ORDINARY DIFFERENTIAL EQUATIONS LAPLACE TRANSFORMS AND NUMERICAL METHODS FOR ENGINEERS

by

Steven J. DESJARDINS

and

Rémi VAILLANCOURT

Notes for the course MAT 2384 C

of Jason LEVY

Winter 2012

Département de mathématiques et de statistique
Department of Mathematics and Statistics
Université d'Ottawa / University of Ottawa
Ottawa, ON, Canada K1N 6N5

2011.04.01

LEVY, Jason
Department of Mathematics and Statistics
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
e-mail: jlevy@uottawa.ca

DESJARDINS, Steven J.
Department of Mathematics and Statistics
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
e-mail: sdesjar2@uottawa.ca
homepage: http://www.mathstat.uottawa.ca/~sdesj740

VAILLANCOURT, Rémi
Département de mathématiques et de statistique
Université d'Ottawa
Ottawa (Ontario), Canada K1N 6N5
courriel: remi@uottawa.ca
page d'accueil: http://www.site.uottawa.ca/~remi

The production of this book benefitted from grants from the Natural Sciences and Engineering Research Council of Canada.

© S. J. Desjardins and R. Vaillancourt, Ottawa 2011

Contents

Part 1. Differential Equations and Laplace Transforms 1

Chapter 1. First-Order Ordinary Differential Equations 3
1.1. Fundamental Concepts 3
1.2. Separable Equations 5
1.3. Equations with Homogeneous Coefficients 7
1.4. Exact Equations 9
1.5. Integrating Factors 16
1.6. First-Order Linear Equations 21
1.7. Orthogonal Families of Curves 23
1.8. Direction Fields and Approximate Solutions 26
1.9. Existence and Uniqueness of Solutions 26

Chapter 2. Second-Order Ordinary Differential Equations 33
2.1. Linear Homogeneous Equations 33
2.2. Homogeneous Equations with Constant Coefficients 33
2.3. Basis of the Solution Space 34
2.4. Independent Solutions 36
2.5. Modeling in Mechanics 39
2.6. Euler–Cauchy Equations 44

Chapter 3. Linear Differential Equations of Arbitrary Order 49
3.1. Homogeneous Equations 49
3.2. Linear Homogeneous Equations 55
3.3. Linear Nonhomogeneous Equations 59
3.4. Method of Undetermined Coefficients 61
3.5. Particular Solution by Variation of Parameters 65
3.6. Forced Oscillations 71

Chapter 4. Systems of Differential Equations 77
4.1. Introduction 77
4.2. Existence and Uniqueness Theorem 79
4.3. Fundamental Systems 80
4.4. Homogeneous Linear Systems with Constant Coefficients 83
4.5. Nonhomogeneous Linear Systems 91

Chapter 5. Laplace Transform 97
5.1. Definition 97
5.2. Transforms of Derivatives and Integrals 102
5.3. Shifts in s and in t 106
5.4. Dirac Delta Function 115

5.5. Derivatives and Integrals of Transformed Functions 117
5.6. Laguerre Differential Equation 120
5.7. Convolution 122
5.8. Partial Fractions 125
5.9. Transform of Periodic Functions 125

Chapter 6. Power Series Solutions 129
6.1. The Method 129
6.2. Foundation of the Power Series Method 131
6.3. Legendre Equation and Legendre Polynomials 139
6.4. Orthogonality Relations for $P_n(x)$ 142
6.5. Fourier–Legendre Series 145
6.6. Derivation of Gaussian Quadratures 148

Part 2. Numerical Methods 153

Chapter 7. Solutions of Nonlinear Equations 155
7.1. Computer Arithmetic 155
7.2. Review of Calculus 158
7.3. The Bisection Method 158
7.4. Fixed Point Iteration 162
7.5. Newton's, Secant, and False Position Methods 167
7.6. Aitken–Steffensen Accelerated Convergence 175
7.7. Horner's Method and the Synthetic Division 177
7.8. Müller's Method 179

Chapter 8. Interpolation and Extrapolation 183
8.1. Lagrange Interpolating Polynomial 183
8.2. Newton's Divided Difference Interpolating Polynomial 185
8.3. Gregory–Newton Forward-Difference Polynomial 189
8.4. Gregory–Newton Backward-Difference Polynomial 191
8.5. Hermite Interpolating Polynomial 192
8.6. Cubic Spline Interpolation 194

Chapter 9. Numerical Differentiation and Integration 197
9.1. Numerical Differentiation 197
9.2. The Effect of Roundoff and Truncation Errors 199
9.3. Richardson's Extrapolation 201
9.4. Basic Numerical Integration Rules 203
9.5. The Composite Midpoint Rule 206
9.6. The Composite Trapezoidal Rule 208
9.7. The Composite Simpson Rule 210
9.8. Romberg Integration for the Trapezoidal Rule 212
9.9. Adaptive Quadrature Methods 213
9.10. Gaussian Quadrature 215

Chapter 10. Numerical Solution of Differential Equations 217
10.1. Initial Value Problems 217
10.2. Euler's and Improved Euler's Methods 218
10.3. Low-Order Explicit Runge–Kutta Methods 221

10.4. Convergence of Numerical Methods 229
10.5. Absolutely Stable Numerical Methods 230
10.6. Stability of Runge–Kutta Methods 231
10.7. Embedded Pairs of Runge–Kutta Methods 234
10.8. Multistep Predictor-Corrector Methods 240
10.9. Stiff Systems of Differential Equations 252

Part 3. Exercises and Solutions 261

Chapter 11. Exercises for Differential Equations and Laplace Transforms 263
Exercises for Chapter 1 263
Exercises for Chapter 2 265
Exercises for Chapter 3 266
Exercises for Chapter 4 268
Exercises for Chapter 5 269
Exercises for Chapter 6 271

Chapter 12. Exercises for Numerical Methods 275
Exercises for Chapter 7 275
Exercises for Chapter 8 277
Exercises for Chapter 9 278
Exercises for Chapter 10 280

Solutions to Starred Exercises 283
Solutions to Exercises from Chapters 1 to 6 283
Solutions to Exercises from Chapter 7 292
Solutions to Exercises for Chapter 8 294
Solutions to Exercises for Chapter 10 295

Part 4. Formulas and Tables 301

Chapter 13. Formulas and Tables 303
13.1. Integrating Factor of $M(x, y)\,dx + N(x, y)\,dy = 0$ 303
13.2. Solution of First-Order Linear Differential Equations 303
13.3. Laguerre Polynomials on $0 \le x < \infty$ 303
13.4. Legendre Polynomials $P_n(x)$ on $[-1, 1]$ 304
13.5. Fourier–Legendre Series Expansion 305
13.6. Table of Integrals 306
13.7. Table of Laplace Transforms 306

Index 309

Part 1

Differential Equations and Laplace Transforms

CHAPTER 1

First-Order Ordinary Differential Equations

1.1. Fundamental Concepts

(a) A *differential equation* is an equation involving an unkonwn function y, derivatives of it and functions of the independent variable.

Here are three *ordinary differential equations*, where $' := \dfrac{d}{dx}$:

(1) $y' = \cos x$,
(2) $y'' + 4y = 0$,
(3) $x^2 y''' y' + 2\, e^x y'' = (x^2 + 2) y^2$.

Here is a *partial differential equation*:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

(b) The *order* of a differential equation is equal to the highest-order derivative that appears in it.

The above equations (1), (2) and (3) are of order 1, 2 and 3, respectively.

(c) An *explicit solution* of a differential equation with independent variable x on $]a, b[$ is a function $y = g(x)$ of x such that the differential equation becomes an identity in x on $]a, b[$ when $g(x)$, $g'(x)$, etc. are substituted for y, y', etc. in the differential equation. The solution $y = g(x)$ describes a curve, or trajectory, in the xy-plane.

We see that the function

$$y(x) = e^{2x}$$

is an explicit solution of the differential equation

$$\frac{dy}{dx} = 2y.$$

In fact, we have

$$\text{L.H.S.} := y'(x) = 2\, e^{2x},$$
$$\text{R.H.S.} := 2y(x) = 2\, e^{2x}.$$

Hence

$$\text{L.H.S.} = \text{R.H.S.}, \qquad \text{for all } x.$$

We thus have an identity in x on $]-\infty, \infty[$. □

(d) An *implicit solution* of a differential equation is a curve which is defined by an equation of the form $G(x, y) = c$ where c is an arbitrary constant.

Note that $G(x, y)$ represents a surface, a 2-dimensional object in 3-dimensional space where x and y are independent variables. By setting $G(x, y) = c$, a relationship is created between x and y.

We remark that an implicit solution always contains an equal sign, "=", followed by a constant, otherwise $z = G(x, y)$ represents a surface and not a curve.

We see that the curve in the xy-plane,

$$x^2 + y^2 - 1 = 0, \qquad y > 0,$$

is an implicit solution of the differential equation

$$yy' = -x, \qquad \text{on } -1 < x < 1.$$

In fact, letting y be a function of x and differentiating the equation of the curve with respect to x,

$$\frac{d}{dx}(x^2 + y^2 - 1) = \frac{d}{dx}(0) = 0,$$

we obtain

$$2x + 2yy' = 0 \quad \text{or} \quad yy' = -x. \quad \square$$

(e) The *general solution* of a differential equation of order n contains n arbitrary constants.

The one-parameter family of functions

$$y(x) = \sin x + c$$

is the general solution of the first-order differential equation

$$y'(x) = \cos x.$$

This infinite family of curves all have the same slope, and hence all members of this familiy are solutions of the differential equation. The *general solution* is written $y(x) = \sin x + c$ (with the arbitrary constant) to represent *all* of the possible solutions.

Putting $c = 1$, we have the *unique solution*,

$$y(x) = \sin x + 1,$$

which goes through the point $(0, 1)$ of $\mathbb{R}^2$. Given an arbitrary point (x_0, y_0) of the plane, there is one and only one curve of the family which goes through that point. (See Fig. 1.1(a)).

Similarly, we see that the one-parameter family of functions

$$y(x) = c\, e^x$$

is the general solution of the differential equation

$$y' = y.$$

Setting $c = -1$, we have the unique solution,

$$y(x) = -e^x,$$

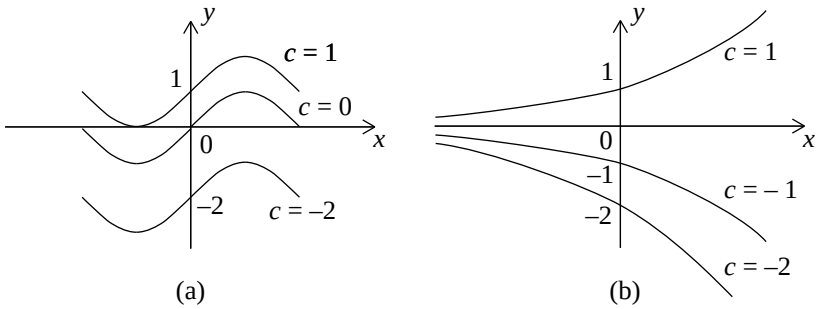

FIGURE 1.1. (a) Two one-parameter families of curves: (a) $y = \sin x + c$; (b) $y(x) = c\exp(x)$.

which goes through the point $(0, -1)$ of $\mathbb{R}^2$. Given an arbitrary point (x_0, y_0) of the plane, there is one and only one curve of the family which goes through that point. (See Fig. 1.1(b)).

1.2. Separable Equations

A differential equation is called *separable* if it can be written in the form

$$g(y)\,\frac{dy}{dx} = f(x). \tag{1.1}$$

We rewrite the equation using the differentials dy and dx and separate it by grouping on the left-hand side all terms containing y and on the right-hand side all terms containing x:

$$g(y)\,dy = f(x)\,dx. \tag{1.2}$$

The solution of a separated equation is obtained by taking the indefinite integral (primitive or antiderivative) of both sides and adding an arbitrary constant:

$$\int g(y)\,dy = \int f(x)\,dx + c, \tag{1.3}$$

that is

$$G(y) = F(x) + c.$$

Only one constant is needed and it is placed on the right-hand side (i.e. on the side with the independent variable). The two forms of the implicit solution,

$$G(y) = F(x) + c, \quad \text{or} \quad K(x, y) = -F(x) + G(y) = c,$$

define y as a function of x or x as a function of y.

Letting $y = y(x)$ be a function of x, we verify that (1.3) is a solution of (1.1):

$$\frac{d}{dx}(\text{LHS}) = \frac{d}{dx}G\left(y(x)\right) = G'\left(y(x)\right)y'(x) = g(y)y',$$
$$\frac{d}{dx}(\text{RHS}) = \frac{d}{dx}[F(x) + c] = F'(x) = f(x). \quad \square$$

EXAMPLE 1.1. Solve $y' = 1 + y^2$.

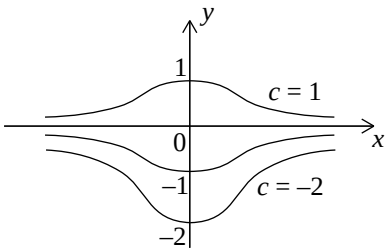

FIGURE 1.2. Three bell functions.

SOLUTION. Since the differential equation is separable, we have

$$\int \frac{dy}{1+y^2} = \int dx + c \implies \arctan y = x + c.$$

Thus

$$y(x) = \tan(x + c)$$

is a general solution, since it contains an arbitrary constant. □

When it is possible to solve for an explicit soltuion (i.e. solving for y), it should be done, as explicit functions are more convenient to work with.

EXAMPLE 1.2. Solve the initial value problem $y' = -2xy$, with $y(0) = y_0$.

SOLUTION. Since the differential equation is separable, the general solution is

$$\int \frac{dy}{y} = -\int 2x\,dx + c_1 \implies \ln|y| = -x^2 + c_1.$$

Taking the exponential of the solution, we have

$$y(x) = e^{-x^2+c_1} = e^{c_1}\, e^{-x^2}$$

which we rewrite in the form

$$y(x) = c\, e^{-x^2}.$$

Note that, since c_1 is an arbitrary constant, e^{c_1} is also an arbitrary constant, which can be denoted by c. We remark that the additive constant c_1 has become a multiplicative constant after exponentiation.

Figure 1.2 shows three bell functions which are members of the one-parameter family of the general solution.

Finally, the solution which satisfies the initial condition, is

$$y(x) = y_0\, e^{-x^2}.$$

This solution is unique (for each y_0). □

EXAMPLE 1.3. According to Newton's Law of Cooling, the rate of change of the temperature $T(t)$ of a body in a surrounding medium of temperature T_0 is proportional to the temperature difference $T(t) - T_0$,

$$\frac{dT}{dt} = -k(T - T_0).$$

Let a copper ball be immersed in a large basin containing a liquid whose constant temperature is 30 degrees. The initial temperature of the ball is 100 degrees. If, after 3 min, the ball's temperature is 70 degrees, when will it be 31 degrees?

SOLUTION. Since the differential equation is separable:

$$\frac{dT}{dt} = -k(T-30) \implies \frac{dT}{T-30} = -k\,dt,$$

then

$$\begin{aligned} \ln|T-30| &= -kt + c_1 && \text{(additive constant)} \\ T-30 &= e^{c_1-kt} = c\,e^{-kt} && \text{(multiplicative constant)} \\ T(t) &= 30 + c\,e^{-kt}. \end{aligned}$$

At $t = 0$,

$$100 = 30 + c \implies c = 70.$$

At $t = 3$,

$$70 = 30 + 70\,e^{-3k} \implies e^{-3k} = \frac{4}{7}.$$

When $T(t) = 31$,

$$31 = 70\left(e^{-3k}\right)^{t/3} + 30 \implies \left(e^{-3k}\right)^{t/3} = \frac{1}{70}.$$

Taking the logarithm of both sides, we have

$$\frac{t}{3}\ln\left(\frac{4}{7}\right) = \ln\left(\frac{1}{70}\right).$$

Hence

$$t = 3\,\frac{\ln(1/70)}{\ln(4/7)} = 3 \times \frac{-4.25}{-0.56} = 22.78\,\text{min} \quad \square$$

1.3. Equations with Homogeneous Coefficients

DEFINITION 1.1. A function $M(x,y)$ is said to be *homogeneous of degree s simultaneously in x and y* if

$$M(\lambda x, \lambda y) = \lambda^s M(x,y), \qquad \text{for all } x, y, \lambda. \tag{1.4}$$

Differential equations with homogeneous coefficients of the same degree are separable as follows.

THEOREM 1.1. *Consider a differential equation with homogeneous coefficients of degree s,*

$$M(x,y)dx + N(x,y)dy = 0. \tag{1.5}$$

Then either substitution $y = xu(x)$ or $x = yu(y)$ makes (1.5) separable.

PROOF. Letting

$$y = xu, \qquad dy = x\,du + u\,dx,$$

and substituting in (1.5), we have

$$\begin{aligned} M(x,xu)\,dx + N(x,xu)[x\,du + u\,dx] &= 0, \\ x^s M(1,u)\,dx + x^s N(1,u)[x\,du + u\,dx] &= 0, \\ [M(1,u) + uN(1,u)]\,dx + xN(1,u)\,du &= 0. \end{aligned}$$

This equation separates,

$$\frac{N(1,u)}{M(1,u)+uN(1,u)}\,du=-\frac{dx}{x}.$$

Note that the left-hand side is a function of u only.

The general solution of this equation is

$$\int\frac{N(1,u)}{M(1,u)+uN(1,u)}\,du=-\ln|x|+c. \quad \square$$

EXAMPLE 1.4. Solve $2xyy'-y^2+x^2=0$.

SOLUTION. We rewrite the equation in differential form:

$$(x^2-y^2)\,dx+2xy\,dy=0.$$

Since the coefficients are homogeneous functions of degree 2 in x and y, let

$$x=yu, \qquad dx=y\,du+u\,dy.$$

Substituting these expressions in the last equation we obtain

$$\begin{gathered}(y^2u^2-y^2)[y\,du+u\,dy]+2y^2u\,dy=0,\\ (u^2-1)[y\,du+u\,dy]+2u\,dy=0,\\ (u^2-1)y\,du+[(u^2-1)u+2u]\,dy=0,\\ \frac{u^2-1}{u(u^2+1)}\,du=-\frac{dy}{y}.\end{gathered}$$

Since integrating the left-hand side of this equation seems difficult (but can be done by Partial Fractions), let us restart with the substitution

$$y=xu, \qquad dy=x\,du+u\,dx.$$

Then,

$$\begin{gathered}(x^2-x^2u^2)\,dx+2x^2u[x\,du+u\,dx]=0,\\ [(1-u^2)+2u^2]\,dx+2ux\,du=0,\\ \int\frac{2u}{1+u^2}\,du=-\int\frac{dx}{x}+c_1.\end{gathered}$$

Integrating this last equation is easy:

$$\begin{gathered}\ln(u^2+1)=-\ln|x|+c_1,\\ \ln|x(u^2+1)|=c_1,\\ x\left[\left(\frac{y}{x}\right)^2+1\right]=e^{c_1}=c.\end{gathered}$$

The general solution is

$$y^2+x^2=cx.$$

Putting $c=2r$ in this formula and adding r^2 to both sides, we have

$$(x-r)^2+y^2=r^2.$$

The general solution describes a one-parameter family of circles with centre $(r,0)$ and radius $|r|$ (see Fig. 1.3).

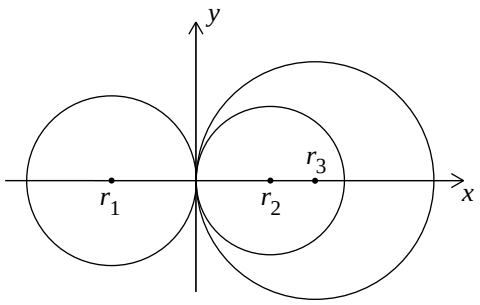

FIGURE 1.3. One-parameter family of circles with centre $(r, 0)$.

Since Theorem 1.1 states that both substitutions will make the differential equation separable, we can try one substitution and if it leads to difficult integrations, we can then try the other to see if the integrations are simpler. □

EXAMPLE 1.5. Solve the differential equation

$$y' = g\left(\frac{y}{x}\right).$$

SOLUTION. Rewriting this equation in differential form,

$$g\left(\frac{y}{x}\right)\,dx - dy = 0,$$

we see that this is an equation with homogeneous coefficients of degree zero in x and y. With the substitution

$$y = xu, \qquad dy = x\,du + u\,dx,$$

the last equation separates:

$$\begin{aligned} g(u)\,dx - x\,du - u\,dx &= 0,\\ x\,du &= [g(u) - u]\,dx,\\ \frac{du}{g(u) - u} &= \frac{dx}{x}. \end{aligned}$$

It can therefore be integrated directly,

$$\int \frac{du}{g(u) - u} = \int \frac{dx}{x} + c.$$

Finally one substitutes $u = y/x$ in the solution after the integration. □

1.4. Exact Equations

DEFINITION 1.2. The first-order differential equation

$$M(x, y)\,dx + N(x, y)\,dy = 0 \tag{1.6}$$

is called *exact* if its left-hand side is the total, or exact, differential

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy \tag{1.7}$$

of some function $u(x, y)$.

If equation (1.6) is exact, then

$$du = 0$$

and by integration we see that its general solution is

$$u(x, y) = c. \tag{1.8}$$

Comparing the expressions (1.6) and (1.7), we see that

$$\frac{\partial u}{\partial x} = M, \qquad \frac{\partial u}{\partial y} = N. \tag{1.9}$$

The following important theorem gives a necessary and sufficient condition for equation (1.6) to be exact.

THEOREM 1.2. *Let $M(x, y)$ and $N(x, y)$ be continuous functions with continuous first-order partial derivatives on a connected and simply connected (that is, of one single piece and without holes) set $\Omega \in \mathbb{R}^2$. Then the differential equation*

$$M(x, y)\,dx + N(x, y)\,dy = 0 \tag{1.10}$$

is exact if and only if

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}, \qquad \textit{for all } (x, y) \in \Omega. \tag{1.11}$$

PROOF. **Necessity:** Suppose (1.10) is exact. Then

$$\frac{\partial u}{\partial x} = M, \qquad \frac{\partial u}{\partial y} = N.$$

Therefore,

$$\frac{\partial M}{\partial y} = \frac{\partial^2 u}{\partial y \partial x} = \frac{\partial^2 u}{\partial x \partial y} = \frac{\partial N}{\partial x},$$

where exchanging the order of differentiation with respect to x and y is allowed by the continuity of the first and last terms.

Sufficiency: Suppose that (1.11) holds. We construct a function $F(x, y)$ such that

$$dF(x, y) = M(x, y)\,dx + N(x, y)\,dy.$$

Let the function $\varphi(x, y) \in C^2(\Omega)$ be such that

$$\frac{\partial \varphi}{\partial x} = M.$$

For example, we may take

$$\varphi(x, y) = \int M(x, y)\,dx, \qquad y \text{ fixed}.$$

Then,

$$\begin{aligned} \frac{\partial^2 \varphi}{\partial y \partial x} &= \frac{\partial M}{\partial y} \\ &= \frac{\partial N}{\partial x}, \qquad \text{by (1.11)}. \end{aligned}$$

Since

$$\frac{\partial^2 \varphi}{\partial y \partial x} = \frac{\partial^2 \varphi}{\partial x \partial y}$$

by the continuity of both sides, we have

$$\frac{\partial^2 \varphi}{\partial x \partial y} = \frac{\partial N}{\partial x}.$$

Integrating with respect to x, we obtain

$$\begin{aligned} \frac{\partial \varphi}{\partial y} &= \int \frac{\partial^2 \varphi}{\partial x \partial y}\, dx = \int \frac{\partial N}{\partial x}\, dx, \qquad y \text{ fixed}, \\ &= N(x, y) + B'(y). \end{aligned}$$

Taking

$$F(x, y) = \varphi(x, y) - B(y),$$

we have

$$\begin{aligned} dF &= \frac{\partial \varphi}{\partial x}\, dx + \frac{\partial \varphi}{\partial y}\, dy - B'(y)\, dy \\ &= M\, dx + N\, dy + B'(y)\, dy - B'(y)\, dy \\ &= M\, dx + N\, dy. \quad \square \end{aligned}$$

A **practical method** for solving exact differential equations will be illustrated by means of examples.

EXAMPLE 1.6. Find the general solution of

$$3x(xy - 2)\, dx + (x^3 + 2y)\, dy = 0,$$

and the solution that satisfies the initial condition $y(1) = -1$. Plot that solution for $1 \le x \le 4$.

SOLUTION. **(a) Analytic solution by the practical method.**— We verify that the equation is exact:

$$\begin{gathered} M = 3x^2 y - 6x, \qquad N = x^3 + 2y, \\ \frac{\partial M}{\partial y} = 3x^2, \qquad \frac{\partial N}{\partial x} = 3x^2, \\ \frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}. \end{gathered}$$

Indeed, it is exact and hence can be integrated. From

$$\frac{\partial u}{\partial x} = M,$$

we have

$$\begin{aligned} u(x, y) &= \int M(x, y)\, dx + T(y), \qquad y \text{ fixed}, \\ &= \int (3x^2 y - 6x)\, dx + T(y) \\ &= x^3 y - 3x^2 + T(y), \end{aligned}$$

and from

$$\frac{\partial u}{\partial y} = N,$$

we have

$$\begin{aligned}\frac{\partial u}{\partial y} &= \frac{\partial}{\partial y}\big(x^3y - 3x^2 + T(y)\big)\\ &= x^3 + T'(y) = N\\ &= x^3 + 2y.\end{aligned}$$

Thus

$$T'(y) = 2y.$$

It is essential that $T'(y)$ be a function of y only; otherwise there is an error somewhere: either the equation is not exact or there is a computational mistake.

We integrate $T'(y)$:

$$T(y) = y^2.$$

An integration constant is not needed at this stage since such a constant will appear in $u(x, y) = c$. Hence, we have the **surface**

$$u(x, y) = x^3y - 3x^2 + y^2.$$

Since $du = 0$, then $u(x, y) = c$, and the (implicit) general solution, containing *an arbitrary constant and an equal sign* "=" (that is, a **curve**), is

$$x^3y - 3x^2 + y^2 = c.$$

Using the initial condition $y(1) = -1$ to determine the value of the constant c, we put $x = 1$ and $y = -1$ in the general solution and get

$$c = -3.$$

Hence the implicit solution which satisfies the initial condition is

$$x^3y - 3x^2 + y^2 = -3.$$

(b) Solution by symbolic Matlab.— The general solution is:

```
>> y = dsolve('(x^3+2*y)*Dy=-3*x*(x*y-2)','x')
y =
[ -1/2*x^3+1/2*(x^6+12*x^2+4*C1)^(1/2)]
[ -1/2*x^3-1/2*(x^6+12*x^2+4*C1)^(1/2)]
```

The solution to the initial value problem is the lower branch with $C1 = -3$, as is seen by inserting the initial condition `'y(1)=-1',` in the preceding command:

```
>> y = dsolve('(x^3+2*y)*Dy=-3*x*(x*y-2)','y(1)=-1','x')
y = -1/2*x^3-1/2*(x^6+12*x^2-12)^(1/2)
```

(c) Solution to I.V.P. by numeric Matlab.— We use the initial condition $y(1) = -1$. The M-file `exp1_6.m` is

```
function yprime = exp1_6(x,y); %MAT 2384, Exp 1.6.
yprime = -3*x*(x*y-2)/(x^3+2*y);
```

The call to the `ode23` solver and the `plot` command are:

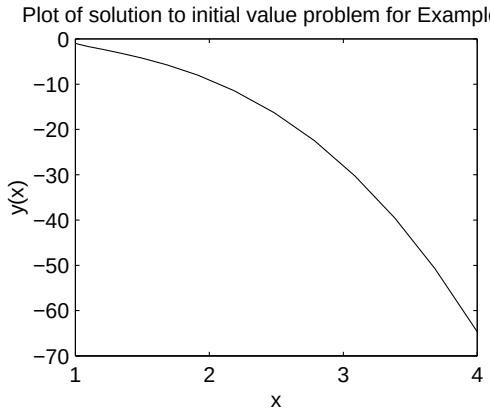

FIGURE 1.4. Graph of solution to Example 1.6.

```
>> xspan = [1 4]; % solution for x=1 to x=4
>> y0 = -1; % initial condition
>> [x,y] = ode23('exp1_6',xspan,y0);%Matlab 2007 format using xspan
>> subplot(2,2,1); plot(x,y);
>> title('Plot of solution to initial value problem for Example 1.6');
>> xlabel('x'); ylabel('y(x)');
>> print Fig.exp1.6
```

□

EXAMPLE 1.7. Find the general solution of

$$(2x^3 - xy^2 - 2y + 3)\,dx - (x^2y + 2x)\,dy = 0$$

and the solution that satisfies the initial condition $y(1) = -1$. Plot that solution for $1 \le x \le 4$.

SOLUTION. **(a) Analytic solution by the practical method.**— First, note the negative sign in $N(x, y) = -(x^2y + 2x)$. Since the left-hand side of the differential equation in standard form is $M\,dx + N\,dy$, the negative sign is part of the function $N(x, y)$. We verify that the equation is exact:

$$\frac{\partial M}{\partial y} = -2xy - 2, \qquad \frac{\partial N}{\partial x} = -2xy - 2,$$
$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}.$$

Hence the equation is exact and can be integrated. From

$$\frac{\partial u}{\partial y} = N,$$

we have

$$\begin{aligned} u(x,y) &= \int N(x,y)\,dy + T(x), \qquad x \text{ fixed}, \\ &= \int (-x^2y - 2x)\,dy + T(x) \\ &= -\frac{x^2y^2}{2} - 2xy + T(x), \end{aligned}$$

and from

$$\frac{\partial u}{\partial x} = M,$$

we have

$$\begin{aligned} \frac{\partial u}{\partial x} &= -xy^2 - 2y + T'(x) = M \\ &= 2x^3 - xy^2 - 2y + 3. \end{aligned}$$

Thus

$$T'(x) = 2x^3 + 3.$$

It is essential that $T'(x)$ be a function of x only; otherwise there is an error somewhere: either the equation is not exact or there is a computational mistake.

We integrate $T'(x)$:

$$T(x) = \frac{x^4}{2} + 3x.$$

An integration constant is not needed at this stage since such a constant will appear in $u(x,y) = c$. Hence, we have the **surface**

$$u(x,y) = -\frac{x^2y^2}{2} - 2xy + \frac{x^4}{2} + 3x.$$

Since $du = 0$, then $u(x,y) = c$, and the (implicit) general solution, containing *an arbitrary constant and an equal sign* "=" (that is, a **curve**), is

$$x^4 - x^2y^2 - 4xy + 6x = c.$$

Putting $x = 1$ and $y = -1$, we have

$$c = 10.$$

Hence the implicit solution which satisfies the initial condition is

$$x^4 - x^2y^2 - 4xy + 6x = 10.$$

(b) Solution by symbolic Matlab.— The general solution is:

```
>> y = dsolve('(x^2*y+2*x)*Dy=(2*x^3-x*y^2-2*y+3)','x')
y =
[ (-2-(4+6*x+x^4+2*C1)^(1/2))/x]
[ (-2+(4+6*x+x^4+2*C1)^(1/2))/x]
```

The solution to the initial value problem is the lower branch with $C1 = -5$,

```
>> y = dsolve('(x^2*y+2*x)*Dy=(2*x^3-x*y^2-2*y+3)','y(1)=-1','x')
y =(-2+(-6+6*x+x^4)^(1/2))/x
```

(c) Solution to I.V.P. by numeric Matlab.— We use the initial condition $y(1) = -1$. The M-file `exp1_7.m` is

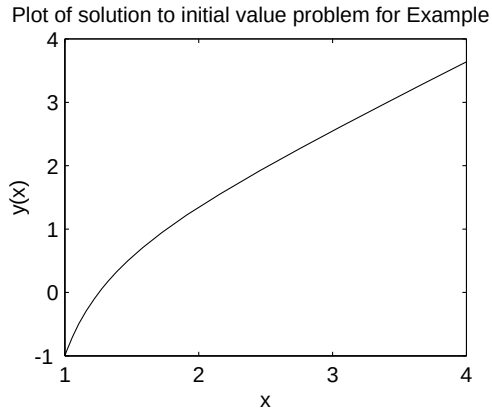

FIGURE 1.5. Graph of solution to Example 1.7.

```
function yprime = exp1_7(x,y); %MAT '2384, Exp 1.7.
yprime = (2*x^3-x*y^2-2*y+3)/(x^2*y+2*x);
```

The call to the `ode23` solver and the `plot` command:

```
>> xspan = [1 4]; % solution for x=1 to x=4
>> y0 = -1; % initial condition
>> [x,y] = ode23('exp1_7',xspan,y0);
>> subplot(2,2,1); plot(x,y);
>> print Fig.exp1.7
```

□

The following convenient notation for partial derivatives will often be used:

$$u_x(x,y) := \frac{\partial u}{\partial x}, \qquad u_y(x,y) := \frac{\partial u}{\partial y}.$$

The following example shows that the practical method of solution breaks down if the equation is not exact.

EXAMPLE 1.8. Solve

$$x\,dy - y\,dx = 0.$$

SOLUTION. We rewrite the equation in standard form:

$$y\,dx - x\,dy = 0.$$

The equation is not exact since

$$M_y = 1 \neq -1 = N_x.$$

Anyway, let us try to solve the inexact equation by the proposed method:

$$u(x,y) = \int u_x\,dx = \int M\,dx = \int y\,dx = yx + T(y),$$
$$u_y(x,y) = x + T'(y) = N = -x.$$

Thus,

$$T'(y) = -2x.$$

But this is impossible since $T(y)$ must be a function of y only. □

EXAMPLE 1.9. Consider the differential equation

$$(ax + by)\,dx + (kx + ly)\,dy = 0.$$

Choose a, b, k, l so that the equation is exact.

SOLUTION.

$$M_y = b, \quad N_x = k \implies k = b.$$

$$u(x,y) = \int u_x(x,y)\,dx = \int M\,dx = \int (ax + by)\,dx = \frac{ax^2}{2} + bxy + T(y),$$

$$u_y(x,y) = bx + T'(y) = N = bx + ly \implies T'(y) = ly \implies T(y) = \frac{ly^2}{2}.$$

Thus,

$$u(x,y) = \frac{ax^2}{2} + bxy + \frac{ly^2}{2}, \quad a, b, l \text{ arbitrary.}$$

The general solution is

$$\frac{ax^2}{2} + bxy + \frac{ly^2}{2} = c_1 \quad \text{or} \quad ax^2 + 2bxy + ly^2 = c. \quad \square$$

1.5. Integrating Factors

If the differential equation

$$M(x,y)\,dx + N(x,y)\,dy = 0 \tag{1.12}$$

is not exact, it can be made exact by multiplication by an integrating factor $\mu(x,y)$,

$$\mu(x,y)M(x,y)\,dx + \mu(x,y)N(x,y)\,dy = 0. \tag{1.13}$$

Rewriting this equation in the form

$$\widetilde{M}(x,y)\,dx + \widetilde{N}(x,y)\,dy = 0,$$

we have

$$\widetilde{M}_y = \mu_y M + \mu M_y, \qquad \widetilde{N}_x = \mu_x N + \mu N_x.$$

and equation (1.13) will be exact if

$$\mu_y M + \mu M_y = \mu_x N + \mu N_x. \tag{1.14}$$

In general, it is difficult to solve the partial differential equation (1.14).

We consider two particular cases, where μ is a function of one variable, that is, $\mu = \mu(x)$ or $\mu = \mu(y)$.

Case 1. If $\mu = \mu(x)$ is a function of x only, then $\mu_x = \mu'(x)$ and $\mu_y = 0$. Thus, (1.14) reduces to an ordinary differential equation:

$$N\mu'(x) = \mu(M_y - N_x). \tag{1.15}$$

If the left-hand side of the following expression

$$\frac{M_y - N_x}{N} = f(x) \tag{1.16}$$

is a function of x only, then (1.15) is separable:

$$\frac{d\mu}{\mu} = \frac{M_y - N_x}{N}\,dx = f(x)\,dx.$$

Integrating this separated equation, we obtain the integration factor

$$\mu(x) = e^{\int f(x)\,dx}. \tag{1.17}$$

Case 2. Similarly, if $\mu = \mu(y)$ is a function of y only, then $\mu_x = 0$ and $\mu_y = \mu'(y)$. Thus, (1.14) reduces to an ordinary differential equation:

$$M\mu'(y) = -\mu(M_y - N_x). \tag{1.18}$$

If the left-hand side of the following expression

$$\frac{M_y - N_x}{M} = g(y) \tag{1.19}$$

is a function of y only, then (1.18) is separable:

$$\frac{d\mu}{\mu} = -\frac{M_y - N_x}{M}\,dy = -g(y)\,dy.$$

Integrating this separated equation, we obtain the integration factor

$$\mu(y) = e^{-\int g(y)\,dy}. \tag{1.20}$$

One has to notice the presence of the negative sign in (1.20) and its absence in (1.17).

EXAMPLE 1.10. Find the general solution of the differential equation

$$(4xy + 3y^2 - x)\,dx + x(x + 2y)\,dy = 0.$$

SOLUTION. **(a) The analytic solution.**— This equation is not exact since

$$M_y = 4x + 6y, \qquad N_x = 2x + 2y$$

and

$$M_y \neq N_x.$$

However, since

$$\frac{M_y - N_x}{N} = \frac{2x + 4y}{x(x + 2y)} = \frac{2(x + 2y)}{x(x + 2y)} = \frac{2}{x} = f(x)$$

is a function of x only, we have the integrating factor

$$\mu(x) = e^{\int (2/x)\,dx} = e^{2\ln x} = e^{\ln x^2} = x^2.$$

Multiplying the differential equation by x^2 produces the exact equation

$$x^2(4xy + 3y^2 - x)\,dx + x^3(x + 2y)\,dy = 0.$$

The exactness of the differential equation should be verified at this point to ensure that the integrating factor is correct (otherwise any solution found cannot be correct).

$$\widetilde{M}(x, y) = 4x^3y + 3x^2y^2 - x^3, \qquad \widetilde{N}(x, y) = x^4 + 2x^3y,$$

$$\widetilde{M}_y = 4x^3 + 6x^2y, \qquad \widetilde{N}_x = 4x^3 + 6x^2y.$$

Thus, $\widetilde{M}_y = \widetilde{N}_x$ and the equation is certainly exact.

This equation is solved by the practical method:

$$\begin{aligned} u(x,y) &= \int (x^4 + 2x^3y)\,dy + T(x) \\ &= x^4y + x^3y^2 + T(x), \\ u_x(x,y) &= 4x^3y + 3x^2y^2 + T'(x) = \mu M \\ &= 4x^3y + 3x^2y^2 - x^3. \end{aligned}$$

Thus,

$$T'(x) = -x^3 \implies T(x) = -\frac{x^4}{4}.$$

No constant of integration is needed here; it will come later. Hence,

$$u(x,y) = x^4y + x^3y^2 - \frac{x^4}{4}$$

and the general solution is

$$x^4y + x^3y^2 - \frac{x^4}{4} = c_1 \quad \text{or} \quad 4x^4y + 4x^3y^2 - x^4 = c.$$

(b) The Matlab symbolic solution.— Matlab does not find the general solution of the nonexact equation:

```
>> y = dsolve('x*(x+2*y)*Dy=-(4*x+3*y^2-x)','x')
Warning: Explicit solution could not be found.
> In HD2:Matlab5.1:Toolbox:symbolic:dsolve.m at line 200
y = [ empty sym ]
```

but it solves the exact equation

```
>> y = dsolve('x^2*(x^3+2*y)*Dy=-3*x^3*(x*y-2)','x')
y =
[ -1/2*x^3-1/2*(x^6+12*x^2+4*C1)^(1/2)]
[ -1/2*x^3+1/2*(x^6+12*x^2+4*C1)^(1/2)]
```

□

EXAMPLE 1.11. Find the general solution of the differential equation

$$y(x + y + 1)\,dx + x(x + 3y + 2)\,dy = 0.$$

SOLUTION. **(a) The analytic solution.**— This equation is not exact since

$$M_y = x + 2y + 1 \neq N_x = 2x + 3y + 2.$$

Since

$$\frac{M_y - N_x}{N} = \frac{-x - y - 1}{x(x + 3y + 2)}$$

is not a function of x only, we try

$$\frac{M_y - N_x}{M} = \frac{-(x + y + 1)}{y(x + y + 1)} = -\frac{1}{y} = g(y),$$

which is a function of y only. The integrating factor is

$$\mu(y) = e^{-\int g(y)\,dy} = e^{\int (1/y)\,dy} = e^{\ln y} = y.$$

Multiplying the differential equation by y produces the exact equation, which should be verified before continuing,

$$(xy^2 + y^3 + y^2)\,dx + (x^2y + 3xy^2 + 2xy)\,dy = 0.$$

This equation is solved by the practical method:

$$\begin{aligned} u(x,y) &= \int (xy^2 + y^3 + y^2)\,dx + T(y) \\ &= \frac{x^2y^2}{2} + xy^3 + xy^2 + T(y), \\ u_y &= x^2y + 3xy^2 + 2xy + T'(y) = \mu N \\ &= x^2y + 3xy^2 + 2xy. \end{aligned}$$

Thus,

$$T'(y) = 0 \implies T(y) = 0$$

since no constant of integration is needed here. Hence,

$$u(x,y) = \frac{x^2y^2}{2} + xy^3 + xy^2$$

and the general solution is

$$\frac{x^2y^2}{2} + xy^3 + xy^2 = c_1 \quad \text{or} \quad x^2y^2 + 2xy^3 + 2xy^2 = c.$$

(b) The Matlab symbolic solution.— The symbolic Matlab command `dsolve` produces a very intricate general solution for both the nonexact and the exact equations. This solution does not simplify with the commands `simplify` and `simple`.

We therefore repeat the practical method having symbolic Matlab do the simple algebraic and calculus manipulations.

```
>> clear
>> syms M N x y u
>> M = y*(x+y+1); N = x*(x+3*y+2);
>> test = diff(M,'y') - diff(N,'x') % test for exactness
 test = -x-y-1 % equation is not exact
>> syms mu g
>> g = (diff(M,'y') - diff(N,'x'))/M
 g = (-x-y-1)/y/(x+y+1)
>> g =  simple(g)
 g = -1/y % a function of y only
>> mu = exp(-int(g,'y')) % integrating factor
 mu = y
>> syms MM NN
>> MM = mu*M; NN = mu*N; % multiply equation by integrating factor
>> u = int(MM,'x') % solution u; arbitrary T(y) not included yet
 u = y^2*(1/2*x^2+y*x+x)
>> syms DT
>> DT = simple(diff(u,'y') - NN)
 DT = 0 % T'(y) = 0 implies T(y) = 0.
>> u = u
 u = y^2*(1/2*x^2+y*x+x) % general solution u = c.
```

The general solution is

$$\frac{x^2y^2}{2} + xy^3 + xy^2 = c_1 \quad \text{or} \quad x^2y^2 + 2xy^3 + 2xy^2 = c. \quad \square$$

REMARK 1.1. Note that a separated equation,

$$f(x)\,dx + g(y)\,dy = 0,$$

is exact. In fact, since $M_y = 0$ and $N_x = 0$, we have the integrating factors

$$\mu(x) = e^{\int 0\,dx} = 1, \qquad \mu(y) = e^{-\int 0\,dy} = 1.$$

Solving this equation by the practical method for exact equations, we have

$$\begin{aligned} u(x,y) &= \int f(x)\,dx + T(y), \\ u_y = T'(y) = g(y) &\implies T(y) = \int g(y)\,dy, \\ u(x,y) &= \int f(x)\,dx + \int g(y)\,dy = c. \end{aligned}$$

This is the solution that was obtained by the earlier method (1.3).

REMARK 1.2. The factor which transforms a separable equation into a separated equation is an integrating factor since the latter equation is exact.

EXAMPLE 1.12. Consider the separable equation

$$y' = 1 + y^2, \qquad \text{that is,} \quad \left(1 + y^2\right)\,dx - dy = 0.$$

Show that the factor $\left(1 + y^2\right)^{-1}$ which separates the equation is an integrating factor.

SOLUTION. We have

$$M_y = 2y, \quad N_x = 0, \qquad \frac{2y - 0}{1 + y^2} = g(y).$$

Hence

$$\begin{aligned} \mu(y) &= e^{-\int (2y)/(1+y^2)\,dy} \\ &= e^{\ln\left[(1+y^2)^{-1}\right]} = \frac{1}{1+y^2}. \quad \square \end{aligned}$$

In the next example, we easily find an integrating factor $\mu(x,y)$ which is a function of x and y.

EXAMPLE 1.13. Consider the separable equation

$$y\,dx + x\,dy = 0.$$

Show that the factor

$$\mu(x,y) = \frac{1}{xy},$$

which makes the equation separable, is an integrating factor.

SOLUTION. The differential equation

$$\mu(x,y)y\,dx + \mu(x,y)x\,dy = \frac{1}{x}\,dx + \frac{1}{y}\,dy = 0$$

is separated; hence it is exact. $\square$

1.6. First-Order Linear Equations

Consider the nonhomogeneous first-order differential equation of the form

$$y' + f(x)y = r(x). \tag{1.21}$$

The left-hand side is a linear expression with respect to the dependent variable y and its first derivative y'. In this case, we say that (1.21) is a *linear* differential equation.

In this section, we solve equation (1.21) by transforming the left-hand side into a total derivative by means of an integrating factor. In Example 3.10, the general solution will be expressed as the sum of a general solution of the homogeneous equation (with right-hand side equal to zero) and a particular solution of the nonhomogeneous equation. Power Series solutions and numerical solutions will be considered in Chapters 6 and 11, respectively.

The first way is to rewrite (1.21) in differential form,

$$f(x)y\,dx + dy = r(x)\,dx, \quad \text{or} \quad \big(f(x)y - r(x)\big)\,dx + dy = 0, \tag{1.22}$$

and make it exact. Since $M_y = f(x)$ and $N_x = 0$, this equation is not exact. As

$$\frac{M_y - N_x}{N} = \frac{f(x) - 0}{1} = f(x)$$

is a function of x only, by (1.17) we have the integration factor

$$\mu(x) = e^{\int f(x)\,dx}.$$

Multiplying (1.21) by $\mu(x)$ makes the left-hand side an exact, or total, derivative. To see this, put

$$u(x, y) = \mu(x)y = e^{\int f(x)\,dx} y.$$

Taking the differential of u we have

$$\begin{aligned} du &= d\left[e^{\int f(x)\,dx} y\right] \\ &= e^{\int f(x)\,dx} f(x)y\,dx + e^{\int f(x)\,dx}\,dy \\ &= \mu[f(x)y\,dx + dy] \end{aligned}$$

which is the left-hand side of (1.21) multiplied by μ, as claimed. Hence

$$d\left[e^{\int f(x)\,dx} y(x)\right] = e^{\int f(x)\,dx} r(x)\,dx.$$

Integrating both sides with respect to x, we have

$$e^{\int f(x)\,dx} y(x) = \int e^{\int f(x)\,dx} r(x)\,dx + c.$$

Solving the last equation for $y(x)$, we see that the general solution of (1.21) is

$$y(x) = e^{-\int f(x)\,dx}\left[\int e^{\int f(x)\,dx} r(x)\,dx + c\right]. \tag{1.23}$$

It is extremely important to note that the arbitrary constant c is also multiplied by $e^{-\int f(x)}\,dx$.

Example 1.14. Solve the linear first-order differential equation

$$x^2 y' + 2xy = \sinh 3x.$$

SOLUTION. Rewriting this equation in standard form, we have

$$y' + \frac{2}{x}\, y = \frac{1}{x^2} \sinh 3x.$$

This equation is linear in y and y', with $f(x) = \frac{2}{x}$ and $r(x) = \frac{1}{x^2} \sinh 3x$. The integrating factor, which makes the left-hand side exact, is

$$\mu(x) = e^{\int (2/x)\, dx} = e^{\ln x^2} = x^2.$$

Thus,

$$\frac{d}{dx}(x^2 y) = \sinh 3x, \qquad \text{that is,} \quad d(x^2 y) = \sinh 3x\, dx.$$

Hence,

$$x^2 y(x) = \int \sinh 3x\, dx + c = \frac{1}{3} \cosh 3x + c,$$

or

$$y(x) = \frac{1}{3x^2} \cosh 3x + \frac{c}{x^2}. \quad \square$$

EXAMPLE 1.15. Solve the linear first-order differential equation

$$y\, dx + (3x - xy + 2)\, dy = 0.$$

SOLUTION. Rewriting this equation in standard form for the dependent variable $x(y)$, we have

$$\frac{dx}{dy} + \left(\frac{3}{y} - 1\right) x = -\frac{2}{y}, \qquad y \neq 0.$$

The integrating factor, which makes the left-hand side exact, is

$$\mu(y) = e^{\int [(3/y) - 1]\, dy} = e^{\ln y^3 - y} = y^3\, e^{-y}.$$

Then

$$d\left(y^3\, e^{-y} x\right) = -2y^2\, e^{-y}\, dy, \qquad \text{that is,} \quad \frac{d}{dy}\left(y^3\, e^{-y} x\right) = -2y^2\, e^{-y}.$$

Hence,

$$\begin{aligned} y^3\, e^{-y} x &= -2 \int y^2\, e^{-y}\, dy + c \\ &= 2y^2\, e^{-y} - 4 \int y e^{-y}\, dy + c \quad \text{(by integration by parts)} \\ &= 2y^2\, e^{-y} + 4y\, e^{-y} - 4 \int e^{-y}\, dy + c \\ &= 2y^2\, e^{-y} + 4y\, e^{-y} + 4\, e^{-y} + c. \end{aligned}$$

The general solution is

$$xy^3 = 2y^2 + 4y + 4 + c\, e^{y}. \quad \square$$

Some nonlinear differential equations can be reduced to a linear form via a change of variable. An example of this is the *Bernoulli equation*,

$$y' + p(x) y = g(x) y^a,$$

where a is (real) constant. Notice that the DE is linear if $a = 0$ or $a = 1$.

We make the substitution

$$u(x) = (y(x))^{1-a}.$$

Then

$$\frac{d}{dx}(u(x)) = u'(x) = \frac{d}{dx}((y(x))^{1-a} = (1-a)(y(x))^{-a}\,\frac{dy}{dx},$$

that is,

$$u' = (1-a)y^{-a}y'.$$

But

$$y' = gy^a - py$$

from the DE, so

$$\begin{aligned} u' &= (1-a)y^{-a}(gy^a - py) \\ &= (1-a)(g - py^{1-a}) \\ &= (1-a)(g - pu) \\ &= (1-a)g - (1-a)pu. \end{aligned}$$

And thus, we have the linear equation

$$u' + (1-a)p(x)u = (1-a)g(x)$$

which can be solved for u, after which we transform back to y.

As an example, we consider the nonlinear differential equation

$$y' + y = -\frac{x}{y} \quad \text{or} \quad y' + (1)y = (-x)y^{-1},$$

which is a Bernoulli equation with $p(x) = 1$, $g(x) = -x$ and $a = -1$. Letting $u(x) = (y(x))^2$, we have the linear DE

$$u' + 2u = -2x.$$

The integrating factor is

$$\mu(x) = e^{\int 2\,dx} = e^{2x}.$$

Thus

$$\begin{aligned} u(x) &= e^{-2x}\left[\int e^{2x}(-2x)\,dx + c\right] \\ &= e^{-2x}\left[-x\,e^{2x} + \frac{1}{2}\,e^{2x} + c\right] \qquad \text{(by integration by parts)} \\ &= -x + \frac{1}{2} + c\,e^{-2x}. \end{aligned}$$

Therefore, the general solution of the original DE is

$$y^2 = \frac{1}{2} - x + c\,e^{-2x}.$$

1.7. Orthogonal Families of Curves

A one-parameter family of curves can be given by an equation

$$u(x, y) = c,$$

where the parameter c is explicit, or by an equation

$$F(x, y, c) = 0,$$

which is implicit with respect to c.

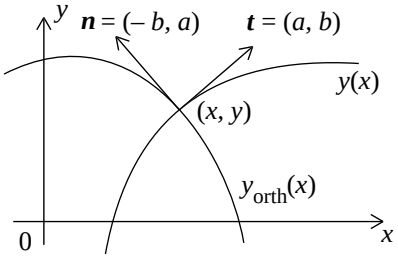

FIGURE 1.6. Two curves orthogonal at the point (x, y).

In the first case, the curves satisfy the differential equation

$$u_x\,dx + u_y\,dy = 0, \quad \text{or} \quad \frac{dy}{dx} = -\frac{u_x}{u_y} = m,$$

where m is the slope of the curve at the point (x, y). Note that this differential equation does not contain the parameter c.

In the second case we have

$$F_x(x, y, c)\,dx + F_y(x, y, c)\,dy = 0.$$

To eliminate c from this differential equation we solve the equation $F(x, y, c) = 0$ for c as a function of x and y,

$$c = H(x, y),$$

and substitute this function in the differential equation,

$$\frac{dy}{dx} = -\frac{F_x(x, y, c)}{F_y(x, y, c)} = -\frac{F_x\big(x, y, H(x, y)\big)}{F_y\big(x, y, H(x, y)\big)} = m.$$

Let $\boldsymbol{t} = (a, b)$ be the tangent and $\boldsymbol{n} = (-b, a)$ be the normal to the given curve $y = y(x)$ at the point (x, y) of the curve. Then, the slope, $y'(x)$, of the tangent is

$$y'(x) = \frac{b}{a} = m \tag{1.24}$$

and the slope, $y'_{\text{orth}}(x)$, of the curve $y_{\text{orth}}(x)$ which is orthogonal to the curve $y(x)$ at (x, y) is

$$y'_{\text{orth}}(x) = -\frac{a}{b} = -\frac{1}{m}. \tag{1.25}$$

(see Fig. 1.6). Thus, the orthogonal family satisfies the differential equation

$$y'_{\text{orth}}(x) = -\frac{1}{m(x)}.$$

EXAMPLE 1.16. Consider the one-parameter family of circles

$$x^2 + (y - c)^2 = c^2 \tag{1.26}$$

with centre $(0, c)$ on the y-axis and radius $|c|$. Find the differential equation for this family and the differential equation for the orthogonal family. Solve the latter equation and plot a few curves of both families on the same graph.

SOLUTION. We obtain the differential equation of the given family by differentiating (1.26) with respect to x,

$$2x + 2(y-c)y' = 0 \implies y' = -\frac{x}{y-c},$$

and solving (1.26) for c we have

$$x^2 + y^2 - 2yc + c^2 = c^2 \implies c = \frac{x^2+y^2}{2y}.$$

Substituting this value for c in the differential equation, we have

$$y' = -\frac{x}{y - \frac{x^2+y^2}{2y}} = -\frac{2xy}{2y^2 - x^2 - y^2} = \frac{2xy}{x^2 - y^2}.$$

The differential equation of the orthogonal family is

$$y'_{\text{orth}} = -\frac{x^2 - y^2_{\text{orth}}}{2xy_{\text{orth}}}.$$

Rewriting this equation in differential form $M\,dx + N\,dy = 0$, and omitting the subscript "orth", we have

$$(x^2 - y^2)\,dx + 2xy\,dy = 0.$$

Since $M_y = -2y$ and $N_x = 2y$, this equation is not exact, but

$$\frac{M_y - N_x}{N} = \frac{-2y - 2y}{2xy} = -\frac{2}{x} = f(x)$$

is a function of x only. Hence

$$\mu(x) = e^{-\int (2/x)\,dx} = x^{-2}$$

is an integrating factor. We multiply the differential equation by $\mu(x)$,

$$\left(1 - \frac{y^2}{x^2}\right)\,dx + 2\,\frac{y}{x}\,dy = 0,$$

and solve by the practical method:

$$u(x,y) = \int 2\,\frac{y}{x}\,dy + T(x) = \frac{y^2}{x} + T(x),$$

$$u_x(x,y) = -\frac{y^2}{x^2} + T'(x) = 1 - \frac{y^2}{x^2},$$

$$T'(x) = 1 \implies T(x) = x,$$

$$u(x,y) = \frac{y^2}{x} + x = c_1,$$

that is, the solution

$$x^2 + y^2 = c_1 x$$

is a one-parameter family of circles. We may rewrite this equation in the more explicit form:

$$x^2 - 2\,\frac{c_1}{2}\,x + \frac{c_1^2}{4} + y^2 = \frac{c_1^2}{4},$$

$$\left(x - \frac{c_1}{2}\right)^2 + y^2 = \left(\frac{c_1}{2}\right)^2,$$

$$(x-k)^2 + y^2 = k^2.$$

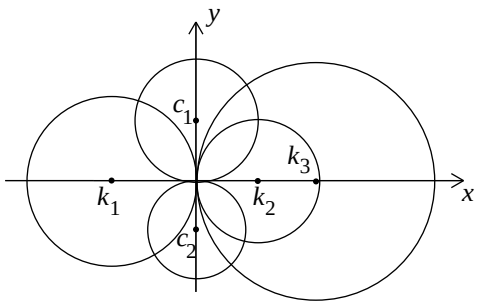

FIGURE 1.7. A few curves of both orthogonal families.

The orthogonal family is a family of circles with centre $(k, 0)$ on the x-axis and radius $|k|$. A few curves of both orthogonal families are plotted in Fig. 1.7. □

1.8. Direction Fields and Approximate Solutions

Approximate solutions of a differential equation are of practical interest if the equation has no explicit exact solution formula or if that formula is too complicated to be of practical value. In that case, one can use a numerical method (see Chapter 11), or use the Method of Direction Fields. By this latter method, one can sketch many solution curves at the same time, without actually solving the equation.

The Method of Direction Fields can be applied to any differential equation of the form

$$y' = f(x, y). \tag{1.27}$$

The idea is to take y' as the slope of the unknown solution curve. The curve that passes through the point (x_0, y_0) has the slope $f(x_0, y_0)$ at that point. Hence one can draw *lineal elements* at various points, that is, short segments indicating the tangent directions of solution curves as determined by (1.27) and then fit solution curves through this field of tangent directions.

First draw curves of constant slopes, $f(x, y) = \text{const}$, called *isoclines*. Second, draw along each isocline $f(x, y) = k$ many lineal elements of slope k. Thus one gets a direction field. Third, sketch approximate solutions curves of (1.27).

EXAMPLE 1.17. Graph the direction field of the first-order differential equation

$$y' = xy \tag{1.28}$$

and an approximation to the solution curve through the point $(1, 2)$.

SOLUTION. The isoclines are the equilateral hyperbolae $xy = k$ together with the two coordinate axes as shown in Fig. 1.8 □

1.9. Existence and Uniqueness of Solutions

DEFINITION 1.3. A function $f(y)$ is said to be *Lipschitz continuous* on the open interval $]c, d[$ if there exists a constant $M > 0$, called the Lipschitz constant, such that

$$|f(z) - f(y)| \leq M|z - y|, \qquad \text{for all} \quad y, z \in]c, d[. \tag{1.29}$$

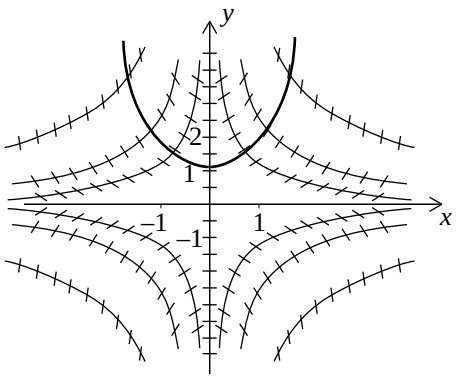

FIGURE 1.8. Direction field for Example 1.17.

We note that condition (1.29) implies the existence of left and right derivatives of $f(y)$ of the first order, but not their equality. Geometrically, the slope of the curve $f(y)$ is bounded on $]c, d[$.

We state, without proof, the following Existence and Uniqueness Theorem.

THEOREM 1.3 (Existence and Uniqueness Theorem). *Consider the initial value problem*

$$y' = f(x, y), \qquad y(x_0) = y_0. \tag{1.30}$$

If the function $f(x, y)$ is continuous and bounded,

$$|f(x, y)| \leq K,$$

on the rectangle

$$R: \quad |x - x_0| < a, \quad |y - y_0| < b,$$

and Lipschitz continuous with respect to y on R, then (1.30) *admits one and only one solution for all x such that*

$$|x - x_0| < \alpha, \quad \textit{where} \quad \alpha = \min\{a, b/K\}.$$

Theorem 1.3 is applied to the following example.

EXAMPLE 1.18. Solve the initial value problem

$$yy' + x = 0, \qquad y(0) = -2$$

and plot the solution.

SOLUTION. **(a) The analytic solution.**— We rewrite the differential equation in standard form, that is, $y' = f(x, y)$,

$$y' = -\frac{x}{y}.$$

Since the function

$$f(x, y) = -\frac{x}{y}$$

is not continuous at $y = 0$, there will be a solution for $y < 0$ and another solution for $y > 0$. We separate the equation and integrate:

$$\int x\,dx + \int y\,dy = c_1,$$
$$\frac{x^2}{2} + \frac{y^2}{2} = c_1,$$
$$x^2 + y^2 = r^2.$$

The general solution is a one-parameter family of circles with centre at the origin and radius r. The two solutions are

$$y_\pm(x) = \begin{cases} \sqrt{r^2 - x^2}, & \text{if } y > 0, \\ -\sqrt{r^2 - x^2}, & \text{if } y < 0. \end{cases}$$

Since $y(0) = -2$, we need to take the second solution. We determine the value of r by means of the initial condition:

$$0^2 + (-2)^2 = r^2 \implies r = 2.$$

Hence the solution, which is unique, is

$$y(x) = -\sqrt{4 - x^2}, \qquad -2 < x < 2.$$

We see that the slope $y'(x)$ of the solution tends to $\pm\infty$ as $y \to 0\pm$. To have a continuous solution in a neighbourhood of $y = 0$, we solve for $x = x(y)$.

(b) The Matlab symbolic solution.—

```
dsolve('y*Dy=-x','y(0)=-2','x')
y = -(-x^2+4)^(1/2)
```

(c) The Matlab numeric solution.— The numerical solution of this initial value problem is a little tricky because the general solution $y_\pm$ has two branches. We need a function M-file to run the Matlab ode solver. The M-file `halfcircle.m` is

```
function yprime = halfcircle(x,y);
yprime = -x/y;
```

To handle the lower branch of the general solution, we call the `ode23` solver and the `plot` command as follows.

```
xspan1 = [0 -2]; % span from x = 0 to x = -2
xspan2 = [0 2]; % span from x = 0 to x = 2
y0 = [0; -2]; % initial condition
[x1,y1] = ode23('halfcircle',xspan1,y0);
[x2,y2] = ode23('halfcircle',xspan2,y0);
plot(x1,y1(:,2),x2,y2(:,2))
axis('equal')
xlabel('x')
ylabel('y')
title('Plot of solution')
```

The numerical solution is plotted in Fig. 1.9. □

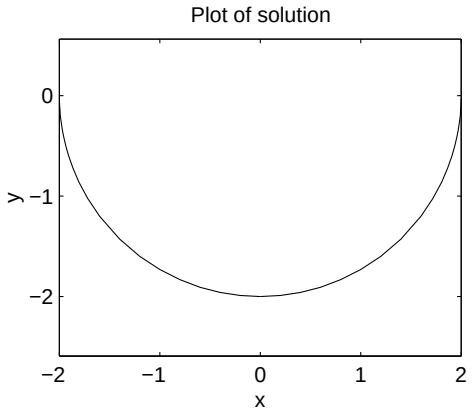

FIGURE 1.9. Graph of solution of the differential equation in Example 1.18.

In the following two examples, we find an approximate solution to a differential equation by Picard's method and by the method of Section 1.6. In Example 6.4, we shall find a series solution to the same equation. One will notice that the three methods produce the same series solution. Also, in Example 10.9, we shall solve this equation numerically.

EXAMPLE 1.19. Use Picard's recursive method to solve the initial value problem

$$y' = xy + 1, \qquad y(0) = 1.$$

SOLUTION. Since the function $f(x, y) = 1 + xy$ has a bounded partial derivative of first-order with respect to y,

$$\partial_y f(x, y) = x,$$

on any bounded interval $0 \le x \le a < \infty$, Picard's recursive formula (1.31),

$$y_n(x) = y_0 + \int_{x_0}^{x} f\big(t, y_{n-1}(t)\big)\, dt, \qquad n = 1, 2, \ldots,$$

converges to the solution $y(x)$. Here $x_0 = 0$ and $y_0 = 1$. Hence,

$$\begin{aligned}
y_1(x) &= 1 + \int_0^x (1 + t)\, dt \\
&= 1 + x + \frac{x^2}{2}, \\
y_2(x) &= 1 + \int_0^x \left(1 + t + t^2 + \frac{t^3}{2}\right) dt \\
&= 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8}, \\
y_3(x) &= 1 + \int_0^x \big(1 + t y_2(t)\big)\, dt,
\end{aligned}$$

and so on. □

EXAMPLE 1.20. Use the method of Section 1.6 for linear first-order differential equations to solve the initial value problem

$$y' - xy = 1, \qquad y(0) = 1.$$

SOLUTION. An integrating factor that makes the left-hand side an exact derivative is

$$\mu(x) = e^{-\int x\,dx} = e^{-x^2/2}.$$

Multiplying the equation by $\mu(x)$, we have

$$\frac{d}{dx}\left(e^{-x^2/2}y\right) = e^{-x^2/2},$$

and integrating from 0 to x, we obtain

$$e^{-x^2/2}y(x) = \int_0^x e^{-t^2/2}\,dt + c.$$

Putting $x = 0$ and $y(0) = 1$, we see that $c = 1$. Hence,

$$y(x) = e^{x^2/2}\left[1 + \int_0^x e^{-t^2/2}\,dt\right].$$

Since the integral cannot be expressed in closed form, we expand the two exponential functions in convergent power series, integrate the second series term by term and multiply the resulting series term by term:

$$\begin{aligned} y(x) &= e^{x^2/2}\left[1 + \int_0^x \left(1 - \frac{t^2}{2} + \frac{t^4}{8} - \frac{t^6}{48} + - \dots\right) dt\right] \\ &= e^{x^2/2}\left(1 + x - \frac{x^3}{6} + \frac{x^5}{40} - \frac{x^7}{336} - + \dots\right) \\ &= \left(1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{48} + \dots\right)\left(1 + x - \frac{x^3}{6} + \frac{x^5}{40} - \frac{x^7}{336} - + \dots\right) \\ &= 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} + \dots. \end{aligned}$$

As expected, the symbolic Matlab command `dsolve` produces the solution in terms of the Maple error function `erf(x)`:

```
>> dsolve('Dy=x*y+1','y(0)=1','x')
y=1/2*exp(1/2*x^2)*pi^(1/2)*2^(1/2)*erf(1/2*2^(1/2)*x)+exp(1/2*x^2)
```

□

Under the conditions of Theorem 1.3, the solution of problem (1.30) can be obtained by means of Picard's method, that is, the sequence $y_0, y_1, \dots, y_n, \dots$, defined by the Picard iteration formula,

$$y_n(x) = y_0 + \int_{x_0}^x f\big(t, y_{n-1}(t)\big)\,dt, \qquad n = 1, 2, \dots, \tag{1.31}$$

converges to the solution $y(x)$.

The following example shows that with continuity, but without Lispschitz continuity of the function $f(x, y)$ in $y' = f(x, y)$, the solution may not be unique.

EXAMPLE 1.21. Show that the initial value problem

$$y' = 3y^{2/3}, \qquad y(x_0) = y_0,$$

has non-unique solutions.

SOLUTION. The right-hand side of the equation is continuous for all y and because it is independent of x, it is continuous on the whole xy-plane. However, it is not Lipschitz continuous in y at $y = 0$ since $f_y(x, y) = 2y^{-1/3}$ is not even defined at $y = 0$. It is seen that $y(x) \equiv 0$ is a solution of the differential equation. Moreover, for $a \leq b$,

$$y(x) = \begin{cases} (x-a)^3, & t < a, \\ 0, & a \leq x \leq b, \\ (x-b)^3, & t > b, \end{cases}$$

is also a solution. By properly choosing the value of the parameter a or b, a solution curve can be made to satisfy the initial conditions. By varying the other parameter, one gets a family of solutions to the initial value problem. Hence the solution is not unique. □

CHAPTER 2

Second-Order Ordinary Differential Equations

In this chapter, we introduce basic concepts for linear second-order differential equations. We solve linear constant coefficients equations and Euler–Cauchy equations. Further theory on linear nonhomogeneous equations of arbitrary order will be developed in Chapter 3.

2.1. Linear Homogeneous Equations

Consider the *second-order linear nonhomogeneous differential equation*

$$y'' + f(x)y' + g(x)y = r(x). \tag{2.1}$$

The equation is linear with respect to y, y' and y''. It is *nonhomogeneous* if the right-hand side, $r(x)$, is not identically zero, i.e. $r(x) \not\equiv 0$.

The capital letter L will often be used to denote a linear differential operator of the form

$$L := a_n(x)D^n + a_{n-1}(x)D^{n-1} + \cdots + a_1(x)D + a_0(x), \qquad D = \ ' \ = \frac{d}{dx}.$$

Specifically, let $L := D^2 + f(x)D + g(x)$.

If the right-hand side of (2.1) is zero, we say that the equation

$$Ly := y'' + f(x)y' + g(x)y = 0, \tag{2.2}$$

is *homogeneous*.

THEOREM 2.1. *The solutions of* (2.2) *form a vector space.*

PROOF. We shall demonstrate that the space of solutions is closed under linear combination. Let y_1 and y_2 be two solutions of (2.2). The result follows from the linearity of L:

$$L(\alpha y_1 + \beta y_2) = \alpha L y_1 + \beta L y_2 = 0, \qquad \alpha, \beta \in \mathbb{R}. \quad \square$$

Since the general solution of a differential equation must represent all possible solutions and since the solutions of (2.2) form a vector space, the general solution of (2.2) must span the vector space of solutions.

2.2. Homogeneous Equations with Constant Coefficients

Consider the second-order linear homogeneous differential equation with *constant coefficients*:

$$y'' + ay' + by = 0. \tag{2.3}$$

What kind of functions would satisfy this equation? Notice that the differential equation requires that a linear combnation of y, y' and y'' be equal to zero for all x. This suggests that y, y' and y'' must be all the same kind of function.

To solve this equation we suppose that a solution is of exponential form,

$$y(x) = e^{\lambda x}.$$

Inserting this function in (2.3), we have

$$\lambda^2 e^{\lambda x} + a\lambda e^{\lambda x} + b e^{\lambda x} = 0, \tag{2.4}$$

$$e^{\lambda x}\left(\lambda^2 + a\lambda + b\right) = 0. \tag{2.5}$$

Since $e^{\lambda x}$ is never zero, we obtain the *characteristic equation*

$$\lambda^2 + a\lambda + b = 0 \tag{2.6}$$

for λ and the *eigenvalues* or roots

$$\lambda_{1,2} = \frac{-a \pm \sqrt{a^2 - 4b}}{2}. \tag{2.7}$$

If $\lambda_1 \neq \lambda_2$, we have two distinct solutions,

$$y_1(x) = e^{\lambda_1 x}, \qquad y_2(x) = e^{\lambda_2 x}.$$

In this case, the general solution, which contains two arbitrary constants, is

$$y = c_1 y_1 + c_2 y_2.$$

EXAMPLE 2.1. Find the general solution of the linear homogeneous differential equation with constant coefficients

$$y'' + 5y' + 6y = 0.$$

SOLUTION. The characteristic equation is

$$\lambda^2 + 5\lambda + 6 = (\lambda + 2)(\lambda + 3) = 0.$$

Hence $\lambda_1 = -2$ and $\lambda_2 = -3$, and the general solution is

$$y(x) = c_1 e^{-2x} + c_2 e^{-3x}. \quad \square$$

2.3. Basis of the Solution Space

We generalize to functions the notion of linear independence for two vectors of $\mathbb{R}^n$.

DEFINITION 2.1. The functions $f_1(x)$ and $f_2(x)$ are said to be *linearly independent* on the interval $[a, b]$ if the identity

$$c_1 f_1(x) + c_2 f_2(x) \equiv 0, \qquad \text{for all } x \in [a, b], \tag{2.8}$$

implies that

$$c_1 = c_2 = 0.$$

Otherwise the functions are said to be *linearly dependent*.

If $f_1(x)$ and $f_2(x)$ are linearly dependent on $[a, b]$, then there exist two numbers $(c_1, c_2) \neq (0, 0)$ such that, if, say, $c_1 \neq 0$, we have

$$\frac{f_1(x)}{f_2(x)} \equiv -\frac{c_2}{c_1} = \text{const.} \tag{2.9}$$

If

$$\frac{f_1(x)}{f_2(x)} \neq \text{const.} \qquad \text{on } [a, b], \tag{2.10}$$

then f_1 and f_2 are linearly independent on $[a, b]$. This characterization of linear independence of two functions will often be used.

DEFINITION 2.2. The general solution of the homogeneous equation (2.2) spans the vector space of solutions of (2.2).

THEOREM 2.2. *Let $y_1(x)$ and $y_2(x)$ be two solutions of* (2.2) *on* $[a, b]$. *Then, the solution*

$$y(x) = c_1 y_1(x) + c_2 y_2(x)$$

is a general solution of (2.2) *if and only if y_1 and y_2 are linearly independent on* $[a, b]$.

PROOF. The proof will be given in Chapter 3 for equations of order n. □

Notice that we are saying that the dimension of the vector space of solutions is 2. This follows from the fact that the characteristic equation is a quadratic and hence has 2 roots.

The next example illustrates the use of the general solution.

EXAMPLE 2.2. Solve the following initial value problem:

$$y'' + y' - 2y = 0, \qquad y(0) = 4, \quad y'(0) = 1.$$

SOLUTION. **(a) The analytic solution.**— The characteristic equation is

$$\lambda^2 + \lambda - 2 = (\lambda - 1)(\lambda + 2) = 0.$$

Hence $\lambda_1 = 1$ and $\lambda_2 = -2$. The two solutions

$$y_1(x) = e^x, \qquad y_2(x) = e^{-2x}$$

are linearly independent since

$$\frac{y_1(x)}{y_2(x)} = e^{3x} \neq \text{const.}$$

Thus, the general solution is

$$y = c_1\, e^x + c_2\, e^{-2x}.$$

The constants are determined by the initial conditions,

$$\begin{aligned} y(0) &= c_1 + c_2 = 4, \\ y'(x) &= c_1\, e^x - 2c_2\, e^{-2x}, \\ y'(0) &= c_1 - 2c_2 = 1. \end{aligned}$$

We therefore have the linear system

$$\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad \text{that is,} \quad A\boldsymbol{c} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}.$$

Since

$$\det A = -3 \neq 0,$$

the solution $\boldsymbol{c}$ is unique. This solution is easily computed by Cramer's rule,

$$c_1 = \frac{1}{-3} \begin{vmatrix} 4 & 1 \\ 1 & -2 \end{vmatrix} = \frac{-9}{-3} = 3, \qquad c_2 = \frac{1}{-3} \begin{vmatrix} 1 & 4 \\ 1 & 1 \end{vmatrix} = \frac{-3}{-3} = 1.$$

The solution of the initial value problem is

$$y(x) = 3\, e^x + e^{-2x}.$$

This solution is unique.

(b) The Matlab symbolic solution.—

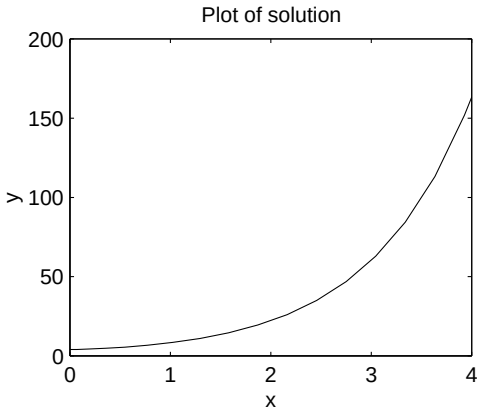

FIGURE 2.1. Graph of solution of the linear equation in Example 2.2.

```
dsolve('D2y+Dy-2*y=0','y(0)=4','Dy(0)=1','x')
y = 3*exp(x)+exp(-2*x)
```

(c) The Matlab numeric solution.— To rewrite the second-order differential equation as a system of first-order equations, we put

$$y_1 = y,$$
$$y_2 = y',$$

Thus, we have

$$y_1' = y_2,$$
$$y_2' = 2y_1 - y_2.$$

The M-file `exp22.m`:

```
function yprime = exp22(x,y);
yprime = [y(2); 2*y(1)-y(2)];
```

The call to the ode23 solver and the plot command:

```
xspan = [0 4]; % solution for x=0 to x=4
y0 = [4; 1]; % initial conditions
[x,y] = ode23('exp22',xspan,y0);
subplot(2,2,1); plot(x,y(:,1))
```

The numerical solution is plotted in Fig. 2.1. □

2.4. Independent Solutions

The form of independent solutions of a homogeneous equation,

$$Ly := y'' + ay' + by = 0, \tag{2.11}$$

depends on the form of the roots

$$\lambda_{1,2} = \frac{-a \pm \sqrt{a^2 - 4b}}{2} \tag{2.12}$$

of the characteristic equation

$$\lambda^2 + a\lambda + b = 0. \tag{2.13}$$

Let $\Delta = a^2 - 4b$ be the discriminant of equation (2.13). There are three cases: $\lambda_1 \neq \lambda_2$ real if $\Delta > 0$, $\lambda_2 = \bar{\lambda}_1$ complex if $\Delta < 0$, and $\lambda_1 = \lambda_2$ real if $\Delta = 0$.

Case I. In the case of two real distinct eigenvalues, $\lambda_1 \neq \lambda_2$, it was seen in Section 2.3 that the two solutions,

$$y_1(x) = e^{\lambda_1 x}, \qquad y_2(x) = e^{\lambda_2 x},$$

are independent since

$$\frac{y_2(x)}{y_1(x)} = \frac{e^{\lambda_2 x}}{e^{\lambda_1 x}} = e^{(\lambda_2 - \lambda_1)x} \neq \text{ constant}$$

since $\lambda_2 - \lambda_1 \neq 0$.

Therefore, the general solution is

$$y(x) = c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x}. \tag{2.14}$$

Case II. In the case of two distinct complex conjugate eigenvalues, we have

$$\lambda_1 = \alpha + i\beta, \qquad \lambda_2 = \alpha - i\beta = \bar{\lambda}_1, \qquad \text{where} \quad i = \sqrt{-1}.$$

By means of Euler's identity,

$$e^{i\theta} = \cos\theta + i\sin\theta, \tag{2.15}$$

the two complex solutions can be written in the form

$$u_1(x) = e^{(\alpha+i\beta)x} = e^{\alpha x}(\cos\beta x + i\sin\beta x),$$
$$u_2(x) = e^{(\alpha-i\beta)x} = e^{\alpha x}(\cos\beta x - i\sin\beta x) = \overline{u_1(x)}.$$

Since $\lambda_1 \neq \lambda_2$, the solutions u_1 and u_2 are independent. To have two real independent solutions, we use the following change of basis, or, equivalently we take the real and imaginary parts of u_1 since a and b are real and (2.11) is homogeneous (since the real and imaginary parts of a complex solution of a homogeneous linear equation with real coefficients are also solutions, or since any linear combination of solutions is still a solution). Thus,

$$y_1(x) = \Re u_1(x) = \frac{1}{2}[u_1(x) + u_2(x)] = e^{\alpha x}\cos\beta x, \tag{2.16}$$

$$y_2(x) = \Im u_1(x) = \frac{1}{2i}[u_1(x) - u_2(x)] = e^{\alpha x}\sin\beta x. \tag{2.17}$$

It is clear that y_1 and y_2 are independent. Therefore, the general solution is

$$y(x) = c_1\, e^{\alpha x}\cos\beta x + c_2\, e^{\alpha x}\sin\beta x. \tag{2.18}$$

Case III. In the case of real double eigenvalues we have

$$\lambda = \lambda_1 = \lambda_2 = -\frac{a}{2}$$

and equation (2.11) admits a solution of the form

$$y_1(x) = e^{\lambda x}. \tag{2.19}$$

To obtain a second independent solution, we use the *method of variation of parameters*, which is described in greater detail in Section 3.5. Thus, we put

$$y_2(x) = u(x)y_1(x). \tag{2.20}$$

It is important to note that the parameter u is a function of x and that y_1 is a solution of (2.11). We substitute y_2 in (2.11). This amounts to adding the following three equations,

$$\begin{aligned} by_2(x) &= bu(x)y_1(x) \\ ay_2'(x) &= au(x)y_1'(x) + ay_1(x)u'(x) \\ y_2''(x) &= u(x)y_1''(x) \ + 2y_1'(x)u'(x) + y_1(x)u''(x) \end{aligned}$$

to get

$$Ly_2 = u(x)Ly_1 + [ay_1(x) + 2y_1'(x)]u'(x) + y_1(x)u''(x).$$

The left-hand side is zero since y_2 is assumed to be a solution of $Ly = 0$. The first term on the right-hand side is also zero since y_1 is a solution of $Ly = 0$.

The second term on the right-hand side is zero since

$$\lambda = -\frac{a}{2} \in \mathbb{R},$$

and $y_1'(x) = \lambda y_1(x)$, that is,

$$ay_1(x) + 2y_1'(x) = a\,e^{-ax/2} - a\,e^{-ax/2} = 0.$$

It follows that

$$u''(x) = 0,$$

whence

$$u'(x) = k_1$$

and

$$u(x) = k_1x + k_2.$$

We therefore have

$$y_2(x) = k_1x\,e^{\lambda x} + k_2\,e^{\lambda x}.$$

We may take $k_2 = 0$ since the second term on the right-hand side is already contained in the linear span of y_1. Moreover, we may take $k_1 = 1$ since the general solution contains an arbitrary constant multiplying y_2.

