

3.9 Exercises

- 3.1. Show that the backward Euler method and the trapezoidal method are 0-stable.
- 3.2. To draw a circle of radius r on a graphics screen, one may proceed to evaluate pairs of values $x = r \cos \theta$, $y = r \sin \theta$ for a succession of values θ . But this is expensive. A cheaper method may be obtained by considering the ODE

$$\begin{aligned} \dot{x} &= -y, & x(0) &= r, \\ \dot{y} &= x, & y(0) &= 0, \end{aligned}$$

where $\dot{x} = \frac{dx}{d\theta}$, and approximating this using a simple discretization method. However, care must be taken so as to ensure that the obtained approximate solution looks right, i.e., that the approximate curve closes rather than spirals.

For each of the three discretization methods introduced in this chapter, namely, forward Euler, backward Euler, and trapezoidal methods, carry out this integration using a uniform step size $h = .02$ for $0 \leq \theta \leq 120$. Determine if the solution spirals in, spirals out, or forms an approximate circle as desired. Explain the observed results. [Hint: This has to do with a certain invariant function of x and y , rather than with the order of the methods.]

- 3.3. The following ODE system:

$$\begin{aligned} y_1' &= \alpha - y_1 - \frac{4y_1y_2}{1+y_1^2}, \\ y_2' &= \beta y_1 \left(1 - \frac{y_2}{1+y_1^2} \right), \end{aligned}$$

where α and β are parameters, represents a simplified approximation to a chemical reaction [92]. There is a parameter value $\beta_c = \frac{3\alpha}{5} - \frac{25}{\alpha}$ such that for $\beta > \beta_c$ solution trajectories decay in amplitude and spiral in phase space into a stable fixed point, whereas for $\beta < \beta_c$ trajectories oscillate without damping and are attracted to a stable limit cycle. [This is called a *Hopf bifurcation*.]

- (a) Set $\alpha = 10$ and use any of the discretization methods introduced in this chapter with a fixed step size $h = .01$ to approximate the solution starting at $y_1(0) = 0$, $y_2(0) = 2$, for $0 \leq t \leq 20$. Do this for the parameter values $\beta = 2$ and $\beta = 4$. For each case plot y_1 vs. t and y_2 vs. y_1 . Describe your observations.
- (b) Investigate the situation closer to the critical value $\beta_c = 3.5$. [You may have to increase the length of the integration interval b to get a better look.]
- 3.4. When deriving the trapezoidal method, we proceeded to replace $y'(t_{n-1/2})$ in (3.30) by an average and then use the ODE (3.1). If instead we first use the ODE, replacing $y'(t_{n-1/2})$ by $\mathbf{f}(t_{n-1/2}, \mathbf{y}(t_{n-1/2}))$, and then average \mathbf{y} , we obtain the implicit *midpoint method*,

$$\mathbf{y}_n = \mathbf{y}_{n-1} + h_n \mathbf{f} \left(t_{n-1/2}, \frac{1}{2} (\mathbf{y}_n + \mathbf{y}_{n-1}) \right). \quad (3.33)$$

- (a) Show that this method is symmetric, second-order, and A-stable. How does it relate to the trapezoidal method for the constant-coefficient ODE (3.18)?
- (b) Show that even if we allow λ to vary in t , i.e., we consider the scalar ODE

$$y' = \lambda(t)y$$

in place of the test equation, what corresponds to A-stability holds; namely, using the midpoint method,

$$|y_n| \leq |y_{n-1}| \quad \text{if} \quad \operatorname{Re}(\lambda) \leq 0$$

(this property is called *AN-stability* [24]). Show that the same cannot be said about the trapezoidal method: the latter is not AN-stable.

