# 1.2 Ordinary Differential Equations

*This project builds on material covered in the Part IA lectures on Computational Projects, see* `http://www.maths.cam.ac.uk/undergrad/catam/part-ia-lectures`*. The Part IA Differential Equations and Part IB Methods courses are also relevant.*

## 1 Background Theory

This project is concerned with the numerical step-by-step integration of ordinary differential equations (ODEs) of the form

$$\frac{d\mathbf{y}}{dx} = \mathbf{f}(x, \mathbf{y})\,, \tag{1a}$$

where $\mathbf{y}$ and $\mathbf{f}$ are vectors of length $m$. An initial condition is specified at, say, $x = x_0$, i.e.

$$\mathbf{y}\,(x_0) = \mathbf{y}_0\,. \tag{1b}$$

In the first part of the project the performance of two different numerical methods will be examined. A first-order equation ($m = 1$) has been chosen which has an analytic solution for comparison. In the second part, one of the methods is extended to solve a second-order problem.

The numerical methods, presented for $m = 1$, to be investigated are as follows.

(a) The **Euler** method, or more precisely the **forward Euler** method, for scalar ODEs is the simple scheme

$$Y_{n+1} = Y_n + hf(x_n, Y_n)\,, \tag{2}$$

where $Y_n$ denotes the numerical solution at $x_n \equiv x_0 + nh$, that is, at the $n$th step with step length $h$. The exact solution at $x_n$ is denoted by $y(x_n)$.

*Definition.* The *global error* after the $n$th step is defined as

$$E_n = Y_n - y(x_n)\,. \tag{3a}$$

*Definition.* The *local error* of the first step is defined as $e_1 = Y_1 - y(x_1)$. For subsequent steps the *local error* is defined as

$$e_n = Y_n - w(x_n)\,, \tag{3b}$$

where $w(x_n)$ is the exact solution to (1a) at $x = x_n$ starting from $x = x_{n-1}$ and $y = Y_{n-1}$. Note that, in general, $Y_{n-1} \neq y(x_{n-1})$ for $n > 1$.

For the Euler method it can be shown that $e_n$ is $O(h^2)$ as $h \to 0$. As a result the Euler method is said to have *first-order accuracy.*

(b) The fourth-order **Runge–Kutta** (RK4) method employs the scheme:

$$Y_{n+1} = Y_n + \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)\,, \tag{4a}$$

where

$$k_1 = hf(x_n,\, Y_n)\,, \tag{4b}$$
$$k_2 = hf(x_n + \tfrac{1}{2}h,\, Y_n + \tfrac{1}{2}k_1)\,, \tag{4c}$$
$$k_3 = hf(x_n + \tfrac{1}{2}h,\, Y_n + \tfrac{1}{2}k_2)\,, \tag{4d}$$
$$k_4 = hf(x_n + h,\, Y_n + k_3)\,. \tag{4e}$$

The RK4 method has *fourth-order accuracy*, i.e. $e_n$ is $O(h^5)$ as $h \to 0$. When the RK4 method is used on coupled ODEs, each of $f$, $Y_n$, $k_1$, $k_2$, $k_3$ and $k_4$ become vectors of the same dimension as the number of coupled ODEs.

 

The theoretical background for the stability and accuracy of these methods is set out in, for example, *An Introduction to Numerical Methods and Analysis* by J.F.Epperson, *An Introduction to Numerical Methods* by A.Kharab and R.B.Guenther and *Numerical Recipes* by Press *et al.*

# 2 Stability and accuracy of the numerical methods

The example to be studied in detail in this section is the scalar version of equation (1a) with

$$f(x,y) = -16y + 15\mathrm{e}^{-x} \tag{5a}$$

and initial condition

$$y(0) = 0\ . \tag{5b}$$

This has the exact solution

$$y(x) = \mathrm{e}^{-x} - \mathrm{e}^{-16x}\ . \tag{6}$$

**Programming Task:** Write program(s) to apply each of the methods (a) and (b) to this problem.

## 2.1 Stability

**Question 1** Using the Euler method, starting with $Y_0 = 0$, compute $Y_n$ for $x$ up to $x = 6$ with $h = 0.6$, i.e. for $n$ up to $6/h = 10$. Tabulate the values of $x_n$, the numerical solution $Y_n$, the analytic solution $y(x_n)$ from (6), and the global error $E_n \equiv Y_n - y(x_n)$. You should find that the numerical result is unstable: the error oscillates with a magnitude that ultimately grows proportional to $\mathrm{e}^{\gamma x}$, where the 'growth rate' $\gamma$ is a positive constant which you should estimate.