It is clear that the solutions

$$y_1(x) = e^{\lambda x}, \qquad y_2(x) = x\,e^{\lambda x},$$

are linearly independent. The general solution is

$$y(x) = c_1\,e^{\lambda x} + c_2x\,e^{\lambda x}. \tag{2.21}$$

EXAMPLE 2.3. Consider the following three problems.

i) Find the general solution of the homogeneous equation with constant coefficients:

$$y'' - 7y' + 12y = 0.$$

The characteristic equation is

$$\lambda^2 - 7\lambda + 12 = (\lambda - 3)(\lambda - 4) = 0.$$

So $\lambda_1 = 3$ and $\lambda_2 = 4$ (Case I). Therefore, the general solution is

$$y(x) = c_1\,e^{3x} + c_2\,e^{4x}.$$

ii) Find the general solution of the homogeneous equation with constant coefficients:

$$y'' + 4y' + 10y = 0.$$

The characteristic equation is

$$\lambda^2 + 4\lambda + 10 = 0.$$

So

$$\lambda_{1,2} = \frac{-4 \pm \sqrt{(4)^2 - 4(10)}}{2} = \frac{-4 \pm \sqrt{-24}}{2} = -2 \pm i\sqrt{6}.$$

This is Case II and the general solution is

$$y(x) = c_1\, e^{-2x} \cos(\sqrt{6}\, x) + c_2\, e^{-2x} \sin(\sqrt{6}\, x).$$

iii) Solve the initial value problem

$$y'' - 4y' + 4y = 0, \qquad y(0) = 0, \quad y'(0) = 3.$$

The characteristic equation is

$$\lambda^2 - 4\lambda + 4 = (\lambda - 2)^2 = 0.$$

So $\lambda_1 = \lambda_2 = 2$ (Case III) and the general solution is

$$y(x) = c_1\, e^{2x} + c_2 x\, e^{2x}.$$

By the first initial condition,

$$y(0) = c_1\, e^0 + c_2(0)\, e0 = c_1 = 0.$$

So $y(x) = c_2 x\, e^{2x}$. Then

$$y'(x) = c_2\, e^{2x} + 2c_2 x\, e^{2x}$$

and, at $x = 0$,

$$y'(0) = c_2\, e^0 + 2c_2(0)\, e^0 = c_2 = 3.$$

Therefore, the unique solution is

$$y(x) = 3x\, e^{2x}.$$

2.5. Modeling in Mechanics

We consider elementary models of mechanics.

Example 2.4 (Free Oscillation). Consider a vertical spring attached to a rigid beam. The spring resists both extension and compression with Hooke's constant equal to k. Study the problem of the free vertical oscillation of a mass of m kg which is attached to the lower end of the spring.

Solution. Let the positive Oy axis point downward. Let s_0 m be the extension of the spring due to the force of gravity acting on the mass at rest at $y = 0$. (See Fig. 2.2).

We neglect friction. The force due to gravity is

$$F_1 = mg, \qquad \text{where} \quad g = 9.8 \text{ m/sec}^2.$$

The restoration force exerted by the spring is

$$F_2 = -k\, s_0.$$

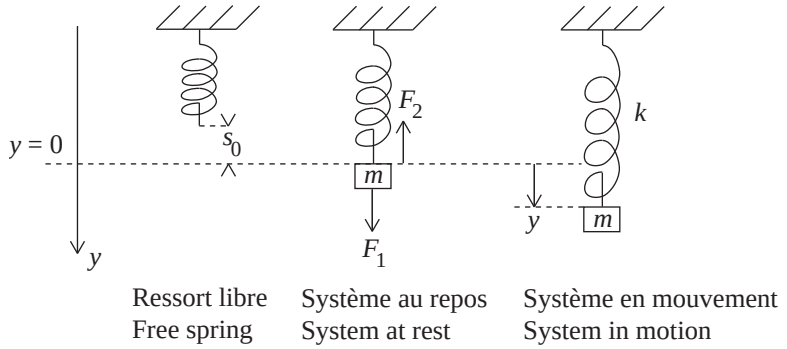

FIGURE 2.2. Undamped System.

By Newton's Second Law of Motion, when the system is at rest, at position $y = 0$, the resultant is zero,

$$F_1 + F_2 = 0.$$

Now consider the system in motion in position y. By the same law, the resultant is

$$m\,a = -k\,y.$$

Since the acceleration is $a = y''$, then

$$my'' + ky = 0, \qquad \text{or} \quad y'' + \omega^2 y = 0, \qquad \omega = \sqrt{\frac{k}{m}},$$

where $\omega/2\pi$ Hz is the frequency of the system. The characteristic equation of this differential equation,

$$\lambda^2 + \omega^2 = 0,$$

admits the pure imaginary eigenvalues

$$\lambda_{1,2} = \pm i\omega.$$

Hence, the general solution is

$$y(t) = c_1 \cos\omega t + c_2 \sin\omega t.$$

We see that the system oscillates freely without any loss of energy. □

The **amplitude**, A, and **period**, p, of the previous system are

$$A = \sqrt{c_1^2 + c_2^2}, \qquad p = \frac{2\pi}{\omega}.$$

The amplitude can be obtained by rewriting $y(t)$ with phase shift φ as follows:

$$\begin{aligned} y(t) &= A(\cos\omega t + \varphi) \\ &= A\cos\varphi\cos\omega t - A\sin\varphi\sin\omega t \\ &= c_1\cos\omega t + c_2\sin\omega t. \end{aligned}$$

Then, identifying coefficients we have

$$c_1^1 + c_2^2 = (A\cos\varphi)^2 + (A\sin\varphi)^2 = A^2.$$

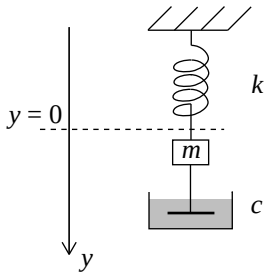

FIGURE 2.3. Damped system.

EXAMPLE 2.5 (Damped System). Consider a vertical spring attached to a rigid beam. The spring resists extension and compression with Hooke's constant equal to k. Study the problem of the damped vertical motion of a mass of m kg which is attached to the lower end of the spring. (See Fig. 2.3). The damping constant is equal to c.

SOLUTION. Let the positive Oy axis point downward. Let s_0 m be the extension of the spring due to the force of gravity on the mass at rest at $y = 0$. (See Fig. 2.2).

The force due to gravity is

$$F_1 = mg, \qquad \text{where} \quad g = 9.8 \text{ m/sec}^2.$$

The restoration force exerted by the spring is

$$F_2 = -k\, s_0.$$

By Newton's Second Law of Motion, when the system is at rest, the resultant is zero,

$$F_1 + F_2 = 0.$$

Since damping opposes motion, by the same law, the resultant for the system in motion is

$$m\, a = -c\, y' - k\, y.$$

Since the acceleration is $a = y''$, then

$$my'' + cy' + ky = 0, \qquad \text{or} \quad y'' + \frac{c}{m}\, y' + \frac{k}{m}\, y = 0.$$

The characteristic equation of this differential equation,

$$\lambda^2 + \frac{c}{m}\, \lambda + \frac{k}{m} = 0,$$

admits the eigenvalues

$$\lambda_{1,2} = -\frac{c}{2m} \pm \frac{1}{2m}\sqrt{c^2 - 4mk} =: -\alpha \pm \beta, \qquad \alpha > 0.$$

There are three cases to consider.

Case I: Overdamping. If $c^2 > 4mk$, the system is overdamped. Both eigenvalues are real and negative since

$$\lambda_1 = -\frac{c}{2m} - \frac{1}{2m}\sqrt{c^2 - 4mk} < 0, \qquad \lambda_1 \lambda_2 = \frac{k}{m} > 0.$$

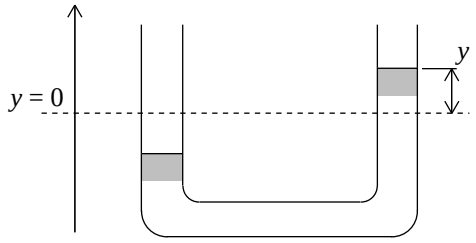

FIGURE 2.4. Vertical movement of a liquid in a U tube.

The general solution,

$$y(t) = c_1\, e^{\lambda_1 t} + c_2\, e^{\lambda_2 t},$$

decreases exponentially to zero without any oscillation of the system.

Case II: Underdamping. If $c^2 < 4mk$, the system is underdamped. The two eigenvalues are complex conjugate to each other,

$$\lambda_{1,2} = -\frac{c}{2m} \pm \frac{i}{2m}\sqrt{4mk - c^2} =: -\alpha \pm i\beta, \qquad \text{with} \quad \alpha > 0.$$

The general solution,

$$y(t) = c_1\, e^{-\alpha t} \cos \beta t + c_2\, e^{-\alpha t} \sin \beta t,$$

oscillates while decreasing exponentially to zero.

Case III: Critical damping. If $c^2 = 4mk$, the system is critically damped. Both eigenvalues are real and equal,

$$\lambda_{1,2} = -\frac{c}{2m} = -\alpha, \qquad \text{with} \quad \alpha > 0.$$

The general solution,

$$y(t) = c_1\, e^{-\alpha t} + c_2 t\, e^{-\alpha t} = (c_1 + c_2 t)\, e^{-\alpha t},$$

decreases exponentially to zero with an initial increase in $y(t)$ if $c_2 > 0$. □

EXAMPLE 2.6 (Oscillation of water in a tube in a U form).
Find the frequency of the oscillatory movement of 2 L of water in a tube in a U form. The diameter of the tube is 0.04 m.

SOLUTION. We neglect friction between the liquid and the tube wall. The mass of the liquid is $m = 2\,\text{kg}$. The volume, of height $h = 2y$, responsible for the restoring force is

$$\begin{aligned} V = \pi r^2 h &= \pi(0.02)^2 2y\,\text{m}^3 \\ &= \pi(0.02)^2 2000y\,\text{L} \end{aligned}$$

(see Fig. 2.4). The mass of volume V is

$$M = \pi(0.02)^2 2000y\,\text{kg}$$

and the restoration force is

$$Mg = \pi(0.02)^2 9.8 \times 2000y\,\text{N}, \qquad g = 9.8\,\text{m/s}^2.$$

By Newton's Second Law of Motion,

$$m\,y'' = -Mg,$$

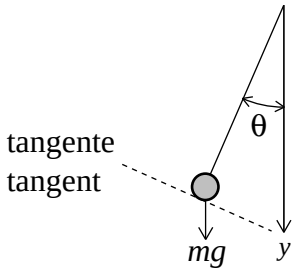

FIGURE 2.5. Pendulum in motion.

that is,

$$y'' + \frac{\pi(0.02)^2 9.8 \times 2000}{2}\, y = 0, \quad \text{or} \quad y'' + \omega_0^2 y = 0,$$

where

$$\omega_0^2 = \frac{\pi(0.02)^2 9.8 \times 2000}{2} = 12.3150.$$

Hence, the frequency is

$$\frac{\omega_0}{2\pi} = \frac{\sqrt{12.3150}}{2\pi} = 0.5585\,\text{Hz}. \quad \square$$

EXAMPLE 2.7 (Oscillation of a pendulum). Find the frequency of the oscillations of small amplitude of a pendulum of mass m kg and length $L = 1$ m.

SOLUTION. We neglect air resistance and the mass of the rod. Let θ be the angle, in radian measure, made by the pendulum measured from the vertical axis. (See Fig. 2.5).

The tangential force is

$$m\,a = mL\theta''.$$

Since the length of the rod is fixed, the orthogonal component of the force is zero. Hence it suffices to consider the tangential component of the restoring force due to gravity, that is,

$$mL\theta'' = -mg\sin\theta \approx -mg\theta, \qquad g = 9.8,$$

where $\sin\theta \approx \theta$ if θ is sufficiently small. Thus,

$$\theta'' + \frac{g}{L}\,\theta = 0, \quad \text{or} \quad \theta'' + \omega_0^2\theta = 0, \quad \text{where} \quad \omega_0^2 = \frac{g}{L} = 9.8.$$

Therefore, the frequency is

$$\frac{\omega_0}{2\pi} = \frac{\sqrt{9.8}}{2\pi} = 0.498\,\text{Hz}. \quad \square$$

$\square$

2.6. Euler–Cauchy Equations

Consider the homogeneous Euler–Cauchy equation

$$Ly := x^2y'' + axy' + by = 0, \qquad x > 0. \tag{2.22}$$

Because of the particular form of the differential operator of this equation with variable coefficients,

$$L = x^2D^2 + axD + bI, \qquad D = {}' = \frac{d}{dx},$$

where each term is of the form $a_kx^kD^k$, with a_k a constant, we can solve (2.22) by setting

$$y(x) = x^m \tag{2.23}$$

In fact, if $y(x) = x^m$, then $D^ky = m(m-1)\cdots(m-k+1)x^{m-k}$ and so $x^kD^ky = m(m-1)\cdots(m-k+1)x^m$, i.e. all terms will have the same power of x and hence a linear combination of them can be zero for all x.

In (2.22),

$$m(m-1)x^m + amx^m + bx^m = x^m[m(m-1) + am + b] = 0.$$

We can divide by x^m if $x > 0$. We thus obtain the characteristic equation

$$m^2 + (a-1)m + b = 0. \tag{2.24}$$

The eigenvalues or roots are

$$m_{1,2} = \frac{1-a}{2} \pm \frac{1}{2}\sqrt{(a-1)^2 - 4b}. \tag{2.25}$$

There are three cases: $m_1 \neq m_2$ real, m_1 and $m_2 = \overline{m}_1$ complex and distinct, and $m_1 = m_2$ real.

Case I. If both roots are real and distinct, the general solution of (2.22) is

$$y(x) = c_1x^{m_1} + c_2x^{m_2}, \tag{2.26}$$

because $y_1(x) = x^{m_1}$ and $y_2(x) = x^{m_2}$ are linearly independent as

$$\frac{y_2(x)}{y_1(x)} = \frac{x^{m_2}}{x^{m_1}} = x^{m_2-m_1} \neq \text{ constant}$$

since $m_2 - m_1 \neq 0$.

Case II. If the roots are complex conjugates of one another,

$$m_1 = \alpha + i\beta, \qquad m_2 = \alpha - i\beta, \qquad \beta \neq 0,$$

we have two independent complex solutions of the form

$$u_1 = x^\alpha x^{i\beta} = x^\alpha\, e^{i\beta\ln x} = x^\alpha[\cos(\beta\ln x) + i\sin(\beta\ln x)]$$

and

$$u_2 = x^\alpha x^{-i\beta} = x^\alpha\, e^{-i\beta\ln x} = x^\alpha[\cos(\beta\ln x) - i\sin(\beta\ln x)].$$

For $x > 0$, we obtain two real independent solutions by adding and subtracting u_1 and u_2, and dividing the sum and the difference by 2 and $2i$, respectively, or, equivalently, by taking the real and imaginary parts of u_1 since a and b are real and (2.22) is linear and homogeneous:

$$y_1(x) = x^\alpha\cos(\beta\ln x), \qquad y_2(x) = x^\alpha\sin(\beta\ln x).$$

Clearly, $y_1(x)$ and $y_2(x)$ are independent. The general solution of (2.22) is

$$y(x) = c_1 x^\alpha \cos(\beta \ln x) + c_2 x^\alpha \sin(\beta \ln x). \tag{2.27}$$

Case III. If both roots are real and equal,

$$m = m_1 = m_2 = \frac{1-a}{2},$$

one solution is of the form

$$y_1(x) = x^m.$$

We find a second independent solution by variation of parameters by putting

$$y_2(x) = u(x) y_1(x)$$

in (2.22). Adding the left- and right-hand sides of the following three expressions, we have

$$\begin{aligned} by_2(x) &= bu(x)y_1(x) \\ axy_2'(x) &= axu(x)y_1'(x) + axy_1(x)u'(x) \\ x^2y_2''(x) &= x^2u(x)y_1''(x) + 2x^2y_1'(x)u'(x) + x^2y_1(x)u''(x) \end{aligned}$$

to get

$$Ly_2 = u(x)Ly_1 + [axy_1(x) + 2x^2y_1'(x)]u'(x) + x^2y_1(x)u''(x).$$

The left-hand side is zero since y_2 is assumed to be a solution of $Ly = 0$. The first term on the right-hand side is also zero since y_1 is a solution of $Ly = 0$.

The coefficient of u' is

$$\begin{aligned} axy_1(x) + 2x^2y_1'(x) &= axx^m + 2mx^2x^{m-1} = ax^{m+1} + 2mx^{m+1} \\ &= (a+2m)x^{m+1} = \left[a + 2\left(\frac{1-a}{2}\right)\right] x^{m+1} = x^{m+1}. \end{aligned}$$

Hence we have

$$x^2y_1(x)u'' + x^{m+1}u' = x^{m+1}(xu'' + u') = 0, \qquad x > 0,$$

and dividing by x^{m+1}, we have

$$xu'' + u' = 0.$$

Since u is absent from this differential equation, we can reduce the order by putting

$$v = u', \qquad v' = u''.$$

The resulting equation is separable,

$$x\frac{dv}{dx} + v = 0, \qquad \text{that is,} \quad \frac{dv}{v} = -\frac{dx}{x},$$

and can be integrated,

$$\ln|v| = \ln x^{-1} \implies u' = v = \frac{1}{x} \implies u = \ln x.$$

No constant of integration is needed here. The second independent solution is

$$y_2 = (\ln x)x^m.$$

Therefore, the general solution of (2.22) is

$$y(x) = c_1x^m + c_2(\ln x)x^m. \tag{2.28}$$

EXAMPLE 2.8. Find the general solution of the Euler–Cauchy equation

$$x^2y'' - 6y = 0.$$

SOLUTION. **(a) The analytic solution.**— Putting $y = x^m$ in the differential equation, we have

$$m(m-1)x^m - 6x^m = 0.$$

The characteristic equation is

$$m^2 - m - 6 = (m-3)(m+2) = 0.$$

The eigenvalues,

$$m_1 = 3, \qquad m_2 = -2,$$

are real and distinct. The general solution is

$$y(x) = c_1x^3 + c_2x^{-2}.$$

(b) The Matlab symbolic solution.—

```
dsolve('x^2*D2y=6*y','x')
y = (C1+C2*x^5)/x^2
```

□

EXAMPLE 2.9. Solve the initial value problem

$$x^2y'' - 6y = 0, \qquad y(1) = 2, \qquad y'(1) = 1.$$

SOLUTION. **(a) The analytic solution.**— The general solution as found in Example 2.8 is

$$y(x) = c_1x^3 + c_2x^{-2}.$$

From the initial conditions, we have the linear system in c_1 and c_2:

$$\begin{aligned} y(1) &= \ c_1 + \ c_2 = 2 \\ y'(1) &= 3c_1 - 2c_2 = 1 \end{aligned}$$

whose solution is

$$c_1 = 1, \qquad c_2 = 1.$$

Hence the unique solution is

$$y(x) = x^3 + x^{-2}.$$

(b) The Matlab symbolic solution.—

```
dsolve('x^2*D2y=6*y','y(1)=2','Dy(1)=1','x')
y = (1+x^5)/x^2
```

(c) The Matlab numeric solution.— To rewrite the second-order differential equation as a system of first-order equations, we put

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

with initial conditions at $x = 1$:

$$y_1(1) = 2, \qquad y_2(1) = 1.$$

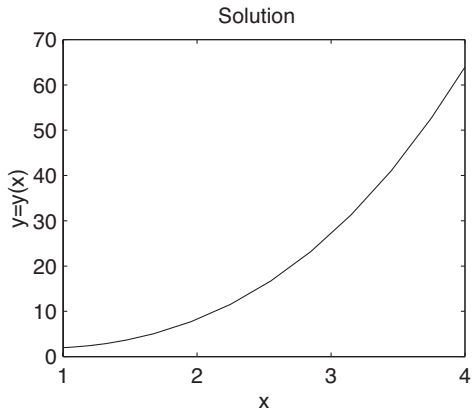

FIGURE 2.6. Graph of solution of the Euler–Cauchy equation in Example 2.9.

Thus, we have

$$y_1' = y_2,$$
$$y_2' = 6y_1/x^2.$$

The M-file `euler2.m`:

```
function yprime = euler2(x,y);
yprime = [y(2); 6*y(1)/x^2];
```

The call to the `ode23` solver and the plot command:

```
xspan = [1 4]; % solution for x=1 to x=4
y0 = [2; 1]; % initial conditions
[x,y] = ode23('euler2',xspan,y0);
subplot(2,2,1); plot(x,y(:,1))
```

The numerical solution is plotted in Fig. 2.6. □

EXAMPLE 2.10. Find the general solution of the Euler–Cauchy equation

$$x^2y'' + 7xy' + 9y = 0.$$

SOLUTION. The characteristic equation

$$m^2 + 6m + 9 = (m+3)^2 = 0$$

admits a double root $m = -3$. Hence the general solution is

$$y(x) = (c_1 + c_2 \ln x)x^{-3}. \quad \square$$

EXAMPLE 2.11. Find the general solution of the Euler–Cauchy equation

$$x^2y'' + 1.25y = 0.$$

SOLUTION. The characteristic equation

$$m^2 - m + 1.25 = 0$$

admits a pair of complex conjuguate roots

$$m_1 = \frac{1}{2} + i, \qquad m_2 = \frac{1}{2} - i.$$

Hence the general solution is

$$y(x) = x^{1/2}[c_1 \cos(\ln x) + c_2 \sin(\ln x)]. \quad \square$$

The existence and uniqueness of solutions of initial value problems of order greater than 1 will be considered in the next chapter.

CHAPTER 3

Linear Differential Equations of Arbitrary Order

3.1. Homogeneous Equations

Consider the linear *nonhomogeneous* differential equation of order n,

$$y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = r(x), \tag{3.1}$$

with variable coefficients, $a_0(x), a_1(x), \ldots, a_{n-1}(x)$. Let L denote the differential operator on the left-hand side,

$$L := D^n + a_{n-1}(x)D^{n-1} + \cdots + a_1(x)D + a_0(x), \qquad D := \ ' \ = \frac{d}{dx}. \tag{3.2}$$

Then the nonhomogeneous equation (3.1) is written in the form

$$Lu = r(x).$$

If $r \equiv 0$, equation (3.1) is said to be *homogeneous*,

$$y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = 0, \tag{3.3}$$

that is,

$$Ly = 0.$$

DEFINITION 3.1. A *solution* of (3.1) or (3.3) on the interval $]a, b[$ is a function $y(x)$, n times continuously differentiable on $]a, b[$, which satisfies identically the differential equation.

Theorem 2.1 proved in the previous chapter generalizes to linear homogeneous equations of arbitrary order n.

THEOREM 3.1. *The solutions of the homogeneous equation* (3.3) *form a vector space.*

PROOF. Let $y_1, y_2, \ldots, y_k$ be k solutions of $Ly = 0$. The linearity of the operator L implies that

$$L(c_1y_1 + c_2y_2 + \cdots + c_ky_k) = c_1Ly_1 + c_2Ly_2 + \cdots + c_kLy_k = 0, \qquad c_i \in \mathbb{R}. \quad \square$$

DEFINITION 3.2. We say that n *functions*, $f_1, f_2, \ldots, f_n$, are *linearly dependent* on the interval $]a, b[$ if and only if there exist n constants not all zero, i.e.

$$(k_1, k_2, \ldots, k_n) \neq (0, 0, \ldots, 0),$$

such that

$$k_1f_1(x) + k_2f_2(x) + \cdots + k_nf_n(x) = 0, \qquad \text{for all} \quad x \in]a, b[. \tag{3.4}$$

Otherwise, they are said to be *linearly independent.*

REMARK 3.1. Let $f_1, f_2, \ldots, f_n$ be n linearly dependent *functions*. Without loss of generality, we may suppose that $k_1 \neq 0$ in (3.4). Then f_1 is a linear combination of $f_2, f_3, \ldots, f_n$.

$$f_1(x) = -\frac{1}{k_1}\,[k_2 f_2(x) + \cdots + k_n f_n(x)].$$

We have the following Existence and Uniqueness Theorem.

THEOREM 3.2 (Existence and Uniqueness). *If the functions $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ are continuous on the interval $]a, b[$ and $x_0 \in]a, b[$, then the initial value problem*

$$Ly = 0, \qquad y(x_0) = k_1, \quad y'(x_0) = k_2, \quad \ldots, \quad y^{(n-1)}(x_0) = k_n, \tag{3.5}$$

admits one and only one solution.

PROOF. One can prove the theorem by reducing the differential equation of order n to a system of n differential equations of the first order. To do this, define the n dependent variables

$$u_1 = y, \quad u_2 = y', \quad \ldots, \quad u_n = y^{(n-1)}.$$

Then the initial value problem becomes

$$\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix}' = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix}, \quad \begin{bmatrix} u_1(x_0) \\ u_2(x_0) \\ \vdots \\ u_{n-1}(x_0) \\ u_n(x_0) \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_{n-1} \\ k_n \end{bmatrix},$$

or, in matrix and vector notation,

$$\boldsymbol{u}'(x) = A(x)\boldsymbol{u}(x), \qquad \boldsymbol{u}(x_0) = \boldsymbol{k}.$$

We say that the matrix A is a companion matrix because the determinant $|A - \lambda I|$ is the characteristic polynomial of the homogeneous differential equation,

$$|A - \lambda I| = (-1)^n(\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0 = (-1)^n p_n(\lambda).$$

Using Picard's method, one can show that this system admits one and only one solution. Picard's iterative procedure is as follows:

$$\boldsymbol{u}^{[n]}(x) = \boldsymbol{u}^{[0]}(x_0) + \int_{x_0}^{x} A(t)\boldsymbol{u}^{[n-1]}(t)\,dt, \qquad \boldsymbol{u}^{[0]}(x_0) = \boldsymbol{k}. \quad \square$$

DEFINITION 3.3. The Wronskian of n *functions*, $f_1(x), f_2(x), \ldots, f_n(x)$, $n-1$ times differentiable on the interval $]a, b[$, is the following determinant of order n:

$$W(f_1, f_2, \ldots, f_n)(x) := \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & & & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}. \tag{3.6}$$

The linear dependence of n *solutions* of the linear homogeneous differential equation (3.3) is characterized by means of their Wronskian.

First, let us prove Abel's Lemma.

LEMMA 3.1 (Abel). *Let $y_1, y_2, \dots, y_n$ be n solutions of (3.3) on the interval $]a, b[$. Then the Wronskian $W(x) = W(y_1, y_2, \dots, y_n)(x)$ satisfies the following identity:*

$$W(x) = W(x_0)\, e^{-\int_{x_0}^{x} a_{n-1}(t)\, dt}, \qquad x_0 \in]a, b[. \tag{3.7}$$

PROOF. For simplicity of writing, let us take $n = 3$; the general case is treated as easily. Let $W(x)$ be the Wronskian of three solutions y_1, y_2, y_3. The derivative $W'(x)$ of the Wronskian is of the form

$$\begin{aligned}
W'(x) &= \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1'' & y_2'' & y_3'' \end{vmatrix}' \\
&= \begin{vmatrix} y_1' & y_2' & y_3' \\ y_1' & y_2' & y_3' \\ y_1'' & y_2'' & y_3'' \end{vmatrix} + \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1'' & y_2'' & y_3'' \\ y_1'' & y_2'' & y_3'' \end{vmatrix} + \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1''' & y_2''' & y_3''' \end{vmatrix} \\
&= \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1''' & y_2''' & y_3''' \end{vmatrix} \\
&= \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ -a_0 y_1 - a_1 y_1' - a_2 y_1'' & -a_0 y_2 - a_1 y_2' - a_2 y_2'' & -a_0 y_3 - a_1 y_3' - a_2 y_3'' \end{vmatrix},
\end{aligned}$$

since the first two determinants of the second line are zero because two rows are equal, and in the last determinant we have used the fact that y_k, $k = 1, 2, 3$, is a solution of the homogeneous equation (3.3).

Adding to the third row a_0 times the first row and a_1 times the second row, we obtain the separable differential equation

$$W'(x) = -a_2(x) W(x),$$

namely,

$$\frac{dW}{W} = -a_2(x)\, dx.$$

The solution is

$$\ln |W| = -\int a_2(x)\, dx + c,$$

that is

$$W(x) = W(x_0)\, e^{-\int_{x_0}^{x} a_2(t)\, dt}, \qquad x_0 \in]a, b[. \quad \square$$

THEOREM 3.3. *If the coefficients $a_0(x), a_1(x), \dots, a_{n-1}(x)$ of (3.3) are continuous on the interval $]a, b[$, then n solutions, $y_1, y_2, \dots, y_n$, of (3.3) are linearly dependent if and only if their Wronskian is zero at a point $x_0 \in]a, b[$,*

$$W(y_1, y_2, \dots, y_n)(x_0) := \begin{vmatrix} y_1(x_0) & \cdots & y_n(x_0) \\ y_1'(x_0) & \cdots & y_n'(x_0) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x_0) & \cdots & y_n^{(n-1)}(x_0) \end{vmatrix} = 0. \tag{3.8}$$

PROOF. If the solutions are linearly dependent, then by Definition 3.2 there exist n constants not all zero,

$$(k_1, k_2, \dots, k_n) \neq (0, 0, \dots, 0),$$

such that

$$k_1y_1(x) + k_2y_2(x) + \cdots + k_ny_n(x) = 0, \qquad \text{for all} \quad x \in]a, b[.$$

Differentiating this identity $n-1$ times, we obtain

$$\begin{aligned} k_1y_1(x) + k_2y_2(x) + \cdots + k_ny_n(x) &= 0, \\ k_1y_1'(x) + k_2y_2'(x) + \cdots + k_ny_n'(x) &= 0, \\ &\vdots \\ k_1y_1^{(n-1)}(x) + k_2y_2^{(n-1)}(x) + \cdots + k_ny_n^{(n-1)}(x) &= 0. \end{aligned}$$

We rewrite this homogeneous algebraic linear system, in the n unknowns $k_1, k_2, \ldots, k_n$ in matrix form,

$$\begin{bmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{3.9}$$

that is,

$$A\boldsymbol{k} = 0.$$

Since, by hypothesis, the solution $\boldsymbol{k}$ is nonzero, the determinant of the system must be zero,

$$\det A = W(y_1, y_2, \ldots, y_n)(x) = 0, \qquad \text{for all} \quad x \in]a, b[.$$

On the other hand, if the Wronskian of n *solutions* is zero at an arbitrary point x_0,

$$W(y_1, y_2, \ldots, y_n)(x_0) = 0,$$

then it is zero for all $x \in]a, b[$ by Abel's Lemma 3.1. Hence the determinant $W(x)$ of system (3.9) is zero for all $x \in]a, b[$. Therefore this system admits a nonzero solution $\boldsymbol{k}$. Consequently, the solutions, $y_1, y_2, \ldots, y_n$, of (3.3) are linearly dependent. □

REMARK 3.2. The Wronskian of n linearly dependent functions, which are sufficiently differentiable on $]a, b[$, is necessarily zero on $]a, b[$, as can be seen from the first part of the proof of Theorem 3.3. But for functions which *are not solutions* of the same linear homogeneous differential equation, a zero Wronskian on $]a, b[$ is not a sufficient condition for the linear dependence of these functions. For instance, $u_1 = x^3$ and $u_2 = |x|^3$ are of class C^1 in the interval $[-1, 1]$ and are linearly independent, but satisfy $W(x^3, |x|^3) = 0$ identically.

COROLLARY 3.1. *If the coefficients $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ of* (3.3) *are continuous on $]a, b[$, then n solutions, $y_1, y_2, \ldots, y_n$, of* (3.3) *are linearly independent if and only if their Wronskian is not zero at a single point $x_0 \in]a, b[$.*

This follows from Theorem 3.3.

COROLLARY 3.2. *Suppose $f_1(x), f_2(x), \ldots, f_n(x)$ are n functions which possess continuous nth-order derivatives on a real interval I, and $W(f_1, \ldots, f_n)(x) \neq 0$ on I. Then there exists a unique homogeneous differential equation of order n*

(with the coefficient of $y^{(n)}$ equal to one) for which these functions are linearly independent solutions, namely,

$$(-1)^n \frac{W(y, f_1, \ldots, f_n)}{W(f_1, \ldots, f_n)} = 0.$$

EXAMPLE 3.1. Show that the functions

$$y_1(x) = \cosh x \quad \text{and} \quad y_2(x) = \sinh x$$

are linearly independent.

SOLUTION. Since y_1 and y_2 are twice continuously differentiable and

$$W(y_1, y_2)(x) = \begin{vmatrix} \cosh x & \sinh x \\ \sinh x & \cosh x \end{vmatrix} = \cosh^2 x - \sinh^2 x = 1,$$

for all x, then, by Corollary 3.2, y_1 and y_2 are linearly independent. Incidently, it is easy to see that y_1 and y_2 are solutions of the differential equation

$$y'' - y = 0.$$

□

REMARK 3.3. In the previous solution we have used the following identity:

$$\begin{aligned} \cosh^2 x - \sinh^2 x &= \left(\frac{e^x + e^{-x}}{2}\right)^2 - \left(\frac{e^x - e^{-x}}{2}\right)^2 \\ &= \frac{1}{4}\left(e^{2x} + e^{-2x} + 2\,e^x\,e^{-x} - e^{2x} - e^{-2x} + 2\,e^x\,e^{-x}\right) \\ &= 1. \end{aligned}$$

EXAMPLE 3.2. Use the Wronskian of the functions

$$y_1(x) = x^m \quad \text{and} \quad y_2(x) = x^m \ln x$$

to show that they are linearly independent for $x > 0$ and construct a second-order differential equation for which they are solutions.

SOLUTION. We verify that the Wronskian of y_1 and y_2 does not vanish for $x > 0$:

$$\begin{aligned} W(y_1, y_2)(x) &= \begin{vmatrix} x^m & x^m \ln x \\ mx^{m-1} & mx^{m-1} \ln x + x^{m-1} \end{vmatrix} \\ &= x^m x^{m-1} \begin{vmatrix} 1 & \ln x \\ m & m \ln x + 1 \end{vmatrix} \\ &= x^{2m-1}(1 + m\,\ln x - m\,\ln x) = x^{2m-1} \neq 0, \qquad \text{for all } x > 0. \end{aligned}$$

Hence, by Corollary 3.2, y_1 and y_2 are linearly independent. By the same corollary

$$\begin{aligned} W(y, x^m, x^m \ln x)(x) &= \begin{vmatrix} y & x^m & x^m \ln x \\ y' & mx^{m-1} & mx^{m-1} \ln x + x^{m-1} \\ y'' & m(m-1)x^{m-2} & m(m-1)x^{m-2} \ln x + (2m-1)x^{m-2} \end{vmatrix} \\ &= 0. \end{aligned}$$

Multiplying the second and third rows by x and x^2, respectively, dividing the second and third columns by x^m, subtracting m times the first row from the

second row and $m(m-1)$ times the first row from the third row, one gets the equivalent simplified determinantal equation

$$\begin{vmatrix} y & 1 & \ln x \\ xy' - my & 0 & 1 \\ x^2y'' - m(m-1)y & 0 & 2m-1 \end{vmatrix} = 0,$$

which upon expanding by the second column produces the Euler–Cauchy equation

$$x^2y'' + (1-2m)xy' + m^2y = 0. \quad \square$$

DEFINITION 3.4. We say that n linearly independent solutions, $y_1, y_2, \ldots, y_n$, of the homogeneous equation (3.3) on $]a,b[$ form a *fundamental system* or *basis* on $]a,b[$.

DEFINITION 3.5. Let $y_1, y_2, \ldots, y_n$ be a fundamental system for (3.3). A solution of (3.3) on $]a,b[$ of the form

$$y(x) = c_1y_1(x) + c_2y_2(x) + \cdots + c_ny_n(x), \tag{3.10}$$

where $c_1, c_2, \ldots, c_n$ are n arbitrary constants, is said to be a *general solution* of (3.3) on $]a,b[$.

Recall that we need the general solution to span the vector space of solutions as it must represent all possible solutions of the equation.

THEOREM 3.4. *If the functions $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ are continuous on $]a,b[$, then the linear homogeneous equation* (3.3) *admits a general solution on $]a,b[$.*

PROOF. By Theorem 3.2, for each i, $i = 1, 2, \ldots, n$, the initial value problem (3.5),

$$Ly = 0, \qquad \text{with} \quad y^{(i-1)}(x_0) = 1, \quad y^{(j-1)}(x_0) = 0, \quad j \neq i,$$

admits one (and only one) solution $y_i(x)$ such that

$$y_i^{(i-1)}(x_0) = 1, \qquad y_i^{(j-1)}(x_0) = 0, \quad j = 1, 2, \ldots, i-1, i+1, \ldots, n.$$

Then the Wronskian W satisfies the following relation

$$W(y_1, y_2, \ldots, y_n)(x_0) = \begin{vmatrix} y_1(x_0) & \cdots & y_n(x_0) \\ y_1'(x_0) & \cdots & y_n'(x_0) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x_0) & \cdots & y_n^{(n-1)}(x_0) \end{vmatrix} = |I_n| = 1,$$

where I_n is the identity matrix of order n. It follows from Corollary 3.1 that the solutions are independent. $\square$

THEOREM 3.5. *If the functions $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ are continuous on $]a,b[$, then the solution of the initial value problem* (3.5) *on $]a,b[$ is obtained from a general solution.*

PROOF. Let

$$y = c_1y_1 + c_2y_2 + \cdots + c_ny_n$$

be a general solution of (3.3). The system

$$\begin{bmatrix} y_1(x_0) & \cdots & y_n(x_0) \\ y_1'(x_0) & \cdots & y_n'(x_0) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x_0) & \cdots & y_n^{(n-1)}(x_0) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}$$

admits a unique solution $\boldsymbol{c}$ since the determinant of the system is nonzero. □

3.2. Linear Homogeneous Equations

Consider the linear homogeneous differential equation of order n,

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1y' + a_0y = 0, \tag{3.11}$$

with constant coefficients, $a_0, a_1, \ldots, a_{n-1}$. Let L denote the differential operator on the left-hand side,

$$L := D^n + a_{n-1}D^{n-1} + \cdots + a_1D + a_0, \qquad D := {}' = \frac{d}{dx}. \tag{3.12}$$

Putting $y(x) = e^{\lambda x}$ in (3.11), we obtain the characteristic equation

$$p(\lambda) := \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0. \tag{3.13}$$

If the n roots of $p(\lambda) = 0$ are distinct, we have n independent solutions,

$$y_1(x) = e^{\lambda_1 x}, \qquad y_2(x) = e^{\lambda_2 x}, \qquad \ldots, \qquad y_n(x) = e^{\lambda_n x}, \tag{3.14}$$

and the general solution is of the form

$$y(x) = c_1\, e^{\lambda_1 x} + c_2\, e^{\lambda_2 x} + \cdots + c_n\, e^{\lambda_n x}. \tag{3.15}$$

If (3.13) has a double root, say, $\lambda_1 = \lambda_2$, we have two independent solutions of the form

$$y_1(x) = e^{\lambda_1 x}, \qquad y_2(x) = x\, e^{\lambda_1 x}.$$

Similarly, if there is a triple root, say, $\lambda_1 = \lambda_2 = \lambda_3$, we have three independent solutions of the form

$$y_1(x) = e^{\lambda_1 x}, \qquad y_2(x) = x\, e^{\lambda_1 x}, \qquad y_3(x) = x^2\, e^{\lambda_1 x}.$$

We prove the following theorem.

THEOREM 3.6. *Let μ be a root of multiplicity m of the characteristic equation* (3.13). *Then the differential equation* (3.11) *has m independent solutions of the form*

$$y_1(x) = e^{\mu x}, \qquad y_2(x) = x\, e^{\mu x}, \qquad \ldots, \qquad y_m(x) = x^{m-1}\, e^{\mu x}. \tag{3.16}$$

PROOF. Let us write

$$p(D)y = (D^n + a_{n-1}D^{n-1} + \cdots + a_1D + a_0)y = 0.$$

Since, by hypothesis,

$$p(\lambda) = q(\lambda)(\lambda - \mu)^m,$$

and the coefficients are constant, the differential operator $p(D)$ can be factored in the form

$$p(D) = q(D)(D - \mu)^m.$$

We see by recurrence that the m functions (3.16),

$$x^k\, e^{\mu x}, \qquad k = 0, 1, \ldots, m-1,$$

satisfy the following equations:

$$\begin{aligned}(D-\mu)\left(x^k\,e^{\mu x}\right) &= kx^{k-1}\,e^{\mu x}+\mu x^k\,e^{\mu x}-\mu x^k\,e^{\mu x}\\ &= kx^{k-1}\,e^{\mu x},\\ (D-\mu)^2\left(x^k\,e^{\mu x}\right) &= (D-\mu)\left(kx^{k-1}\,e^{\mu x}\right)\\ &= k(k-1)x^{k-2}\,e^{\mu x},\\ &\vdots\\ (D-\mu)^k\left(x^k\,e^{\mu x}\right) &= k!\,e^{\mu x},\\ (D-\mu)^{k+1}\left(x^k\,e^{\mu x}\right) &= k!\,(e^{\mu x}-e^{\mu x})=0.\end{aligned}$$

Since $m \geq k+1$, we have

$$(D-\mu)^m\left(x^k\,e^{\mu x}\right)=0, \qquad k=0,1,\ldots,m-1.$$

Hence, by Lemma 3.2 below, the functions (3.16) are m independent solutions of (3.11). □

LEMMA 3.2. *Let*

$$y_1(x)=e^{\mu x}, \qquad y_2(x)=x\,e^{\mu x}, \qquad \ldots, \qquad y_m(x)=x^{m-1}\,e^{\mu x},$$

be m solutions of a linear homogeneous differential equation. Then they are independent.

PROOF. By Corollary 3.1, it suffices to show that the Wronskian of the solutions is nonzero at $x=0$. We have seen, in the proof of the preceding theorem, that

$$(D-\mu)^k\left(x^k\,e^{\mu x}\right)=k!\,e^{\mu x},$$

that is,

$$D^k\left(x^k\,e^{\mu x}\right)=k!\,e^{\mu x}+ \text{ terms in } x^l\,e^{\mu x}, \qquad l=1,2,\ldots,k-1.$$

Hence

$$D^k\left(x^k\,e^{\mu x}\right)\big|_{x=0}=k!, \qquad D^k\left(x^{k+l}\,e^{\mu x}\right)\big|_{x=0}=0, \quad l\geq 1.$$

It follows that the matrix M of the Wronskian at $x=0$ is lower triangular with $m_{i,i}=(i-1)!$,

$$W(0)=\begin{vmatrix} 0! & 0 & 0 & \ldots & 0\\ \times & 1! & 0 & & 0\\ \times & \times & 2! & 0 & \vdots\\ \vdots & & \ddots & \ddots & 0\\ \times & \times & \ldots & \times & (m-1)!\end{vmatrix}=\prod_{k=0}^{m-1}k!\neq 0. \quad \square$$

EXAMPLE 3.3. Find the general solution of

$$(D^4-13D^2+36I)y=0.$$

SOLUTION. The characteristic polynomial is easily factored,

$$\begin{aligned}\lambda^4-13\lambda^2+36 &= (\lambda^2-9)(\lambda^2-4)\\ &= (\lambda+3)(\lambda-3)(\lambda+2)(\lambda-2).\end{aligned}$$

Hence,

$$y(x)=c_1\,e^{-3x}+c_2\,e^{3x}+c_3\,e^{-2x}+c_4\,e^{2x}.$$

The Matlab polynomial solver.— To find the zeros of the characteristic polynomial

$$\lambda^4 - 13\lambda^2 + 36$$

with Matlab, one represents the polynomial by the vector of its coefficients,

$$p = \begin{bmatrix} 1 & 0 & -13 & 0 & 36 \end{bmatrix}$$

and uses the command `roots` on p.

```
>> p = [1 0 -13 0 36]
 p =  1     0   -13     0    36
>> r = roots(p)
 r =
    3.0000
   -3.0000
    2.0000
   -2.0000
```

In fact the command `roots` constructs a matrix C of p (see the proof of Theorem 3.2)

$$C = \begin{bmatrix} 0 & 13 & 0 & -36 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and uses the QR algorithm to find the eigenvalues of C which, in fact, are the zeros of p.

```
>> p = [1 0 -13 0 36];
>> C = compan(p)
 C =
     0    13     0   -36
     1     0     0     0
     0     1     0     0
     0     0     1     0
>> eigenvalues = eig(C)'
 eigenvalues = 3.0000   -3.0000    2.0000   -2.0000
```

□

EXAMPLE 3.4. Find the general solution of the differential equation

$$(D - I)^3 y = 0.$$

SOLUTION. The characteristic polynomial $(\lambda - 1)^3$ admits a triple zero:

$$\lambda_1 = \lambda_2 = \lambda_3 = 1,$$

Hence:

$$y(x) = c_1\, e^x + c_2 x\, e^x + c_3 x^2\, e^x. \quad \square$$

If the characteristic equation (3.13) has complex roots, say $\lambda_{1,2} = \alpha \pm i\beta$, then we have two independent solutions

$$y_1(x) = e^{\alpha x} \cos(\beta x), \qquad y_2(x) = e^{\alpha x} \sin(\beta x),$$

as in Chapter 2.

If the order of the differential equation is $n \geq 4$, there are other possibilities for complex roots. Suppose that there are two distinct conjugate pairs, i.e. $\lambda_{1,2} = \alpha_1 \pm i\beta_1$ and $\lambda_{3,4} = \alpha_2 \pm i\beta_2$ where $(\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)$, then we have the independent solutions

$$y_1(x) = e^{\alpha_1 x}\cos(\beta_1 x), \qquad y_2(x) = e^{\alpha_1 x}\sin(\beta_1 x),$$
$$y_3(x) = e^{\alpha_2 x}\cos(\beta_2 x), \qquad y_4(x) = e^{\alpha_2 x}\sin(\beta_2 x).$$

But we could also have repeated complex roots. Suppose that $\lambda_1 = \lambda_2 = \alpha + i\beta$ and $\lambda_3 = \lambda_4 = \alpha - i\beta$. Then we have the independent solutions

$$y_1(x) = e^{\alpha x}\cos(\beta x), \qquad y_2(x) = e^{\alpha x}\sin(\beta x),$$
$$y_3(x) = x\,e^{\alpha x}\cos(\beta x), \qquad y_4(x) = x\,e^{\alpha x}\sin(\beta x),$$

but the order in which we write them does not matter.

EXAMPLE 3.5. Find the general solution of the differential equation

$$y^{(4)} + 13y'' + 36y = 0.$$

SOLUTION. The characteristic equation is

$$\lambda^4 + 13\lambda^2 + 36 = (\lambda^2 + 4)(\lambda^2 + 9) = 0.$$

So the roots are $\lambda_{1,2} = \pm 2i$ and $\lambda_{3,4} = \pm 3i$, and the general solution is

$$y(x) = c_1\cos(2x) + c_2\sin(2x) + c_3\cos(3x) + c_4\sin(3x). \quad \square$$

EXAMPLE 3.6. Find the general solution of the differential equation

$$y^{(4)} + 8y''' + 26y'' + 40y' + 25y = 0.$$

SOLUTION. The characteristic equation is

$$\lambda^4 + 8\lambda^3 + 26\lambda^2 + 40\lambda + 25 = (\lambda^2 + 4\lambda + 5)^2 = 0.$$

So the roots are $\lambda_{1,2} = -2 + i$ and $\lambda_{3,4} = -2 - i$, and the general solution is

$$y(x) = c_1\,e^{-2x}\cos x + c_2\,e^{-2x}\sin x + c_3 x\,e^{-2x}\cos x + c_4 x\,e^{-2x}\sin x. \quad \square$$

EXAMPLE 3.7. Find the general solution of the Euler–Cauchy equation

$$x^3y''' - 3x^2y'' + 6xy' - 6y = 0.$$

SOLUTION. **(a) The analytic solution.**— Putting

$$y(x) = x^m$$

in the differential equation, we have

$$m(m-1)(m-2)x^m - 3m(m-1)x^m + 6mx^m - 6x^m = 0,$$

and dividing by x^m, we obtain the characteristic equation,

$$m(m-1)(m-2) - 3m(m-1) + 6m - 6 = 0.$$

Noting that $m - 1$ is a common factor, we have

$$\begin{aligned}(m-1)[m(m-2) - 3m + 6] &= (m-1)(m^2 - 5m + 6)\\ &= (m-1)(m-2)(m-3) = 0.\end{aligned}$$

Thus,

$$y(x) = c_1\,x + c_2\,x^2 + c_3\,x^3.$$

(b) The Matlab symbolic solution.—

```
dsolve('x^3*D3y-3*x^2*D2y+6*x*Dy-6*y=0','x')
y = C1*x+C2*x^2+C3*x^3
```

□

If $m_1 = m_2 = m$ is a double root of the characteristic equation of an Euler–Cauchy equation, we have from Chapter 2 that $y_1(x) = x^m$ and $y_2(x) = x^m \ln x$ are independent solutions.

Suppose that $m_1 = m_2 = m_3 = m$ is a triple root, then we have three independent solutions

$$y_1(x) = x^m, \quad y_2(x) = x^m \ln x, \quad y_3(x) = x^m (\ln x)^2.$$

EXAMPLE 3.8. Find the general solution of the Euler–Cauchy equation

$$x^3 y''' + 6x^2 y'' + 7xy' - y = 0.$$

SOLUTION. The characteristic equation is

$$m(m-1)(m-2) + 6m(m-1) + 7m + 1 = m^3 + 3m^2 + 3m + 1 = (m+1)^3 = 0,$$

and so the general solution is

$$y(x) = c_1 x^{-1} + c_2 x^{-1} \ln x + c_3 x^{-1} (\ln x)^2. \quad \square$$

3.3. Linear Nonhomogeneous Equations

Consider the linear *nonhomogeneous* differential equation of order n,

$$Ly := y^{(n)} + a_{n-1}(x) y^{(n-1)} + \cdots + a_1(x) y' + a_0(x) y = r(x). \tag{3.17}$$

Let

$$y_h(x) = c_1 y_1(x) + c_2 y_2(x) + \cdots + c_n y_n(x), \tag{3.18}$$

be a general solution of the corresponding homogeneous equation

$$Ly = 0.$$

Moreover, let $y_p(x)$ be a *particular solution* of the nonhomogeneous equation (3.17). Then,

$$y_g(x) = y_h(x) + y_p(x)$$

is a general solution of (3.17). In fact,

$$Ly_g = Ly_h + Ly_p = 0 + r(x).$$

As $y_g(x)$ is a solution of the nonhomogeneous equation (3.17) which contains n arbitrary constants, it must be the general solution.

EXAMPLE 3.9. Find a general solution $y_g(x)$ of

$$y'' - y = 3\, e^{2x}$$

if

$$y_p(x) = e^{2x}$$

is a particular solution.

SOLUTION. **(a) The analytic solution.**— It is easy to see that e^{2x} is a particular solution. Since

$$y'' - y = 0 \implies \lambda^2 - 1 = 0 \implies \lambda = \pm 1,$$

a general solution to the homogeneous equation is

$$y_h(x) = c_1\, e^x + c_2\, e^{-x}$$

and a general solution of the nonhomogeneous equation is

$$y_g(x) = c_1\, e^x + c_2\, e^{-x} + e^{2x}.$$

(b) The Matlab symbolic solution.—

```
dsolve('D2y-y-3*exp(2*x)','x')
y = (exp(2*x)*exp(x)+C1*exp(x)^2+C2)/exp(x)
z = expand(y)
z = exp(x)^2+exp(x)*C1+1/exp(x)*C2
```

□

Here is a second method for solving linear first-order differential equations treated in Section 1.6.

EXAMPLE 3.10. Find the general solution of the first-order linear nonhomogeneous equation

$$Ly := y' + f(x)y = r(x). \tag{3.19}$$

SOLUTION. The homogeneous equation $Ly = 0$ is separable:

$$\frac{dy}{y} = -f(x)\,dx \implies \ln|y| = -\int f(x)\,dx \implies y_h(x) = e^{-\int f(x)\,dx}.$$

No arbitrary constant is needed here. To find a particular solution by variation of parameters we put

$$y_p(x) = u(x)y_h(x)$$

in the nonhomogeneous equation $Ly = r(x)$:

$$\begin{aligned} y_p' &= uy_h' + u'y_h \\ f(x)y_p &= uf(x)y_h. \end{aligned}$$

Adding the left- and right-hand sides of these expressions we have

$$\begin{aligned} Ly_p &= uLy_h + u'y_h \\ &= u'y_h \\ &= r(x). \end{aligned}$$

Since the differential equation $u'y_h = r$ is separable,

$$du = e^{\int f(x)\,dx} r(x)\,dx,$$

it can be integrated directly,

$$u(x) = \int e^{\int f(x)\,dx} r(x)\,dx.$$

No arbitrary constant is needed here. Thus,

$$y_p(x) = e^{-\int f(x)\,dx} \int e^{\int f(x)\,dx} r(x)\,dx.$$

Hence, the general solution of (3.19) is

$$\begin{aligned} y(x) &= cy_h(x) + y_p(x) \\ &= e^{-\int f(x)\,dx}\left[\int e^{\int f(x)\,dx} r(x)\,dx + c\right], \end{aligned}$$

which agrees with what we saw in Chapter 1. □

In the next two sections we present two methods to find particular solutions, namely, the Method of Undetermined Coefficients and the Method of Variation of Parameters. The first method, which is more restrictive than the second, does not always require the general solution of the homogeneous equation, but the second always does.

3.4. Method of Undetermined Coefficients

Consider the linear nonhomogeneous differential equation of order n,

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1 y' + a_0 y = r(x), \tag{3.20}$$

with *constant coefficients*, $a_0, a_1, \ldots, a_{n-1}$.

If the dimension of the space spanned by the derivatives of the functions on the right-hand side of (3.20) is *finite*, we can use the Method of Undetermined Coefficients.

Here is a list of usual functions $r(x)$ which have a finite number of linearly independent derivatives. We indicate the dimension of the space of derivatives.

$$\begin{aligned} r(x) &= x^2 + 2x + 1, \qquad r'(x) = 2x + 2, \qquad r''(x) = 2, \\ r^{(k)}(x) &= 0,\ k = 3, 4, \ldots, \qquad \Longrightarrow \text{ dim. } = 3; \\ r(x) &= \cos 2x + \sin 2x, \qquad r'(x) = -2\sin 2x + 2\cos 2x, \\ r''(x) &= -4r(x), \qquad \Longrightarrow \text{ dim. } = 2; \\ r(x) &= x\,e^x, \qquad r'(x) = e^x + x\,e^x, \\ r''(x) &= 2r'(x) - r(x), \qquad \Longrightarrow \text{ dim. } = 2. \end{aligned}$$

More specifically, the functions that have a finite number of independent derivatives are polynomials, exponentials, sine, cosine, hyperbolic sine, hyperbolic cosine, and sums and products of them.

The Method of Undetermined Coefficients consists in choosing for a particular solution a linear combination,

$$y_p(x) = c_1 p_1(x) + c_2 p_2(x) + \cdots + c_l p_l(x), \tag{3.21}$$

of the independent derivatives of the function $r(x)$ on the right-hand side. We determine the coefficients c_k by substituting $y_p(x)$ in (3.20) and equating coefficients. A bad choice or a mistake leads to a contradiction.

Example 3.11. Find a general solution $y_g(x)$ of

$$Ly := y'' + y = 3x^2$$

by the Method of Undetermined Coefficients.

SOLUTION. **(a) The analytic solution.**— Since $r(x) = 3x^2$ is a quadratic, put

$$y_p(x) = ax^2 + bx + c$$

in the differential equation and add the terms on the left- and the right-hand sides, respectively,

$$\begin{aligned} y_p &= ax^2 + bx + c \\ y_p'' &= 2a \\ Ly_p &= ax^2 + bx + (2a + c) \\ &= 3x^2. \end{aligned}$$

Identifying the coefficients of 1, x and x^2 on both sides, we have

$$a = 3, \qquad b = 0, \qquad c = -2a = -6.$$

The general solution of $Ly = 0$ is

$$y_h(x) = A\cos x + B\sin x.$$

Hence, the general solution of $Ly = 3x^2$ is

$$y_g(x) = A\cos x + B\sin x + 3x^2 - 6.$$

(b) The Matlab symbolic solution.—

```
dsolve('D2y+y=3*x^2','x')
y = -6+3*x^2+C1*sin(x)+C2*cos(x)
```

□

Important remark. If for a chosen term $p_j(x)$ in (3.21), $x^k p_j(x)$ is a solution of the homogeneous equation, but $x^{k+1}p_j(x)$ is not, then $p_j(x)$ must be replaced by $x^{k+1}p_j(x)$. Naturally, we exclude from y_p the terms which are in the space of solution of the homogeneous equation since they contribute zero to the right-hand side.

EXAMPLE 3.12. Find the form of a particular solution for solving the equation

$$y'' - 4y' + 4y = 3\,e^{2x} + 32\sin x$$

by the Method of Undetermined Coefficients.

SOLUTION. Since the general solution of the homogeneous equation is

$$y_h(x) = c_1\,e^{2x} + c_2 x\,e^{2x},$$

a particular solution is of the form

$$y_p(x) = ax^2\,e^{2x} + b\cos x + c\sin x.$$

Since $r(x) = 3\,e^{2x} + 32\sin x$, the exponential part $3\,e^{2x}$ would contribute $a\,e^{2x}$, but this and $x\,e^{2x}$ appear in $y_h(x)$, so we have $ax^2\,e^{2x}$ and the trigonometric part $32\sin x$ contributes $b\cos x$ and $c\sin x$. We must be careful to note that whenever sine or cosine appears in $r(x)$, we shall have both of them in $y_p(x)$. □

EXAMPLE 3.13. Solve the initial value problem

$$y''' - y' = 4\,e^{-x} + 3\,e^{2x}, \qquad y(0) = 0, \quad y'(0) = -1, \quad y''(0) = 2$$

and plot the solution.

SOLUTION. **(a) The analytic solution.**— The characteristic equation is

$$\lambda^3 - \lambda = \lambda(\lambda - 1)(\lambda + 1) = 0.$$

The general solution of the homogeneous equation is

$$y_h(x) = c_1 + c_2\, e^{x} + c_3\, e^{-x}.$$

Considering that e^{-x} is contained in the right-hand side of the differential equation, we take a particular solution of the form

$$y_p(x) = ax\, e^{-x} + b\, e^{2x}.$$

Then,

$$\begin{aligned} y_p'(x) &= -ax\, e^{-x} + a\, e^{-x} + 2b\, e^{2x}, \\ y_p''(x) &= \quad ax\, e^{-x} - 2a\, e^{-x} + 4b\, e^{2x}, \\ y_p'''(x) &= -ax\, e^{-x} + 3a\, e^{-x} + 8b\, e^{2x}. \end{aligned}$$

Hence,

$$\begin{aligned} y_p'''(x) - y_p'(x) &= 2a\, e^{-x} + 6b\, e^{2x} \\ &= 4\, e^{-x} + 3\, e^{2x}, \qquad \text{for all} \quad x. \end{aligned}$$

Identifying the coefficients of e^{-x} and e^{2x}, we have

$$a = 2, \qquad b = \frac{1}{2}.$$

Thus, a particular solution of the nonhomogeneous equation is

$$y_p(x) = 2x\, e^{-x} + \frac{1}{2}\, e^{2x}$$

and the general solution of the nonhomogeneous equation is

$$y(x) = c_1 + c_2\, e^{x} + c_3\, e^{-x} + 2x\, e^{-x} + \frac{1}{2}\, e^{2x}.$$

The arbitrary constants c_1, c_2 and c_3 are determined by the initial conditions:

$$\begin{aligned} y(0) &= c_1 + c_2 + c_3 + \frac{1}{2} = \quad 0, \\ y'(0) &= \qquad c_2 - c_3 \; + 3 = -1, \\ y''(0) &= \qquad c_2 + c_3 \; - 2 = \quad 2, \end{aligned}$$

yielding the linear algebraic system

$$\begin{aligned} c_1 + c_2 + c_3 &= -\frac{1}{2}, \\ c_2 - c_3 &= -4, \\ c_2 + c_3 &= \quad 4, \end{aligned}$$

whose solution is

$$c_1 = -\frac{9}{2}, \qquad c_2 = 0, \qquad c_3 = 4,$$

and the unique solution is

$$y(x) = -\frac{9}{2} + 4\, e^{-x} + 2x\, e^{-x} + \frac{1}{2}\, e^{2x}.$$

(b) The Matlab symbolic solution.—

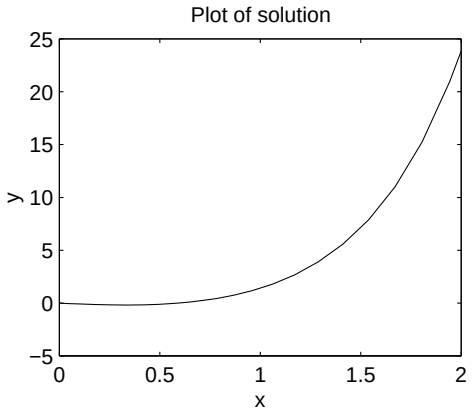

FIGURE 3.1. Graph of solution of the linear equation in Example 3.13.

```
dsolve('D3y-Dy=4*exp(-x)+3*exp(2*x)','y(0)=0','Dy(0)=-1','D2y(0)=2','x')
y = 1/2*(8+exp(3*x)+4*x-9*exp(x))/exp(x)
z =expand(y)
z = 4/exp(x)+1/2*exp(x)^2+2/exp(x)*x-9/2
```

(c) The Matlab numeric solution.— To rewrite the third-order differential equation as a system of first-order equations, we put

$$\begin{aligned} y(1) &= y, \\ y(2) &= y', \\ y(3) &= y''. \end{aligned}$$

Thus, we have

$$\begin{aligned} y(1)' &= y(2), \\ y(2)' &= y(3), \\ y(3)' &= y(2) + 4 * \exp(-x) + 3 * \exp(2 * x). \end{aligned}$$

The M-file `exp39.m`:

```
function yprime = exp39(x,y);
yprime=[y(2); y(3); y(2)+4*exp(-x)+3*exp(2*x)];
```

The call to the ode23 solver and the plot command:

```
xspan = [0 2]; % solution for x=0 to x=2
y0 = [0;-1;2]; % initial conditions
[x,y] = ode23('exp39',xspan,y0);
plot(x,y(:,1))
```

The numerical solution is plotted in Fig. 3.1. □

3.5. Particular Solution by Variation of Parameters

Consider the linear *nonhomogeneous* differential equation of order n,

$$Ly := y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = r(x), \tag{3.22}$$

in *standard form*, that is, the coefficient of $y^{(n)}$ is equal to 1.

Let

$$y_h(x) = c_1y_1(x) + c_2y_2(x) + \cdots + c_ny_n(x), \tag{3.23}$$

be a general solution of the corresponding homogeneous equation

$$Ly = 0.$$

For simplicity, we derive the Method of Variation of Parameters in the case $n = 3$,

$$Ly := y''' + a_2(x)y'' + a_1(x)y' + a_0(x)y = r(x), \tag{3.24}$$

The general case follows in the same way.

Following an idea due to Lagrange, we take a particular solution of the form

$$y_p(x) = c_1(x)y_1(x) + c_2(x)y_2(x) + c_3(x)y_3(x), \tag{3.25}$$

where we let the parameters c_1, c_2 and c_3 of the general solution y_h vary, thus giving us three degrees of freedom.

We differentiate $y_p(x)$:

$$\begin{aligned} y_p'(x) &= \big[c_1'(x)y_1(x) + c_2'(x)y_2(x) + c_3'(x)y_3(x)\big] \\ &\quad + c_1(x)y_1'(x) + c_2(x)y_2'(x) + c_3(x)y_3'(x) \\ &= c_1(x)y_1'(x) + c_2(x)y_2'(x) + c_3(x)y_3'(x), \end{aligned}$$

where, using one degree of freedom, we let the term in square brackets be zero,

$$c_1'(x)y_1(x) + c_2'(x)y_2(x) + c_3'(x)y_3(x) = 0. \tag{3.26}$$

We differentiate $y_p'(x)$:

$$\begin{aligned} y_p''(x) &= \big[c_1'(x)y_1'(x) + c_2'(x)y_2'(x) + c_3'(x)y_3'(x)\big] \\ &\quad + c_1(x)y_1''(x) + c_2(x)y_2''(x) + c_3(x)y_3''(x) \\ &= c_1(x)y_1''(x) + c_2(x)y_2''(x) + c_3(x)y_3''(x), \end{aligned}$$

where, using another degree of freedom, we let the term in square brackets be zero,

$$c_1'(x)y_1'(x) + c_2'(x)y_2'(x) + c_3'(x)y_3'(x) = 0. \tag{3.27}$$

Lastly, we differentiate $y_p''(x)$:

$$\begin{aligned} y_p'''(x) &= \big[c_1'(x)y_1''(x) + c_2'(x)y_2''(x) + c_3'(x)y_3''(x)\big] \\ &\quad + \big[c_1(x)y_1'''(x) + c_2(x)y_2'''(x) + c_3(x)y_3'''(x)\big]. \end{aligned}$$

Using the expressions obtained for y_p, y_p', y_p'' and y_p''', we have

$$\begin{aligned} Ly_p &= y_p''' + a_2y_p'' + a_1y_p' + a_0y_p \\ &= c_1'y_1'' + c_2'y_2'' + c_3'y_3'' + \big[c_1Ly_1 + c_2Ly_2 + c_3Ly_3\big] \\ &= c_1'y_1'' + c_2'y_2'' + c_3'y_3'' \\ &= r(x), \end{aligned}$$

since y_1, y_2 and y_3 are solutions of $Ly = 0$ and hence the term in square brackets is zero. Moreover, we want y_p to satisfy $Ly_p = r(x)$, which we can do by our third and last degree of freedom. Hence we have

$$c_1' y_1'' + c_2' y_2'' + c_3' y_3'' = r(x). \tag{3.28}$$

We rewrite the three equations (3.26)–(3.28) in the unknowns $c_1'(x)$, $c_2'(x)$ and $c_3'(x)$ in matrix form,

$$\begin{bmatrix} y_1(x) & y_2(x) & y_3(x) \\ y_1'(x) & y_2'(x) & y_3'(x) \\ y_1''(x) & y_2''(x) & y_3''(x) \end{bmatrix} \begin{bmatrix} c_1'(x) \\ c_2'(x) \\ c_3'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ r(x) \end{bmatrix}, \tag{3.29}$$

that is,

$$M(x)\boldsymbol{c}'(x) = \begin{bmatrix} 0 \\ 0 \\ r(x) \end{bmatrix}.$$

Since y_1, y_2 and y_3 form a fundamental system, by Corollary 3.1 their Wronskian does not vanish,

$$W(y_1, y_2, y_3) = \det M \neq 0.$$

We solve the linear system for $\boldsymbol{c}'(x)$ and integrate the solution

$$\boldsymbol{c}(x) = \int \boldsymbol{c}'(x)\, dx.$$

No constants of integrations are needed here since the general solution will contain three arbitrary constants. The general solution is (3.24) is

$$y_g(x) = Ay_1 + By_2 + Cy_3 + c_1(x)y_1 + c_2(x)y_2 + c_3(x)y_3. \tag{3.30}$$

Because of the particular form of the right-hand side of system (3.29), Cramer's rule leads to nice formulae for the solution of this system in two and three dimensions. In 2D, we have

$$y_p(x) = -y_1(x) \int \frac{y_2(x)r(x)}{W(x)}\, dx + y_2(x) \int \frac{y_1(x)r(x)}{W(x)}\, dx. \tag{3.31}$$

In 3D, solve (3.29) for c_1', c_2' and c_3' by Cramer's rule:

$$c_1'(x) = \frac{r}{W} \begin{vmatrix} y_2 & y_3 \\ y_2' & y_3' \end{vmatrix}, \quad c_2'(x) = -\frac{r}{W} \begin{vmatrix} y_1 & y_3 \\ y_1' & y_3' \end{vmatrix}, \quad c_3'(x) = \frac{r}{W} \begin{vmatrix} y_1 & y_2 \\ y_1' & y_1' \end{vmatrix}, \tag{3.32}$$

integrate the c_i' with respect to x and form $y_p(x)$ as in (3.25).

REMARK 3.4. If the coefficient $a_n(x)$ of $y^{(n)}$ is not equal to 1, we must divide the right-hand side of (3.29) by $a_n(x)$, that is, replace $r(x)$ by $r(x)/a_n(x)$. This is important to remember when solving Euler–Cauchy equations which are usually not written in standard form.

EXAMPLE 3.14. Find the general solution of the differential equation

$$y'' + y = \sec x \tan x,$$

by the Method of Variation of Parameters.

SOLUTION. **(a) The analytic solution.**— We have that the general solution of the homogeneous equation $y'' + y = 0$ is

$$y_h(x) = c_1 \cos x + c_2 \sin x.$$

Since the term on the right-hand side, $r(x) = \sec x \tan x$, does not have a finite number of independent derivatives, we must use Variation of Parameters, i.e. we must solve

$$\begin{aligned} c_1'(x)y_1 + c_2'(x)y_2 &= 0, \\ c_1'(x)y_1' + c_2'(x)y_2' &= r, \end{aligned}$$

or

$$\begin{aligned} c_1'(x)\cos x + c_2' \sin x &= 0, && \text{(a)} \\ -c_1'(x)\sin x + c_2' \cos x &= \sec x \tan x. && \text{(b)} \end{aligned}$$

Multiply (a) by $\sin x$ and (b) by $\cos x$ to get

$$\begin{aligned} c_1'(x)\sin x \cos x + c_2' \sin^2 x &= 0, && \text{(c)} \\ -c_1'(x)\sin x \cos x + c_2' \cos^2 x &= \tan x. && \text{(d)} \end{aligned}$$

Then (c)+(d) gives

$$c_2' = \tan x \Longrightarrow c_2(x) = -\ln|\cos x| = \ln|\sec x|,$$

and then from (c)

$$c_1' = -\frac{c_2' \sin x}{\cos x} = -\tan^2 x = 1 - \sec^2 x \Longrightarrow c_1(x) = x - \tan x.$$

So the particular solution is

$$\begin{aligned} y_p(x) &= c_1(x)\cos x + c_2(x)\sin x \\ &= (x - \tan x)\cos x + (\ln|\sec x|)\sin x. \end{aligned}$$

Finally, the general solution is

$$\begin{aligned} y(x) &= y_h(x) + y_p(x) \\ &= A\cos x + B\sin x + (x - \tan x)\cos x + (\ln|\sec x|)\sin x. \end{aligned}$$

Note that the term $-\tan x \cos x = -\sin x$ can be absorbed in y_h.

(b) The Matlab symbolic solution.—

```
dsolve('D2y+y=sec(x)*tan(x)','x')
y = -log(cos(x))*sin(x)-sin(x)+x*cos(x)+C1*sin(x)+C2*cos(x)
```

□

EXAMPLE 3.15. Find the general solution of the differential equation

$$y''' - y' = \cosh x$$

by the Method of Variation of Parameters.

SOLUTION. The characteristic equation is

$$\lambda^3 - \lambda = \lambda(\lambda^2 - 1) = 0 \implies \lambda_1 = 0,\ \lambda_2 = 1,\ \lambda_3 = -1.$$

The general solution of the homogeneous equation is

$$y_h(x) = c_1 + c_2\, e^x + c_3\, e^{-x}.$$

By the Method of Variation of Parameters, the particular solution of the nonhomogeneous equation is

$$y_p(x) = c_1(x) + c_2(x)\, e^x + c_3(x)\, e^{-x}.$$

Thus, we have the system

$$\begin{bmatrix} 1 & e^x & e^{-x} \\ 0 & e^x & -e^{-x} \\ 0 & e^x & e^{-x} \end{bmatrix} \begin{bmatrix} c_1'(x) \\ c_2'(x) \\ c_3'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cosh x \end{bmatrix}.$$

We solve this system by Gaussian elimination:

$$\begin{bmatrix} 1 & e^x & e^{-x} \\ 0 & e^x & -e^{-x} \\ 0 & 0 & 2e^{-x} \end{bmatrix} \begin{bmatrix} c_1'(x) \\ c_2'(x) \\ c_3'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cosh x \end{bmatrix}.$$

Hence

$$\begin{aligned} c_3' &= \frac{1}{2}\, e^x \cosh x = \frac{1}{2}\, e^x \left(\frac{e^x + e^{-x}}{2} \right) = \frac{1}{4} \left(e^{2x} + 1 \right), \\ c_2' &= e^{-2x} c_3' = \frac{1}{4} \left(1 + e^{-2x} \right), \\ c_1' &= -e^x c_2' - e^{-x} c_3' = -\frac{1}{2} \left(e^x + e^{-x} \right) = -\cosh x, \end{aligned}$$

and after integration, we have

$$\begin{aligned} c_1 &= -\sinh x \\ c_2 &= \frac{1}{4} \left(x - \frac{1}{2}\, e^{-2x} \right) \\ c_3 &= \frac{1}{4} \left(\frac{1}{2}\, e^{2x} + x \right). \end{aligned}$$

The particular solution is

$$\begin{aligned} y_p(x) &= -\sinh x + \frac{1}{4} \left(x\, e^x - \frac{1}{2}\, e^{-x} \right) + \frac{1}{4} \left(\frac{1}{2}\, e^x + x\, e^{-x} \right) \\ &= -\sinh x + \frac{1}{4} x \left(e^x + e^{-x} \right) + \frac{1}{8} \left(e^x - e^{-x} \right) \\ &= \frac{1}{2} x \cosh x - \frac{3}{4} \sinh x. \end{aligned}$$

The general solution of the nonhomogeneous equation is

$$\begin{aligned} y_g(x) &= A + B'\, e^x + C'\, e^{-x} + \frac{1}{2}\, x \cosh x - \frac{3}{4} \sinh x \\ &= A + B\, e^x + C\, e^{-x} + \frac{1}{2}\, x \cosh x, \end{aligned}$$

where we have used the fact that the function

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

is already contained in the general solution y_h of the homogeneous equation. Symbolic Matlab does not produce a general solution in such a simple form. □

If one uses the Method of Undetermined Coefficients to solve this problem, one has to take a particular solution of the form

$$y_p(x) = ax\cosh x + bx\sinh x,$$

since $\cosh x$ and $\sinh x$ are linear combinations of e^x and e^{-x} which are solutions of the homogeneous equation. In fact, putting

$$y_p(x) = ax\cosh x + bx\sinh x$$

in the equation $y''' - y' = \cosh x$, we obtain

$$\begin{aligned} y_p''' - y_p' &= 2a\cosh x + 2b\sinh x \\ &= \cosh x, \end{aligned}$$

whence

$$a = \frac{1}{2} \quad \text{and} \quad b = 0.$$

EXAMPLE 3.16. Find the general solution of the differential equation

$$Ly := y'' + 3y' + 2y = \frac{1}{1+e^x}.$$

SOLUTION. Since the dimension of the space of derivatives of the right-hand side is infinite, one has to use the Method of Variation of Parameters.

It is to be noted that the symbolic Matlab command `dsolve` produces a several-line-long solution that is unusable. We therefore follow the theoretical method of Lagrange but do the simple algebraic and calculus manipulations by symbolic Matlab.

The characteristic polynomial of the homogeneous equation $Ly = 0$ is

$$\lambda^2 + 3\lambda + 2 = (\lambda+1)(\lambda+2) = 0 \implies \lambda_1 = -1, \quad \lambda_2 = -2.$$

Hence, the general solution $y_h(x)$ to $Ly = 0$ is

$$y_h(x) = c_1\, e^{-x} + c_2\, e^{-2x}.$$

By the Method of Variation of Parameters, a particular solution of the inhomogeneous equation is searched in the form

$$y_p(x) = c_1(x)\, e^{-x} + c_2(x)\, e^{-2x}.$$

The functions $c_1(x)$ and $c_2(x)$ are the integrals of the solutions $c_1'(x)$ and $c_2'(x)$ of the algebraic system $A\boldsymbol{c}' = \boldsymbol{b}$,

$$\begin{bmatrix} e^{-x} & e^{-2x} \\ -e^{-x} & -2\,e^{-2x} \end{bmatrix} \begin{bmatrix} c_1' \\ c_2' \end{bmatrix} = \begin{bmatrix} 0 \\ 1/(1+e^x) \end{bmatrix}.$$

We use symbolic Matlab to solve this simple system.

```
>> clear
>> syms x real; syms c dc A b yp;
>> A = [exp(-x) exp(-2*x); -exp(-x) -2*exp(-2*x)];
>> b=[0 1/(1+exp(x))]';
>> dc = A\b % solve for c'(x)
dc =
 [    1/exp(-x)/(1+exp(x))]
 [ -1/exp(-2*x)/(1+exp(x))]
>> c = int(dc) % get c(x) by integrating c'(x)
c =
 [         log(1+exp(x))]
 [ -exp(x)+log(1+exp(x))]
>>  yp=c'*[exp(-x) exp(-2*x)]'
 yp =
 log(1+exp(x))*exp(-x)+(-exp(x)+log(1+exp(x)))*exp(-2*x)
```

Since $-e^{-x}$ is contained in $y_h(x)$, the general solution of the inhomogeneous equation is

$$y(x) = A\,e^{-x} + B\,e^{-2x} + [\ln(1+e^x)]\;e^{-x} + [\ln(1+e^x)]\;e^{-2x}. \quad \square$$

EXAMPLE 3.17. Solve the nonhomogeneous Euler–Cauchy differential equation with given initial values:

$$Ly := 2x^2y'' + xy' - 3y = x^{-3}, \qquad y(1) = 0, \qquad y'(1) = 2.$$

SOLUTION. Putting $y = x^m$ in the homogeneous equation $Ly = 0$ we obtain the characteristic polynomial:

$$2m^2 - m - 3 = 0 \implies m_1 = \frac{3}{2}, \quad m_2 = -1.$$

Thus, general solution, $y_h(x)$, of $Ly = 0$ is

$$y_h(x) = c_1x^{3/2} + c_2x^{-1}.$$

To find a particular solution, $y_p(x)$, to the nonhomogeneous equation, we use the Method of Variation of Parameters since the dimension of the space of derivatives of the right-hand side is infinite and the left-hand side is Euler–Cauchy. We put

$$y_p(x) = c_1(x)x^{3/2} + c_2(x)x^{-1}.$$

We need to solve the linear system

$$\begin{bmatrix} x^{3/2} & x^{-1} \\ \frac{3}{2}\,x^{1/2} & -x^{-2} \end{bmatrix} \begin{bmatrix} c_1' \\ c_2' \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2}\,x^{-5} \end{bmatrix},$$

where the right-hand side of the linear system has been divided by the coefficient $2x^2$ of y'' to have the equation in standard form with the coefficient of y'' equal to 1. Solving this system for c_1' and c_2', we obtain

$$c_1' = \frac{1}{5}\,x^{-11/2}, \qquad c_2' = -\frac{1}{5}\,x^{-3}.$$

Thus, after integration,

$$c_1(x) = -\frac{2}{45}\,x^{-9/2}, \qquad c_2(x) = \frac{1}{10}\,x^{-2},$$

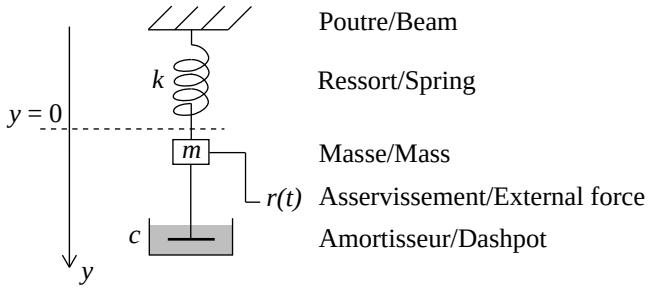

FIGURE 3.2. Forced damped system.

and the general solution is

$$\begin{aligned} y(x) &= Ax^{3/2} + Bx^{-1} - \frac{2}{45}\,x^{-3} + \frac{1}{10}\,x^{-3} \\ &= Ax^{3/2} + Bx^{-1} + \frac{1}{18}\,x^{-3}. \end{aligned}$$

The constants A and B are uniquely determined by the initial conditions. For this we need the derivative, $y'(x)$, of $y(x)$,

$$y'(x) = \frac{3}{2}\,Ax^{1/2} - Bx^{-2} - \frac{1}{6}\,x^{-4}.$$

Thus,

$$\begin{aligned} y(1) &= A + B + \frac{1}{18} = 0, \\ y'(1) &= \frac{3}{2}\,A - B - \frac{1}{6} = 2. \end{aligned}$$

Solving for A and B, we have

$$A = \frac{38}{45}, \qquad B = -\frac{9}{10}.$$

The (unique) solution is

$$y(x) = \frac{38}{45}\,x^{3/2} - \frac{9}{10}\,x^{-1} + \frac{1}{18}\,x^{-3}. \quad \square$$

3.6. Forced Oscillations

We present two examples of forced vibrations of mechanical systems.

Consider a vertical spring attached to a rigid beam. The spring resists both extension and compression with Hooke's constant equal to k. Study the problem of the forced damped vertical oscillation of a mass of m kg which is attached at the lower end of the spring. (See Fig. 3.2). The damping constant is c and the external force is $r(t)$.

We refer to Example 2.5 for the derivation of the differential equation governing the nonforced system, and simply add the external force to the right-hand side,

$$y'' + \frac{c}{m}\,y' + \frac{k}{m}\,y = \frac{1}{m}\,r(t).$$

EXAMPLE 3.18 (Forced oscillation without resonance). Solve the initial value problem with external force

$$Ly := y'' + 9y = 8\sin t, \qquad y(0) = 1, \qquad y'(0) = 1,$$

and plot the solution.

