutta method are

$$\sum_i b_i = 1 \qquad (1)$$

$$\sum_i b_i c_i = \frac{1}{2} \qquad (2)$$

$$\sum_i b_i c_i^2 = \frac{1}{3} \qquad (3)$$

$$\sum_{i,j} b_i a_{ij} c_j = \frac{1}{6} \qquad (4)$$

$$\sum_i b_i c_i^3 = \frac{1}{4} \qquad (5)$$

$$\sum_{i,j} b_i c_i a_{ij} c_j = \frac{1}{8} \qquad (6)$$

$$\sum_{i,j} b_i a_{ij} c_j^2 = \frac{1}{12} \qquad (7)$$

$$\sum_{i,j,k} b_i a_{ij} a_{jk} c_k = \frac{1}{24} \qquad (8)$$

To make sure that the given method satisfies these conditions, we have to insert the values from the Butcher-tableau. For instance, (5) and (7) yield

$$\sum_i b_i c_i^3 = b_2 c_2^3 + b_3 c_3^3 + b_4 c_4^3 = \tfrac{1}{3}\left(\tfrac{1}{2}\right)^3 + \tfrac{1}{3}\left(\tfrac{1}{2}\right)^3 + \tfrac{1}{6}\cdot 1^3 = \frac{1}{4}$$

and

$$\sum_{i,j} b_i a_{ij} c_j^2 = b_3 a_{32} c_2^2 + b_4\left(a_{42}c_2^2 + a_{43}c_3^2\right) = \tfrac{1}{3}\tfrac{1}{2}\left(\tfrac{1}{2}\right)^2 + \tfrac{1}{6}\left(0\cdot\left(\tfrac{1}{2}\right)^2 + 1\cdot\left(\tfrac{1}{2}\right)^2\right) = \frac{1}{12}.$$

b) A captivating thought would be to find another set of weights $\{\widehat{b}_1, \widehat{b}_2, \widehat{b}_3, \widehat{b}_4\}$, such that the associated method has order 3. This we can use for error estimates and stepsize control. Try to find such a set of weights.

Solution. We seek a set of weights $\widehat{b}_i$ such that the conditions (1)–(4) hold, that is,

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & \frac{1}{4} & \frac{1}{4} & 1 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \widehat{b}_1 \\ \widehat{b}_2 \\ \widehat{b}_3 \\ \widehat{b}_4 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{6} \end{bmatrix}.$$

This system has a unique solution, namely the weights in the originial Kutta's method, $\widehat{b}_1 = \widehat{b}_4 = \frac{1}{6}$ and $\widehat{b}_2 = \widehat{b}_3 = \frac{1}{3}$. Hence, there is no set of weights which makes the associated method of order 3.

$\boxed{2}$ a) Show that an explicit Runge–Kutta method of order 3 with 3 stages has to satisfy

$$3\,a_{32}\,c_2^2 - 2\,a_{32}\,c_2 - c_2\,c_3 + c_3^2 = 0. \tag{9}$$

Solution. The four order conditions for methods of order 3 with 3 stages are

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & c_2 & c_3 \\ 0 & c_2^2 & c_3^2 \\ 0 & 0 & a_{32}c_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{6} \end{bmatrix},$$

which is a linear system in b_1, b_2 and b_3. Since b_1 is automatically given by b_2 and b_3 from the first equation, we can focus on the reduced system

$$\begin{bmatrix} c_2 & c_3 \\ c_2^2 & c_3^2 \\ 0 & a_{32}c_2 \end{bmatrix} \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{6} \end{bmatrix}.$$

In order for this system to have a solution, which must be unique, the right-hand side needs to be in the column space of the matrix. As such, the 3×3 matrix we obtain by adding the right-hand side as a third column must be singular. We hence require

$$\begin{vmatrix} c_2 & c_3 & \frac{1}{2} \\ c_2^2 & c_3^2 & \frac{1}{3} \\ 0 & a_{32}c_2 & \frac{1}{6} \end{vmatrix} = 0, \qquad \text{implying that} \qquad c_2\left(3\,a_{32}c_2^2 - 2\,a_{32}c_2 - c_2c_3 + c_3^2\right) = 0.$$

Note that $c_2 \neq 0$, because if not, the first column would then be exact zero and the second would not be proportional to the third. This would violate the last order condition. We therefore conclude that a necessary and sufficient condition is given by

$$3\,a_{32}c_2^2 - 2\,a_{32}c_2 - c_2c_3 + c_3^2 = 0.$$

b) Characterize all 3rd-order explicit Runge–Kutta methods with 3 stages which satisfy $a_{31} = 0$, that is, which satisfy $a_{32} = c_3$. How many free parameters are there?