Repeat with $h = 0.4$, 0.2, 0.125 and 0.1, presenting only a judicious selection of output to illustrate the behaviour. What effect does reducing $h$ have on the instability and its growth rate?

**Question 2**

(i) Find the analytic solution of the Euler *difference* equation

$$Y_{n+1} = Y_n + h\left(-16Y_n + 15\left(\mathrm{e}^{-h}\right)^n\right) \quad \text{with} \quad Y_0 = 0\,. \tag{7}$$

(ii) Hence explain why and when instability occurs, and with what growth rate.

(iii) Show that in the limit $h \to 0$, $n \to \infty$ with $x_n \equiv nh$ fixed, the solution of the difference-equation problem (7) converges to the solution (6) of the differential-equation problem specified by (1a), (5a) and (5b).

## 2.2 Accuracy

**Question 3** Integrate the ODE numerically with $h = 0.05$ from $x = 0$ to $x = 4$ using both the Euler and RK4 methods, in each case tabulating the numerical solution $Y_n$ against $x_n$ (although not necessarily at every step), and plotting it with the exact solution (6) superposed.

**Question 4** For both the Euler and the RK4 methods, tabulate the global error $E_n$ at $x_n = 0.1$ against $h \equiv 0.1/n$ for $n = 2^k$ with $k = 0, 1, 2, \ldots, 15$, and plot a log–log graph of $|E_n|$ against $h$ over this range.

Comment on the relationship of your results to the theoretical accuracy of the methods.

 

### 2.3 More on the growth of errors

Now consider the scalar version of equation (1a) with

$$f(x,y) = 4y - 5\mathrm{e}^{-x} \tag{8a}$$

and initial condition

$$y(0) = 1\ , \tag{8b}$$

for which the exact solution is

$$y(x) = \mathrm{e}^{-x}\ . \tag{8c}$$

> **Question 5** Use the Euler method with $h = 0.001$ to integrate from $x = 0$ to $x = 10$. You should find that the magnitude of the global error ultimately grows *exponentially*. Solve the Euler difference-equation problem analytically; hence explain the reason for this behaviour, and identify the growth rate. Could the error be suppressed by using a smaller value of $h$? Why, or why not? What if RK4 were used instead of Euler?

## 3 Numerical solutions of second-order ODEs

This section is concerned with small 'normal-mode' oscillations of a non-uniform string under uniform positive tension $T_0$, with ends fixed at $x = 0$ and $x = 1$. The string's transverse displacement, $\eta(x,t)$, a function of longitudinal distance $x$ and time $t$, is assumed to satisfy the equation of motion

$$m(x)\frac{\partial^2 \eta}{\partial t^2} = T_0 \frac{\partial^2 \eta}{\partial x^2}\,, \tag{9a}$$

where $m(x) \equiv m_0\mu(x)$ is the mass per unit length of the string, $m_0$ is a positive constant and $\mu(0) = 1$. The corresponding boundary conditions are

$$\eta(0,t) = \eta(1,t) = 0 \quad \text{for all } t. \tag{9b}$$

The problem admits 'normal-mode' solutions of the form

$$\eta(x,t) = y(x)\cos(\omega t + \theta)\,, \tag{10}$$

where the angular frequency, $\omega$, and phase, $\theta$, are constants. From (9a) and (9b), it follows that $y(x)$ satisfies

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + p^2\mu(x)y = 0 \quad \text{and} \quad y(0) = y(1) = 0\,, \tag{11}$$

where $p = \omega\sqrt{m_0/T_0}$ is a constant and, without loss of generality, $p \geqslant 0$. The system (11) is an example of a *Sturm-Liouville eigenproblem*, i.e. only for a discrete set of values of $p$, i.e. the *eigenvalues*, are there non-zero *eigenfunction* solutions for $y$.