SOLUTION. **(a) The analytic solution.**— The general solution of $Ly = 0$ is

$$y_h(t) = A\cos 3t + B\sin 3t.$$

Following the Method of Undetermined Coefficients, we choose y_p of the form

$$y_p(t) = a\cos t + b\sin t.$$

Substituting this in $Ly = 8\sin t$ we obtain

$$\begin{aligned} y_p'' + 9y_p &= (-a + 9a)\cos t + (-b + 9b)\sin t \\ &= 8\sin t. \end{aligned}$$

Identifying coefficients on both sides, we have

$$a = 0, \qquad b = 1.$$

The general solution of $Ly = 8\sin t$ is

$$y(t) = A\cos 3t + B\sin 3t + \sin t.$$

We determine A and B by means of the initial conditions:

$$\begin{aligned} y(0) &= A = 1, \\ y'(t) &= -3A\sin 3t + 3B\cos 3t + \cos t, \\ y'(0) &= 3B + 1 = 1 \implies B = 0. \end{aligned}$$

The (unique) solution is

$$y(t) = \cos 3t + \sin t.$$

(b) The Matlab symbolic solution.—

```
dsolve('D2y+9*y=8*sin(t)','y(0)=1','Dy(0)=1','t')
y = sin(t)+cos(3*t)
```

(c) The Matlab numeric solution.— To rewrite the second-order differential equation as a system of first-order equations, we put

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

Thus, we have

$$\begin{aligned} y_1' &= y_2, \\ y_2' &= -9y_1 + 8\sin t. \end{aligned}$$

The M-file `exp312.m`:

```
function yprime = exp312(t,y);
yprime = [y(2); -9*y(1)+8*sin(t)];
```

The call to the ode23 solver and the plot command:

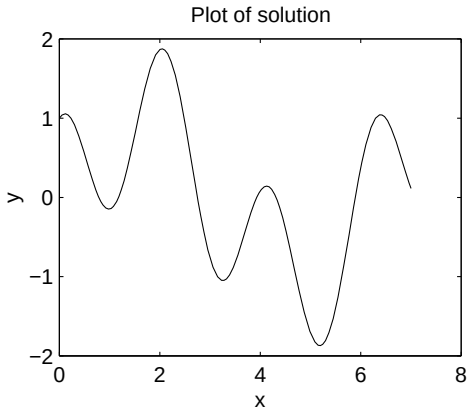

FIGURE 3.3. Graph of solution of the linear equation in Example 3.18.

```
tspan = [0 7]; % solution for t=0 to t=7
y0 = [1; 1]; % initial conditions
[x,y] = ode23('exp312',tspan,y0);
plot(x,y(:,1))
```

The numerical solution is plotted in Fig. 3.3. □

EXAMPLE 3.19 (Forced oscillation with resonance). Solve the initial value problem with external force

$$Ly := y'' + 9y = 6\sin 3t, \qquad y(0) = 1, \qquad y'(0) = 2,$$

and plot the solution.

SOLUTION. **(a) The analytic solution.**— The general solution of $Ly = 0$ is

$$y_h(t) = A\cos 3t + B\sin 3t.$$

Since the right-hand side of $Ly = 6\sin 3t$ is contained in the solution y_h, following the Method of Undetermined Coefficients, we choose y_p of the form

$$y_p(t) = at\cos 3t + bt\sin 3t.$$

Then we obtain

$$\begin{aligned} y_p'' + 9y_p &= -6a\sin 3t + 6b\cos 3t \\ &= 6\sin 3t. \end{aligned}$$

Identifying coefficients on both sides, we have

$$a = -1, \qquad b = 0.$$

The general solution of $Ly = 6\sin 3t$ is

$$y(t) = A\cos 3t + B\sin 3t - t\cos 3t.$$

We determine A and B by means of the initial conditions,

$$\begin{aligned} y(0) &= A = 1, \\ y'(t) &= -3A\sin 3t + 3B\cos 3t - \cos 3t + 3t\sin 3t, \\ y'(0) &= 3B - 1 = 2 \implies B = 1. \end{aligned}$$

The (unique) solution is

$$y(t) = \cos 3t + \sin 3t - t\cos 3t.$$

The term $-t\cos 3t$, whose amplitude is increasing, comes from the resonance of the system because the frequency of the external force coincides with the natural frequency of the system.

(b) The Matlab symbolic solution.—

```
dsolve('D2y+9*y=6*sin(3*t)','y(0)=1','Dy(0)=2','t')
y = sin(3*t)-cos(3*t)*t+cos(3*t)
```

(c) The Matlab numeric solution.— To rewrite the second-order differential equation as a system of first-order equations, we put

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

Thus, we have

$$\begin{aligned} y_1' &= y_2, \\ y_2' &= -9y_1 + 6\sin 3t. \end{aligned}$$

The M-file `exp313.m`:

```
function yprime = exp313(t,y);
yprime = [y(2); -9*y(1)+6*sin(3*t)];
```

The call to the ode23 solver and the plot command:

```
tspan = [0 7]; % solution for t=0 to t=7
y0 = [1; 1]; % initial conditions
[x,y] = ode23('exp313',tspan,y0);
plot(x,y(:,1))
```

The numerical solution is plotted in Fig. 3.4. □

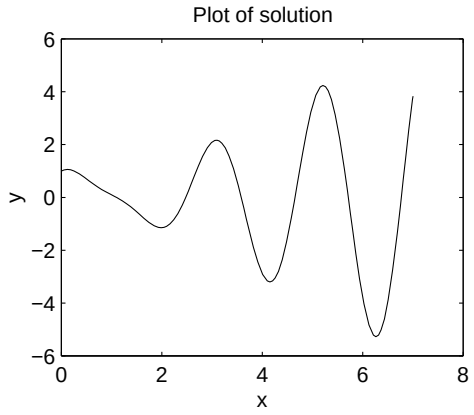

FIGURE 3.4. Graph of solution of the linear equation in Example 3.19.

CHAPTER 4

Systems of Differential Equations

4.1. Introduction

In Section 3.1, it was seen that a linear differential equation of order n,

$$y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = r(x),$$

can be written as a linear system of n first-order equations in the form

$$\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix}' = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ r(x) \end{bmatrix},$$

where the dependent variables are defined as

$$u_1 = y, \quad u_2 = y', \quad \ldots, \quad u_n = y^{(n-1)}.$$

In this case, the n initial values,

$$y(x_0) = k_1, \quad y'(x_0) = k_2, \quad \ldots, \quad y^{(n-1)}(x_0) = k_n,$$

and the right-hand side, $r(x)$, becomes

$$\begin{bmatrix} u_1(x_0) \\ u_2(x_0) \\ \vdots \\ u_{n-1}(x_0) \\ u_n(x_0) \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_{n-1} \\ k_n \end{bmatrix}, \qquad \boldsymbol{g}(x) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ r(x) \end{bmatrix},$$

respectively. In matrix and vector notation, this system is written as

$$\boldsymbol{u}'(x) = A(x)\boldsymbol{u}(x) + \boldsymbol{g}(x), \qquad \boldsymbol{u}(x_0) = \boldsymbol{k},$$

where the matrix $A(x)$ is a *companion matrix*.

If $\boldsymbol{g}(x) = \boldsymbol{0}$, the system is said to be *homogeneous*. If $\boldsymbol{g}(x) \neq \boldsymbol{0}$, it is *nonhomogeneous*.

EXAMPLE 4.1. Write the differential equation

$$y'' + 5y' - y = e^x \tag{4.1}$$

as a system of two equations.

SOLUTION. Let

$$u_1 = y, \qquad u_2 = y' = u_1'.$$

Then $y'' = u_2'$ and (4.1) becomes

$$u_2' + 5u_2 - u_1 = e^x,$$

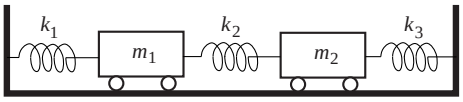

FIGURE 4.1. Mechanical system for Example 4.2.

or

$$u_2' = u_1 - 5u_2 + e^x.$$

If

$$\boldsymbol{u}(x) = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad \text{then} \quad \boldsymbol{u}'(x) = \begin{bmatrix} u_1' \\ u_2' \end{bmatrix} = \begin{bmatrix} u_2 \\ u_1 - 5u_2 + e^x \end{bmatrix}$$

or

$$\boldsymbol{u}'(x) = \begin{bmatrix} 0 & 1 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 0 \\ e^x \end{bmatrix} = A\boldsymbol{u}(x) + \boldsymbol{g}(x). \quad \square$$

In this chapter, we shall consider linear systems of n equations where the matrix $A(x)$ is a general $n \times n$ matrix, not necessarily of the form of a companion matrix. An example of such systems follows.

EXAMPLE 4.2. Set up a system of differential equations for the mechanical system shown in Fig. 4.1

SOLUTION. Consider a mechanical system in which two masses m_1 and m_2 are connected to each other by three springs as shown in Fig. 4.1 with Hooke's constants k_1, k_2 and k_3, respectively. Let $x_1(t)$ and $x_2(t)$ be the positions of the centers of mass of m_1 and m_2 away from their points of equilibrium, the positive x-direction pointing to the right. Then, $x_1''(t)$ and $x_2''(t)$ measure the acceleration of each mass. The resulting force acting on each mass is exerted on it by the springs that are attached to it, each force being proportional to the distance the spring is stretched or compressed. For instance, when mass m_1 has moved a distance x_1 to the right of its equilibrium position, the spring to the left of m_1 exerts a restoring force $-k_1x_1$ on this mass, attempting to return the mass back to its equilibrium position. The spring to the right of m_1 exerts a restoring force $-k_2(x_2 - x_1)$ on it; the part k_2x_1 reflects the compression of the middle spring due to the movement of m_1, while $-k_2x_2$ is due to the movement of m_2 and its influence on the same spring. Following Newton's Second Law of Motion, we arrive at the two coupled second-order equations:

$$m_1x_1'' = -k_1x_1 + k_2(x_2 - x_1), \qquad m_2x_2'' = -k_2(x_2 - x_1) - k_3x_2. \tag{4.2}$$

We convert each equation in (4.2) to a first-order system of equations by introducing two new variables y_1 and y_2 representing the velocities of each mass:

$$y_1 = x_1', \qquad y_2 = x_2'. \tag{4.3}$$

Using these new dependent variables, we rewrite (4.2) as the following four simultaneous equations in the four unknowns x_1, y_1, x_2 and y_2:

$$\begin{aligned} x_1' &= y_1, \\ y_1' &= \frac{-k_1x_1 + k_2(x_2 - x_1)}{m_1}, \\ x_2' &= y_2, \\ y_2' &= \frac{-k_2(x_2 - x_1) - k_3x_2}{m_2}, \end{aligned} \tag{4.4}$$

which, in matrix form, become

$$\begin{bmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{k_1+k_2}{m_1} & 0 & \frac{k_2}{m_1} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{k_2}{m_2} & 0 & -\frac{k_2+k_3}{m_2} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix}. \tag{4.5}$$

Using the following notation for the unknown vector, the coefficient matrix and given initial conditions,

$$\boldsymbol{u} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix}, \qquad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{k_1+k_2}{m_1} & 0 & \frac{k_2}{m_1} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{k_2}{m_2} & 0 & -\frac{k_2+k_3}{m_2} & 0 \end{bmatrix} \qquad \boldsymbol{u}_0 = \begin{bmatrix} x_1(0) \\ y_1(0) \\ x_2(0) \\ y_2(0) \end{bmatrix},$$

the initial value problem becomes

$$\boldsymbol{u}' = A\boldsymbol{u}, \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0. \tag{4.6}$$

It is to be noted that the matrix A is not in the form of a companion matrix. □

4.2. Existence and Uniqueness Theorem

In this section, we recall results which have been quoted for systems in the previous chapters. In particular, the Existence and Uniqueness Theorem 1.3 holds for general first-order systems of the form

$$\boldsymbol{y}' = \boldsymbol{f}(x, \boldsymbol{y}), \qquad \boldsymbol{y}(x_0) = \boldsymbol{y}_0, \tag{4.7}$$

provided, in Definition 1.3, norms replace absolute values in the Lipschitz condition

$$\|\boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{f}(\boldsymbol{y})\| \leq M\|\boldsymbol{z} - \boldsymbol{y}\|, \qquad \text{for all} \quad \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n,$$

and in the statement of the theorem.

A similar remark holds for the Existence and Uniqueness Theorem 3.2 for linear systems of the form

$$\boldsymbol{y}' = A(x)\boldsymbol{y} + \boldsymbol{g}(x), \qquad \boldsymbol{y}(x_0) = \boldsymbol{y}_0, \tag{4.8}$$

provided the matrix $A(x)$ and the vector-valued function $\boldsymbol{f}(x)$ are continuous on the interval (x_0, x_f). The Picard iteration method used in the proof of this theorem has been stated for systems of differential equations and needs no change for the present systems.

4.3. Fundamental Systems

It is readily seen that the solutions to the linear homogeneous system

$$\boldsymbol{y}' = A(x)\boldsymbol{y}, \qquad x \in]a, b[, \tag{4.9}$$

form a vector space since differentiation and matrix multiplication are linear operators.

As before, m vector-valued functions, $\boldsymbol{y}_1(x), \boldsymbol{y}_2(x), \ldots, \boldsymbol{y}_m(x)$, are said to be *linearly independent* on an interval $]a, b[$ if the identity

$$c_1\boldsymbol{y}_1(x) + c_2\boldsymbol{y}_2(x) + \cdots + \boldsymbol{y}_m(x) = \mathbf{0}, \qquad \text{for all} \quad x \in]a, b[,$$

implies that

$$c_1 = c_2 = \cdots = c_m = 0.$$

Otherwise, this set of functions is said to be *linearly dependent*.

For general systems, the determinant $W(x)$ of n column-vector functions, $\boldsymbol{y}_1(x), \boldsymbol{y}_2(x), \ldots, \boldsymbol{y}_n(x)$, with values in $\mathbb{R}^n$,

$$W(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)(x) = \det \begin{bmatrix} y_{11}(x) & y_{12}(x) & \cdots & y_{1n}(x) \\ y_{21}(x) & y_{22}(x) & \cdots & y_{2n}(x) \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1}(x) & y_{n2}(x) & \cdots & y_{nn}(x) \end{bmatrix},$$

is a generalization of the Wronskian for a linear scalar equation.

We restate and prove Liouville's or Abel's Lemma 3.1 for general linear systems. For this purpose, we define the *trace* of a matrix A, denoted by $\operatorname{tr} A$, to be the sum of the diagonal elements, a_{ii}, of A,

$$\operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn}.$$

LEMMA 4.1 (Abel). *Let $\boldsymbol{y}_1(x), \boldsymbol{y}_2(x), \ldots, \boldsymbol{y}_n(x)$, be n solutions of the system $\boldsymbol{y}' = A(x)\boldsymbol{y}$ on the interval $]a, b[$. Then the determinant $W(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)(x)$ satisfies the following identity:*

$$W(x) = W(x_0)\, e^{-\int_{x_0}^{x} tr\, A(t)\, dt}, \qquad x_0 \in]a, b[. \tag{4.10}$$

PROOF. For simplicity of writing, let us take $n = 3$; the general case is treated as easily. Let $W(x)$ be the determinant of three solutions $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, $\boldsymbol{y}_3$. Then its derivative $W'(x)$ is of the form

$$\begin{aligned} W'(x) &= \begin{vmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{vmatrix}' \\ &= \begin{vmatrix} y'_{11} & y'_{12} & y'_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{vmatrix} + \begin{vmatrix} y_{11} & y_{12} & y_{13} \\ y'_{21} & y'_{22} & y'_{23} \\ y_{31} & y_{32} & y_{33} \end{vmatrix} + \begin{vmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y'_{31} & y'_{32} & y'_{33} \end{vmatrix}. \end{aligned}$$

We consider the first of the last three determinants. We see that the first row of the differential system

$$\begin{bmatrix} y'_{11} & y'_{12} & y'_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix}$$

is

$$\begin{aligned} y'_{11} &= a_{11}y_{11} + a_{12}y_{21} + a_{13}y_{31}, \\ y'_{12} &= a_{11}y_{12} + a_{12}y_{22} + a_{13}y_{32}, \\ y'_{13} &= a_{11}y_{13} + a_{12}y_{23} + a_{13}y_{33}. \end{aligned}$$

Substituting these expressions in the first row of the first determinant and subtracting a_{12} times the second row and a_{13} times the third row from the first row, we obtain $a_{11}W(x)$. Similarly, for the second and third determinants we obtain $a_{22}W(x)$ and $a_{33}W(x)$, respectively. Thus $W(x)$ satisfies the separable equation

$$W'(x) = \operatorname{tr}\big(A(x)\big)W(x)$$

whose solution is

$$W(x) = W(x_0)\, e^{\int_{x_0}^{x} \operatorname{tr} A(t)\, dt}. \quad \square$$

The following corollary follows from Abel's lemma.

COROLLARY 4.1. *If n solutions to the homogeneous differential system* (4.9) *are independent at one point, then they are independent on the interval $]a, b[$. If, on the other hand, these solutions are linearly dependent at one point, then their determinant, $W(x)$, is identically zero, and hence they are everywhere dependent.*

REMARK 4.1. It is worth emphasizing the difference between linear independence of vector-valued functions and solutions of linear systems. For instance, the two vector-valued functions

$$\boldsymbol{f}_1(x) = \begin{bmatrix} x \\ 0 \end{bmatrix}, \qquad \boldsymbol{f}_2(x) = \begin{bmatrix} 1+x \\ 0 \end{bmatrix},$$

are linearly independent. Their determinant, however, is zero. This does not contradict Corollary 4.1 since $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ cannot be solutions to a system (4.9).

DEFINITION 4.1. A set of n linearly independent solutions of a linear homogeneous system $\boldsymbol{y}' = A(x)\boldsymbol{y}$ is called a *fundamental system*, and the corresponding invertible matrix

$$Y(x) = \begin{bmatrix} y_{11}(x) & y_{12}(x) & \cdots & y_{1n}(x) \\ y_{21}(x) & y_{22}(x) & \cdots & y_{2n}(x) \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1}(x) & y_{n2}(x) & \cdots & y_{nn}(x) \end{bmatrix},$$

is called a *fundamental matrix*.

LEMMA 4.2. *If $Y(x)$ is a fundamental matrix, then $Z(x) = Y(x)Y^{-1}(x_0)$ is also a fundamental matrix such that $Z(x_0) = I$.*

PROOF. Let C be any constant matrix. Since $Y' = AY$, it follows that $(YC)' = Y'C = (AY)C = A(YC)$. The lemma follows by letting $C = Y^{-1}(x_0)$. Obviously, $Z(x_0) = I$. $\square$

In the following, we shall often assume that a fundamental matrix satisfies the condition $Y(x_0) = I$. We have the following theorem for linear homogeneous systems.

THEOREM 4.1. *Let $Y(x)$ be a fundamental matrix for $\boldsymbol{y}' = A(x)\boldsymbol{y}$. Then the general solution is*

$$\boldsymbol{y}(x) = Y(x)\boldsymbol{c},$$

where $\boldsymbol{c}$ is an arbitrary vector. If $Y(x_0) = I$, then

$$\boldsymbol{y}(x) = Y(x)\boldsymbol{y}_0$$

is the unique solution of the initial value problem

$$\boldsymbol{y}' = A(x)\boldsymbol{y}, \qquad \boldsymbol{y}(x_0) = \boldsymbol{y}_0.$$

PROOF. The proof of both statements relies on the Uniqueness Theorem. To prove the first part, let $Y(x)$ be a fundamental matrix and $\boldsymbol{z}(x)$ be any solution of the system. Let x_0 be in the domain of $\boldsymbol{z}(x)$ and define $\boldsymbol{c}$ by

$$\boldsymbol{c} = Y^{-1}(x_0)\boldsymbol{z}(x_0).$$

Define $\boldsymbol{y}(x) = Y(x)\boldsymbol{c}$. Since both $\boldsymbol{y}(x)$ and $\boldsymbol{z}(x)$ satisfy the same differential equation and the same initial conditions, they must be the same solution by the Uniqueness Theorem. The proof of the second part is similar. □

The following lemma will be used to obtain a formula for the solution of the initial value problem (4.8) in terms of a fundamental solution.

LEMMA 4.3. *Let $Y(x)$ be a fundamental matrix for the system* (4.9). *Then, $(Y^T)^{-1}(x)$ is a fundamental solution for the adjoint system*

$$\boldsymbol{y}' = -A^T(x)\boldsymbol{y}. \tag{4.11}$$

PROOF. Differentiating the identity

$$Y^{-1}(x)Y(x) = I,$$

we have

$$(Y^{-1})'(x)Y(x) + Y^{-1}(x)Y'(x) = 0.$$

Since the matrix $Y(x)$ is a solution of (4.9), we can replace $Y'(x)$ in the previous identity with $A(x)Y(x)$ and obtain

$$(Y^{-1})'(x)Y(x) = -Y^{-1}(x)A(x)Y(x).$$

Multiplying this equation on the right by $Y^{-1}(x)$ and taking the transpose of both sides lead to (4.11). □

THEOREM 4.2 (Solution formula). *Let $Y(x)$ be a fundamental solution matrix of the homogeneous linear system* (4.9). *Then the unique solution to the initial value problem* (4.8) *is*

$$\boldsymbol{y}(x) = Y(x)Y^{-1}(x_0)\boldsymbol{y}_0 + Y(x)\int_{x_0}^{x} Y^{-1}(t)\boldsymbol{g}(t)\,dt. \tag{4.12}$$

PROOF. Multiply both sides of (4.8) by $Y^{-1}(x)$ and use the result of Lemma 4.3 to get

$$\big(Y^{-1}(x)\boldsymbol{y}(x)\big)' = Y^{-1}(x)\boldsymbol{g}(x).$$

The proof of the theorem follows by integrating the previous expression with respect to x from x_0 to x. □

4.4. Homogeneous Linear Systems with Constant Coefficients

A *homogeneous linear system with constant coefficients* has the form

$$\boldsymbol{y}' = A\boldsymbol{y}, \tag{4.13}$$

where the $n \times n$ matrix A is constant, i.e. all entries of A are constants. We shall assume that all entries of A are real. The form of equation (4.13) is very reminiscent of the (scalar) differential equation $y' = ay$ and so we seek solutions of the form

$$\boldsymbol{y}(x) = e^{\lambda x}\boldsymbol{v},$$

where λ and $\boldsymbol{v}$ are constants ($\boldsymbol{v} \neq \boldsymbol{0}$).

If $\boldsymbol{y}(x) = e^{\lambda x}\boldsymbol{v}$, then $\boldsymbol{y}'(x) = \lambda\, e^{\lambda x}\boldsymbol{v}$, and system (4.13) becomes

$$\lambda\, e^{\lambda x}\boldsymbol{v} = A\, e^{\lambda x}\boldsymbol{v} \quad \text{or} \quad e^{\lambda x}\lambda\boldsymbol{v} = e^{\lambda x}A\boldsymbol{v} \quad \text{or} \quad \lambda\boldsymbol{v} = A\boldsymbol{v},$$

which can be rewritten as

$$(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}, \qquad \boldsymbol{v} \neq \boldsymbol{0}. \tag{4.14}$$

Now, since $\boldsymbol{v} \neq \boldsymbol{0}$ (otherwise $\boldsymbol{y}(x) = \boldsymbol{0}$), we are seeking a nontrivial solution of the homogeneous system (4.13). For such a solution to exist, the matrix $A - \lambda I$ cannot be invertible and hence it must have determinant equal to 0, i.e. $\det(A - \lambda I) = 0$, which is called the *characteristic equation* of the matrix A and it will imply that a polynomial of degree n must be zero. The n roots of this equation, $\lambda_1, \lambda_2, \ldots, \lambda_n$, are called the *eigenvalues* of the matrix A. The corresponding (nonzero) vectors $\boldsymbol{v}_i$ which satisfy $(A - \lambda_i I)\boldsymbol{v}_i = \boldsymbol{0}$ are called the *eigenvectors* of A.

It is known that for each distinct eigenvalue, A has a corresponding eigenvector and the set of such eigenvectors are linearly independent. If A is symmetric, $A^T = A$, that is, A and its transpose are equal, then the eigenvalues are real and A has n eigenvectors which can be chosen to be orthonormal.

EXAMPLE 4.3. Find the general solution of the symmetric system $\boldsymbol{y}' = A\boldsymbol{y}$:

$$\boldsymbol{y}' = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \boldsymbol{y}.$$

SOLUTION. The eigenvalues are obtained from the characteristic polynomial of A,

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = \lambda^2 - 4\lambda + 3 = (\lambda - 1)(\lambda - 3) = 0.$$

Hence the eigenvalues are

$$\lambda_1 = 1, \qquad \lambda_2 = 3.$$

The eigenvector corresponding to λ_1 is obtained from the singular system

$$(A - I)\boldsymbol{u} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0.$$

We are free to take any nonzero solution. Taking $u_1 = 1$ we have the eigenvector

$$\boldsymbol{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Similarly, the eigenvector corresponding to λ_2 is obtained from the singular system

$$(A - 3I)\boldsymbol{v} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0.$$

Taking $v_1 = 1$ we have the eigenvector

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Since $\lambda_1 \neq \lambda_2$, we have two independent solutions

$$\boldsymbol{y}_1 = e^x \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \boldsymbol{y}_2 = e^{3x} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

and the fundamental system and general solution are

$$Y(x) = \begin{bmatrix} e^x & e^{3x} \\ -e^x & e^{3x} \end{bmatrix}, \qquad \boldsymbol{y} = Y(x)\boldsymbol{c}.$$

The general solution can also be written as

$$\begin{aligned} \boldsymbol{y}(x) &= c_1\boldsymbol{y}_1(x) + c_2\boldsymbol{y}_2(x) \\ &= c_1\, e^x \begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_2\, e^{3x} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned}$$

The Matlab solution is

```
A = [2 1; 1 2];
[Y,D] = eig(A);
syms x c1 c2
z = Y*diag(exp(diag(D*x)))*[c1; c2]
 z =
   [  1/2*2^(1/2)*exp(x)*c1+1/2*2^(1/2)*exp(3*x)*c2]
   [ -1/2*2^(1/2)*exp(x)*c1+1/2*2^(1/2)*exp(3*x)*c2]
```

Note that Matlab normalizes the eigenvectors in the l_2 norm. Hence, the matrix Y is orthogonal since the matrix A is symmetric. The solution $\boldsymbol{y}$

```
y = simplify(sqrt(2)*z)
 y =
   [  exp(x)*c1+exp(3*x)*c2]
   [ -exp(x)*c1+exp(3*x)*c2]
```

is produced by the nonnormalized eigenvectors $\boldsymbol{u}$ and $\boldsymbol{v}$. □

If the constant matrix A of the system $\boldsymbol{y}' = A\boldsymbol{y}$ has a full set of independent eigenvectors, then it is diagonalizable

$$Y^{-1}AY = D,$$

where the columns of the matrix Y are eigenvectors of A and the corresponding eigenvalues are the diagonal elements of the diagonal matrix D. This fact can be used to solve the initial value problem

$$\boldsymbol{y}' = A\boldsymbol{y}, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0.$$

Set

$$\boldsymbol{y} = Y\boldsymbol{x}, \quad \text{or} \quad \boldsymbol{x} = Y^{-1}\boldsymbol{y}.$$

Since A is constant, then Y is constant and $\boldsymbol{x}' = Y^{-1}\boldsymbol{y}'$. Hence the given system $\boldsymbol{y}' = A\boldsymbol{y}$ becomes

$$Y^{-1}\boldsymbol{y}' = Y^{-1}AYY^{-1}\boldsymbol{y},$$

that is,

$$\boldsymbol{x}' = D\boldsymbol{x}.$$

Componentwise, we have

$$x_1'(t) = \lambda_1 x_1(t), \quad x_2'(t) = \lambda_2 x_2(t), \quad \ldots, \quad x_n(t)' = \lambda_n x_n(t),$$

with solutions

$$x_1(t) = c_1\, e^{\lambda_1 t}, \quad x_2(t) = c_2\, e^{\lambda_2 t}, \quad \ldots, \quad x_n(t) = c_n\, e^{\lambda_n t},$$

where the constants $c_1, \ldots, c_n$ are determined by the initial conditions. Since

$$\boldsymbol{y}_0 = Y\boldsymbol{x}(0) = Y\boldsymbol{c},$$

it follows that

$$\boldsymbol{c} = Y^{-1}\boldsymbol{y}_0.$$

These results are used in the following example.

EXAMPLE 4.4. Solve the system of Example 4.2 with

$$m_1 = 10, \qquad m_2 = 20, \qquad k_1 = k_2 = k_3 = 1,$$

and initial values

$$x_1 = 0.8, \qquad x_2 = y_0 = y_1 = 0,$$

and plot the solution.

SOLUTION. The matrix A takes the form

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -0.2 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.05 & 0 & -0.1 & 0 \end{bmatrix}$$

The Matlab solution for $x(t)$ and $y(t)$ and their plot are

```
A = [0 1 0 0; -0.2 0 0.1 0; 0 0 0 1; 0.05 0 -0.1 0];
y0 = [0.8 0 0 0]';
[Y,D] = eig(A);
t = 0:1/5:60; c = inv(Y)*y0; y = y0;
for i = 1:length(t)-1
yy = Y*diag(exp(diag(D)*t(i+1)))*c;
y = [y,yy];
end
ry = real(y); % the solution is real; here the imaginary part is zero
subplot(2,2,1); plot(t,ry(1,:),t,ry(3,:),'--');
```

The Matlab `ode45` command from the ode suite produces the same numerical solution. Using the M-file `spring.m`,

```
function yprime = spring(t,y); % MAT 2331, Example 3a.4.2.
A = [0 1 0 0; -0.2 0 0.1 0; 0 0 0 1; 0.05 0 -0.1 0];
yprime = A*y;
```

we have

```
y0 = [0.8 0 0 0];  tspan=[0 60];
[t,y]=ode45('spring',tspan,y0);
subplot(2,2,1); plot(t,y(:,1),t,y(:,3));
```

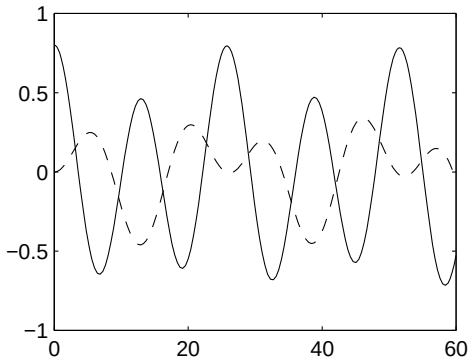

FIGURE 4.2. Graph of solution $x_1(t)$ (solid line) and $x_2(t)$ (dashed line) to Example 4.4.

The plot is shown in Fig. 4.2. □

The case of multiple eigenvalues may lead to a lack of eigenvectors in the construction of a fundamental solution. In this situation, one has recourse to generalized eigenvectors.

DEFINITION 4.2. Let A be an $n \times n$ matrix. We say that λ is a *deficient eigenvalue* of A if it has algebraic multiplicity $m > 1$ and fewer than m eigenvectors associated with it. If there are $k < m$ linearly independent eigenvectors associated with λ, then the integer

$$r = m - k$$

is called the *degree of deficiency* of λ. A vector $\boldsymbol{u}$ is called a *generalized eigenvector* of A associated with λ if there is an integer $s > 0$ such that

$$(A - \lambda I)^s \boldsymbol{u} = \boldsymbol{0},$$

but

$$(A - \lambda I)^{s-1} \boldsymbol{u} \neq \boldsymbol{0}.$$

In general, given a matrix A with an eigenvalue λ of degree of deficiency r and corresponding eigenvector $\boldsymbol{u}_1$, we construct a set of generalized eigenvectors $\{\boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ as solutions of the systems

$$(A - \lambda I)\boldsymbol{u}_2 = \boldsymbol{u}_1, \quad (A - \lambda I)\boldsymbol{u}_3 = \boldsymbol{u}_2, \quad \ldots, \quad (A - \lambda I)\boldsymbol{u}_r = \boldsymbol{u}_{r-1}.$$

The eigenvector $\boldsymbol{u}_1$ and the set of generalized eigenvectors, in turn, generate the following set of linearly independent solutions of (4.13):

$$\boldsymbol{y}_1(x) = e^{\lambda x}\boldsymbol{u}_1, \quad \boldsymbol{y}_2(x) = e^{\lambda x}(x\boldsymbol{u}_1 + \boldsymbol{u}_2), \quad \boldsymbol{y}_3(x) = e^{\lambda x}\left(\frac{x^2}{2}\boldsymbol{u}_1 + x\boldsymbol{u}_2 + \boldsymbol{u}_3\right), \quad \ldots.$$

It is a result of linear algebra that any $n \times n$ matrix has n linearly independent generalized eigenvectors.

Let us look at the details of the 2×2 situation. Suppose we have $\boldsymbol{y}' = A\boldsymbol{y}$ where the eigenvalue is repeated, $\lambda_1 = \lambda_2 = \lambda$. Then one solution is

$$\boldsymbol{y}_1(x) = e^{\lambda x}\boldsymbol{v}.$$

From Chapter 2, we would suspect that the second independent solution should be

$$\boldsymbol{y}_2(x) = x\, e^{\lambda x}\boldsymbol{v},$$

but then

$$\boldsymbol{y}_2'(x) = e^{\lambda x}\boldsymbol{v} + \lambda x\, e^{\lambda x}\boldsymbol{v},$$

whereas

$$A\boldsymbol{y}_2(x) = Ax\, e^{\lambda x}\boldsymbol{v} = x\, e^{\lambda x}A\boldsymbol{v} = x\, e^{\lambda x}\boldsymbol{v}$$

and so this $\boldsymbol{y}_2(x)$ is not a solution ($\boldsymbol{y}_2' \neq A\boldsymbol{y}_2$). Instead, we seek a solution of the form

$$\boldsymbol{y}_2(x) = e^{\lambda x}(x\boldsymbol{v} + \boldsymbol{u}).$$

Then

$$\boldsymbol{y}_2'(x) = e^{\lambda x}\boldsymbol{v} + \lambda x\, e^{\lambda x}\boldsymbol{v} + \lambda e^{\lambda x}\boldsymbol{u}$$

and

$$A\boldsymbol{y}_2 = A(x\, e^{\lambda x}\boldsymbol{v} + e^{\lambda x}\boldsymbol{u}) = x\lambda\, e^{\lambda x}\boldsymbol{v} + e^{\lambda x}A\boldsymbol{u};$$

so, by succesive simplifications, $\boldsymbol{y}_2' = A\boldsymbol{y}_2$ requires that

$$\begin{aligned} e^{\lambda x}\boldsymbol{v} + \lambda x\, e^{\lambda x}\boldsymbol{v} + \lambda e^{\lambda x}\boldsymbol{u} &= x\lambda\, e^{\lambda x}\boldsymbol{v} + e^{\lambda x}A\boldsymbol{u} \\ e^{\lambda x}\boldsymbol{v} + \lambda e^{\lambda x}\boldsymbol{u} &= e^{\lambda x}A\boldsymbol{u} \\ \boldsymbol{v} + \lambda\boldsymbol{u} &= A\boldsymbol{u}, \end{aligned}$$

that is,

$$(A - \lambda I)\boldsymbol{u} = \boldsymbol{v}, \tag{4.15}$$

as seen above. So provided $\boldsymbol{u}$ satisfies (4.15),

$$\boldsymbol{y}_2(x) = e^{\lambda x}(x\boldsymbol{v} + \boldsymbol{u})$$

is the second independent solution.

EXAMPLE 4.5. Find the general solution of the system

$$\boldsymbol{y}' = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\boldsymbol{y}.$$

SOLUTION. Since

$$\det(A - \lambda I) = \left|\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right| = \begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 = 0,$$

we have a repeated eigenvalue $\lambda_1 = \lambda_2 = 1$. The eigenvector $\boldsymbol{v}$ must satisfy the homogeneous system $(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}$, which is

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\boldsymbol{v} = \boldsymbol{0}$$

Taking

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

we have the first solution

$$\boldsymbol{y}_1(x) = e^{\lambda x}\boldsymbol{v} = e^x\begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

To get the second solution, we must solve the nonhomogeneous system

$$(A - \lambda I)\boldsymbol{u} = \boldsymbol{v}$$

for $\boldsymbol{u}$, or

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \boldsymbol{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We take

$$\boldsymbol{u} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

so the second solution is

$$\boldsymbol{y}_2(x) = e^{\lambda x}(x\boldsymbol{v} + \boldsymbol{u}) = e^x \left(x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

The general solution can be written as

$$\boldsymbol{y}_g(x) = c_1\boldsymbol{y}_1(x) + c_2\boldsymbol{y}_2(x) = c_1\, e^x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2\, e^x \left(x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right). \quad \square$$

EXAMPLE 4.6. Solve the system $\boldsymbol{y}' = A\boldsymbol{y}$:

$$\boldsymbol{y}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix} \boldsymbol{y}.$$

SOLUTION. One finds that the matrix A has a triple eigenvalue $\lambda = 1$. Row-reducing the matrix $A - I$, we obtain a matrix of rank 2; hence $A - I$ admits a single eigenvector $\boldsymbol{u}_1$:

$$A - I \sim \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Thus, one solution is

$$\boldsymbol{y}_1(x) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} e^x.$$

To construct a first generalized eigenvector, we solve the equation

$$(A - I)\boldsymbol{u}_2 = \boldsymbol{u}_1.$$

Thus,

$$\boldsymbol{u}_2 = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}$$

and

$$\boldsymbol{y}_2(x) = (x\boldsymbol{u}_1 + \boldsymbol{u}_2)\, e^x = \left(x \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} \right) e^x$$

is a second linearly independent solution.

To construct a second generalized eigenvector, we solve the equation

$$(A - I)\boldsymbol{u}_3 = \boldsymbol{u}_2.$$

Thus,

$$\boldsymbol{u}_3 = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$$

and

$$\boldsymbol{y}_3(x) = \left(\frac{x^2}{2}\boldsymbol{u}_1 + x\boldsymbol{u}_2 + \boldsymbol{u}_3\right) e^x = \left(\frac{x^2}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x\begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}\right) e^x$$

is a third linearly independent solution. □

In the previous example, the invariant subspace associated with the triple eigenvalue is one-dimensional. Hence the construction of two generalized eigenvectors is straightforward. In the next example, this invariant subspace associated with the triple eigenvalue is two-dimensional. Hence the construction of a generalized eigenvector is a bit more complex.

EXAMPLE 4.7. Solve the system $\boldsymbol{y}' = A\boldsymbol{y}$:

$$\boldsymbol{y}' = \begin{bmatrix} 1 & 2 & 1 \\ -4 & 7 & 2 \\ 4 & -4 & 1 \end{bmatrix} \boldsymbol{y}.$$

SOLUTION. **(a) The analytic solution.**— One finds that the matrix A has a triple eigenvalue $\lambda = 3$. Row-reducing the matrix $A - 3I$, we obtain a matrix of rank 1; hence $A - 3I$ two independent eigenvectors, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$:

$$A - 3I \sim \begin{bmatrix} -2 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

Thus, two independent solutions are

$$\boldsymbol{y}_1(x) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} e^{3x}, \qquad \boldsymbol{y}_2(x) = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} e^{3x}.$$

To obtain a third independent solution, we construct a generalized eigenvector by solving the equation

$$(A - 3I)\boldsymbol{u}_3 = \alpha\boldsymbol{u}_1 + \beta\boldsymbol{u}_2,$$

where the parameters α and β are to be chosen so that the right-hand side,

$$\boldsymbol{u}_4 = \alpha\boldsymbol{u}_1 + \beta\boldsymbol{u}_2 = \begin{bmatrix} \alpha + \beta \\ \alpha \\ 2\beta \end{bmatrix},$$

is in the space V spanned by the columns of the matrix $(A - 3I)$. Since $\operatorname{rank}(A - 3I) = 1$, then

$$V = \operatorname{span}\begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

and we may take

$$\begin{bmatrix} \alpha + \beta \\ \alpha \\ 2\beta \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}.$$

Thus, $\alpha = 2$ and $\beta = -1$, It follows that

$$\boldsymbol{u}_4 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \qquad \boldsymbol{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and

$$\boldsymbol{y}_3(x) = (x\boldsymbol{u}_4 + \boldsymbol{u}_3)\, e^{3x} = \left(x \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) e^{3x}$$

is a third linearly independent solution.

(b) The Matlab symbolic solution.— To solve the problem with symbolic Matlab, one uses the Jordan normal form, $J = X^{-1}AX$, of the matrix A. If we let

$$\boldsymbol{y} = X\boldsymbol{w},$$

the equation simplifies to

$$\boldsymbol{w}' = J\boldsymbol{w}.$$

```
A =  1     2     1
    -4     7     2
     4    -4     1

[X,J] = jordan(A)

X = -2.0000    1.5000    0.5000
    -4.0000         0         0
     4.0000    1.0000    1.0000

J = 3     1     0
    0     3     0
    0     0     3
```

The matrix $J - 3I$ admits the two eigenvectors

$$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

and the generalized eigenvector

$$\boldsymbol{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

the latter being a solution of the equation

$$(J - 3I)\boldsymbol{u}_2 = \boldsymbol{u}_1.$$

Thus three independent solutions are

$$\boldsymbol{y}_1 = e^{3x}X\boldsymbol{u}_1, \qquad \boldsymbol{y}_2 = e^{3x}X(x\boldsymbol{u}_1 + \boldsymbol{u}_2), \qquad \boldsymbol{y}_3 = e^{3x}X\boldsymbol{u}_3,$$

that is

```
u1=[1 0 0]'; u2=[0 1 0]'; u3=[0 0 1]';
syms x; y1 = exp(3*x)*X*u1

y1 = [ -2*exp(3*x)]
     [ -4*exp(3*x)]
     [  4*exp(3*x)]
```

```
y2 = exp(3*x)*X*(x*u1+u2)

y2 = [ -2*exp(3*x)*x+3/2*exp(3*x)]
     [               -4*exp(3*x)*x]
     [      4*exp(3*x)*x+exp(3*x)]

y3 = exp(3*x)*X*u3

y3 = [ 1/2*exp(3*x)]
     [             0]
     [     exp(3*x)]
```

□

4.5. Nonhomogeneous Linear Systems

In Chapter 3, the Method of Undetermined Coefficients and the Method of Variation of Parameters have been used for finding particular solutions of nonhomogeneous differential equations. In this section, we generalize these methods to linear systems of the form

$$\boldsymbol{y}' = A\boldsymbol{y} + \boldsymbol{f}(x). \tag{4.16}$$

We recall that once a particular solution $\boldsymbol{y}_p$ of this system has been found, the general solution is the sum of $\boldsymbol{y}_p$ and the solution $\boldsymbol{y}_h$ of the homogeneous system

$$\boldsymbol{y}' = A\boldsymbol{y}.$$

4.5.1. Method of Undetermined Coefficients. The Method of Undetermined Coefficients can be used when the matrix A in (4.16) is constant and the dimension of the vector space spanned by the derivatives of the vector function $\boldsymbol{f}(x)$ of (4.16) is finite. This is the case when the components of $\boldsymbol{f}(x)$ are combinations of cosines, sines, exponentials, hyperbolic sines and cosines, and polynomials. For such problems, the appropriate choice of $\boldsymbol{y}_p$ is a linear combination of vectors in the form of the functions that appear in $\boldsymbol{f}(x)$ together with all their independent derivatives.

Example 4.8. Find the general solution of the nonhomogeneous linear system

$$\boldsymbol{y}' = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \boldsymbol{y} + \begin{bmatrix} 4\,e^{-3x} \\ e^{-2x} \end{bmatrix} := A\boldsymbol{y} + \boldsymbol{f}(x).$$

Solution. The eigenvalues of the matrix A of the system are

$$\lambda_1 = i, \qquad \lambda_2 = -i,$$

and the corresponding eigenvectors are

$$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}, \qquad \boldsymbol{u}_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

Hence the general solution of the homogeneous system is

$$\boldsymbol{y}_h(x) = k_1\, e^{ix} \begin{bmatrix} 1 \\ i \end{bmatrix} + k_2\, e^{-ix} \begin{bmatrix} 1 \\ -i \end{bmatrix},$$

where k_1 and k_2 are complex constants. To obtain real independent solutions we use the fact that the real and imaginary parts of a solution of a real homogeneous

linear equation are solutions. We see that the real and imaginary parts of the first solution,

$$\begin{aligned}\boldsymbol{u}_1 &= \begin{bmatrix} 1 \\ i \end{bmatrix} e^{ix} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} + i \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) (\cos x + i \sin x) \\ &= \begin{bmatrix} \cos x \\ -\sin x \end{bmatrix} + i \begin{bmatrix} \sin x \\ \cos x \end{bmatrix},\end{aligned}$$

are independent solutions. Hence, we obtain the following real-valued general solution of the homogeneous system

$$\boldsymbol{y}_h(x) = c_1 \begin{bmatrix} \cos x \\ -\sin x \end{bmatrix} + c_2 \begin{bmatrix} \sin x \\ \cos x \end{bmatrix}.$$

The function $\boldsymbol{f}(x)$ can be written in the form

$$\boldsymbol{f}(x) = 4 \begin{bmatrix} 1 \\ 0 \end{bmatrix} e^{-3x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} e^{-2x} = 4\boldsymbol{e}_1 \, e^{-3x} + \boldsymbol{e}_2 \, e^{-2x}$$

with obvious definitions for $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. Note that $\boldsymbol{f}(x)$ and $\boldsymbol{y}_h(x)$ do not have any part in common. We therefore choose $\boldsymbol{y}_p(x)$ in the form

$$\boldsymbol{y}_p(x) = \boldsymbol{a} \, e^{-3x} + \boldsymbol{b} \, e^{-2x}.$$

Substituting $\boldsymbol{y}_p(x)$ in the given system, we obtain

$$\boldsymbol{0} = (3\boldsymbol{a} + A\boldsymbol{a} + 4\boldsymbol{e}_1) \, e^{-3x} + (2\boldsymbol{b} + A\boldsymbol{b} + \boldsymbol{e}_2) \, e^{-2x}.$$

Since the functions e^{-3x} and e^{-2x} are linearly independent, their coefficients must be zero, from which we obtain two equations for $\boldsymbol{a}$ and $\boldsymbol{b}$,

$$(A + 3I)\boldsymbol{a} = -4\boldsymbol{e}_1, \qquad (A + 2I)\boldsymbol{b} = -\boldsymbol{e}_2.$$

Hence,

$$\begin{aligned}\boldsymbol{a} &= -4(A + 3I)^{-1}\boldsymbol{e}_1 = -\frac{1}{5} \begin{bmatrix} 6 \\ 2 \end{bmatrix} \\ \boldsymbol{b} &= -(A + 2I)^{-1}\boldsymbol{e}_2 = \frac{1}{5} \begin{bmatrix} 1 \\ -2 \end{bmatrix}.\end{aligned}$$

Finally,

$$\boldsymbol{y}_h(x) = -\begin{bmatrix} \dfrac{6}{5} \, e^{-3x} - \dfrac{1}{5} \, e^{-2x} \\ \\ \dfrac{2}{5} \, e^{-3x} + \dfrac{2}{5} \, e^{-2x} \end{bmatrix}.$$

The general solution is

$$\boldsymbol{y}_g(x) = \boldsymbol{y}_h(x) + \boldsymbol{y}_p(x).$$

□

EXAMPLE 4.9. Find the general solution of the nonhomogeneous system

$$\boldsymbol{y}' = \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix} \boldsymbol{y} + \begin{bmatrix} 0 \\ 9x - 24 \end{bmatrix}.$$

SOLUTION. The corresponding homogeneous equation is

$$\boldsymbol{y}' = \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix} \boldsymbol{y}.$$

Since

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & -1 \\ 1 & 4-\lambda \end{vmatrix} = (2-\lambda)(4-\lambda) = \lambda^2 - 6\lambda + 9 = (\lambda - 3)^2 = 0,$$

we have repeated eigenvalues $\lambda_1 = \lambda_2 = 3$. To find the first eigenvector, we solve $(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}$ which is

$$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \boldsymbol{v} = \boldsymbol{0}.$$

So we take

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

To find the second solution, which is a generalized eigenvector, we solve $(A - \lambda I)\boldsymbol{u} = \boldsymbol{v}$ which is

$$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \boldsymbol{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

so we take

$$\boldsymbol{u} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Thus the general solution of the homogeneous system is

$$\boldsymbol{y}_h = c_1\, e^{3x} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_2\, e^{3x} \left(x \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \end{bmatrix} \right).$$

Now,

$$\boldsymbol{f}(x) = \begin{bmatrix} 0 \\ 9x - 24 \end{bmatrix} = \begin{bmatrix} 0 \\ 9 \end{bmatrix} x + \begin{bmatrix} 0 \\ -24 \end{bmatrix}.$$

So the guess for the particular solution is $\boldsymbol{y}_p(x) = \boldsymbol{a}x + \boldsymbol{b}$. Thus

$$\boldsymbol{y}_p'(x) = \boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

and

$$\begin{aligned} A\boldsymbol{y}_p + \boldsymbol{f}(x) &= \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix} [\boldsymbol{a}x + \boldsymbol{b}] + \begin{bmatrix} 0 \\ 9x - 24 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} a_1 x + b_1 \\ a_2 x + b_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 9x - 24 \end{bmatrix} \\ &= \begin{bmatrix} (2a_1 - a_2)x + 2b_1 - b_2 \\ (a_1 + 4a_2)x + b_1 + 4b_2 + 9x - 24 \end{bmatrix} \\ &= \begin{bmatrix} (2a_1 - a_2)x + 2b_1 - b_2 \\ (a_1 + 4a_2 + 9)x + b_1 + 4b_2 - 24 \end{bmatrix}. \end{aligned}$$

So $\boldsymbol{y}_p'(x) = A\boldsymbol{y}_p(x) + \boldsymbol{f}$ means that

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} (2a_1 - a_2)x + 2b_1 - b_2 \\ (a_1 + 4a_2 + 9)x + b_1 + 4b_2 - 24 \end{bmatrix}.$$

Equating the components, which are polynomials, and hence we must equate coefficients, yields

$$a_1 = (2a_1 - a_2)x + 2b_1 - b_2 \quad \Rightarrow 2a_1 - a_2 = 0, \quad 2b_1 - b_2 = a_1,$$

and

$$a_2 = (a_1 + 4a_2 + 9)x + b_1 + 4b_2 - 24 \quad \Rightarrow a_1 + 4a_2 + 9 = 0, \quad b_1 + 4b_2 - 24 = a2.$$

Solving for the constants gives

$$a_1 = -1, \quad a_2 = -2, \quad b_1 = 2, \quad b_2 = 5.$$

Thus, the particular solution is

$$\boldsymbol{y}_p(x) = \begin{bmatrix} -x + 2 \\ -2x + 5 \end{bmatrix}$$

and the general solution is $\boldsymbol{y}_g(x) = \boldsymbol{y}_h(x) + \boldsymbol{y}_p(x)$.

It is important to note that even though the first component of $\boldsymbol{f}(x)$ was 0, that is not the case in $\boldsymbol{y}_p(x)$.

□

4.5.2. Method of Variation of Parameters. The Method of Variation of Parameters can be applied, at least theoretically, to nonhomogeneous systems with nonconstant matrix $A(x)$ and general vector function $\boldsymbol{f}(x)$. A fundamental matrix solution

$$Y(x) = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n],$$

of the homogeneous system

$$\boldsymbol{y}' = A(x)\boldsymbol{y}$$

satisfies the equation

$$Y'(x) = A(x)Y(x).$$

Since the columns of $Y(x)$ are linearly independent, the general solution $\boldsymbol{y}_h(x)$ of the homogeneous system is a linear combinations of these columns,

$$\boldsymbol{y}_h(x) = Y(x)\boldsymbol{c},$$

where $\boldsymbol{c}$ is an arbitrary n-vector. The Method of Variation of Parameters seeks a particular solution $\boldsymbol{y}_p(x)$ to the nonhomogeneous system

$$\boldsymbol{y}' = A(x)\boldsymbol{y} + \boldsymbol{f}(x)$$

in the form

$$\boldsymbol{y}_p(x) = Y(x)\boldsymbol{c}(x).$$

Substituting this expression in the nonhomogeneous system, we obtain

$$Y'\boldsymbol{c} + Y\boldsymbol{c}' = AY\boldsymbol{c} + \boldsymbol{f}.$$

Since $Y' = AY$, therefore $Y'\boldsymbol{c} = AY\boldsymbol{c}$. Thus, the previous expression reduces to

$$Y\boldsymbol{c}' = \boldsymbol{f}.$$

The fundamental matrix solution being invertible, we have

$$\boldsymbol{c}'(x) = Y^{-1}(x)\boldsymbol{f}(x), \quad \text{or} \quad \boldsymbol{c}(x) = \int_0^x Y^{-1}(s)\boldsymbol{f}(s)\,ds.$$

It follows that

$$\boldsymbol{y}_p(x) = Y(x)\int_0^x Y^{-1}(s)\boldsymbol{f}(s)\,ds.$$

In the case of an initial value problem with

$$\boldsymbol{y}(0) = \boldsymbol{y}_0,$$

the unique solution is

$$\boldsymbol{y}(x) = Y(x)Y^{-1}(0)\boldsymbol{y}_0 + Y(x)\int_0^x Y^{-1}(s)\boldsymbol{f}(s)\,ds.$$

It is left to the reader to solve Example 4.9 by the Method of Variation of Parameters.

CHAPTER 5

Laplace Transform

5.1. Definition

DEFINITION 5.1. Let $f(t)$ be a function defined on $[0, \infty)$. The *Laplace Transform* $F(s)$ of $f(t)$ is defined by the integral

$$\mathcal{L}\{f\}(s) := F(s) = \int_0^\infty e^{-st} f(t)\, dt, \tag{5.1}$$

provided the integral exists for $s > \gamma$. Note that the integral defining the Laplace transform is improper and, hence, it need not converge. Theorem 5.2 will specify conditions on $f(t)$ for the integral to exist, that is, for $F(s)$ to exist. If the integral exists, we say that $f(t)$ is *transformable* and that it is the *original* of $F(s)$.

We see that the function

$$f(t) = e^{t^2}$$

is not transformable since the integral (5.1) does not exist for any $s > 0$.

We illustrate the definition of Laplace transform by means of a few examples.

EXAMPLE 5.1. Find the Laplace transform of the function $f(t) = 1$.

SOLUTION. **(a) The analytic solution.**—

$$\begin{aligned}\mathcal{L}\{1\}(s) &= \int_0^\infty e^{-st}\, dt, \qquad s > 0,\\ &= -\frac{1}{s}\, e^{-st}\Big|_0^\infty = -\frac{1}{s}\,(0 - 1)\\ &= \frac{1}{s}.\end{aligned}$$

The condition $s > 0$ is required here for the integral to converge as

$$\lim_{t\to\infty} e^{-st} = 0 \quad \text{only if } s > 0.$$

(b) The Matlab symbolic solution.—

```
>> f = sym('Heaviside(t)');
>> F = laplace(f)
F = 1/s
```

The function `Heaviside` is a Maple function. Help for Maple functions is obtained by the command `mhelp`. □

EXAMPLE 5.2. Show that

$$\mathcal{L}\left\{e^{at}\right\}(s) = \frac{1}{s-a}, \qquad s > a. \tag{5.2}$$

SOLUTION. **(a) The analytic solution.**— Assuming that $s > a$, we have

$$\begin{aligned}\mathcal{L}\left\{e^{at}\right\}(s) &= \int_0^\infty e^{-st}\, e^{at}\, dt \\ &= \int_0^\infty e^{-(s-a)t}\, dt \\ &= -\frac{1}{s-a}\Big[e^{-(s-a)t}\Big|_0^\infty \\ &= \frac{1}{s-a}.\end{aligned}$$

Again, the condition $s - a > 0$ is required here for the integral to converge as

$$\lim_{t\to\infty} e^{-(s-a)t} = 0 \quad \text{only if } s - a > 0.$$

(b) The Matlab symbolic solution.—

```
>> syms a t;
>> f = exp(a*t);
>> F = laplace(f)
F = 1/(s-a)
```

□

THEOREM 5.1. *The Laplace transform*

$$\mathcal{L} : f(t) \;\mapsto\; F(s)$$

is a linear operator.

PROOF.

$$\begin{aligned}\mathcal{L}\{af + bg\} &= \int_0^\infty e^{-st}\big[af(t) + bg(t)\big]\, dt \\ &= a\int_0^\infty e^{-st} f(t)\, dt + b\int_0^\infty e^{-st} g(t)\, dt \\ &= a\mathcal{L}\{f\}(s) + b\mathcal{L}\{g\}(s). \quad \square\end{aligned}$$

EXAMPLE 5.3. Find the Laplace transform of the function $f(t) = \cosh at$.

SOLUTION. **(a) The analytic solution.**— Since

$$\cosh at = \frac{1}{2}\left(e^{at} + e^{-at}\right),$$

we have

$$\begin{aligned}\mathcal{L}\{\cosh at\}(s) &= \frac{1}{2}\left[\mathcal{L}\left\{e^{at}\right\} + \mathcal{L}\left\{e^{-at}\right\}\right] \\ &= \frac{1}{2}\left[\frac{1}{s-a} + \frac{1}{s+a}\right] \\ &= \frac{s}{s^2 - a^2}.\end{aligned}$$

(b) The Matlab symbolic solution.—

```
>> syms a t;
>> f = cosh(a*t);
>> F = laplace(f)
F = s/(s^2-a^2)
```

□

EXAMPLE 5.4. Find the Laplace transform of the function $f(t) = \sinh at$.

SOLUTION. **(a) The analytic solution.**— Since

$$\sinh at = \frac{1}{2}\left(e^{at} - e^{-at}\right),$$

we have

$$\begin{aligned}\mathcal{L}\{\sinh at\}(s) &= \frac{1}{2}\left[\mathcal{L}\left\{e^{at}\right\} - \mathcal{L}\left\{e^{-at}\right\}\right] \\ &= \frac{1}{2}\left[\frac{1}{s-a} - \frac{1}{s+a}\right] \\ &= \frac{a}{s^2 - a^2}.\end{aligned}$$

(b) The Matlab symbolic solution.—

```
>> syms a t;
>> f = sinh(a*t);
>> F = laplace(f)
F = a/(s^2-a^2)
```

□

REMARK 5.1. We see that $\mathcal{L}\{\cosh at\}(s)$ is an even function of a and $\mathcal{L}\{\sinh at\}(s)$ is an odd function of a.

EXAMPLE 5.5. Find the Laplace transform of the function $f(t) = t^n$.

SOLUTION. We proceed by induction. Suppose that

$$\mathcal{L}\{t^{n-1}\}(s) = \frac{(n-1)!}{s^n}.$$

This formula is true for $n = 1$,

$$\mathcal{L}\{1\}(s) = \frac{0!}{s^1} = \frac{1}{s}.$$

If $s > 0$, by integration by parts, we have

$$\begin{aligned}\mathcal{L}\{t^n\}(s) &= \int_0^\infty e^{-st} t^n \, dt \\ &= -\frac{1}{s}\left[t^n \, e^{-st}\Big|_0^\infty + \frac{n}{s}\int_0^\infty e^{-st} t^{n-1} \, dt\right. \\ &= \frac{n}{s}\,\mathcal{L}\{t^{n-1}\}(s).\end{aligned}$$

Now, the induction hypothesis gives

$$\begin{aligned}\mathcal{L}\{t^n\}(s) &= \frac{n}{s}\,\frac{(n-1)!}{s^n}\\ &= \frac{n!}{s^{n+1}}, \qquad s>0. \quad \square\end{aligned}$$

Symbolic Matlab finds the Laplace transform of, say, t^5 by the commands

```
>> syms t
>> f = t^5;
>> F = laplace(f)
F = 120/s^6
```

or

```
>> F = laplace(sym('t^5'))
F = 120/s^6
```

or

```
>> F = laplace(sym('t')^5)
F = 120/s^6
```

EXAMPLE 5.6. Find the Laplace transform of the functions $\cos\omega t$ and $\sin\omega t$.

SOLUTION. **(a) The analytic solution.**— Using Euler's identity,

$$e^{i\omega t} = \cos\omega t + i\sin\omega t, \qquad i=\sqrt{-1},$$

and assuming that $s>0$, we have

$$\begin{aligned}\mathcal{L}\left\{e^{i\omega t}\right\}(s) &= \int_0^\infty e^{-st}\,e^{i\omega t}\,dt \qquad (s>0)\\ &= \int_0^\infty e^{-(s-i\omega)t}\,dt\\ &= -\frac{1}{s-i\omega}\left[e^{-(s-i\omega)t}\right|_0^\infty\\ &= -\frac{1}{s-i\omega}\left[e^{-st}e^{i\omega t}\right|_{t\to\infty} - 1\Big]\\ &= \frac{1}{s-i\omega} = \frac{1}{s-i\omega}\,\frac{s+i\omega}{s+i\omega}\\ &= \frac{s+i\omega}{s^2+\omega^2}.\end{aligned}$$

By the linearity of $\mathcal{L}$, we have

$$\begin{aligned}\mathcal{L}\left\{e^{i\omega t}\right\}(s) &= \mathcal{L}\{\cos\omega t + i\sin\omega t\}\\ &= \mathcal{L}\{\cos\omega t\} + i\mathcal{L}\{\sin\omega t\}\\ &= \frac{s}{s^2+\omega^2} + i\frac{\omega}{s^2+\omega^2}\end{aligned}$$

Hence,

$$\mathcal{L}\{\cos\omega t\} = \frac{s}{s^2+\omega^2}, \tag{5.3}$$

which is an even function of ω, and

$$\mathcal{L}\{\sin\omega t\} = \frac{\omega}{s^2+\omega^2}, \tag{5.4}$$

which is an odd function of ω.

(b) The Matlab symbolic solution.—

```
>> syms omega t;
>> f = cos(omega*t);
>> g = sin(omega*t);
>> F = laplace(f)
F = s/(s^2+omega^2)
>> G = laplace(g)
G = omega/(s^2+omega^2)
```

□

In the sequel, we shall implicitly assume that the Laplace transforms of the functions considered in this chapter exist and can be differentiated and integrated under additional conditions. The bases of these assumptions are found in the following definition and theorem. The general formula for the inverse transform, which is not introduced in this chapter, also requires the following results.

DEFINITION 5.2. A function $f(t)$ is said to be of *exponential type of order* γ if there are constants γ, $M > 0$ and $T > 0$, such that

$$|f(t)| \leq M\, e^{\gamma t}, \qquad \text{for all } t > T. \tag{5.5}$$

The least upper bound γ_0 of all values of γ for which (5.5) holds is called the *abscissa of convergence of* $f(t)$.

THEOREM 5.2. *If the function $f(t)$ is piecewise continuous on the interval $[0, \infty)$ and if γ_0 is the abscissa of convergence of $f(t)$, then the integral*

$$\int_0^\infty e^{-st} f(t)\, dt$$

is absolutely and uniformly convergent for all $s > \gamma_0$.

PROOF. We prove only the absolute convergence:

$$\left| \int_0^\infty e^{-st} f(t)\, dt \right| \leq \int_0^\infty M\, e^{-(s-\gamma_0)t}\, dt$$
$$= -\frac{M}{s-\gamma_0}\, e^{-(s-\gamma_0)t} \Big|_0^\infty = \frac{M}{s-\gamma_0}. \quad \square$$

The integral formula for the Inverse Laplace Transform $f(t) = \mathcal{L}^{-1}\{F(s)\}$ is a path integral over a complex variable and, as such, we will not use it. Rather, we will find inverse transforms using tables (see Chapter 13 and the two pages of formulas at the end of these Notes).

EXAMPLE 5.7. Use tables to find the Laplace transform of the two given functions.

SOLUTION.

$$(i)\quad \mathcal{L}\{3t^2-\sqrt{2}\,t+2\,e^{-4t}\} = 3\mathcal{L}\{t^2\}-\sqrt{2}\,\mathcal{L}\{t\}+2\mathcal{L}\{e^{-4t}\}$$
$$= 3\left(\frac{2}{s^3}\right)-\sqrt{2}\left(\frac{1}{s^2}\right)+2\left(\frac{1}{s+4}\right)$$
$$= \frac{6}{s^3}+\frac{\sqrt{2}}{s^2}+\frac{2}{s+4}.$$

$$(ii)\quad \mathcal{L}^{-1}\left\{\frac{2}{s^4}-\frac{5}{s^2+9}\right\} = 2\mathcal{L}^{-1}\left\{\frac{1}{s^4}\right\}-5\mathcal{L}^{-1}\left\{\frac{1}{s^2+9}\right\}$$
$$= \frac{2}{6}\mathcal{L}^{-1}\left\{\frac{6}{s^4}\right\}-\frac{5}{3}\mathcal{L}^{-1}\left\{\frac{3}{s^2+9}\right\}$$
$$= \frac{1}{3}\,t^3-\frac{5}{3}\,\sin(3t). \quad \square$$

5.2. Transforms of Derivatives and Integrals

In view of applications to ordinary differential equations, one needs to know how to transform the derivative of a function.

THEOREM 5.3.

$$\mathcal{L}\{f'\}(s) = s\mathcal{L}\{f\}-f(0), \quad s>0. \tag{5.6}$$

PROOF. Integrating by parts, we have

$$\mathcal{L}\{f'\}(s) = \int_0^\infty e^{-st}f'(t)\,dt$$
$$= e^{-st}f(t)\Big|_0^\infty-(-s)\int_0^\infty e^{-st}f(t)\,dt$$
$$= s\mathcal{L}\{f\}(s)-f(0)$$

since $e^{-st}f(t)\to 0$ as $t\to\infty$ by assumption of the existence of $\mathcal{L}\{f\}$. $\square$

REMARK 5.2. The following formulae are obtained by induction.

$$\mathcal{L}\{f''\}(s) = s^2\mathcal{L}\{f\}-sf(0)-f'(0), \tag{5.7}$$
$$\mathcal{L}\{f'''\}(s) = s^3\mathcal{L}\{f\}-s^2f(0)-sf'(0)-f''(0). \tag{5.8}$$

In fact,

$$\mathcal{L}\{f''\}(s) = s\mathcal{L}\{f'\}(s)-f'(0)$$
$$= s[s\mathcal{L}\{f\}(s)-f(0)]-f'(0)$$
$$= s^2\mathcal{L}\{f\}-sf(0)-f'(0)$$

and

$$\mathcal{L}\{f'''\}(s) = s\mathcal{L}\{f''\}(s)-f''(0)$$
$$= s[s^2\mathcal{L}\{f\}-sf(0)-f'(0)]-f''(0)$$
$$= s^3\mathcal{L}\{f\}-s^2f(0)-sf'(0)-f''(0). \quad \square$$

The following general theorem follows by induction.

THEOREM 5.4. *Let the functions $f(t), f'(t), \ldots, f^{(n-1)}(t)$ be continuous for $t \geq 0$ and $f^{(n)}(t)$ be transformable for $s \geq \gamma$. Then*

$$\mathcal{L}\left\{f^{(n)}\right\}(s) = s^n\mathcal{L}\{f\} - s^{n-1}f(0) - s^{n-2}f'(0) - \cdots - f^{(n-1)}(0). \tag{5.9}$$

PROOF. The proof is by induction as in Remark 5.2 for the cases $n = 2$ and $n = 3$. □

EXAMPLE 5.8. Use the Laplace transform to solve the following initial value problem for a damped oscillator:

$$y'' + 4y' + 3y = 0, \qquad y(0) = 3, \qquad y'(0) = 1,$$

and plot the solution.

SOLUTION. **(a) The analytic solution.**— Letting

$$\mathcal{L}\{y\}(s) = Y(s),$$

we transform the differential equation,

$$\begin{aligned}\mathcal{L}\{y''\} + 4\mathcal{L}\{y'\} + 3\mathcal{L}\{y\} &= s^2Y(s) - sy(0) - y'(0) + 4[sY(s) - y(0)] + 3Y(s)\\ &= \mathcal{L}\{0\} = 0.\end{aligned}$$

Then we have

$$(s^2 + 4s + 3)Y(s) - (s + 4)y(0) - y'(0) = 0,$$

in which we replace $y(0)$ and $y'(0)$ by their values,

$$\begin{aligned}(s^2 + 4s + 3)Y(s) &= (s + 4)y(0) + y'(0)\\ &= 3(s + 4) + 1\\ &= 3s + 13.\end{aligned}$$

We solve for the unknown $Y(s)$ and expand the right-hand side in partial fractions,

$$\begin{aligned}Y(s) &= \frac{3s + 13}{s^2 + 4s + 3}\\ &= \frac{3s + 13}{(s + 1)(s + 3)}\\ &= \frac{A}{s + 1} + \frac{B}{s + 3}.\end{aligned}$$

To compute A and B, we get rid of denominators by rewriting the last two expressions in the form

$$\begin{aligned}3s + 13 &= (s + 3)A + (s + 1)B\\ &= (A + B)s + (3A + B).\end{aligned}$$

We rewrite the first and third terms as a linear system,

$$\begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 3 \\ 13 \end{bmatrix} \implies \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix},$$

which one can solve. However, in this simple case, we can get the values of A and B by first setting $s = -1$ and then $s = -3$ in the identity

$$3s + 13 = (s + 3)A + (s + 1)B.$$

Thus

$$-3 + 13 = 2A \implies A = 5, \qquad -9 + 13 = -2B \implies B = -2.$$

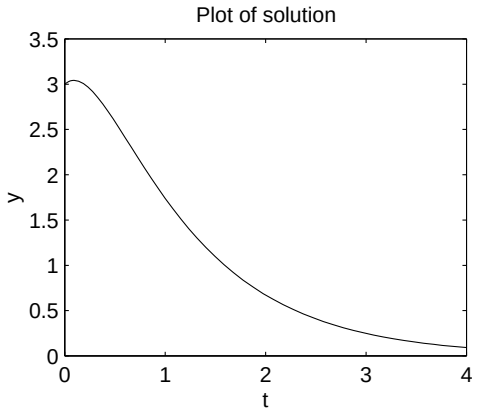

FIGURE 5.1. Graph of solution of the differential equation in Example 5.8.

Therefore

$$Y(s) = \frac{5}{s+1} - \frac{2}{s+3}.$$

We find the original by means of the inverse Laplace transform given by formula (5.2):

$$\begin{aligned} y(t) = \mathcal{L}^{-1}\{Y\} &= 5\mathcal{L}^{-1}\left\{\frac{1}{s+1}\right\} - 2\mathcal{L}^{-1}\left\{\frac{1}{s+3}\right\} \\ &= 5\,e^{-t} - 2\,e^{-3t}. \end{aligned}$$

(b) The Matlab symbolic solution.— Using the expression for $Y(s)$, we have

```
>> syms s t
>> Y = (3*s+13)/(s^2+4*s+3);
>> y = ilaplace(Y,s,t)
y = -2*exp(-3*t)+5*exp(-t)
```

(c) The Matlab numeric solution.— The function M-file `exp77.m` is

```
function yprime = exp77(t,y);
yprime = [y(2); -3*y(1)-4*y(2)];
```

and numeric Matlab solver `ode45` produces the solution.

```
>> tspan = [0 4];
>> y0 = [3;1];
>> [t,y] = ode45('exp77',tspan,y0);
>> subplot(2,2,1); plot(t,y(:,1));
>> xlabel('t'); ylabel('y'); title('Plot of solution')
```

The command `subplot` is used to produce Fig. 5.1 which, after reduction, still has large enough lettering. □

REMARK 5.3. We notice that the characteristic polynomial of the original homogeneous differential equation multiplies the function $Y(s)$ of the transformed equation when the original equation has constant coefficients.

REMARK 5.4. Solving a differential equation by the Laplace transform involves the initial values and, hence, we are solving for the unique solution directly, that is, we do not have to find the general soution first and then plug in the initial conditions. Also, if the equation is nonhomogeneous, we shall see that the method will find the particular solution as well.

Since integration is the inverse of differentiation and the Laplace transform of $f'(t)$ is essentially the transform of $f(t)$ multiplied by s, one can foresee that the transform of the indefinite integral of $f(t)$ will be the transform of $f(t)$ divided by s since division is the inverse of multiplication.

THEOREM 5.5. *Let $f(t)$ be transformable for $s \geq \gamma$. Then*

$$\mathcal{L}\left\{\int_0^t f(\tau)\,d\tau\right\} = \frac{1}{s}\mathcal{L}\{f\}, \tag{5.10}$$

or, in terms of the inverse Laplace transform,

$$\mathcal{L}^{-1}\left\{\frac{1}{s}F(s)\right\} = \int_0^t f(\tau)\,d\tau. \tag{5.11}$$

PROOF. Letting

$$g(t) = \int_0^t f(\tau)\,d\tau,$$

we have

$$\mathcal{L}\big\{f(t)\big\} = \mathcal{L}\big\{g'(t)\big\} = s\mathcal{L}\big\{g(t)\big\} - g(0).$$

Since $g(0) = 0$, we have $\mathcal{L}\{f\} = s\mathcal{L}\{g\}$, whence (5.10). □

EXAMPLE 5.9. Find $f(t)$ if

$$\mathcal{L}\{f\} = \frac{1}{s(s^2+\omega^2)}.$$

SOLUTION. **(a) The analytic solution.**— Since

$$\mathcal{L}^{-1}\left\{\frac{1}{s^2+\omega^2}\right\} = \frac{1}{\omega}\sin\omega t,$$

by (5.11) we have

$$\mathcal{L}^{-1}\left\{\frac{1}{s}\left(\frac{1}{s^2+\omega^2}\right)\right\} = \frac{1}{\omega}\int_0^t \sin\omega\tau\,d\tau = \frac{1}{\omega^2}\,(1-\cos\omega t).$$

(b) The Matlab symbolic solution.—

```
>> syms s omega t
>> F = 1/(s*(s^2+omega^2));
>> f = ilaplace(F)
f = 1/omega^2-1/omega^2*cos(omega*t)
```

Note that this example can also be done using partial fractions. □

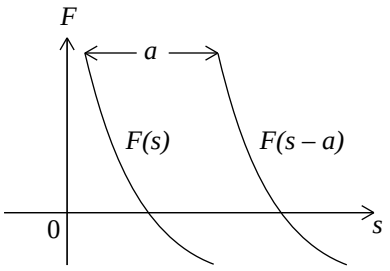

FIGURE 5.2. Function $F(s)$ and shifted function $F(s-a)$ for $a>0$.

5.3. Shifts in s and in t

In the applications, we need the original of $F(s-a)$ and the transform of $u(t-a)f(t-a)$ where $u(t)$ is the *Heaviside function*, or *unit step function*,

$$u(t)=\begin{cases}0, & \text{if } t<0,\\ 1, & \text{if } t>0.\end{cases} \tag{5.12}$$

THEOREM 5.6 (The First Shifting Theorem). *Let*

$$\mathcal{L}\{f\}(s)=F(s), \qquad s>\gamma.$$

Then

$$\mathcal{L}\big\{e^{at}f(t)\big\}(s)=F(s-a), \qquad s-a>\gamma. \tag{5.13}$$

PROOF. (See Fig. 5.2)

$$\begin{aligned}F(s-a)&=\int_0^\infty e^{-(s-a)t}f(t)\,dt\\ &=\int_0^\infty e^{-st}\left[e^{at}f(t)\right]\,dt\\ &=\mathcal{L}\big\{e^{at}f(t)\big\}(s). \quad \square\end{aligned}$$

EXAMPLE 5.10. Apply Theorem 5.6 to the three simple functions t^n, $\cos\omega t$ and $\sin\omega t$.

SOLUTION. **(a) The analytic solution.**— The results are obvious and are presented in the form of a table.

$f(t)$	$F(s)$	$e^{at}f(t)$	$F(s-a)$
t^n	$\dfrac{n!}{s^{n+1}}$	$e^{at}t^n$	$\dfrac{n!}{(s-a)^{n+1}}$
$\cos\omega t$	$\dfrac{s}{s^2+\omega^2}$	$e^{at}\cos\omega t$	$\dfrac{(s-a)}{(s-a)^2+\omega^2}$
$\sin\omega t$	$\dfrac{\omega}{s^2+\omega^2}$	$e^{at}\sin\omega t$	$\dfrac{\omega}{(s-a)^2+\omega^2}$

(b) The Matlab symbolic solution.— For the second and third functions, Matlab gives:

```
>> syms a t omega s;
>> f = exp(a*t)*cos(omega*t);
>> g = exp(a*t)*sin(omega*t);
>> F = laplace(f,t,s)
F = (s-a)/((s-a)^2+omega^2)
>> G = laplace(g,t,s)
G = omega/((s-a)^2+omega^2)
```

□

EXAMPLE 5.11. Find the solution of the damped system:

$$y'' + 2y' + 5y = 0, \qquad y(0) = 2, \qquad y'(0) = -4,$$

by means of Laplace transform.

SOLUTION. Setting

$$\mathcal{L}\{y\}(s) = Y(s),$$

we have

$$s^2Y(s) - sy(0) - y'(0) + 2[sY(s) - y(0)] + 5Y(s) = \mathcal{L}\{0\} = 0.$$

We group the terms containing $Y(s)$ on the left-hand side,

$$\begin{aligned}(s^2 + 2s + 5)Y(s) &= sy(0) + y'(0) + 2y(0)\\ &= 2s - 4 + 4\\ &= 2s.\end{aligned}$$

We solve for $Y(s)$ and rearrange the right-hand side,

$$\begin{aligned}Y(s) &= \frac{2s}{s^2 + 2s + 1 + 4}\\ &= \frac{2(s+1) - 2}{(s+1)^2 + 2^2}\\ &= \frac{2(s+1)}{(s+1)^2 + 2^2} - \frac{2}{(s+1)^2 + 2^2}.\end{aligned}$$

Hence, the solution is

$$y(t) = 2\,e^{-t}\cos 2t - e^{-t}\sin 2t. \quad \square$$

DEFINITION 5.3. The translate $u_a(t) = u(t-a)$ of the Heaviside function $u(t)$, called a unit step function, is the function (see Fig. 5.3)

$$u_a(t) := u(t-a) = \begin{cases} 0, & \text{if } t < a,\\ 1, & \text{if } t > a, \end{cases} \qquad a \geq 0. \tag{5.14}$$

The notation $\alpha(t)$ or $H(t)$ is also used for $u(t)$. In symbolic Matlab, the Maple Heaviside function is accessed by the commands

```
>> sym('Heaviside(t)')
>> u = sym('Heaviside(t)')
u = Heaviside(t)
```

Help to Maple functions is obtained by the command `mhelp`.

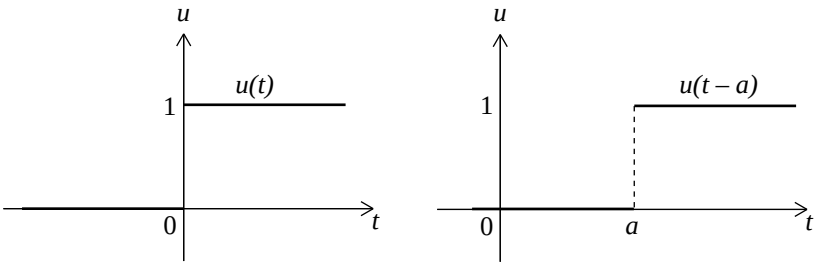

FIGURE 5.3. The Heaviside function $u(t)$ and its translate $u(t-a)$, $a>0$.

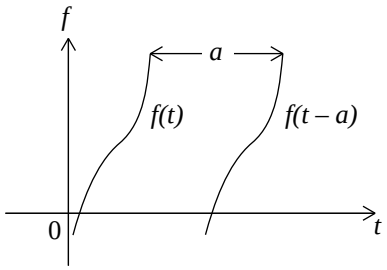

FIGURE 5.4. Shift $f(t-a)$ of the function $f(t)$ for $a>0$.

THEOREM 5.7 (The Second Shifting Theorem). *Let*

$$\mathcal{L}\{f\}(s)=F(s).$$

Then

$$\mathcal{L}^{-1}\left\{e^{-as}F(s)\right\}=u(t-a)f(t-a), \tag{5.15}$$

that is,

$$\mathcal{L}\big\{u(t-a)f(t-a)\big\}(s)=e^{-as}F(s), \tag{5.16}$$

or, equivalently,

$$\mathcal{L}\big\{u(t-a)f(t)\big\}(s)=e^{-as}\mathcal{L}\big\{f(t+a)\big\}(s). \tag{5.17}$$

PROOF. (See Fig. 5.4)

$$\begin{aligned}
e^{-as}F(s) &= e^{-as}\int_0^\infty e^{-s\tau}f(\tau)\,d\tau \\
&= \int_0^\infty e^{-s(\tau+a)}f(\tau)\,d\tau \\
&\qquad \text{(setting } \tau+a=t,\ d\tau=dt\text{)} \\
&= \int_a^\infty e^{-st}f(t-a)\,dt \\
&= \int_0^a e^{-st}0f(t-a)\,dt+\int_a^\infty e^{-st}1f(t-a)\,dt \\
&= \int_0^\infty e^{-st}u(t-a)f(t-a)\,dt
\end{aligned}$$

$$= \mathcal{L}\big\{u(t-a)f(t-a)\big\}(s).$$

The equivalent formula (5.17) is obtained by a similar change of variable:

$$\begin{aligned}
\mathcal{L}\big\{u(t-a)f(t)\big\}(s) &= \int_0^\infty e^{-st}u(t-a)f(t)\,dt \\
&= \int_a^\infty e^{-st}f(t)\,dt \\
&\qquad \text{(setting } t=\tau+a,\ d\tau=dt\text{)} \\
&= \int_0^\infty e^{-s(\tau+a)}f(\tau+a)\,d\tau \\
&= e^{-as}\int_0^\infty e^{-st}f(t+a)\,dt \\
&= e^{-as}\mathcal{L}\big\{f(t+a)\big\}(s). \quad \square
\end{aligned}$$

The equivalent formula (5.17) may simplify computation as will be seen in some of the following examples.