Solution. Insertion of $a_{32} = c_3$ into (9) yields the condition

$$c_3\left[c_3 - 3c_2(1-c_2)\right] = 0.$$

We cannot have $c_3 = 0$, since then $a_{32} = 0$, which violates the last order condition. Instead we require

$$c_3 = 3c_2(1-c_2),$$

and so the method is characterized by the single free parameter c_2, where $c_2 = 0$ and $c_2 = 1$ must be excluded. Also, we find that

$$b_1 = \frac{c_2^3 - \frac{3}{2}c_2^2 + \frac{2}{3}c_2 - \frac{1}{9}}{c_2^2(c_2-1)}, \qquad b_2 = \frac{c_2 - \frac{1}{3}}{2c_2^2} \qquad \text{and} \qquad b_3 = \frac{1}{18c_2^2(1-c_2)}.$$

Another way is to set up the four order conditions again. There are four equations and five unknowns, which means that there is one free parameter.

c) Find all explicit methods of order 2 which have the same A matrix as the method(s) in b), but with weights $\{\widehat{b}_1, \widehat{b}_2, \widehat{b}_3\}$ that also satisfy $\widehat{b}_3 = 0$.

Solution. Since $c_1 = 0$ (explicit method) and $\widehat{b}_3 = 0$, the order conditions become

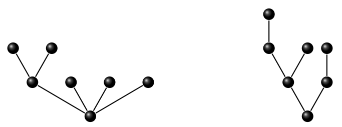
$$\widehat{b}_1 + \widehat{b}_2 = 1 \qquad \text{and} \qquad \widehat{b}_2 c_2 = \frac{1}{2}.$$

We require $c_2 \notin \{0,1\}$ from b), and so

$$\widehat{b}_1 = 1 - \frac{1}{2c_2} \qquad \text{and} \qquad \widehat{b}_2 = \frac{1}{2c_2}.$$

Hence, we again get a family of methods with the free parameter c_2.

$\boxed{3}$ Write down the order conditions corresponding to the following rooted trees:

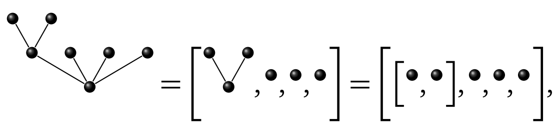


Also, write down the order of the trees.

Solution. The order condition corresponding to a rooted tree τ is

$$\widehat{\phi}(\tau) = \frac{1}{\gamma(\tau)},$$

where $\widehat{\phi}(\tau)$ is the "elementary weight" and $\gamma(\tau)$ is the "density" of τ; see the notes on the webpage on how to compute these. For example,



so that $\gamma = 7\cdot(3\cdot 1\cdot 1)\cdot 1\cdot 1\cdot 1 = 21.$

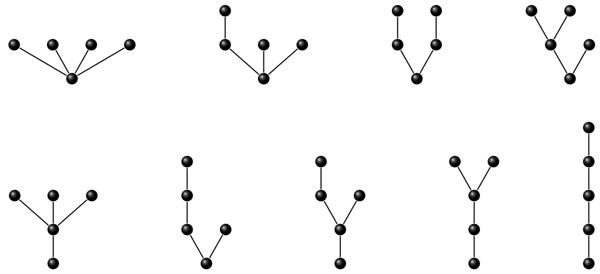
In total we get

$$\sum_{i,j} b_i c_i^3 a_{ij} c_j^2 = \frac{1}{21} \qquad \text{and} \qquad \sum_{i,j,k,l} b_i a_{ij} c_j a_{jk} c_k a_{il} c_l = \frac{1}{112}$$

for the left and right trees, respectively. The order $\rho(\tau)$ equals the number of nodes in τ, which in both cases are 7.

4 Draw all rooted trees of order 5 (there are 9 different ones).

Solution.



5 a) Implement an adaptive Runge–Kutta solver based on the Bogacki-Shampine pair:

0				
1/2	1/2			
3/4	0	3/4		
1	2/9	1/3	4/9	
	2/9	1/3	4/9	0
	7/24	1/4	1/3	1/8

Use $P = 0.9$ as a pessimist factor and measure the local error estimate in the 2-norm.

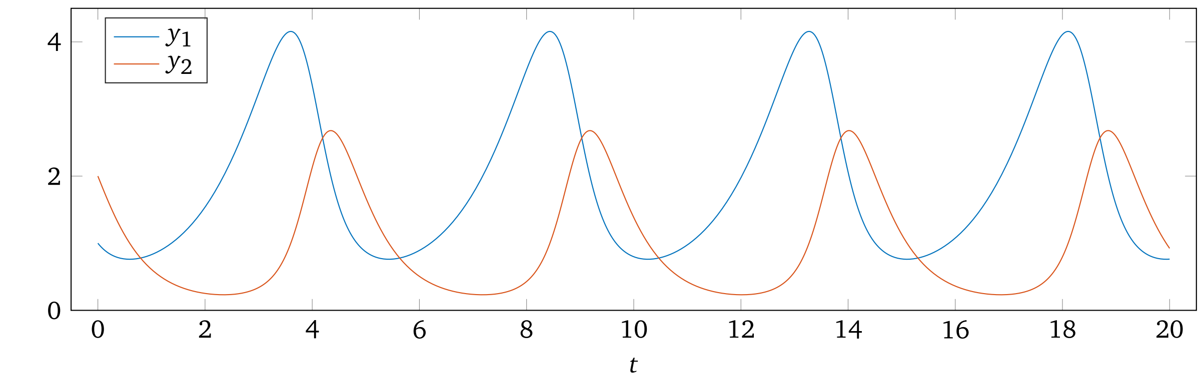
Solution. The implementation can be found in the file `rkbs.m` on the webpage. The Bogacki-Shampine method is a 3rd order advancing and 2nd order estimating method. We have used EPS stepsize selection. In addition, the solver includes a rough preallocation of the solution data, which is key for large problems.

