

**Problem 1)**

Write three functions that numerically solves a first order ODE by using:

- (a) Explicit Euler method: This is a very simple but educational exercise to get feeling about the method.
- (b) Modified Euler method: The method is a bridge to methods used in particle simulations.
- (c) The 5<sup>th</sup> order Runge-Kutta-Fehlberg (RKF45). This is a very practical exercise leading to use a quite powerful method with an error estimate.

(a) The explicit Euler method is perhaps the most simple method for numerically solving differential equation, typically introduced in a first-year calculus course. A generic first-order ordinary differential equation takes the form

$$\frac{dx}{dt} = f(t, x(t)), \quad (1)$$

where  $t$  is our independent variable (e.g. time) and  $x(t)$  is the function of interest. Of course, we can naively approximate the derivative at  $t$  as the slope of the secant line between  $t$  and  $t + h$ , where  $h$  is small step. Doing, so we have

$$\frac{dx}{dt} \approx \frac{x(t+h) - x(t)}{h} = f(t, x(t)). \quad (2)$$

From this, we can approximate  $x(t+h)$  by having some initial value for  $x(t)$  through

$$x(t+h) \approx x(t) + hf(t, x(t)), \quad (3)$$

and therefore, we can define a recursive sequence of approximations by first specifying an initial condition  $x(t_0) = x_0$  and computing

$$x_{n+1} = x_n + hf(t_n, x_n), \quad (4)$$

where  $t_n = t_0 + nh$  and  $x_n = x(t_n)$ , until we reach some final time  $t_1$ . Typically, the step-size is derived by specifying the number of steps  $N$  between  $t_0$  and  $t_1$  as  $h = (t_1 - t_0)/N$ .

(b) As one probably knows from a differential equations class, the Euler method is only  $\mathcal{O}(h^2)$  accurate, modulated by the size of the second-derivative of  $x$ , and hence suffers from some significant systematic defects unless our step size  $h \ll (t_1 - t_0)$  (i.e.  $N \gg 1$ ). There are two ways to make these errors smaller. Firstly, we can devise methods such that the local truncation error of order  $\mathcal{O}(h^n)$  is such that  $n$  is larger, or secondly, we can devise a method which adjusts the size of  $h$  at each step. A first improvement to Euler's method is the modified Euler (or Heun's) method, which aims to mitigate the error as follows. If we adopt notation consistent with the general class of Runge-Kutta methods,

we define our recursive sequence as follows:

$$k = x_n + hf(t_n, x_n) \quad (5)$$

$$x_{n+1} = x_n + \frac{h}{2} \left[ f(t_n, x_n) + f(t_{n+1}, k) \right]. \quad (6)$$

Indeed, by averaging the slopes in the above way, we increase by one power of  $h$  the local truncation error to  $\mathcal{O}(h^3)$ .

(c) As alluded to in part (b) another way to obtain more accurate numerical schemes is to devise one such that our step size is adjusted at each step so that at each step our prescribed error is less than some tolerance. In the RKF45 scheme, we use the 5<sup>th</sup> order Runge-Kutta scheme for our recursive sequence. Generically, an  $N^{\text{th}}$  Runge-Kutta method is given by

$$k_1 = hf(t_n, x_n) \quad (7)$$

$$k_2 = hf(t_n + b_1h, x_n + A_{11}k_1) \quad (8)$$

$$\vdots \quad (9)$$

$$k_{N+1} = hf(t_n + b_Nh, x_n + A_{Ni}k_i) \quad (10)$$

$$x_{n+1} = x_n + hc_i^{(N)}k_i, \quad (11)$$

where the coefficients  $b_i$ ,  $c_i^{(N)}$ , and  $A_{ij}$  are method-specific. Note that for brevity we utilize Einstein summation convention where repeated indices are summed implicitly. Although we do not have the exact solution with which to compute an error, we can define an error as the difference between the  $N^{\text{th}}$  and  $(N-1)^{\text{th}}$  RK method:

$$e_{n+1} = \left[ c_i^{(N)} - c_i^{(N-1)} \right] k_i, \quad (12)$$

where  $d_i$  is again a vector of coefficients dependent on the method. With this estimate, if our error  $e_{n+1} > \epsilon$ , where  $\epsilon$  is some tolerance, then we redefine our step size as

$$h = \min\{0.9h(\epsilon/|e_{n+1}|)^{1/(N+1)}, 0.5h\} \quad (13)$$

and repeat the step. Alternatively, if  $e_{n+1} \leq \epsilon$ , then we accept the step and redefine our step size as

$$h = \min\{0.9h(\epsilon/|e_{n+1}|)^{1/(N+1)}, 2h\}. \quad (14)$$

Typically, the coefficients are specified by a Butcher tableau

$$\begin{array}{c|c} b & A \\ \hline & c^{(N)} \\ & c^{(N-1)} \end{array}$$

For the RKF45 method, the Butcher tableau is as follows:

1/4	1/4					
3/8	3/32	9/32				
12/13	1932/2197	-7200/2197	7296/2197			
1	439/216	-8	3680/513	-845/4104		
1/2	-8/27	2	-3544/2565	1859/4104	-11/40	
	16/135	0	6656/12825	28561/56430	-9/50	2/55
	25/216	0	1408/2565	2197/4104	-1/5	0

### Problem 2)

Solve numerically following ODEs (employing all three methods), and study the effect of step size by calculating the relative error either by comparing with analytic solution or with RKF45 (if analytic solution is not attainable).

- (a)  $x' = x + e^{-t}$  with  $x(0) = 0$  in the interval  $t \in [0, 1]$ .
- (b)  $x' = x + 2 \cos t$  with  $x(0) = 1$  in the interval  $t \in [0, 1]$ .
- (c)  $x' = tx^2$  with  $x(0) = 1$  in the interval  $t \in [0, 1]$ .
- (d)  $x' = 1.5 \sin 2x - x \cos t$  with  $x(0) = 1$  in the interval  $t \in [0, 10]$ .

We can obtain analytic solutions for parts (a) – (c), which are as follows:

$$(a) : x(t) = \sinh t \quad (15)$$

$$(b) : x(t) = 2e^t + \sin t - \cos t \quad (16)$$

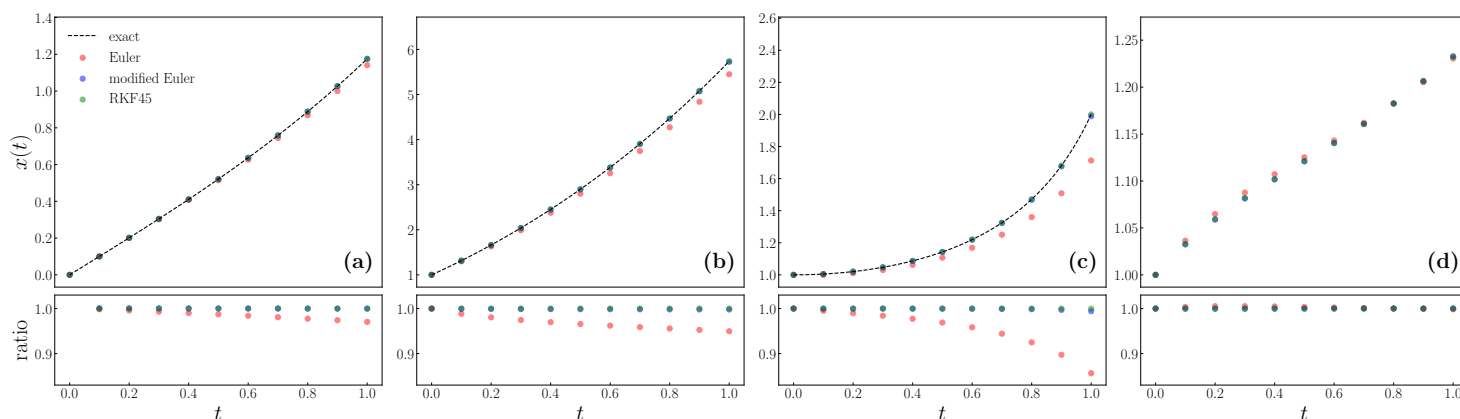
$$(c) : x(t) = \frac{2}{2 - t^2}. \quad (17)$$

Numeric results are displayed in Fig. 1. Across the board, we can see that the RKF45 and modified Euler methods perform quite well at a step size of  $h = 0.1$ , being within 1% of the analytic values, while the Euler method tends to start diverging from the exact or RKF45 solution closer to  $t = 1$ , being as much as roughly 15% from the analytic result in part (c).

