Chapter 5 Exercises

From: *Finite Difference Methods for Ordinary and Partial Differential Equations*
by R. J. LeVeque, SIAM, 2007.   `http://www.amath.washington.edu/~rjl/fdmbook`

Exercise 5.1 *(Uniqueness for an ODE)*

Prove that the ODE

$$u'(t) = \frac{1}{t^2 + u(t)^2}, \quad \text{for } t \geq 1$$

has a unique solution for all time from any initial value $u(1) = \eta$.

Exercise 5.2 *(Lipschitz constant for an ODE)*

Let $f(u) = \log(u)$.

(a) Determine the best possible Lipschitz constant for this function over $2 \leq u < \infty$.

(b) Is $f(u)$ Lipschitz continuous over $0 < u < \infty$?

(c) Consider the initial value problem

$$u'(t) = \log(u(t)),$$
$$u(0) = 2.$$

Explain why we know that this problem has a unique solution for all $t \geq 0$ based on the existence and uniqueness theory described in Section 5.2.1. (Hint: Argue that f is Lipschitz continuous in a domain that the solution never leaves, though the domain is not symmetric about $\eta = 2$ as assumed in the theorem quoted in the book.)

Exercise 5.3 *(Lipschitz constant for a system of ODEs)*

Consider the system of ODEs

$$u_1' = 3u_1 + 4u_2,$$
$$u_2' = 5u_1 - 6u_2.$$

Determine the Lipschitz constant for this system in the max-norm $\|\cdot\|_\infty$ and the 1-norm $\|\cdot\|_1$. (See Appendix A.3.)

Exercise 5.4 *(Duhamel's principle)*

Check that the solution $u(t)$ given by (5.8) satisfies the ODE (5.6) and initial condition. Hint: To differentiate the matrix exponential you can differentiate the Taylor series (D.31) (in Appendix D) term by term.

Exercise 5.5 *(matrix exponential form of solution)*

The initial value problem

$$v''(t) = -4v(t), \qquad v(0) = v_0, \quad v'(0) = v_0'$$

has the solution $v(t) = v_0 \cos(2t) + \frac{1}{2}v_0' \sin(2t)$. Determine this solution by rewriting the ODE as a first order system $u' = Au$ so that $u(t) = e^{At}u(0)$ and then computing the matrix exponential using (D.30) in Appendix D.

Exercise 5.6 *(matrix exponential form of solution)*

Consider the IVP

$$u_1' = 2u_1,$$
$$u_2' = 3u_1 - u_2,$$

with initial conditions specified at time $t = 0$. Solve this problem in two different ways:

(a) Solve the first equation, which only involves u_1, and then insert this function into the second equation to obtain a nonhomogeneous linear equation for u_2. Solve this using (5.8).

(b) Write the system as $u' = Au$ and compute the matrix exponential using (D.30) to obtain the solution.

Exercise 5.7 *(matrix exponential for a defective matrix)*

Consider the IVP

$$u_1' = 2u_1,$$
$$u_2' = 3u_1 + 2u_2,$$

with initial conditions specified at time $t = 0$. Solve this problem in two different ways:

(a) Solve the first equation, which only involves u_1, and then insert this function into the second equation to obtain a nonhomogeneous linear equation for u_2. Solve this using (5.8).

(b) Write the system as $u' = Au$ and compute the matrix exponential using (D.35) to obtain the solution. (See Appendix C.3 for a discussion of the Jordan Canonical form in the defective case.)

Exercise 5.8 *(Use of* `ode113` *and* `ode45`*)*

This problem can be solved by a modifying the m-files odesample.m and odesampletest.m available from the webpage.

Consider the third order initial value problem

$$v'''(t) + v''(t) + 4v'(t) + 4v(t) = 4t^2 + 8t - 10,$$
$$v(0) = -3, \quad v'(0) = -2, \quad v''(0) = 2.$$

(a) Verify that the function
$$v(t) = -\sin(2t) + t^2 - 3$$
is a solution to this problem. How do you know it is the unique solution?