As a particular case, we see that

$$\mathcal{L}\big\{u(t-a)\big\} = \frac{e^{-as}}{s}, \qquad s>0.$$

In this example, $f(t-a)=f(t)=1$ and we know that $\mathcal{L}\{1\}=1/s$. This formula is a direct consequence of the definition,

$$\begin{aligned}
\mathcal{L}\big\{u(t-a)\big\} &= \int_0^\infty e^{-st}u(t-a)\,dt \\
&= \int_0^a e^{-st}\,0\,dt + \int_a^\infty e^{-st}\,1\,dt \\
&= -\frac{1}{s}\,e^{-st}\Big|_a^\infty = \frac{e^{-as}}{s}.
\end{aligned}$$

The Heaviside function will allow us to write piecewise defined functions as simple expressions. Since

$$u(t-a) = \begin{cases} 0, & \text{if } t<a, \\ 1, & \text{if } t>a, \end{cases}$$

we will have that

$$u(t-a)g(t) = \begin{cases} 0, & \text{if } t<a, \\ g(t), & \text{if } t>a, \end{cases}$$

for any function $g(t)$. See Fig. 5.5(a).

Also,

$$1-u(t-a) = \begin{cases} 1, & \text{if } t<a, \\ 0, & \text{if } t>a, \end{cases}$$

shown in Fig. 5.5(b), and so

$$[1-u(t-a)]g(t) = \begin{cases} g(t), & \text{if } t<a, \\ 0, & \text{if } t>a. \end{cases}$$

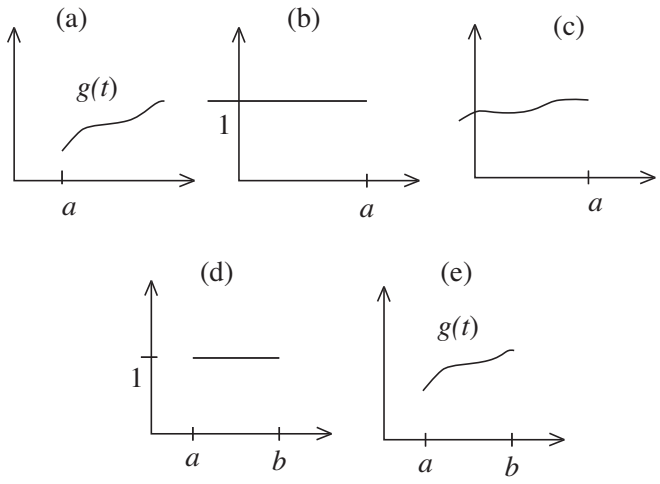

FIGURE 5.5. Figures involving the Heaviside function.

See Fig. 5.5(c). And, if $a < b$,

$$u(t-a) - u(t-b) = \begin{cases} 0, & \text{if } t < a, \\ 1, & \text{if } a < t < b, \\ 0, & \text{if } t > b, \end{cases}$$

shown in Fig. 5.5(d), and hence

$$[u(t-a) - u(t-b)]g(t) = \begin{cases} 0, & \text{if } t < a, \\ g(t), & \text{if } a < t < b, \\ 0, & \text{if } t > b, \end{cases}$$

shown in Fig. 5.5(e).

EXAMPLE 5.12. Find $F(s)$ if

$$f(t) = \begin{cases} 2, & \text{if } 0 < t < \pi, \\ 0, & \text{if } \pi < t < 2\pi, \\ \sin t, & \text{if } 2\pi < t. \end{cases}$$

(See Fig. 5.6).

SOLUTION. We rewrite $f(t)$ using the Heaviside function and the 2π-periodicity of $\sin t$:

$$f(t) = 2 - 2u(t-\pi) + u(t-2\pi)\sin(t-2\pi).$$

Then

$$\begin{aligned} F(s) &= 2\mathcal{L}\{1\} - 2\mathcal{L}\big\{u(t-\pi)1(t-\pi)\big\} + \mathcal{L}\big\{u(t-2\pi)\sin(t-2\pi)\big\} \\ &= \frac{2}{s} - e^{-\pi s}\frac{2}{s} + e^{-2\pi s}\frac{1}{s^2+1}. \quad \square \end{aligned}$$

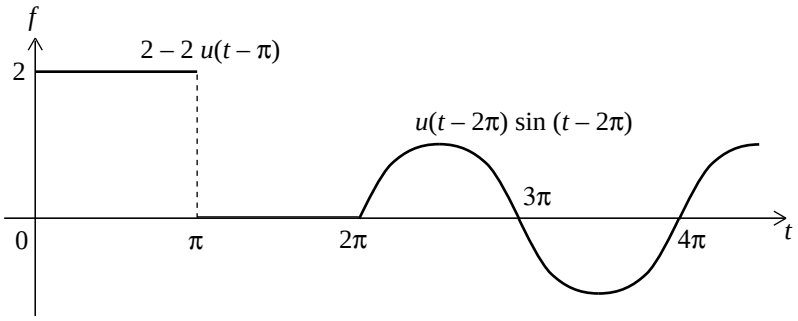

FIGURE 5.6. The function $f(t)$ of example 5.12.

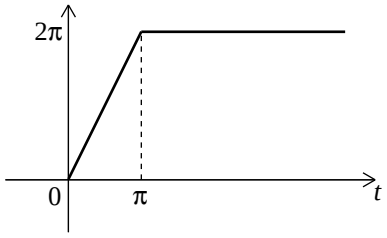

FIGURE 5.7. The function $f(t)$ of example 5.13.

EXAMPLE 5.13. Find $F(s)$ if

$$f(t) = \begin{cases} 2t, & \text{if } 0 < t < \pi, \\ 2\pi, & \text{if } \pi < t. \end{cases}$$

(See Fig. 5.7).

SOLUTION. We rewrite $f(t)$ using the Heaviside function:

$$\begin{aligned} f(t) &= 2t - u(t-\pi)(2t) + u(t-\pi)2\pi \\ &= 2t - 2u(t-\pi)(t-\pi). \end{aligned}$$

Then, by (5.16)

$$F(s) = \frac{2 \times 1!}{s^2} - 2\,e^{-\pi s}\,\frac{1}{s^2}. \quad \square$$

EXAMPLE 5.14. Find $F(s)$ if

$$f(t) = \begin{cases} 0, & \text{if } 0 \le t < 2, \\ t, & \text{if } 2 < t. \end{cases}$$

(See Fig. 5.8).

SOLUTION. We rewrite $f(t)$ using the Heaviside function:

$$\begin{aligned} f(t) &= u(t-2)t \\ &= u(t-2)(t-2) + u(t-2)2. \end{aligned}$$

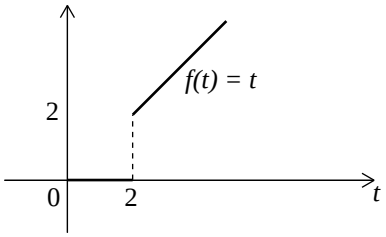

FIGURE 5.8. The function $f(t)$ of example 5.14.

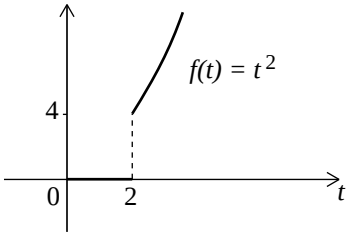

FIGURE 5.9. The function $f(t)$ of example 5.15.

Then, by (5.16),

$$\begin{aligned} F(s) &= e^{-2s}\,\frac{1!}{s^2} + 2\,e^{-2s}\,\frac{0!}{s} \\ &= e^{-2s}\left[\frac{1}{s^2} + \frac{2}{s}\right]. \end{aligned}$$

Equivalently, by (5.17),

$$\begin{aligned} \mathcal{L}\big\{u(t-2)f(t)\big\}(s) &= e^{-2s}\mathcal{L}\big\{f(t+2)\big\}(s) \\ &= e^{-2s}\mathcal{L}\big\{t+2\big)\}(s) \\ &= e^{-2s}\left[\frac{1}{s^2} + \frac{2}{s}\right]. \quad \square \end{aligned}$$

EXAMPLE 5.15. Find $F(s)$ if

$$f(t) = \begin{cases} 0, & \text{if } 0 \le t < 2, \\ t^2, & \text{if } 2 < t. \end{cases}$$

(See Fig. 5.9).

SOLUTION. **(a) The analytic solution.**— We rewrite $f(t)$ using the Heaviside function:

$$\begin{aligned} f(t) &= u(t-2)t^2 \\ &= u(t-2)\big[(t-2)+2\big]^2 \\ &= u(t-2)\left[(t-2)^2 + 4(t-2) + 4\right]. \end{aligned}$$

Then, by (5.16),

$$F(s) = e^{-2s}\left[\frac{2!}{s^3} + \frac{4}{s^2} + \frac{4}{s}\right].$$

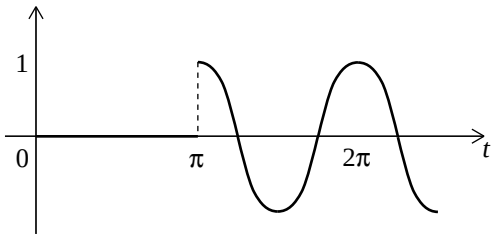

FIGURE 5.10. The function $f(t)$ of example 5.16.

Equivalently, by (5.17),

$$\begin{aligned}\mathcal{L}\big\{u(t-2)f(t)\big\}(s) &= e^{-2s}\mathcal{L}\big\{f(t+2)\big\}(s)\\ &= e^{-2s}\mathcal{L}\big\{(t+2)^2\big\}(s)\\ &= e^{-2s}\mathcal{L}\big\{t^2+4t+4\big\}(s)\\ &= e^{-2s}\left[\frac{2!}{s^3}+\frac{4}{s^2}+\frac{4}{s}\right].\end{aligned}$$

(b) The Matlab symbolic solution.—

```
syms s t
F = laplace('Heaviside(t-2)'*((t-2)^2+4*(t-2)+4))
F = 4*exp(-2*s)/s+4*exp(-2*s)/s^2+2*exp(-2*s)/s^3
```

□

EXAMPLE 5.16. Find $f(t)$ if

$$F(s) = e^{-\pi s}\,\frac{s}{s^2+4}.$$

SOLUTION. **(a) The analytic solution.**— We see that

$$\begin{aligned}\mathcal{L}^{-1}\big\{F(s)\big\}(t) &= u(t-\pi)\cos\big(2(t-\pi)\big)\\ &= \begin{cases}0, & \text{if } 0\le t<\pi,\\ \cos\big(2(t-\pi)\big)=\cos 2t, & \text{if } \pi<t.\end{cases}\end{aligned}$$

We plot $f(t)$ in figure 5.10.

(b) The Matlab symbolic solution.—

```
>> syms s;
>> F = exp(-pi*s)*s/(s^2+4);
>> f = ilaplace(F)
f = Heaviside(t-pi)*cos(2*t)
```

□

EXAMPLE 5.17. Solve the following initial value problem:

$$y''+4y = g(t) = \begin{cases}t, & \text{if } 0\le t<\pi/2,\\ \pi/2, & \text{if } \pi/2<t,\end{cases}$$

$$y(0)=0, \qquad y'(0)=0,$$

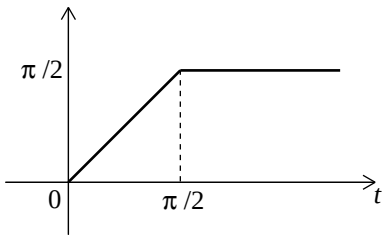

FIGURE 5.11. The function $g(t)$ of example 5.17.

by means of Laplace transform.

SOLUTION. Setting

$$\mathcal{L}\{y\} = Y(s) \quad \text{and} \quad \mathcal{L}\{g\} = G(s)$$

we have

$$\begin{aligned}\mathcal{L}\{y'' + 4y\} &= s^2Y(s) - sy(0) - y'(0) + 4Y(s)\\ &= (s^2+4)Y(s)\\ &= G(s),\end{aligned}$$

where we have used the given values of $y(0)$ and $y'(0)$. Thus

$$Y(s) = \frac{G(s)}{s^2+4}.$$

Using the Heaviside function, we rewrite $g(t)$, shown in Fig. 5.11, in the form

$$\begin{aligned}g(t) &= t - u(t-\pi/2)t + u(t-\pi/2)\frac{\pi}{2}\\ &= t - u(t-\pi/2)(t-\pi/2),\end{aligned}$$

Thus, the Laplace transform of $g(t)$ is

$$G(s) = \frac{1}{s^2} - e^{-(\pi/2)s}\,\frac{1}{s^2} = \left[1 - e^{-(\pi/2)s}\right]\,\frac{1}{s^2}.$$

It follows that

$$Y(s) = \left[1 - e^{-(\pi/2)s}\right]\,\frac{1}{(s^2+4)s^2}.$$

We expand the second factor on the right-hand side in partial fractions,

$$\frac{1}{(s^2+4)s^2} = \frac{A}{s} + \frac{B}{s^2} + \frac{Cs+D}{s^2+4}.$$

Partial fractions can be used when the polynnomial in the numerator in the original expression has a degree smaller than the polynomial in the denominator (which always happens with Laplace transform). When doing the partial fraction expansion, we always take the numerators in the expansion to be one degree smaller than their respective denominators.

Ignoring denominators,

$$\begin{aligned}1 &= (s^2+4)sA + (s^2+4)B + s^2(Cs+D)\\ &= (A+C)s^3 + (B+D)s^2 + 4As + 4B.\end{aligned}$$

and identify coefficients,

$$\begin{aligned} 4A = 0 &\implies A = 0, \\ 4B = 1 &\implies B = \frac{1}{4}, \\ B + D = 0 &\implies D = -\frac{1}{4}, \\ A + C = 0 &\implies C = 0, \end{aligned}$$

whence

$$\frac{1}{(s^2+4)s^2} = \frac{1}{4}\,\frac{1}{s^2} - \frac{1}{4}\,\frac{1}{s^2+4}.$$

Thus

$$Y(s) = \frac{1}{4}\,\frac{1}{s^2} - \frac{1}{8}\,\frac{2}{s^2+2^2} - \frac{1}{4}\,e^{-(\pi/2)s}\frac{1}{s^2} + \frac{1}{8}\,e^{-(\pi/2)s}\frac{2}{s^2+2^2}$$

and, taking the inverse Laplace transform, we have

$$y(t) = \frac{1}{4}\,t - \frac{1}{8}\sin 2t - \frac{1}{4}\,u(t-\pi/2)(t-\pi/2) + \frac{1}{8}\,u(t-\pi/2)\sin\bigl(2[t-\pi/2]\bigr). \quad \square$$

A second way of finding the inverse Laplace transform of the function

$$Y(s) = \frac{1}{2}\left[1 - e^{-(\pi/2)s}\right]\,\frac{2}{(s^2+4)s^2}$$

of previous example 5.17 is a double integration by means of formula (5.11) of Theorem 5.5, that is,

$$\begin{aligned} \mathcal{L}^{-1}\left\{\frac{1}{s}\,\frac{2}{s^2+2^2}\right\} &= \int_0^t \sin 2\tau\,d\tau = \frac{1}{2} - \frac{1}{2}\,\cos(2t), \\ \mathcal{L}^{-1}\left\{\frac{1}{s}\,\left[\frac{1}{s}\frac{2}{s^2+2^2}\right]\right\} &= \frac{1}{2}\int_0^t (1-\cos 2\tau)\,d\tau = \frac{t}{2} - \frac{1}{4}\,\sin(2t). \end{aligned}$$

The inverse Laplace transform $y(t)$ is obtained by (5.15) of Theorem 5.7.

5.4. Dirac Delta Function

Consider the function, called a *unit impulse*,

$$f_k(t;a) = \begin{cases} 1/k, & \text{if } a \le t \le a+k, \\ 0, & \text{otherwise.} \end{cases} \tag{5.18}$$

We see that the integral of $f_k(t;a)$ is equal to 1,

$$I_k = \int_0^\infty f_k(t;a)\,dt = \int_a^{a+k} \frac{1}{k}\,dt = 1. \tag{5.19}$$

We denote by

$$\delta(t-a)$$

the limit of $f_k(t;a)$ as $k \to 0$ and call this limit *Dirac's delta function*.

We can represent $f_k(t;a)$ by means of the difference of two Heaviside functions,

$$f_k(t;a) = \frac{1}{k}\,[u(t-a) - u(t-(a+k))].$$

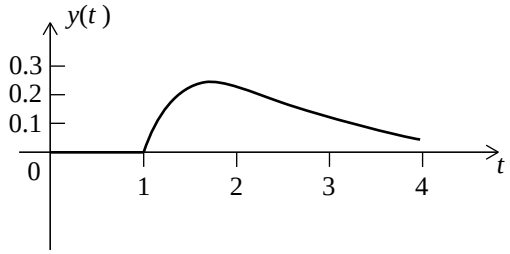

FIGURE 5.12. Solution $y(t)$ of example 5.18.

From (5.17) we have

$$\mathcal{L}\{f_k(t;a)\} = \frac{1}{ks}\left[e^{-as} - e^{-(a+k)s}\right] = e^{-as}\,\frac{1-e^{-ks}}{ks}. \tag{5.20}$$

The quotient in the last term tends to 1 as $k \to 0$ as one can see by L'Hôpital's rule. Thus,

$$\mathcal{L}\{\delta(t-a)\} = e^{-as}. \tag{5.21}$$

Symbolic Matlab produces the Laplace transform of the symbolic function $\delta(t)$ by the following commands.

```
>> syms t a
>> f = sym('Dirac(t-a)');
>> F = laplace(f)
F = [PIECEWISE(exp(-s*a), 0 <= a], [0, otherwise])
```

EXAMPLE 5.18. Solve the damped system

$$y'' + 3y' + 2y = \delta(t-a), \qquad y(0) = 0, \qquad y'(0) = 0,$$

at rest for $0 \le t < a$ and hit at time $t = a$.

SOLUTION. By (5.21), the transform of the differential equation is

$$s^2Y + 3sY + 2Y = e^{-as}.$$

We solve for $Y(s)$,

$$Y(s) = e^{-as}F(s),$$

where

$$F(s) = \frac{1}{(s+1)(s+2)} = \frac{1}{s+1} - \frac{1}{s+2}.$$

Then

$$f(t) = \mathcal{L}^{-1}(F) = e^{-t} - e^{-2t}.$$

Hence, by (5.15), we have

$$\begin{aligned} y(t) &= \mathcal{L}^{-1}\left\{e^{-as}F(s)\right\} \\ &= u(t-a)f(t-a) \\ &= \begin{cases} 0, & \text{if } 0 \le t < a, \\ e^{-(t-a)} - e^{-2(t-a)}, & \text{if } t > a. \end{cases} \end{aligned}$$

The solution for $a = 1$ is shown in figure 5.12. □

5.5. Derivatives and Integrals of Transformed Functions

We derive the following formulae.

THEOREM 5.8. *If $F(s) = \mathcal{L}\{f(t)\}(s)$, then*

$$\mathcal{L}\{tf(t)\}(s) = -F'(s), \tag{5.22}$$

or, in terms of the inverse Laplace transform,

$$\mathcal{L}^{-1}\{F'(s)\} = -tf(t). \tag{5.23}$$

Moreover, if the limit

$$\lim_{t\to 0+} \frac{f(t)}{t}$$

exists, then

$$\mathcal{L}\left\{\frac{f(t)}{t}\right\}(s) = \int_s^\infty F(\tilde{s})\, d\tilde{s}, \tag{5.24}$$

or, in terms of the inverse Laplace transform,

$$\mathcal{L}^{-1}\left\{\int_s^\infty F(\tilde{s})\, d\tilde{s}\right\} = \frac{1}{t}\, f(t). \tag{5.25}$$

PROOF. Let

$$F(s) = \int_0^\infty e^{-st} f(t)\, dt.$$

Then, by Theorem 5.2, (5.22) follows by differentiation,

$$\begin{aligned} F'(s) &= -\int_0^\infty e^{-st}[tf(t)]\, dt \\ &= -\mathcal{L}\{tf(t)\}(s). \end{aligned}$$

On the other hand, by Theorem 5.2, (5.24) follows by integration,

$$\begin{aligned} \int_s^\infty F(\tilde{s})\, d\tilde{s} &= \int_s^\infty \int_0^\infty e^{-\tilde{s}t} f(t)\, dt\, d\tilde{s} \\ &= \int_0^\infty f(t)\left[\int_s^\infty e^{-\tilde{s}t} d\tilde{s}\right] dt \\ &= \int_0^\infty f(t)\left[-\frac{1}{t}\, e^{-\tilde{s}t}\right]\Bigg|_{\tilde{s}=s}^{\tilde{s}=\infty} dt \\ &= \int_0^\infty e^{-st}\left[\frac{1}{t}\, f(t)\right] dt \\ &= \mathcal{L}\left\{\frac{1}{t}\, f(t)\right\}. \quad \square \end{aligned}$$

The following theorem generalizes formula (5.22).

THEOREM 5.9. *If $t^n f(t)$ is transformable, then*

$$\mathcal{L}\{t^n f(t)\}(s) = (-1)^n F^{(n)}(s), \tag{5.26}$$

or, in terms of the inverse Laplace transform,

$$\mathcal{L}^{-1}\{F^{(n)}(s)\} = (-1)^n t^n f(t). \tag{5.27}$$

EXAMPLE 5.19. Use (5.23) to obtain the original of $\dfrac{1}{(s+1)^2}$.

SOLUTION. Setting

$$\frac{1}{(s+1)^2} = -\frac{d}{ds}\left(\frac{1}{s+1}\right) =: -F'(s),$$

by (5.23) we have

$$\begin{aligned}\mathcal{L}^{-1}\{-F'(s)\} = tf(t) = t\,\mathcal{L}^{-1}\left\{\frac{1}{s+1}\right\}\\ = t\,e^{-t}. \quad \square\end{aligned}$$

EXAMPLE 5.20. Use (5.22) to find $F(s)$ for the given functions $f(t)$.

$f(t)$	$F(s)$	
$\frac{1}{2\beta^3}\left[\sin\beta t - \beta t\cos\beta t\right]$	$\frac{1}{(s^2+\beta^2)^2}$,	(5.28)
$\frac{t}{2\beta}\sin\beta t$	$\frac{s}{(s^2+\beta^2)^2}$,	(5.29)
$\frac{1}{2\beta}\left[\sin\beta t + \beta t\cos\beta t\right]$	$\frac{s^2}{(s^2+\beta^2)^2}$.	(5.30)

SOLUTION. **(a) The analytic solution.**— We apply (5.22) to the first term of (5.29),

$$\begin{aligned}\mathcal{L}\big\{t\sin\beta t\big\}(s) &= -\frac{d}{ds}\left[\frac{\beta}{s^2+\beta^2}\right]\\ &= \frac{2\beta s}{(s^2+\beta^2)^2},\end{aligned}$$

whence, after division by 2β, we obtain the second term of (5.29).

Similarly, using (5.22) we have

$$\begin{aligned}\mathcal{L}\big\{t\cos\beta t\big\}(s) &= -\frac{d}{ds}\left[\frac{s}{s^2+\beta^2}\right]\\ &= -\frac{s^2+\beta^2-2s^2}{(s^2+\beta^2)^2}\\ &= \frac{s^2-\beta^2}{(s^2+\beta^2)^2}.\end{aligned}$$

Then

$$\begin{aligned}\mathcal{L}\left\{\frac{1}{\beta}\,\sin\beta t \pm t\cos\beta t\right\}(s) &= \frac{1}{s^2+\beta^2} \pm \frac{s^2-\beta^2}{(s^2+\beta^2)^2}\\ &= \frac{(s^2+\beta^2)\pm(s^2-\beta^2)}{(s^2+\beta^2)^2}.\end{aligned}$$

Taking the + sign and dividing by two, we obtain (5.30). Taking the − sign and dividing by $2\beta^2$, we obtain (5.28).

(b) The Matlab symbolic solution.—

```
>> syms t beta s
>> f = (sin(beta*t)-beta*t*cos(beta*t))/(2*beta^3);
>> F = laplace(f,t,s)
F = 1/2/beta^3*(beta/(s^2+beta^2)-beta*(-1/(s^2+beta^2)+2*s^2/(s^2+beta^2)^2))
```

```
>> FF = simple(F)
FF = 1/(s^2+beta^2)^2

>> g = t*sin(beta*t)/(2*beta);
>> G = laplace(g,t,s)
G = 1/(s^2+beta^2)^2*s

>> h = (sin(beta*t)+beta*t*cos(beta*t))/(2*beta);
>> H = laplace(h,t,s)
H = 1/2/beta*(beta/(s^2+beta^2)+beta*(-1/(s^2+beta^2)+2*s^2/(s^2+beta^2)^2))
>> HH = simple(H)
HH = s^2/(s^2+beta^2)^2
```

□

EXAMPLE 5.21. Find

$$\mathcal{L}^{-1}\left\{\ln\left(1+\frac{\omega^2}{s^2}\right)\right\}(t).$$

SOLUTION. **(a) The analytic solution.**— We have

$$\begin{aligned}
-\frac{d}{ds}\ln\left(1+\frac{\omega^2}{s^2}\right) &= -\frac{d}{ds}\ln\left(\frac{s^2+\omega^2}{s^2}\right)\\
&= -\frac{s^2}{s^2+\omega^2}\,\frac{2s^3-2s(s^2+\omega^2)}{s^4}\\
&= \frac{2\omega^2}{s(s^2+\omega^2)}\\
&= 2\,\frac{(\omega^2+s^2)-s^2}{s(s^2+\omega^2)}\\
&= \frac{2}{s}-2\,\frac{s}{s^2+\omega^2}\\
&=: F(s).
\end{aligned}$$

Thus

$$f(t)=\mathcal{L}^{-1}(F)=2-2\cos\omega t.$$

Since

$$\frac{f(t)}{t}=2\omega\,\frac{1-\cos\omega t}{\omega t}\to 0 \quad \text{as} \quad t\to 0,$$

and using the fact that

$$\begin{aligned}
\int_s^\infty F(\tilde{s})\,d\tilde{s} &= -\int_s^\infty \frac{d}{d\tilde{s}}\ln\left(1+\frac{\omega^2}{\tilde{s}^2}\right)\,d\tilde{s}\\
&= -\ln\left(1+\frac{\omega^2}{\tilde{s}^2}\right)\Big|_s^\infty\\
&= -\ln 1+\ln\left(1+\frac{\omega^2}{\tilde{s}^2}\right)\\
&= \ln\left(1+\frac{\omega^2}{\tilde{s}^2}\right),
\end{aligned}$$

by (5.25) we have

$$\begin{aligned}\mathcal{L}^{-1}\left\{\ln\left(1+\frac{\omega^2}{s^2}\right)\right\} &= \mathcal{L}^{-1}\left\{\int_s^\infty F(\tilde{s})\,d\tilde{s}\right\}\\ &= \frac{1}{t}\,f(t)\\ &= \frac{2}{t}\,(1-\cos\omega t).\end{aligned}$$

(b) Alternate solution.— As

$$G(s) = \ln\left(1+\frac{\omega^2}{s^2}\right) = \ln\left(\frac{s^2+\omega^2}{s^2}\right) = \ln(s^2+\omega^2) - 2\ln s$$

then

$$G'(s) = \frac{2s}{s^2+\omega^2} - \frac{2}{s}$$

and

$$\mathcal{L}\{G'(s)\} = 2\cos(\omega t) - 2;$$

but

$$\begin{aligned}\mathcal{L}^{-1}\{G(s)\} &= -\frac{1}{t}\mathcal{L}^{-1}\{G'(s)\}\\ &= -\frac{1}{t}\,(2\cos(\omega t)-2)\\ &= -\frac{2}{t}\,(\cos(\omega t)-1).\end{aligned}$$

(c) The Matlab symbolic solution.—

```
>> syms omega t s
>> F = log(1+(omega^2/s^2));
>> f = ilaplace(F,s,t)
f = 2/t-2/t*cos(omega*t)
```

□

5.6. Laguerre Differential Equation

We can solve differential equations with variable coefficients of the form $at+b$ by means of Laplace transform. In fact, by (5.22), (5.6) and (5.7), we have

$$\begin{aligned}\mathcal{L}\big\{ty'(t)\big\} &= -\frac{d}{ds}[sY(s)-y(0)]\\ &= -Y(s)-sY'(s), \qquad (5.31)\\ \mathcal{L}\big\{ty''(t)\big\} &= -\frac{d}{ds}[s^2Y(s)-sy(0)-y'(0)]\\ &= -2sY(s)-s^2Y'(s)+y(0). \qquad (5.32)\end{aligned}$$

EXAMPLE 5.22. Find the polynomial solutions $L_n(t)$ of the *Laguerre equation*

$$ty''+(1-t)y'+ny=0, \qquad n=0,1,\ldots. \qquad (5.33)$$

SOLUTION. The Laplace transform of equation (5.33) is

$$\begin{aligned} -2sY(s) - s^2Y'(s) + y(0) + sY(s) - y(0) + Y(s) + sY'(s) + nY(s) \\ = (s - s^2)Y'(s) + (n + 1 - s)Y(s) = 0. \end{aligned}$$

This equation is separable:

$$\begin{aligned} \frac{dY}{Y} &= \frac{n+1-s}{(s-1)s}\,ds \\ &= \left(\frac{n}{s-1} - \frac{n+1}{s}\right) ds, \end{aligned}$$

whence its solution is

$$\begin{aligned} \ln|Y(s)| &= n\ln|s-1| - (n+1)\ln s \\ &= \ln\left|\frac{(s-1)^n}{s^{n+1}}\right|, \end{aligned}$$

that is,

$$Y(s) = \frac{(s-1)^n}{s^{n+1}}.$$

Note that an additive constant of integration in $\ln|Y(s)|$ would result in a multiplicative constant in $Y(s)$ which would still be a solution since Laguerre equation (5.33) is homogeneous. Set

$$L_n(t) = \mathcal{L}^{-1}\{Y\}(t),$$

where, exceptionally, capital L in L_n is a function of t. In fact, $L_n(t)$ denotes the Laguerre polynomial of degree n. We show that

$$L_0(t) = 1, \qquad L_n(t) = \frac{e^t}{n!}\,\frac{d^n}{dt^n}\left(t^n\,e^{-t}\right), \qquad n = 1, 2, \ldots.$$

We see that $L_n(t)$ is a polynomial of degree n since the exponential functions cancel each other after differentiation. Since by Theorem 5.4,

$$\mathcal{L}\left\{f^{(n)}\right\}(s) = s^nF(s) - s^{n-1}f(0) - s^{n-2}f'(0) - \cdots - f^{(n-1)}(0),$$

we have

$$\mathcal{L}\left\{\left(t^n\,e^{-t}\right)^{(n)}\right\}(s) = s^n\,\frac{n!}{(s+1)^{n+1}},$$

and, consequently,

$$\begin{aligned} \mathcal{L}\left\{\frac{e^t}{n!}\,\left(t^n\,e^{-t}\right)^{(n)}\right\} &= \frac{n!}{n!}\,\frac{(s-1)^n}{s^{n+1}} \\ &= Y(s) \\ &= \mathcal{L}\{L_n\}. \quad \square \end{aligned}$$

The first four Laguerre polynomials are (see Fig. 5.13):

$$\begin{aligned} &L_0(x) = 1, && L_1(x) = 1 - x, \\ &L_2(x) = 1 - 2x + \frac{1}{2}x^2, && L_3(x) = 1 - 3x + \frac{3}{2}x^2 - \frac{1}{6}x^3. \end{aligned}$$

We can obtain $L_n(x)$ by the recurrence formula

$$(n+1)L_{n+1}(x) = (2n + 1 - x)L_n(x) - nL_{n-1}(x).$$

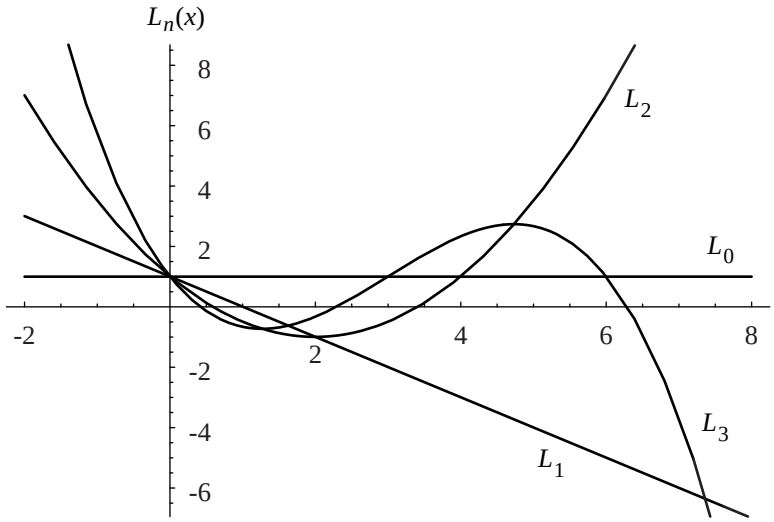

FIGURE 5.13. The first four Laguerre polynomials.

The Laguerre polynomials satisfy the following orthogonality relations with the weight $p(x) = e^{-x}$:

$$\int_0^\infty e^{-x} L_m(x) L_n(x)\, dx = \begin{cases} 0, & m \neq n, \\ 1, & m = n. \end{cases}$$

Symbolic Matlab obtains the Laguerre polynomials as follows.

```
>>L0 = dsolve('t*D2y+(1-t)*Dy=0','y(0)=1','t')
L0 = 1

>>L1 = dsolve('t*D2y+(1-t)*Dy+y=0','y(0)=1','t');
>> L1 = simple(L1)
L1 = 1-t

>> L2 = dsolve('t*D2y+(1-t)*Dy+2*y=0','y(0)=1','t');
>> L2 = simple(L2)
L2 = 1-2*t+1/2*t^2
```

and so on. The symbolic Matlab command `simple` has the mathematically unorthodox goal of finding a simplification of an expression that has the fewest number of characters.

5.7. Convolution

The original of the product of two transforms is the convolution of the two originals.

DEFINITION 5.4. The convolution of $f(t)$ with $g(t)$, denoted by $(f * g)(t)$, is the function

$$h(t) = \int_0^t f(\tau) g(t - \tau)\, d\tau. \tag{5.34}$$

We say "$f(t)$ convolved with $g(t)$".

We verify that convolution is commutative:

$$\begin{aligned}(f*g)(t) &= \int_0^t f(\tau)g(t-\tau)\,d\tau \\ &\qquad \text{(setting } t-\tau=\sigma,\ d\tau=-d\sigma) \\ &= -\int_t^0 f(t-\sigma)g(\sigma)\,d\sigma \\ &= \int_0^t g(\sigma)f(t-\sigma)\,d\sigma \\ &= (g*f)(t).\end{aligned}$$

THEOREM 5.10. *Let*

$$F(s)=\mathcal{L}\{f\},\quad G(s)=\mathcal{L}\{g\},\quad H(s)=F(s)G(s),\quad h(t)=\{L\}^{-1}(H).$$

Then

$$h(t)=(f*g)(t)=\mathcal{L}^{-1}\big(F(s)G(s)\big). \tag{5.35}$$

PROOF. By definition and by (5.16), we have

$$\begin{aligned}e^{-s\tau}G(s) &= \mathcal{L}\big\{g(t-\tau)u(t-\tau)\big\} \\ &= \int_0^\infty e^{-st}g(t-\tau)u(t-\tau)\,dt \\ &= \int_\tau^\infty e^{-st}g(t-\tau)\,dt, \qquad s>0.\end{aligned}$$

Whence, by the definition of $F(s)$ we have

$$\begin{aligned}F(s)G(s) &= \int_0^\infty e^{-s\tau}f(\tau)G(s)\,d\tau \\ &= \int_0^\infty f(\tau)\left[\int_\tau^\infty e^{-st}g(t-\tau)\,dt\right]d\tau, \qquad (s>\gamma) \\ &= \int_0^\infty e^{-st}\left[\int_0^t f(\tau)g(t-\tau)\,d\tau\right]dt \\ &= \mathcal{L}\big\{(f*g)(t)\big\}(s) \\ &= \mathcal{L}\{h\}(s).\end{aligned}$$

Figure 5.14 shows the region of integration in the $t\tau$-plane used in the proof of Theorem 5.10. □

EXAMPLE 5.23. Find $(1*1)(t)$.

SOLUTION.

$$(1*1)(t)=\int_0^t 1\times 1\,d\tau = t. \quad \square$$

EXAMPLE 5.24. Find $e^t * e^t$.

SOLUTION.

$$\begin{aligned}e^t*e^t &= \int_0^t e^\tau\, e^{t-\tau}\,d\tau \\ &= \int_0^t e^t\,d\tau = t\,e^t. \quad \square\end{aligned}$$

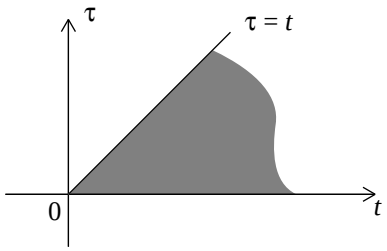

FIGURE 5.14. Region of integration in the $t\tau$-plane used in the proof of Theorem 5.10.

EXAMPLE 5.25. Find the original of

$$\frac{1}{(s-a)(s-b)}, \qquad a \neq b,$$

by means of convolution.

SOLUTION.

$$\begin{aligned}
\mathcal{L}^{-1}\left\{\frac{1}{(s-a)(s-b)}\right\} &= e^{at} * e^{bt} \\
&= \int_0^t e^{a\tau}\, e^{b(t-\tau)}\, d\tau \\
&= e^{bt} \int_0^t e^{(a-b)\tau}\, d\tau \\
&= e^{bt}\, \frac{1}{a-b}\, e^{(a-b)\tau}\Big|_0^t \\
&= \frac{e^{bt}}{a-b}\left[e^{(a-b)t} - 1\right] \\
&= \frac{e^{at} - e^{bt}}{a-b}. \quad \square
\end{aligned}$$

Some integral equations can be solved by means of Laplace transform.

EXAMPLE 5.26. Solve the integral equation

$$y(t) = t + \int_0^t y(\tau) \sin(t-\tau)\, d\tau. \tag{5.36}$$

SOLUTION. Since the last term of (5.36) is a convolution, then

$$y(t) = t + y * \sin t.$$

Hence

$$Y(s) = \frac{1}{s^2} + Y(s)\, \frac{1}{s^2+1},$$

whence

$$Y(s) = \frac{s^2+1}{s^4} = \frac{1}{s^2} + \frac{1}{s^4}.$$

Thus,

$$y(t) = t + \frac{1}{6}\, t^3. \quad \square$$

5.8. Partial Fractions

Expanding a rational function in partial fractions has been studied in elementary calculus.

We only mention that if $p(\lambda)$ is the characteristic polynomial of a differential equation, $Ly = r(t)$ with constant coefficients, the factorization of $p(\lambda)$ needed to find the zeros of p and consequently the independent solutions of $Ly = 0$, is also needed to expand $1/p(s)$ in partial fractions when one uses the Laplace transform.

Resonance corresponds to multiple zeros.

The extended symbolic toolbox of the professional Matlab gives access to the complete Maple kernel. In this case, partial fractions can be obtained by using the Maple `convert` command. This command is referenced by entering `mhelp convert[parfrac]`.

5.9. Transform of Periodic Functions

DEFINITION 5.5. A function $f(t)$ defined for all $t > 0$ is said to be periodic of period p, $p > 0$, if

$$f(t + p) = f(t), \qquad \text{for all } t > 0. \tag{5.37}$$

THEOREM 5.11. *Let $f(t)$ be a periodic function of period p. Then*

$$\mathcal{L}\{f\}(s) = \frac{1}{1 - e^{-ps}} \int_0^p e^{-st} f(t)\, dt, \qquad s > 0. \tag{5.38}$$

PROOF. To use the periodicity of $f(t)$, we write

$$\begin{aligned}\mathcal{L}\{f\}(s) &= \int_0^\infty e^{-st} f(t)\, dt \\ &= \int_0^p e^{-st} f(t)\, dt + \int_p^{2p} e^{-st} f(t)\, dt + \int_{2p}^{3p} e^{-st} f(t)\, dt + \dots .\end{aligned}$$

Substituting

$$t = \tau + p, \quad t = \tau + 2p, \quad \dots,$$

in the second, third integrals, etc., changing the limits of integration to 0 and p, and using the periodicity of $f(t)$, we have

$$\begin{aligned}\mathcal{L}\{f\}(s) &= \int_0^p e^{-st} f(t)\, dt + \int_0^p e^{-s(t+p)} f(t)\, dt + \int_0^p e^{-s(t+2p)} f(t)\, dt + \dots \\ &= \left(1 + e^{-sp} + e^{-2sp} + \cdots\right) \int_0^p e^{-st} f(t)\, dt \\ &= \frac{1}{1 - e^{-ps}} \int_0^p e^{-st} f(t)\, dt.\end{aligned}$$

Note that the series $1 + e^{-sp} + e^{-2sp} + \cdots$ in front of the integral in the second last expression is a convergent geometric series with ratio e^{-ps}. □

EXAMPLE 5.27. Find the Laplace transform of the half-wave rectification of the sine function

$$\sin \omega t$$

(see Fig. 5.15).

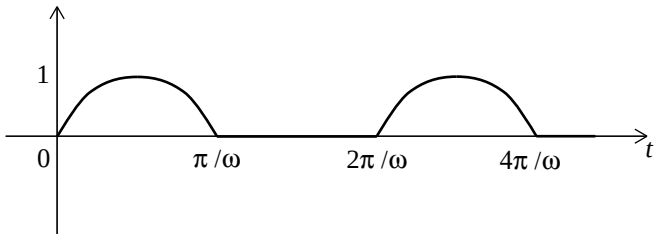

FIGURE 5.15. Half-wave rectifier of example 5.27.

SOLUTION. **(a) The analytic solution.**— The half-rectified wave of period $p = 2\pi/\omega$ is

$$f(t) = \begin{cases} \sin \omega t, & \text{if } 0 < t < \pi/\omega, \\ 0, & \text{if } \pi/\omega < t < 2\pi/\omega, \end{cases} \qquad f(t + 2\pi/\omega) = f(t).$$

By (5.38),

$$\mathcal{L}\{f\}(s) = \frac{1}{1 - e^{-2\pi s/\omega}} \int_0^{\pi/\omega} e^{-st} \sin \omega t \, dt.$$

Integrating by parts or, more simply, noting that the integral is the imaginary part of the following integral, we have

$$\int_0^{\pi/\omega} e^{(-s+i\omega)t} \, dt = \frac{1}{-s + i\omega} \, e^{(-s+i\omega)t} \Big|_0^{\pi/\omega} = \frac{-s - i\omega}{s^2 + \omega^2} \left(-e^{-s\pi/\omega} - 1 \right).$$

Using the formula

$$1 - e^{-2\pi s/\omega} = \left(1 + e^{-\pi s/\omega} \right) \left(1 - e^{-\pi s/\omega} \right),$$

we have

$$\mathcal{L}\{f\}(s) = \frac{\omega \left(1 + e^{-\pi s/\omega} \right)}{(s^2 + \omega^2) \left(1 - e^{-2\pi s/\omega} \right)} = \frac{\omega}{(s^2 + \omega^2) \left(1 - e^{-\pi s/\omega} \right)}.$$

(b) The Matlab symbolic solution.—

```
syms pi s t omega
G = int(exp(-s*t)*sin(omega*t),t,0,pi/omega)
G = omega*(exp(-pi/omega*s)+1)/(s^2+omega^2)
F = 1/(1-exp(-2*pi*s/omega))*G
F = 1/(1-exp(-2*pi/omega*s))*omega*(exp(-pi/omega*s)+1)/(s^2+omega^2)
```

□

EXAMPLE 5.28. Find the Laplace transform of the full-wave rectification of

$$f(t) = \sin \omega t$$

(see Fig. 5.16).

SOLUTION. The fully rectified wave of period $p = 2\pi/\omega$ is

$$f(t) = |\sin \omega t| = \begin{cases} \sin \omega t, & \text{if } 0 < t < \pi\omega, \\ -\sin \omega t, & \text{if } \pi < t < 2\pi\omega, \end{cases} \qquad f(t + 2\pi/\omega) = f(t).$$

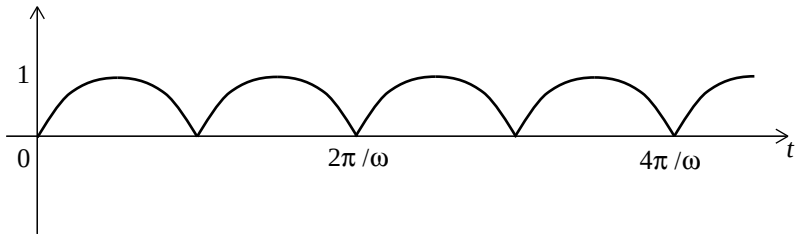

FIGURE 5.16. Full-wave rectifier of example 5.28.

By the method used in example 5.27, we have

$$\mathcal{L}\{f\}(s) = \frac{\omega}{s^2+\omega^2} \coth \frac{\pi s}{2\omega},$$

where coth is the hyperbolic cotangent. □

CHAPTER 6

Power Series Solutions

6.1. The Method

We illustrate the Power Series Method by a very simple example.

EXAMPLE 6.1. Find a power series solution of the form

$$y(x) = \sum_{m=0}^{\infty} a_m x^m = a_0 + a_1 x + a_2 x^2 + \dots$$

to the initial value problem

$$y'' + 25y = 0, \qquad y(0) = 3, \quad y'(0) = 13.$$

SOLUTION. In this simple case, we already know the general solution,

$$y(x) = a\cos 5x + b\sin 5x,$$

of the ordinary differential equation. The arbitrary constants a and b are determined by the initial conditions,

$$y(0) = a = 3 \implies a = 3,$$
$$y'(0) = 5b = 13 \implies b = \frac{13}{5}.$$

We also know that

$$\cos 5x = \sum_{m=0}^{\infty} (-1)^m \frac{(5x)^{2m}}{(2m)!} = 1 - \frac{(5x)^2}{2!} + \frac{(5x)^4}{4!} - \frac{(5x)^6}{6!} + - \dots$$

and

$$\sin 5x = \sum_{m=0}^{\infty} (-1)^m \frac{(5x)^{2m+1}}{(2m+1)!} = 5x - \frac{(5x)^3}{3!} + \frac{(5x)^5}{5!} - \frac{(5x)^7}{7!} + - \dots$$

To obtain the series solution, we substitute

$$y(x) = \sum_{m=0}^{\infty} a_m x^m = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$$

into the differential equation. If

$$y(x) = \sum_{m=0}^{\infty} a_m x^m = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots,$$

then

$$y'(x) = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots = \sum_{m=0}^{\infty} m a_m x^{m-1}$$

and

$$\begin{aligned} y''(x) &= 2a_2 + 6a_3x + 12a_4x^2 + 20a_5x^3 + \dots \\ &= \sum_{m=0}^{\infty} m(m-1)a_m x^{m-2}. \end{aligned}$$

Substituting this into the differential equation, we have

$$y'' + 25y = \sum_{m=0}^{\infty} m(m-1)a_m x^{m-2} + 25\sum_{m=0}^{\infty} a_m x^{m-1} = 0.$$

Since the sum of two power series is another power series, we can combine the terms into a single series. To ensure that we are combining the series correctly, we insist that the range of the summation indices and the powers of x agree. In our example, the powers of x are not the same, x^{m-2} and x^m. But notice that

$$\begin{aligned} y''(x) &= \sum_{m=0}^{\infty} m(m-1)a_m x^{m-2} = 0 + 0 + 2a_2 + 6a_3x + 12a_4x^2 + \dots \\ &= \sum_{m=2}^{\infty} m(m-1)a_m x^{m-2} \quad \text{(since the first two terms are zero)} \\ &= \sum_{m=0}^{\infty} (m+1)(m+2)a_{m+2}x^m \quad \text{(reset the index } \quad m \to m+2), \end{aligned}$$

and so,

$$\begin{aligned} y''(x) &= \sum_{m=0}^{\infty} (m+1)(m+2)a_{m+2}x^m + 25\sum_{m=0}^{\infty} a_m x^m \\ &= \sum_{m=0}^{\infty} \left[(m+1)(m+2)a_{m+2} + 25a_m\right] x^m \\ &= 0, \quad \text{for all } x. \end{aligned}$$

The only way that a power series can be identically zero is if all of the coefficients are zero. So it must be that

$$(m+1)(m+2)a_{m+2} + 25a_m = 0, \quad \text{for all } m,$$

or

$$a_{m+2} = -\frac{25a_m}{(m+1)(m+2)},$$

which is called the coefficient *recurrence relation*. Then

$$a_2 = -\frac{25a_0}{1\cdot 2},\ a_4 = -\frac{25a_2}{3\cdot 4} = \frac{(25^2)a_6}{1\cdot 2\cdot 3\cdot 4} = \frac{5^4 a_0}{4!},\ a_6 = -\frac{25a_4}{5\cdot 6} = -\frac{5^6 a_0}{6!},$$

and so on, that is,

$$a_{2k} = \frac{(-1)^k 5^{2k} a_0}{(2k)!}.$$

We also have

$$a_3 = -\frac{25a_1}{2\cdot 3} = -\frac{5^2 a_1}{3!}, \qquad a_5 = -\frac{25a_3}{4\cdot 5} = \frac{(25^4)a_1}{5!},$$

and so on, that is,

$$a_{2k+1} = \frac{(-1)^k 5^{2k} a_1}{(2k+1)!} = \frac{1}{5}(-1)^k \frac{5^{2k+1} a_1}{(2k+1)!}.$$

Therefore, the general solution is

$$\begin{aligned} y(x) &= \sum_{m=0}^{\infty} a_m x^m \\ &= \sum_{k=0}^{\infty} a_{2k} x^{2k} + \sum_{k=0}^{\infty} a_{2k+1} x^{2k+1} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k 5^{2k} a_0 x^{2k}}{(2k)!} + \sum_{k=0}^{\infty} \frac{1}{5} \frac{(-1)^k 5^{2k+1} a_1 x^{2k+1}}{(2k+1)!} \\ &= a_0 \sum_{k=0}^{\infty} \frac{(-1)^k ((5x)^{2k}}{(2k)!} + \frac{a_1}{5} \sum_{k=0}^{\infty} \frac{(-1)^k (5x)^{2k+1}}{(2k+1)!} \\ &= a_0 \cos(5x) + \frac{a_1}{5} \sin(5x). \end{aligned}$$

The parameter a_0 is determined by the initial condition $y(0) = 3$,

$$a_0 = 3.$$

To determine a_1, we differentiate $y(x)$,

$$y'(x) = -5a_0 \sin(5x) + a_1 \cos(5x),$$

and set $x = 0$. Thus, we have

$$y'(0) = a_1 = 13$$

by the initial condition $y'(0) = 13$. □

6.2. Foundation of the Power Series Method

It will be convenient to consider power series in the complex plane. We recall that a point z in the complex plane $\mathbb{C}$ admits the following representations:

- *Cartesian* or *algebraic*:

$$z = x + iy, \qquad i^2 = -1,$$

- *trigonometric*:

$$z = r(\cos\theta + i \sin\theta),$$

- *polar* or *exponential*:

$$z = r\, e^{i\theta},$$

where

$$r = \sqrt{x^2 + y^2}, \qquad \theta = \arg z = \arctan \frac{y}{x}.$$

As usual, $\bar{z} = x - iy$ denotes the complex conjugate of z and

$$|z| = \sqrt{x^2 + y^2} = \sqrt{z\bar{z}} = r$$

denotes the modulus of z (see Fig. 6.1).

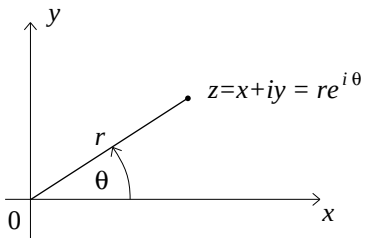

FIGURE 6.1. A point $z = x + iy = r\,e^{i\theta}$ in the complex plane $\mathbb{C}$.

EXAMPLE 6.2. Extend the function

$$f(x) = \frac{1}{1-x},$$

to the complex plane and expand it in power series with centres at $z_0 = 0$, $z_0 = -1$, and $z_0 = i$, respectively.

SOLUTION. We extend the function

$$f(x) = \frac{1}{1-x}, \qquad x \in \mathbb{R} \setminus \{1\},$$

of a real variable to the complex plane

$$f(z) = \frac{1}{1-z}, \qquad z = x + iy \in \mathbb{C}.$$

This is a rational function with a simple pole at $z = 1$. We say that $z = 1$ is a pole of $f(z)$ since $|f(z)|$ tends to $+\infty$ as $z \to 1$. Moreover, $z = 1$ is a simple pole since $1 - z$ appears to the first power in the denominator.

We expand $f(z)$ in a Taylor series around 0 (sometimes called a *Maclaurin series*),

$$f(z) = \sum_{m=0}^{\infty} \frac{1}{m!}\, f^{(m)}(0) z^m = f(0) + \frac{1}{1!} f'(0) z + \frac{1}{2!} f''(0) z^2 + \dots$$

Since

$$\begin{aligned}
f(z) &= \frac{1}{(1-z)} && \implies f(0) = 1, \\
f'(z) &= \frac{1!}{(1-z)^2} && \implies f'(0) = 1!, \\
f''(z) &= \frac{2!}{(1-z)^3} && \implies f''(0) = 2!, \\
&\vdots \\
f^{(n)}(z) &= \frac{n!}{(1-z)^{n+1}} && \implies f^{(n)}(0) = n!,
\end{aligned}$$

it follows that

$$\begin{aligned} f(z) &= \frac{1}{1-z} \\ &= 1 + z + z^2 + z^3 + \dots \\ &= \sum_{n=0}^{\infty} z^n. \end{aligned}$$

Note that the last series is a geometric series. The series converges *absolutely* for $|z| \equiv \sqrt{x^2+y^2} < 1$, that is,

$$\sum_{n=0}^{\infty} |z|^n < \infty, \qquad \text{for all } |z| < 1,$$

and *uniformly* for $|z| \le \rho < 1$, that is, given $\epsilon > 0$ there exists N_ϵ such that

$$\left| \sum_{n=N}^{\infty} z^n \right| < \epsilon, \qquad \text{for all } N > N_\epsilon \text{ and all } |z| \le \rho < 1.$$

Thus, the radius of convergence R of the series $\sum_{n=0}^{\infty} z^n$ is 1.

Now, we expand $f(z)$ in a neighbourhood of $z = -1$,

$$\begin{aligned} f(z) &= \frac{1}{1-z} = \frac{1}{1-(z+1-1)} \\ &= \frac{1}{2-(z+1)} = \frac{1}{2}\,\frac{1}{1-(\frac{z+1}{2})} \\ &= \frac{1}{2}\left\{1 + \frac{z+1}{2} + \left(\frac{z+1}{2}\right)^2 + \left(\frac{z+1}{2}\right)^3 + \left(\frac{z+1}{2}\right)^4 + \dots\right\}. \end{aligned}$$

The series converges absolutely for

$$\left|\frac{z+1}{2}\right| < 1, \quad \text{that is } |z+1| < 2, \quad \text{or } |z-(-1)| < 2.$$

The centre of the disk of convergence is $z = -1$ and the radius of convergence is $R = 2$.

Finally, we expand $f(z)$ in a neighbourhood of $z = i$,

$$\begin{aligned} f(z) &= \frac{1}{1-z} = \frac{1}{1-(z-i+i)} \\ &= \frac{1}{(1-i)-(z-i)} = \frac{1}{1-i}\,\frac{1}{1-(\frac{z-i}{1-i})} \\ &= \frac{1}{1-i}\left\{1 + \frac{z-i}{1-i} + \left(\frac{z-i}{1-i}\right)^2 + \left(\frac{z-i}{1-i}\right)^3 + \left(\frac{z-i}{1-i}\right)^4 + \dots\right\}. \end{aligned}$$

The series converges absolutely for

$$\left|\frac{z-i}{1-i}\right| < 1, \quad \text{that is} \quad |z-i| < |1-i| = \sqrt{2}.$$

We see that the centre of the disk of convergence is $z = i$ and the radius of convergence is $R = \sqrt{2} = |1-i|$ (see Fig. 6.2). □

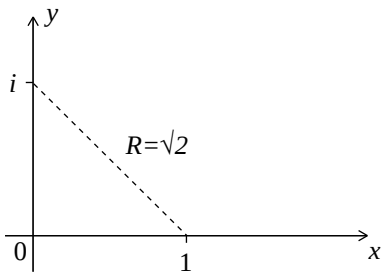

FIGURE 6.2. Distance from the centre $a = i$ to the pole $z = 1$.

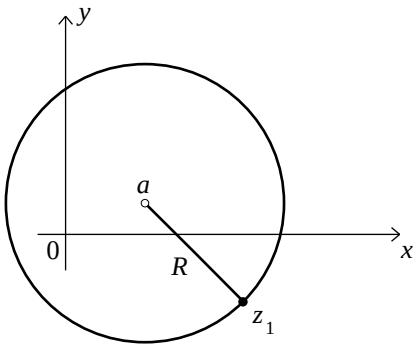

FIGURE 6.3. Distance R from the centre a to the nearest singularity z_1.

This example shows that the Taylor series expansion of a function $f(z)$, with centre $z = a$ and radius of convergence R, stops being convergent as soon as $|z - a| \geq R$, that is, as soon as $|z - a|$ is bigger than the distance from a to the nearest singularity z_1 of $f(z)$ (see Fig. 6.3).

We shall use the following result.

THEOREM 6.1 (Convergence Criteria). *The reciprocal of the radius of convergence R of a power series with centre $z = a$,*

$$\sum_{m=0}^{\infty} a_m (z - a)^m, \tag{6.1}$$

is equal to the following limit superior,

$$\frac{1}{R} = \limsup_{m \to \infty} |a_m|^{1/m}. \tag{6.2}$$

The following criterion also holds,

$$\frac{1}{R} = \lim_{m \to \infty} \left| \frac{a_{m+1}}{a_m} \right|, \tag{6.3}$$

if this limit exists.

PROOF. The Root Test, also called Cauchy's Criterion, states that the series

$$\sum_{m=0}^{\infty} c_m$$

converges if

$$\lim_{m\to\infty} |c_m|^{1/m} < 1.$$

By the Root Test, the power series converges if

$$\lim_{m\to\infty} |a_m(z-a)^m|^{1/m} = \lim_{m\to\infty} |a_m|^{1/m}|z-a| < 1.$$

Let R be maximum of $|z-a|$ such that the equality

$$\lim_{m\to\infty} |a_m|^{1/m} R = 1$$

is satisfied. If there are several limits, one must take the limit superior, that is the largest limit. This establishes criterion (6.2).

The second criterion follows from the Ratio Test, also called d'Alembert's Criterion, which states that the series

$$\sum_{m=0}^{\infty} c_m$$

converges if

$$\lim_{m\to\infty} \frac{|c_{m+1}|}{|c_m|} < 1.$$

By the Ratio Test, the power series converges if

$$\lim_{m\to\infty} \frac{|a_{m+1}(z-a)^{m+1}|}{|a_m(z-a)^m|} = \lim_{m\to\infty} \frac{|a_{m+1}|}{|a_m|}\,|z-a| < 1.$$

Let R be maximum of $|z-a|$ such that the equality

$$\lim_{m\to\infty} \frac{|a_{m+1}|}{|a_m|}\,R = 1$$

is satisfied. This establishes criterion (6.3). □

EXAMPLE 6.3. Find the radius of convergence of the series

$$\sum_{m=0}^{\infty} \frac{1}{k^m} x^{3m}$$

and of its first term-by-term derivative.

SOLUTION. By the Root Test,

$$\frac{1}{R} = \limsup_{m\to\infty} |a_m|^{1/m} = \lim_{m\to\infty} \left|\frac{1}{k^m}\right|^{1/3m} = \frac{1}{|k|^{1/3}}.$$

Hence the radius of convergence of the series is

$$R = |k|^{1/3}.$$

To use the Ratio Test, we put

$$w = z^3$$

in the series, which becomes

$$\sum_0^\infty \frac{1}{k^m} w^m.$$

Then the radius of convergence, R_1, of the new series is given by

$$\frac{1}{R_1} = \lim_{m\to\infty} \left| \frac{k^m}{k^{m+1}} \right| = \left| \frac{1}{k} \right|.$$

Therefore the original series converges for

$$|z^3| = |w| < |k|, \quad \text{that is} \quad |z| < |k|^{1/3}.$$

The radius of convergence R' of the differentiated series,

$$\sum_{m=0}^\infty \frac{3m}{k^m} x^{3m-1},$$

is obtained in a similar way:

$$\begin{aligned}
\frac{1}{R'} &= \lim_{m\to\infty} \left| \frac{3m}{k^m} \right|^{1/(3m-1)} \\
&= \lim_{m\to\infty} |3m|^{1/(3m-1)} \lim_{m\to\infty} \left| \frac{1}{k^m} \right|^{(1/m)(m/(3m-1)} \\
&= \lim_{m\to\infty} \left(\frac{1}{|k|} \right)^{1/(3-1/m)} \\
&= \frac{1}{|k|^{1/3}},
\end{aligned}$$

since

$$\lim_{m\to\infty} |3m|^{1/(3m-1)} = 1. \quad \square$$

One sees by induction that all term-by-term derivatives of a given series have the same radius of convergence R.

DEFINITION 6.1. We say that a function $f(z)$ is *analytic* inside a disk $D(a, R)$, of centre a and radius $R > 0$, if it has a power series with centre a,

$$f(z) = \sum_{n=0}^\infty a_n (z-a)^n,$$

which is *uniformly* convergent in every closed subdisk strictly contained inside $D(a, R)$.

The following theorem follows from the previous definition.

THEOREM 6.2. *A function $f(z)$ analytic in $D(a, R)$ admits the power series representation*

$$f(z) = \sum_{n=0}^\infty \frac{f^{(n)}(a)}{n!} (z-a)^n$$

uniformly and absolutely convergent in $D(a, R)$. Moreover $f(z)$ is infinitely often differentiable, the series is termwise infinitely often differentiable, and

$$f^{(k)}(z) = \sum_{n=k}^\infty \frac{f^{(n)}(a)}{(n-k)!} (z-a)^{n-k}, \qquad k = 0, 1, 2, \ldots,$$

in $D(a, R)$.

PROOF. Since the radius of convergence of the termwise differentiated series is still R, the result follows from the facts that the differentiated series converges uniformly in every closed disk strictly contained inside $D(a, R)$ and $f(z)$ is differentiable in $D(a, R)$. □

The following general theorem holds for ordinary differential equations with analytic coefficients.

THEOREM 6.3 (Existence of Series Solutions). *Consider the second-order ordinary differential equation in standard form*

$$y'' + f(x)y' + g(x)y = r(x),$$

where $f(z)$, $g(z)$ *and* $r(z)$ *are analytic functions in a circular neighbourhood of the point* a. *If* R *is equal to the minimum of the radii of convergence of the power series expansions of* $f(z)$, $g(z)$ *and* $r(z)$ *with centre* $z = a$, *then the differential equation admits an analytic solution in a disk of centre* a *and radius of convergence* R. *This general solution contains two undetermined coefficients.*

PROOF. The proof makes use of majorizing series in the complex plane $\mathbb{C}$. This method consists in finding a series with nonnegative coefficients which converges absolutely in $D(a, R)$,

$$\sum_{n=0}^{\infty} b_n(x-a)^n, \qquad b_n \geq 0,$$

and whose coefficients majorizes the absolute value of the coefficients of the solution,

$$y(x) = \sum_{n=0}^{\infty} a_n(x-a)^n,$$

that is,

$$|a_n| \leq b_n. \quad \square$$

We shall use Theorems 6.2 and 6.3 to obtain power series solutions of ordinary differential equations. In the next two sections, we shall obtain the power series solution of the Legendre equation and prove the orthogonality relation satisfied by the Legendre polynomials $P_n(x)$.

In closing this section, we revisit Examples 1.19 and 1.20.

EXAMPLE 6.4. Use the power series method to solve the initial value problem

$$y' - xy - 1 = 0, \qquad y(0) = 1.$$

SOLUTION. Putting

$$y(x) = \sum_{m=0}^{\infty} a_m x^m \quad \text{and} \quad y'(x) = \sum_{m=0}^{\infty} m a_m x^{m-1}$$

into the differential equation, we have

$$\begin{aligned}
y' - xy - 1 &= \sum_{m=0}^{\infty} m a_m x^{m-1} - x \sum_{m=0}^{\infty} a_m x^m - 1 \\
&= \sum_{m=0}^{\infty} (m+1) a_{m+1} x^m - \sum_{m=0}^{\infty} a_m x^{m+1} - 1 \\
&= a_1 - 1 + \sum_{m=1}^{\infty} (m+1) a_{m+1} x^m - \sum_{m=0}^{\infty} a_m x^{m+1} \\
&= a_1 - 1 + \sum_{m=0}^{\infty} (m+2) a_{m+2} x^{m+1} - \sum_{m=0}^{\infty} a_m x^{m+1} \\
&= a_1 - 1 + \sum_{m=0}^{\infty} \left[(m+2) a_{m+2} - a_m \right] x^{m+1} \\
&= 0 \quad \text{for all } x.
\end{aligned}$$

This requires that $a_1 - 1 = 0$ and

$$(m+2) a_{m+2} - a_m = 0 \quad \text{for all } m.$$

So me must have $a_1 = 1$ and $a_{m+2} = a_m/(m+2)$, that is,

$$a_2 = \frac{a_0}{2}, \quad a_4 = \frac{a_2}{4} = \frac{a_0}{2 \cdot 4}, \quad a_6 = \frac{a_4}{6} = \frac{a_0}{2 \cdot 4 \cdot 6} = \frac{a_0}{2^3 (1 \cdot 2 \cdot 3)}$$

$$a_8 = \frac{a_6}{8} = \frac{a_0}{2^4 (1 \cdot 2 \cdot 3 \cdot 4)} = \frac{a_0}{2^4 4!},$$

and so on, that is

$$a_{2k} = \frac{a_0}{2^k k!},$$

and

$$a_3 = \frac{a_1}{3} = \frac{1}{3}, \quad a_5 = \frac{a_3}{5} = \frac{1}{5 \cdot 5}, \quad a_7 = \frac{a_5}{7} = \frac{1}{1 \cdot 3 \cdot 5 \cdot 7}$$

and so on, that is

$$a_{2k+1} = \frac{1}{1 \cdot 3 \cdot 5 \cdot 7 \cdots (2k+1)} \frac{2 \cdot 4 \cdot 6 \cdot 8 \cdots (2k)}{2 \cdot 4 \cdot 6 \cdot 8 \cdots (2k)} = \frac{2^k k!}{(2k+1)!}.$$

The general solution is

$$\begin{aligned}
y_g(x) &= \sum_{m=0}^{\infty} a_m x^m \\
&= \sum_{k=0}^{\infty} a_{2k} x^{2k} + \sum_{k=0}^{\infty} a_{2k+1} x^{2k+1} \\
&= \sum_{k=0}^{\infty} \frac{a_0}{2^k k!} x^{2k} + \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!} \\
&= a_0 \sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} + \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}.
\end{aligned}$$

But $y(0) = 1$ means that $y(0) = a_0 = 1$, Thus, the unique solution is

$$\begin{aligned} y(x) &= \sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} + \sum_{k=0}^{\infty} \frac{x^{2k+1}}{1 \cdot 3 \cdot 5 \cdot 7 \cdots (2k+1)} \\ &= 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{48} + \cdots + x + \frac{x^3}{3} + \frac{x^5}{15} + \frac{x^7}{105} + \ldots, \end{aligned}$$

which coincides with the solutions of Examples 1.19 and 1.20. □

6.3. Legendre Equation and Legendre Polynomials

We look for the general solution of the *Legendre equation*:

$$(1 - x^2)y'' - 2xy' + n(n+1)y = 0, \qquad -1 < x < 1, \tag{6.4}$$

in the form of a power series with centre $a = 0$. We rewrite the equation in standard form $y'' + f(x)y' + g(x)y = r(x)$,

$$y'' - \frac{2x}{1 - x^2}y' + \frac{n(n+1)}{1 - x^2}y = 0.$$

Since the coefficients,

$$f(x) = \frac{2x}{(x-1)(x+1)}, \qquad g(x) = -\frac{n(n+1)}{(x-1)(x+1)},$$

have simple poles at $x = \pm 1$, they have convergent power series with centre $a = 0$ and radius of convergence $R = 1$:

$$\begin{aligned} f(x) &= -\frac{2x}{1 - x^2} = -2x[1 + x^2 + x^4 + x^6 + \ldots], \quad -1 < x < 1, \\ g(x) &= \frac{n(n+1)}{1 - x^2} = n(n+1)[1 + x^2 + x^4 + x^6 + \ldots], \quad -1 < x < 1. \end{aligned}$$

Moreover

$$r(x) = 0, \qquad -\infty < x < \infty.$$

Hence, we see that $f(x)$ and $g(x)$ are analytic for $-1 < x < 1$, and $r(x)$ is everywhere analytic.

By Theorem 6.3, we know that (6.4) has two linearly independent analytic solutions for $-1 < x < 1$.

Set

$$y(x) = \sum_{m=0}^{\infty} a_m x^m \tag{6.5}$$

and substitute in (6.4) to get

$$\begin{aligned}
&(1-x^2)y'' - 2xy' + n(n+1)y \\
&\quad = y'' - x^2y'' - 2xy' + n(n-1)y \\
&\quad = \sum_{m=0}^{\infty}(m)(m-1)a_m x^{m-2} - x^2\sum_{m=0}^{\infty} m(m-1)a_m x^{m-2} \\
&\qquad - 2x\sum_{m=0}^{\infty} m a_m x^{m-1} + n(n+1)\sum_{m=0}^{\infty} a_m x^m \\
&\quad = \sum_{m=0}^{\infty}(m+1)(m+2)a_{m+2}x^m - \sum_{m=0}^{\infty} m(m-1)a_m x^m \\
&\qquad - \sum_{m=0}^{\infty} 2m a_m x^m + n(n+1)\sum_{m=0}^{\infty} a_m x^m \\
&\quad = \sum_{m=0}^{\infty}\{(m+1)(m+2)a_{m+2} - [m(m-1)+2m-n(n-1)]a_m\}x^m \\
&\quad = \sum_{m=0}^{\infty}\{(m+1)(m+2)a_{m+2} - [m(m+1)-n(n-1)]a_m\}x^m \\
&\quad = 0, \quad \text{for all } x,
\end{aligned}$$

and so

$$a_{m+2} = \frac{m(m+1)-n(n+1)}{(m+1)(m+2)}\,a_m.$$

Therefore,

$$a_2 = -\frac{n(n+1)}{2!}\,a_0, \qquad a_3 = -\frac{(n-1)(n+2)}{3!}\,a_1, \tag{6.6}$$

$$a_4 = \frac{(n-2)n(n+1)(n+3)}{4!}\,a_0, \quad a_5 = \frac{(n-3)(n-1)(n+2)(n+4)}{5!}\,a_1, \tag{6.7}$$

etc. The solution can be written in the form

$$y(x) = a_0 y_1(x) + a_1 y_2(x), \tag{6.8}$$

where

$$\begin{aligned}
y_1(x) &= 1 - \frac{n(n+1)}{2!}\,x^2 + \frac{(n-2)n(n+1)(n+3)}{4!}\,x^4 - + \ldots, \\
y_2(x) &= x - \frac{(n-1)(n+2)}{3!}\,x^3 + \frac{(n-3)(n-1)(n+2)(n+4)}{5!}\,x^5 - + \ldots
\end{aligned}$$

Each series converges for $|x| < R = 1$. We remark that $y_1(x)$ is even and $y_2(x)$ is odd. Since

$$\frac{y_1(x)}{y_2(x)} \neq \text{ const},$$

it follows that $y_1(x)$ and $y_2(x)$ are two independent solutions and (6.8) is the general solution.

COROLLARY 6.1. *For n even, $y_1(x)$ is an even polynomial,*

$$y_1(x) = k_n P_n(x).$$

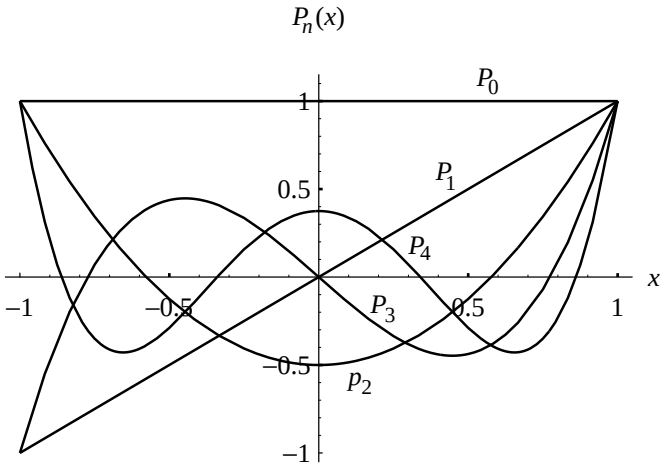

FIGURE 6.4. The first five Legendre polynomials.

Similarly, for n odd, $y_2(x)$ is an odd polynomial,

$$y_2(x) = k_n P_n(x),$$

The polynomial $P_n(x)$ is the Legendre polynomial of degree n, normalized such that $P_n(1) = 1$.

The first six Legendre polynomials are:

$$P_0(x) = 1, \qquad P_1(x) = x,$$
$$P_2(x) = \frac{1}{2}\left(3x^2 - 1\right), \qquad P_3(x) = \frac{1}{2}\left(5x^3 - 3x\right),$$
$$P_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right), \qquad P_5(x) = \frac{1}{8}\left(63x^5 - 70x^3 + 15x\right).$$

The graphs of the first five $P_n(x)$ are shown in Fig. 6.4.

We notice that the n zeros of the polynomial $P_n(x)$, of degree n, lie in the open interval $]-1, 1[$. These zeros are simple and interlace the $n-1$ zeros of $P_{n-1}(x)$, two properties that are ordinarily possessed by the zeros of orthogonal functions.

REMARK 6.1. It can be shown that the series for $y_1(x)$ and $y_2(x)$ diverge at $x = \pm 1$ if $n \neq 0, 2, 4, \ldots$, and $n \neq 1, 3, 5, \ldots$, respectively.

Symbolic Matlab can be used to obtain the Legendre polynomials if we use the condition $P_n(1) = 1$ as follows.

```
>> dsolve('(1-x^2)*D2y-2*x*Dy=0','y(1)=1','x')
y = 1

>> dsolve('(1-x^2)*D2y-2*x*Dy+2*y=0','y(1)=1','x')
y = x

>> dsolve('(1-x^2)*D2y-2*x*Dy+6*y=0','y(1)=1','x')
y = -1/2+3/2*x^2
```

and so on. With the Matlab extended symbolic toolbox, the Legendre polynomials $P_n(x)$ can be obtained from the full Maple kernel by using the command `orthopoly[P](n,x)`, which is referenced by the command `mhelp orthopoly[P]`.

6.4. Orthogonality Relations for $P_n(x)$

THEOREM 6.4. *The Legendre polynomials $P_n(x)$ satisfy the following orthogonality relation,*

$$\int_{-1}^{1} P_m(x)P_n(x)\,dx = \begin{cases} 0, & m \neq n, \\ \frac{2}{2n+1}, & m = n. \end{cases} \tag{6.9}$$

PROOF. We give below two proofs of the second part ($m = n$) of the orthogonality relation. The first part ($m \neq n$) follows simply from the Legendre equation

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0,$$

rewritten in divergence form,

$$L_n y := [(1-x^2)y']' + n(n+1)y = 0.$$

Since $P_m(x)$ and $P_n(x)$ are solutions of $L_m y = 0$ and $L_n y = 0$, respectively, we have

$$P_n(x)L_m(P_m) = 0, \qquad P_m(x)L_n(P_n) = 0.$$

Integrating these two expressions from -1 to 1, we have

$$\int_{-1}^{1} P_n(x)[(1-x^2)P_m'(x)]'\,dx + m(m+1)\int_{-1}^{1} P_n(x)P_m(x)\,dx = 0,$$
$$\int_{-1}^{1} P_m(x)[(1-x^2)P_n'(x)]'\,dx + n(n+1)\int_{-1}^{1} P_m(x)P_n(x)\,dx = 0.$$

Now integrating by parts the first term of these expressions, we have

$$P_n(x)(1-x^2)P_m'(x)\Big|_{-1}^{1} - \int_{-1}^{1} P_n'(x)(1-x^2)P_m'(x)\,dx + m(m+1)\int_{-1}^{1} P_n(x)P_m(x)\,dx = 0,$$
$$P_m(x)(1-x^2)P_n'(x)\Big|_{-1}^{1} - \int_{-1}^{1} P_m'(x)(1-x^2)P_n'(x)\,dx + n(n+1)\int_{-1}^{1} P_m(x)P_n(x)\,dx = 0.$$

The integrated terms are zero and the next term is the same in both equations. Hence, subtracting these equations, we obtain the orthogonality relation

$$[m(m+1) - n(n+1)]\int_{-1}^{1} P_m(x)P_n(x)\,dx = 0$$
$$\Longrightarrow \int_{-1}^{1} P_m(x)P_n(x)\,dx = 0 \qquad \text{for} \quad m \neq n.$$

The second part ($m = n$) follows from *Rodrigues' formula*:

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left[\left(x^2-1\right)^n\right]. \tag{6.10}$$

In fact,

$$\begin{aligned}
\int_{-1}^{1} P_n^2(x)\,dx &= \frac{1}{2^n}\times\frac{1}{n!}\times\frac{1}{2^n}\times\frac{1}{n!}\int_{-1}^{1}\left[\frac{d^n}{dx^n}\left(x^2-1\right)^n\right]\left[\frac{d^n}{dx^n}\left(x^2-1\right)^n\right]dx \\
&\qquad \{\text{and integrating by parts } n \text{ times}\} \\
&= \frac{1}{2^n}\times\frac{1}{n!}\times\frac{1}{2^n}\times\frac{1}{n!}\left[\frac{d^{n-1}}{dx^{n-1}}\left(x^2-1\right)^n\frac{d^n}{dx^n}\left(x^2-1\right)\Big|_{-1}^{1}\right. \\
&\quad \left.+(-1)^1\int_{-1}^{1}\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n\frac{d^{n+1}}{dx^{n+1}}\left(x^2-1\right)^n\,dx\right] \\
&\quad +\ldots \\
&= \frac{1}{2^n}\times\frac{1}{n!}\times\frac{1}{2^n}\times\frac{1}{n!}(-1)^n\int_{-1}^{1}\left(x^2-1\right)^n\frac{d^{2n}}{dx^{2n}}\left(x^2-1\right)^n\,dx \\
&\qquad \{\text{and differentiating } 2n \text{ times}\} \\
&= \frac{1}{2^n}\times\frac{1}{n!}\times\frac{1}{2^n}\times\frac{1}{n!}(-1)^n(2n)!\int_{-1}^{1}1\times\left(x^2-1\right)^n\,dx \\
&\qquad \{\text{and integrating by parts } n \text{ times}\} \\
&= \frac{(-1)^n(2n)!}{2^n n!2^n n!}\left[\frac{x}{1}\left(x^2-1\right)^n\Big|_{-1}^{1}+\frac{(-1)^1}{1!}2n\int_{-1}^{1}x^2\left(x^2-1\right)^{n-1}\,dx\right] \\
&\quad +\ldots \\
&= \frac{(-1)^n(2n)!}{2^n n!2^n n!}(-1)^n\frac{2n2(n-1)2(n-2)\cdots 2(n-(n-1))}{1\times 3\times 5\times\cdots\times(2n-1)}\int_{-1}^{1}x^{2n}\,dx \\
&= \frac{(-1)^n(-1)^n(2n)!}{2^n n!2^n n!}\frac{2^n n!}{1\times 3\times 5\times\cdots\times(2n-1)}\frac{1}{(2n+1)}x^{2n+1}\Big|_{-1}^{1} \\
&= \frac{2}{2n+1}. \quad \square
\end{aligned}$$

REMARK 6.2. Rodrigues' formula can be obtained by direct computation with $n = 0, 1, 2, 3, \ldots$, or otherwise. We compute $P_4(x)$ using Rodrigues' formula with the symbolic Matlab command `diff`.

```
>> syms x f p4
>> f = (x^2-1)^4
 f = (x^2-1)^4
>> p4 = (1/(2^4*prod(1:4)))*diff(f,x,4)
 p4 = x^4+3*(x^2-1)*x^2+3/8*(x^2-1)^2
>> p4 = expand(p4)
 p4 = 3/8-15/4*x^2+35/8*x^4
```

We finally present a second proof of the formula for the norm of P_n,

$$\|P_n\|^2 := \int_{-1}^{1}[P_n(x)]^2\,dx = \frac{2}{2n+1},$$

by means of the *generating function* for $P_n(x)$,

$$\sum_{k=0}^{\infty} P_k(x)t^k = \frac{1}{\sqrt{1-2xt+t^2}}. \tag{6.11}$$

PROOF. Squaring both sides of (6.11),

$$\sum_{k=0}^{\infty} P_k^2(x)t^{2k} + \sum_{j\neq k} P_j(x)P_k(x)t^{j+k} = \frac{1}{1-2xt+t^2},$$

and integrating with respect to x from -1 to 1, we have

$$\sum_{k=0}^{\infty}\left[\int_{-1}^{1} P_k^2(x)\,dx\right]t^{2k} + \sum_{j\neq k}\left[\int_{-1}^{1} P_j(x)P_k(x)\,dx\right]t^{j+k} = \int_{-1}^{1}\frac{dx}{1-2xt+t^2}.$$

Since $P_j(x)$ and $P_k(x)$ are orthogonal for $j \neq k$, the second term on the left-hand side is zero. Hence, after integration of the right-hand side, we obtain

$$\begin{aligned}\sum_{k=0}^{\infty}\|P_k\|^2t^{2k} &= -\frac{1}{2t}\ln\bigl(1-2xt+t^2\bigr)\Big|_{x=-1}^{x=1}\\ &= -\frac{1}{t}\bigl[\ln(1-t)-\ln(1+t)\bigr].\end{aligned}$$

Multiplying by t,

$$\sum_{k=0}^{\infty}\|P_k\|^2t^{2k+1} = -\ln(1-t)+\ln(1+t)$$

and differentiating with respect to t, we have

$$\begin{aligned}\sum_{k=0}^{\infty}(2k+1)\|P_k\|^2t^{2k} &= \frac{1}{1-t}+\frac{1}{1+t}\\ &= \frac{2}{1-t^2}\\ &= 2\bigl(1+t^2+t^4+t^6+\dots\bigr)\qquad \text{for all } t, |t|<1.\end{aligned}$$

Since we have an identity in t, we can identify the coefficients of t^{2k} on both sides,

$$(2k+1)\|P_k\|^2 = 2 \implies \|P_k\|^2 = \frac{2}{2k+1}. \quad \square$$

REMARK 6.3. The generating function (6.11) can be obtained by expanding its right-hand side in a Taylor series in t, as is easily done with symbolic Matlab by means of the command `taylor`.

```
>> syms t x; f = 1/(1-2*x*t+t^2)^(1/2);
>> g = taylor(f,3,t)
 g = 1+t*x+(-1/2+3/2*x^2)*t^2
        +(-3/2*x+5/2*x^3)*t^3+(3/8-15/4*x^2+35/8*x^4)*t^4
```

6.5. Fourier–Legendre Series

The Fourier–Legendre series of $f(x)$ on $-1 < x < 1$ is

$$f(x) = \sum_{m=0}^{\infty} a_m P_m(x), \quad -1 < x < 1,$$

where the coefficients are given by

$$a_m = \frac{2m+1}{2} \int_{-1}^{1} f(x) P_m(x)\, dx.$$

The value of a_m follows from the orthogonality relations (6.9) of the $P_m(x)$ on $-1 < x < 1$:

$$\begin{aligned} \int_{-1}^{1} f(x) P_n(x)\, dx &= \sum_{m=0}^{\infty} a_m \int_{-1}^{1} P_m(x) P_n(x)\, dx \\ &= a_m \int_{-1}^{1} P_m(x) P_m(x)\, dx = \frac{2}{2m+1}\, a_m. \end{aligned}$$

For polynomials $p(x)$ of degree k, we obtain a finite expansion

$$f(x) = \sum_{m=0}^{k} a_m P_m(x)$$

on $-\infty < x < \infty$, without integration, by the simple change of bases from x^m to $P_m(x)$ for $m = 1, 2, \ldots, k$.