### Problem 3)

Using the error estimation in the 5<sup>th</sup> order Runge-Kutta-Fehlberg method, you can upgrade your code to include adaptive step control. Apply your upgraded code to problem 2.4. Report your observations when you start either with a very small step or very large step and compare with the results when you do not employ adaptive step size.

From Fig. 2, we can see the performance of our RKF45 with static and adaptive step sizes for three different initial step sizes:  $10^{-2}$ ,  $10^{-1}$ , 1. It is clear that the solutions cannot be



**Figure 1:** In the top row of panels, a comparison of the exact (when an analytic solution is possible to obtain), Euler, modified Euler, and RKF45 (without adaptive step sizes) solutions to the respective differential equation given in problem 2 is shown, and in the lower panel, we plot the ratio of the results of the numerical scheme to the value provided by the exact solution (or RKF45 result when an analytic solution is not available).

properly reconstructed with one static step, but it is not too bad when we take 10 steps and certainly a more resolved construction when we take 100 static steps. When one permits the step-size to change with a tolerance of  $\epsilon = 10^{-4}$ , we can see that a step size of about  $1/6$  is approximately optimal across the time interval to construct the solution, and therefore there is only a slight adjustment of the step sizes when  $h_0 = 0.1$ . When  $h_0 = 1$ , the first step is not accepted until a much smaller step of roughly  $h = 0.3$  is attained and decrease subsequently. On the other hand, while the steps are accepted initiall for  $h_0 = 0.01$ , the code makes the subsequent steps larger because such small steps are not necessary to accurately produce the result.

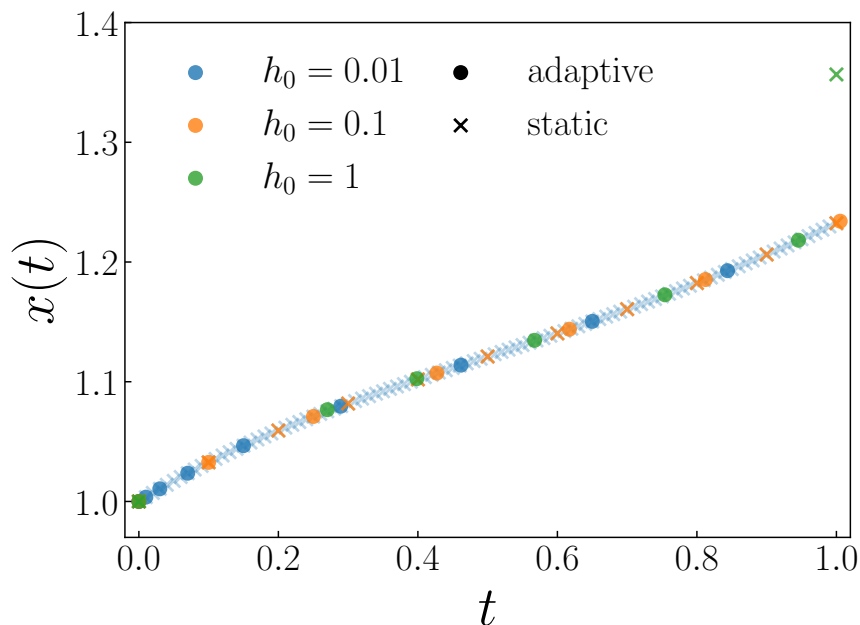
#### Problem 4)

- (a) Under certain conditions the rate of change of the temperature of a body, immersed in a medium whose temperature (kept constant) differs from it, is proportional to the difference in temperature between it and the medium. In mathematical symbols, this statement is written as

$$\frac{dT}{dt} = -k(T - T_M), \quad (18)$$

where  $k > 0$  is a proportionality constant,  $T$  is the temperature of the body at any time  $t$ , and  $T_M$  is the constant temperature of the medium. Let's say you have a cup of coffee at  $90^\circ\text{C}$  initially ( $T_0 = 90$ ) and its placed in a room where the temperature is  $20^\circ\text{C}$  initially ( $T_M = 20$ ). The constant  $k$  is given as  $0.05$  (per minute). We want to find the temperature of the coffee after 10 minutes. Think, how can you test your code?

- (b) A projectile of mass  $m = 1$  kg is thrown vertically upward from the ground



**Figure 2:** Comparison of solutions using the RKF45 method with and without adaptive step sizes for an array of initial step sizes  $h_0$ .

with an initial velocity  $v_0 = 20$  m/s. The projectile is the subject to linear air resistance so that the force of air resistance is given  $f_d = -bv$  with  $b = 1$  kg/s. Find its velocity as a function of time. For how long will it rise?