(b) Rewrite this problem as a first order system of the form $u'(t) = f(u(t), t)$ where $u(t) \in \mathbb{R}^3$. Make sure you also specify the initial condition $u(0) = \eta$ as a 3-vector.

(c) Use the MATLAB function ode113 to solve this problem over the time interval $0 \le t \le 2$. Plot the true and computed solutions to make sure you've done this correctly.

(d) Test the MATLAB solver by specifying different tolerances spanning several orders of magnitude. Create a table showing the maximum error in the computed solution for each tolerance and the number of function evaluations required to achieve this accuracy.

(e) Repeat part (d) using the MATLAB function ode45, which uses an embedded pair of Runge-Kutta methods instead of Adams-Bashforth-Moulton methods.

Exercise 5.9 *(truncation errors)*

Compute the leading term in the local truncation error of the following methods:

(a) the trapezoidal method (5.22),

(b) the 2-step BDF method (5.25),

(c) the Runge-Kutta method (5.30).

Exercise 5.10 *(Derivation of Adams-Moulton)*

Determine the coefficients β_0, β_1, β_2 for the third order, 2-step Adams-Moulton method. Do this in two different ways:

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

(b) Using the relation
$$u(t_{n+2}) = u(t_{n+1}) + \int_{t_{n+1}}^{t_{n+2}} f(u(s))\, ds.$$

Interpolate a quadratic polynomial $p(t)$ through the three values $f(U^n)$, $f(U^{n+1})$ and $f(U^{n+2})$ and then integrate this polynomial exactly to obtain the formula. The coefficients of the polynomial will depend on the three values $f(U^{n+j})$. It's easiest to use the "Newton form" of the interpolating polynomial and consider the three times $t_n = -k$, $t_{n+1} = 0$, and $t_{n+2} = k$ so that $p(t)$ has the form
$$p(t) = A + B(t+k) + C(t+k)t$$

where A, B, and C are the appropriate divided differences based on the data. Then integrate from 0 to k. (The method has the same coefficients at any time, so this is valid.)

Exercise 5.11 *(Characteristic polynomials)*

Determine the characteristic polynomials $\rho(\zeta)$ and $\sigma(\zeta)$ for the following linear multistep methods. Verify that (5.48) holds in each case.

(a) The 3-step Adams-Bashforth method,

(b) The 3-step Adams-Moulton method,

(c) The 2-step Simpson's method of Example 5.16.

Exercise 5.12 *(predictor-corrector methods)*

(a) Verify that the predictor-corrector method (5.51) is second order accurate.

(b) Show that the predictor-corrector method obtained by predicting with the 2-step Adams-Bashforth method followed by correcting with the 2-step Adams-Moulton method is third order accurate.

Exercise 5.13 *(Order of accuracy of Runge-Kutta methods)*

Consider the Runge-Kutta methods defined by the tableaux below. In each case show that the method is third order accurate in two different ways: First by checking that the order conditions (5.35), (5.37), and (5.38) are satisfied, and then by applying one step of the method to $u' = \lambda u$ and verifying that the Taylor series expansion of $e^{k\lambda}$ is recovered to the expected order.

(a) Runge's 3rd order method:

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

(b) Heun's 3rd order method:

$$
\begin{array}{c|ccc}
0 & & & \\
1/3 & 1/3 & & \\
2/3 & 0 & 2/3 & \\
\hline
 & 1/4 & 0 &
\end{array}
$$

Exercise 5.14 *(accuracy of TR-ZBDF2)*

4

Use the approach suggested in the Remark at the bottom of page 129 to test the accuracy of the TR-BDF2 method (5.36).