b) Test your program on the Lotka-Volterra equation

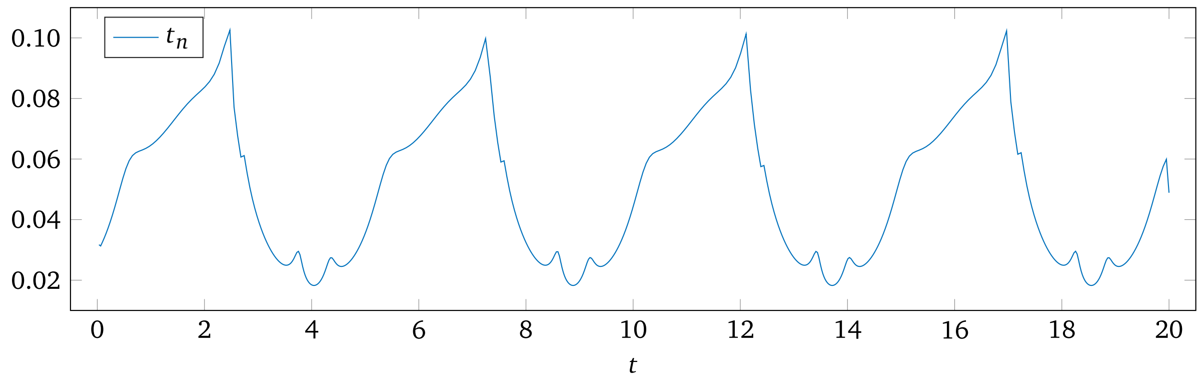
$$\begin{aligned} y_1' &= y_1 - y_1 y_2 \\ y_2' &= y_1 y_2 - 2y_2, \end{aligned}$$

on the interval $[0, 20]$ with initial conditions $y_1(0) = 1$ and $y_2(0) = 2$. Let $h_0 = 1/20$ be the starting stepsize and choose the tolerance 1×10^{-5}. Make a plot of the solution as well as of the stepsizes h_n. Do h_n behave as expected?

Solution. The solution is plotted in Figure 1a, while Figure 1b shows the stepsizes. We notice that the stepsizes are largest when y_1 and y_2 change little, and *vice versa*, that small stepsizes correspond to where y_1 and y_2 change much.



(a) Solution.



(b) Stepsizes.

Figure 1: Solution of the Lotka-Volterra equation using the adaptive Bogacki-Shampine pair, with a display of the stepsizes.

c) Plot the ratio $\|e_N\|_2/\mathtt{TOL}$ between the tolerance and the global error, where e_N is the global error at the end of the interval. For example, you can choose a series of tolerances 2^{-i}, where $i = 1, \ldots, 20$. In order to compute the global error, compare with an "exact" reference solution using MATLABs ODE45 integrator. What do you observe?

Solution. In Figure 2 we have plotted the various ratios. As is observed, there is proportionality between the global error and the tolerance, which is expected for EPS stepsize selection together with local extrapolation. In this case, TOL underestimates the global error by a factor of ≈ 10.

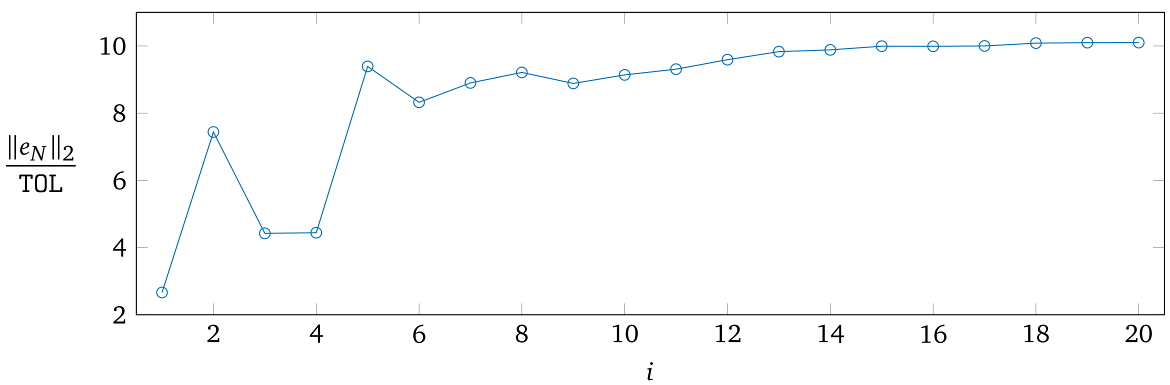


Figure 2: The behavior of $\|e_N\|_2/\mathtt{TOL}$ with $\mathtt{TOL} = 2^{-i}$ and $i = 1, \ldots, 20$.