Example 6.5. Expand the polynomial

$$p(x) = x^3 - 2x^2 + 4x + 1$$

over $[-1, 1]$ in terms of the Legendre polynomials $P_0(x), P_1(x), \ldots$

Solution. We express the powers of x in terms of the basis of Legendre polynomials:

$$\begin{aligned} P_0(x) = 1 &\implies 1 = P_0(x), \\ P_1(x) = x &\implies x = P_1(x), \\ P_2(x) = \frac{1}{2}(3x^2 - 1) &\implies x^2 = \frac{2}{3} P_2(x) + \frac{1}{3} P_0(x), \\ P_3(x) = \frac{1}{2}(5x^3 - 3x) &\implies x^3 = \frac{2}{5} P_3(x) + \frac{3}{5} P_1(x). \end{aligned}$$

This way, one avoids computing integrals. Thus

$$\begin{aligned} p(x) &= \frac{2}{5} P_3(x) + \frac{3}{5} P_1(x) - \frac{4}{3} P_2(x) - \frac{2}{3} P_0(x) + 4P_1(x) + P_0(x) \\ &= \frac{2}{5} P_3(x) - \frac{4}{3} P_2(x) + \frac{23}{5} P_1(x) + \frac{1}{3} P_0(x). \quad \square \end{aligned}$$

.

Example 6.6. Expand the polynomial

$$p(x) = 2 + 3x + 5x^2$$

over $[3, 7]$ in terms of the Legendre polynomials $P_0(x), P_1(x), \ldots$

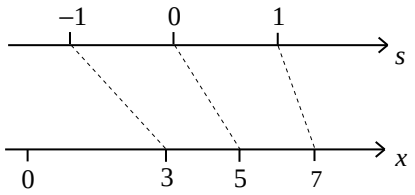

FIGURE 6.5. Affine mapping of $x \in [3,7]$ onto $s \in [-1,1]$.

SOLUTION. To map the segment $x \in [3,7]$ onto the segment $s \in [-1,1]$ (see Fig. 6.5) we consider the affine transformation

$$s \mapsto x = \alpha s + \beta, \qquad \text{such that} \quad -1 \mapsto 3 = -\alpha + \beta, \quad 1 \mapsto 7 = \alpha + \beta.$$

Solving for α and β, we have

$$x = 2s + 5. \tag{6.12}$$

Then

$$\begin{aligned} p(x) &= p(2s+5) \\ &= 2 + 3(2s+5) + 5(2s+5)^2 \\ &= 142 + 106s + 20s^2 \\ &= 142P_0(s) + 106P_1(s) + 20\left[\frac{2}{3}P_2(s) + \frac{1}{3}P_0(s)\right]; \end{aligned}$$

consequently, we have

$$p(x) = \left(142 + \frac{20}{3}\right) P_0\left(\frac{x-5}{2}\right) + 106P_1\left(\frac{x-5}{2}\right) + \frac{40}{3}P_2\left(\frac{x-5}{2}\right). \quad \square$$

.

EXAMPLE 6.7. Compute the first three terms of the Fourier–Legendre expansion of the function

$$f(x) = \begin{cases} 0, & -1 < x < 0, \\ x, & 0 < x < 1. \end{cases}$$

SOLUTION. Putting

$$f(x) = \sum_{m=0}^{\infty} a_m P_m(x), \quad -1 < x < 1,$$

we have

$$a_m = \frac{2m+1}{2} \int_{-1}^{1} f(x) P_m(x)\, dx.$$

Hence

$$\begin{aligned} a_0 &= \frac{1}{2}\int_{-1}^{1} f(x)P_0(x)\,dx = \frac{1}{2}\int_0^1 x\,dx = \frac{1}{4}, \\ a_1 &= \frac{3}{2}\int_{-1}^{1} f(x)P_1(x)\,dx = \frac{3}{2}\int_0^1 x^2\,dx = \frac{1}{2}, \\ a_2 &= \frac{5}{2}\int_{-1}^{1} f(x)P_2(x)\,dx = \frac{5}{2}\int_0^1 x\,\frac{1}{2}\,(3x^2-1)\,dx = \frac{5}{16}. \end{aligned}$$

Thus we have the approximation

$$f(x) \approx \frac{1}{4}\,P_0(x) + \frac{1}{2}\,P_1(x) + \frac{5}{16}\,P_2(x). \quad \square$$

EXAMPLE 6.8. Compute the first three terms of the Fourier–Legendre expansion of the function

$$f(x) = e^x, \qquad 0 \le x \le 1.$$

SOLUTION. To use the orthogonality of the Legendre polynomials, we transform the domain of $f(x)$ from $[0,1]$ to $[-1,1]$ by the substitution

$$s = 2\left(x - \frac{1}{2}\right), \quad \text{that is} \quad x = \frac{s}{2} + \frac{1}{2}.$$

Then

$$f(x) = e^x = e^{(1+s)/2} = \sum_{m=0}^{\infty} a_m P_m(s), \quad -1 \le s \le 1,$$

where

$$a_m = \frac{2m+1}{2} \int_{-1}^{1} e^{(1+s)/2} P_m(s)\, ds.$$

We first compute the following three integrals by recurrence:

$$\begin{aligned}
I_0 &= \int_{-1}^{1} e^{s/2}\, ds = 2\left(e^{1/2} - e^{-1/2}\right), \\
I_1 &= \int_{-1}^{1} s\, e^{s/2}\, ds = 2s\, e^{s/2}\Big|_{-1}^{1} - 2\int_{-1}^{1} e^{s/2}\, ds \\
&= 2\left(e^{1/2} + e^{-1/2}\right) - 2I_0 \\
&= -2\, e^{1/2} + 6\, e^{-1/2}, \\
I_2 &= \int_{-1}^{1} s^2\, e^{s/2}\, ds = 2s^2\, e^{s/2}\Big|_{-1}^{1} - 4\int_{-1}^{1} s\, e^{s/2}\, ds \\
&= 2\left(e^{1/2} - e^{-1/2}\right) - 4I_1 \\
&= 10\, e^{1/2} - 26\, e^{-1/2}.
\end{aligned}$$

Thus

$$\begin{aligned}
a_0 &= \frac{1}{2}\, e^{1/2} I_0 = e - 1 \approx 1.7183, \\
a_1 &= \frac{3}{2}\, e^{1/2} I_1 = -3\, e + 9 \approx 0.8452, \\
a_2 &= \frac{5}{2}\, e^{1/2} \frac{1}{2}(3I_2 - I_0) = 35\, e - 95 \approx 0.1399.
\end{aligned}$$

We finally have the approximation

$$f(x) \approx 1.7183 P_0(2x-1) + 0.8452 P_1(2x-1) + 0.1399 P_2(2x-1). \quad \square$$

6.6. Derivation of Gaussian Quadratures

We easily obtain the n-point Gaussian Quadrature formula by means of the Legendre polynomials. We restrict ourselves to the cases $n = 2$ and $n = 3$. We immediately remark that the number of points n refers to the n points at which we need to evaluate the integrand over the interval $[-1, 1]$, and not to the numbers of subintervals into which one usually breaks the whole interval of integration $[a, b]$ in order to have a smaller error in the numerical value of the integral.

EXAMPLE 6.9. Determine the four parameters of the two-point Gaussian Quadrature formula,

$$\int_{-1}^{1} f(x)\,dx = af(x_1) + bf(x_2).$$

SOLUTION. By symmetry, it is expected that the nodes will be negative to each other, $x_1 = -x_2$, and the weights will be equal, $a = b$. Since there are four free parameters, the formula will be exact for polynomials of degree three or less. By Example 6.5, it suffices to consider the polynomials $P_0(x), \ldots, P_3(x)$. Since $P_0(x) = 1$ is orthogonal to $P_n(x)$, $n = 1, 2, \ldots$, we have

$$2 = \int_{-1}^{1} P_0(x)\,dx = aP_0(x_1) + bP_0(x_2) = a + b, \tag{6.13}$$

$$0 = \int_{-1}^{1} 1 \times P_1(x)\,dx = aP_1(x_1) + bP_1(x_2) = ax_1 + bx_2, \tag{6.14}$$

$$0 = \int_{-1}^{1} 1 \times P_2(x)\,dx = aP_2(x_1) + bP_2(x_2), \tag{6.15}$$

$$0 = \int_{-1}^{1} 1 \times P_3(x)\,dx = aP_3(x_1) + bP_3(x_2), \tag{6.16}$$

To satisfy (6.15) we choose x_1 and x_2 such that

$$P_2(x_1) = P_2(x_2) = 0,$$

that is,

$$P_2(x) = \frac{1}{2}(3x^2 - 1) = 0 \Rightarrow -x_1 = x_2 = \frac{1}{\sqrt{3}} = 0.577\,350\,27.$$

Hence, by (6.14), we have

$$a = b.$$

Moreover, (6.16) is automatically satisfied since $P_3(x)$ is odd. Finally, by (6.13), we have

$$a = b = 1.$$

Thus, the two-point Gaussian Quadrature formula is

$$\int_{-1}^{1} f(x)\,dx = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right). \quad \square \tag{6.17}$$

EXAMPLE 6.10. Determine the six parameters of the three-point Gaussian Quadrature formula,

$$\int_{-1}^{1} f(x)\,dx = af(x_1) + bf(x_2) + cf(x_3).$$

SOLUTION. By symmetry, it is expected that the two extremal nodes are negative to each other, $x_1 = -x_3$, and the middle node is at the origin, $x_2 = 0$, Moreover, the extremal weights should be equal, $a = c$, and the central one be larger that the other two, $b > a = c$. Since there are six free parameters, the formula will be exact for polynomials of degree five or less. By Example 6.5, it suffices to consider the basis $P_0(x), \ldots, P_5(x)$. Thus,

$$2 = \int_{-1}^{1} P_0(x)\,dx = aP_0(x_1) + bP_0(x_2) + cP_0(x_3), \tag{6.18}$$

$$0 = \int_{-1}^{1} P_1(x)\,dx = aP_1(x_1) + bP_1(x_2) + cP_1(x_3), \tag{6.19}$$

$$0 = \int_{-1}^{1} P_2(x)\,dx = aP_2(x_1) + bP_2(x_2) + cP_2(x_3), \tag{6.20}$$

$$0 = \int_{-1}^{1} P_3(x)\,dx = aP_3(x_1) + bP_3(x_2) + cP_3(x_3), \tag{6.21}$$

$$0 = \int_{-1}^{1} P_4(x)\,dx = aP_4(x_1) + bP_4(x_2) + cP_4(x_3), \tag{6.22}$$

$$0 = \int_{-1}^{1} P_5(x)\,dx = aP_5(x_1) + bP_5(x_2) + cP_5(x_3). \tag{6.23}$$

To satisfy (6.21), we let x_1, x_2, x_3 be the three zeros of

$$P_3(x) = \frac{1}{2}(5x^3 - 3x) = \frac{1}{2}x(5x^2 - 3)$$

that is,

$$-x_1 = x_3 = \sqrt{\frac{3}{5}} = 0.774\,596\,7, \qquad x_2 = 0.$$

Hence (6.19) implies

$$-\sqrt{\frac{3}{5}}\,a + \sqrt{\frac{3}{5}}\,c = 0 \Rightarrow a = c.$$

We immediately see that (6.23) is satisfied since $P_5(x)$ is odd. Moreover, by substituting $a = c$ in (6.20), we have

$$a\,\frac{1}{2}\left(3 \times \frac{3}{5} - 1\right) + b\left(-\frac{1}{2}\right) + a\,\frac{1}{2}\left(3 \times \frac{3}{5} - 1\right) = 0,$$

that is,

$$4a - 5b + 4a = 0 \quad \text{or} \quad 8a - 5b = 0. \tag{6.24}$$

Now, it follows from (6.18) that

$$2a + b = 2 \quad \text{or} \quad 10a + 5b = 10. \tag{6.25}$$

Adding the second expressions in (6.24) and (6.25), we have

$$a = \frac{10}{18} = \frac{5}{9} = 0.555.$$

Thus

$$b = 2 - \frac{10}{9} = \frac{8}{9} = 0.888.$$

Finally, we verify that (6.22) is satisfied. Since

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3),$$

we have

$$\begin{aligned} 2 \times \frac{5 \times 1}{9 \times 8}\left(35 \times \frac{9}{25} - 30 \times \frac{3}{5} + 3\right) + \frac{8}{9} \times \frac{3}{8} &= \frac{2 \times 5}{9 \times 8}\left(\frac{315 - 450 + 75}{25}\right) + \frac{8}{9} \times \frac{3}{8} \\ &= \frac{2 \times 5}{9 \times 8} \times \frac{(-60)}{25} + \frac{8 \times 3}{9 \times 8} \\ &= \frac{-24 + 24}{9 \times 8} = 0. \end{aligned}$$

Therefore, the three-point Gaussian Quadrature formula is

$$\int_{-1}^{1} f(x)\,dx = \frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right). \quad \square \tag{6.26}$$

REMARK 6.4. The interval of integration in the Gaussian Quadrature formulas is normalized to $[-1, 1]$. To integrate over the interval $[a, b]$ we use the change of independent variable from $x \in [a, b]$ to $t \in [-1, 1]$ (see Example 6.8):

$$t \mapsto x = \alpha t + \beta, \qquad \text{such that} \quad -1 \mapsto a = -\alpha + \beta, \quad 1 \mapsto b = \alpha + \beta.$$

Solving for α and β, we have

$$x = \frac{(b-a)t + b + a}{2}, \qquad dx = \left(\frac{b-a}{2}\right) dt.$$

Thus, the integral becomes

$$\int_a^b f(x)\,dx = \frac{b-a}{2} \int_{-1}^{1} f\left(\frac{(b-a)t + b + a}{2}\right) dt.$$

EXAMPLE 6.11. Evaluate

$$I = \int_0^{\pi/2} \sin x\,dx$$

by applying the two-point Gaussian Quadrature formula once over the interval $[0, \pi/2]$ and over the half-intervals $[0, \pi/4]$ and $[\pi/4, \pi/2]$.

SOLUTION. Let

$$x = \frac{(\pi/2)t + \pi/2}{2}, \qquad dx = \frac{\pi}{4}\,dt.$$

At $t = -1$, $x = 0$ and, at $t = 1$, $x = \pi/2$. Hence

$$\begin{aligned} I &= \frac{\pi}{4} \int_{-1}^{1} \sin\left(\frac{\pi t + \pi}{4}\right) dt \\ &\approx \frac{\pi}{4}\left[1.0 \times \sin\left(0.105\,66\pi\right) + 1.0 \times \sin\left(0.394\,34\pi\right)\right] \\ &= 0.998\,47. \end{aligned}$$

The error is 1.53×10^{-3}. Over the half-intervals, we have

$$\begin{aligned} I &= \frac{\pi}{8}\int_{-1}^{1} \sin\left(\frac{\pi t+\pi}{8}\right) dt + \frac{\pi}{8}\int_{-1}^{1} \sin\left(\frac{\pi t+3\pi}{8}\right) dt \\ &\approx \frac{\pi}{8}\left[\sin\frac{\pi}{8}\left(-\frac{1}{\sqrt{3}}+1\right) + \sin\frac{\pi}{8}\left(\frac{1}{\sqrt{3}}+1\right)\right. \\ &\qquad \left. + \sin\frac{\pi}{8}\left(-\frac{1}{\sqrt{3}}+3\right) + \sin\frac{\pi}{8}\left(\frac{1}{\sqrt{3}}+3\right)\right] \\ &= 0.999\,910\,166\,769\,89. \end{aligned}$$

The error is 8.983×10^{-5}. The Matlab solution is as follows. For generality, it is convenient to set up a function M-file `exp5_10.m`,

```
function f=exp5_10(t)
f=sin(t); % evaluate the function f(t)
```

The two-point Gaussian Quadrature is programmed as follows.

```
>> clear
>> a = 0; b = pi/2; c = (b-a)/2; d= (a+b)/2;
>> weight = [1 1]; node = [-1/sqrt(3) 1/sqrt(3)];
>> syms x t
>> x = c*node+d;
>> nv1 = c*weight*exp5_10(x)' % numerical value of integral
 nv1 = 0.9985
>> error1 = 1 - nv1 % error in solution
 error1 = 0.0015
```

The other part is done in a similar way. □

We evaluate the integral of Example 6.11 by Matlab's adapted Simpson's rule (`quad`) and adapted 8-panel Newton–Cotes' method (`quad8`).

```
>> v1 = quad('sin',0,pi/2)
 v1 = 1.00000829552397
>> v2 = quad8('sin',0,pi/2)
 v2 = 1.00000000000000
```

respectively, within a relative error of 10^{-3}.

REMARK 6.5. The Gaussian Quadrature formulae are the most accurate integration formulae for a given number of nodes. The error in the n-point formula is

$$E_n(f) = \frac{2}{(2n+1)!}\left[\frac{2^n(n!)^2}{(2n)!}\right]^2 f^{(2n)}(\xi), \qquad -1 < \xi < 1.$$

This formula is therefore exact for polynomials of degree $2n-1$ or less.

The nodes of the four- and five-point Gaussian Quadratures can be expressed in terms of radicals. See Exercises 6.35 and 6.37.

Part 2

Numerical Methods

CHAPTER 7

Solutions of Nonlinear Equations

7.1. Computer Arithmetic

7.1.1. Definitions. The following notation and terminology will be used.

(1) If a is the exact value of a computation and $\tilde{a}$ is an approximate value for the same computation, then

$$\epsilon = \tilde{a} - a$$

is the *error* in $\tilde{a}$ and $|\epsilon|$ is the *absolute error*. If $a \neq 0$,

$$\epsilon_r = \frac{\tilde{a} - a}{a} = \frac{\epsilon}{a}$$

is the *relative error* in $\tilde{a}$.

(2) *Upper bounds* for the absolute and relative errors in $\tilde{a}$ are numbers B_a and B_r such that

$$|\epsilon| = |\tilde{a} - a| < B_a, \qquad |\epsilon_r| = \left|\frac{\tilde{a} - a}{a}\right| < B_r,$$

respectively.

(3) A *roundoff error* occurs when a computer approximates a real number by a number with only a finite number of digits to the right of the decimal point (see Subsection 7.1.2).

(4) In scientific computation, the *floating point representation* of a number c of length d in the base β is

$$c = \pm 0.b_1 b_2 \cdots b_d \times \beta^N,$$

where $b_1 \neq 0$, $0 \leq b_i < \beta$. We call $b_1 b_2 \cdots b_d$ the *mantissa* or *decimal part* and N the *exponent* of c. For instance, with $d = 5$ and $\beta = 10$,

$$0.27120 \times 10^2, \qquad -0.31224 \times 10^3.$$

(5) The number of *significant digits* of a floating point number is the number of digits counted from the first to the last nonzero digits. For example, with $d = 4$ and $\beta = 10$, the number of significant digits of the three numbers:

$$0.1203 \times 10^2, \qquad 0.1230 \times 10^{-2}, \qquad 0.1000 \times 10^3,$$

is 4, 3, and 1, respectively.

(6) The term *truncation error* is used for the error committed when an infinite series is truncated after a finite number of terms.

REMARK 7.1. For simplicity, we shall often write floating point numbers without exponent and with zeros immediately to the right of the decimal point or with nonzero numbers to the left of the decimal point:

$$0.001203, \quad 12300.04$$

7.1.2. Rounding and chopping numbers. Real numbers are rounded away from the origin. The floating-point number, say in base 10,

$$c = \pm 0.b_1 b_2 \dots b_d \times 10^N$$

is rounded to k digits as follows:

(i) If $0.b_{k+1}b_{k+2}\dots b_m \geq 0.5$, round c to

$$(0.b_1 b_2 \dots b_{k-1}b_k + 0.1 \times 10^{-k+1}) \times 10^N.$$

(ii) If $0.b_{k+1}b_{k+2}\dots b_m < 0.5$, round c to

$$0.b_1 b_2 \dots b_{k-1}b_k \times 10^N.$$

EXAMPLE 7.1. Numbers rounded to three digits:

$$1.9234542 \approx 1.92$$
$$2.5952100 \approx 2.60$$
$$1.9950000 \approx 2.00$$
$$-4.9850000 \approx -4.99$$

Floating-point numbers are chopped to k digits by replacing the digits to the right of the kth digit by zeros.

7.1.3. Cancellation in computations. Cancellation due to the subtraction of two almost equal numbers leads to a loss of significant digits. It is better to avoid cancellation than to try to estimate the error due to cancellation. Example 7.2 illustrates these points.

EXAMPLE 7.2. Use 10-digit rounded arithmetic to solve the quadratic equation

$$x^2 - 1634x + 2 = 0.$$

SOLUTION. The usual formula yields

$$x = 817 \pm \sqrt{2\,669\,948}.$$

Thus,

$$x_1 = 817 + 816.998\,776\,0 = 1.633\,998\,776 \times 10^3,$$
$$x_2 = 817 - 816.998\,776\,0 = 1.224\,000\,000 \times 10^{-3}.$$

Four of the six zeros at the end of the fractional part of x_2 are the result of cancellation and thus are meaningless. A more accurate result for x_2 can be obtained if we use the relation

$$x_1 x_2 = 2.$$

In this case

$$x_2 = 1.223\,991\,125 \times 10^{-3},$$

where all digits are significant. □

From Example 7.2, it is seen that a numerically stable formula for solving the quadratic equation

$$ax^2 + bx + c = 0, \qquad a \neq 0,$$

is

$$x_1 = \frac{1}{2a}\left[-b - \text{sign}\,(b)\sqrt{b^2 - 4ac}\right], \qquad x_2 = \frac{c}{ax_1},$$

where the signum function is

$$\text{sign}\,(x) = \begin{cases} +1, & \text{if} \quad x \geq 0, \\ -1, & \text{if} \quad x < 0. \end{cases}$$

EXAMPLE 7.3. If the value of x rounded to three digits is 4.81 and the value of y rounded to five digits is 12.752, find the smallest interval which contains the exact value of $x - y$.

SOLUTION. Since

$$4.805 \leq x < 4.815 \quad \text{and} \quad 12.7515 \leq y < 12.7525,$$

then

$$4.805 - 12.7525 < x - y < 4.815 - 12.7515 \Leftrightarrow -7.9475 < x - y < -7.9365. \quad \square$$

EXAMPLE 7.4. Find the error and the relative error in the commonly used rational approximations 22/7 and 355/113 to the transcendental number π and express your answer in three-digit floating point numbers.

SOLUTION. The error and the relative error in 22/7 are

$$\epsilon = 22/7 - \pi, \qquad \epsilon_r = \epsilon/\pi,$$

which Matlab evaluates as

```
pp = pi
   pp = 3.14159265358979
r1 = 22/7.
   r1 = 3.14285714285714
abserr1 = r1 -pi
   abserr1 =  0.00126448926735
relerr1 = abserr1/pi
   relerr1 =   4.024994347707008e-04
```

Hence, the error and the relative error in 22/7 rounded to three digits are

$$\epsilon = 0.126 \times 10^{-2} \quad \text{and} \quad \epsilon_r = 0.402 \times 10^{-3},$$

respectively. Similarly, Matlab computes the error and relative error in 355/113 as

```
r2 = 355/113.
   r2 = 3.14159292035398
abserr2 = r2 - pi
   abserr2 = 2.667641894049666e-07
relerr2 = abserr2/pi
   relerr2 = 8.491367876740610e-08
```

Hence, the error and the relative error in 355/113 rounded to three digits are

$$\epsilon = 0.267 \times 10^{-6} \quad \text{and} \quad \epsilon_r = 0.849 \times 10^{-7}. \quad \square$$

7.2. Review of Calculus

The following results from elementary calculus are needed to justify the methods of solution presented here.

THEOREM 7.1 (Intermediate Value Theorem). *Let $a < b$ and $f(x)$ be a continuous function on $[a, b]$. If w is a number strictly between $f(a)$ and $f(b)$, then there exists a number c such that $a < c < b$ and $f(c) = w$.*

COROLLARY 7.1. *Let $a < b$ and $f(x)$ be a continuous function on $[a, b]$. If $f(a)f(b) < 0$, then there exists a zero of $f(x)$ in the open interval $]a, b[$.*

PROOF. Since $f(a)$ and $f(b)$ have opposite signs, 0 lies between $f(a)$ and $f(b)$. The result follows from the intermediate value theorem with $w = 0$. $\square$

THEOREM 7.2 (Extreme Value Theorem). *Let $a < b$ and $f(x)$ be a continuous function on $[a, b]$. Then there exist numbers $\alpha \in [a, b]$ and $\beta \in [a, b]$ such that, for all $x \in [a, b]$, we have*

$$f(\alpha) \le f(x) \le f(\beta).$$

THEOREM 7.3 (Mean Value Theorem). *Let $a < b$ and $f(x)$ be a continuous function on $[a, b]$ which is differentiable on $]a, b[$. Then there exists a number c such that $a < c < b$ and*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

THEOREM 7.4 (Mean Value Theorem for Integrals). *Let $a < b$ and $f(x)$ be a continuous function on $[a, b]$. If $g(x)$ is an integrable function on $[a, b]$ which does not change sign on $[a, b]$, then there exists a number c such that $a < c < b$ and*

$$\int_a^b f(x)\, g(x)\, dx = f(c) \int_a^b g(x)\, dx.$$

A similar theorem holds for sums.

THEOREM 7.5 (Mean Value Theorem for Sums). *Let $\{w_i\}$, $i = 1, 2, \ldots, n$, be a set of n distinct real numbers and let $f(x)$ be a continuous function on an interval $[a, b]$. If the numbers w_i all have the same sign and all the points $x_i \in [a, b]$, then there exists a number $c \in [a, b]$ such that*

$$\sum_{i=1}^{n} w_i f(x_i) = f(c) \sum_{i=1}^{n} w_i.$$

7.3. The Bisection Method

The bisection method constructs a sequence of intervals of decreasing length which contain a root p of $f(x) = 0$. If

$$f(a)\, f(b) < 0 \quad \text{and} \quad f \quad \text{is continuous on } [a, b],$$

then, by Corollary 7.1, $f(x) = 0$ has a root between a and b. The root is either between

$$a \quad \text{and} \quad \frac{a+b}{2}, \quad \text{if} \quad f(a)\, f\left(\frac{a+b}{2}\right) < 0,$$

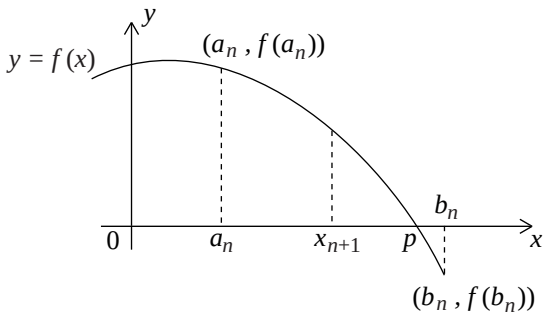

FIGURE 7.1. The nth step of the bisection method.

or between

$$\frac{a+b}{2} \quad \text{and} \quad b, \quad \text{if} \quad f\left(\frac{a+b}{2}\right) f(b) < 0,$$

or exactly at

$$\frac{a+b}{2}, \quad \text{if} \quad f\left(\frac{a+b}{2}\right) = 0.$$

The nth step of the bisection method is shown in Fig. 7.1.

The algorithm of the *bisection method* is as follows.

ALGORITHM 7.1 (Bisection Method). Given that $f(x)$ is continuous on $[a, b]$ and $f(a)\, f(b) < 0$:

(1) Choose $a_0 = a$, $b_0 = b$; tolerance TOL; maximum number of iteration N_0.
(2) For $n = 0, 1, 2, \ldots, N_0$, compute
$$x_{n+1} = \frac{a_n + b_n}{2}.$$
(3) If $f(x_{n+1}) = 0$ or $(b_n - a_n)/2 < TOL$, then output p $(= x_{n+1})$ and stop.
(4) Else if $f(x_{n+1})$ and $f(a_n)$ have opposite signs, set $a_{n+1} = a_n$ and $b_{n+1} = x_{n+1}$.
(5) Else set $a_{n+1} = x_{n+1}$ and $b_{n+1} = b_n$.
(6) Repeat (2), (3), (4) and (5).
(7) Ouput 'Method failed after N_0 iterations' and stop.

Other stopping criteria are described in Subsection 7.4.1. The rate of convergence of the bisection method is low but the method always converges.

The bisection method is programmed in the following Matlab function M-file which is found in `ftp://ftp.cs.cornell.edu/pub/cv`.

```
  function root = Bisection(fname,a,b,delta)
%
% Pre:
%    fname   string that names a continuous function  f(x) of
%            a single variable.
%
%      a,b   define an interval [a,b]
%            f is continuous, f(a)f(b) < 0
```

```
%
%    delta   non-negative real number.
%
% Post:
%     root   the midpoint of an interval [alpha,beta]
%            with the property that f(alpha)f(beta)<=0 and
%            |beta-alpha| <= delta+eps*max(|alpha|,|beta|)
%
   fa = feval(fname,a);
   fb = feval(fname,b);
   if fa*fb > 0
      disp('Initial interval is not bracketing.')
          return
   end
   if nargin==3
      delta = 0;
   end
   while abs(a-b) > delta+eps*max(abs(a),abs(b))
      mid = (a+b)/2;
      fmid = feval(fname,mid);
      if fa*fmid<=0
             % There is a root in [a,mid].
         b  = mid;
                fb = fmid;
      else
             % There is a root in [mid,b].
         a  = mid;
                fa = fmid;
      end
   end
   root = (a+b)/2;
```

EXAMPLE 7.5. Find an approximation to $\sqrt{2}$ using the bisection method. Stop iterating when $|x_{n+1} - x_n| < 10^{-2}$.

SOLUTION. We need to find a root of $f(x) = x^2 - 2 = 0$. Choose $a_0 = 1$ and $b_0 = 2$, and obtain recursively

$$x_{n+1} = \frac{a_n + b_n}{2}$$

by the bisection method. The results are listed in Table 7.1. The answer is $\sqrt{2} \approx 1.414063$ with an accuracy of 10^{-2}. Note that a root lies in the interval $[1.414063, 1.421875]$. □

EXAMPLE 7.6. Show that the function $f(x) = x^3 + 4\,x^2 - 10$ has a unique root in the interval $[1, 2]$ and give an approximation to this root using eight iterations of the bisection method. Give a bound for the absolute error.

SOLUTION. Since

$$f(1) = -5 < 0 \quad \text{and} \quad f(2) = 14 > 0,$$

TABLE 7.1. Results of Example 7.5.

n	x_n	a_n	b_n	$\lvert x_{n-1} - x_n \rvert$	$f(x_n)$	$f(a_n)$
0		1	2			−
1	1.500000	1	1.500000	.500000	+	−
2	1.250000	1.250000	1.500000	.250000	−	−
3	1.375000	1.375000	1.500000	.125000	−	−
4	1.437500	1.375000	1.437500	.062500	+	−
5	1.406250	1.406250	1.437500	.031250	−	−
6	1.421875	1.406250	1.421875	.015625	+	−
7	1.414063	1.414063	1.421875	.007812	−	−

TABLE 7.2. Results of Example 7.6.

n	x_n	a_n	b_n	$f(x_n)$	$f(a_n)$
0		1	2		−
1	1.500000000	1	1.500000000	+	−
2	1.250000000	1.250000000	1.500000000	−	−
3	1.375000000	1.250000000	1.375000000	+	−
4	1.312500000	1.312500000	1.375000000	−	−
5	1.343750000	1.343750000	1.375000000	−	−
6	1.359375000	1.359375000	1.375000000	−	−
7	1.367187500	1.359375000	1.367187500	+	−
8	1.363281250	1.363281250	1.367187500	−	−

then $f(x)$ has a root, p, in $[1, 2]$. This root is unique since $f(x)$ is strictly increasing on $[1, 2]$; in fact

$$f'(x) = 3\,x^2 + 4\,x > 0 \quad \text{for all } x \text{ between 1 and 2.}$$

The results are listed in Table 7.2.

After eight iterations, we find that p lies between 1.363281250 and 1.367187500. Therefore, the absolute error in p is bounded by

$$1.367187500 - 1.363281250 = 0.00390625. \quad \square$$

EXAMPLE 7.7. Find the number of iterations needed in Example 7.6 to have an absolute error less than 10^{-4}.

SOLUTION. Since the root, p, lies in each interval $[a_n, b_n]$, after n iterations the error is at most $b_n - a_n$. Thus, we want to find n such that $b_n - a_n < 10^{-4}$. Since, at each iteration, the length of the interval is halved, it is easy to see that

$$b_n - a_n = (2 - 1)/2^n.$$

Therefore, n satisfies the inequality

$$2^{-n} < 10^{-4},$$

that is,

$$\ln 2^{-n} < \ln 10^{-4}, \quad \text{or} \quad -\,n \ln 2 < -4 \ln 10.$$

Thus,

$$n > 4\ln 10/\ln 2 = 13.28771238 \implies n = 14.$$

Hence, we need 14 iterations. □

7.4. Fixed Point Iteration

Let $f(x)$ be a real-valued function of a real variable x. In this section, we present iterative methods for solving equations of the form

$$f(x) = 0. \tag{7.1}$$

A *root* of the equation $f(x) = 0$, or a *zero* of $f(x)$, is a number p such that $f(p) = 0$.

To find a root of equation (7.1), we rewrite this equation in an equivalent form

$$x = g(x), \tag{7.2}$$

for instance, $g(x) = x - f(x)$. Hence, we say that p is a fixed point of g.

We say that (7.1) and (7.2) are *equivalent* (on a given interval) if any root of (7.1) is a fixed point for (7.2) and vice-versa.

Conversely, if, for a given initial value x_0, the sequence x_0, x_1, ..., defined by the recurrence

$$x_{n+1} = g(x_n), \qquad n = 0, 1, \ldots, \tag{7.3}$$

converges to a number p, we say that the fixed point method converges. If $g(x)$ is continuous, then $p = g(p)$. This is seen by taking the limit in equation (7.3) as $n \to \infty$. The number p is called a *fixed point* for the function $g(x)$ of the fixed point iteration (7.2).

It is easily seen that the two equations

$$x^3 + 9\,x - 9 = 0, \qquad x = (9 - x^3)/9$$

are equivalent. The problem is to choose a suitable function $g(x)$ and a suitable initial value x_0 to have convergence. To treat this question we need to define the different types of fixed points.

Definition 7.1. A fixed point, $p = g(p)$, of an iterative scheme

$$x_{n+1} = g(x_n),$$

is said to be *attractive*, *repulsive* or *indifferent* if the *multiplier*, $g'(p)$, of $g(x)$ at p satisfies

$$|g'(p)| < 1, \quad |g'(p)| > 1, \quad \text{or} \quad |g'(p)| = 1,$$

respectively.

Theorem 7.6 (Fixed Point Theorem). *Let $g(x)$ be a real-valued function satisfying the following conditions:*

1. *$g(x) \in [a, b]$ for all $x \in [a, b]$.*
2. *$g(x)$ is differentiable on $[a, b]$.*
3. *There exists a number K, $0 < K < 1$, such that $|g'(x)| \le K$ for all $x \in]a, b[$.*

Then $g(x)$ has a unique attractive fixed point $p \in [a, b]$. Moreover, for arbitrary $x_0 \in [a, b]$, the sequence x_0, x_1, x_2, ... defined by

$$x_{n+1} = g(x_n), \qquad n = 0, 1, 2, \ldots,$$

converges to p.

PROOF. If $g(a) = a$ or $g(b) = b$, the existence of an attractive fixed point is obvious. Suppose not, then it follows that $g(a) > a$ and $g(b) < b$. Define the auxiliary function

$$h(x) = g(x) - x.$$

Then h is continuous on $[a, b]$ and

$$h(a) = g(a) - a > 0, \qquad h(b) = g(b) - b < 0.$$

By Corollary 7.1, there exists a number $p \in]a, b[$ such that $h(p) = 0$, that is, $g(p) = p$ and p is a fixed point for $g(x)$.

To prove uniqueness, suppose that p and q are distinct fixed points for $g(x)$ in $[a, b]$. By the Mean Value Theorem 7.3, there exists a number c between p and q (and hence in $[a, b]$) such that

$$|p - q| = |g(p) - g(q)| = |g'(c)|\,|p - q| \le K|p - q| < |p - q|,$$

which is a contradiction. Thus $p = q$ and the attractive fixed point in $[a, b]$ is unique.

We now prove convergence. By the Mean Value Theorem 7.3, for each pair of numbers x and y in $[a, b]$, there exists a number c between x and y such that

$$g(x) - g(y) = g'(c)(x - y).$$

Hence,

$$|g(x) - g(y)| \le K|x - y|.$$

In particular,

$$|x_{n+1} - p| = |g(x_n) - g(p)| \le K|x_n - p|.$$

Repeating this procedure $n + 1$ times, we have

$$|x_{n+1} - p| \le K^{n+1}|x_0 - p| \to 0, \quad \text{as} \quad n \to \infty,$$

since $0 < K < 1$. Thus the sequence $\{x_n\}$ converges to p. □

EXAMPLE 7.8. Find a root of the equation

$$f(x) = x^3 + 9x - 9 = 0$$

in the interval $[0, 1]$ by a fixed point iterative scheme.

SOLUTION. Solving this equation is equivalent to finding a fixed point for

$$g(x) = (9 - x^3)/9.$$

Since

$$f(0)f(1) = -9 < 0,$$

Corollary 7.1 implies that $f(x)$ has a root, p, between 0 and 1. Condition (3) of Theorem 7.6 is satisfied with $K = 1/3$ since

$$|g'(x)| = |-x^2/3| \le 1/3$$

for all x between 0 and 1. The other conditions are also satisfied.

Five iterations are performed with Matlab starting with $x_0 = 0.5$. The function M-file `exp8_8.m` is

```
function x1 = exp8_8(x0); % Example 8.8.
x1 = (9-x0^3)/9;
```

TABLE 7.3. Results of Example 7.8.

n	x_n	error ϵ_n	$\epsilon_n/\epsilon_{n-1}$
0	0.50000000000000	−0.41490784153366	1.00000000000000
1	0.98611111111111	0.07120326957745	−0.17161225325185
2	0.89345451579409	−0.02145332573957	−0.30129691890395
3	0.92075445888550	0.00584661735184	−0.27252731920515
4	0.91326607850598	−0.00164176302768	−0.28080562295762
5	0.91536510274262	0.00045726120896	−0.27851839836463

The exact solution

$$0.914\,907\,841\,533\,66$$

is obtained by means of some 30 iterations by the following iterative procedure.

```
xexact =  0.91490784153366;
N = 5; x=zeros(N+1,4);
x0 = 0.5; x(1,:) = [0 x0 (x0-xexact), 1];
for i = 1:N
xt=exp8_8(x(i,2));
x(i+1,:) = [i xt (xt-xexact), (xt-xexact)/x(i,3)];
end
```

The iterates, their errors and the ratios of successive errors are listed in Table 7.3. One sees that the ratios of successive errors are nearly constant; therefore the order of convergence, defined in Subsection 7.4.2, is one. □

In Example 7.9 below, we shall show that the convergence of an iterative scheme $x_{n+1} = g(x_n)$ to an attractive fixed point depends upon a judicious rearrangement of the equation $f(x) = 0$ to be solved.

Besides fixed points, an iterative scheme may have cycles which are defined in Definition 7.2, where $g^2(x) = g(g(x))$, $g^3(x) = g(g^2(x))$, etc.

DEFINITION 7.2. Given an iterative scheme

$$x_{n+1} = g(x_n),$$

a *k-cycle* of $g(x)$ is a set of k distinct points,

$$x_0, \quad x_1, \quad x_2, \quad \ldots, \quad x_{k-1},$$

satisfying the relations

$$x_1 = g(x_0), \quad x_2 = g^2(x_0), \quad \ldots, \quad x_{k-1} = g^{k-1}(x_0), \quad x_0 = g^k(x_0).$$

The *multiplier* of a k cycle is

$$(g^k)'(x_j) = g'(x_{k-1}) \cdots g'(x_0), \qquad j = 0, 1, \ldots, k-1.$$

A k-cycle is *attractive*, *repulsive*, or *indifferent* as

$$|(g^k)'(x_j)| < 1, \quad > 1, \quad = 1.$$

A fixed point is a 1-cycle.

The multiplier of a cycle is seen to be the same at every point of the cycle.

EXAMPLE 7.9. Find a root of the equation

$$f(x) = x^3 + 4\,x^2 - 10 = 0$$

in the interval $[1, 2]$ by fixed point iterative schemes and study their convergence properties.

SOLUTION. Since $f(1)f(2) = -70 < 0$, the equation $f(x) = 0$ has a root in the interval $[1, 2]$. The exact roots are given by the Matlab command `roots`

```
p=[1 4 0 -10]; % the polynomial f(x)
r =roots(p)
 r =
    -2.68261500670705 + 0.35825935992404i
    -2.68261500670705 - 0.35825935992404i
     1.36523001341410
```

There is one real root, which we denote by x_∞, in the interval $[1, 2]$, and a pair of complex conjugate roots.

Six iterations are performed with the following five rearrangements $x = g_j(x)$, $j = 1, 2, 3, 4, 5$, of the given equation $f(x) = 0$. The derivative of $g_j'(x)$ is evaluated at the real root $x_\infty \approx 1.365$.

$$\begin{aligned}
x = g_1(x) &=: 10 + x - 4x^2 - x^3, && g_1'(x_\infty) \approx -15.51,\\
x = g_2(x) &=: \sqrt{(10/x) - 4x}, && g_2'(x_\infty) \approx -3.42,\\
x = g_3(x) &=: \frac{1}{2}\sqrt{10 - x^3}, && g_3'(x_\infty) \approx -0.51,\\
x = g_4(x) &=: \sqrt{10/(4 + x)}, && g_4'(x_\infty) \approx -0.13\\
x = g_5(x) &=: x - \frac{x^3 + 4x^2 - 10}{3x^2 + 8x}, && g_5'(x_\infty) = 0.
\end{aligned}$$

The Matlab function M-file `exp1_9.m` is

```
function y = exp1_9(x); % Example 1.9.
y = [10+x(1)-4*x(1)^2-x(1)^3; sqrt((10/x(2))-4*x(2));
 sqrt(10-x(3)^3)/2; sqrt(10/(4+x(4)));
 x(5)-(x(5)^3+4*x(5)^2-10)/(3*x(5)^2+8*x(5))]';
```

The following iterative procedure is used.

```
N = 6; x=zeros(N+1,5);
x0 = 1.5; x(1,:) = [0 x0 x0 x0 x0];
for i = 1:N
xt=exp1_9(x(i,2:5));
x(i+1,:) = [i xt];
end
```

The results are summarized in Table 7.4. We see from the table that x_∞ is an attractive fixed point of $g_3(x)$, $g_4(x)$ and $g_5(x)$. Moreover, $g_4(x_n)$ converges more quickly to the root 1.365 230 013 than $g_3(x_n)$, and $g_5(x)$ converges even faster. In fact, these three fixed point methods need 30, 15 and 4 iterations, respectively,

TABLE 7.4. Results of Example 7.9.

	$g_1(x)$	$g_2(x)$	$g_3(x)$	$g_4(x)$	$g_5(x)$
n	$10+x-4x^2-x^3$	$\sqrt{(10/x)-4x}$	$0.5\sqrt{10-x^3}$	$\sqrt{10/(4+x)}$	$x-\frac{x^3+4x^2-10}{2x^2+8x}$
0	1.5	1.5	1.5	1.5	1.5
1	-0.8750	0.816	1.286953	1.348399	1.373333333
2	6.732421875	2.996	1.402540	1.367376	1.365262015
3	-4.6972001×10^2	$0.00-2.94\,i$	1.345458	1.364957	1.365230014
4	1.0275×10^8	$2.75-2.75\,i$	1.375170	1.365264	1.365230013
5	-1.08×10^{24}	$1.81-3.53\,i$	1.360094	1.365225	
6	1.3×10^{72}	$2.38-3.43\,i$	1.367846	1.365230	

to produce a 10-digit correct answer. On the other hand, the sequence $g_2(x_n)$ is trapped in an attractive two-cycle,

$$z_\pm = 2.27475487839820 \pm 3.60881272309733\,i,$$

with multiplier

$$g_2'(z_+)g_2'(z_-) = 0.19790433047378$$

which is smaller than one in absolute value. Once in an attractive cycle, an iteration cannot converge to a fixed point. Finally x_∞ is a repulsive fixed point of $g_1(x)$ and $x_{n+1} = g(x_n)$ diverges to $-\infty$. □

REMARK 7.2. An iteration started in the basin of attraction of an attractive fixed point (or cycle) will converge to that fixed point (or cycle). An iteration started near a repulsive fixed point (or cycle) will not converge to that fixed point (or cycle). Convergence to an indifferent fixed point is very slow, but can be accelerated by different acceleration processes.

7.4.1. Stopping criteria. Three usual criteria that are used to decide when to stop an iteration procedure to find a zero of $f(x)$ are:

(1) Stop after N iterations (for a given N).
(2) Stop when $|x_{n+1}-x_n| < \epsilon$ (for a given ϵ).
(3) Stop when $|f(x_n)| < \eta$ (for a given η).

The usefulness of any of these criteria is problem dependent.

7.4.2. Order and rate of convergence of an iterative method. We are often interested in the rate of convergence of an iterative scheme. Suppose that the function $g(x)$ for the iterative method

$$x_{n+1} = g(x_n)$$

has a Taylor expansion about the fixed point p ($p = g(p)$) and let

$$\epsilon_n = x_n - p.$$

Then, we have

$$\begin{aligned} x_{n+1} = g(x_n) = g(p+\epsilon_n) &= g(p) + g'(p)\epsilon_n + \frac{g''(p)}{2!}\,\epsilon_n^2 + \dots \\ &= p + g'(p)\epsilon_n + \frac{g''(p)}{2!}\,\epsilon_n^2 + \dots. \end{aligned}$$

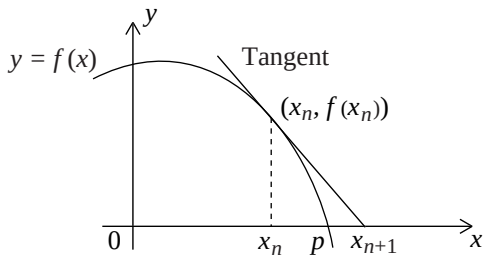

FIGURE 7.2. The nth step of Newton's method.

Hence,

$$\epsilon_{n+1} = x_{n+1} - p = g'(p)\epsilon_n + \frac{g''(p)}{2!}\,\epsilon_n^2 + \dots. \tag{7.4}$$

DEFINITION 7.3. The *order of convergence of an iterative method* $x_{n+1} = g(x_n)$ is the order of the first non-zero derivative of $g(x)$ at p. A method of order p is said to have a *rate of convergence* p.

In Example 7.9, the iterative schemes $g_3(x)$ and $g_4(x)$ converge to first order, while $g_5(x)$ converges to second order.

Note that, for a second-order iterative scheme, we have

$$\frac{\epsilon_{n+1}}{\epsilon_n^2} \approx \frac{g''(p)}{2} = \text{constant}.$$

7.5. Newton's, Secant, and False Position Methods

7.5.1. Newton's method. Let x_n be an approximation to a root, p, of $f(x) = 0$. Draw the tangent line

$$y = f(x_n) + f'(x_n)(x - x_n)$$

to the curve $y = f(x)$ at the point $(x_n, f(x_n))$ as shown in Fig. 7.2. Then x_{n+1} is determined by the point of intersection, $(x_{n+1}, 0)$, of this line with the x-axis,

$$0 = f(x_n) + f'(x_n)\,(x_{n+1} - x_n).$$

If $f'(x_n) \neq 0$, solving this equation for x_{n+1} we obtain *Newton's method*, also called the *Newton–Raphson method*,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{7.5}$$

Note that Newton's method is a fixed point method since it can be rewritten in the form

$$x_{n+1} = g(x_n), \qquad \text{where} \quad g(x) = x - \frac{f(x)}{f'(x)}.$$

EXAMPLE 7.10. Approximate $\sqrt{2}$ by Newton's method. Stop when $|x_{n+1} - x_n| < 10^{-4}$.

SOLUTION. We wish to find a root to the equation

$$f(x) = x^2 - 2 = 0.$$

TABLE 7.5. Results of Example 7.10.

n	x_n	$\lvert x_n - x_{n-1}\rvert$
0	2	
1	1.5	0.5
2	1.416667	0.083333
3	1.414216	0.002451
4	1.414214	0.000002

TABLE 7.6. Results of Example 7.11.

n	x_n	$\lvert x_n - x_{n-1}\rvert$
0	1.5	
1	1.37333333333333	0.126667
2	1.36526201487463	0.00807132
3	1.36523001391615	0.000032001
4	1.3652300134141	5.0205×10^{-10}
5	1.3652300134141	2.22045×10^{-16}
6	1.3652300134141	2.22045×10^{-16}

In this case, Newton's method becomes

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - 2}{2\,x_n} = \frac{x_n^2 + 2}{2x_n}.$$

With $x_0 = 2$, we obtain the results listed in Table 7.5. Therefore,

$$\sqrt{2} \approx 1.414214.$$

Note that the number of zeros in the errors roughly doubles as it is the case with methods of second order. □

EXAMPLE 7.11. Use six iterations of Newton's method to approximate a root $p \in [1, 2]$ of the polynomial

$$f(x) = x^3 + 4\,x^2 - 10 = 0$$

given in Example 7.9.

SOLUTION. In this case, Newton's method becomes

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^3 + 4x_n^2 - 10}{3x_n^2 + 8x_n} = \frac{2(x_n^3 + 2x_n^2 + 5)}{3x_n^2 + 8x_n}.$$

We take $x_0 = 1.5$. The results are listed in Table 7.6. □

EXAMPLE 7.12. Use Newton's method to approximate the solution of

$$\ln x = \cos x$$

to six decimal places.

SOLUTION. Let $f(x) = \ln x - \cos x$; thus $f(x) = 0$ when $\ln x = \cos x$. From the graph, it is easy to see that the solution is between $x = 1$ and $x = \pi/2$, so we will use $x_0 = 1$.

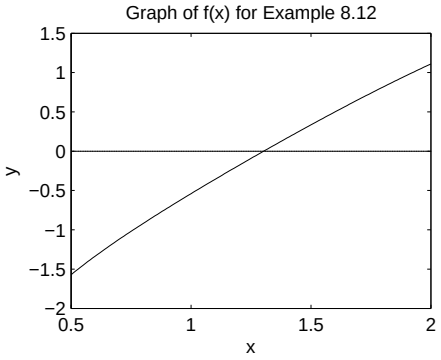

FIGURE 7.3. Graph of $\ln x - \cos x$ for Example 7.12

We have

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{\ln x_n - \cos x_n}{(1/x_n) + \sin x_n}.$$

Hence,

$$\begin{aligned} x_0 &= 1, \\ x_1 &= 1 - \frac{\ln 1 - \cos 1}{(1/1) + \sin 1} = 1.293\,408. \\ x_2 &= 1.302\,956 \\ x_3 &= 1.302\,964, \\ x_4 &= 1.302\,964, \\ &\text{stop} \end{aligned}$$

We can easily verify that the answer is correct:

$$\begin{aligned} \ln 1.302\,964 &\approx 0.264\,641\,6, \\ \cos 1.302\,964 &\approx 0.264\,641\,6. \end{aligned}$$

Therefore the solution, to six decimal places, is 1.302 964. □

THEOREM 7.7. *Let p be a simple root of $f(x) = 0$, that is, $f(p) = 0$ and $f'(p) \neq 0$. If $f''(p)$ exists, then Newton's method is at least of second order near p.*

PROOF. Differentiating the function

$$g(x) = x - \frac{f(x)}{f'(x)}$$

we have

$$\begin{aligned} g'(x) &= 1 - \frac{(f'(x))^2 - f(x)\,f''(x)}{(f'(x))^2} \\ &= \frac{f(x)\,f''(x)}{(f'(x))^2}. \end{aligned}$$

Since $f(p) = 0$, we have

$$g'(p) = 0.$$

TABLE 7.7. Results of Example 7.13.

	Newton's Method		ModifiedNewton	
n	x_n	$\epsilon_{n+1}/\epsilon_n$	x_n	$\epsilon_{n+1}/\epsilon_n^2$
0	0.000		0.00000000000000	
1	0.400	0.600	0.80000000000000	−0.2000
2	0.652	2.245	0.98461538461538	−0.3846
3	0.806	0.143	0.99988432620012	−0.4887
4	0.895	0.537	0.99999999331095	−0.4999
5	0.945	0.522	1	
6	0.972	0.512	1	

Therefore, Newton's method is of order two near a simple zero of f. □

REMARK 7.3. Taking the second derivative of $g(x)$ in Newton's method, we have

$$g''(x) = \frac{(f'(x))^2 f''(x) + f(x)f'(x)f'''(x) - 2f(x)(f''(x))^2}{(f'(x))^3}.$$

If $f'''(p)$ exists, we obtain

$$g''(p) = -\frac{f''(p)}{f'(p)}.$$

Thus, by (7.4), the successive errors satisfy the approximate relation

$$\epsilon_{n+1} \approx -\frac{1}{2}\,\frac{f''(p)}{f'(p)}\,\epsilon_n^2,$$

which explains the doubling of the number of leading zeros in the error of Newton's method near a simple root of $f(x) = 0$.

EXAMPLE 7.13. Use six iterations of the ordinary and modified Newton's methods

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad x_{n+1} = x_n - 2\,\frac{f(x_n)}{f'(x_n)}$$

to approximate the double root, $x = 1$, of the polynomial

$$f(x) = (x-1)^2(x-2).$$

SOLUTION. The two methods have iteration functions

$$g_1(x) = x - \frac{(x-1)(x-2)}{2(x-2)+(x-1)}, \qquad g_2(x) = x - \frac{(x-1)(x-2)}{(x-2)+(x-1)},$$

respectively. We take $x_0 = 0$. The results are listed in Table 7.7. One sees that Newton's method has first-order convergence near a double zero of $f(x)$, but one can verify that the modified Newton method has second-order convergence. In fact, near a root of multiplicity m the modified Newton method

$$x_{n+1} = x_n - m\,\frac{f(x_n)}{f'(x_n)}$$

has second-order convergence. □

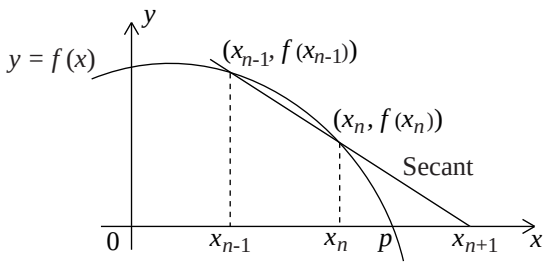

FIGURE 7.4. The nth step of the secant method.

In general, Newton's method may converge to the desired root, to another root, or to an attractive cycle, especially in the complex plane. The location of our initial guess x_0 relative to the root p, other roots and local extrema can affect the outcome.

7.5.2. The secant method. Let x_{n-1} and x_n be two approximations to a root, p, of $f(x) = 0$. Draw the secant to the curve $y = f(x)$ through the points $(x_{n-1}, f(x_{n-1}))$ and $(x_n, f(x_n))$. The equation of this secant is

$$y = f(x_n) + \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}\,(x - x_n).$$

The $(n+1)$st iterate x_{n+1} is determined by the point of intersection $(x_{n+1}, 0)$ of the secant with the x-axis as shown in Fig. 7.4,

$$0 = f(x_n) + \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}\,(x_{n+1} - x_n).$$

Solving for x_{n+1}, we obtain the *secant method*:

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}\,f(x_n). \tag{7.6}$$

The algorithm for the secant method is as follows.

ALGORITHM 7.2 (Secant Method). Given that $f(x)$ is continuous on $[a, b]$ and has a root in $[a, b]$.

(1) Choose x_0 and x_1 near the root p that is sought.
(2) Given x_{n-1} and x_n, x_{n+1} is obtained by the formula

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}\,f(x_n),$$

provided that $f(x_n) - f(x_{n-1}) \neq 0$. If $f(x_n) - f(x_{n-1}) = 0$, try other starting values x_0 and x_1.
(3) Repeat (2) until the selected stopping criterion is satisfied (see Subsection 7.4.1).

This method converges to a simple root to order 1.618 and may not converge to a multiple root. Thus it is generally slower than Newton's method. However, it does not require the derivative of $f(x)$. In general applications of Newton's method, the derivative of the function $f(x)$ is approximated numerically by the slope of a secant to the curve.

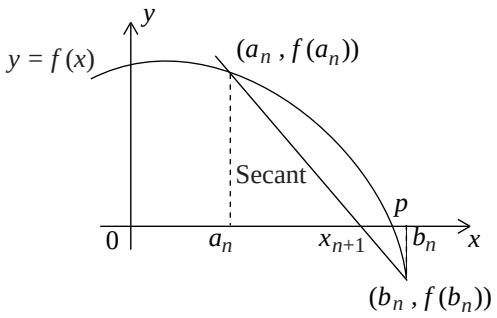

FIGURE 7.5. The nth step of the method of false position.

7.5.3. The method of false position. The *method of false position*, also called *regula falsi*, is similar to the secant method, but with the additional condition that, for each $n = 0, 1, 2, \ldots$, the pair of approximate values, a_n and b_n, to the root, p, of $f(x) = 0$ be such that $f(a_n)\, f(b_n) < 0$. The next iterate, x_{n+1}, is determined by the intersection of the secant passing through the points $(a_n, f(a_n))$ and $(b_n, f(b_n))$ with the x-axis.

The equation for the secant through $(a_n, f(a_n))$ and $(b_n, f(b_n))$, shown in Fig. 7.5, is

$$y = f(a_n) + \frac{f(b_n) - f(a_n)}{b_n - a_n}\,(x - a_n).$$

Hence, x_{n+1} satisfies the equation

$$0 = f(a_n) + \frac{f(b_n) - f(a_n)}{b_n - a_n}\,(x_{n+1} - a_n),$$

which leads to the *method of false position*:

$$x_{n+1} = \frac{a_n\, f(b_n) - b_n\, f(a_n)}{f(b_n) - f(a_n)}. \tag{7.7}$$

The algorithm for the method of false position is as follows.

ALGORITHM 7.3 (False Position Method). Given that $f(x)$ is continuous on $[a, b]$ and that $f(a)\, f(b) < 0$.

(1) Pick $a_0 = a$ and $b_0 = b$.
(2) Given a_n and b_n such that $f(a_n) f(b_n) < 0$, compute
$$x_{n+1} = \frac{a_n\, f(b_n) - b_n\, f(a_n)}{f(b_n) - f(a_n)}.$$
(3) If $f(x_{n+1}) = 0$, stop.
(4) Else if $f(x_{n+1})$ and $f(a_n)$ have opposite signs, set $a_{n+1} = a_n$ and $b_{n+1} = x_{n+1}$;
(5) Else set $a_{n+1} = x_{n+1}$ and $b_{n+1} = b_n$.
(6) Repeat (2)–(5) until the selected stopping criterion is satisfied (see Subsection 7.4.1).

This method is generally slower than Newton's method, but it does not require the derivative of $f(x)$ and it always converges to a nested root. If the approach

TABLE 7.8. Results of Example 7.14.

n	x_n	a_n	b_n	$\lvert x_{n-1} - x_n \rvert$	$f(x_n)$	$f(a_n)$
0		1	2			$-$
1	1.333333	1.333333	2		$-$	$-$
2	1.400000	1.400000	2	0.066667	$-$	$-$
3	1.411765	1.411765	2	0.011765	$-$	$-$
4	1.413793	1.413793	2	0.002028	$-$	$-$
5	1.414141	1.414141	2	0.000348	$-$	$-$

to the root is one-sided, convergence can be accelerated by replacing the value of $f(x)$ at the stagnant end position with $f(x)/2$.

EXAMPLE 7.14. Approximate $\sqrt{2}$ by the method of false position. Stop iterating when $|x_{n+1} - x_n| < 10^{-3}$.

SOLUTION. This problem is equivalent to the problem of finding a root of the equation

$$f(x) = x^2 - 2 = 0.$$

We have

$$x_{n+1} = \frac{a_n\,(b_n^2 - 2) - b_n\,(a_n^2 - 2)}{(b_n^2 - 2) - (a_n^2 - 2)} = \frac{a_n\,b_n + 2}{a_n + b_n}.$$

Choose $a_0 = 1$ and $b_0 = 2$. Notice that $f(1) < 0$ and $f(2) > 0$. The results are listed in Table 7.8. Therefore, $\sqrt{2} \approx 1.414141$. □

7.5.4. A global Newton-bisection method. The many difficulties that can occur with Newton's method can be handled with success by combining the Newton and bisection ideas in a way that captures the best features of each framework. At the beginning, it is assumed that we have a bracketing interval $[a, b]$ for $f(x)$, that is, $f(a)f(b) < 0$, and that the initial value x_c is one of the endpoints. If

$$x_+ = x_c - \frac{f(x_c)}{f'(x_c)} \in [a, b],$$

we proceed with either $[a, x_+]$ or $[x_+, b]$, whichever is bracketing. The new x_c equals x_+. If the Newton step falls out of $[a, b]$, we take a bisection step setting the new x_c to $(a + b)/2$. In a typical situation, a number of bisection steps are taken before the Newton iteration takes over. This globalization of the Newton iteration is programmed in the following Matlab function M-file which is found in `ftp://ftp.cs.cornell.edu/pub/cv`.

```
function [x,fx,nEvals,aF,bF] = ...
                GlobalNewton(fName,fpName,a,b,tolx,tolf,nEvalsMax)

% Pre:
%   fName       string that names a function f(x).
%   fpName      string that names the derivative function f'(x).
%   a,b         A root of f(x) is sought in the interval [a,b]
%               and f(a)*f(b)<=0.
%   tolx,tolf   Nonnegative termination criteria.
%   nEvalsMax   Maximum number of derivative evaluations.
```

```
%
% Post:
%   x          An approximate zero of f.
%   fx         The value of f at x.
%   nEvals     The number of derivative evaluations required.
%   aF,bF      The final bracketing interval is [aF,bF].
%
% Comments:
%   Iteration terminates as soon as x is within tolx of a true zero
%   or if |f(x)|<= tolf or after nEvalMax f-evaluations

   fa  = feval(fName,a);
   fb  = feval(fName,b);
   if fa*fb>0
      disp('Initial interval not bracketing.')
          return
   end
   x   = a;
   fx  = feval(fName,x);
   fpx = feval(fpName,x);
   disp(sprintf('%20.15f  %20.15f    %20.15f',a,x,b))

   nEvals = 1;
   while (abs(a-b) > tolx ) & (abs(fx) > tolf) &
                    ((nEvals<nEvalsMax) | (nEvals==1))
      %[a,b] brackets a root and x = a or x = b.
      if StepIsIn(x,fx,fpx,a,b)
              %Take Newton Step
                  disp('Newton')
              x   = x-fx/fpx;
          else
              %Take a Bisection Step:
                  disp('Bisection')
                  x = (a+b)/2;
          end
      fx  = feval(fName,x);
      fpx = feval(fpName,x);
          nEvals = nEvals+1;
      if fa*fx<=0
             % There is a root in [a,x]. Bring in right endpoint.
         b  = x;
                 fb = fx;
      else
             % There is a root in [x,b]. Bring in left endpoint.
         a  = x;
                 fa = fx;
      end
          disp(sprintf('%20.15f  %20.15f    %20.15f',a,x,b))
```

```
end
aF = a;
bF = b;
```

7.5.5. The Matlab `fzero` function. The MATLAB `fzero` function is a general-purpose root finder that does not require derivatives. A simple call involves only the name of the function and a starting value x_0. For example

```
aroot = fzero('function_name', x0)
```

The value returned is near a point where the function changes sign, or `NaN` if the search fails. Other options are described in `help fzero`.

7.6. Aitken–Steffensen Accelerated Convergence

The linear convergence of an iterative method, $x_{n+1} = g(x_n)$, can be accelerated by Aitken's process. Suppose that the sequence $\{x_n\}$ converges to a fixed point p to first order. Then the following ratios are approximately equal:

$$\frac{x_{n+1} - p}{x_n - p} \approx \frac{x_{n+2} - p}{x_{n+1} - p}.$$

We make this an equality by substituting a_n for p,

$$\frac{x_{n+1} - a_n}{x_n - a_n} = \frac{x_{n+2} - a_n}{x_{n+1} - a_n}$$

and solve for a_n which, after some algebraic manipulation, becomes

$$a_n = x_n - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n}.$$

This is Aitken's process which accelerates convergence in the sense that

$$\lim_{n\to\infty} \frac{a_n - p}{x_n - p} = 0.$$

If we introduce the first- and second-order forward differences:

$$\Delta x_n = x_{n+1} - x_n, \qquad \Delta^2 x_n = \Delta(\Delta x_n) = x_{n+2} - 2x_{n+1} + x_n,$$

then *Aitken's process* becomes

$$a_n = x_n - \frac{(\Delta x_n)^2}{\Delta^2 x_n}. \tag{7.8}$$

Steffensen's process assumes that $s_1 = a_0$ is a better value than x_2. Thus $s_0 = x_0$, $z_1 = g(s_0)$ and $z_2 = g(z_1)$ are used to produce s_1. Next, s_1, $z_1 = g(s_1)$ and $z_2 = g(z_2)$ are used to produce s_2. And so on. The algorithm is as follows.

ALGORITHM 7.4 (Steffensen's Algorithm). Set

$$s_0 = x_0,$$

and, for $n = 0, 1, 2, \ldots$,

$$z_1 = g(s_n),$$
$$z_2 = g(z_1),$$
$$s_{n+1} = s_n - \frac{(z_1 - s_n)^2}{z_2 - 2z_1 + s_n}.$$

Steffensen's process applied to a first-order fixed point method produces a second-order method.

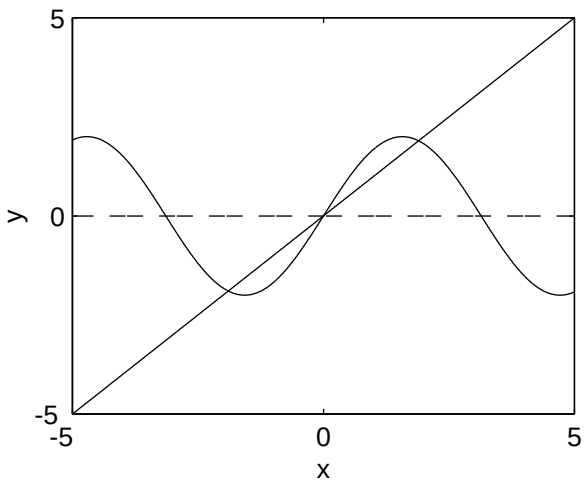

FIGURE 7.6. The three real roots of $x = 2\sin x$ in Example 7.15.

TABLE 7.9. Results of Example 7.15.

n	x_n	a_n	s_n	$\epsilon_{n+1}/\epsilon_n^2$
0	1.00000000000000	2.23242945471637	1.00000000000000	−0.2620
1	1.68294196961579	1.88318435428750	2.23242945471637	0.3770
2	1.98743653027215	1.89201364327283	1.83453173271065	0.3560
3	1.82890755262358	1.89399129067379	1.89422502453561	0.3689
4	1.93374764234016	1.89492839486397	1.89549367325365	0.3691
5	1.86970615363078	1.89525656226218	1.89549426703385	
6	1.91131617912526		1.89549426703398	
7	1.88516234821223		NaN	

EXAMPLE 7.15. Consider the fixed point iteration $x_{n+1} = g(x_n)$:

$$x_{n+1} = 2\sin x_n, \qquad x_0 = 1.$$

Do seven iterations and perform Aitken's and Steffensen's accelerations.

SOLUTION. The three real roots of $x = 2\sin x$ can be seen in Fig. 7.6. The Matlab function `fzero` produces the fixed point near $x = 1$:

```
p = fzero('x-2*sin(x)',1.)
  p = 1.89549426703398
```

The convergence is linear since

$$g'(p) = -0.63804504828524 \neq 0.$$

The following Matlab M function and script produce the results listed in Table 7.9. The second, third, and fourth columns are the iterates x_n, Aitken's and Steffensen's accelerated sequences a_n and s_n, respectively. The fifth column, which lists $\epsilon_{n+1}/\epsilon_n^2 = (s_{n+2} - s_{n+1})/(s_{n+1} - s_n)^2$ tending to a constant, indicates that the Steffensen sequence s_n converges to second order.

The M function function is:

```
function f = twosine(x);
f = 2*sin(x);
```

The M script function is:

```
n = 7;
x = ones(1,n+1);
x(1) = 1.0;
for k = 1:n
x(k+1)=twosine(x(k)); % iterating x(k+1) = 2*sin(x(k))
end

a = ones(1,n-1);
for k = 1:n-1
a(k) = x(k) - (x(k+1)-x(k))^2/(x(k+2)-2*x(k+1)+x(k)); % Aitken
end

s = ones(1,n+1);
s(1) = 1.0;
for k = 1:n
z1=twosine(s(k));
z2=twosine(z1);
s(k+1) = s(k) - (z1-s(k))^2/(z2-2*z1+s(k)); % Steffensen
end

d = ones(1,n-2);
for k = 1:n-2
d(k) = (s(k+2)-s(k+1))/(s(k+1)-s(k))^2; % 2nd order convergence
end
```

Note that the Matlab program produced `NaN` (not a number) for s_7 because of a division by zero. □

7.7. Horner's Method and the Synthetic Division

7.7.1. Horner's method. To reduce the number of products in the evaluation of polynomials, these should be expressed in nested form. For instance,

$$\begin{aligned} p(x) &= a_3x^3 + a_2x^2 + a_1x + a_0 \\ &= \big((a_3x + a_2)x + a_1\big)x + a_0. \end{aligned}$$

In this simple case, the reduction is from 8 to 3 products.

The Matlab command `horner` transforms a symbolic polynomial into its Horner, or nested, representation.

```
syms x
p = x^3-6*x^2+11*x-6
   p = x^3-6*x^2+11*x-6
hp = horner(p)
   hp = -6+(11+(-6+x)*x)*x
```

Horner's method incorporates this nesting technique.

THEOREM 7.8 (Horner's Method). *Let*

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0.$$

If $b_n = a_n$ *and*

$$b_k = a_k + b_{k+1}x_0, \quad \textit{for} \quad k = n-1, n-2, \ldots, 1, 0,$$

then

$$b_0 = p(x_0).$$

Moreover, if

$$q(x) = b_n x^{n-1} + b_{n-1}x^{n-2} + \cdots + b_2 x + b_1,$$

then

$$p(x) = (x - x_0)q(x) + b_0.$$

PROOF. By the definition of $q(x)$,

$$\begin{aligned}
(x - x_0)q(x) + b_0 &= (x - x_0)(b_n x^{n-1} + b_{n-1}x^{n-2} + \cdots + b_2 x + b_1) + b_0 \\
&= (b_n x^n + b_{n-1}x^{n-1} + \cdots + b_2 x^2 + b_1 x) \\
&\quad - (b_n x_0 x^{n-1} + b_{n-1}x_0 x^{n-2} + \cdots + b_2 x_0 x + b_1 x_0) + b_0 \\
&= b_n x^n + (b_{n-1} - b_n x_0)x^{n-1} + \cdots + (b_1 - b_2 x_0)x + (b_0 - b_1 x_0) \\
&= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \\
&= p(x)
\end{aligned}$$

and

$$b_0 = p(x_0). \quad \square$$

7.7.2. Synthetic division. Evaluating a polynomial at $x = x_0$ by Horner's method is equivalent to applying the synthetic division as shown in Example 7.16.

EXAMPLE 7.16. Find the value of the polynomial

$$p(x) = 2x^4 - 3x^2 + 3x - 4$$

at $x_0 = -2$ by Horner's method.

SOLUTION. By successively multiplying the elements of the third line of the following tableau by $x_0 = -2$ and adding to the first line, one gets the value of $p(-2)$.

$a_4 = 2$	$a_3 = 0$	$a_2 = -3$	$a_1 = 3$	$a_0 = -4$
	-4	8	-10	14
$b_4 = 2$	$b_3 = -4$	$b_2 = 5$	$b_1 = -7$	$b_0 = 10$

Thus

$$p(x) = (x + 2)(2x^3 - 4x^2 + 5x - 7) + 10$$

and

$$p(-2) = 10. \quad \square$$

Horner's method can be used efficiently with Newton's method to find zeros of a polynomial $p(x)$. Differentiating

$$p(x) = (x - x_0)q(x) + b_0$$

we obtain

$$p'(x) = (x - x_0)q'(x) + q(x).$$

Hence

$$p'(x_0) = q(x_0).$$

Putting this in Newton's method we have

$$\begin{aligned} x_n &= x_{n-1} - \frac{p(x_{n-1})}{p'(x_{n-1})} \\ &= x_{n-1} - \frac{p(x_{n-1})}{q(x_{n-1})}. \end{aligned}$$

This procedure is shown in Example 7.17.

EXAMPLE 7.17. Compute the value of the polynomial

$$p(x) = 2x^4 = 3x^3 + 3x - 4$$

and of its derivative $p'(x)$ at $x_0 = -2$ by Horner's method and apply the results to Newton's method to find the first iterate x_1.

SOLUTION. By successively multiplying the elements of the third line of the following tableau by $x_0 = -2$ and adding to the first line, one gets the value of $p(-2)$. Then by successively multiplying the elements of the fifth line of the tableau by $x_0 = -2$ and adding to the third line, one gets the value of $p'(-2)$.

2	0	-3	$= 3$	-4
	-4	8	-10	14
2	-4	5	-7	$10 = p(-2)$
	-4	16	-42	
2	-8	21	$-49 = p'(-2)$	

Thus

$$p(-2) = 10, \qquad p'(-2) = -49,$$

and

$$x_1 = -2 - \frac{10}{-49} \approx -1.7959. \quad \square$$

7.8. Müller's Method

Müller's, or the parabola, method finds the real or complex roots of an equation

$$f(x) = 0.$$

This method uses three initial approximations, x_0, x_1, and x_2, to construct a parabola,

$$p(x) = a(x - x_2)^2 + b(x - x_2) + c,$$

through the three points $(x_0, f(x_0))$, $(x_1, f(x_1))$, and $(x_2, f(x_2))$ on the curve $f(x)$ and determines the next approximation x_3 as the point of intersection of the parabola with the real axis closer to x_2.

The coefficients a, b and c defining the parabola are obtained by solving the linear system

$$\begin{aligned} f(x_0) &= a(x_0 - x_2)^2 + b(x_0 - x_2) + c, \\ f(x_1) &= a(x_1 - x_2)^2 + b(x_1 - x_2) + c, \\ f(x_2) &= c. \end{aligned}$$

We immediately have

$$c = f(x_2)$$

and obtain a and b from the linear system

$$\begin{bmatrix} (x_0 - x_2)^2 & (x_0 - x_2) \\ (x_1 - x_2)^2 & (x_1 - x_2) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} f(x_0) - f(x_2) \\ f(x_1) - f(x_2) \end{bmatrix}.$$

Then, we set

$$p(x_3) = a(x_3 - x_2)^2 + b(x_3 - x_2) + c = 0$$

and solve for $x_3 - x_2$:

$$\begin{aligned} x_3 - x_2 &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \times \frac{-b \mp \sqrt{b^2 - 4ac}}{-b \mp \sqrt{b^2 - 4ac}} \\ &= \frac{-2c}{b \pm \sqrt{b^2 - 4ac}}. \end{aligned}$$

To find x_3 closer to x_2, we maximize the denominator:

$$x_3 = x_2 - \frac{2c}{b + \text{sign}(b)\sqrt{b^2 - 4ac}}.$$

Müller's method converges approximately to order 1.839 to a simple or double root. It may not converge to a triple root.

EXAMPLE 7.18. Find the four zeros of the polynomial

$$16x^4 - 40x^3 + 5x^2 + 20x + 6,$$

whose graph is shown in Fig. 7.7, by means of Müller's method.

SOLUTION. The following Matlab commands do one iteration of Müller's method on the given polynomial which is transformed into its nested form:

```
syms x
pp = 16*x^4-40*x^3+5*x^2+20*x+6
 pp = 16*x^4-40*x^3+5*x^2+20*x+6
pp = horner(pp)
 pp = 6+(20+(5+(-40+16*x)*x)*x)*x
```

The polynomial is evaluated by the Matlab M function:

```
function pp = mullerpol(x);
pp = 6+(20+(5+(-40+16*x)*x)*x)*x;
```

Müller's method obtains x_3 with the given three starting values:

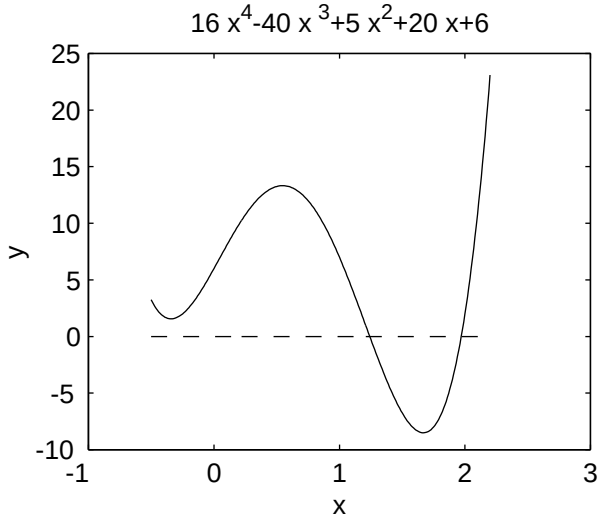

FIGURE 7.7. The graph of the polynomial $16x^4 - 40x^3 + 5x^2 + 20x + 6$ for Example 7.18.

```
x0 = 0.5; x1 = -0.5; x2 = 0; % starting values
m = [(x0-x2)^2 x0-x2; (x1-x2)^2 x1-x2];
rhs = [mullerpol(x0)-mullerpol(x2); mullerpol(x1)- mullerpol(x2)];
ab = m\rhs; a = ab(1); b = ab(2); % coefficients a and b
c = mullerpol(x2); % coefficient c
x3 = x2 -(2*c)/(b+sign(b)*sqrt(b^2-4*a*c))
  x3 = -0.5556 + 0.5984i
```

The method is iterated until convergence. The four roots of this polynomial are

```
rr = roots([16 -40 5 20 6])'
  rr = 1.9704  1.2417  -0.3561 - 0.1628i  -0.3561 + 0.1628i
```

The two real roots can be obtained by Müller's method with starting values $[0.5, 1.0, 1.5]$ and $[2.5, 2.0, 2.25]$, respectively. □

CHAPTER 8

Interpolation and Extrapolation

Quite often, experimental results provide only a few values of an unknown function $f(x)$, say,

$$(x_0, f_0), \quad (x_1, f_1), \quad (x_2, f_2), \quad \ldots, \quad (x_n, f_n), \tag{8.1}$$

where f_i is the observed value for $f(x_i)$. We would like to use these data to approximate $f(x)$ at an arbitrary point $x \neq x_i$.