- 3.5. (a) Show that the trapezoidal step (3.32) can be viewed as a half-step of forward Euler followed by a half-step of backward Euler.
- (b) Show that the midpoint step (3.33) can be viewed as a half-step of backward Euler followed by a half-step of forward Euler.
- (c) Consider an autonomous system $y' = f(y)$ and a fixed step size, $h_n = h$, $n = 1, \dots, N$. Show that the trapezoidal method applied N times is equivalent to applying first a half-step of forward Euler (i.e., forward Euler with step size $h/2$), followed by $N-1$ midpoint steps, finishing off with a half-step of backward Euler. Conclude that these two symmetric methods are *dynamically equivalent* [34]; i.e., for h small enough their performance is very similar independently of N , even over a very long time: $b = Nh \gg 1$.
- (d) However, if h is not small enough (compared to the problem's small parameter, say λ^{-1}) then these methods do not necessarily perform similarly. Construct an example where one of these methods blows up (error $> 10^5$, say) while the other yields an error below 10^{-5} . [Do not program anything: this is a (nontrivial!) pen-and-paper question.]

- 3.6. Consider the method of lines applied to the simple heat equation in one space dimension,

$$u_t = au_{xx},$$

with $a > 0$ a constant, $u = 0$ at $x = 0$, $x = 1$ for $t \geq 0$, and $u(x, 0) = g(x)$ given as well. Formulate the method of lines, as in Example 1.3, to arrive at a system of the form (3.18) with A symmetric. Find the eigenvalues of A and show that, when using the forward Euler discretization for the time variable, the resulting method is stable if

$$h \leq \frac{1}{2a} \Delta x^2.$$

(This is a rather restrictive condition on the time step.) On the other hand, if we discretize in time using the trapezoidal method (the resulting method, second-order in both space and time, is called Crank-Nicolson) or the backward Euler method, then no stability restriction for the time step arises. [Hint: To find the eigenvalues, try eigenvectors v^k in the form $v_i^k = \sin(ik\pi\Delta x)$, $i = 1, \dots, m$, for $1 \leq k \leq m$.]

- 3.7. Consider the same question as the previous one, but this time the heat equation is in two space variables on a unit square,

$$u_t = a(u_{xx} + u_{yy}), \quad 0 \leq x, y \leq 1, \quad t \geq 0.$$

The boundary conditions are $u = 0$ around the square, and $u(x, y, 0) = g(x, y)$ is given as well.

Formulate a system (3.18) using a uniform grid with spacing Δx on the unit square. Conclude again that no restrictions on the time step arise when using the implicit methods which we have presented for time discretization. What happens with the forward Euler method? [Hint: Don't try this exercise before you have done the previous one.]

- 3.8. Consider the ODE

$$\frac{dy}{dt} = f(t, y), \quad 0 \leq t \leq b,$$

where $b \gg 1$.

- (a) Apply the *stretching* transformation $t = \tau b$ to obtain the equivalent ODE

$$\frac{dy}{d\tau} = b f(\tau b, y), \quad 0 \leq \tau \leq 1.$$

(Strictly speaking, y in these two ODEs is not quite the same function. Rather, it stands in each case for the unknown function.)

- (b) Show that applying any of the discretization methods in this chapter to the ODE in t with step size $h = \Delta t$ is equivalent to applying the same method to the ODE in τ with step size $\Delta\tau$ satisfying $\Delta t = b\Delta\tau$. In other words, the same stretching transformation can be equivalently applied to the discretized problem.

- 3.9. Write a short program which uses the forward Euler, the backward Euler, and the trapezoidal *or* midpoint methods to integrate a linear, scalar ODE with a known solution, using a fixed step size $h = b/N$, and which finds the maximum error. Apply your program to the following problem:

$$\frac{dy}{dt} = (\cos t)y, \quad 0 \leq t \leq b,$$

$y(0) = 1$. The exact solution is

$$y(t) = e^{\sin t}.$$

Verify those entries given in Table 3.2 and complete the missing ones. Make as many (useful) observations as you can on the results in the complete table. Attempt to provide explanations. [Hint: Plotting these solution curves for $b = 20$, $N = 10b$, say, may help.]

b	N	Forward Euler	Backward Euler	Trapezoidal	Midpoint
1	10	.35e-1	.36e-1	.29e-2	.22e-2
	20	.18e-1	.18e-1	.61e-3	.51e-3
10	100				
	200				
100	1000	2.46	25.90	.42e-2	.26e-2
	2000				
1000	1000				
	10000	2.72	1.79e+11	.42e-2	.26e-2
	20000				
	100000	2.49	29.77	.42e-4	.26e-4

Table 3.2: Maximum errors for long interval integration of $y' = (\cos t)y$.