Note: Newton's second law reads

$$m \frac{dv}{dt} = -mg - bv, \quad (19)$$

where  $m$  is the mass,  $g$  is the free-fall acceleration, and  $b$  is the linear drag coefficient. Think, how can you test your code?

- (c) A projectile of mass  $m = 1$  kg is thrown vertically upward from the ground with an initial velocity  $v_0 = 20$  m/s. The projectile is the subject to linear air resistance so that the force of air resistance is given  $f_d = -cv^2$  with  $c = 0.1$  kg/m. Find its velocity as a function of time. For how long will it rise?

Note: Newton's second law reads

$$m \frac{dv}{dt} = -mg - cv^2, \quad (20)$$

where  $m$  is the mass,  $g$  is the free-fall acceleration, and  $b$  is the linear drag coefficient. Think, how can you test your code?

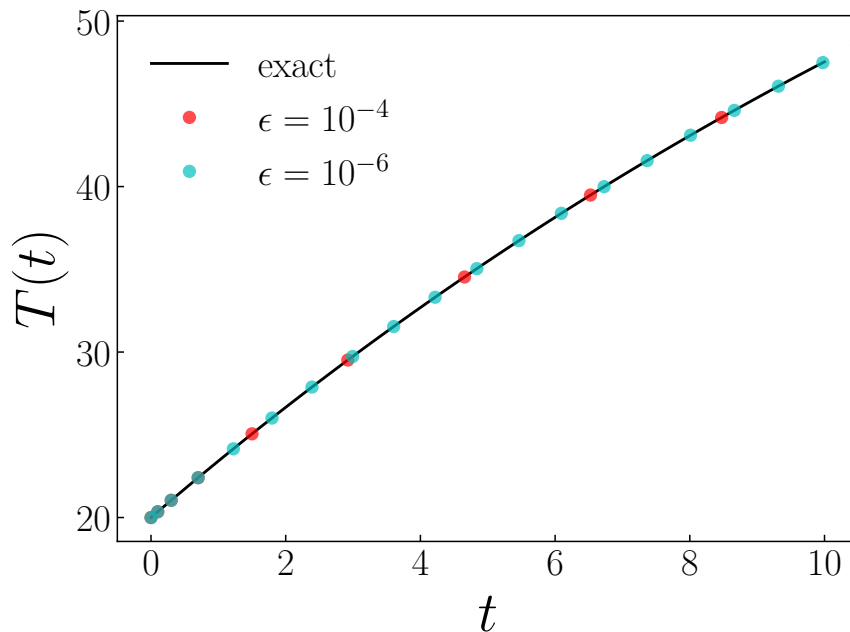
(a) The exact solution to the differential equation is

$$T(t) = T_0 + T_M(1 - e^{-kt}), \quad (21)$$

so

$$T(10 \text{ min}) \approx 47.543^\circ\text{C}. \quad (22)$$

With the RKF45 method with  $\epsilon = 10^{-4}$ , we obtain  $T(10) \approx 47.504^\circ\text{C}$ . However, with  $\epsilon = 10^{-6}$ , we have  $T(10) \approx 47.542^\circ\text{C}$ . Although our result is quite accurate in both cases, a lesson to be wary of is that the error can propagate to a noticeable and perhaps significant level over a few steps, so one must always balance computational efficiency and accuracy.



**Figure 3:** Results using RKF45 method to solve Newton's law of cooling equation with the prescribed initial conditions and physical constants.