When we want to estimate $f(x)$ for x between two of the x_i's, we talk about *interpolation* of $f(x)$ at x. When x is not between two of the x_i's, we talk about *extrapolation* of $f(x)$ at x.

The idea is to construct an interpolating polynomial, $p_n(x)$, of degree n whose graph passes through the $n+1$ points listed in (8.1). This polynomial will be used to estimate $f(x)$.

We will make use of an important fact. If we have a set of $n+1$ data points (8.1), where the *nodes*, x_i, are distinct, there is a unique polynomial of degree less than or equal to n, $p_n(x)$, that passes through the points, that is, $p_n(x_j) = f_j$ for all j. For, suppose $p_n(x)$ and $q_n(x)$ of degree n both interpolate $f(x)$ at $n+1$ distinct points, then

$$p_n(x) - q_n(x)$$

is a polynomial of degree n which admits $n+1$ distinct zeros, hence it is identically zero.

8.1. Lagrange Interpolating Polynomial

The Lagrange interpolating polynomial, $p_n(x)$, of degree n through the $n+1$ points $\big(x_k, f(x_k)\big)$, $k = 0, 1, \ldots, n$, is expressed in terms of the following Lagrange basis:

$$L_k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}.$$

Clearly, $L_k(x)$ is a polynomial of degree n and

$$L_k(x) = \begin{cases} 1, & x = x_k, \\ 0, & x = x_j, \quad j \neq k. \end{cases}$$

Then the *Lagrange interpolating polynomial* of $f(x)$ is

$$p_n(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + \cdots + f(x_n)L_n(x). \tag{8.2}$$

It is of degree up to n and interpolates $f(x)$ at the points listed in (8.1).

EXAMPLE 8.1. Interpolate $f(x) = 1/x$ at the nodes $x_0 = 2$, $x_1 = 2.5$ and $x_2 = 4$ with the Lagrange interpolating polynomial of degree 2.

SOLUTION. The Lagrange basis, in nested form, is

$$\begin{aligned} L_0(x) &= \frac{(x-2.5)(x-4)}{(2-2.5)(2-4)} &&= (x-6.5)x+10, \\ L_1(x) &= \frac{(x-2)(x-4)}{(2.5-2)(2.5-4)} &&= \frac{(-4x+24)x-32}{3}, \\ L_2(x) &= \frac{(x-2)(x-2.5)}{(4-2)(4-2.5)} &&= \frac{(x-4.5)x+5}{3}. \end{aligned}$$

Thus,

$$\begin{aligned} p(x) &= \frac{1}{2}\left[(x-6.5)x+10\right] + \frac{1}{2.5}\,\frac{(-4x+24)x-32}{3} + \frac{1}{4}\,\frac{(x-4.5)x+5}{3} \\ &= (0.05x-0.425)x+1.15. \quad \square \end{aligned}$$

THEOREM 8.1. *Suppose $x_0, x_1, \ldots, x_n$ are $n+1$ distinct points in the interval $[a,b]$ and $f \in C^{n+1}[a,b]$. Then there exits a number $\xi(x) \in [a,b]$ such that*

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}\,(x-x_0)\,(x-x_1)\,\cdots\,(x-x_n), \tag{8.3}$$

where $p_n(x)$ is the Lagrange interpolating polynomial. In particular, if

$$m_{n+1} = \min_{a\le x\le b} |f^{(n+1)}(x)| \quad \textit{and} \quad M_{n+1} = \max_{a\le x\le b} |f^{(n+1)}(x)|,$$

then the absolute error in $p_n(x)$ is bounded by the inequalities:

$$\begin{aligned} \frac{m_{n+1}}{(n+1)!}\,|(x-x_0)\,(x-x_1)\,\cdots(x-x_n)| &\le |f(x)-p_n(x)| \\ &\le \frac{M_{n+1}}{(n+1)!}\,|(x-x_0)\,(x-x_1)\,\cdots\,(x-x_n)| \end{aligned}$$

for $a \le x \le b$.

PROOF. First, note that the error is 0 at $x = x_0, x_1, \ldots, x_n$ since

$$p_n(x_k) = f(x_k), \qquad k = 0, 1, \ldots, n,$$

from the interpolating property of $p_n(x)$. For $x \ne x_k$, define the auxiliary function

$$\begin{aligned} g(t) &= f(t) - p_n(t) - [f(x)-p_n(x)]\,\frac{(t-x_0)(t-x_1)\cdots(t-x_n)}{(x-x_0)(x-x_1)\cdots(x-x_n)} \\ &= f(t) - p_n(t) - [f(x)-p_n(x)]\prod_{i=0}^{n}\frac{t-x_i}{x-x_i}. \end{aligned}$$

For $t = x_k$,

$$g(x_k) = f(x_k) - p_n(x_k) - [f(x)-p_n(x)] \times 0 = 0$$

and for $t = x$,

$$g(x) = f(x) - p_n(x) - [f(x)-p_n(x)] \times 1 = 0.$$

Thus $g \in C^{n+1}[a,b]$ and it has $n+2$ zeros in $[a,b]$. By the generalized Rolle theorem, $g'(t)$ has $n+1$ zeros in $[a,b]$, $g''(t)$ has n zeros in $[a,b]$, $\ldots$, $g^{(n+1)}(t)$

has 1 zero, $\xi \in [a,b]$,

$$\begin{aligned} g^{(n+1)}(\xi) &= f^{(n+1)}(\xi) - p_n^{(n+1)}(\xi) - [f(x) - p_n(x)] \frac{d^{n+1}}{dt^{n+1}} \left[\prod_{i-0}^{n} \frac{t - x_i}{x - x_i} \right|_{t=\xi} \\ &= f^{(n+1)}(\xi) - 0 - [f(x) - p_n(x)] \frac{(n+1)!}{\prod_{i=0}^{n}(x - x_i)} \\ &= 0 \end{aligned}$$

since $p_n(x)$ is a polynomial of degree n so that its $(n+1)$st derivative is zero and only the top term, t^{n+1}, in the product $\prod_{i=0}^{n}(t - x_i)$ contributes to $(n+1)!$ in its $(n+1)$st derivative. Hence

$$f(x) = p_n(x) + \frac{f^{(n+1)}(\xi(x))}{(n+1)!} (x - x_0)(x - x_1) \cdots (x - x_n). \quad \square$$

From a computational point of view, (8.2) is not the best representation of $p_n(x)$ because it is computationally costly and has to be redone from scratch if we want to increase the degree of $p_n(x)$ to improve the interpolation.

EXAMPLE 8.2. Suppose we have the data points $(1.2, 3.1)$, $(1.7, 5.2)$ and $(2.1, 7.4)$. Use Lagrange's interpolating polynomial to estimate the value of the function at $x = 1.5$.

SOLUTION. Since we have 3 data points, we will be using

$$\begin{aligned} p_2(x) &= L_0(x) f_0 + L_1(x) f_1 + L_2(x) f_2 \\ &= \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_0)} f_1 + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} f_2. \end{aligned}$$

So,

$$\begin{aligned} p_2(1.5) &= \frac{(1.5 - 1.7)(1.5 - 2.1)}{(1.2 - 1.7)(1.2 - 2.1)} (3.1) + \frac{(1.5 - 1.2)(1.5 - 2.1)}{(1.7 - 1.2)(1.7 - 2.1)} (5.2) \\ &\quad + \frac{(1.5 - 1.2)(1.5 - 1.7)}{(2.1 - 1.2)(2.1 - 1.7)} (7.4) \\ &= \frac{0.12}{0.45} (3.1) + \frac{0.18}{0.20)} (5.2) + \frac{-0.06}{0.36} (7.4) \\ &= 4.2733, \end{aligned}$$

and so

$$f(1.5) \approx p_2(1.5) = 4.2733. \quad \square$$

8.2. Newton's Divided Difference Interpolating Polynomial

Newton's divided difference interpolating polynomials, $p_n(x)$, of degree n use a factorial basis in the form

$$p_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \cdots + a_n(x - x_0)(x - x_1) \cdots (x - x_{n-1}).$$

The values of the coefficients a_k are determined by recurrence. We denote

$$f_k = f(x_k).$$

Let $x_0 \neq x_1$ and consider the two data points: (x_0, f_0) and (x_1, f_1). Then the interpolating property of the polynomial

$$p_1(x) = a_0 + a_1(x - x_0)$$

implies that

$$p_1(x_0) = a_0 = f_0, \qquad p_1(x_1) = f_0 + a_1(x_1 - x_0) = f_1.$$

Solving for a_1 we have

$$a_1 = \frac{f_1 - f_0}{x_1 - x_0}.$$

If we let

$$f[x_0, x_1] = \frac{f_1 - f_0}{x_1 - x_0}$$

be the *first divided difference*, then the divided difference interpolating polynomial of degree one is

$$p_1(x) = f_0 + (x - x_0)\, f[x_0, x_1].$$

EXAMPLE 8.3. Consider a function $f(x)$ which passes through the points $(2.2, 6.2)$ and $(2.5, 6.7)$. Find the divided difference interpolating polynomial of degree one for $f(x)$ and use it to interpolate f at $x = 2.35$.

SOLUTION. Since

$$f[2.2, 2.5] = \frac{6.7 - 6.2}{2.5 - 2.2} = 1.6667,$$

then

$$p_1(x) = 6.2 + (x - 2.2) \times 1.6667 = 2.5333 + 1.6667\, x.$$

In particular, $p_1(2.35) = 6.45$. □

EXAMPLE 8.4. Approximate $\cos 0.2$ linearly using the values of $\cos 0$ and $\cos \pi/8$.

SOLUTION. We have the points

$$(0, \cos 0) = (0, 1) \quad \text{and} \quad \left(\frac{\pi}{8}, \cos\frac{\pi}{8}\right) = \left(\frac{\pi}{8}, \frac{1}{2}\sqrt{\sqrt{2} + 2}\right)$$

(Substitute $\theta = \pi/8$ into the formula

$$\cos^2\theta = \frac{1 + \cos(2\,\theta)}{2}$$

to get

$$\cos\frac{\pi}{8} = \frac{1}{2}\sqrt{\sqrt{2} + 2}$$

since $\cos(\pi/4) = \sqrt{2}/2$.) Thus

$$f[0, \pi/8] = \frac{\left(\sqrt{\sqrt{2} + 2}\right)/2 - 1}{\pi/8 - 0} = \frac{4}{\pi}\left(\sqrt{\sqrt{2} + 2} - 2\right).$$

This leads to

$$p_1(x) = 1 + \frac{4}{\pi}\left(\sqrt{\sqrt{2} + 2} - 2\right) x.$$

In particular,

$$p_1(0.2) = 0.96125.$$

Note that $\cos 0.2 = 0.98007$ (rounded to five digits). The absolute error is 0.01882. □

Consider the three data points

$$(x_0, f_0), \quad (x_1, f_1), \quad (x_2, f_2), \qquad \text{where} \quad x_i \neq x_j \quad \text{for} \quad i \neq j.$$

Then the divided difference interpolating polynomial of degree two through these points is

$$p_2(x) = f_0 + (x - x_0)\, f[x_0, x_1] + (x - x_0)\,(x - x_1)\, f[x_0, x_1, x_2]$$

where

$$f[x_0, x_1] := \frac{f_1 - f_0}{x_1 - x_0} \quad \text{and} \quad f[x_0, x_1, x_2] := \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$$

are the *first* and *second* divided differences, respectively.

EXAMPLE 8.5. Interpolate a given function $f(x)$ through the three points

$$(2.2, 6.2), \qquad (2.5, 6.7), \qquad (2.7, 6.5),$$

by means the divided difference interpolating polynomial of degree two, $p_2(x)$, and interpolate $f(x)$ at $x = 2.35$ by means of $p_2(2.35)$.

SOLUTION. We have

$$f[2.2, 2.5] = 1.6667, \qquad f[2.5, 2.7] = -1$$

and

$$f[2.2, 2.5, 2.7] = \frac{f[2.5, 2.7] - f[2.2, 2.5]}{2.7 - 2.2} = \frac{-1 - 1.6667}{2.7 - 2.2} = -5.3334.$$

Therefore,

$$p_2(x) = 6.2 + (x - 2.2) \times 1.6667 + (x - 2.2)\,(x - 2.5) \times (-5.3334).$$

In particular, $p_2(2.35) = 6.57$. □

EXAMPLE 8.6. Construct the divided difference interpolating polynomial of degree two for $\cos x$ using the values $\cos 0$, $\cos \pi/8$ and $\cos \pi/4$, and approximate $\cos 0.2$.

SOLUTION. It was seen in Example 8.4 that

$$\cos \frac{\pi}{8} = \frac{1}{2}\sqrt{\sqrt{2} + 2}.$$

Hence, from the three data points

$$(0, 1), \qquad \left(\frac{\pi}{8}, \frac{1}{2}\sqrt{\sqrt{2} + 2}\right), \qquad \left(\frac{\pi}{4}, \frac{\sqrt{2}}{2}\right),$$

we obtain the divided differences

$$f[0, \pi/8] = \frac{4}{\pi}\left(\sqrt{\sqrt{2} + 2} - 2\right), \qquad f[\pi/8, \pi/4] = \frac{4}{\pi}\left(\sqrt{2} - \sqrt{\sqrt{2} + 2}\right),$$

TABLE 8.1. Ddivided difference table

x	$f(x)$	First divided differences	Second divided differences	Third divided differences
x_0	$f[x_0]$			
		$f[x_0, x_1]$		
x_1	$f[x_1]$		$f[x_0, x_1, x_2]$	
		$f[x_1, x_2]$		$f[x_0, x_1, x_2, x_3]$
x_2	$f[x_2]$		$f[x_1, x_2, x_3]$	
		$f[x_2, x_3]$		$f[x_1, x_2, x_3, x_4]$
x_3	$f[x_3]$		$f[x_2, x_3, x_4]$	
		$f[x_3, x_4]$		$f[x_2, x_3, x_4, x_5]$
x_4	$f[x_4]$		$f[x_3, x_4, x_5]$	
		$f[x_4, x_5]$		
x_5	$f[x_5]$			

and

$$\begin{aligned} f[0, \pi/8, \pi/4] &= \frac{f[\pi/8, \pi/4] - f[0, \pi/8]}{\pi/4 - 0} \\ &= \frac{4}{\pi}\left[\frac{\sqrt{2}/2 - (\sqrt{\sqrt{2}+2})/2}{\pi/4 - \pi/8} - \frac{4\sqrt{\sqrt{2}+2} - 8}{\pi}\right] \\ &= \frac{16}{\pi^2}\left(\sqrt{2} - 2\sqrt{\sqrt{2}+2}\right). \end{aligned}$$

Hence,

$$p_2(x) = 1 + x\,\frac{4}{\pi}\left(\sqrt{\sqrt{2}+2} - 2\right) + x\left(x - \frac{\pi}{8}\right)\frac{16}{\pi^2}\left(\sqrt{2} - 2\sqrt{\sqrt{2}+2}\right).$$

Evaluating this polynomial at $x = 0.2$, we obtain

$$p_2(0.2) = 0.97881.$$

The absolute error is 0.00189. □

In general, given $n+1$ data points

$$(x_0, f_0), \quad (x_1, f_1), \quad \ldots, \quad (x_n, f_n),$$

where $x_i \neq x_j$ for $i \neq j$, Newton's divided difference interpolating polynomial of degree n is

$$\begin{aligned} p_n(x) = f_0 + (x - x_0)\, f[x_0, x_1] + (x - x_0)\,(x - x_1)\, f[x_0, x_1, x_2] + \cdots \\ + (x - x_0)\,(x - x_1)\cdots(x - x_{n-1})\, f[x_0, x_1, \ldots, x_n], \end{aligned} \quad (8.4)$$

where, by definition,

$$f[x_j, x_{j+1}, \ldots, x_k] = \frac{f[x_{j+1}, \ldots, x_k] - f[x_j, x_{j+1}, \ldots, x_{k-1}]}{x_k - x_j}$$

is a $(k-j)$*th divided difference*. This formula can be obtained by recurrence.

A divided difference table is shown in Table 8.1.

EXAMPLE 8.7. Construct the cubic interpolating polynomial through the four unequally spaced points

$$(1.0, 2.4), \quad (1.3, 2.2), \quad (1.5, 2.3), \quad (1.7, 2.4),$$

on the graph of a certain function $f(x)$ and approximate $f(1.4)$.

SOLUTION. Newton's divided difference table is

x_i	$f(x_i)$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$
1.0	**2.4**			
		−0.66667		
1.3	2.2		**2.33333**	
		0.500000		**−3.33333**
1.5	2.3		0.00000	
		0.500000		
1.7	2.4			

Therefore,

$$p_3(x) = 2.4 + (x - 1.0)\,(-0.66667) + (x - 1.0)\,(x - 1.3) \times 2.33333 + (x - 1.0)\,(x - 1.3)\,(x - 1.5)\,(-3.33333).$$

The approximation to $f(1.4)$ is

$$p_3(1.4) = 2.2400. \quad \square$$

Since the interpolating polynomial, $p_n(x)$, is unique, it does not matter how we find it. We can see from the examples that Newton's divided difference is a more efficient way of calculating $p_n(x)$ than Lagrange's idea, but the two methods will give the same polynomial (up to round-off errors in the coefficients).

8.3. Gregory–Newton Forward-Difference Polynomial

We rewrite (8.4) in the special case where the nodes x_i are equidistant,

$$x_i = x_0 + i\,h.$$

The *first* and *second forward differences* of $f(x)$ at x_j are

$$\Delta f_j := f_{j+1} - f_j, \qquad \Delta^2 f_j := \Delta f_{j+1} - \Delta f_j,$$

respectively, and in general, the *kth forward difference* of $f(x)$ at x_j is

$$\Delta^k f_j := \Delta^{k-1} f_{j+1} - \Delta^{k-1} f_j.$$

It is seen by mathematical induction that

$$f[x_0, \ldots, x_k] := \frac{1}{k!\,h^k}\,\Delta^k f_0.$$

If we set

$$r = \frac{x - x_0}{h},$$

then, for equidistant nodes,

$$x - x_k = x - x_0 - (x_k - x_0) = hr - hk = h(r - k)$$

and

$$(x-x_0)(x-x_1)\cdots(x-x_{k-1}) = h^k r(r-1)(r-2)\cdots(r-k+1).$$

Thus (8.4) becomes

$$\begin{aligned} p_n(r) &= f_0 + \sum_{k=1}^{n} \frac{r\,(r-1)\,\cdots\,(r-k+1)}{k!}\,\Delta^k f_0 \\ &= \sum_{k=0}^{n} \binom{r}{k} \Delta^k f_0, \end{aligned} \tag{8.5}$$

where

$$\binom{r}{k} = \begin{cases} \dfrac{r\,(r-1)\,\cdots\,(r-k+1)}{k!}, & \text{if} \quad k>0, \\ 1, & \text{if} \quad k=0. \end{cases}$$

Polynomial (8.5) is the *Gregory–Newton forward-difference interpolating polynomial*.

EXAMPLE 8.8. Suppose that we are given the following equally spaced data:

x	1988	1989	1990	1991	1992	1993
y	35000	36000	36500	37000	37800	39000

Extrapolate the value of y in year 1994.

SOLUTION. The forward difference table is

i	x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	$\Delta^5 y_i$
0	1988	**35000**					
			1000				
1	1989	36000		**−500**			
			500		**500**		
2	1990	36500		0		**−200**	
			500		300		**0**
3	1991	37000		300		−200	
			800		100		
4	1992	37800		400			
			1200				
5	1993	39000					

Setting $r = (x-1988)/1$, we have

$$\begin{aligned} p_5(r) = 35000 &+ r\,(1000) + \frac{r(r-1)}{2}\,(-500) + \frac{r(r-1)(r-2)}{6}\,(500) \\ &+ \frac{r(r-1)(r-2)(r-3)}{24}\,(-200) + \frac{r(r-1)(r-2)(r-3)(r-4)}{120}\,(0). \end{aligned}$$

Extrapolating the data at 1994 we have $r = 6$ and

$$p_5(6) = 40500.$$

An iterative use of the Matlab `diff(y,n)` command produces a difference table.

```
y = [35000 36000 36500 37000 37800 39000]
dy = diff(y);
  dy = 1000          500          500          800         1200
d2y = diff(y,2)
  d2y = -500     0   300   400
d3y = diff(y,3)
  d3y = 500   300   100
d4y = diff(y,4)
  d4y = -200  -200
d5y = diff(y,5)
  d5y = 0
```

□

EXAMPLE 8.9. Use the following equally spaced data to approximate $f(1.5)$.

x	1.0	1.3	1.6	1.9	2.2
$f(x)$	0.7651977	0.6200860	0.4554022	0.2818186	0.1103623

SOLUTION. The forward difference table is

i	x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
0	1.0	0.7651977				
			−0.145112			
1	1.3	0.6200860		−0.0195721		
			-0.164684		0.0106723	
2	1.6	0.4554022		-0.0088998		0.0003548
			-0.173584		0.0110271	
3	1.9	0.2818186		0.0021273		
			-0.170856			
4	2.2	0.1103623				

Setting $r = (x - 1.0)/0.3$, we have

$$p_4(r) = 0.7651977 + r\,(-0.145112) + \frac{r(r-1)}{2}\,(-0.0195721) + \frac{r(r-1)(r-2)}{6}\,(0.0106723) + \frac{r(r-1)(r-2)(r-3)}{24}\,(0.0003548).$$

Interpolating $f(x)$ at $x = 1$, we have $r = 5/3$ and and

$$p_4(5/3) = 0.511819. \quad \square$$

8.4. Gregory–Newton Backward-Difference Polynomial

To interpolate near the bottom of a difference table with equidistant nodes, one uses the Gregory–Newton backward-difference interpolating polynomial for the data

$$(x_{-n}, f_{-n}), \quad (x_{-n+1}, f_{-n+1}), \quad \ldots, \quad (x_0, f_0).$$

If we set

$$r = \frac{x - x_0}{h},$$

then, for equidistant nodes,

$$x - x_{-k} = x - x_0 - (x_{-k} - x_0) = hr + hk = h(r + k)$$

and

$$(x - x_0)(x - x_{-1}) \cdots (x - x_{-(k-1)}) = h^k r(r + 1)(r + 2) \cdots (r + k - 1).$$

Thus (8.5) becomes

$$\begin{aligned} p_n(r) &= f_0 + \sum_{k=1}^{n} \frac{r\,(r+1)\cdots(r+k-1)}{k!}\,\Delta^k f_{-k} \\ &= \sum_{k=0}^{n} \binom{r+k-1}{k} \Delta^k f_{-k}, \end{aligned} \tag{8.6}$$

The polynomial (8.6) is the *Gregory–Newton backward-difference interpolating polynomial*.

EXAMPLE 8.10. Interpolate the equally spaced data of Example 8.9 at $x = 2.1$

SOLUTION. The difference table is

i	x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
−4	1.0	0.7651977				
			-0.145112			
−3	1.3	0.6200860		-0.0195721		
			-0.164684		0.0106723	
−2	1.6	0.4554022		-0.0088998		0.0003548
			-0.173584		0.0110271	
−1	1.9	0.2818186		0.0021273		
			−0.170856			
0	2.2	0.1103623				

Setting $r = (x - 2.2)/0.3$, we have

$$\begin{aligned} p_4(r) = 0.1103623 &+ r\,(-0.170856) + \frac{r(r+1)}{2}\,(0.0021273) \\ &+ \frac{r(r+1)(r+2)}{6}\,(0.0110271) + \frac{r(r+1)(r+2)(r+3)}{24}\,(0.0003548). \end{aligned}$$

Since

$$r = \frac{2.1 - 2.2}{0.3} = -\frac{1}{3},$$

then

$$p_4(-1/3) = 0.115904. \quad \square$$

8.5. Hermite Interpolating Polynomial

Given $n + 1$ distinct nodes x_0, x_1,...,x_n and $2n + 2$ values $f_k = f(x_k)$ and $f'_k = f'(x_k)$, the Hermite interpolating polynomial $p_{2n+1}(x)$ of degree $2n + 1$,

$$p_{2n+1}(x) = \sum_{m=0}^{n} h_m(x) f_m + \sum_{m=0}^{n} \widehat{h}_m(x) f'_m,$$

takes the values

$$p_{2n+1}(x_k) = f_k, \qquad p'_{2n+1}(x_k) = f'_k, \qquad k = 0, 1, \ldots, n.$$

We look for polynomials $h_m(x)$ and $\widehat{h}_m(x)$ of degree at most $2n+1$ satisfying the following conditions:

$$\begin{aligned} h_m(x_k) = h'_m(x_k) &= 0, \qquad k \neq m, \\ h_m(x_m) &= 1, \\ h'_m(x_m) &= 0, \end{aligned}$$

and

$$\begin{aligned} \widehat{h}_m(x_k) = \widehat{h}'_m(x_k) &= 0, \qquad k \neq m, \\ \widehat{h}_m(x_m) &= 0, \\ \widehat{h}'_m(x_m) &= 1. \end{aligned}$$

These conditions are satisfied by the polynomials

$$h_m(x) = [1 - 2(x - x_m)L'_m(x_m)]L^2_m(x)$$

and

$$\widehat{h}_m(x) = (x - x_m)L^2_m(x),$$

where

$$L_m(x) = \prod_{k=0, k\neq m}^{n} \frac{x - x_k}{x_m - x_k}$$

are the elements of the Lagrange basis of degree n.

A practical method of constructing a Hermite interpolating polynomial over the $n+1$ distinct nodes x_0, x_1,...,x_n is to set

$$z_{2i} = z_{2i+1} = x_i, \qquad i = 0, 1, \ldots, n,$$

and take

$$f'(x_0) \text{ for } f[z_0, z_1], \quad f'(x_1) \text{ for } f[z_2, z_3], \quad \ldots, \quad f'(x_j) \text{ for } f[z_{2n} z_{2n+1}]$$

in the divided difference table for the Hermite interpolating polynomial of degree $2n+1$. Thus,

$$p_{2n+1}(x) = f[z_0] + \sum_{k=1}^{2n+1} f[z_0, z_1, \ldots, z_k](x - z_0)(x - z_1)\cdots(x - z_{k-1}).$$

A divided difference table for a Hermite interpolating polynomial is as follows.

z	$f(z)$	First divided differences	Second divided differences	Third divided differences
$z_0 = x_0$	$f[z_0] = f(x_0)$			
		$f[z_0, z_1] = f'(x_0)$		
$z_1 = x_0$	$f[z_1] = f(x_0)$		$f[z_0, z_1, z_2]$	
		$f[z_1, z_2]$		$f[z_0, z_1, z_2, z_3]$
$z_2 = x_1$	$f[z_2] = f(x_1)$		$f[z_1, z_2, z_3]$	
		$f[z_2, z_3] = f'(x_1)$		$f[z_1, z_2, z_3, z_4]$
$z_3 = x_1$	$f[z_3] = f(x_1)$		$f[z_2, z_3, z_4]$	
		$f[z_3, z_4]$		$f[z_2, z_3, z_4, z_5]$
$z_4 = x_2$	$f[z_4] = f(x_2)$		$f[z_3, z_4, z_5]$	
		$f[z_4, z_5] = f'(x_2)$		
$z_5 = x_2$	$f[z_5] = f(x_2)$			

EXAMPLE 8.11. Interpolate the underlined data, given in the table below, at $x = 1.5$ by a Hermite interpolating polynomial of degree five.

SOLUTION. In the difference table the underlined entries are the given data. The remaining entries are generated by standard divided differences.

1.3	0.6200860					
		-0.5220232				
1.3	0.6200860		-0.0897427			
		-0.5489460		0.0663657		
1.6	0.4554022		-0.0698330		0.0026663	
		-0.5698959		0.0679655		-0.0027738
1.6	0.4554022		-0.0290537		0.0010020	
		-0.5786120		0.0685667		
1.9	0.2818186		-0.0084837			
		-0.5811571				
1.9	0.2818186					

Taking the elements along the top downward diagonal, we have

$$\begin{aligned} P(1.5) &= 0.6200860 + (1.5 - 1.3)(-0.5220232) + (1.5 - 1.3)^2(-0.0897427) \\ &\quad + (1.5 - 1.3)^2(1.5 - 1.6)(0.0663657) + (1.5 - 1.3)^2(1.5 - 1.6)^2(0.0026663) \\ &\quad + (1.5 - 1.3)^2(1.5 - 1.6)^2(1.5 - 1.9)(-0.0027738) \\ &= 0.5118277. \quad \square \end{aligned}$$

8.6. Cubic Spline Interpolation

In this section, we interpolate functions by piecewise cubic polynomials which satisfy some global smoothness conditions. Piecewise polynomials avoid the oscillatory nature of high-degree polynomials over a large interval as a polynomial of degree n will habve up to $n - 1$ local extrema or turning points.

DEFINITION 8.1. Given a function $f(x)$ defined on the interval $[a, b]$ and a set of nodes

$$a = x_0 < x_1 < \cdots < x_n = b,$$

a cubic spline interpolant S for f is a piecewise cubic polynomial that satisfies the following conditions:

(a) $S(x)$ is a cubic polynomial, denoted $S_j(x)$, on the subinterval $[x_j, x_{j+1}]$ for each $j = 0, 1, \ldots, n-1$;
(b) $S(x_j) = f(x_j)$ for each $j = 0, 1, \ldots, n$;
(c) $S_{j+1}(x_{j+1}) = S_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;
(d) $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;
(e) $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;
(f) One of the sets of boundary conditions is satisfied:
  (i) $S''(x_0) = S''(x_n) = 0$ (free or natural boundary);
  (ii) $S'(x_0) = f'(x_0)$ and $S'(x_n) = f'(x_n)$ (clamped boundary).

Other boundary conditions can be used in the definition of splines. When free or clamped boundary conditions occur, the spline is called a *natural spline* or a *clamped spline*, respectively.

To construct the cubic spline interpolant for a given function f, the conditions in the definition are applied to the cubic polynomials

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3,$$

for each $j = 0, 1, \ldots, n-1$.

The following existence and uniqueness theorems hold for natural and clamped spline interpolants, respectively.

THEOREM 8.2 (Natural Spline). *If f is defined at $a = x_0 < x_1 < \cdots < x_n = b$, then f has a unique natural spline interpolant S on the nodes $x_0, x_1, \ldots, x_n$ with boundary conditions $S''(a) = 0$ and $S''(b) = 0$.*

THEOREM 8.3 (Clamped Spline). *If f is defined at $a = x_0 < x_1 < \cdots < x_n = b$ and is differentiable at a and b, then f has a unique clamped spline interpolant S on the nodes $x_0, x_1, \ldots, x_n$ with boundary conditions $S'(a) = f'(a)$ and $S'(b) = f'(b)$.*

The following Matlab commands generate a sine curve and sample the spline over a finer mesh:

```
x = 0:10;  y = sin(x);
xx = 0:0.25:10;
yy = spline(x,y,xx);
subplot(2,2,1); plot(x,y,'o',xx,yy);
```

The result is shown in Fig 8.1.

The following Matlab commands illustrate the use of clamped spline interpolation where the end slopes are prescribed. Zero slopes at the ends of an interpolant to the values of a certain distribution are enforced:

```
x = -4:4; y = [0 .15 1.12 2.36 2.36 1.46 .49 .06 0];
cs = spline(x,[0 y 0]);
xx = linspace(-4,4,101);
plot(x,y,'o',xx,ppval(cs,xx),'-');
```

The result is shown in Fig 8.2.

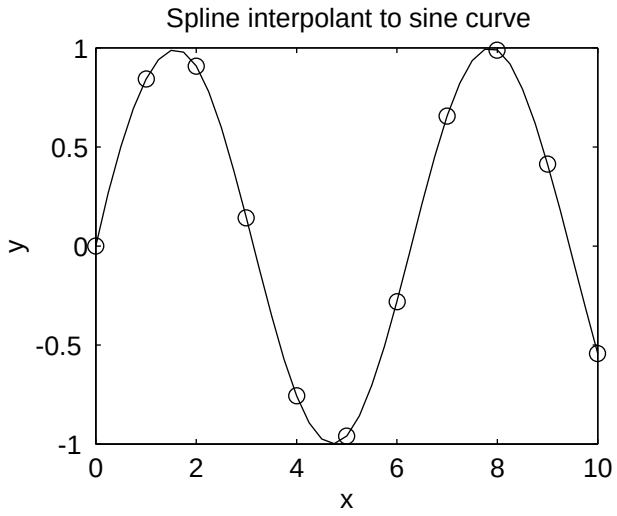

FIGURE 8.1. Spline interpolant of sine curve.

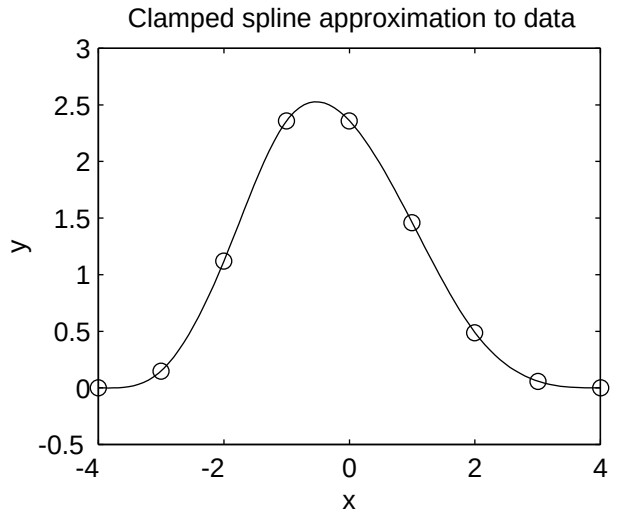

FIGURE 8.2. Clamped spline approximation to data.

CHAPTER 9

Numerical Differentiation and Integration

9.1. Numerical Differentiation

9.1.1. Two-point formula for $f'(x)$. The Lagrange interpolating polynomial of degree 1 for $f(x)$ at x_0 and $x_1 = x_0 + h$ is

$$f(x) = f(x_0)\,\frac{x - x_1}{-h} + f(x_1)\,\frac{x - x_0}{h} + \frac{(x - x_0)(x - x_1)}{2!}\,f''(\xi(x)), \qquad x_0 < \xi(x) < x_0 + h.$$

Differentiating this polynomial, we have

$$f'(x) = f(x_0)\,\frac{1}{-h} + f(x_1)\,\frac{1}{h} + \frac{(x - x_1) + (x - x_0)}{2!}\,f''(\xi(x)) + \frac{(x - x_0)(x - x_1)}{2!}\,\frac{d}{dx}\big[f''(\xi(x))\big].$$

Putting $x = x_0$ in $f'(x)$, we obtain the first-order two-point formula

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0)}{h} - \frac{h}{2}\,f''(\xi). \tag{9.1}$$

If $h > 0$, this is a forward difference formula and, if $h < 0$, this is a backward difference formula.

9.1.2. Three-point formula for $f'(x)$. The Lagrange interpolating polynomial of degree 2 for $f(x)$ at x_0, $x_1 = x_0 + h$ and $x_2 = x_0 + 2h$ is

$$f(x) = f(x_0)\,\frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f(x_1)\,\frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + f(x_2)\,\frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} + \frac{(x - x_0)(x - x_1)(x - x_2)}{3!}\,f'''(\xi(x)),$$

where $x_0 < \xi(x) < x_2$. Differentiating this polynomial and substituting $x = x_j$, we have

$$f'(x_j) = f(x_0)\,\frac{2x_j - x_1 - x_2}{(x_0 - x_1)(x_0 - x_2)} + f(x_1)\,\frac{2x_j - x_0 - x_2}{(x_1 - x_0)(x_1 - x_2)} + f(x_2)\,\frac{2x_j - x_0 - x_1}{(x_2 - x_0)(x_2 - x_1)} + \frac{1}{6}f'''(\xi(x_j))\prod_{k=0,k\neq j}^{2}(x_j - x_k).$$

With $j = 0, 1, 2$, $f'(x_j)$ gives three second-order three-point formulae:

$$\begin{aligned} f'(x_0) &= f(x_0)\,\frac{-3h}{2h^2} + f(x_1)\,\frac{-2h}{-h^2} + f(x_2)\,\frac{-h}{2h^2} + \frac{2h^2}{6}f'''(\xi_0) \\ &= \frac{1}{h}\left[-\frac{3}{2}\,f(x_0) + 2f(x_1) - \frac{1}{2}\,f(x_2)\right] + \frac{h^2}{3}f'''(\xi_0), \end{aligned}$$

$$\begin{aligned} f'(x_1) &= f(x_0)\,\frac{-h}{2h^2} + f(x_1)\,\frac{0}{-h^2} + f(x_2)\,\frac{h}{2h^2} - \frac{h^2}{6}f'''(\xi_1) \\ &= \frac{1}{h}\left[-\frac{1}{2}\,f(x_0) + \frac{1}{2}\,f(x_2)\right] - \frac{h^2}{6}f'''(\xi_1), \end{aligned}$$

and, similarly,

$$f'(x_2) = \frac{1}{h}\left[\frac{1}{2}\,f(x_0) - 2f(x_1) + \frac{3}{2}\,f(x_2)\right] + \frac{h^2}{3}f'''(\xi_2).$$

These three-point formulae are usually written at x_0:

$$f'(x_0) = \frac{1}{2h}\,[-3f(x_0) + 4f(x_0 + h) - f(x_0 + 2h)] + \frac{h^2}{3}f'''(\xi_0), \tag{9.2}$$

$$f'(x_0) = \frac{1}{2h}\,[f(x_0 + h) - f(x_0 - h)] - \frac{h^2}{6}f'''(\xi_1). \tag{9.3}$$

The third formula is obtained from (9.2) by replacing h with $-h$. It is to be noted that the centred formula (9.3) is more precise than (9.2) since its error coefficient is half the error coefficient of the other formula.

EXAMPLE 9.1. Consider the data points $(0, 1)$, $(0.25, 0.97)$ and $(0.5, 0.88)$. Estimate the derivative of the function at the three points.

SOLUTION. Clearly, $h = \Delta x = 0.25$. So

$$\begin{aligned} f'(0) &\approx \frac{1}{h}\left[-\frac{3}{2}\,f(0) + 2f(0.25) - \frac{1}{2}\,f(0.5)\right] \\ &= \frac{1}{0.25}\left[-\frac{3}{2}\,(1) + 2(0.97) - \frac{1}{2}\,(0.88)\right] \\ &= 0, \\ f'(0.25) &\approx \frac{1}{h}\left[-\frac{1}{2}\,f(0) + \frac{1}{2}\,f(0.5)\right] \\ &= \frac{1}{0.25}\left[-\frac{1}{2}\,(1) + \frac{1}{2}\,(0.88)\right] \\ &= -0.24, \\ f'(0.5) &\approx \frac{1}{h}\left[\frac{1}{2}\,f(0) - 2f(0.25) + \frac{3}{2}\,f(0.5)\right] \\ &= \frac{1}{0.25}\left[\frac{1}{2}\,(1) - 2(0.97) + \frac{3}{2}\,(0.88)\right] \\ &= -0.48. \end{aligned}$$

The true function is $f(x) = \cos x$, so $f'(x) = -\sin x$ and the true values are

$$f'(0) = -\sin 0 = 0,$$
$$f'(0.25) = -\sin 0.25 = -0.2474,$$
$$f'(0.5) = \sin 0.5 = -0.4794.$$

It is seen that the aproximations are actually quite good. □

9.1.3. Three-point centered difference formula for $f''(x)$. We use truncated Taylor's expansions for $f(x+h)$ and $f(x-h)$:

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{6}f'''(x_0)h^3 + \frac{1}{24}f^{(4)}(\xi_0)h^4,$$
$$f(x_0-h) = f(x_0) - f'(x_0)h + \frac{1}{2}f''(x_0)h^2 - \frac{1}{6}f'''(x_0)h^3 + \frac{1}{24}f^{(4)}(\xi_1)h^4.$$

Adding these expansions, we have

$$f(x_0+h) + f(x_0-h) = 2f(x_0) + f''(x_0)h^2 + \frac{1}{24}\left[f^{(4)}(\xi_0) + f^{(4)}(\xi_1)\right]h^4.$$

Solving for $f''(x_0)$, we have

$$f''(x_0) = \frac{1}{h^2}\left[f(x_0-h) - 2f(x_0) + f(x_0+h)\right] - \frac{1}{24}\left[f^{(4)}(\xi_0) + f^{(4)}(\xi_1)\right]h^2.$$

By the Mean Value Theorem 7.5 for sums, there is a value ξ, $x_0 - h < \xi < x_0 + h$, such that

$$\frac{1}{2}\left[f^{(4)}(\xi_0) + f^{(4)}(\xi_1)\right] = f^{(4)}(\xi).$$

We thus obtain the three-point second-order centered difference formula

$$f''(x_0) = \frac{1}{h^2}\left[f(x_0-h) - 2f(x_0) + f(x_0+h)\right] - \frac{h^2}{12}f^{(4)}(\xi). \tag{9.4}$$

9.2. The Effect of Roundoff and Truncation Errors

The presence of the stepsize h in the denominator of numerical differentiation formulae may produce large errors due to roundoff errors. We consider the case of the two-point centred formula (9.3) for $f'(x)$. Other cases are treated similarly.

Suppose that the roundoff error in the evaluated value $\widetilde{f}(x_j)$ for $f(x_j)$ is $e(x_j)$. Thus,

$$f(x_0+h) = \widetilde{f}(x_0+h) + e(x_0+h), \qquad f(x_0-h) = \widetilde{f}(x_0-h) + e(x_0-h).$$

Subtituting these values in (9.3), we have the total error, which is the sum of the roundoff and the truncation errors,

$$f'(x_0) - \frac{\widetilde{f}(x_0+h) - \widetilde{f}(x_0-h)}{2h} = \frac{e(x_0+h) - e(x_0-h)}{2h} - \frac{h^2}{6}f^{(3)}(\xi).$$

Taking the absolute value of the right-hand side and applying the triangle inequality, we have

$$\left|\frac{e(x_0+h) - e(x_0-h)}{2h} - \frac{h^2}{6}f^{(3)}(\xi)\right| \le \frac{1}{2h}\left(|e(x_0+h)| + |e(x_0-h)|\right) + \frac{h^2}{6}|f^{(3)}(\xi)|.$$

If

$$|e(x_0 \pm h)| \le \varepsilon, \qquad |f^{(3)}(x)| \le M,$$

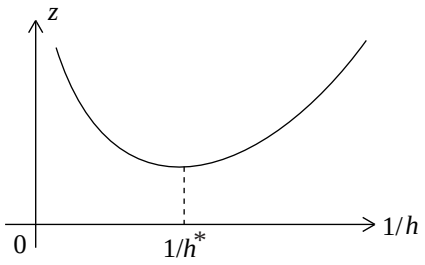

FIGURE 9.1. Truncation and roundoff error curve as a function of $1/h$.

then

$$\left| f'(x_0) - \frac{\widetilde{f}(x_0+h) - \widetilde{f}(x_0-h)}{2h} \right| \leq \frac{\varepsilon}{h} + \frac{h^2}{6}\, M.$$

We remark that the expression

$$z(h) = \frac{\varepsilon}{h} + \frac{h^2}{6}\, M$$

first decreases and afterwards increases as $1/h$ increases, as shown in Fig. 9.1. The term $Mh^2/6$ is due to the trunctation error and the term ε/h is due to roundoff errors.

EXAMPLE 9.2. (a) Given the function $f(x)$ and its first derivative $f'(x)$:

$$f(x) = \cos x, \qquad f'(x) = -\sin x,$$

approxminate $f'(0.7)$ with $h = 0.1$ by the five-point formula, without the truncation error term,

$$f'(x) = \frac{1}{12h}\left[-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)\right] + \frac{h^4}{30}\, f^{(5)}(\xi),$$

where ξ, in the truncaction error, satisfies the inequalities $x - 2h \leq \xi \leq x + 2h$.
(b) Given that the roundoff error in each evaluation of $f(x)$ is bounded by $\epsilon = 5 \times 10^{-7}$, find a bound for the total error in $f'(0.7)$ by adding bounds for the roundoff and the truncation errors).
(c) Finally, find the value of h that minimizes the total error.

SOLUTION. (a) A simple computation with the given formula, without the truncation error, gives the approximation

$$f'(0.7) \approx -0.644\,215\,542.$$

(b) Since

$$f^{(5)}(x) = -\sin x$$

is negative and decreasing on the interval $0.5 \leq x \leq 0.9$, then

$$M = \max_{0.5 \leq x \leq 0.9} |-\sin x| = \sin 0.9 = 0.7833.$$

Hence, a bound for the total error is

$$\begin{aligned}\text{Total error} &\leq \frac{1}{12\times 0.1}(1+8+8+1)\times 5\times 10^{-7}+\frac{(0.1)^4}{30}\times 0.7833\\ &= 7.5000\times 10^{-6}+2.6111\times 10^{-6}\\ &= 1.0111\times 10^{-5}.\end{aligned}$$

(c) The minimum of the total error, as a function of h,

$$\frac{90\times 10^{-7}}{12h}+\frac{0.7833}{30}\,h^4,$$

will be attained at a zero of its derivative with respect to h, that is,

$$\frac{d}{dh}\left(\frac{90\times 10^{-7}}{12h}+\frac{0.7833}{30}\,h^4\right)=0.$$

Performing the derivative and multiplying both sides by h^2, we obtain a quintic equation for h:

$$-7.5\times 10^{-7}+\frac{4\times 0.7833}{30}\,h^5=0.$$

Hence,

$$\begin{aligned}h &= \left(\frac{7.5\times 10^{-7}\times 30}{4\times 0.7833}\right)^{1/5}\\ &= 0.0936\end{aligned}$$

minimizes the total error. □

9.3. Richardson's Extrapolation

Suppose it is known that a numerical formula, $N(h)$, approximates an exact value M with an error in the form of a series in h^j,

$$M=N(h)+K_1h+K_2h^2+K_3h^3+\ldots,$$

where the constants K_j are independent of h. Then computing $N(h/2)$, we have

$$M=N\left(\frac{h}{2}\right)+\frac{1}{2}\,K_1h+\frac{1}{4}\,K_2h^2+\frac{1}{8}\,K_3h^3+\ldots.$$

Subtracting the first expression from twice the second, we eliminate the error in h:

$$M=N\left(\frac{h}{2}\right)+\left[N\left(\frac{h}{2}\right)-N(h)\right]+\left(\frac{h^2}{2}-h^2\right)\,K_2+\left(\frac{h^3}{4}-h^3\right)\,K_3+\ldots.$$

If we put

$$\begin{aligned}N_1(h) &= N(h),\\ N_2(h) &= N_1\left(\frac{h}{2}\right)+\left[N_1\left(\frac{h}{2}\right)-N_1(h)\right],\end{aligned}$$

the last expression for M becomes

$$M=N_2(h)-\frac{1}{2}\,K_2h^2-\frac{3}{4}\,K_3h^3-\ldots.$$

Now with $h/4$, we have

$$M = N_2\left(\frac{h}{2}\right) - \frac{1}{8}K_2h^2 - \frac{3}{32}K_3h^3 + \dots.$$

Subtracting the second last expression for M from 4 times the last one and dividing the result by 3, we elininate the term in h^2:

$$M = \left[N_2\left(\frac{h}{2}\right) + \frac{N_2(h/2) - N_2(h)}{3}\right] + \frac{1}{8}K_3h^3 + \dots.$$

Now, putting

$$N_3(h) = N_2\left(\frac{h}{2}\right) + \frac{N_2(h/2) - N_2(h)}{3},$$

we have

$$M = N_3(h) + \frac{1}{8}K_3h^3 + \dots.$$

The presence of the number $2^{j-1} - 1$ in the denominator of the second term of $N_j(h)$ ensures convergence. It is clear how to continue this process which is called *Richardson's extrapolation*.

An important case of Richardson's extrapolation is when $N(h)$ is the centred difference formula (9.3) for $f'(x)$, that is,

$$f'(x_0) = \frac{1}{2h}\big[f(x_0 + h) - f(x_0 - h)\big] - \frac{h^2}{6}f'''(x_0) - \frac{h^4}{120}f^{(5)}(x_0) - \dots.$$

Since, in this case, the error term contains only even powers of h, the convergence of Richardson's extrapolation is very fast. Putting

$$N_1(h) = N(h) = \frac{1}{2h}\big[f(x_0 + h) - f(x_0 - h)\big],$$

the above formula for $f'(x_0)$ becomes

$$f'(x_0) = N_1(h) - \frac{h^2}{6}f'''(x_0) - \frac{h^4}{120}f^{(5)}(x_0) - \dots.$$

Replacing h with $h/2$ in this formula gives the approximation

$$f'(x_0) = N_1\left(\frac{h}{2}\right) - \frac{h^2}{24}f'''(x_0) - \frac{h^4}{1920}f^{(5)}(x_0) - \dots.$$

Subtracting the second last formula for $f'(x_0)$ from 4 times the last one and dividing by 3, we have

$$f'(x_0) = N_2(h) - \frac{h^4}{480}f^{(5)}(x_0) + \dots,$$

where

$$N_2(h) = N_1\left(\frac{h}{2}\right) + \frac{N_1(h/2) - N_1(h)}{3}.$$

The presence of the number $4^{j-1} - 1$ in the denominator of the second term of $N_j(h)$ provides fast convergence.

EXAMPLE 9.3. Let

$$f(x) = x\,e^x.$$

Apply Richardson's extrapolation to the centred difference formula to compute $f'(x)$ at $x_0 = 2$ with $h = 0.2$.

TABLE 9.1. Richardson's extrapolation to the derivative of $x\,e^x$.

$N_1(0.2) = 22.414\,160$		
$N_1(0.1) = 22.228\,786$	$N_2(0.2) = 22.166\,995$	
$N_1(0,05) = 22.182\,564$	$N_2(0.1) = 22.167\,157$	$N_3(0.2) = 22.167\,168$

SOLUTION. We have

$$N_1(0.2) = N(0.2) = \frac{1}{0.4}\,[f(2.2) - f(1.8)] = 22.414\,160,$$
$$N_1(0.1) = N(0.1) = \frac{1}{0.2}\,[f(2.1) - f(1.9)] = 22.228\,786,$$
$$N_1(0.05) = N(0.05) = \frac{1}{0.1}\,[f(2.05) - f(1.95)] = 22.182\,564.$$

Next,

$$N_2(0.2) = N_1(0.1) + \frac{N_1(0.1) - N_1(0.2)}{3} = 22.166\,995,$$
$$N_2(0.1) = N_1(0.05) + \frac{N_1(0.05) - N_1(0.1)}{3} = 22.167\,157.$$

Finally,

$$N_3(0.2) = N_2(0.1) + \frac{N_2(0.1) - N_2(0.2)}{15} = 22.167\,168,$$

which is correct to all 6 decimals. The results are listed in Table 9.1. One sees the fast convergence of Richarson's extrapolation for the centred difference formula. □

9.4. Basic Numerical Integration Rules

To approximate the value of the definite integral

$$\int_a^b f(x)\,dx,$$

where the function $f(x)$ is smooth on $[a, b]$ and $a < b$, we subdivide the interval $[a, b]$ into n subintervals of equal length $h = (b - a)/n$. The function $f(x)$ is approximated on each of these subintervals by an interpolating polynomial and the polynomials are integrated.

For the *midpoint rule*, $f(x)$ is interpolated on each subinterval $[x_{i-1}, x_1]$ by $f([x_{i-1} + x_1]/2)$, and the integral of $f(x)$ over a subinterval is estimated by the area of a rectangle (see Fig. 9.2). This corresponds to making a piecewise constant approximation of the function.

For the *trapezoidal rule*, $f(x)$ is interpolated on each subinterval $[x_{i-1}, x_1]$ by a polynomial of degree one, and the integral of $f(x)$ over a subinterval is estimated by the area of a trapezoid (see Fig. 9.3).This corresponds to making a piecewise linear approximation of the function.

For *Simpson's rule*, $f(x)$ is interpolated on each pair of subintervals, $[x_{2i}, x_{2i+1}]$ and $[x_{2i+1}, x_{2i+2}]$, by a polynomial of degree two (parabola), and the integral of $f(x)$ over such pair of subintervals is estimated by the area under the parabola

(see Fig. 9.4). This corresponds to making a piecewise quadratic approximation of the function.

9.4.1. Midpoint rule. The midpoint rule,

$$\int_{x_0}^{x_1} f(x)\,dx = hf(x_1^*) + \frac{1}{24}\,f''(\xi)h^3, \qquad x_0 < \xi < x_1, \tag{9.5}$$

approximates the integral of $f(x)$ on the interval $x_0 \le x \le x_1$ by the area of a rectangle with height $f(x_1^*)$ and base $h = x_1 - x_0$, where x_1^* is the midpoint of the interval $[x_0, x_1]$,

$$x_1^* = \frac{x_0 + x_1}{2},$$

(see Fig. 9.2).

To derive formula (9.5), we expand $f(x)$ in a truncated Taylor series with center at $x = x_1^*$,

$$f(x) = f(x_1^*) + f'(x_1^*)(x - x_1^*) + \frac{1}{2}\,f''(\xi)(x - x_1^*)^2, \qquad x_0 < \xi < x_1.$$

Integrating this expression from x_0 to x_1, we have

$$\begin{aligned}\int_{x_0}^{x_1} f(x)\,dx &= hf(x_1^*) + \int_{x_0}^{x_1} f'(x_1^*)(x - x_1^*)\,dx + \frac{1}{2}\int_{x_0}^{x_1} f''(\xi(x))(x - x_1^*)^2\,dx \\ &= hf(x_1^*) + \frac{1}{2}\,f''(\xi)\int_{x_0}^{x_1} (x - x_1^*)^2\,dx.\end{aligned}$$

where the integral over the linear term $(x - x_1^*)$ is zero because this term is an odd function with respect to the midpoint $x = x_1^*$ and the Mean Value Theorem 7.4 for integrals has been used in the integral of the quadratic term $(x - x_0^*)^2$ which does not change sign over the interval $[x_0, x_1]$. The result follows from the value of the integral

$$\frac{1}{2}\int_{x_0}^{x_1} (x - x_1^*)^2\,dx = \frac{1}{6}\big[(x - x_1^*)^3\big|_{x_0}^{x_1} = \frac{1}{24}\,h^3.$$

9.4.2. Trapezoidal rule. The trapezoidal rule,

$$\int_{x_0}^{x_1} f(x)\,dx = \frac{h}{2}\,[f(x_0) + f(x_1)] - \frac{1}{12}\,f''(\xi)h^3, \qquad x_0 < \xi < x_1, \tag{9.6}$$

approximates the integral of $f(x)$ on the interval $x_0 \le x \le x_1$ by the area of a trapezoid with heights $f(x_0)$ and $f(x_1)$ and base $h = x_1 - x_0$ (see Fig. 9.3).

To derive formula (9.6), we interpolate $f(x)$ at $x = x_0$ and $x = x_1$ by the linear Lagrange polynomial

$$p_1(x) = f(x_0)\frac{x - x_1}{x_0 - x_1} + f(x_1)\frac{x - x_0}{x_1 - x_0}.$$

Thus,

$$f(x) = p_1(x) + \frac{f''(\xi)}{2}(x - x_0)(x - x_1), \qquad x_0 < \xi < x_1.$$

Since

$$\int_{x_0}^{x_1} p_1(x)\,dx = \frac{h}{2}\big[f(x_0) + f(x_1)\big],$$

we have

$$\begin{aligned}
&\int_{x_0}^{x_1} f(x)\,dx - \frac{h}{2}\,[f(x_0)+f(x_1)] \\
&\quad = \int_{x_0}^{x_1} [f(x)-p_1(x)]\,dx \\
&\quad = \int_{x_0}^{x_1} \frac{f''(\xi(x))}{2}\,(x-x_0)(x-x_1)\,dx \\
&\qquad \text{by the Mean Value Theorem 7.4 for integrals} \\
&\qquad \text{since } (x-x_0)(x-x_1)\le 0 \text{ over } [x_0,x_1] \\
&\quad = \frac{f''(\xi)}{2}\int_{x_0}^{x_1} (x-x_0)(x-x_1)\,dx \\
&\qquad \text{with } x-x_0=s, dx=ds, x-x_1=(x-x_0)-(x_1-x_0)=s-h \\
&\quad = \frac{f''(\xi)}{2}\int_0^h s(s-h)\,ds \\
&\quad = \frac{f''(\xi)}{2}\left[\frac{s^3}{3}-\frac{h}{2}\,s^2\right]_0^h \\
&\quad = -\frac{f''(\xi)}{12}\,h^3,
\end{aligned}$$

.

9.4.3. Simpson's rule. Simpson's rule

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}\left[f(x_0)+4f(x_1)+f(x_2)\right] - \frac{h^5}{90}\,f^{(4)}(\xi), \qquad x_0<\xi<x_2, \quad (9.7)$$

approximates the integral of $f(x)$ on the interval $x_0 \le x \le x_2$ by the area under a parabola which interpolates $f(x)$ at $x=x_0$, x_1 and x_2 (see Fig. 9.4).

To derive formula (9.7), we expand $f(x)$ in a truncated Taylor series with center at $x=x_1$,

$$f(x) = f(x_1)+f'(x_1)(x-x_1)+\frac{f''(x_1)}{2}\,(x-x_1)^2+\frac{f'''(x_1)}{6}\,(x-x_1)^3+\frac{f^{(4)}(\xi(x))}{24}\,(x-x_1)^4.$$

Integrating this expression from x_0 to x_2 and noticing that the odd terms $(x-x_1)$ and $(x-x_1)^3$ are odd functions with respect to the point $x=x_1$ so that their integrals vanish, we have

$$\begin{aligned}
\int_{x_0}^{x_2} f(x)\,dx &= \left[f(x_1)x+\frac{f''(x_1)}{6}\,(x-x_1)^3+\frac{f^{(4)}(\xi_1)}{120}\,(x-x_1)^5\right|_{x_0}^{x_2} \\
&= 2hf(x_1)+\frac{h^3}{3}\,f''(x_1)+\frac{f^{(4)}(\xi_1)}{60}\,h^5,
\end{aligned}$$

where the Mean Value Theorem 7.4 for integrals was used in the integral of the error term because the factor $(x-x_1)^4$ does not change sign over the interval $[x_0,x_2]$.

Substituting the three-point centered difference formula (9.4) for $f''(x_1)$ in terms of $f(x_0)$, $f(x_1)$ and $f(x_2)$:

$$f''(x_1) = \frac{1}{h^2}\,[f(x_0)-2f(x_1)+f(x_2)] - \frac{1}{12}\,f^{(4)}(\xi_2)h^2,$$

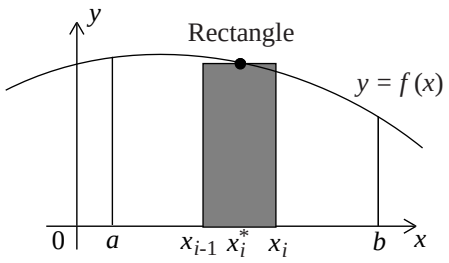

FIGURE 9.2. The ith panel of the midpoint rule.

we obtain

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}\left[f(x_0) + 4f(x_1) + f(x_2)\right] - \frac{h^5}{12}\left[\frac{1}{3}\,f^{(4)}(\xi_2) - \frac{1}{5}\,f^{(4)}(\xi_2)\right].$$

In this case, we cannot apply the Mean Value Theorem 7.5 for sums to express the error term in the form of $f^{(4)}(\xi)$ evaluated at one point since the weights $1/3$ and $-1/5$ have different signs. However, since the formula is exact for polynomials of degree less than or equal to 4, to obtain the factor $1/90$ it suffices to apply the formula to the monomial $f(x) = x^4$ and, for simplicity, integrate from $-h$ to h:

$$\begin{aligned}\int_{-h}^{h} x^4\,dx &= \frac{h}{3}\left[(-h)^4 + 4(0)^4 + h^4\right] + kf^{(4)}(\xi)\\ &= \frac{2}{3}\,h^5 + 4!k = \frac{2}{5}\,h^5,\end{aligned}$$

where the last term is the exact value of the integral. It follows that

$$k = \frac{1}{4!}\left[\frac{2}{5} - \frac{2}{3}\right] h^5 = -\frac{1}{90}h^5,$$

which yields (9.7).

9.5. The Composite Midpoint Rule

We subdivide the interval $[a, b]$ into n subintervals of equal length $h = (b - a)/n$ with end-points

$$x_0 = a, \qquad x_1 = a + h, \qquad \ldots, \qquad x_i = a + ih, \qquad \ldots, \qquad x_n = b.$$

On the subinterval $[x_{i-1}, x_i]$, the integral of $f(x)$ is approximated by the signed area of the rectangle with base $[x_{i-1}, x_i]$ and height $f(x_i^*)$, where

$$x_i^* = \frac{1}{2}\,(x_{i-1} + x_i)$$

is the mid-point of the segment $[x_{i-1}, x_i]$, as shown in Fig. 9.2 Thus, on the interval $[x_{i-1}, x_i]$, by the basic midpoint rule (9.5) we have

$$\int_{x_{i-1}}^{x_i} f(x)\,dx = hf(x_i^*) + \frac{1}{24}\,f''(\xi_i)h^3, \qquad x_{i-1} < \xi_i < x_i.$$

Summing over all the subintervals, we have

$$\int_a^b f(x)\,dx = h\sum_{i=1}^{n} f(x_i^*) + \frac{h^3}{24}\sum_{i=1}^{n} f''(\xi_i).$$

Multiplying and dividing the error term by n, applying the Mean Value Theorem 7.5 for sums to this term and using the fact that $nh = b - a$, we have

$$\frac{nh^3}{24}\sum_{i=1}^{n}\frac{1}{n}f''(\xi_i) = \frac{(b-a)h^2}{24}f''(\xi), \qquad a < \xi < b.$$

Thus, we obtain the *composite midpoint rule*:

$$\int_a^b f(x)\,dx = h\big[f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)\big] + \frac{(b-a)h^2}{24}f''(\xi), \qquad a < \xi < b. \quad (9.8)$$

We see that the composite midpoint rule is a method of order $O(h^2)$, which is exact for polynomials of degree smaller than or equal to 1.

EXAMPLE 9.4. Use the composite midpoint rule to approximate the integral

$$I = \int_0^1 e^{x^2}\,dx$$

with step size h such that the absolute truncation error is bounded by 10^{-4}.

SOLUTION. Since

$$f(x) = e^{x^2} \quad \text{and} \quad f''(x) = (2 + 4\,x^2)\,e^{x^2},$$

then

$$0 \le f''(x) \le 6\,e \quad \text{for} \quad x \in [0, 1].$$

Therefore, a bound for the absolute truncation error is

$$|\epsilon_M| \le \frac{1}{24}\,6\,e(1-0)h^2 = \frac{1}{4}\,eh^2 < 10^{-4}.$$

Thus

$$h < 0.0121 \qquad \frac{1}{h} = 82.4361.$$

We take $n = 83 \ge 1/h = 82.4361$ and $h = 1/83$. The approximate value of I is

$$\begin{aligned} I &\approx \frac{1}{83}\left[e^{(0.5/83)^2} + e^{(1.5/83)^2} + \cdots + e^{(13590.5/83)^2} + e^{(82.5/83)^2}\right] \\ &\approx 1.46262 \quad \square \end{aligned}$$

The following Matlab commands produce the midpoint integration.

```
x = 0.5:82.5; y = exp((x/83).^2);
z = 1/83*sum(y)
  z = 1.4626
```

$\square$

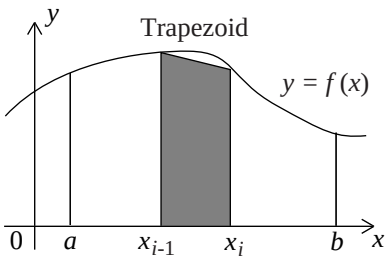

FIGURE 9.3. The ith panel of the trapezoidal rule.

9.6. The Composite Trapezoidal Rule

We divide the interval $[a, b]$ into n subintervals of equal length $h = (b-a)/n$, with end-points

$$x_0 = a, \quad x_1 = a + h, \quad \ldots, \quad x_i = a + ih, \quad \ldots, \quad x_n = b.$$

On each subinterval $[x_{i-1}, x_i]$, the integral of $f(x)$ is approximated by the signed area of the trapezoid with vertices

$$(x_{i-1}, 0), \quad (x_i, 0), \quad (x_i, f(x_i)), \quad (x_{i-1}, f(x_{i-1})),$$

as shown in Fig. 9.3. Thus, by the basic trapezoidal rule (9.6),

$$\int_{x_{i-1}}^{x_1} f(x)\,dx = \frac{h}{2}\left[f(x_{i-1}) + f(x_i)\right] - \frac{h^3}{12} f''(\xi_i).$$

Summing over all the subintervals, we have

$$\int_a^b f(x)\,dx = \frac{h}{2}\sum_{i=1}^{n}\left[f(x_{i-1}) + f(x_i)\right] - \frac{h^3}{12}\sum_{i=1}^{n} f''(\xi_i).$$

Multiplying and dividing the error term by n, applying the Mean Value Theorem 7.5 for sums to this term and using the fact that $nh = b - a$, we have

$$-\frac{nh^3}{12}\sum_{i=1}^{n}\frac{1}{n} f''(\xi_i) = -\frac{(b-a)h^2}{12} f''(\xi), \qquad a < \xi < b.$$

Thus, we obtain the *composite trapezoidal rule*:

$$\begin{aligned}\int_a^b f(x)\,dx = \frac{h}{2}\big[&f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-2}) \\ &+ 2f(x_{n-1}) + f(x_n)\big] - \frac{(b-a)h^2}{12} f''(\xi), \qquad a < \xi < b. \end{aligned} \tag{9.9}$$

We see that the composite trapezoidal rule is a method of order $O(h^2)$, which is exact for polynomials of degree smaller than or equal to 1. Its absolute truncation error is twice the absolute truncation error of the midpoint rule.

EXAMPLE 9.5. Use the composite trapezoidal rule to approximate the integral

$$I = \int_0^1 e^{x^2}\,dx$$

with step size h such that the absolute truncation error is bounded by 10^{-4}. Compare with Examples 9.4 and 9.7.

SOLUTION. Since

$$f(x) = e^{x^2} \quad \text{and} \quad f''(x) = (2 + 4\,x^2)\,e^{x^2},$$

then

$$0 \le f''(x) \le 6\,e \quad \text{for} \quad x \in [0, 1].$$

Therefore,

$$|\epsilon_T| \le \frac{1}{12}\,6\,e(1-0)h^2 = \frac{1}{2}\,eh^2 < 10^{-4}, \quad \text{that is,} \quad h < 0.008\,577\,638.$$

We take $n = 117 \ge 1/h = 116.6$ (compared to 83 for the composite midpoint rule). The approximate value of I is

$$\begin{aligned} I &\approx \frac{1}{117 \times 2}\left[e^{(0/117)^2} + 2\,e^{(1/117)^2} + 2\,e^{(2/117)^2} + \cdots \right. \\ &\qquad \left. + 2\,e^{(115/117)^2} + 2\,e^{(116/117)^2} + e^{(117/117)^2}\right] \\ &= 1.46268. \end{aligned}$$

The following Matlab commands produce the trapesoidal integration of numerical values y_k at nodes $k/117$, $k = 0, 1, \ldots, 117$, with stepsize $h = 1/117$.

```
x = 0:117; y = exp((x/117).^2);
z = trapz(x,y)/117
  z = 1.4627
```

□

EXAMPLE 9.6. How many subintervals are necessary for the composite trapezoidal rule to approximate the integral

$$I = \int_1^2 \left[x^2 - \frac{1}{12}(x - 1.5)^4\right] dx$$

with step size h such that the absolute truncation error is bounded by 10^{-3} ?

SOLUTION. Denote the integrand by

$$f(x) = x^2 - \frac{1}{12}(x - 1.5)^4.$$

Then

$$f''(x) = 2 - (x - 1.5)^2.$$

It is clear that

$$M = \max_{1 \le x \le 2} f''(x) = f(1.5) = 2.$$

To bound the absolute truncation error by 10^{-3}, we need

$$\begin{aligned} \left|\frac{(b-a)h^2}{12} f''(\xi)\right| &\le \frac{h^2}{12}\,M \\ &= \frac{h^2}{6} \\ &\le 10^{-3}. \end{aligned}$$

This gives

$$h \le \sqrt{6 \times 10^{-3}} = 0.0775$$

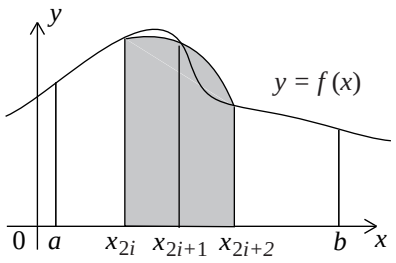

FIGURE 9.4. A double panel of Simpson's rule.

and

$$\frac{1}{h} = 12.9099 \le n = 13.$$

Thus it suffices to take

$$h = \frac{1}{13}, \qquad n = 13. \quad \square$$

9.7. The Composite Simpson Rule

We subdivide the interval $[a, b]$ into an even number, $n = 2m$, of subintervals of equal length, $h = (b-a)/(2m)$, with end-points

$$x_0 = a, \quad x_1 = a + h, \quad \ldots, \quad x_i = a + i\,h, \quad \ldots, \quad x_{2m} = b.$$

On the subinterval $[x_{2i}, x_{2i+2}]$, the function $f(x)$ is interpolated by the quadratic polynomial $p_2(x)$ which passes through the points

$$\big(x_{2i}, f(x_{2i})\big), \quad \big(x_{2i+1}, f(x_{2i+1})\big), \quad \big(x_{2i+2}, f(x_{2i+2})\big),$$

as shown in Fig. 9.4.

Thus, by the basic Simpson rule (9.7),

$$\int_{x_{2i}}^{x_{2i+2}} f(x)\,dx = \frac{h}{3}\left[f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})\right] - \frac{h^5}{90}\,f^{(4)}(\xi_i), \quad x_{2i} < \xi < x_{2i+2}.$$

Summing over all the subintervals, we have

$$\int_a^b f(x)\,dx = \frac{h}{3}\sum_{i=1}^{m}\big[f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})\big] - \frac{h^5}{90}\sum_{i=1}^{m} f^{(4)}(\xi_i).$$

Multiplying and dividing the error term by m, applying the Mean Value Theorem 7.5 for sums to this term and using the fact that $2mh = nh = b - a$, we have

$$-\frac{2mh^5}{2\times 90}\sum_{i=1}^{m}\frac{1}{m}f^{(4)}(\xi_i) = -\frac{(b-a)h^4}{180}f^{(4)}(\xi), \qquad a < \xi < b.$$

Thus, we obtain the *composite Simpson rule*:

$$\int_a^b f(x)\,dx = \frac{h}{3}\big[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots$$

$$+ 2f(x_{2m-2}) + 4f(x_{2m-1}) + f(x_{2m})\big] - \frac{(b-a)h^4}{180}f^{(4)}(\xi), \qquad a < \xi < b. \quad (9.10)$$

We see that the composite Simpson rule is a method of order $O(h^4)$, which is exact for polynomials of degree smaller than or equal to 3.

EXAMPLE 9.7. Use the composite Simpson rule to approximate the integral

$$I = \int_0^1 e^{x^2}\, dx$$

with stepsize h such that the absolute truncation error is bounded by 10^{-4}. Compare with Examples 9.4 and 9.5.

SOLUTION. We have

$$f(x) = e^{x^2} \qquad \text{and} \qquad f^{(4)}(x) = 4\, e^{x^2} \left(3 + 12x^2 + 4x^4\right).$$

Thus

$$0 \le f^{(4)}(x) \le 76\, e \quad \text{on} \quad [0, 1].$$

The absolute truncation error is thus less than or equal to $\frac{76}{180}\, e(1-0)h^4$. Hence, h must satisfy the inequality

$$\frac{76}{180}\, eh^4 < 10^{-4}, \quad \text{that is,} \quad h < 0.096\,614\,232.$$

To satisfy the inequality

$$2m \ge \frac{1}{h} = 10.4$$

we take

$$n = 2m = 12 \quad \text{and} \quad h = \frac{1}{12}.$$

The approximation is

$$\begin{aligned} I &\approx \frac{1}{12 \times 3} \left[e^{(0/12)^2} + 4\, e^{(1/12)^2} + 2\, e^{(2/12)^2} + \cdots + 2\, e^{(10/12)^2} + 4\, e^{(11/12)^2} + e^{(12/12)^2} \right] \\ &= 1.46267. \end{aligned}$$

We obtain a value which is similar to those found in Examples 9.4 and 9.5. However, the number of arithmetic operations is much less when using Simpson's rule (hence cost and truncation errors are reduced). In general, Simpson's rule is preferred to the midpoint and trapezoidal rules. □

EXAMPLE 9.8. Use the composite Simpson rule to approximate the integral

$$I = \int_0^2 \sqrt{1 + \cos^2 x}\, dx$$

within an accuracy of 0.0001.

SOLUTION. We must determine the step size h such that the absolute truncation error, $|\epsilon_S|$, will be bounded by 0.0001. For

$$f(x) = \sqrt{1 + \cos^2 x},$$

we have

$$\begin{aligned} f^{(4)}(x) = {} & \frac{-3\cos^4(x)}{(1+\cos^2(x))^{3/2}} + \frac{4\cos^2(x)}{\sqrt{1+\cos^2(x)}} - \frac{18\cos^4(x)\sin^2(x)}{(1+\cos^2(x))^{5/2}} \\ & + \frac{22\cos^2(x)\sin^2(x)}{(1+\cos^2(x))^{3/2}} - \frac{4\sin^2(x)}{\sqrt{1+\cos^2(x)}} - \frac{15\cos^4(x)\,\sin^4(x)}{(1+\cos^2(x))^{7/2}} \\ & + \frac{18\cos^2(x)\,\sin^4(x)}{(1+\cos^2(x))^{5/2}} - \frac{3\sin^4(x)}{(1+\cos^2(x))^{3/2}}. \end{aligned}$$

Since every denominator is greater than one, we have

$$|f^{(4)}(x)| \le 3 + 4 + 18 + 22 + 4 + 15 + 18 + 3 = 87.$$

Therefore, we need

$$|\epsilon_S| < \frac{87}{180}(2-0)h^4.$$

Hence,

$$h < 0.100\,851\,140, \qquad \frac{1}{h} > 9.915\,604\,269.$$

To have $2m \ge 2/h = 2 \times 9.9$ we take $n = 2m = 20$ and $h = 1/10$. The approximation is

$$\begin{aligned} I &\approx \frac{1}{20 \times 3}\left[\sqrt{1+\cos^2(0)} + 4\sqrt{1+\cos^2(0.1)} + 2\sqrt{1+\cos^2(0.2)} + \cdots \right. \\ &\quad \left. + 2\sqrt{1+\cos^2(1.8)} + 4\sqrt{1+\cos^2(1.9)} + \sqrt{1+\cos^2(2)}\right] \\ &= 2.35169. \quad \square \end{aligned}$$

9.8. Romberg Integration for the Trapezoidal Rule

Romberg integration uses Richardson's extrapolation to improve the trapezoidal rule approximation, $R_{k,1}$, with step size h_k, to an integral

$$I = \int_a^b f(x)\,dx.$$

It can be shown that

$$I = R_{k,1} + K_1 h_k^2 + K_2 h_k^4 + K_3 h_k^6 + \ldots,$$

where the constants K_j are independent of h_k. With step sizes

$$h_1 = h, \quad h_2 = \frac{h}{2}, \quad h_3 = \frac{h}{2^2}, \quad \ldots, \quad h_k = \frac{h}{2^{k-1}}, \quad \ldots,$$

one can cancel errors of order h^2, h^4, etc. as follows. Suppose $R_{k,1}$ and $R_{k+1,1}$ have been computed, then we have

$$I = R_{k,1} + K_1 h_k^2 + K_2 h_k^4 + K_3 h_k^6 + \ldots$$

and

$$I = R_{k+1,1} + K_1 \frac{h_k^2}{4} + K_2 \frac{h_k^4}{16} + K_3 \frac{h_k^6}{64} + \ldots.$$

Subtracting the first expression for I from 4 times the second expression and dividing by 3, we obtain

$$I = \left[R_{k+1,1} + \frac{R_{k+1,1} - R_{k,1}}{3}\right] + \frac{K_2}{3}\left[\frac{1}{4} - 1\right] h_k^4 + \frac{K_3}{3}\left[\frac{1}{16} - 1\right] h_k^4 + \ldots.$$

Put

$$R_{k,2} = R_{k,1} + \frac{R_{k,1} - R_{k-1,1}}{3}$$

and, in general,

$$R_{k,j} = R_{k,j-1} + \frac{R_{k,j-1} - R_{k-1,j-1}}{4^{j-1} - 1}.$$

Then $R_{k,j}$ is a better approximation to I than $R_{k,j-1}$ and $R_{k-1,j-1}$. The relations between the $R_{k,j}$ are shown in Table 9.2.

TABLE 9.2. Romberg integration table with n levels

$$
\begin{array}{ccccccccc}
R_{1,1} & & & & & & & & \\
 & \searrow & & & & & & & \\
R_{2,1} & \rightarrow & R_{2,2} & & & & & & \\
 & \searrow & & \searrow & & & & & \\
R_{3,1} & \rightarrow & R_{3,2} & \rightarrow & R_{3,3} & & & & \\
 & \searrow & & \searrow & & \searrow & & & \\
R_{4,1} & \rightarrow & R_{4,2} & \rightarrow & R_{4,3} & \rightarrow & R_{4,4} & & \\
\vdots & & \vdots & \vdots & \vdots & & & & \\
 & \searrow & & \searrow & & \searrow & & \searrow & \\
R_{n,1} & \rightarrow & R_{n,2} & \rightarrow & R_{n,3} & \rightarrow & R_{n,4} & \cdots \rightarrow & R_{n,n}
\end{array}
$$

EXAMPLE 9.9. Use 6 levels of Romberg integration, with $h_1 = h = \pi/4$, to approximate the integral

$$I = \int_0^{\pi/4} \tan x \, dx.$$

SOLUTION. The following results are obtained by a simple Matlab program.

```
Romberg integration table:

0.39269908
0.35901083  0.34778141
0.34975833  0.34667417  0.34660035
0.34737499  0.34658054  0.34657430  0.34657388
0.34677428  0.34657404  0.34657360  0.34657359  0.34657359
0.34662378  0.34657362  0.34657359  0.34657359  0.34657359  0.34657359
```

□

9.9. Adaptive Quadrature Methods

Uniformly spaced composite rules that are exact for degree d polynomials are efficient if the $(d+1)$st derivative $f^{(d+1)}$ is uniformly behaved across the interval of integration $[a, b]$. However, if the magnitude of this derivative varies widely across this interval, the error control process may result in an unnecessary number of function evaluations. This is because the number n of nodes is determined by an interval-wide derivative bound M_{d+1}. In regions where $f^{(d+1)}$ is small compared to this value, the subintervals are (possibly) much shorter than necessary. *Adaptive quadrature* methods address this problem by discovering where the integrand is *ill-behaved* and shortening the subintervals accordingly.

We take Simpson's rule as a typical example:

$$I =: \int_a^b f(x)\,dx = S(a,b) - \frac{h^5}{90} f^{(4)}(\xi), \qquad 0 < \xi < b,$$

where

$$S(a,b) = \frac{h}{3}\,[f(a) + 4f(a+h) + f(b)], \qquad h = \frac{b-a}{2}.$$

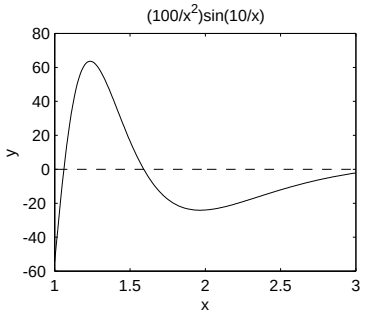

FIGURE 9.5. A fast varying function for adaptive quadrature.

The aim of adaptive quadrature is to take h large over regions where $|f^{(4)}(x)|$ is small and take h small over regions where $|f^{(4)}(x)|$ is large to have a uniformly small error. A simple way to estimate the error is to use h and $h/2$ as follows:

$$I = S(a,b) - \frac{h^5}{90} f^{(4)}(\xi_1), \tag{9.11}$$

$$I = S\left(a, \frac{a+b}{2}\right) + S\left(\frac{a+b}{2}, b\right) - \frac{2}{32}\frac{h^5}{90} f^{(4)}(\xi_2). \tag{9.12}$$

Assuming that

$$f^{(4)}(\xi_2) \approx f^{(4)}(\xi_1)$$

and subtracting the second expression for I from the first we have an expression for the error term:

$$\frac{h^5}{90} f^{(4)}(\xi_1) \approx \frac{16}{15}\left[S(a,b) - S\left(a, \frac{a+b}{2}\right) - S\left(\frac{a+b}{2}, b\right)\right].$$

Putting this expression in (9.12), we obtain an estimate for the absolute error:

$$\left|I - S\left(a, \frac{a+b}{2}\right) - S\left(\frac{a+b}{2}, b\right)\right| \approx \frac{1}{15}\left|S(a,b) - S\left(a, \frac{a+b}{2}\right) - S\left(\frac{a+b}{2}, b\right)\right|.$$

If the right-hand side of this estimate is smaller than a given tolerance, then

$$S\left(a, \frac{a+b}{2}\right) + S\left(\frac{a+b}{2}, b\right)$$

is taken as a good approximation to the value of I.

The adaptive quadrature for Simpson's rule is often better than the composite Simpson rule. For example, in integrating the function

$$f(x) = \frac{100}{x^2}\sin\left(\frac{10}{x}\right), \qquad 1 \le x \le 3,$$

shown in Fig. 9.5, with toleralance 10^{-4}, the adaptive quadrature uses 23 subintervals and requires 93 evaluations of f. On the other hand, the composite Simpson rule uses a constant value of $h = 1/88$ and requires 177 evaluations of f. It is seen from the figure that f varies quickly over the interval $[1, 1.5]$. The adaptive quadrature needs 11 subintervals on the short interval $[1, 1.5]$ and only 12 on the longer interval $[1.5, 3]$.