Exercise 5.15 *(Embedded Runge-Kutta method)*

Consider the embedded Runge-Kutta method defined by the tableau below. Apply one step of this method to $u' = \lambda u$ to obtain both U^{n+1} and $\hat{U}^{n+1}$. What is the order of accuracy of each method? What error estimate would be obtained when using this embedded method on this problem? How does it compare to the actual error in the lower order method? (Assume U^n is the exact value at time t_n).

$$
\begin{array}{c|ccc}
0 & & & \\
1 & 1 & & \\
1/2 & 1/4 & 1/4 & \\
\hline
& 1/2 & 1/2 & 0 \\
\hline
& 1/6 & 1/6 & 4/6
\end{array}
$$

Exercise 5.16 *(accuracy of a Runge-Kutta method)*

(a) Determine the leading term of the truncation error (i.e., the $\mathcal{O}(k^2)$ term) for the Runge-Kutta method (5.30) of Example 5.11.

(b) Do the same for the method (5.32) for the non-autonomous case.

Exercise 5.17 *(R(z) for the trapezoidal method)*

(a) Apply the trapezoidal method to the equation $u' = \lambda u$ and show that

$$
U^{n+1} = \left(\frac{1 + z/2}{1 - z/2} \right) U^n,
$$

where $z = \lambda k$.

(b) Let

$$
R(z) = \frac{1 + z/2}{1 - z/2}.
$$

Show that $R(z) = e^z + \mathcal{O}(z^3)$ and conclude that the one-step error of the trapezoidal method on this problem is $\mathcal{O}(k^3)$ (as expected since the method is second order accurate).

Hint: One way to do this is to use the "Neumann series" expansion

$$
\frac{1}{1 - z/2} = 1 + \frac{z}{2} + \left(\frac{z}{2} \right)^2 + \left(\frac{z}{2} \right)^3 + \cdots
$$

and then multiply this series by $(1 + z/2)$. A more general approach to checking the accuracy of rational approximations to e^z is explored in the next exercises.

Exercise 5.18 *(R(z) for Runge-Kutta methods)*

Any r-stage Runge-Kutta method applied to $u' = \lambda u$ will give an expression of the form

$$U^{n+1} = R(z)U^n$$

where $z = \lambda k$ and $R(z)$ is a rational function, a ratio of polynomials in z each having degree at most r. For an explicit method $R(z)$ will simply be a polynomial of degree r and for an implicit method it will be a more general rational function.

Since $u(t_{n+1}) = e^z u(t_n)$ for this problem, we expect that a pth order accurate method will give a function $R(z)$ satisfying

$$R(z) = e^z + \mathcal{O}(z^{p+1}) \quad \text{as } z \to 0, \tag{Ex5.18a}$$

as discussed in the Remark on page 129. The rational function $R(z)$ also plays a role in stability analysis as discussed in Section 7.6.2.

One can determine the value of p in (Ex5.18a). by expanding e^z in a Taylor series about $z = 0$, writing the $\mathcal{O}(z^{p+1})$ term as

$$Cz^{p+1} + \mathcal{O}(z^{p+2}),$$

multiplying through by the denominator of $R(z)$, and then collecting terms. For example, for the trapezoidal method of Exercise 5.17,

$$\frac{1 + z/2}{1 - z/2} = \left(1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \cdots\right) + Cz^{p+1} + \mathcal{O}(z^{p+2})$$

gives

$$1 + \frac{1}{2}z = \left(1 - \frac{1}{2}z\right)\left(1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \cdots\right) + Cz^{p+1} + \mathcal{O}(z^{p+2})$$

$$= 1 + \frac{1}{2}z - \frac{1}{12}z^3 + \cdots + Cz^{p+1} + \mathcal{O}(z^{p+2})$$

and so

$$Cz^{p+1} = \frac{1}{12}z^3 + \cdots,$$

from which we conclude that $p = 2$.

(a) Let

$$R(z) = \frac{1 + \frac{1}{3}z}{1 - \frac{2}{3}z + \frac{1}{6}z^2}.$$

Determine p for this rational function as an approximation to e^z.