- 3.10. Consider two linear harmonic oscillators (recall Example 2.6), one fast and one slow, $u_1'' = -\varepsilon^{-2}(u_1 - \bar{u}_1)$ and $u_2'' = -(u_2 - \bar{u}_2)$. The parameter is small: $0 < \varepsilon \ll 1$. We write this as a first-order system

$$\mathbf{u}' = \begin{pmatrix} \varepsilon^{-1} & 0 \\ 0 & 1 \end{pmatrix} \mathbf{v},$$

$$\mathbf{v}' = - \begin{pmatrix} \varepsilon^{-1} & 0 \\ 0 & 1 \end{pmatrix} (\mathbf{u} - \bar{\mathbf{u}}),$$

where $\mathbf{u}(t)$, $\mathbf{v}(t)$, and the given constant vector $\bar{\mathbf{u}}$ each have two components. It is easy to see that $E_F = \frac{1}{2\varepsilon}(v_1^2 + (u_1 - \bar{u}_1)^2)$ and $E_S = \frac{1}{2}(v_2^2 + (u_2 - \bar{u}_2)^2)$ remain constant for all t (see Section 2.5).

Next, we apply the following time-dependent linear transformation:

$$\mathbf{u} = Q\mathbf{x}, \quad \mathbf{v} = Q\mathbf{z}, \quad Q(t) = \begin{pmatrix} \cos \omega t & \sin \omega t \\ -\sin \omega t & \cos \omega t \end{pmatrix},$$

$$K = \dot{Q}^T Q = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

where $\omega \geq 0$ is another parameter. This yields the coupled system

$$\mathbf{x}' = Q^T \begin{pmatrix} \varepsilon^{-1} & 0 \\ 0 & 1 \end{pmatrix} Q\mathbf{z} + \omega K\mathbf{x}, \quad (3.34a)$$

$$\mathbf{z}' = -Q^T \begin{pmatrix} \varepsilon^{-1} & 0 \\ 0 & 1 \end{pmatrix} Q(\mathbf{x} - \bar{\mathbf{x}}) + \omega K\mathbf{z}, \quad (3.34b)$$

where $\bar{\mathbf{x}} = Q^T \bar{\mathbf{u}}$. We can write the latter system in our usual notation as a system of order 4,

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{q}(t).$$

- (a) Show that the eigenvalues of the matrix A are all purely imaginary for all ω . [Hint: Show that $A^T = -A$.]
- (b) Using the values $\bar{\mathbf{u}} = (1, \pi/4)^T$, $\mathbf{u}(0) = \bar{\mathbf{u}}$, $\mathbf{v}(0)^T = (1, -1)/\sqrt{2}$, and $b = 20$, apply the midpoint method with a constant step size h to the system (3.34) for the following parameter combinations: $\varepsilon = 0.001$, $k = 0.1, 0.05, 0.001$, and $\omega = 0, 1, 10$ (a total of nine runs). Compute the error indicators $\max_t |E_F(t) - E_F(0)|$ and $\max_t |E_S(t) - E_S(0)|$. Discuss your observations.
- (c) Attempt to show that the midpoint method is unstable for this problem if $h > 2\sqrt{\varepsilon/\omega}$ (see [10]). Conclude that A-stability and AN-stability do not automatically extend to ODE systems.

3.11. Consider the implicit ODE

$$M(\mathbf{y})\mathbf{y}' = \mathbf{f}(t, \mathbf{y}),$$

where $M(\mathbf{y})$ is nonsingular for all \mathbf{y} . The need to integrate IVPs of this type typically arises in robotics. When the system size m is large, the cost of inverting M may dominate the entire solution cost. Also, $\frac{\partial M}{\partial \mathbf{y}}$ is complicated to evaluate, but it is given that its norm is not large, say $O(1)$.

- (a) Extend the forward Euler and the backward Euler discretizations for this case (without inverting M). Justify.
- (b) Propose a method for solving the nonlinear system of equations resulting at each time step when using backward Euler, for the case where $\|\partial \mathbf{f} / \partial \mathbf{y}\|$ is very large.