The MATLAB quadrature routines `quad`, `quadl` and `dblquad` are adaptive routines.

Matlab's adaptive Simpson's rule `quad` and adaptive Newton–Cotes 8-panel rule `quad8` evaluate the integral

$$I = \int_0^{\pi/2} \sin x \, dx$$

as follows.

```
>> v1 = quad('sin',0,pi/2)
 v1 = 1.00000829552397
>> v2 = quad8('sin',0,pi/2)
 v2 = 1.00000000000000
```

respectively, within a relative error of 10^{-3}.

9.10. Gaussian Quadrature

The Gaussian Quadrature formulae are the most accurate integration formulae for a given number of nodes. The n-point Gaussian Quadrature formula approximate the integral of $f(x)$ over the standardized interval $-1 \le x \le 1$ by the formula

$$\int_{-1}^{1} f(x) \, dx \approx \sum_{i=1}^{n} w_i f(t_i) \tag{9.13}$$

where the nodes x_i are the zeros of the Legendre polynomial $P_n(x)$ of degree n.

The two-point Gaussian Quadrature formula is

$$\int_{-1}^{1} f(x) \, dx = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right).$$

The three-point Gaussian Quadrature formula is

$$\int_{-1}^{1} f(x) \, dx = \frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right).$$

The nodes x_i, weights w_i and precision $2n-1$ of n points Gaussian Quadratures, are listed in Table 9.3 for $n = 1, 2, \ldots, 5$.

The error in the n-point formula is

$$E_n(f) = \frac{2}{(2n+1)!} \left[\frac{2^n (n!)^2}{(2n)!}\right]^2 f^{(2n)}(\xi), \qquad -1 < \xi < 1.$$

This formula is therefore exact for polynomials of degree $2n-1$ or less.

Gaussian Quadratures are derived in Section 6.6 by means of the orthogonality relations of the Legendre polynomials. These quadratures can also be obtained by means of the integrals of the Lagrange basis on $-1 \le x \le 1$ for the nodes x_i taken as the zeros of the Legendre polynomials:

$$w_i = \int_{-1}^{1} \prod_{j=1, j\neq i}^{n} \frac{x - x_j}{x_i - x_j} \, dx.$$

Examples can be found in Section 6.6 and exercises in Exercises for Chapter 6.

In the applications, the interval $[a, b]$ of integration is split into smaller intervals and a Gaussian Quadrature is used on each subinterval with an appropriate change of variable as in Example 6.11.

TABLE 9.3. Nodes x_i, weights w_i and precision $2n-1$ of n points Gaussian quadratures.

n	x_i	w_i	Precision $2n-1$
2	$-1/\sqrt{3}$	1	3
	$1/\sqrt{3}$	1	
3	$-\sqrt{3/5}$	$5/9$	5
	0	$8/9$	
	$\sqrt{3/5}$	$5/9$	
4	$-0.861\,136\,311\,6$	$0.347\,854\,845\,1$	7
	$-0.339\,981\,043\,6$	$0.652\,145\,154\,9$	
	$0.339\,981\,043\,6$	$0.652\,145\,154\,9$	
	$0.861\,136\,311\,6$	$0.347\,854\,845\,1$	
5	$-0.906\,179\,845\,9$	$0.236\,926\,885\,1$	9
	$-0.538\,469\,310\,1$	$0.478\,628\,670\,5$	
	0	$0{,}568\,888\,888\,9$	
	$0.538\,469\,310\,1$	$0.478\,628\,670\,5$	
	$0.906\,179\,845\,9$	$0.236\,926\,885\,1$	

CHAPTER 10

Numerical Solution of Differential Equations

10.1. Initial Value Problems

Consider the first-order initial value problem:

$$y' = f(x, y), \qquad y(x_0) = y_0. \tag{10.1}$$

To find an approximation to the solution $y(x)$ of (10.1) on the interval $a \le x \le b$, we choose $N + 1$ distinct points, $x_0, x_1, \ldots, x_N$, such that $a = x_0 < x_1 < x_2 < \ldots < x_N = b$, and construct approximations y_n to $y(x_n)$, $n = 0, 1, \ldots, N$.

It is important to know whether or not a *small perturbation* of (10.1) shall lead to a *large variation* in the solution. If this is the case, it is extremely unlikely that we will be able to find a good approximation to (10.1). Truncation errors, which occur when computing $f(x, y)$ and evaluating the initial condition, can be identified with perturbations of (10.1). The following theorem gives sufficient conditions for an initial value problem to be *well-posed*.

Definition 10.1. Problem (10.1) is said to be *well-posed* in the sense of Hadamard if it has one, and only one, solution and any small perturbation of the problem leads to a correspondingly small change in the solution.

Theorem 10.1. *Let*

$$D = \{(x, y) : a \le x \le b \text{ and } -\infty < y < \infty\}.$$

If $f(x, y)$ is continuous on D and satisfies the Lipschitz condition

$$|f(x, y_1) - f(x, y_2)| \le L|y_1 - y_2| \tag{10.2}$$

for all (x, y_1) and (x, y_2) in D, where L is the Lipschitz constant, then the initial value problem (10.1) *is well-posed, that is, we are assuming that the problem satisfies the Existence and Uniqueness Theorem requirements as seen in Chapter 1.*

In the sequel, we shall assume that the conditions of Theorem 10.1 hold and (10.1) is well-posed. Moreover, we shall suppose that $f(x, y)$ has mixed partial derivatives of arbitrary order.

In considering numerical methods for the solution of (10.1) we shall use the following notation:

- $h > 0$ denotes the integration *step size*
- $x_n = x_0 + nh$ is the n-th *node*
- $y(x_n)$ is the *exact solution* at x_n
- y_n is the *numerical solution* at x_n
- $f_n = f(x_n, y_n)$ is the numerical value of $f(x, y)$ at (x_n, y_n)

A function, $g(x)$, is said to be of *order* p as $x \to x_0$, written $g \in O(|x - x_0|^p)$ if

$$|g(x)| < M|x - x_0|^p, \qquad M \quad \text{a constant},$$

for all x near x_0.

10.2. Euler's and Improved Euler's Methods

We begin with the simplest explicit method.

10.2.1. Euler's method. To find an approximation to the solution $y(x)$ of (10.1) on the interval $a \le x \le b$, we choose $N+1$ distinct points, $x_0, x_1, \ldots, x_N$, such that $a = x_0 < x_1 < x_2 < \ldots < x_N = b$ and set $h = (x_N - x_0)/N$. From Taylor's Theorem we get

$$y(x_{n+1}) = y(x_n) + y'(x_n)\,(x_{n+1} - x_n) + \frac{y''(\xi_n)}{2}\,(x_{n+1} - x_n)^2$$

with ξ_n between x_n and x_{n+1}, $n = 0, 1, \ldots, N$. Since $y'(x_n) = f(x_n, y(x_n))$ and $x_{n+1} - x_n = h$, it follows that

$$y(x_{n+1}) = y(x_n) + f\big(x_n, y(x_n)\big)h + \frac{y''(\xi_n)}{2}\,h^2.$$

We obtain Euler's method,

$$y_{n+1} = y_n + hf(x_n, y_n), \tag{10.3}$$

by deleting the term of order $O(h^2)$,

$$\frac{y''(\xi_n)}{2}\,h^2,$$

called the *local truncation error*. This corresponds to making a piecewise linear approximation to the solution as we are assuming constant slope on each subinterval $[x_i, x_{i+1}]$, based on the left endpoint.

The algorithm for *Euler's method* is as follows.

(1) Choose h such that $N = (x_N - x_0)/h$ is an integer.
(2) Given y_0, for $n = 0, 1, \ldots, N$, iterate the scheme

$$y_{n+1} = y_n + hf(x_0 + nh, y_n). \tag{10.4}$$

Then, y_n is as an approximation to $y(x_n)$.

EXAMPLE 10.1. Use Euler's method with $h = 0.1$ to approximate the solution to the initial value problem

$$y'(x) = 0.2xy, \qquad y(1) = 1, \tag{10.5}$$

on the interval $1 \le x \le 1.5$.

SOLUTION. We have

$$x_0 = 1, \qquad x_N = 1.5, \qquad y_0 = 1, \qquad f(x, y) = 0.2xy.$$

Hence

$$x_n = x_0 + hn = 1 + 0.1n, \qquad N = \frac{1.5 - 1}{0.1} = 5,$$

and

$$y_{n+1} = y_n + 0.1 \times 0.2(1 + 0.1n)y_n, \quad \text{with } y_0 = 1,$$

for $n = 0, 1, \ldots, 4$. The numerical results are listed in Table 10.1. Note that the differential equation in (10.5) is separable. The (unique) solution of (10.5) is

$$y(x) = e^{(0.1x^2 - 0.1)}.$$

This formula has been used to compute the exact values $y(x_n)$ in the table. □

TABLE 10.1. Numerical results of Example 10.1.

n	x_n	y_n	$y(x_n)$	Absolute error	Relative error
0	1.00	1.0000	1.0000	0.0000	0.00
1	1.10	1.0200	1.0212	0.0012	0.12
2	1.20	1.0424	1.0450	0.0025	0.24
3	1.30	1.0675	1.0714	0.0040	0.37
4	1.40	1.0952	1.1008	0.0055	0.50
5	1.50	1.1259	1.1331	0.0073	0.64

TABLE 10.2. Numerical results of Example 10.2.

n	x_n	y_n	$y(x_n)$	Absolute error	Relative error
0	1.00	1.0000	1.0000	0.0000	0.00
1	1.10	1.2000	1.2337	0.0337	2.73
2	1.20	1.4640	1.5527	0.0887	5.71
3	1.30	1.8154	1.9937	0.1784	8.95
4	1.40	2.2874	2.6117	0.3244	12.42
5	1.50	2.9278	3.4904	0.5625	16.12

The next example illustrates the limitations of Euler's method. In the next subsections, we shall see more accurate methods than Euler's method.

EXAMPLE 10.2. Use Euler's method with $h = 0.1$ to approximate the solution to the initial value problem

$$y'(x) = 2xy, \qquad y(1) = 1, \tag{10.6}$$

on the interval $1 \le x \le 1.5$.

SOLUTION. As in the previous example, we have

$$x_0 = 1, \quad x_N = 1.5, \quad y_0 = 1, \quad x_n = x_0 + hn = 1 + 0.1n, \quad N = \frac{1.5 - 1}{0.1} = 5.$$

With $f(x, y) = 2xy$, Euler's method is

$$y_{n+1} = y_n + 0.1 \times 2(1 + 0.1n)y_n, \qquad y_0 = 1,$$

for $n = 0, 1, 2, 3, 4$. The numerical results are listed in Table 10.2. The relative errors show that our approximations are not very good. □

Because Euler's method is assuming constant slope on each subinterval, the results are not very good in general.

DEFINITION 10.2. The *local truncation error* of a method of the form

$$y_{n+1} = y_n + h\,\phi(x_n, y_n), \tag{10.7}$$

is defined by the expression

$$\tau_{n+1} = \frac{1}{h}\big[y(x_{n+1}) - y(x_n)\big] - \phi(x_n, y(x_n)) \qquad \text{for} \quad n = 0, 1, 2, \ldots, N-1.$$

The method (10.7) is of *order* k if $|\tau_j| \le M\,h^k$ for some constant M and for all j.

An equivalent definition is found in Section 10.4

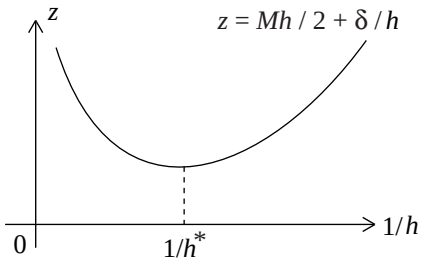

FIGURE 10.1. Truncation and roundoff error curve as a function of $1/h$.

EXAMPLE 10.3. The local truncation error of Euler's method is

$$\tau_{n+1} = \frac{1}{h}\big[y(x_{n+1}) - y(x_n)\big] - f\big(x_n, y(x_n)\big) = \frac{h}{2}\, y''(\xi_n)$$

for some ξ_n between x_n and x_{n+1}. If

$$M = \max_{x_0 \le x \le x_N} |y''(x)|,$$

then $|\tau_n| \le \frac{h}{2} M$ for all n. Hence, Euler's method is of order one.

REMARK 10.1. It is generally incorrect to say that by taking h sufficiently small one can obtain any desired level of precision, that is, get y_n as close to $y(x_n)$ as one wants. As the step size h decreases, at first the truncation error of the method decreases, but as the number of steps increases, the number of arithmetic operations increases, and, hence, the roundoff errors increase as shown in Fig. 10.1.

For instance, let y_n be the computed value for $y(x_n)$ in (10.4). Set

$$e_n = y(x_n) - y_n, \quad \text{for} \quad n = 0, 1, \ldots, N.$$

If

$$|e_0| < \delta_0$$

and the precision in the computations is bounded by δ, then it can be shown that

$$|e_n| \le \frac{1}{L}\left(\frac{Mh}{2} + \frac{\delta}{h}\right)\left(e^{L(x_n - x_0)} - 1\right) + \delta_0\, e^{L(x_n - x_0)},$$

where L is the Lipschitz constant defined in Theorem 10.1,

$$M = \max_{x_0 \le x \le x_N} |y''(x)|,$$

and $h = (x_N - x_0)/N$.

We remark that the expression

$$z(h) = \frac{Mh}{2} + \frac{\delta}{h}$$

first decreases and afterwards increases as $1/h$ increases, as shown in Fig. 10.1. The term $Mh/2$ is due to the trunctation error and the term δ/h is due to the roundoff errors.

TABLE 10.3. Numerical results of Example 10.4.

n	x_n	y_n^{P}	y_n^{C}	$y(x_n)$	Absolute error	Relative error
0	1.00		1.0000	1.0000	0.0000	0.00
1	1.10	1.200	1.2320	1.2337	0.0017	0.14
2	1.20		1.5479	1.5527	0.0048	0.31
3	1.30		1.9832	1.9937	0.0106	0.53
4	1.40		2.5908	2.6117	0.0209	0.80
5	1.50		3.4509	3.4904	0.0344	1.13

10.2.2. Improved Euler's method. The improved Euler's method takes the average of the slopes at the left and right ends of each step. It is, here, formulated in terms of a predictor and a corrector:

$$y_{n+1}^{\mathrm{P}} = y_n^{\mathrm{C}} + hf(x_n, y_n^{\mathrm{C}}),$$
$$y_{n+1}^{\mathrm{C}} = y_n^{\mathrm{C}} + \frac{1}{2}\, h \left[f(x_n, y_n^{\mathrm{C}}) + f(x_{n+1}, y_{n+1}^{\mathrm{P}}) \right].$$

This method is of order 2.

Notice that the second formula or the corrector is an implicit function of y_{n+1}. However the formula has order two and will give better results than Euler's method and so we would like to use it. To get around the problem "we have to know y_{n+1} to calculate y_{n+1}", we use Euler's method to predict a value of y_{n+1} which is then plugged into the implicit formula to calculate a better or corrected value.

EXAMPLE 10.4. Use the improved Euler method with $h = 0.1$ to approximate the solution to the initial value problem of Example 10.2.

$$y'(x) = 2xy, \qquad y(1) = 1,$$

$1 \le x \le 1.5$.

SOLUTION. We have

$$x_n = x_0 + hn = 1 + 0.1n, \qquad n = 0, 1, \ldots, 5.$$

The approximation y_n to $y(x_n)$ is given by the predictor-corrector scheme

$$y_0^{\mathrm{C}} = 1,$$
$$y_{n+1}^{\mathrm{P}} = y_n^{\mathrm{C}} + 0.2\, x_n\, y_n,$$
$$y_{n+1}^{\mathrm{C}} = y_n^{\mathrm{C}} + 0.1 \left(x_n\, y_n^{\mathrm{C}} + x_{n+1}\, y_{n+1}^{\mathrm{P}} \right)$$

for $n = 0, 1, \ldots, 4$. The numerical results are listed in Table 10.3. These results are much better than those listed in Table 10.2 for Euler's method. □

We need to develop methods of order greater than one, which, in general, are more precise than Euler's method.

10.3. Low-Order Explicit Runge–Kutta Methods

Runge–Kutta methods are one-step multistage methods.

10.3.1. Second-order Runge–Kutta method. Two-stage explicit Runge–Kutta methods are given by the formula (left) and, conveniently, in the form of a Butcher tableau (right):

$$\begin{aligned} k_1 &= hf(x_n, y_n) \\ k_2 &= hf\left(x_n + c_2 h, y_n + a_{21} k_1\right) \\ y_{n+1} &= y_n + b_1 k_1 + b_2 k_2 \end{aligned}$$

	$\mathbf{c}$	A	
k_1	0	0	
k_2	c_2	a_{21}	0
y_{n+1}	$\mathbf{b}^T$	b_1	b_2

In a Butcher tableau, the components of the vector $\mathbf{c}$ are the increments of x_n and the entries of the matrix A are the multipliers of the approximate slopes which, after multiplication by the step size h, increments y_n. The components of the vector $\mathbf{b}$ are the weights in the combination of the intermediary values k_j. The left-most column of the tableau is added here for the reader's convenience.

To attain second order, $\mathbf{c}$, A and $\mathbf{b}$ have to be chosen judiciously. We proceed to derive two-stage second-order Runge-Kutta methods.

By Taylor's Theorem, we have

$$y(x_{n+1}) = y(x_n) + y'(x_n)(x_{n+1} - x_n) + \frac{1}{2} y''(x_n)(x_{n+1} - x_n)^2 + \frac{1}{6} y'''(\xi_n)(x_{n+1} - x_n)^3 \quad (10.8)$$

for some ξ_n between x_n and x_{n+1} and $n = 0, 1, \ldots, N-1$. From the differential equation

$$y'(x) = f\big(x, y(x)\big),$$

and its first total derivative with respect to x, we obtain expressions for $y'(x_n)$ and $y''(x_n)$,

$$\begin{aligned} y'(x_n) &= f\big(x_n, y(x_n)\big), \\ y''(x_n) &= \frac{d}{dx} f\big(x, y(x)\big)\big|_{x=x_n} \\ &= f_x\big(x_n, y(x_n)\big) + f_y\big(x_n, y(x_n)\big)\, f\big(x_n, y(x_n)\big). \end{aligned}$$

Therefore, putting $h = x_{n+1} - x_n$ and substituting these expressions in (10.8), we have

$$\begin{aligned} y(x_{n+1}) = y(x_n) &+ f\big(x_n, y(x_n)\big)\, h \\ &+ \frac{1}{2}\left[f_x\big(x_n, y(x_n)\big) + f_y\big(x_n, y(x_n)\big) f\big(x_n, y(x_n)\big)\right] h^2 \\ &+ \frac{1}{6}\, y'''(\xi_n) h^3 \end{aligned} \quad (10.9)$$

for $n = 0, 1, \ldots, N-1$.

Our goal is to replace the expression

$$f\big(x_n, y(x_n)\big)h + \frac{1}{2}\big[f_x\big(x_n, y(x_n)\big) + f_y\big(x_n, y(x_n)\big)\, f\big(x_n, y(x_n)\big)\big]h + O(h^2)$$

by an expression of the form

$$af\big(x_n, y(x_n)\big)h + bf\big(x_n + \alpha h, y(x_n) + \beta h f(x_n, y(x_n))\big)\, h + O(h^2). \quad (10.10)$$

The constants a, b, α and β are to be determined. This last expression is simpler to evaluate than the previous one since it does not involve partial derivatives.

Using Taylor's Theorem for functions of two variables, we get

$$f\big(x_n+\alpha h, y(x_n)+\beta h f(x_n,y(x_n))\big) = f\big(x_n,y(x_n)\big)+\alpha h f_x\big(x_n,y(x_n)\big) + \beta h f\big(x_n,y(x_n)\big) f_y\big(x_n,y(x_n)\big)+O(h^2).$$

In order for the expressions (10.8) and (10.9) to be equal to order h^3, we must have

$$a+b=1, \qquad \alpha b = 1/2, \qquad \beta b = 1/2.$$

Thus, we have three equations in four unknowns. This gives rise to a one-parameter family of solutions. Identifying the parameters:

$$c_1=\alpha, \quad a_{21}=\beta, \quad b_1=a, \quad b_2=b,$$

we obtain second-order Runge–Kutta methods.

Here are some two-stage second-order Runge–Kutta methods.

The Improved Euler's method can be written in the form of a two-stage explicit Runge–Kutta method (left) with its Butcher tableau (right):

$$\begin{aligned} k_1 &= hf(x_n,y_n)\\ k_2 &= hf\left(x_n+h, y_n+k_1\right)\\ y_{n+1} &= y_n+\frac{1}{2}(k_1+k_2) \end{aligned}$$

	$\mathbf{c}$	A	
k_1	0	0	
k_2	1	1	0
y_{n+1}	$\mathbf{b}^T$	1/2	1/2

This is Heun's method of order 2.

Other two-stage second-order methods are the mid-point method:

$$\begin{aligned} k_1 &= hf(x_n,y_n)\\ k_2 &= hf\left(x_n+\frac{1}{2}h, y_n+\frac{1}{2}k_1\right)\\ y_{n+1} &= y_n+k_2 \end{aligned}$$

	$\mathbf{c}$	A	
k_1	0	0	
k_2	1/2	1/2	0
y_{n+1}	$\mathbf{b}^T$	0	1

and Heun's method:

$$\begin{aligned} k_1 &= hf(x_n,y_n)\\ k_2 &= hf\left(x_n+\frac{2}{3}h, y_n+\frac{2}{3}k_1\right)\\ y_{n+1} &= y_n+\frac{1}{4}k_1+\frac{3}{4}k_2 \end{aligned}$$

	$\mathbf{c}$	A	
k_1	0	0	
k_2	2/3	2/3	0
y_{n+1}	$\mathbf{b}^T$	1/4	3/4

10.3.2. Third-order Runge–Kutta method. We list two common three-stage third-order Runge–Katta methods in their Butcher tableau, namely Heun's third-order formula and Kutta's third-order rule.

	$\mathbf{c}$	A		
k_1	0	0		
k_2	1/3	1/3	0	
k_3	2/3	0	2/3	0
y_{n+1}	$\mathbf{b}^T$	1/4	0	3/4

Butcher tableau of Heun's third-order formula.

	$\mathbf{c}$	A		
k_1	0	0		
k_2	1/2	1/2	0	
k_3	1	-1	2	0
y_{n+1}	$\mathbf{b}^T$	1/6	2/3	1/6

Butcher tableau of Kutta's third-order rule.

10.3.3. Fourth-order Runge–Kutta method. The fourth-order Runge–Kutta method (also known as the classic Runge–Kutta method or sometimes just as the Runge–Kutta method) is very popular amongst the explicit one-step methods.

By Taylor's Theorem, we have

$$y(x_{n+1}) = y(x_n)+y'(x_n)(x_{n+1}-x_n)+\frac{y''(x_n)}{2!}\,(x_{n+1}-x_n)^2+\frac{y^{(3)}(x_n)}{3!}\,(x_{n+1}-x_n)^3 + \frac{y^{(4)}(x_n)}{4!}\,(x_{n+1}-x_n)^4 + \frac{y^{(5)}(\xi_n)}{5!}\,(x_{n+1}-x_n)^5$$

for some ξ_n between x_n and x_{n+1} and $n = 0,\ 1,\ \ldots,\ N-1$. To obtain the fourth-order Runge–Kutta method, we can proceed as we did for the second-order Runge–Kutta methods. That is, we seek values of a, b, c, d, α_j and β_j such that

$$y'(x_n)(x_{n+1}-x_n) + \frac{y''(x_n)}{2!}\,(x_{n+1}-x_n)^2 + \frac{y^{(3)}(x_n)}{3!}\,(x_{n+1}-x_n)^3 + \frac{y^{(4)}(x_n)}{4!}\,(x_{n+1}-x_n)^4 + O(h^5)$$

is equal to

$$ak_1 + bk_2 + ck_3 + dk_4 + O(h^5),$$

where

$$\begin{aligned} k_1 &= hf(x_n, y_n),\\ k_2 &= hf(x_n + \alpha_1\, h, y_n + \beta_1 k_1),\\ k_3 &= hf(x_n + \alpha_2\, h, y_n + \beta_2 k_2),\\ k_4 &= hf(x_n + \alpha_3\, h, y_n + \beta_3 k_3). \end{aligned}$$

This follows from the relations

$$\begin{aligned} x_{n+1} - x_n &= h,\\ y'(x_n) &= f(x_n, y(x_n)),\\ y''(x_n) &= \frac{d}{d\,x}\, f(x, y(x))|_{t=x_n}\\ &= f_x(x_n, y(x_n)) + f_y(x_n, y(x_n))\, f(x_n, y(x_n)), \ldots, \end{aligned}$$

and Taylor's Theorem for functions of two variables. The lengthy computation is omitted.

The (classic) four-stage Runge–Kutta method of order 4 given by its formula (left) and, conveniently, in the form of a Butcher tableau (right).

TABLE 10.4. Numerical results for Example 10.5.

x_n	y_n	$y(x_n)$	Absolute error	Relative error
1.00	1.0000	1.0000	0.0000	0.0
1.10	1.2337	1.2337	0.0000	0.0
1.20	1.5527	1.5527	0.0000	0.0
1.30	1.9937	1.9937	0.0000	0.0
1.40	2.6116	2.6117	0.0001	0.0
1.50	3.4902	3.4904	0.0002	0.0

$$
\begin{aligned}
k_1 &= hf(x_n, y_n) \\
k_2 &= hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right) \\
k_3 &= hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_2\right) \\
k_4 &= hf\left(x_n + h, y_n + k_3\right) \\
y_{n+1} &= y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)
\end{aligned}
$$

	$\mathbf{c}$	A			
k_1	0	0			
k_2	1/2	1/2	0		
k_3	1/2	0	1/2	0	
k_4	1	0	0	1	0
y_{n+1}	$\mathbf{b}^T$	1/6	2/6	2/6	1/6

Essentially, the method is making four Δy estimates based on slopes at the left end, midpoint and right end of the subinterval. A weighted average of these Δy estimates, the two midpoint estimates weighted more than those at the left and right ends, is added to the previous y value.

The next example shows that the fourth-order Runge–Kutta method yields better results for (10.6) than the previous methods.

EXAMPLE 10.5. Use the fourth-order Runge–Kutta method with $h = 0.1$ to approximate the solution to the initial value problem of Example 10.2,

$$y'(x) = 2xy, \qquad y(1) = 1,$$

on the interval $1 \le x \le 1.5$.

SOLUTION. We have $f(x, y) = 2xy$ and

$$x_n = 1.0 + 0.1n, \qquad \text{for} \quad n = 0, 1, \ldots, 5.$$

With the starting value $y_0 = 1.0$, the approximation y_n to $y(x_n)$ is given by the scheme

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2\,k_2 + 2\,k_3 + k_4)$$

where

$$
\begin{aligned}
k_1 &= 0.1 \times 2(1.0 + 0.1n)y_n, \\
k_2 &= 0.1 \times 2(1.05 + 0.1n)(y_n + k_1/2), \\
k_3 &= 0.1 \times 2(1.05 + 0.1n)(y_n + k_2/2), \\
k_4 &= 0.1 \times 2(1.0 + 0.1(n+1))(y_n + k_3),
\end{aligned}
$$

and $n = 0, 1, 2, 3, 4$. The numerical results are listed in Table 10.4. These results are much better than all those previously obtained. □

EXAMPLE 10.6. Consider the initial value problem

$$y' = (y - x - 1)^2 + 2, \qquad y(0) = 1.$$

Compute y_4 by means of Runge–Kutta's method of order 4 with step size $h = 0.1$.

SOLUTION. The solution is given in tabular form.

n	x_n	y_n	Exact value $y(x_n)$	Global error $y(x_n) - y_n$
0	0.0	1.000 000 000	1.000 000 000	0.000 000 000
1	0.1	1.200 334 589	1.200 334 672	0.000 000 083
2	0.2	1.402 709 878	1.402 710 036	0.000 000 157
3	0.3	1.609 336 039	1.609 336 250	0.000 000 181
4	0.4	1.822 792 993	1.822 793 219	0.000 000 226

□

EXAMPLE 10.7. Use the Runge–Kutta method of order 4 with $h = 0.01$ to obtain a six-decimal approximation for the initial value problem

$$y' = x + \arctan y, \quad y(0) = 0,$$

on $0 \le x \le 1$. Print every tenth value and plot the numerical solution.

SOLUTION. **The Matlab numeric solution.**— The M-file `exp5_7` for Example 10.7 is

```
function yprime = exp5_7(x,y); % Example 5.7.
yprime = x+atan(y);
```

The Runge–Kutta method of order 4 is applied to the given differential equation:

```
clear
h = 0.01; x0= 0; xf= 1; y0 = 0;
n = ceil((xf-x0)/h); % number of steps
%
count = 2; print_time = 10; % when to write to output
x = x0; y = y0; % initialize x and y
output = [0 x0 y0];
for i=1:n
 k1 = h*exp5_7(x,y);
 k2 = h*exp5_7(x+h/2,y+k1/2);
 k3 = h*exp5_7(x+h/2,y+k2/2);
 k4 = h*exp5_7(x+h,y+k3);
 z  = y + (1/6)*(k1+2*k2+2*k3+k4);
 x = x + h;
   if count > print_time
      output = [output; i x z];
      count = count - print_time;
   end
 y = z;
count = count + 1;
end
```

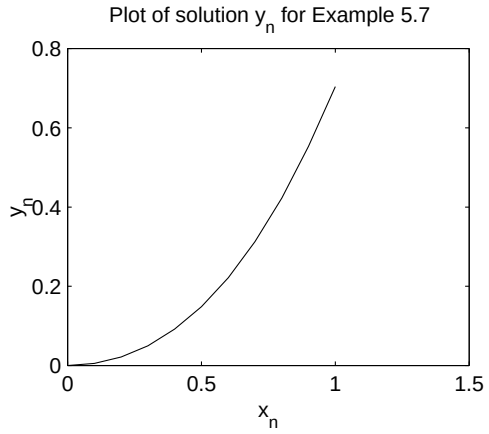

FIGURE 10.2. Graph of numerical solution of Example 10.7.

```
output
save output %for printing the graph
```

The command `output` prints the values of n, x, and y.

```
      n         x         y
        0         0         0
   10.0000    0.1000    0.0052
   20.0000    0.2000    0.0214
   30.0000    0.3000    0.0499
   40.0000    0.4000    0.0918
   50.0000    0.5000    0.1486
   60.0000    0.6000    0.2218
   70.0000    0.7000    0.3128
   80.0000    0.8000    0.4228
   90.0000    0.9000    0.5531
  100.0000    1.0000    0.7040
```

The following commands print the output.

```
load output;
subplot(2,2,1); plot(output(:,2),output(:,3));
title('Plot of solution y_n for Example 5.7');
xlabel('x_n'); ylabel('y_n');
```

□

In the next example, the Runge–Kutta method of order 4 is used to solve the van der Pol system of two equations. This system is also solved by means of the Matlab `ode23` code and the graphs of the two solutions are compared.

EXAMPLE 10.8. Use the Runge–Kutta method of order 4 with fixed step size $h = 0.1$ to solve the second-order van der Pol equation

$$y'' + \big(y^2 - 1\big)y' + y = 0, \qquad y(0) = 0, \qquad y'(0) = 0.25, \tag{10.11}$$

on $0 \le x \le 20$, print every tenth value, and plot the numerical solution. Also, use the `ode23` code to solve (10.11) and plot the solution.

SOLUTION. We first rewrite problem (10.11) as a system of two first-order differential equations by putting $y_1 = y$ and $y_2 = y_1'$,

$$\begin{aligned} y_1' &= y_2, \\ y_2' &= y_2(1 - y_1^2) - y_1, \end{aligned}$$

with initial conditions $y_1(0) = 0$ and $y_2(0) = 0.25$.

Our MATLAB program will call the MATLAB function M-file `exp1vdp.m`:

```
function yprime = exp1vdp(t,y); % Example 5.8.
yprime = [y(2); y(2).*(1-y(1).^2)-y(1)]; % van der Pol system
```

The following program applies the Runge–Kutta method of order 4 to the differential equation defined in the M-file `exp1vdp.m`:

```
clear
h = 0.1; t0= 0; tf= 21; % step size, initial and final times
y0 = [0 0.25]'; % initial conditions
n = ceil((xf-t0)/h); % number of steps

count = 2; print_control = 10; % when to write to output
t = t0; y = y0; % initialize t and y
output = [t0 y0']; % first row of matrix of printed values
w = [t0, y0']; % first row of matrix of plotted values
for i=1:n
 k1 = h*exp1vdp(x,y);            k2 = h*exp1vdp(x+h/2,y+k1/2);
 k3 = h*exp1vdp(x+h/2,y+k2/2);   k4 = h*exp1vdp(x+h,y+k3);
 z  = y + (1/6)*(k1+2*k2+2*k3+k4);
 t = t + h;
   if count > print_control
      output = [output; t z']; % augmenting matrix of printed values
      count = count - print_control;
   end
y = z;
w = [w; t z']; % augmenting matrix of plotted values
count = count + 1;
end
[output(1:11,:) output(12:22,:)] % print numerical values of solution
save w % save matrix to plot the solution
```

The command `output` prints the values of t, y_1, and y_2.

```
      t        y(1)     y(2)       t        y(1)     y(2)

      0          0    0.2500   11.0000   -1.9923   -0.2797
 1.0000     0.3586    0.4297   12.0000   -1.6042    0.7195
 2.0000     0.6876    0.1163   13.0000   -0.5411    1.6023
 3.0000     0.4313   -0.6844   14.0000    1.6998    1.6113
 4.0000    -0.7899   -1.6222   15.0000    1.8173   -0.5621
 5.0000    -1.6075    0.1456   16.0000    0.9940   -1.1654
```

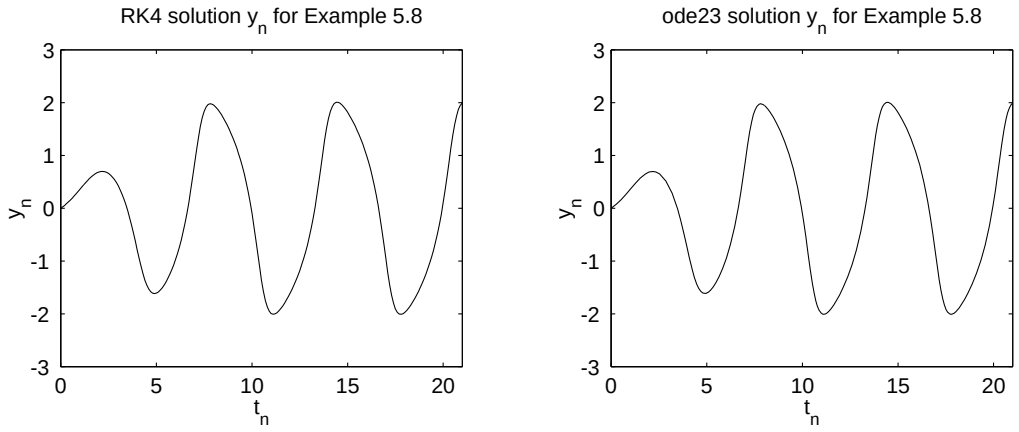

FIGURE 10.3. Graph of numerical solution of Example 10.8.

```
 6.0000   -0.9759    1.0662   17.0000   -0.9519   -2.6628
 7.0000    0.8487    2.5830   18.0000   -1.9688    0.3238
 8.0000    1.9531   -0.2733   19.0000   -1.3332    0.9004
 9.0000    1.3357   -0.8931   20.0000    0.1068    2.2766
10.0000   -0.0939   -2.2615   21.0000    1.9949    0.2625
```

The following commands graph the solution.

```
load w % load values to produce the graph
subplot(2,2,1);  plot(w(:,1),w(:,2));  % plot RK4 solution
title('RK4 solution y_n for Example 5.8');  xlabel('t_n');  ylabel('y_n');
```

We now use the `ode23` code. The command

```
load w % load values to produce the graph
v = [0 21 -3 3 ]; % set t and y axes
subplot(2,2,1);
plot(w(:,1),w(:,2)); % plot RK4 solution
axis(v);
title('RK4 solution y_n for Example 5.8'); xlabel('t_n'); ylabel('y_n');
subplot(2,2,2);
[t,y] = ode23('exp1vdp',[0 21], y0);
plot(x,y(:,1)); % plot ode23 solution
axis(v);
title('ode23 solution y_n for Example 5.8'); xlabel('t_n'); ylabel('y_n');
```

The code `ode23` produces three vectors, namely `t` of (144 unequally-spaced) nodes and corresponding solution values `y(1)` and `y(2)`, respectively. The left and right parts of Fig. 9.3 show the plots of the solutions obtained by RK4 and `ode23`, respectively. It is seen that the two graphs are identical. □

10.4. Convergence of Numerical Methods

In this and the next sections, we introduce the concepts of convergence, consistency and stability of numerical ode solvers.

The numerical methods considered in this chapter can be written in the general form

$$\sum_{n=0}^{k} \alpha_j y_{n+j} = h\varphi_f(y_{n+k}, y_{n+k-1}, \ldots, y_n, x_n; h). \tag{10.12}$$

where the subscript f to φ indicates the dependance of φ on the function $f(x, y)$ of (10.1). We impose the condition that

$$\varphi_{f\equiv 0}(y_{n+k}, y_{n+k-1}, \ldots, y_n, x_n; h) \equiv 0,$$

and note that the Lipschitz continuity of φ with respect to y_{n+j}, $n = 0, 1, \ldots, k$, follows from the Lipschitz continuity (10.2) of f.

DEFINITION 10.3. Method (10.12) with appropriate starting values is said to be *convergent* if, for all initial value problems (10.1), we have

$$y_n - y(x_n) \to 0 \qquad \text{as} \quad h \downarrow 0,$$

where $nh = x$ for all $x \in [a, b]$.

The *local truncation error* of (10.12) is the residual

$$R_{n+k} := \sum_{n=0}^{k} \alpha_j y(x_{n+j}) - h\varphi_f(y(x_{n+k}), y(x_{n+k-1}), \ldots, y(x_n), x_n; h). \tag{10.13}$$

DEFINITION 10.4. Method (10.12) with appropriate starting values is said to be *consistent* if, for all initial value problems (10.1), we have

$$\frac{1}{h} R_{n+k} \to 0 \quad \text{as} \quad h \downarrow 0,$$

where $nh = x$ for all $x \in [a, b]$.

DEFINITION 10.5. Method (10.12) is *zero-stable* if the roots of the characteristic polynomial

$$\sum_{n=0}^{k} \alpha_j r^{n+j}$$

lie inside or on the boundary of the unit disk, and those on the unit circle are simple.

We finally can state the following fundamental theorem.

THEOREM 10.2. *A method is convergent as $h \downarrow 0$ if and only if it is zero-stable and consistent.*

All numerical methods considered in this chapter are convergent.

10.5. Absolutely Stable Numerical Methods

We now turn attention to the application of a consistent and zero-stable numerical solver with small but nonvanishing step size.

For $n = 0, 1, 2, \ldots$, let y_n be the numerical solution of (10.1) at $x = x_n$, and $y^{[n]}(x_{n+1})$ be the exact solution of the *local* problem:

$$y' = f(x, y), \qquad y(x_n) = y_n. \tag{10.14}$$

A numerical method is said to have *local error*,

$$\varepsilon_{n+1} = y_{n+1} - y^{[n]}(x_{n+1}). \tag{10.15}$$

If we assume that $y(x) \in C^{p+1}[x_0, x_N]$ and

$$\varepsilon_{n+1} \approx C_{p+1} h_{n+1}^{p+1} y^{(p+1)}(x_n) + O(h_{n+1}^{p+2}), \tag{10.16}$$

then we say that the local error is of order $p+1$ and C_{p+1} is the error constant of the method. For consistent and zero-stable methods, the global error is of order p whenever the local error is of order $p+1$. In such case, we say that the method is of order p. We remark that a method of order $p \geq 1$ is consistent according to Definition 10.4.

Let us now apply the solver (10.12), with its small nonvanishing parameter h, to the linear test equation

$$y' = \lambda y, \qquad \Re\lambda < 0. \tag{10.17}$$

The *region of absolute stability*, R, is that region in the complex $\widehat{h}$-plane, where $\widehat{h} = h\lambda$, for which the numerical solution y_n of (10.17) goes to zero, as n goes to infinity.

The region of absolute stability of the explicit Euler method is the disk of radius 1 and center $(-1, 0)$, see curve $k = 1$ in Fig. 10.7. The region of stability of the implicit backward Euler method is the outside of the disk of radius 1 and center $(1, 0)$, hence it contains the left half-plane, see curve $k = 1$ in Fig. 10.10.

The region of absolute stability, R, of an explicit method is very roughly a disk or cardioid in the left half-plane (the cardioid overlaps with the right half-plane with a cusp at the origin). The boundary of R cuts the real axis at α, where $-\infty < \alpha < 0$, and at the origin. The interval $[\alpha, 0]$ is called the *interval of absolute stability*. For methods with real coefficients, R is symmetric with respect to the real axis. All methods considered in this work have real coefficients; hence Figs. 10.7, 10.8 and 10.10, below, show only the upper half of R.

The region of stability, R, of implicit methods extends to infinity in the left half-plane, that is $\alpha = -\infty$. The angle subtended at the origin by R in the left half-plane is usually smaller for higher order methods, see Fig. 10.10.

If the region R does not include the whole negative real axis, that is, $-\infty < \alpha < 0$, then the inclusion

$$h\lambda \in R$$

restricts the step size:

$$\alpha \leq h\,\Re\lambda \implies 0 < h \leq \frac{\alpha}{\Re\lambda}.$$

In practice, we want to use a step size h small enough to ensure accuracy of the numerical solution as implied by (10.15)–(10.16), but not too small.

10.6. Stability of Runge–Kutta Methods

There are stable s-stage explicit Runge-Kutta methods of order $p = s$ for $s = 1, 2, 3, 4$. The minimal number of stages of a stable explicit Runge-Kutta method of order 5 is 6.

Applying a Runge-Kutta method to the test equation,

$$y' = \lambda y, \qquad \Re\lambda < 0,$$

with solution $y(x) \to 0$ as $t \to \infty$, one obtains a one-step difference equation of the form

$$y_{n+1} = Q(\widehat{h}) y_n, \qquad \widehat{h} = h\lambda,$$

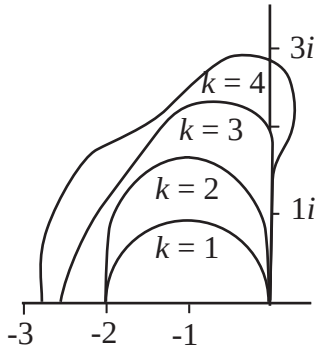

FIGURE 10.4. Region of absolute stability of s-stage explicit Runge–Kutta methods of order $k = s$.

where $Q(\widehat{h})$ is the *stability function* of the method. We see that $y_n \to 0$ as $n \to \infty$ if and only if

$$|Q(\widehat{h})| < 1, \tag{10.18}$$

and the method is *absolutely stable* for those values of $\widehat{h}$ in the complex plane for which (10.18) holds; those values form the *region of absolute stability* of the method. It can be shown that the stability function of explicit s-stage Runge-Kutta methods of order $p = s$, $s = 1, 2, 3, 4$, is

$$R(\widehat{h}) = \frac{y_{n+1}}{y_n} = 1 + \widehat{h} + \frac{1}{2!}\,\widehat{h}^2 + \cdots + \frac{1}{s!}\,\widehat{h}^s.$$

The regions of absolute stability, R, of s-stage explicit Runge–Kutta methods of order $k = s$, for $s = 1, 2, 3, 4$, are the interiors of the closed regions whose upper halves are shown in Fig. 10.4. The left-most point α of R is -2, -2, 2.51 and -2.78 for the methods of order $s = 1$, 2, 3 and 4, respectively

Fixed stepsize Runge–Kutta methods of order 1 to 5 are implemented in the following Matlab function M-files which are found in
`ftp://ftp.cs.cornell.edu/pub/cv`.

```
  function [tvals,yvals] = FixedRK(fname,t0,y0,h,k,n)
%
% Produces approximate solution to the initial value problem
%
%      y'(t) = f(t,y(t))     y(t0) = y0
%
% using a strategy that is based upon a k-th order
% Runge-Kutta method. Stepsize is fixed.
%
% Pre:  fname = string that names the function f.
%       t0 = initial time.
%       y0 = initial condition vector.
%       h  = stepsize.
%       k  = order of method. (1<=k<=5).
%       n  = number of steps to be taken,
%
```

```
% Post: tvals(j) = t0 + (j-1)h, j=1:n+1
%       yvals(:j) = approximate solution at t = tvals(j), j=1:n+1
%
   tc = t0;
   yc = y0;
   tvals = tc;
   yvals = yc;
   fc = feval(fname,tc,yc);
   for j=1:n
          [tc,yc,fc] = RKstep(fname,tc,yc,fc,h,k);
      yvals = [yvals yc ];
      tvals = [tvals tc];
   end

  function [tnew,ynew,fnew] = RKstep(fname,tc,yc,fc,h,k)
%
% Pre:  fname is a string that names a function of the form f(t,y)
%       where t is a scalar and y is a column d-vector.
%
%       yc is an approximate solution to y'(t) = f(t,y(t)) at t=tc.
%
%       fc = f(tc,yc).
%
%       h is the time step.
%
%       k is the order of the Runge-Kutta method used, 1<=k<=5.
%
% Post: tnew=tc+h, ynew is an approximate solution at t=tnew, and
%       fnew = f(tnew,ynew).

   if k==1
      k1 = h*fc;
      ynew = yc + k1;

   elseif k==2
      k1 = h*fc;
      k2 = h*feval(fname,tc+h,yc+k1);
      ynew  = yc + (k1 + k2)/2;

   elseif k==3
      k1 = h*fc;
      k2 = h*feval(fname,tc+(h/2),yc+(k1/2));
      k3 = h*feval(fname,tc+h,yc-k1+2*k2);
      ynew  = yc + (k1 + 4*k2 + k3)/6;

   elseif k==4
      k1 = h*fc;
      k2 = h*feval(fname,tc+(h/2),yc+(k1/2));
      k3 = h*feval(fname,tc+(h/2),yc+(k2/2));
```

```
    k4 = h*feval(fname,tc+h,yc+k3);
    ynew  = yc + (k1 + 2*k2 + 2*k3 + k4)/6;

elseif k==5
    k1 = h*fc;
    k2 = h*feval(fname,tc+(h/4),yc+(k1/4));
    k3 = h*feval(fname,tc+(3*h/8),yc+(3/32)*k1
                          +(9/32)*k2);
    k4 = h*feval(fname,tc+(12/13)*h,yc+(1932/2197)*k1
                     -(7200/2197)*k2+(7296/2197)*k3);
    k5 = h*feval(fname,tc+h,yc+(439/216)*k1
                 - 8*k2 + (3680/513)*k3 -(845/4104)*k4);
    k6 = h*feval(fname,tc+(1/2)*h,yc-(8/27)*k1
         + 2*k2 -(3544/2565)*k3 + (1859/4104)*k4 - (11/40)*k5);
    ynew  = yc + (16/135)*k1 + (6656/12825)*k3 +
                 (28561/56430)*k4 - (9/50)*k5 + (2/55)*k6;
end
tnew = tc+h;
fnew = feval(fname,tnew,ynew);
```

10.7. Embedded Pairs of Runge–Kutta Methods

Thus far, we have only considered a constant step size h. In practice, it is advantageous to let h vary so that h is taken larger when $y(x)$ does not vary rapidly and smaller when $y(x)$ changes rapidly. We turn to this problem.

Embedded pairs of Runge–Kutta methods of orders p and $p+1$ have built-in local error and step-size controls by monitoring the difference between the higher and lower order solutions, $y_{n+1} - \widehat{y}_{n+1}$. Some pairs include an interpolant which is used to interpolate the numerical solution between the nodes of the numerical solution and also, in some cases, to control the step-size.

10.7.1. Matlab's four-stage RK pair ode23. The code `ode23` consists of a four-stage pair of embedded explicit Runge–Kutta methods of orders 2 and 3 with error control. It advances from y_n to y_{n+1} with the third-order method (so called local extrapolation) and controls the local error by taking the difference between the third-order and the second-order numerical solutions. The four stages are:

$$\begin{aligned}
k_1 &= h\, f(x_n, y_n),\\
k_2 &= h\, f(x_n + (1/2)h, y_n + (1/2)k_1),\\
k_3 &= h\, f(x_n + (3/4)h, y_n + (3/4)k_2),\\
k_4 &= h\, f(x_n + h, y_n + (2/9)k_1 + (1/3)k_2 + (4/9)k_3),
\end{aligned}$$

The first three stages produce the solution at the next time step:

$$y_{n+1} = y_n + \frac{2}{9}\, k_1 + \frac{1}{3}\, k_2 + \frac{4}{9}\, k_3,$$

and all four stages give the local error estimate:

$$E = -\frac{5}{72}\, k_1 + \frac{1}{12}\, k_2 + \frac{1}{9}\, k_3 - \frac{1}{8}\, k_4.$$

However, this is really a three-stage method since the first step at x_{n+1} is the same as the last step at x_n, that is $k_1^{[n+1]} = k_4^{[n]}$. Such methods are called FSAL methods.

The natural interpolant used in `ode23` is the two-point Hermite polynomial of degree 3 which interpolates y_n and $f(x_n, y_n)$ at $x = x_n$, and y_{n+1} and $f(x_{n+1}, x_{n+1})$ at $t = x_{n+1}$.

EXAMPLE 10.9. Use Matlab's four-stage FSAL ode23 method with $h = 0.1$ to approximate $y(0.1)$ and $y(0.2)$ to 5 decimal places and estimate the local error for the initial value problem

$$y' = xy + 1, \qquad y(0) = 1.$$

SOLUTION. The right-hand side of the differential equation is

$$f(x, y) = xy + 1.$$

With $n = 0$:

$$\begin{aligned}
k_1 &= 0.1 \times 1 = 0.1 \\
k_2 &= 0.1 \times (0.05 \times 1.05 + 1) = 0.105\,25 \\
k_3 &= 0.1 \times (0.75 \times 1.078\,937\,5 + 1) = 0.108\,092\,031\,25 \\
k_4 &= 0.1 \times (0.1 \times 1.105\,346\,458\,333\,33 + 1) = 0.111\,053\,464\,583\,33 \\
y_1 &= 1.105\,346\,458\,333\,33
\end{aligned}$$

The estimate of the local error is

$$\text{Local error estimate} = -4.506\,848\,958\,333\,448e - 05$$

With $n = 1$:

$$\begin{aligned}
k_1 &= 0.111\,053\,464\,583\,33 \\
k_2 &= 0.117\,413\,097\,859\,37 \\
k_3 &= 0.120\,884\,609\,930\,24 \\
k_4 &= 0.124\,457\,783\,972\,15 \\
y_2 &= 1.222\,889\,198\,607\,30
\end{aligned}$$

The estimate of the local error is

$$\text{Local error estimate} = -5.322\,100\,094\,209\,102e - 05$$

To use the numeric Matlab command `ode23` to solve and plot the given initial value problem on $[0, 1]$, one writes the function M-file `exp5_9.m`:

```
function yprime = exp5_9(x,y)
yprime = x.*y+1;
```

and use the commands

```
clear
xspan = [0 1]; y0 = 1; % xspan and initial value
[x,y] = ode23('exp5_9',xspan,y0);
subplot(2,2,1); plot(x,y); xlabel('x'); ylabel('y');
```

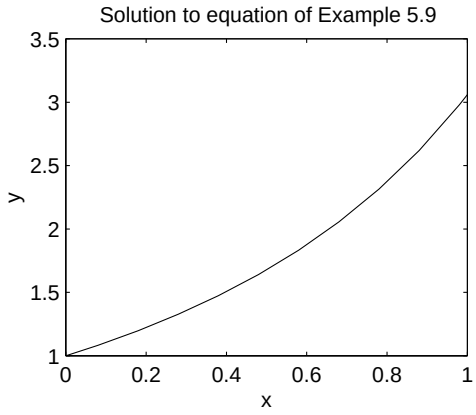

FIGURE 10.5. Graph of numerical solutions of Example 10.9.

```
title('Solution to equation of Example 5.9');
print -deps2 Figexp5_9 % print figure to file Fig.exp5.9
```

□

The Matlab solver `ode23` is an implementation of the explicit Runge–Kutta (2,3) pair of Bogacki and Shampine called BS23. It uses a "free" interpolant of order 3. Local extrapolation is done, that is, the higher-order solution, namely of order 3, is used to avance the solution.

10.7.2. Seven-stage Dormand–Prince pair DP(5,4)7M with interpolant. The seven-stage Dormand–Prince pair DP(5,4)7M with local error estimate and interpolant is presented in a Butcher tableau. The number 5 in the designation DP(5,4)7M means that the solution is advanced with the solution y_{n+1} of order five (a procedure called *local extrapolation*). The number 4 means that the solution $\widehat{y}_{n+1}$ of order four is used to obtain the local error estimate by means of the difference $y_{n+1} - \widehat{y}_{n+1}$. In fact, $\widehat{y}_{n+1}$ is not computed; rather the coefficients in the line $b^T - \widehat{b}^T$ are used to obtain the local error estimate. The number 7 means that the method has seven stages. The letter M means that the constant C_6 in the top-order error term has been minimized, while maintaining stability. Six stages are necessary for the method of order 5. The seventh stage is necessary to have an interpolant. The last line of the tableau is used to produce an interpolant.

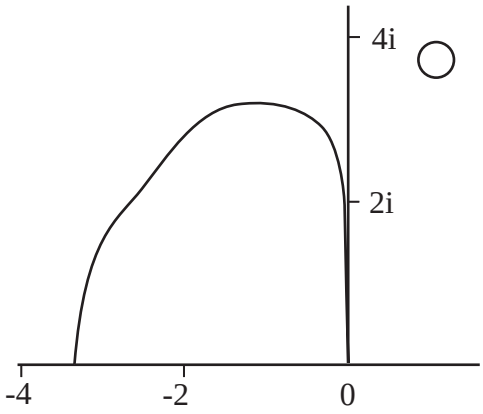

FIGURE 10.6. Region of absolute stability of the Dormand-Prince pair DP(5,4)7M.

	$\mathbf{c}$	A						
k_1	0	0						
k_2	$\frac{1}{5}$	$\frac{1}{5}$	0					
k_3	$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$	0				
k_4	$\frac{4}{5}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$	0			
k_5	$\frac{8}{9}$	$\frac{19372}{6561}$	$-\frac{25360}{2187}$	$\frac{64448}{6561}$	$-\frac{212}{729}$	0		
k_6	1	$\frac{9017}{3168}$	$-\frac{355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$-\frac{5103}{18656}$	0	
k_7	1	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	
$\widehat{y}_{n+1}$	$\widehat{\mathbf{b}}^T$	$\frac{5179}{57600}$	0	$\frac{7571}{16695}$	$\frac{393}{640}$	$-\frac{92097}{339200}$	$\frac{187}{2100}$	$\frac{1}{40}$
y_{n+1}	$\mathbf{b}^T$	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
$\mathbf{b}^T - \widehat{\mathbf{b}}^T$		$\frac{71}{57\,600}$	0	$-\frac{71}{16\,695}$	$\frac{71}{1\,920}$	$-\frac{17\,253}{339\,200}$	$\frac{22}{525}$	$-\frac{1}{40}$
$y_{n+0.5}$		$\frac{5783653}{57600000}$	0	$\frac{466123}{1192500}$	$-\frac{41347}{1920000}$	$\frac{16122321}{339200000}$	$-\frac{7117}{20000}$	$\frac{183}{10000}$

(10.19)

Seven-stage Dormand–Prince pair DP(5,4)7M of order 5 and 4.

This seven-stage FSAL method reduces, in practice, to a six-stage method since $k_1^{[n+1]} = k_7^{[n]}$; in fact the row vector b^T is the same as the 7-th line corresponding to k_7.

The interval of absolute stability of the pair DP(5,4)7M is approximately $(-3.3, 0)$ (see Fig. 10.6).

One notices that the matrix A in the Butcher tableau of an explicit Rung–Kutta method is strictly lower triangular. Semi-explicit methods have a lower triangular matrix. Otherwise, the method is implicit. Solving semi-explicit methods for the vector solution y_{n+1} of a system is much cheaper than solving implicit methods.

Runge–Kutta methods constitute a clever and sensible idea. The unique solution of a well-posed initial value problem is a single curve in $\mathbb{R}^{n+1}$, but due

to truncation and roundoff error, any numerical solution is, in fact, going to wander off that integral curve, and the numerical solution is inevitably going to be affected by the behavior of neighboring curves. Thus, it is the behavior of the family of integral curves, and not just that of the unique solution curve, that is of importance. Runge–Kutta methods deliberately try to gather information about this family of curves, as it is most easily seen in the case of explicit Runge–Kutta methods.

The Matlab solver `ode45` is an implementation of the explicit Runge–Kutta (5,4) pair of Dormand and Prince called variously RK5(4)7FM, DOPRI5, DP(4,5) and DP54. It uses a "free" interpolant of order 4 communicated privately by Dormand and Prince. Local extrapolation is done.

Details on Matlab solvers `ode23`, `ode45` and other solvers can be found in *The MATLAB ODE Suite*, L. F. Shampine and M. W. Reichelt, SIAM Journal on Scientific Computing, **18**(1), 1997.

10.7.3. Six-stage Runge–Kutta–Fehlberg pair RKF(4,5). The six-stage Runge–Kutta–Fehlberg pair RKF(4,5) with local error estimate uses a method of order 4 to advance the numerical value from y_n to y_{n+1}, and a method of order 5 to obtain the auxiliary value $\widehat{y}_{n+1}$ which serves in computing the local error by means of the difference $y_{n+1} - \widehat{y}_{n+1}$. We present this method in a Butcher tableau. The estimated local error is obtained from the last line. The method of order 4 minimizes the local error.

$$
\begin{array}{llllllll}
k_1 & 0 & 0 & & & & & \\
k_2 & \frac{1}{4} & \frac{1}{4} & 0 & & & & \\
k_3 & \frac{3}{8} & \frac{3}{32} & \frac{9}{32} & 0 & & & \\
k_4 & \frac{12}{13} & \frac{1932}{2197} & -\frac{7200}{2197} & \frac{7296}{2197} & 0 & & \\
k_5 & 1 & \frac{439}{216} & -8 & \frac{3680}{513} & -\frac{845}{4104} & 0 & \\
k_6 & \frac{1}{2} & -\frac{8}{27} & 2 & -\frac{3544}{2565} & \frac{1859}{4104} & -\frac{11}{40} & 0 \\
\hline
\frac{2197}{4104} & -\frac{1}{5} & 0 & & & & & \\
\widehat{y}_{n+1} & \widehat{\mathbf{b}}^T & \frac{16}{135} & 0 & \frac{6656}{12825} & \frac{28561}{56430} & -\frac{9}{50} & \frac{2}{55} \\
 & \widehat{\mathbf{b}}^T - \mathbf{b}^T & \frac{1}{360} & 0 & -\frac{128}{4275} & -\frac{2197}{75240} & \frac{1}{50} & \frac{2}{55}
\end{array}
\tag{10.20}
$$

Six-stage Runge–Kutta–Fehlberg pair RKF(4,5) of order 4 and 5.

The interval of absolute stability of the pair RKF(4,5) is approximately $(-3.78, 0)$.

The pair RKF45 of order four and five minimizes the error constant C_5 of the lower order method which is used to advance the solution from y_n to y_{n+1}, that is, without using local extrapolation. The algorithm follows.

ALGORITHM 10.1. Let y_0 be the initial condition. Suppose that the approximation y_n to $y(x_n)$ has been computed and satisfies $|y(x_n) - y_n| < \epsilon$ where ϵ is the desired precision. Let $h > 0$.

(1) Compute two approximations for y_{n+1}: one using the fourth-order method

$$y_{n+1} = y_n + \left(\frac{25}{216} k_1 + \frac{1408}{2565} k_3 + \frac{2197}{4104} k_4 - \frac{1}{5} k_5 \right), \tag{10.21}$$

and the second using the fifth-order method,

$$\widehat{y}_{j+1} = y_n + \left(\frac{16}{135}k_1 + \frac{6656}{12825}k_3 + \frac{28561}{56430}k_4 - \frac{9}{50}k_5 + \frac{2}{55}k_6\right), \tag{10.22}$$

where

$$\begin{aligned}
k_1 &= hf(x_n, y_n),\\
k_2 &= hf(x_n + h/4, y_n + k_1/4),\\
k_3 &= hf(x_n + 3h/8, y_n + 3k_1/32 + 9k_2/32),\\
k_4 &= hf(x_n + 12h/13, y_n + 1932k_1/2197 - 7200k_2/2197 + 7296k_3/2197),\\
k_5 &= hf(x_n + h, y_n + 439k_1/216 - 8k_2 + 3680k_3/513 + 845k_4/4104),\\
k_6 &= hf(x_n + h/2, y_n - 8k_1/27 + 2k_2 + 3544k_3/2565 + 1859k_4/4104 - 11k_5/40).
\end{aligned}$$

(2) If $|\widehat{y}_{j+1} - y_{n+1}| < \epsilon h$, accept y_{n+1} as the approximation to $y(x_{n+1})$. Replace h by qh where

$$q = \left[\epsilon h/(2|\widehat{y}_{j+1} - y_{n+1}|)\right]^{1/4}$$

and go back to step (1) to compute an approximation for y_{j+2}.

(3) If $|\widehat{y}_{j+1} - y_{n+1}| \geq \epsilon h$, replace h by qh where

$$q = \left[\epsilon h/(2|\widehat{y}_{j+1} - y_{n+1}|)\right]^{1/4}$$

and go back to step (1) to compute the next approximation for y_{n+1}.

One can show that the local truncation error for (10.21) is approximately

$$|\widehat{y}_{j+1} - y_{n+1}|/h.$$

At step (2), one requires that this error be smaller than ϵh in order to get $|y(x_n) - y_n| < \epsilon$ for all j (and in particular $|y(x_N) - y_f| < \epsilon$). The formula to compute q in (2) and (3) (and hence a new value for h) is derived from the relation between the local truncation errors of (10.21) and (10.22).

RKF(4,5) overestimate the error in the order-four solution because its local error constant is minimized. The next method, RKV, corrects this fault.

10.7.4. Eight-stage Runge–Kutta–Verner pair RKV(5,6). The eight-stage Runge–Kutta–Verner pair RKV(5,6) of order 5 and 6 is presented in a Butcher tableau. Note that 8 stages are necessary to get order 6. The method attempts to keep the global error proportional to a user-specified tolerance. It is efficient for nonstiff systems where the derivative evaluations are not expensive and where the solution is not required at a large number of finely spaced points (as might be required for graphical output).

	$\mathbf{c}$	A							
k_1	0	0							
k_2	$\frac{1}{6}$	$\frac{1}{6}$	0						
k_3	$\frac{4}{15}$	$\frac{4}{75}$	$\frac{16}{75}$	0					
k_4	$\frac{2}{3}$	$\frac{5}{6}$	$-\frac{8}{3}$	$\frac{5}{2}$	0				
k_5	$\frac{5}{6}$	$-\frac{165}{64}$	$\frac{55}{6}$	$-\frac{425}{64}$	$\frac{85}{96}$	0			
k_6	1	$\frac{12}{5}$	-8	$\frac{4015}{612}$	$-\frac{11}{36}$	$\frac{88}{255}$	0		
k_7	$\frac{1}{15}$	$-\frac{8263}{15000}$	$\frac{124}{75}$	$-\frac{643}{680}$	$-\frac{81}{250}$	$\frac{2484}{10625}$	0		
k_8	1	$\frac{3501}{1720}$	$-\frac{300}{43}$	$\frac{297275}{52632}$	$-\frac{319}{2322}$	$\frac{24068}{84065}$	0	$\frac{3850}{26703}$	
y_{n+1}	$\mathbf{b}^T$	$\frac{13}{160}$	0	$\frac{2375}{5984}$	$\frac{5}{16}$	$\frac{12}{85}$	$\frac{3}{44}$		
$\widehat{y}_{n+1}$	$\widehat{\mathbf{b}}^T$	$\frac{3}{40}$	0	$\frac{875}{2244}$	$\frac{23}{72}$	$\frac{264}{1955}$	0	$\frac{125}{11592}$	$\frac{43}{616}$

(10.23)

Eight-stage Runge–Kutta–Verner pair RKV(5,6) of order 5 and 6.

10.8. Multistep Predictor-Corrector Methods

10.8.1. General multistep methods. Consider the initial value problem

$$y' = f(x, y), \qquad y(a) = \eta, \tag{10.24}$$

where $f(x, y)$ is continuous with respect to x and Lipschitz continuous with respect to y on the strip $[a, b] \times (-\infty, \infty)$. Then, by Theorem 10.1, the exact solution, $y(x)$, exists and is unique on $[a, b]$.

We look for an approximate numerical solution $\{y_n\}$ at the nodes $x_n = a+nh$ where h is the step size and $n = (b - a)/h$.
For this purpose, we consider the *k-step linear method*:

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f_{n+j}, \tag{10.25}$$

where $y_n \approx y(x_n)$ and $f_n := f(x_n, y_n)$. We normalize the method by the condition $\alpha_k = 1$ and insist that the number of steps be exactly k by imposing the condition

$$(\alpha_0, \beta_0) \neq (0, 0).$$

We choose k *starting values* $y_0, y_1, \ldots, y_{k-1}$, say, by means of a Runge–Kutta method of the same order.

The method is *explicit* if $\beta_k = 0$; in this case, we obtain y_{n+1} directly. The method is *implicit* if $\beta_k \neq 0$; in this case, we have to solve for y_{n+k} by the recurrence formula:

$$y_{n+k}^{[s+1]} = h\beta_k f\left(x_{n+k}, y_{n+k}^{[s]}\right) + g, \qquad y_{n+k}^{[0]} \text{ arbitrary}, \qquad s = 0, 1, \ldots, \tag{10.26}$$

where the function

$$g = g(x_n, \ldots, x_{n+k-1}, y_0, \ldots, y_{n+k-1})$$

contains only known values. The recurrence formula (10.26) converges as $s \to \infty$, if $0 \leq M < 1$ where M is the Lipschitz constant of the right-hand side of (10.26) with respect to y_{n+k}. If L is the Lipschitz constant of $f(x, y)$ with respect to y, then

$$M := Lh|\beta_k| < 1 \tag{10.27}$$

and the inequality

$$h < \frac{1}{L|\beta_k|}$$

implies convergence.

Applying (10.25) to the test equation,

$$y' = \lambda y, \qquad \Re\lambda < 0,$$

with solution $y(x) \to 0$ as $t \to \infty$, one finds that the numerical solution $y_n \to 0$ as $n \to \infty$ if the zeros, $r_s(\widehat{h})$, of the stability polynomial

$$\pi(r, \widehat{h}) := \sum_{n=0}^{k} (\alpha_j - \widehat{h}\beta_j) r^j$$

satisfy $|r_s(\widehat{h})| \le 1$, $s = 1, 2, \ldots, k$, $s = 1, 2, \ldots, k$, and $|r_s(\widehat{h})| < 1$ if $r_s(\widehat{h})$ is a multiple zero. In that case, we say that the linear multistep method (10.25) is *absolutely stable* for given $\widehat{h}$. The *region of absolute stability*, R, in the complex plane is the set of values of $\widehat{h}$ for with the method is absolutely stable.

10.8.2. Adams-Bashforth-Moulton linear multistep methods. Popular linear k-step methods are (explicit) Adams–Bashforth (AB) and (implicit) Adams–Moulton (AM) methods,

$$y_{n+1} - y_n = h \sum_{j=0}^{k-1} \beta_j^* f_{n+j-k+1}, \qquad y_{n+1} - y_n = h \sum_{j=0}^{k} \beta_j f_{n+j-k+1},$$

respectively. Tables 10.5 and 10.6 list the AB and AM methods of stepnumber 1 to 6, respectively. In the tables, the coefficients of the methods are to be divided by d, k is the stepnumber, p is the order, and C_{p+1}^* and C_{p+1} are the corresponding error constants of the methods.

TABLE 10.5. Coefficients of Adams–Bashforth methods of stepnumber 1–6.

β_5^*	β_4^*	β_3^*	β_2^*	β_1^*	β_0^*	d	k	p	C_{p+1}^*
					1	1	1	1	1/2
				3	−1	2	2	2	5/12
			23	−16	5	12	3	3	3/8
		55	−59	37	−9	24	4	4	251/720
	1901	−2774	1616	−1274	251	720	5	5	95/288
4277	−7923	9982	−7298	2877	−475	1440	6	6	19 087/60 480

The regions of absolute stability of k-step Adams–Bashforth and Adams–Moulton methods of order $k = 1, 2, 3, 4$, are the interiors of the closed regions whose upper halves are shown in the left and right parts, respectively, of Fig. 10.7. The region of absolute stability of the Adams–Bashforth method of order 3 extends in a small triangular region in the right half-plane. The region of absolute stability of the Adams–Moulton method of order 1 is the whole left half-plane.

TABLE 10.6. Coefficients of Adams–Moulton methods of step-number 1–6.

β_5	β_4	β_3	β_2	β_1	β_0	d	k	p	C_{p+1}
				1	1	2	1	2	$-1/12$
			5	8	-1	12	2	3	$-1/24$
		9	19	-5	1	24	3	4	$-19/720$
	251	646	-264	106	-19	720	4	5	$-3/160$
475	1427	-798	482	-173	27	1440	5	6	$-863/60\,480$

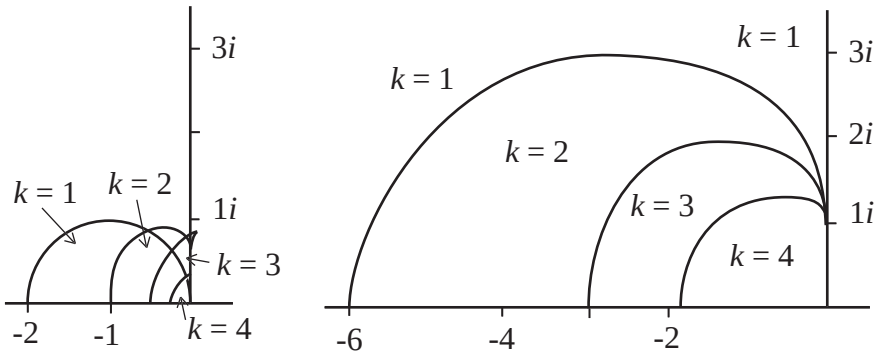

FIGURE 10.7. Left: Regions of absolute stability of k-step Adams–Bashforth methods. Right: Regions of absolute stability of k-step Adams–Moulton methods.

In practice, an AB method is used as a *predictor* to predict the next-step value y^*_{n+1}, which is then inserted in the right-hand side of an AM method used as a *corrector* to obtain the corrected value y_{n+1}. Such combination is called an ABM predictor-corrector which, when of the same order, comes with the Milne estimate for the principal local truncation error

$$\epsilon_{n+1} \approx \frac{C_{p+1}}{C^*_{p+1} - C_{p+1}} (y_{n+1} - y^*_{n+1}).$$

The procedure called *local approximation* improves the higher-order solution y_{n+1} by the addition of the error estimator, namely,

$$y_{n+1} + \frac{C_{p+1}}{C^*_{p+1} - C_{p+1}} (y_{n+1} - y^*_{n+1}).$$

The regions of absolute stability of kth-order Adams–Bashforth–Moulton pairs, for $k = 1, 2, 3, 4$, in *Predictor-Evaluation-Corrector-Evaluation* mode, denoted by PECE, are the interiors of the closed regions whose upper halves are shown in the left part of Fig. 10.8. The regions of absolute stability of kth-order Adams–Bashforth–Moulton pairs, for $k = 1, 2, 3, 4$, in the PECLE mode where L stands for local extrapolation, are the interiors of the closed regions whose upper halves are shown in the right part of Fig. 10.8.

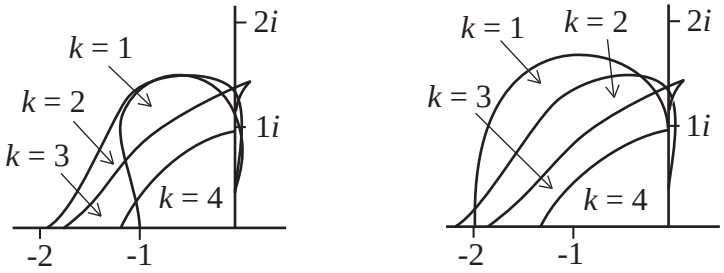

FIGURE 10.8. Regions of absolute stability of k-order Adams–Bashforth–Moulton methods, left in PECE mode, and right in PECLE mode.

10.8.3. Adams–Bashforth–Moulton methods of orders 3 and 4. As a first example of multistep methods, we consider the three-step Adams–Bashforth–Moulton method of order 3, given by the formula pair:

$$y^{\mathrm{P}}_{n+1} = y^{\mathrm{C}}_n + \frac{h}{12}\left(23f^{\mathrm{C}}_n - 16f^{\mathrm{C}}_{n-1} + 5f^{\mathrm{C}}_{n-2}\right), \qquad f^{\mathrm{C}}_k = f\left(x_k, y^{\mathrm{C}}_k\right), \tag{10.28}$$

$$y^{\mathrm{C}}_{n+1} = y^{\mathrm{C}}_n + \frac{h}{12}\left(5f^{\mathrm{P}}_{n+1} + 8f^{\mathrm{C}}_n - f^{\mathrm{C}}_{n-1}\right), \qquad f^{\mathrm{P}}_k = f\left(x_k, y^{\mathrm{P}}_k\right), \tag{10.29}$$

with local error estimate

$$\text{Err.} \approx -\frac{1}{10}\left[y^{\mathrm{C}}_{n+1} - y^{\mathrm{P}}_{n+1}\right]. \tag{10.30}$$

EXAMPLE 10.10. Solve to six decimal places the initial value problem

$$y' = x + \sin y, \qquad y(0) = 0,$$

by means of the Adams–Bashforth–Moulton method of order 3 over the interval $[0, 2]$ with $h = 0.2$. The starting values have been obtained by a high precision method. Use formula (10.30) to estimate the local error at each step.

SOLUTION. The solution is given in a table.

n	x_n	Starting y^C_n	Predicted y^P_n	Corrected y^C_n	$10^5\times$Local Error in y^C_n $\approx -(y^C_n - y^P_n)\times 10^4$
0	0.0	0.000 000 0			
1	0.2	0.021 404 7			
2	0.4	0.091 819 5			
3	0.6		0.221 260	0.221 977	− 7
4	0.8		0.423 703	0.424 064	− 4
5	1.0		0.710 725	0.709 623	11
6	1.2		1.088 004	1.083 447	46
7	1.4		1.542 694	1.533 698	90
8	1.6		2.035 443	2.026 712	87
9	1.8		2.518 039	2.518 431	− 4
10	2.0		2.965 994	2.975 839	−98

□

As a second and better known example of multistep methods, we consider the four-step Adams–Bashforth–Moulton method of order 4.

The *Adams–Bashforth predictor* and the *Adams–Moulton corrector* of order 4 are

$$y_{n+1}^{\mathrm{P}} = y_n^{\mathrm{C}} + \frac{h}{24}\left(55f_n^{\mathrm{C}} - 59f_{n-1}^{\mathrm{C}} + 37f_{n-2}^{\mathrm{C}} - 9f_{n-3}^{\mathrm{C}}\right) \tag{10.31}$$

and

$$y_{n+1}^{\mathrm{C}} = y_n^{\mathrm{C}} + \frac{h}{24}\left(9f_{n+1}^{\mathrm{P}} + 19f_n^{\mathrm{C}} - 5f_{n-1}^{\mathrm{C}} + f_{n-2}^{\mathrm{C}}\right), \tag{10.32}$$

where

$$f_n^{\mathrm{C}} = f(x_n, y_n^{\mathrm{C}}) \quad \text{and} \quad f_n^{\mathrm{P}} = f(x_n, y_n^{\mathrm{P}}).$$

Starting values are obtained with a Runge–Kutta method or otherwise.

The local error is controlled by means of the estimate

$$C_5 h^5 y^{(5)}(x_{n+1}) \approx -\frac{19}{270}\left[y_{n+1}^{\mathrm{C}} - y_{n+1}^{\mathrm{P}}\right]. \tag{10.33}$$

A certain number of past values of y_n and f_n are kept in memory in order to extend the step size if the local error is small with respect to the given tolerance. If the local error is too large with respect to the given tolerance, the step size can be halved by means of the following formulae:

$$y_{n-1/2} = \frac{1}{128}\left(35y_n + 140y_{n-1} - 70y_{n-2} + 28y_{n-3} - y_{n-4}\right), \tag{10.34}$$

$$y_{n-3/2} = \frac{1}{162}\left(-y_n + 24y_{n-1} + 54y_{n-2} - 16y_{n-3} + 3y_{n-4}\right). \tag{10.35}$$

In PECE mode, the Adams–Bashforth–Moulton pair of order 4 has interval of absolute stability equal to $(-1.25, 0)$, that is, the method does not amplify past errors if the step size h is sufficiently small so that

$$-1.25 < h\frac{\partial f}{\partial y} < 0, \quad \text{where} \quad \frac{\partial f}{\partial y} < 0.$$

Example 10.11. Consider the initial value problem

$$y' = x + y, \qquad y(0) = 0.$$

Compute the solution at $x = 2$ by the Adams–Bashforth–Moulton method of order 4 with $h = 0.2$. Use Runge–Kutta method of order 4 to obtain the starting values. Use five decimal places and use the exact solution to compute the global error.

Solution. The global error is computed by means of the exact solution

$$y(x) = e^x - x - 1.$$

We present the solution in the form of a table for starting values, predicted values, corrected values, exact values and global errors in the corrected solution.

n	x_n	Starting y_n^C	Predicted y_n^P	Corrected y_n^C	Exact $y(x_n)$	Error: $10^6\times$ $(y(x_n)-y_n^C)$
0	0.0	0.000 000			0.000 000	0
1	0.2	0.021 400			0.021 403	3
2	0.4	0.091 818			0.091 825	7
3	0.6	0.222 107			0.222 119	12
4	0.8		0.425 361	0.425 529	0.425 541	12
5	1.0		0.718 066	0.718 270	0.718 282	12
6	1.2		1.119 855	1.120 106	1.120 117	11
7	1.4		1.654 885	1.655 191	1.655 200	9
8	1.6		2.352 653	2.353 026	2.353 032	6
9	1.8		3.249 190	3.249 646	3.249 647	1
10	2.0		4.388 505	4.389 062	4.389 056	−6

We see that the method is stable since the error does not grow. □

EXAMPLE 10.12. Solve to six decimal places the initial value problem

$$y' = \arctan x + \arctan y, \qquad y(0) = 0,$$

by means of the Adams–Bashforth–Moulton method of order 3 over the interval $[0, 2]$ with $h = 0.2$. Obtain the starting values by Runge–Kutta 4. Use formula (10.30) to estimate the local error at each step.

SOLUTION. **The Matlab numeric solution.**— The M-file `exp5_12` for Example 10.12 is

```
function yprime = exp5_12(x,y); % Example 5.12.
yprime = atan(x)+atan(y);
```

The initial conditions and the Runge–Kutta method of order 4 is used to obtain the four starting values

```
clear
h = 0.2; x0= 0; xf= 2; y0 = 0;
n = ceil((xf-x0)/h); % number of steps
%
count = 2; print_time = 1; % when to write to output
x = x0; y = y0; % initialize x and y
output = [0 x0 y0 0];
%RK4
for i=1:3
 k1 = h*exp5_12(x,y);
 k2 = h*exp5_12(x+h/2,y+k1/2);
 k3 = h*exp5_12(x+h/2,y+k2/2);
 k4 = h*exp5_12(x+h,y+k3);
 z  = y + (1/6)*(k1+2*k2+2*k3+k4);
  x = x + h;
   if count > print_time
      output = [output; i x z 0];
      count = count - print_time;
   end
 y = z;
```

```
count = count + 1;
end
% ABM4
for i=4:n
 zp = y + (h/24)*(55*exp5_12(output(i,2),output(i,3))-...
                  59*exp5_12(output(i-1,2),output(i-1,3))+...
                  37*exp5_12(output(i-2,2),output(i-2,3))-...
                   9*exp5_12(output(i-3,2),output(i-3,3)) );
  z = y + (h/24)*( 9*exp5_12(x+h,zp)+...
                  19*exp5_12(output(i,2),output(i,3))-...
                   5*exp5_12(output(i-1,2),output(i-1,3))+...
                     exp5_12(output(i-2,2),output(i-2,3)) );
  x = x + h;
   if count > print_time
      errest = -(19/270)*(z-zp);
      output = [output; i x z errest];
      count = count - print_time;
   end
 y = z;
count = count + 1;
end
output
save output %for printing the graph
```

The command `output` prints the values of n, x, and y.

```
   n    x            y           Error estimate

   0    0            0                    0
   1   0.2   0.02126422549044                  0
   2   0.4   0.08962325332457                  0
   3   0.6   0.21103407185113                  0
   4   0.8   0.39029787517821   0.00001007608281
   5   1.0   0.62988482479868   0.00005216829834
   6   1.2   0.92767891924367   0.00004381671342
   7   1.4   1.27663327419538  -0.00003607372725
   8   1.6   1.66738483675693  -0.00008228934754
   9   1.8   2.09110753309673  -0.00005318684309
  10   2.0   2.54068815072267  -0.00001234568256
```

The following commands print the output.

```
load output;
subplot(2,2,1); plot(output(:,2),output(:,3));
title('Plot of solution y_n for Example 5.12');
xlabel('x_n'); ylabel('y_n');
```

□

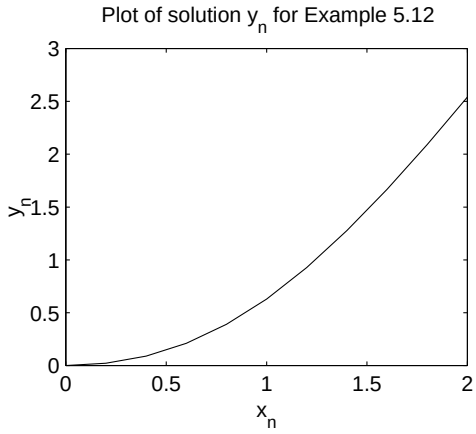

FIGURE 10.9. Graph of the numerical solution of Example 10.12.