The remainder of this project specialises to a mass distribution of the form $\mu(x) = (1+x)^{-\alpha}$ with $\alpha$ a constant, in which case (11) becomes

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + p^2(1+x)^{-\alpha}y = 0\,, \tag{12a}$$

subject to

$$y(0) = y(1) = 0\ . \tag{12b}$$

This has an explicit analytic solution in terms of elementary functions only for special values of $\alpha$, one such being $\alpha = 4$.

 

**Question 6** When $\alpha = 4$, equation (12a) with $p \neq 0$ has general solution

$$y = A(1+x)\sin\left(p(1+x)^{-1} - \phi\right), \tag{13}$$

where $A$ and $\phi$ are arbitrary constants. What if $p = 0$? Write down the particular solution (for all $p$) satisfying

$$y = 0\ , \quad dy/dx = 1 \quad \text{at } x = 0. \tag{14}$$

Deduce that the smallest (non-negative) eigenvalue of (12a)–(12b) with $\alpha = 4$ is $p = 2\pi$, and write down the other eigenvalues and the corresponding eigenfunctions, $y(x)$ (there is no need to normalise the eigenfunctions).

The system (12a)–(12b) can be solved numerically by breaking it down into two first-order ODEs. By letting $z(x) = dy/dx$ and $\mathbf{y} = (y, z)$, we can write (12a) in the form (1a):

$$\frac{dy}{dx} = f_1(x, y, z) \equiv z\ , \quad \frac{dz}{dx} = f_2(x, y, z) \equiv -p^2(1+x)^{-\alpha} y\ . \tag{15}$$

These equations can then be solved for $\mathbf{Y_n} = (Y_n, Z_n)$ using the vector generalisation of the methods discussed for first-order ODEs.

**Programming Task:**

Write a program to compute a numerical approximation to the solution of equations (15) with initial condition (14) by the vector generalisation of the RK4 method.

**Question 7** Taking $\alpha = 4$, run your program with $p = 6$ and $h = 0.1/2^k$ for $k = 0, 1, 2, \ldots, 12$ in turn, tabulating the numerical solution $Y_n$ at $x_n = 1$ and the global error $Y_n - y(1)$ against $h \equiv 1/n$. Repeat with $p = 7$. Do the errors behave as expected?

**Programming Task:** Write a program to search for eigenvalues using the 'false position' method – a variant of the bisection method where if a root of $g(p) = 0$ has been located to the interval $(p_1, p_2)$ with $g(p_1)$ and $g(p_2)$ having opposite signs, an estimate $p_s$ for the root is calculated using the linear interpolant

$$G(p) \equiv g(p_1)\left(\frac{p_2 - p}{p_2 - p_1}\right) + g(p_2)\left(\frac{p - p_1}{p_2 - p_1}\right). \tag{16a}$$

By solving for $G(p) = 0$, the estimate is given by

$$p = p_s \equiv \frac{g(p_2)\,p_1 - g(p_1)\,p_2}{g(p_2) - g(p_1)}\,. \tag{16b}$$

This estimate is accepted if $|g(p_s)| < \epsilon$ where $\epsilon$ is a specified small value; if not, the process is iterated with $p_1$ or $p_2$ replaced by $p_s$ such that $g(p_1)$ and $g(p_2)$ still have opposite signs. For the current task, $g(p)$ is the numerical (RK4) solution $Y_n$ of (12a) and (14) at $x_n = 1$ obtained with a suitably small value of $h \equiv 1/n$.

**Question 8** Taking $\alpha = 4$, run the program to obtain an approximation to the smallest (positive) eigenvalue of (12a)–(12b) with error no more than $\pm 5 \times 10^{-6}$, using $[6, 7]$ as the initial interval; tabulate all the iterates in your report. What values are you using for $\epsilon$ and $h$? Justify these choices.