(b) Determine $R(z)$ and p for the backward Euler method.

(c) Determine $R(z)$ and p for the TR-BDF2 method (5.36).

Exercise 5.19 *(Padé approximations)*

A rational function $R(z) = P(z)/Q(z)$ with degree m in the numerator and degree n in the denominator is called the (m, n) *Padé approximation* to a function $f(z)$ if

$$R(z) - f(z) = \mathcal{O}(z^q)$$

with q as large as possible. The Padé approximation can be uniquely determined by expanding $f(z)$ in a Taylor series about $z = 0$ and then considering the series

$$P(z) - Q(z)f(z),$$

collecting powers of z, and choosing the coefficients of P and Q to make as many terms vanish as possible. Trying to require that they all vanish will give a system of infintely many linear equations for the coefficients. Typically these can not all be satisfied simultaneously, while requiring the maximal number to hold will give a nonsingular linear system. (Note that the $(m, 0)$ Padé approximation is simply the first $m + 1$ terms of the Taylor series.)

For this exercise, consider the exponential function $f(z) = e^z$.

(a) Determine the $(1, 1)$ Padé approximation of the form

$$R(z) = \frac{1 + a_1 z}{1 + b_1 z}.$$

Note that this rational function arises from the trapezoidal method applied to $u' = \lambda u$ (see Exercise 5.18).

(b) Determine the $(1, 2)$ Padé approximation of the form

$$R(z) = \frac{1 + a_1 z}{1 + b_1 z + b_2 z^2}.$$

(c) Determine the $(2, 2)$ Padé approximation of the form

$$R(z) = \frac{1 + a_1 z + a_2 z^2}{1 + b_1 z + b_2 z^2}.$$

You can check your answers at `http://mathworld.wolfram.com/PadeApproximant.html`, for example.

Exercise 5.20 *(R(z) for Runge-Kutta methods)*

Consider a general r-stage Runge-Kutta method with tableau defined by an $r \times r$ matrix A and a row vector b^T of length r. Let

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_r \end{bmatrix}, \qquad e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

and $z = \lambda k$.

(a) Show that if the Runge-Kutta method is applied to the equation $u' = \lambda u$ the formulas (5.34) can be written concisely as

$$Y = U^n e + zAY,$$
$$U^{n+1} = U^n + zb^T Y,$$

and hence

$$U^{n+1} = \left(I + zb^T (I - zA)^{-1} e\right) U^n. \qquad \text{(Ex5.20a)}$$

(b) Recall that by Cramer's rule that if B is an $r \times r$ matrix then the ith element of the vector $y = B^{-1} e$ is given by

$$y_i = \frac{\det(B_i)}{\det(B)},$$

where B_i is the matrix B with the ith column replaced by e, and det denotes the determinant.

In the expression (Ex5.20a), $B = I - zA$ and each element of B is linear in z. From the definition of the determinant it follows that $\det(B)$ will be a polynomial of degree at most r, while $\det(B_i)$ will be a polynomial of degree at most $r - 1$ (since the column vector e does does not involve z).

From these facts, conclude that (Ex5.20a) yields $U^{n+1} = R(z)U^n$ where $R(z)$ is a rational function of degree at most (r, r).

(c) Explain why an explicit Runge-Kutta method (for which A is strictly lower triangular) results in $R(z)$ being a polynomial of degree at most r (i.e., a rational function of degree at most $(r, 0)$).

(d) Use (Ex5.20a) to determine the function $R(z)$ for the TR-BDF2 method (5.36). Note that in this case $I - zA$ is lower triangular and you can compute $(I - zA)^{-1} e$ by forward substitution. You should get the same result as in Exercise 5.18(c).

Exercise 5.21 *(starting values)*

In Example 5.18 it is claimed that generating a value U^1 using the first order accurate forward Euler method and then computing using the midpoint method gives a method that is globally second order accurate. Check that this is true by implementing this approach in MATLAB for a simple ODE.