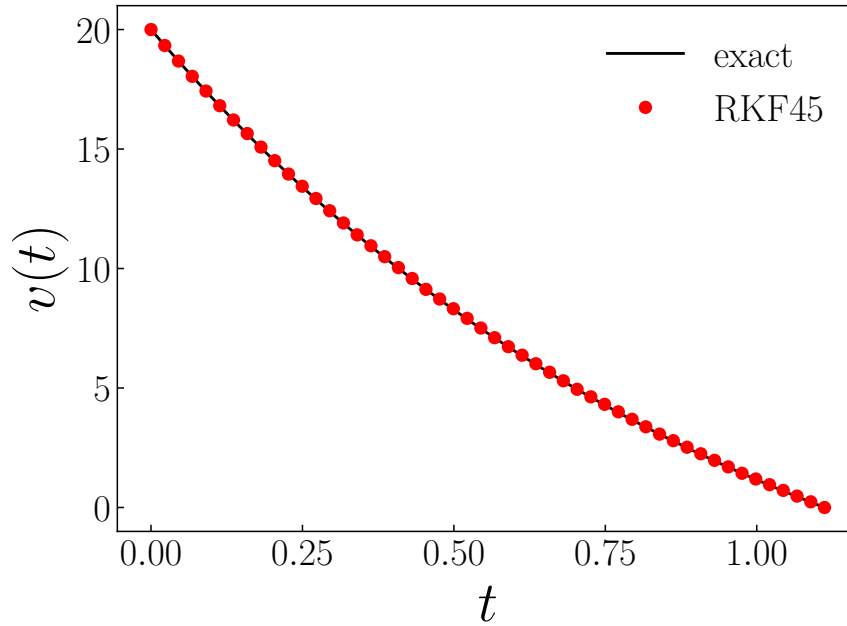
(b) The analytic solution to this differential equation is

$$v(t) = (v_0 + v_t)e^{-t/\tau} - v_t, \quad (23)$$

where  $v_t = mg/b$  and  $\tau = m/b$  are the terminal speed and characteristic damping time of our system. Solving this equation for  $T$  where

$$v(T) = 0 \Rightarrow T = \tau \ln\left(1 + \frac{v_0}{v_t}\right) = 1.112 \text{ s} \quad (24)$$

for the input physical parameters and initial condition.



**Figure 4:** Results using the RKF45 method to solve Newton’s 2<sup>nd</sup> law for a vertically launched projectile under a linear drag force with initial velocity  $v_0 = 20$  m/s.

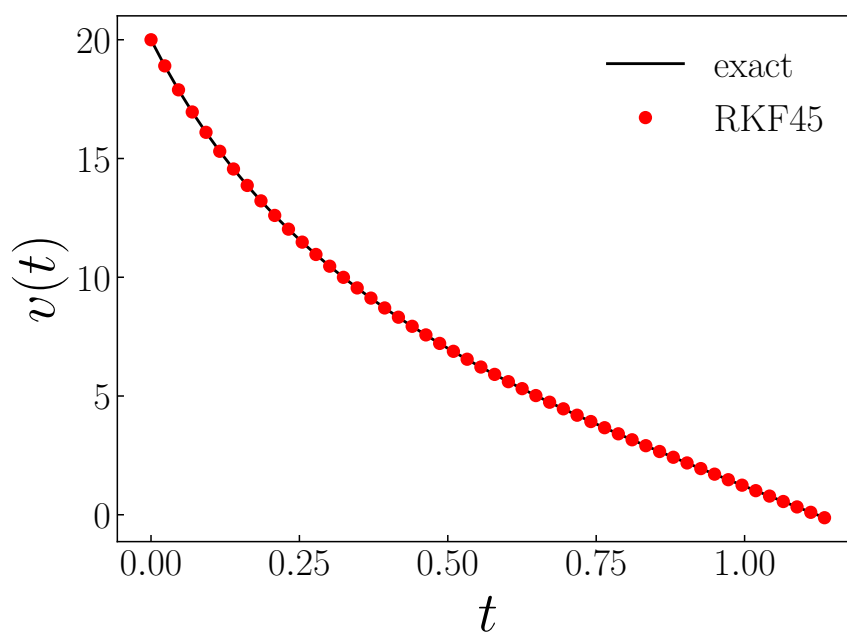
(c) In this case, the analytic solution (for purely upward motion – the sign is flipped on the drag force when the projectile moves down) is

$$v(t) = v_c \tan \left[ \arctan \left( \frac{v_0}{v_c} \right) - \frac{t}{\tau} \right], \quad (25)$$

where  $v_c = \sqrt{mg/c}$  and  $\tau = v_c/g$ . Solving for the time  $T$  where the velocity  $v(T) = 0$  we find

$$T = \tau \arctan \left( \frac{v_0}{v_c} \right) \approx 1.122 \text{ s} \quad (26)$$

with the physical parameters and initial speed as given.



**Figure 5:** Results using the RKF45 method to solve Newton's 2<sup>nd</sup> law for a vertically launched projectile under a quadratic drag force with initial velocity  $v_0 = 20$  m/s.