 

The final question addresses the case $\alpha = 8$, which has no simple analytic solution.

**Question 9** Modify your program to find approximations to the five smallest (positive) eigenvalues $p$ of (12a)–(12b) with $\alpha = 8$ correct to within $\pm 5 \times 10^{-6}$, and plot the corresponding eigenfunctions normalised such that

$$p^2 \int_0^1 (1+x)^{-8} \, [y(x)]^2 \, \mathrm{d}x = 1 \; . \tag{17}$$

If you wish, the integral may be evaluated using a black-box integration routine such as the MATLAB function `trapz`.

Explain carefully why you are satisfied that the eigenvalues found are indeed the smallest, and that they have the required accuracy.

Comment on the results, e.g. on the shape of the eigenfunctions, both from a mathematical point of view and in terms of the physical model. *Hint*: it *may* be instructive to look at the form of the eigensolutions for larger $p$.

## Additional Reference

Boyce, W. E., and DiPrima, R. C., 2001, *Elementary Differential Equations and Boundary Value Problems*, 7th edition. Publ. John Wiley & Sons Inc.

 

**Project 1.2: Ordinary Differential Equations**

**Marking Scheme and additional comments for the Project Report**

*The purpose of these additional comments is to provide guidance on the structure and length of your CATAM report. Use the same concepts to write the rest of the reports. To help you assess where marks have been lost, this marking scheme will be completed and returned to you during Lent Term. You are advised to keep a copy of your write-up in order to correlate your answers to the marks awarded.*

| **Question no.** | **Marks available**[1] | **Marks awarded**[2] |
|---|---|---|
| **Programming task** *Program:* for instructions regarding printouts and what needs to be in the write-up, refer to the introduction to the manual. | | |
| **Question 1** *Tables:* for presentation and layout, refer to the introduction. [*approx. 2 lines*][3] | C1.5 | |
| **Question 2** *Analytic solution:* do not include trivial steps in your working. [*approx. 15 lines*][3] | M3 | |
| **Question 3** *Graphs:* you may use one graph or two. | C1 | |
| **Question 4** *Graphs:* ditto.<br>*Comments:* what can be said about how the global error $E_n$ for each method varies with $h$? How is this reflected in the plots? [*approx. 3 lines*][3] | C1<br>M1 | |
| **Question 5** *Numerical solution:* not all the output needs to be displayed!<br>*Analytic solution:* there is no need to duplicate working from Question 2 [*approx. 5 lines*][3] | C0.5<br>M2 | |
| **Question 6** *Analytic solutions:* do not include trivial steps in your working [*approx. 3 lines*][3]. | M1 | |
| **Question 7** *Analytic solution and numerical solution compared:* The reason for computing an analytic solution is to check that the program is working correctly ('validation'). [*approx. 1 line*][3]. | C1 | |
| **Question 8** *Numerical approximations to the smallest eigenvalue:* Tabulate all the iterates;<br>explain how you have chosen $\epsilon$ and $h$ to ensure that the final approximation has the required accuracy. [*approx. 4 lines*][3] | C1<br><br>M1 | |
| **Question 9** *Numerical solutions*: explain how you have located the five smallest eigenvalues, and computed them to the required accuracy.<br>*Comments:* first identify the salient features of the graphs. Then comment on them using mathematical arguments; link to the theory of the physical system under investigation. [*approx. 30 lines*][3] | C2<br><br>M2 | |
| **Excellence marks** awarded for, *among other things*, mathematical clarity and good, clear output (graphs and tables) — see the introduction to the Project Manual. | E2 | |
| **Total Raw Marks** | 20 | |
| **Total Tripos Marks** | **40** | |

[1] C#, M# and E#: *Computational*, *Mathematical* and *Excellence* marks respectively.
[2] For use by the assessor.
[3] This figure is only meant to be indicative of the length of your answer, rather than the exact number of lines you are expected to write.

 