Fixed stepsize Adams–Bashforth–Moulton methods of order 1 to 5 are implemented in the following Matlab function M-files which are found in `ftp://ftp.cs.cornell.edu/pub/cv`.

```
  function [tvals,yvals] = FixedPC(fname,t0,y0,h,k,n)
%
% Produces an approximate solution to the initial value problem
%
%      y'(t) = f(t,y(t))     y(t0) = y0
%
% using a strategy that is based upon a k-th order
% Adams PC method. Stepsize is fixed.
%
% Pre:  fname = string that names the function f.
%       t0 = initial time.
%       y0 = initial condition vector.
%       h  = stepsize.
%       k  = order of method. (1<=k<=5).
%       n  = number of steps to be taken,
%
% Post: tvals(j) = t0 + (j-1)h, j=1:n+1
%       yvals(:j) = approximate solution at t = tvals(j), j=1:n+1
%

   [tvals,yvals,fvals] = StartAB(fname,t0,y0,h,k);
   tc = tvals(k);
   yc = yvals(:,k);
   fc = fvals(:,k);

   for j=k:n
      % Take a step and then update.
      [tc,yPred,fPred,yc,fc] = PCstep(fname,tc,yc,fvals,h,k);
```

```
      tvals = [tvals tc];
      yvals = [yvals yc];
          fvals = [fc fvals(:,1:k-1)];
   end
```

The starting values are obtained by the following M-file by means of a Runge–Kutta method.

```
  function [tvals,yvals,fvals] = StartAB(fname,t0,y0,h,k)
%
% Uses k-th order Runge-Kutta to generate approximate
% solutions to
%               y'(t) = f(t,y(t))   y(t0) = y0
%
% at t = t0, t0+h, ... , t0 + (k-1)h.
%
% Pre:
%       fname is a string that names the function f.
%       t0 is the initial time.
%       y0 is the initial value.
%       h is the step size.
%       k is the order of the RK method used.
%
% Post:
%       tvals = [ t0, t0+h, ... , t0 + (k-1)h].
%       For j =1:k, yvals(:,j) = y(tvals(j))  (approximately).
%       For j =1:k, fvals(:,j) = f(tvals(j),yvals(j))  .
%
  tc = t0;
  yc = y0;
  fc = feval(fname,tc,yc);
  tvals = tc;
  yvals = yc;
  fvals = fc;

  for j=1:k-1
          [tc,yc,fc] = RKstep(fname,tc,yc,fc,h,k);
      tvals = [tvals tc];
      yvals = [yvals yc];
      fvals = [fc fvals];
  end
```

The function M-file `Rkstep` is found in Subsection 10.6. The Adams-Bashforth predictor step is taken by the following M-file.

```
  function [tnew,ynew,fnew] = ABstep(fname,tc,yc,fvals,h,k)
%
% Pre:  fname is a string that names a function of the form f(t,y)
%       where t is a scalar and y is a column d-vector.
%
%       yc is an approximate solution to y'(t) = f(t,y(t)) at t=tc.
%
```

```
%       fvals is an d-by-k matrix where fvals(:,i) is an approximation
%       to f(t,y) at t = tc +(1-i)h, i=1:k
%
%       h is the time step.
%
%       k is the order of the AB method used, 1<=k<=5.
%
% Post: tnew=tc+h, ynew is an approximate solution at t=tnew, and
%       fnew = f(tnew,ynew).

   if k==1
       ynew = yc + h*fvals;
   elseif k==2
       ynew = yc + (h/2)*(fvals*[3;-1]);
   elseif k==3
       ynew = yc + (h/12)*(fvals*[23;-16;5]);
   elseif k==4
       ynew = yc + (h/24)*(fvals*[55;-59;37;-9]);
   elseif k==5
       ynew = yc + (h/720)*(fvals*[1901;-2774;2616;-1274;251]);
   end
   tnew = tc+h;
   fnew = feval(fname,tnew,ynew);
```

The Adams-Moulton corrector step is taken by the following M-file.

```
  function [tnew,ynew,fnew] = AMstep(fname,tc,yc,fvals,h,k)
%
% Pre:  fname is a string that names a function of the form f(t,y)
%       where t is a scalar and y is a column d-vector.
%
%       yc is an approximate solution to y'(t) = f(t,y(t)) at t=tc.
%
%       fvals is an d-by-k matrix where fvals(:,i) is an approximation
%       to f(t,y) at t = tc +(2-i)h, i=1:k
%
%       h is the time step.
%
%       k is the order of the AM method used, 1<=k<=5.
%
% Post: tnew=tc+h, ynew is an approximate solution at t=tnew, and
%       fnew = f(tnew,ynew).


   if k==1
       ynew = yc + h*fvals;
   elseif k==2
       ynew = yc + (h/2)*(fvals*[1;1]);
   elseif k==3
       ynew = yc + (h/12)*(fvals*[5;8;-1]);
```

```
   elseif k==4
       ynew = yc + (h/24)*(fvals*[9;19;-5;1]);
   elseif k==5
       ynew = yc + (h/720)*(fvals*[251;646;-264;106;-19]);
   end
   tnew = tc+h;
   fnew = feval(fname,tnew,ynew);
```

The predictor-corrector step is taken by the following M-file.

```
  function [tnew,yPred,fPred,yCorr,fCorr] = PCstep(fname,tc,yc,fvals,h,k)
%
% Pre:  fname is a string that names a function of the form f(t,y)
%       where t is a scalar and y is a column d-vector.
%
%       yc is an approximate solution to y'(t) = f(t,y(t)) at t=tc.
%
%       fvals is an d-by-k matrix where fvals(:,i) is an approximation
%       to f(t,y) at t = tc +(1-i)h, i=1:k
%
%       h is the time step.
%
%       k is the order of the Runge-Kutta method used, 1<=k<=5.
%
% Post: tnew=tc+h,
%       yPred is the predicted solution at t=tnew
%       fPred = f(tnew,yPred)
%       yCorr is the corrected solution at t=tnew
%       fCorr = f(tnew,yCorr).


   [tnew,yPred,fPred] = ABstep(fname,tc,yc,fvals,h,k);
   [tnew,yCorr,fCorr] = AMstep(fname,tc,yc,[fPred fvals(:,1:k-1)],h,k);
```

10.8.4. Specification of multistep methods. The left-hand side of Adams methods is of the form

$$y_{n+1} - y_n.$$

Adams–Bashforth methods are explicit and Adams–Moulton methods are implicit. In the following formulae, Adams methods are obtained by taking $a = 0$ and $b = 0$. The integer k is the number of steps of the method. The integer p is the order of the method and the constant C_{p+1} is the constant of the top-order error term.

Explicit Methods

$$\begin{aligned} k = 1: \quad & \\ & \alpha_1 = 1, \\ & \alpha_0 = -1, \quad \beta_0 = 1, \\ & p = 1; \qquad C_{p+1} = \tfrac{1}{2}. \end{aligned}$$

$k = 2$:

$$\begin{aligned} &\alpha_2 = 1, && \\ &\alpha_1 = -1 - a, && \beta_1 = \tfrac{1}{2}(3 - a), \\ &\alpha_0 = a, && \beta_0 = \tfrac{1}{2}(-1 + a), \\ &p = 2; && C_{p+1} = \tfrac{1}{12}(5 + a). \end{aligned}$$

Absolute stability limits the order to 2.

$k = 3$:

$$\begin{aligned} &\alpha_3 = 1, && \\ &\alpha_2 = -1 - a, && \beta_2 = \tfrac{1}{12}(23 - 5a - b), \\ &\alpha_1 = a + b, && \beta_1 = \tfrac{1}{3}(-4 - 2a + 2b), \\ &\alpha_0 = -b, && \beta_0 = \tfrac{1}{12}(5 + a + 5b), \\ &p = 3; && C_{p+1} = \tfrac{1}{24}(9 + a + b). \end{aligned}$$

Absolute stability limits the order to 3.

$k = 4$:

$$\begin{aligned} &\alpha_4 = 1, && \\ &\alpha_3 = -1 - a, && \beta_3 = \tfrac{1}{24}(55 - 9a - b - c), \\ &\alpha_2 = a + b, && \beta_2 = \tfrac{1}{24}(-59 - 19a + 13b - 19c), \\ &\alpha_1 = -b - c, && \beta_1 = \tfrac{1}{24}(37 + 5a + 13b - 19c), \\ &\alpha_0 = c, && \beta_0 = \tfrac{1}{24}(-9 - a - b - 9c), \\ &p = 4; && C_{p+1} = \tfrac{1}{720}(251 + 19a + 11b + 19c). \end{aligned}$$

Absolute stability limits the order to 4.

Implicit Methods

$k = 1$:

$$\begin{aligned} &\alpha_1 = 1, && \beta_1 = \tfrac{1}{2}, \\ &\alpha_0 = -1, && \beta_0 = \tfrac{1}{2}, \\ &p = 2; && C_{p+1} = -\tfrac{1}{12}. \end{aligned}$$

$k = 2$:

$$\begin{aligned} &\alpha_2 = 1, && \beta_2 = \tfrac{1}{12}(5 + a), \\ &\alpha_1 = -1 - a, && \beta_1 = \tfrac{2}{3}(1 - a), \\ &\alpha_0 = a, && \beta_0 = \tfrac{1}{12}(-1 - 5a), \\ \text{If } a \neq -1, \quad &p = 3; && C_{p+1} = -\tfrac{1}{24}(1 + a), \\ \text{If } a = -1, \quad &p = 4; && C_{p+1} = -\tfrac{1}{90}. \end{aligned}$$

$k = 3$:

$$\begin{aligned} &\alpha_3 = 1, && \beta_3 = \tfrac{1}{24}(9 + a + b),\\ &\alpha_2 = -1 - a, && \beta_2 = \tfrac{1}{24}(19 - 13a - 5b),\\ &\alpha_1 = a + b, && \beta_1 = \tfrac{1}{24}(-5 - 13a + 19b),\\ &\alpha_0 = -b, && \beta_0 = \tfrac{1}{24}(1 + a + 9b),\\ &p = 4; && C_{p+1} = -\tfrac{1}{720}(19 + 11a + 19b). \end{aligned}$$

Absolute stability limits the order to 4.

$k = 4$:

$$\begin{aligned} &\alpha_4 = 1, && \beta_4 = \tfrac{1}{720}(251 + 19a + 11b + 19c),\\ &\alpha_3 = -1 - a, && \beta_3 = \tfrac{1}{360}(323 - 173a - 37b - 53c),\\ &\alpha_2 = a + b, && \beta_2 = \tfrac{1}{30}(-11 - 19a + 19b + 11c),\\ &\alpha_1 = -b - c, && \beta_1 = \tfrac{1}{360}(53 + 37a + 173b - 323c),\\ &\alpha_0 = c, && \beta_0 = \tfrac{1}{720}(-19 - 11a - 19b - 251c). \end{aligned}$$

If $27 + 11a + 11b + 27c \neq 0$, then

$$p = 5; \qquad C_{p+1} = -\frac{1}{1440}(27 + 11a + 11b + 27c).$$

If $27 + 11a + 11b + 27c = 0$, then

$$p = 6; \qquad C_{p+1} = -\frac{1}{15\,120}(74 + 10a - 10b - 74c).$$

Absolute stability limits the order to 6.

The Matlab solver `ode113` is a fully variable step size, PECE implementation in terms of modified divided differences of the Adams–Bashforth–Moulton family of formulae of orders 1 to 12. The natural "free" interpolants are used. Local extrapolation is done. Details are to be found in *The MATLAB ODE Suite*, L. F. Shampine and M. W. Reichelt, SIAM Journal on Scientific Computing, **18**(1), 1997.

10.9. Stiff Systems of Differential Equations

In this section, we illustrate the concept of stiff systems of differential equations by means of an example and mention some numerical methods that can handle such systems.

10.9.1. The phenomenon of stiffness. While the intuitive meaning of stiff is clear to all specialists, much controversy is going on about its correct mathematical definition. The most pragmatic opinion is also historically the first one: stiff equations are equations where certain implicit methods, in particular backward differentiation methods, perform much better than explicit ones.

Consider a system of n differential equations,

$$\boldsymbol{y}' = \boldsymbol{f}(x, \boldsymbol{y}),$$

and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of the $n \times n$ Jacobian matrix

$$J = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}} = \left(\frac{\partial f_i}{\partial y_j}\right), \quad i \downarrow 1, \ldots, n, \quad j \rightarrow 1, \ldots, n, \tag{10.36}$$

where Nagumo's matrix index notation has been used. We assume that the n eigenvalues, $\lambda_1, \ldots, \lambda_n$, of the matrix J have negative real parts, $\Re\lambda_j < 0$, and are ordered as follows:

$$\Re\lambda_n \leq \cdots \leq \Re\lambda_2 \leq \Re\lambda_1 < 0. \tag{10.37}$$

The following definition occurs in discussing stiffness.

DEFINITION 10.6. The *stiffness ratio* of the system $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$ is the positive number

$$r = \frac{\Re\lambda_n}{\Re\lambda_1}, \tag{10.38}$$

where the eigenvalues of the Jacobian matrix (10.36) of the system satisfy the relations (10.37).

The phenomenon of stiffness appears under various aspects:

- A linear constant coefficient system is stiff if all of its eigenvalues have negative real parts and the stiffness ratio is large.
- Stiffness occurs when stability requirements, rather than those of accuracy, constrain the step length.
- Stiffness occurs when some components of the solution decay much more rapidly than others.
- A system is said to be stiff in a given interval I containing t if in I the neighboring solution curves approach the solution curve at a rate which is very large in comparison with the rate at which the solution varies in that interval.

A statement that we take as a definition of stiffness is one which merely relates what is observed happening in practice.

DEFINITION 10.7. If a numerical method with a region of absolute stability, applied to a system of differential equation with any initial conditions, is forced to use in a certain interval I of integration a step size which is *excessively small* in relation to the smoothness of the exact solution in I, then the system is said to be *stiff* in I.

Explicit Runge–Kutta methods and predictor-corrector methods, which, in fact, are explicit pairs, cannot handle stiff systems in an economical way, if they can handle them at all. Implicit methods require the solution of nonlinear equations which are almost always solved by some form of Newton's method. Two such implicit methods are in the following two sections.

10.9.2. Backward differentiation formulae. We define a k-step *backward differentiation formula* (BDF) in standard form by

$$\sum_{j=0}^{k} \alpha_j y_{n+j-k+1} = h\beta_k f_{n+1},$$

where $\alpha_k = 1$. BDF's are implicit methods. Table 10.7 lists the BDF's of stepnumber 1 to 6, respectively. In the table, k is the stepnumber, p is the order, C_{p+1} is the error constant, and α is half the angle subtended at the origin by the region of absolute stability R.

TABLE 10.7. Coefficients of the BDF methods.

k	α_6	α_5	α_4	α_3	α_2	α_1	α_0	β_k	p	C_{p+1}	α
1						1	-1	1	1	1	90°
2					1	$-\frac{4}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	2	$-\frac{2}{9}$	90°
3				1	$-\frac{18}{11}$	$\frac{9}{11}$	$=\frac{2}{11}$	$\frac{6}{11}$	3	$-\frac{3}{22}$	86°
4			1	$-\frac{48}{25}$	$\frac{36}{25}$	$-\frac{16}{25}$	$\frac{3}{25}$	$\frac{12}{25}$	4	$-\frac{12}{125}$	73°
5		1	$-\frac{300}{137}$	$\frac{300}{137}$	$-\frac{200}{137}$	$\frac{75}{137}$	$-\frac{12}{137}$	$\frac{60}{137}$	5	$-\frac{110}{137}$	51°
6	1	$-\frac{360}{147}$	$\frac{450}{147}$	$-\frac{400}{147}$	$\frac{225}{147}$	$-\frac{72}{147}$	$\frac{10}{147}$	$\frac{60}{147}$	6	$-\frac{20}{343}$	18°

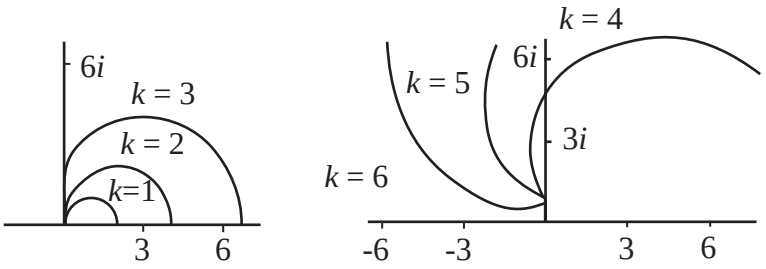

FIGURE 10.10. Left: Regions of absolute stability for k-step BDF for $k = 1, 2 \ldots, 6$. These regions include the negative real axis.

The left part of Fig. 10.10 shows the upper half of the region of absolute stability of the 1-step BDF, which is the exterior of the unit disk with center 1, and the regions of absolute stability of the 2- and 3-step BDF's which are the exterior of closed regions in the right-hand plane. The angle subtended at the origin is $\alpha = 90°$ in the first two cases and $\alpha = 88°$ in the third case. The right part of Fig. 10.10 shows the regions of absolute stability of the 4-, 5-, and 6-steps BDF's which include the negative real axis and make angles subtended at the origin of 73°, 51°, and 18°, respectively.

BDF methods are used to solve stiff systems.

10.9.3. Numerical differentiation formulae. Numerical differentiation formulae (NDF) are a modification of BDF's. Letting

$$\nabla y_n = y_n - y_{n-1}$$

denote the backward difference of y_n, we rewrite the k-step BDF of order $p = k$ in the form

$$\sum_{m=1}^{k} \frac{1}{m} \nabla^m y_{n+1} = h f_{n+1}.$$

The algebraic equation for y_{n+1} is solved with a simplified Newton (chord) iteration. The iteration is started with the predicted value

$$y_{n+1}^{[0]} = \sum_{m=0}^{k} \frac{1}{m} \nabla^m y_n.$$

Then the k-step NDF of order $p = k$ is

$$\sum_{m=1}^{k} \frac{1}{m} \nabla^m y_{n+1} = h f_{n+1} + \kappa \gamma_k \left(y_{n+1} - y_{n+1}^{[0]} \right),$$

where κ is a scalar parameter and $\gamma_k = \sum_{j=1}^{k} 1/j$. The NDF of order 1 to 5 are given in Table 10.8.

TABLE 10.8. Coefficients of the NDF methods.

k	κ	α_5	α_4	α_3	α_2	α_1	α_0	β_k	p	C_{p+1}	α
1	$-37/200$					1	-1	1	1	1	90°
2	$-1/9$				1	$-\frac{4}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	2	$-\frac{2}{9}$	90°
3	-0.0823			1	$-\frac{18}{11}$	$\frac{9}{11}$	$-\frac{2}{11}$	$\frac{6}{11}$	3	$-\frac{3}{22}$	80°
4	-0.0415		1	$-\frac{48}{25}$	$\frac{36}{25}$	$-\frac{16}{25}$	$\frac{3}{25}$	$\frac{12}{25}$	4	$-\frac{12}{125}$	66°
5	0	1	$-\frac{300}{137}$	$\frac{300}{137}$	$-\frac{200}{137}$	$\frac{75}{137}$	$-\frac{12}{137}$	$\frac{60}{137}$	5	$-\frac{110}{137}$	51°

The choice of the number κ is a compromise made in balancing efficiency in step size and stability angle. Compared with the BDF's, there is a step ratio gain of 26% in NDF's of order 1, 2, and 3, 12% in NDF of order 4, and no change in NDF of order 5. The percent change in the stability angle is 0%, 0%, -7%, -10%, and 0%, respectively. No NDF of order 6 is considered because, in this case, the angle α is too small.

10.9.4. The effect of a large stiffness ratio. In the following example, we analyze the effect of the large stiffness ratio of a simple decoupled system of two differential equations with constant coefficients on the step size of the five methods of the ODE Suite. Such problems are called *pseudo-stiff* since they are quite tractable by implicit methods.

Consider the initial value problem

$$\begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix}' = \begin{bmatrix} 1 & 0 \\ 0 & 10^q \end{bmatrix} \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix}, \qquad \begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{10.39}$$

or

$$\boldsymbol{y}' = A\boldsymbol{y}, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0.$$

Since the eigenvalues of A are

$$\lambda_1 = -1, \qquad \lambda_2 = -10^q,$$

the stiffness ratio (10.38) of the system is

$$r = 10^q.$$

The solution is

$$\begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} e^{-x} \\ e^{-10^q x} \end{bmatrix}.$$

Even though the second part of the solution containing the fast decaying factor $\exp(-10^q t)$ for large q numerically disappears quickly, the large stiffness ratio continues to restrict the step size of any explicit schemes, including predictor-corrector schemes.

EXAMPLE 10.13. Study the effect of the stiffness ratio on the number of steps used by the five MATLAB ode codes in solving problem (10.39) with $q = 1$ and $q = 5$.

SOLUTION. The function M-file `exp5_13.m` is

```
function uprime = exp5_13(x,u);  % Example 5.13
global q % global variable
A=[-1 0;0 -10^q]; % matrix A
uprime = A*u;
```

The following commands solve the non-stiff initial value problem with $q = 1$, and hence $r = e^{10}$, with relative and absolute tolerances equal to 10^{-12} and 10^{-14}, respectively. The option `stats on` requires that the code keeps track of the number of function evaluations.

```
clear;
global q; q=1;
tspan = [0 1]; y0 = [1 1]';
options = odeset('RelTol',1e-12,'AbsTol',1e-14,'Stats','on');
[x23,y23] = ode23('exp5_13',tspan,y0,options);
[x45,y45] = ode45('exp5_13',tspan,y0,options);
[x113,y113] = ode113('exp5_13',tspan,y0,options);
[x23s,y23s] = ode23s('exp5_13',tspan,y0,options);
[x15s,y15s] = ode15s('exp5_13',tspan,y0,options);
```

Similarly, when $q = 5$, and hence $r = \exp(10^5)$, the program solves a pseudo-stiff initial value problem (10.39). Table 10.9 lists the number of steps used with $q = 1$ and $q = 5$ by each of the five methods of the ODE suite.

TABLE 10.9. Number of steps used by each method with $q = 1$ and $q = 5$ with default relative and absolute tolerances $RT = 10^{-3}$ and $AT = 10^{-6}$ respectively, and tolerances 10^{-12} and 10^{-14}, respectively.

(RT, AT)	$(10^{-3}, 10^{-6})$		$(10^{-12}, 10^{-14})$	
q	1	5	1	5
ode23	29	39 823	24 450	65 944
ode45	13	30 143	601	30 856
ode113	28	62 371	132	64 317
ode23s	37	57	30 500	36 925
ode15s	43	89	773	1 128

It is seen from the table that nonstiff solvers are hopelessly slow and very expensive in solving pseudo-stiff equations. □

We consider another example of a second-order equation, with one real parameter q, which we first solve analytically. We shall obtain a coupled system in this case.

EXAMPLE 10.14. Solve the initial value problem

$$y'' + (10^q + 1)y' + 10^q y = 0 \qquad \text{on} \quad [0, 1],$$

with initial conditions

$$y(0) = 2, \qquad y'(0) = -10^q - 1,$$

and real parameter q.

SOLUTION. Substituting

$$y(x) = e^{\lambda x}$$

in the differential equation, we obtain the characteristic polynomial and eigenvalues:

$$\lambda^2 + (10^q + 1)\lambda + 10^q = (\lambda + 10^q)(\lambda + 1) = 0 \implies \lambda_1 = -10^q, \quad \lambda_2 = -1.$$

Two independent solutions are

$$y_1 = e^{-10^q x}, \qquad y_2(x) = e^{-x}.$$

The general solution is

$$y(x) = c_1\, e^{-10^q x} + c_2\, e^{-x}.$$

Using the initial conditions, one finds that $c_1 = 1$ and $c_2 = 1$. Thus the unique solution is

$$y(x) = e^{-10^q x} + e^{-x}. \quad \square$$

In view of solving the problem in Example 10.14 with numeric Matlab, we reformulate it into a system of two first-order equations.

EXAMPLE 10.15. Reformulate the initial value problem

$$y'' + (10^q + 1)y' + 10^q y = 0 \qquad \text{on} \quad [0, 1],$$

with initial conditions

$$y(0) = 2, \qquad y'(0) = -10^q - 1,$$

and real parameter q, into a system of two first-order equations and find its vector solution.

SOLUTION. Set

$$u_1 = y, \qquad u_2 = y'.$$

Hence,

$$u_2 = u_1', \qquad u_2' = y'' = -10^q u_1 - (10^q + 1)u_2.$$

Thus we have the system $\boldsymbol{u}' = A\boldsymbol{u}$,

$$\begin{bmatrix} u_1(x) \\ u_2(x) \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ -10^q & -(10^q + 1) \end{bmatrix} \begin{bmatrix} u_1(x) \\ u_2(x) \end{bmatrix}, \quad \text{with} \begin{bmatrix} u_1(0) \\ u_2(0) \end{bmatrix} = \begin{bmatrix} 2 \\ -10^q - 1 \end{bmatrix}.$$

Substituting the vector function

$$\boldsymbol{u}(x) = \boldsymbol{c}\, e^{\lambda x}$$

in the differential system, we obtain the matrix eigenvalue problem

$$(A - \lambda I)\boldsymbol{c} = \begin{bmatrix} -\lambda & 1 \\ -10^q & -(10^q + 1) - \lambda \end{bmatrix} \boldsymbol{c} = 0,$$

This problem has a nonzero solution $\boldsymbol{c}$ if and only if

$$\det(A - \lambda I) = \lambda^2 + (10^q + 1)\lambda + 10^q = (\lambda + 10^q)(\lambda + 1) = 0.$$

Hence the eigenvalues are

$$\lambda_1 = -10^q, \qquad \lambda_2 = -1.$$

The eigenvectors are found by solving the linear systems

$$(A - \lambda_i I)\boldsymbol{v}_i = 0.$$

Thus,

$$\begin{bmatrix} 10^q & 1 \\ -10^q & -1 \end{bmatrix} \boldsymbol{v}_1 = 0 \implies \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -10^q \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 1 \\ -10^q & -10^q \end{bmatrix} \boldsymbol{v}_2 = 0 \implies \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The general solution is

$$\boldsymbol{u}(x) = c_1\, e^{-10^q x}\boldsymbol{v}_1 + c_2\, e^{-x}\boldsymbol{v}_2.$$

The initial conditions implies that $c_1 = 1$ and $c_2 = 1$. Thus the unique solution is

$$\begin{bmatrix} u_1(x) \\ u_2(x) \end{bmatrix} = \begin{bmatrix} 1 \\ -10^q \end{bmatrix} e^{-10^q x} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} e^{-x}. \quad \square$$

We see that the stiffness ratio of the equation in Example 10.15 is

$$10^q.$$

EXAMPLE 10.16. Use the five Matlab ode solvers to solve the pseudo-stiff differential equation

$$y'' + (10^q + 1)y' + 10^q y = 0 \qquad \text{on} \quad [0, 1],$$

with initial conditions

$$y(0) = 2, \qquad y'(0) = -10^q - 1,$$

for $q = 1$ and compare the number of steps used by the solvers.

SOLUTION. The function M-file `exp5_16.m` is

```
function uprime = exp5_16(x,u)
global q
A=[0 1;-10^q -1-10^q];
uprime = A*u;
```

The following commands solve the initial value problem.

```
>> clear
>> global q; q = 1;
>> xspan = [0 1]; u0 = [2 -(10^q + 1)]';
>> [x23,u23] = ode23('exp5_16',xspan,u0);
>> [x45,u45] = ode45('exp5_16',xspan,u0);
>> [x113,u113] = ode113('exp5_16',xspan,u0);
>> [x23s,u23s] = ode23s('exp5_16',xspan,u0);
>> [x15s,u15s] = ode15s('exp5_16',xspan,u0);
>> whos
  Name        Size         Bytes  Class

  q           1x1              8  double array (global)
```

```
  u0          2x1             16  double array
  u113       26x2            416  double array
  u15s       32x2            512  double array
  u23        20x2            320  double array
  u23s       25x2            400  double array
  u45        49x2            784  double array
  x113       26x1            208  double array
  x15s       32x1            256  double array
  x23        20x1            160  double array
  x23s       25x1            200  double array
  x45        49x1            392  double array
  xspan       1x2             16  double array

Grand total is 461 elements using 3688 bytes
```

From the table produced by the command `whos` one sees that the nonstiff ode solvers `ode23`, `ode45`, `ode113`, and the stiff ode solvers `ode23s`, `ode15s`, use 20, 49, 26, and 25, 32 steps, respectively. □

EXAMPLE 10.17. Use the five Matlab ode solvers to solve the pseudo-stiff differential equation

$$y'' + (10^q + 1)y' + 10^q y = 0 \qquad \text{on} \quad [0, 1],$$

with initial conditions

$$y(0) = 2, \qquad y'(0) = -10^q - 1,$$

for $q = 5$ and compare the number of steps used by the solvers.

SOLUTION. Setting the value $q = 5$ in the program of Example 10.16, we obtain the following results for the `whos` command.

```
clear
global q; q = 5;
xspan = [0 1]; u0 = [2 -(10^q + 1)]';
[x23,u23] = ode23('exp5_16',xspan,u0);
[x45,u45] = ode45('exp5_16',xspan,u0);
[x113,u113] = ode113('exp5_16',xspan,u0);
[x23s,u23s] = ode23s('exp5_16',xspan,u0);
[x15s,u15s] = ode15s('exp5_16',xspan,u0);
whos
  Name        Size         Bytes  Class

  q           1x1              8  double array (global)
  u0          2x1             16  double array
  u113    62258x2         996128  double array
  u15s      107x2           1712  double array
  u23     39834x2         637344  double array
  u23s       75x2           1200  double array
  u45    120593x2        1929488  double array
  x113    62258x1         498064  double array
  x15s      107x1            856  double array
```

```
  x23     39834x1          318672  double array
  x23s       75x1             600  double array
  x45    120593x1          964744  double array
  xspan       1x2              16  double array

Grand total is 668606 elements using 5348848 bytes
```

From the table produced by the command `whos`, one sees that the nonstiff ode solvers `ode23`, `ode45`, `ode113`, and the stiff ode solvers `ode23s`, `ode15s`, use 39 834, 120 593, 62 258, and 75, 107 steps, respectively. It follows that nonstiff solvers are hopelessly slow and expensive to solve stiff equations. □

Numeric MATLAB has four solvers with "free" interpolants for stiff systems. The first three are low order solvers.

- The code `ode23s` is an implementation of a new modified Rosenbrock (2,3) pair. Local extrapolation is not done. By default, Jacobians are generated numerically.
- The code `ode23t` is an implementation of the trapezoidal rule.
- The code `ode23tb` is an in an implicit two-stage Runge–Kutta formula.
- The variable-step variable-order Matlab solver `ode15s` is a quasi-constant step size implementation in terms of backward differences of the Klopfenstein–Shampine family of Numerical Differentiation Formulae of orders 1 to 5. Local extrapolation is not done. By default, Jacobians are generated numerically.

Details on these methods are to be found in *The MATLAB ODE Suite*, L. F. Shampine and M. W. Reichelt, SIAM Journal on Scientific Computing, **18**(1), 1997.

Part 3

Exercises and Solutions

Starred exercises have solutions in Chapter 12.

CHAPTER 11

Exercises for Differential Equations and Laplace Transforms

Exercises for Chapter 1

Solve the following separable differential equations.

1.1. $y' = 2xy^2$.

1.2. $y' = \dfrac{xy}{x^2 - 1}$.

***1.3.** $(1 + x^2)y' = \cos^2 y$.

1.4. $(1 + e^x)yy' = e^x$.

1.5. $y' \sin x = y \ln y$.

1.6. $(1 + y^2)\,dx + (1 + x^2)\,dy = 0$.

Solve the following initial-value problems and plot the solutions.

1.7. $y' \sin x - y \cos x = 0, \quad y(\pi/2) = 1$.

1.8. $x \sin y\,dx + (x^2 + 1)\cos y\,dy = 0, \quad y(1) = \pi/2$.

Solve the following differential equations.

1.9. $(x^2 - 3y^2)\,dx + 2xy\,dy = 0$.

1.10. $(x + y)\,dx - x\,dy = 0$.

***1.11.** $xy' = y + \sqrt{y^2 - x^2}$.

1.12. $xy' = y + x\cos^2(y/x)$.

Solve the following initial-value problems.

1.13. $(2x - 5y)\,dx + (4x - y)\,dy = 0, \quad y(1) = 4$.

1.14. $(3x^2 + 9xy + 5y^2)\,dx - (6x^2 + 4xy)\,dy = 0, \quad y(2) = -6$.

1.15. $yy' = -(x + 2y), \quad y(1) = 1$.

1.16. $(x^2 + y^2)\,dx - 2xy\,dy = 0, \quad y(1) = 2$.

Solve the following differential equations.

1.17. $x(2x^2 + y^2) + y(x^2 + 2y^2)y' = 0$.

1.18. $(3x^2y^2 - 4xy)y' + 2xy^3 - 2y^2 = 0$.

1.19. $(\sin xy + xy\cos xy)\,dx + x^2\cos xy\,dy = 0.$

1.20. $\left(\dfrac{\sin 2x}{y} + x\right)\,dx + \left(y - \dfrac{\sin^2 x}{y^2}\right)\,dy = 0.$

Solve the following initial-value problems.

***1.21.** $(2xy - 3)\,dx + (x^2 + 4y)\,dy = 0, \quad y(1) = 2.$

1.22. $\dfrac{2x}{y^3}\,dx + \dfrac{y^2 - 3x^2}{y^4}\,dy = 0, \quad y(1) = 1.$

1.23. $(y\,e^x + 2\,e^x + y^2)\,dx + (e^x + 2xy)\,dy = 0, \quad y(0) = 6.$

1.24. $(2x\cos y + 3x^2y)\,dx + (x^3 - x^2\sin y - y)\,dy = 0, \quad y(0) = 2.$

Solve the following differential equations.

***1.25.** $(x + y^2)\,dx - 2xy\,dy = 0.$

1.26. $(x^2 - 2y)\,dx + x\,dy = 0.$

1.27. $(x^2 - y^2 + x)\,dx + 2xy\,dy = 0.$

1.28. $(1 - x^2y)\,dx + x^2(y - x)\,dy = 0.$

1.29. $(1 - xy)y' + y^2 + 3xy^3 = 0.$

1.30. $(2xy^2 - 3y^3)\,dx + (7 - 3xy^2)\,dy = 0.$

1.31. $(2x^2y - 2y + 5)\,dx + (2x^3 + 2x)\,dy = 0.$

1.32. $(x + \sin x + \sin y)\,dx + \cos y\,dy = 0.$

1.33. $y' + \dfrac{2}{x}y = 12.$

1.34. $y' + \dfrac{2x}{x^2 + 1}y = x.$

1.35. $x(\ln x)y' + y = 2\ln x.$

1.36. $xy' + 6y = 3x + 1.$

Solve the following initial-value problems.

1.37. $y' + 3x^2y = x^2, \quad y(0) = 2.$

1.38. $xy' - 2y = 2x^4, \quad y(2) = 8.$

***1.39.** $y' + y\cos x = \cos x, \quad y(0) = 1.$

1.40. $y' - y\tan x = \dfrac{1}{\cos^3 x}, \quad y(0) = 0.$

Find the orthogonal trajectories of each given family of curves. In each case sketch several members of the family and several of the orthogonal trajectories on the same set of axes.

1.41. $x^2 + y^2/4 = c.$

1.42. $y = e^x + c.$

1.43. $y^2 + 2x = c.$

1.44. $y = \arctan x + c$.

1.45. $x^2 - y^2 = c^2$.

1.46. $y^2 = cx^3$.

1.47. $e^x \cos y = c$.

1.48. $y = \ln x + c$.

In each case draw direction fields and sketch several approximate solution curves.

1.49. $y' = 2y/x$.

1.50. $y' = -x/y$.

1.50. $y' = -xy$.

1.51. $9yy' + x = 0$.

Exercises for Chapter 2

Solve the following differential equations.

2.1. $y'' - 3y' + 2y = 0$.

2.2. $y'' + 2y' + y = 0$.

***2.3.** $y'' - 9y' + 20y = 0$.

Solve the following initial-value problems, with initial conditions $y(x_0) = y_0$, and plot the solutions $y(x)$ for $x \geq x_0$.

2.4. $y'' + y' + \dfrac{1}{4}\, y = 0, \quad y(2) = 1, \quad y'(2) = 1.$

2.5. $y'' + 9y = 0, \quad y(0) = 0, \quad y'(0) = 1.$

2.6. $y'' - 4y' + 3y = 0, \quad y(0) = 6, \quad y'(0) = 0.$

2.7. $y'' - 2y' + 3y = 0, \quad y(0) = 1, \quad y'(0) = 3.$

2.8. $y'' + 2y' + 2y = 0, \quad y(0) = 2, \quad y'(0) = -3.$

For the undamped oscillator equations below, find the amplitude and period of the motion.

2.9. $y'' + 4y = 0, \quad y(0) = 1, \quad y'(0) = 2.$

2.10. $y'' + 16y = 0, \quad y(0) = 0, \quad y'(0) = 1.$

For the critically damped oscillator equations, find a value $T \geq 0$ for which $|y(T)|$ is a maximum, find that maximum, and plot the solutions $y(x)$ for $x \geq 0$.

2.11. $y'' + 2y' + y = 0, \quad y(0) = 1, \quad y'(0) = 1.$

2.12. $y'' + 6y' + 9y = 0, \quad y(0) = 0, \quad y'(0) = 2.$

Solve the following Euler–Cauchy differential equations.

***2.13.** $x^2y'' + 3xy' - 3y = 0$.

2.14. $x^2y'' - xy' + y = 0.$

2.15. $4x^2y'' + y = 0.$

2.16. $x^2y'' + xy' + 4y = 0.$

Solve the following initial-value problems, with initial conditions $y(x_0) = y_0$, and plot the solutions $y(x)$ for $x \geq x_0$.

2.17. $x^2y'' + 4xy' + 2y = 0, \quad y(1) = 1, \quad y'(1) = 2.$

2.18. $x^2y'' + 5xy' + 3y = 0, \quad y(1) = 1, \quad y'(1) = -5.$

2.19. $x^2y'' - xy' + y = 0, \quad y(1) = 1, \quad y'(1) = 0.$

2.20. $x^2y'' + \frac{7}{2}xy' - \frac{3}{2}y = 0, \quad y(4) = 1, \quad y'(4) = 0.$

Exercises for Chapter 3

Solve the following constant coefficient differential equations.

***3.1.** $y''' + 6y'' = 0.$

3.2. $y''' + 3y'' - 4y' - 12y = 0.$

3.3. $y''' - y = 0.$

3.4. $y^{(4)} + y''' - 3y'' - y' + 2y = 0.$

Solve the following initial-value problems and plot the solutions $y(x)$ for $x \geq 0$.

3.5. $y''' + 12y'' + 36y' = 0, \quad y(0) = 0, \quad y'(0) = 1, \quad y''(0) = -7.$

3.6. $y^{(4)} - y = 0, \quad y(0) = 0, \quad y'(0) = 0, \quad y''(0) = 0, \quad y'''(0) = 1.$

3.7. $y''' - y'' - y' + y = 0, \quad y(0) = 0, \quad y'(0) = 5, \quad y''(0) = 2.$

3.8. $y''' - 2y'' + 4y' - 8y = 0, \quad y(0) = 2, \quad y'(0) = 0, \quad y''(0) = 0.$

Determine whether the given functions are linearly dependent or independent on $-\infty < x < +\infty$.

***3.9.** $y_1(x) = x, \quad y_2(x) = x^2, \quad y_3(x) = 2x - 5x^2.$

3.10. $y_1(x) = 1 + x, \quad y_2(x) = x, \quad y_3(x) = x^2.$

3.11. $y_1(x) = 2, \quad y_2(x) = \sin^2 x, \quad y_3(x) = \cos^2 x.$

3.12. $y_1(x) = e^x, \quad y_2(x) = e^{-x}, \quad y_3(x) = \cosh x.$

Show by computing the Wronskian that the given functions are linearly independent on the indicated interval.

***3.13.** $e^x, \quad e^{2x}, \quad e^{-x}, \qquad -\infty < x < +\infty.$

3.14. $x + 2, \quad x^2, \qquad -\infty < x < +\infty.$

3.15. $x^{1/3}, \quad x^{1/4}, \qquad 0 < x < +\infty.$

3.16. $x, \quad x\ln x, \quad x^2\ln x, \qquad 0 < x < +\infty.$

3.17 Show that the functions

$$f_1(x) = x^2, \qquad f_2(x) = x|x| = \begin{cases} x^2, & x \geq 0, \\ -x^2, & x < 0 \end{cases}$$

are linearly independent on $[-1, 1]$ and compute their Wronskian. Explain your result.

Find a second solution of each differential equation if $y_1(x)$ is a solution.

3.18. $xy'' + y' = 0, \quad y_1(x) = \ln x.$

3.19. $x(x-2)y'' - (x^2-2)y' + 2(x-1)y = 0, \quad y_1(x) = e^x.$

3.20. $(1-x^2)y'' - 2xy' = 0, \quad y_1(x) = 1.$

3.21. $(1+2x)y'' + 4xy' - 4y = 0, \quad y_1(x) = e^{-2x}.$

Solve the following differential equations.

3.22. $y'' + 3y' + 2y = 5\,e^{-2x}.$

3.23. $y'' + y' = 3x^2.$

3.24. $y'' - y' - 2y = 2x\,e^{-x} + x^2.$

***3.25.** $y'' - y' = e^x \sin x.$

Solve the following initial-value problems and plot the solutions $y(x)$ for $x \geq 0$.

3.26. $y'' + y = 2\cos x, \quad y(0) = 1, \quad y'(0) = 0.$

3.27. $y^{(4)} - y = 8\,e^x, \quad y(0) = 0, \quad y'(0) = 2, \quad y''(0) = 4, \quad y'''(0) = 6.$

3.28. $y''' + y' = x, \quad y(0) = 0, \quad y'(0) = 1, \quad y''(0) = 0.$

3.29. $y'' + y = 3x^2 - 4\sin x, \quad y(0) = 0, \quad y'(0) = 1.$

Solve the following differential equations.

3.30. $y'' + y = \dfrac{1}{\sin x}.$

3.31. $y'' + y = \dfrac{1}{\cos x}.$

3.32. $y'' + 6y' + 9y = \dfrac{e^{-3x}}{x^3}.$

3.33. $y'' - 2y'\tan x = 1.$

3.34. $y'' - 2y' + y = \dfrac{e^x}{x}.$

***3.35.** $y'' + 3y' + 2y = \dfrac{1}{1+e^x}.$

Solve the following initial-value problems, with initial conditions $y(x_0) = y_0$, and plot the solutions $y(x)$ for $x \geq x_0$.

3.36. $y'' + y = \tan x, \quad y(0) = 1, \quad y'(0) = 0.$

3.37. $y'' - 2y' + y = \dfrac{e^x}{x}, \quad y(1) = e, \quad y'(1) = 0.$

3.38. $2x^2y'' + xy' - 3y = x^{-2}, \quad y(1) = 0, \quad y'(1) = 2.$

3.39. $2x^2y'' + xy' - 3y = 2x^{-3}, \quad y(1) = 0, \quad y'(1) = 3.$

Exercises for Chapter 4

Solve the following systems of differential equations $\boldsymbol{y}' = A\boldsymbol{y}$ for given matrices A.

4.1. $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}.$

4.2. $A = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix}.$

***4.3.** $A = \begin{bmatrix} -1 & 1 \\ 4 & -1 \end{bmatrix}.$

4.4. $A = \begin{bmatrix} -1 & 1 & 4 \\ -2 & 2 & 4 \\ -1 & 0 & 4 \end{bmatrix}.$

4.5. $A = \begin{bmatrix} 1 & 1 \\ -4 & 1 \end{bmatrix}.$

Solve the following systems of differential equations $\boldsymbol{y}' = A\boldsymbol{y} + \boldsymbol{f}(x)$ for given matrices A and vectors $\boldsymbol{f}$.

4.6. $A = \begin{bmatrix} -3 & -2 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{f}(x) = \begin{bmatrix} 2\,e^{-x} \\ -e^{-x} \end{bmatrix}.$

4.7. $A = \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix}, \quad \boldsymbol{f}(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$

4.8. $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}, \quad \boldsymbol{f}(x) = \begin{bmatrix} 2\,e^{-x} \\ 3x \end{bmatrix}.$

4.9. $A = \begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix}, \quad \boldsymbol{f}(x) = \begin{bmatrix} e^x \\ -e^x \end{bmatrix}.$

4.10. $A = \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}, \quad \boldsymbol{f}(x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$

Solve the initial value problem $\boldsymbol{y}' = A\boldsymbol{y}$ with $\boldsymbol{y}(0) = \boldsymbol{y}_0$, for given matrices A and vectors $\boldsymbol{y}_0$.

4.11. $A = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix}, \quad \boldsymbol{y}_0 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$

4.12. $A = \begin{bmatrix} -3 & 2 \\ -1 & -1 \end{bmatrix}, \quad \boldsymbol{y}_0 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$

4.13. $A = \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}, \quad \boldsymbol{y}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$

Exercises for Chapter 5

Find the Laplace transform of the given functions.

5.1. $f(t) = -3t + 2.$

5.2. $f(t) = t^2 + at + b.$

5.3. $f(t) = \cos(\omega t + \theta).$

5.4. $f(t) = \sin(\omega t + \theta).$

***5.5.** $f(t) = \cos^2 t.$

5.6. $f(t) = \sin^2 t.$

5.7. $f(t) = 3\cosh 2t + 4\sinh 5t.$

5.8. $f(t) = 2\,e^{-2t}\sin t.$

5.9. $f(t) = e^{-2t}\,\cosh t.$

5.10. $f(t) = \left(1 + 2e^{-t}\right)^2.$

5.11. $f(t) = u(t-1)(t-1).$

***5.12.** $f(t) = u(t-1)t^2.$

5.13. $f(t) = u(t-1)\cosh t.$

5.14. $f(t) = u(t-\pi/2)\sin t.$

Find the inverse Laplace transform of the given functions.

5.15. $F(s) = \dfrac{4(s+1)}{s^2 - 16}.$

5.16. $F(s) = \dfrac{2s}{s^2 + 3}.$

5.17. $F(s) = \dfrac{2}{s^2 + 3}.$

5.18. $F(s) = \dfrac{4}{s^2 - 9}.$

5.19. $F(s) = \dfrac{4s}{s^2 - 9}.$

5.20. $F(s) = \dfrac{3s - 5}{s^2 + 4}.$

5.21. $F(s) = \dfrac{1}{s^2 + s - 20}.$

5.22. $F(s) = \dfrac{1}{(s-2)(s^2 + 4s + 3)}.$

***5.23.** $F(s) = \dfrac{2s + 1}{s^2 + 5s + 6}.$

5.24. $F(s) = \dfrac{s^2 - 5}{s^3 + s^2 + 9s + 9}$.

5.25. $F(s) = \dfrac{3s^2 + 8s + 3}{(s^2 + 1)(s^2 + 9)}$.

5.26. $F(s) = \dfrac{s - 1}{s^2(s^2 + 1)}$.

5.27. $F(s) = \dfrac{1}{s^4 - 9}$.

5.28. $F(s) = \dfrac{(1 + e^{-2s})^2}{s + 2}$.

***5.29.** $F(s) = \dfrac{e^{-3s}}{s^2(s - 1)}$.

5.30. $F(s) = \dfrac{\pi}{2} - \arctan \dfrac{s}{2}$.

5.31. $F(s) = \ln \dfrac{s^2 + 1}{s^2 + 4}$.

Find the Laplace transform of the given functions.

5.32. $f(t) = \begin{cases} t, & 0 \le t < 1, \\ 1, & t \ge 1. \end{cases}$

5.33. $f(t) = \begin{cases} 2t + 3, & 0 \le t < 2, \\ 0, & t \ge 2. \end{cases}$

5.34. $f(t) = t \sin 3t$.

5.35. $f(t) = t \cos 4t$.

***5.36.** $f(t) = e^{-t}\, t \cos t$.

5.37. $f(t) = \displaystyle\int_0^t \tau e^{t-\tau}\, d\tau$.

5.38. $f(t) = 1 * e^{-2t}$.

5.39. $f(t) = e^{-t} * e^t \cos t$.

5.40. $f(t) = \dfrac{e^t - e^{-t}}{t}$.

Use Laplace transforms to solve the given initial value problems and plot the solution.

5.41. $y'' - 6y' + 13y = 0, \quad y(0) = 0, \quad y'(0) = -3$.

5.42. $y'' + y = \sin 3t, \quad y(0) = 0, \quad y'(0) = 0$.

5.43. $y'' + y = \sin t, \quad y(0) = 0, \quad y'(0) = 0$.

5.44. $y'' + y = t, \quad y(0) = 0, \quad y'(0) = 0$.

5.45. $y'' + 5y' + 6y = 3\, e^{-2t}, \quad y(0) = 0, \quad y'(0) = 1$.

5.46. $y'' + 2y' + 5y = 4t, \quad y(0) = 0, \quad y'(0) = 0$.

5.47. $y'' - 4y' + 4y = t^3\, e^{2t}, \quad y(0) = 0, \quad y'(0) = 0$.

5.48. $y'' + 4y = \begin{cases} 1, & 0 \le t < 1 \\ 0, & t \ge 1 \end{cases}, \quad y(0) = 0, \quad y'(0) = -1.$

5.49. $y'' - 5y' + 6y = \begin{cases} t, & 0 \le t < 1 \\ 0, & t \ge 1 \end{cases}, \quad y(0) = 0, \quad y'(0) = 1.$

5.50. $y'' + 4y' + 3y = \begin{cases} 4\,e^{1-t}, & 0 \le t < 1 \\ 4, & t \ge 1 \end{cases}, \quad y(0) = 0, \quad y'(0) = 0.$

***5.51.** $y'' + 4y' = u(t-1), \quad y(0) = 0, \quad y'(0) = 0.$

5.52. $y'' + 3y' + 2y = 1 - u(t-1), \quad y(0) = 0, \quad y'(0) = 1.$

5.53. $y'' - y = \sin t + \delta(t - \pi/2), \qquad y(0) = 3.5, \quad y'(0) = -3.5.$

5.54. $y'' + 5y' + 6y = u(t-1) + \delta(t-2), \qquad y(0) = 0, \quad y'(0) = 1.$

Using Laplace transforms solve the given integral equations and plot the solutions.

5.55. $y(t) = 1 + \displaystyle\int_0^t y(\tau)\, d\tau.$

5.56. $y(t) = \sin t + \displaystyle\int_0^t y(\tau) \sin(t-\tau)\, d\tau.$

5.57. $y(t) = \cos 3t + 2\displaystyle\int_0^t y(\tau) \cos 3(t-\tau)\, d\tau.$

5.58. $y(t) = t + e^t + \displaystyle\int_0^t y(\tau) \cosh(t-\tau)\, d\tau.$

5.59. $y(t) = t\,e^t + 2\,e^t \displaystyle\int_0^t e^{-\tau} y(\tau)\, d\tau.$

Sketch the following 2π-periodic functions over three periods and find their Laplace transforms.

5.60. $f(t) = \pi - t, \qquad 0 < t < 2\pi.$

5.61. $f(t) = 4\pi^2 - t^2, \qquad 0 < t < 2\pi.$

5.62. $f(t) = e^{-t}, \qquad 0 < t < 2\pi.$

5.63. $f(t) = \begin{cases} t, & \text{if } 0 < t < \pi, \\ \pi - t, & \text{if } \pi < t < 2\pi. \end{cases}$

5.64. $f(t) = \begin{cases} 0, & \text{if } 0 < t < \pi, \\ t - \pi, & \text{if } \pi < t < 2\pi. \end{cases}$

Exercises for Chapter 6

Find the interval of convergence of the given series and of the term by term first derivative of the series.

6.1. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{2n+1}\, x^n.$

6.2. $\displaystyle\sum_{n=1}^{\infty} \frac{2^n}{n3^{n+3}}\, x^n.$

6.3. $\displaystyle\sum_{n=2}^{\infty} \frac{\ln n}{n}\, x^n.$

6.4. $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2+1}\, (x+1)^n.$

6.5. $\displaystyle\sum_{n=3}^{\infty} \frac{n(n-1)(n-2)}{4^n}\, x^n.$

6.6. $\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n}{k^n}\, x^{2n}.$

6.7. $\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n}{k^n}\, x^{3n}.$

6.8. $\displaystyle\sum_{n=1}^{\infty} \frac{(4n)!}{(n!)^4}\, x^n.$

Find the power series solutions of the following ordinary differential equations.

6.9. $y'' - 3y' + 2y = 0.$

6.10. $(1-x^2)y'' - 2xy' + 2y = 0.$

6.11. $y'' + x^2y' + xy = 0.$

***6.12.** $y'' - xy' - y = 0.$

6.13. $(x^2-1)y'' + 4xy' + 2y = 0.$

6.14. $(1-x)y'' - y' + xy = 0.$

6.15. $y'' - 4xy' + (4x^2-2)y = 0.$

6.16. $y'' - 2(x-1)y' + 2y = 0.$

6.17. Show that the equation

$$\sin\theta \frac{d^2y}{d\theta^2} + \cos\theta \frac{dy}{d\theta} + n(n+1)(\sin\theta)y = 0$$

can be transformed into Legendre's equation by means of the substitution $x = \cos\theta$.

6.18. Derive Rodrigues' formula (6.10).

6.19. Derive the generating function (6.11).

6.20. Let A_1 and A_2 be two points in space (see Fig. 11.1). By means of (6.9) derive the formula

$$\frac{1}{r} = \frac{1}{\sqrt{r_1^2 + r_2^2 - 2r_1r_2\cos\theta}} = \frac{1}{r_2} \sum_{m=0}^{\infty} P_m(\cos\theta) \left(\frac{r_1}{r_2}\right)^m,$$

which is important in potential theory.

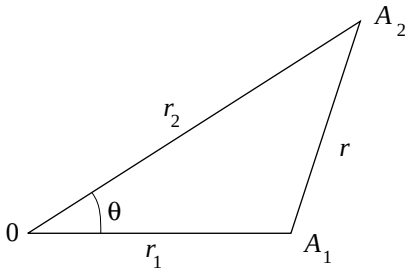

FIGURE 11.1. Distance r from point A_1 to point A_2.

6.21. Derive Bonnet recurrence formula,

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x), \quad n = 1, 2, \ldots \tag{11.1}$$

(Hint. Differentiate the generating function (6.11) with respect to t, substitute (6.11) in the differentiated formula, and compare the coefficients of t^n.)

6.22. Compare the value of $P_4(0.7)$ obtained by means of the three-point recurrence formula (11.1) of the previous exercise with the value obtained by evaluating $P_4(x)$ directly at $x = 0.7$.

6.23. For nonnegative integers m and n, with $m \le n$, let

$$p_n^m(x) = \frac{d^m}{dx^n} P_n(x).$$

Show that the function $p_n^m(x)$ is a solution of the differential equation

$$(1-x^2)y'' + 2(m+1)xy' + (n-m)(n+m+1)y = 0.$$

Express the following polynomials in terms of Legendre polynomials

$$P_0(x), \quad P_1(x), \quad \ldots$$

6.24. $p(x) = 5x^3 + 4x^2 + 3x + 2, \qquad -1 \le x \le 1.$

6.25. $p(x) = 10x^3 + 4x^2 + 6x + 1, \qquad -1 \le x \le 1.$

6.26. $p(x) = x^3 - 2x^2 + 4x + 1, \qquad -1 \le x \le 2.$

Find the first three coefficients of the Fourier–Legendre expansion of the following functions and plot $f(x)$ and its Fourier–Legendre approximation on the same graph.

6.27. $f(x) = e^x, \qquad -1 < x < 1.$

6.28. $f(x) = e^{2x}, \qquad -1 < x < 1.$

6.29. $f(x) = \begin{cases} 0 & -1 < x < 0, \\ 1 & 0 < x < 1. \end{cases}$

6.30. Integrate numerically

$$I = \int_{-1}^{1} (5x^5 + 4x^4 + 3x^3 + 2x^2 + x + 1)\, dx,$$

by means of the three-point Gaussian Quadrature formula. Moreover, find the exact value of I and compute the error in the numerical value.

***6.31.** Evaluate

$$I = \int_{0.2}^{1.5} e^{-x^2}\, dx,$$

by the three-point Gaussian Quadrature formula.

6.32. Evaluate

$$I = \int_{0.3}^{1.7} e^{-x^2}\, dx,$$

by the three-point Gaussian Quadrature formula.

6.33. Derive the four-point Gaussian Quadrature formula.

6.34. Obtain $P_4(x)$ by means of Bonnet's formula of Exercise 6.21 or otherwise.

6.35. Find the zeros of $P_4(x)$ in radical form.

Hint: Put $t = x^2$ in the even quartic polynomial $P_4(x)$ and solve the quadratic equation.

6.36. Obtain $P_5(x)$ by means of Bonnet's formula of Exercise 6.21 or otherwise.

5.37. Find the zeros of $P_5(x)$ in radical form.

Hint: Write $P_5(x) = xQ_4(x)$. Then put $t = x^2$ in the even quartic polynomial $Q_4(x)$ and solve the quadratic equation.

CHAPTER 12

Exercises for Numerical Methods

Angles are always in radian measure.

Exercises for Chapter 7

7.1. Use the bisection method to find x_3 for $f(x) = \sqrt{x} - \cos x$ on $[0, 1]$. Angles in radian measure.

7.2. Use the bisection method to find x_3 for

$$f(x) = 3(x + 1)(x - \frac{1}{2})(x - 1)$$

on the following intervals:

$$[-2, 1.5], \qquad [-1.25, 2.5].$$

7.3. Use the bisection method to find a solution accurate to 10^{-3} for $f(x) = x - \tan x$ on $[4, 4.5]$. Angles in radian measure.

7.4. Use the bisection method to find an approximation to $\sqrt{3}$ correct to within 10^{-4}. [*Hint*: Consider $f(x) = x^2 - 3$.]

7.5. Show that the fixed point iteration

$$x_{n+1} = \sqrt{2x_n + 3}$$

for the solving the equation $f(x) = x^2 - 2x - 3 = 0$ converges in the interval $[2, 4]$.

7.6. Use a fixed point iteration method, other than Newton's method, to determine a solution accurate to 10^{-2} for $f(x) = x^3 - x - 1 = 0$ on $[1, 2]$. Use $x_0 = 1$.

7.7. Use Newton's method to approximate $\sqrt{3}$ to 10^{-4}. Start with en $x_0 = 1$. Compare your result and the number of iterations required with the answer obtained in Exercise 8.4.

7.8. Do five iterations of the fixed point method $g(x) = \cos(x - 1)$. Take $x_0 = 2$. Use at least 6 decimals. Find the order of convergence of the method. Angles in radian measure.

7.9. Do five iterations of the fixed point method $g(x) = 1 + \sin^2 x$. Take $x_0 = 1$. Use at least 6 decimals. Find the order of convergence of the method. Angles in radian measure.

7.10. Sketch the function $f(x) = 2x - \tan x$ and compute a root of the equation $f(x) = 0$ to six decimals by means of Newton's method with $x_0 = 1$. Find the order of convergence of the method.

***7.11.** Sketch the function $f(x) = e^{-x} - \tan x$ and compute a root of the equation $f(x) = 0$ to six decimals by means of Newton's method with $x_0 = 1$. Find the order of convergence of the method.

7.12 Compute a root of the equation $f(x) = 2x - \tan x$ given in Exercise 8.10 with the secant method with starting values $x_0 = 1$ and $x_1 = 0.5$. Find the order of convergence to the root.

7.13. Repeat Exercise 8.12 with the method of false position. Find the order of convergence of the method.

7.14. Repeat Exercise 8.11 with the secant method with starting values $x_0 = 1$ and $x_1 = 0.5$. Find the order of convergence of the method.

7.15. Repeat Exercise 8.14 with the method of false position. Find the order of convergence of the method.

7.16. Consider the fixed point method of Exercise 8.5:

$$x_{n+1} = \sqrt{2x_n + 3}.$$

Complete the table:

n	x_n	Δx_n	$\Delta^2 x_n$
1	$x_1 = 4.000$		
2	$x_2 =$		
3	$x_3 =$		

Accelerate convergence by Aitken.

$$a_1 = x_1 - \frac{(\Delta x_1)^2}{\Delta^2 x_1} =$$

7.17. Apply Steffensen's method to the result of Exercise 8.9. Find the order of convergence of the method.

7.18. Use Müller's method to find the three zeros of

$$f(x) = x^3 + 3x^2 - 1.$$

7.19. Use Müller's method to find the four zeros of

$$f(x) = x^4 + 2x^2 - x - 3.$$

7.20. Sketch the function $f(x) = x - \tan x$. Find the multiplicity of the zero $x = 0$. Compute the root $x = 0$ of the equation $f(x) = 0$ to six decimals by means of a modified Newton method wich takes the multiplicity of the rood into account. Start at $x_0 = 1$. Find the order of convergence of the modified Newton method that was used.

***7.21.** Sketch the function $f(x) = x - \tan x$. Find the multiplicity of the zero $x = 0$. Compute the root $x = 0$ of the equation $f(x) = 0$ to six decimals by means of the secant method. Start at $x_0 = 1$ and $x_1 = 0.5$. Find the order of convergence of the method.

Exercises for Chapter 8

8.1. Given the function $f(x) = \ln(x+1)$ and the points $x_0 = 0$, $x_1 = 0.6$ and $x_2 = 0.9$. Construct the Lagrange interpolating polynomials of degrees exactly one and two to approximate $f(0.45)$ and find the actual errors.

8.2. Consider the data

$$f(8.1) = 16.94410, \quad f(8.3) = 17.56492, \quad f(8.6) = 18.50515, \quad f(8.7) = 18.82091.$$

Interpolate $f(8.4)$ by Lagrange interpolating polynomials of degree one, two and three.

8.3. Construct the Lagrange interpolating polynomial of degree 2 for the function $f(x) = e^{2x}\cos 3x$, using the values of f at the points $x_0 = 0$, $x_1 = 0.3$ and $x_2 = 0.6$.

***8.4.** The three points

$$(0.1, 1.0100502), \qquad (0.2, 1.04081077), \qquad (0.4, 1.1735109)$$

lie on the graph of a certain function $f(x)$. Use these points to estimate $f(0.3)$.

8.5. Complete the following table of divided differences:

i	x_i	$f[x_i]$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$
0	3.2	22.0			
			8.400		
1	2.7	17.8		2.856	
					−0.528
2	1.0	14.2			
3	4.8	38.3			
4	5.6	5.17			

Write the interpolating polynomial of degree 3 that fits the data at all points from $x_0 = 3.2$ to $x_3 = 4.8$.

8.6. Interpolate the data

$$(-1, 2), \quad (0, 0), \quad (1.5, -1), \quad (2, 4),$$

by means of Newton's divided difference interpolating polynomial of degree three. Plot the data and the interpolating polynomial on the same graph.

8.7. Repeat Exercise 2.1 using Newton's divided difference interpolating polynomials.

8.8. Repeat Exercise 2.2 using Newton's divided difference interpolating polynomials.

8.9. Interpolate the data

$$(-1, 2), \quad (0, 0), \quad (1, -1), \quad (2, 4),$$

by means of Gregory–Newton's interpolating polynomial of degree three.

8.10. Interpolate the data

$$(-1, 3), \quad (0, 1), \quad (1, 0), \quad (2, 5),$$

by means of Gregory–Newton's interpolating polynomial of degree three.

8.11. Approximate $f(0.05)$ using the following data and Gregory–Newton's forward interpolating polynomial of degree four.

x	0.0	0.2	0.4	0.6	0.8
$f(x)$	1.00000	1.22140	1.49182	1.82212	2.22554

***8.12.** Approximate $f(0.65)$ using the data in Exercise 2.11 and Gregory–Newton's backward interpolating polynomial of degree four.

8.13. Construct a Hermite interpolating polynomial of degree three for the data

x	$f(x)$	$f'(x)$
8.3	17.56492	3.116256
8.6	18.50515	3.151762

Exercises for Chapter 9

9.1. Consider the numerical differentiation formulae

$$(ND.4) \quad f'(x_0) = \frac{1}{2h}[-3f(x_0) + 4f(x_0 + h) - f(x_0 + 2h)] + \frac{h^2}{3}f^{(3)}(\xi),$$

$$(ND.5) \quad f'(x_0) = \frac{1}{2h}[f(x_0 + h) - f(x_0 - h)] - \frac{h^2}{6}f^{(3)}(\xi),$$

$$(ND.6) \quad f'(x_0) = \frac{1}{12h}[f(x_0 - 2h) - 8f(x_0 - h) + 8f(x_0 + h) - f(x_0 + 2h)] + \frac{h^4}{30}f^{(5)}(\xi),$$

$$(ND.7) \quad f'(x_0) = \frac{1}{12h}[-25f(x_0) + 48f(x_0 + h) - 36f(x_0 + 2h) + 16f(x_0 + 3h) - 3f(x_0 + 4h)] + \frac{h^4}{5}f^{(5)}(\xi),$$

and the table $\{x_n, f(x_n)\}$:

```
x = 1:0.1:1.8; format long; table = [x',(cosh(x)-sinh(x))']
table =
   1.00000000000000   0.36787944117144
   1.10000000000000   0.33287108369808
   1.20000000000000   0.30119421191220
   1.30000000000000   0.27253179303401
   1.40000000000000   0.24659696394161
   1.50000000000000   0.22313016014843
   1.60000000000000   0.20189651799466
   1.70000000000000   0.18268352405273
   1.80000000000000   0.16529888822159
```

For each of the four formulae (DN.4)–(DN.7) with $h = 0.1$,

(a) compute the numerical values

$$ndf = f'(1.2)$$

(deleting the error term in the formulae),

(b) compute the exact value at $x = 1.2$ of the derivative $df = f'(x)$ of the given function

$$f(x) = \cosh x - \sinh x,$$

(c) compute the error

$$\varepsilon = ndf - df;$$

(d) verify that $|\varepsilon|$ is bounded by the absolute value of the error term.

9.2. Use Richardson's extrapolation with $h = 0.4$, $h/2$ and $h/4$ to improve the value $f'(1.4)$ obtained by formula (ND.5) for $f(x) = x^2\, e^{-x}$.

9.3. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+x}$ by the composite trapezoidal rule with $n = 10$.

9.4. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+x}$ by the composite Simpson's rule with $n = 2m = 10$.

9.5. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+2x^2}$ by the composite trapezoidal rule with $n = 10$.

9.6. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+2x^2}$ by the composite Simpson's rule with $n = 2m = 10$.

9.7. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+x^3}$ by the composite trapezoidal rule with h for an error of 10^{-4}.

9.8. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+x^3}$ by the composite Simpson's rule with with h for an error of 10^{-6}.

9.9. Determine the values of h and n to approximate

$$\int_1^3 \ln x\, dx$$

to 10^{-3} by the following composite rules: trapezoidal, Simpson's, and midpoint.

9.10. Same as Exercise 10.9 with

$$\int_0^2 \frac{1}{x+4}\, dx$$

to 10^{-5}.

9.11. Use Romberg integration to compute $R_{3,3}$ for the integral

$$\int_1^{1.5} x^2 \ln x\, dx.$$

9.12. Use Romberg integration to compute $R_{3,3}$ for the integral

$$\int_1^{1.6} \frac{2x}{x^2-4}\, dx.$$

9.13. Apply Romberg integration to the integral

$$\int_0^1 x^{1/3}\,dx$$

until $R_{n-1,n-1}$ and $R_{n,n}$ agree to within 10^{-4}.

Exercises for Chapter 10

Use Euler's method with $h = 0.1$ to obtain a four-decimal approximation for each initial value problem on $0 \le x \le 1$ and plot the numerical solution.

10.1. $y' = e^{-y} - y + 1, \quad y(0) = 1.$

10.2. $y' = x + \sin y, \quad y(0) = 0.$

***10.3.** $y' = x + \cos y, \quad y(0) = 0.$

10.4. $y' = x^2 + y^2, \quad y(0) = 1.$

10.5. $y' = 1 + y^2, \quad y(0) = 0.$

Use the improved Euler method with $h = 0.1$ to obtain a four-decimal approximation for each initial value problem on $0 \le x \le 1$ and plot the numerical solution.

10.6. $y' = e^{-y} - y + 1, \quad y(0) = 1.$

10.7. $y' = x + \sin y, \quad y(0) = 0.$

***10.8.** $y' = x + \cos y, \quad y(0) = 0.$

10.9. $y' = x^2 + y^2, \quad y(0) = 1.$

10.10. $y' = 1 + y^2, \quad y(0) = 0.$

Use the Runge–Kutta method of order 4 with $h = 0.1$ to obtain a six-decimal approximation for each initial value problem on $0 \le x \le 1$ and plot the numerical solution.

10.11. $y' = x^2 + y^2, \quad y(0) = 1.$

10.12. $y' = x + \sin y, \quad y(0) = 0.$

***10.13.** $y' = x + \cos y, \quad y(0) = 0.$

10.14. $y' = e^{-y}, \quad y(0) = 0.$

10.15. $y' = y^2 + 2y - x, \quad y(0) = 0.$

Use the Matlab `ode23` embedded pair of order 3 with $h = 0.1$ to obtain a six-decimal approximation for each initial value problem on $0 \le x \le 1$ and estimate the local truncation error by means of the given formula.

10.16. $y' = x^2 + 2y^2, \quad y(0) = 1.$

10.17. $y' = x + 2\sin y, \quad y(0) = 0.$

10.18. $y' = x + 2\cos y, \quad y(0) = 0.$

10.19. $y' = e^{-y}, \quad y(0) = 0.$

10.20. $y' = y^2 + 2y - x, \quad y(0) = 0.$

Use the Adams–Bashforth–Moulton three-step predictor-corrector method with $h = 0.1$ to obtain a six-decimal approximation for each initial value problem on $0 \le x \le 1$, estimate the local error at $x = 0.5$, and plot the numerical solution.

10.21. $y' = x + \sin y, \quad y(0) = 0.$

10.22. $y' = x + \cos y, \quad y(0) = 0.$

10.23. $y' = y^2 - y + 1, \quad y(0) = 0.$

Use the Adams–Bashforth–Moulton four-step predictor-corrector method with $h = 0.1$ to obtain a six-decimal approximation for each initial value problem on $0 \le x \le 1$, estimate the local error at $x = 0.5$, and plot the numerical solution.

10.24. $y' = x + \sin y, \quad y(0) = 0.$

***10.25.** $y' = x + \cos y, \quad y(0) = 0.$

10.26. $y' = y^2 - y + 1, \quad y(0) = 0.$

Solutions to Starred Exercises

Solutions to Exercises from Chapters 1 to 6

Ex. 1.3. Solve $(1+x^2)y' = \cos^2 y$.

SOLUTION. Separate the variables,

$$\frac{dy}{\cos^2 y} = \frac{dx}{1+x^2},$$

and integrate,

$$\int \frac{dy}{\cos^2 y} = \int \frac{dx}{1+x^2} + c,$$

to get

$$\tan y = \arctan x + c,$$

so the general solution is $y = \arctan(\arctan x + c)$. □

Ex. 1.11. Solve $xy' = y + \sqrt{y^2 - x^2}$.

SOLUTION. Rewrite the equation as

$$x\,\frac{dy}{dx} = y + \sqrt{y^2 - x^2}$$

or

$$\left(y + \sqrt{y^2 - x^2}\right) dx - x\,dy = 0,$$

so

$$M(x,y) = y + \sqrt{y^2 - x^2} \quad \text{and} \quad N(x,y) = -x,$$

which are both homogeneous of degree 1. So let

$$y = ux \quad \text{and} \quad dy = u\,dx + x\,dy$$

to get

$$\left(ux + \sqrt{u^2x^2 - x^2}\right) dx - x(u\,dx + x\,du) = 0,$$

or

$$ux\,dx + x\sqrt{u^2 - 1}\,dx - xu\,dx - x^2\,du = 0,$$

or

$$x\sqrt{u^2 - 1}\,dx - x^2\,du = 0,$$

or

$$\frac{du}{\sqrt{u^2 - 1}} = \frac{dx}{x}.$$

Integrate

$$\int \frac{du}{\sqrt{u^2 - 1}} = \int \frac{dx}{x} + c$$

to get

$$\operatorname{arccosh} u = \ln|x| + c,$$

and so

$$u = \cosh(\ln|x| + c)$$

and the general solution is $y = x\cosh(\ln|x| + c)$. □

Ex. 1.21. Solve the initial value problem

$$(2xy - 3)\,dx + (x^2 + 4y)\,dy = 0, \quad y(1) = 2.$$

SOLUTION.

$$M(x,y) = 2xy - 3 \quad \text{and} \quad N(x,y) = x^2 + 4y,$$
$$M_y = 2x \quad \text{and} \quad N_x = 2x,$$
$$M_y = N_x,$$

so the differential equation is exact. Then

$$\begin{aligned} u(x,y) &= \int M(x,y)\,dx + T(y) \\ &= \int (2xy - 3)\,dx + T(y) \\ &= x^2y - 3x + T(y). \end{aligned}$$

But

$$\frac{\partial u}{\partial y} = \frac{\partial}{\partial y}\left(x^2y - 3x + T(y)\right) = x^2 + T'(y) = N(x,y) = x^2 + 4y,$$

so

$$T'(y) = 4y \quad \Rightarrow \quad T(y) = 2y^2,$$

and

$$u(x,y) = x^2y - 3x + 2y^2.$$

Thus, the general solution is

$$x^2y - 3x + 2y^2 = c.$$

But $y(1) = 2$, so

$$(1)^2(2) - 3(1) + 2(2^2) = c \quad \Rightarrow \quad c = 7.$$

Therefore the unique solution is $x^2y - 3x + 2y^2 = 7$. □

Ex. 1.25. Find the general solution of $(x + y^2)\,dx - 2xy\,dy = 0$.

SOLUTION.

$$M(x,y) = x + y^2 \quad \text{and} \quad N(x,y) = -2xy,$$
$$M_y = 2y \quad \text{and} \quad N_x = -2y,$$
$$M_y \neq N_x,$$

so the differential equation is not exact. Let

$$\frac{M_y - N_x}{N} = \frac{4y}{-2xy} = -\frac{2}{x},$$

a function of x only. So

$$\mu(x) = e^{\int -\frac{2}{x}\,dx} = e^{-2\ln x} = e^{\ln x^{-2}} = x^{-2},$$

and the differential equation becomes

$$\left(x^{-1} + x^{-2}y^2\right)\,dx - 2x^{-1}y\,dy = 0.$$

Now,

$$M^*(x,y) = x^{-1} + x^{-2}y^2 \quad \text{and} \quad N^*(x,y) = -2x^{-1}y,$$
$$M^*_y = 2x^{-2}y \quad \text{and} \quad N^*_x = 2x^{-2}y,$$
$$M^*_y = N^*_x,$$

so the differential equation is now exact. Then

$$\begin{aligned} u(x,y) &= \int N^*(x,y)\,dy + T(x) \\ &= \int -2x^{-1}y\,dy + T(x) \\ &= -x^{-1}y^2 + T(x). \end{aligned}$$

But

$$\frac{\partial u}{\partial x} = \frac{\partial}{\partial x}\left(-x^{-1}y^2 + T(x)\right) = x^{-2}y^2 + T'(x) = M^*(x,y) = x^{-1} + x^{-2}y^2,$$

so

$$T'(x) = x^{-1} \quad \Rightarrow \quad T(x) = \ln|x|,$$

and then

$$u(x,y) = \ln|x| - x^{-1}y^2.$$

Thus, the general solution is

$$\ln|x| - x^{-1}y^2 = c,$$

or $x^{-1}y^2 = c + \ln|x|$, or $y^2 = cx + x\ln|x|$. □

Ex. 1.39. Solve the initial-value problem $y' + y\cos x = \cos x, \quad y(0) = 1.$

Solution. This is a first-order linear differential equation of the form $y' + f(x)y = r(x)$ with $f(x) = \cos x$ and $r(x) = \cos x$, so the general solution is

$$\begin{aligned} y(x) &= e^{-\int \cos x\,dx}\left[\int e^{\int \cos x\,dx}\cos x\,dx + c\right] \\ &= e^{-\sin x}\left[\int e^{\sin x}\cos x\,dx + c\right] \\ &= e^{-\sin x}\left[e^{\sin x} + c\right] \\ &= 1 + c\,e^{-\sin x}. \end{aligned}$$

Then,

$$y(0) = 1 \quad \Rightarrow \quad 1 = 1 + c\,e^{-\sin 0} = 1 + c \quad \Rightarrow \quad c = 0,$$

and the unique solution is $y(x) = 1$. □

Ex. 2.3. Solve the differential equation $y'' - 9y' + 20y = 0.$

Solution. The characteristic equation is

$$\lambda^2 - 9\lambda + 20 = (\lambda - 4)(\lambda - 5) = 0,$$

and the general solution is $y(x) = c_1\,e^{4x} + c_2\,e^{5x}$. □

Ex. 2.13. Solve the Euler–Cauchy differential equation $x^2y'' + 3xy' - 3y = 0.$

SOLUTION. The characteristic equation is

$$m(m-1)+3m-3 = m^2+2m-3 = (m+3)(m-1) = 0,$$

so the general solution is $y(x) = c_1 x + c_2 x^{-3}$. □

Ex. 3.1. Solve the constant coefficient differential equation $y''' + 6y'' = 0$.

SOLUTION. The characteristic equation is

$$\lambda^3 + 6\lambda^2 = \lambda^2(\lambda+6) = 0,$$

so the general solution is $y(x) = c_1 + c_2 x + c_3\, e^{-6x}$. □

Ex. 3.9. Determine whether the functions

$$y_1(x) = x, \quad y_2(x) = x^2, \quad y_3(x) = 2x - 5x^2$$

are linearly dependent or independent on $-\infty < x < +\infty$.

SOLUTION. Since

$$y_3(x) = 2x - 5x^2 = 2y_1(x) - 5y_2(x),$$

the function are linearly dependent. □

Ex. 3.13. Show by computing the Wronskian that the functions e^x, e^{2x}, e^{-x} are linearly independent on the interval $-\infty < x < +\infty$.

SOLUTION.

$$\begin{aligned}
\begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1'' & y_2' & y_3'' \end{vmatrix} &= \begin{vmatrix} e^x & e^{2x} & e^{-x} \\ e^x & 2\,e^{2x} & -e^{-x} \\ e^x & 4\,e^{2x} & e^{-x} \end{vmatrix} \\
&= (e^x)(e^{2x})(e^{-x}) \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & 4 & 1 \end{vmatrix} \\
&= e^{2x}\left[\begin{vmatrix} 2 & -1 \\ 4 & 1 \end{vmatrix} - \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 2 \\ 1 & 4 \end{vmatrix} \right] \\
&= e^{2x}\big[(2+4) - (1+1) + (4-2)\big] \\
&= 6\,e^{2x} \neq 0
\end{aligned}$$

for any x. Since the three functions have continuous derivatives up to order 3, by Corollary 3.2 they are solutions of the same differential equation; therefore they are linearly independent on $-\infty < x < +\infty$. □

Ex. 3.25. Solve the nonhomogeneous differential equation $y'' - y' = e^x \sin x$.

SOLUTION. The corresponding homogeneous equation, $y'' - y' = 0$, has characteristic equation

$$\lambda^2 - \lambda = \lambda(\lambda - 1) = 0$$

and general solution

$$y_h(x) = c_1 + c_2\, e^x.$$

The right-hand side is $r(x) = e^x \sin x$, so we can use Undetermined Coefficients and our guess for the particular solution is

$$y_p(x) = a\, e^x \cos x + b\, e^x \sin x.$$

Then

$$y_p'(x) = a\,e^x\cos x - a\,e^x\sin x + b\,e^x\sin x + b\,e^x\cos x$$

and

$$\begin{aligned} y_p''(x) &= a\,e^x\cos x - 2a\,e^x\sin x - a\,e^x\cos x + b\,e^x\sin x + 2b\,e^x\cos x - b\,e^x\sin x \\ &= -2a\,e^x\sin x + 2b\,e^x\cos x. \end{aligned}$$

Therefore,

$$\begin{aligned} y_p''(x) - y_p'(x) &= -2a\,e^x\sin x + 2b\,e^x\cos x \\ &\quad - (a\,e^x\cos x - a\,e^x\sin x + b\,e^x\sin x + b\,e^x\cos x) \\ &= -a\,e^x\cos x - a\,e^x\sin x + b\,e^x\cos x - b\,e^x\sin x \\ &= (b-a)\,e^x\cos x + (-a-b)\,e^x\sin x \\ &= r(x) = e^x\sin x. \end{aligned}$$

So $b - a = 0$ and $-a - b = 1 \Rightarrow a = b = -1/2$. Thus, the particular solution is

$$y_p(x) = -\frac{1}{2}\,e^x\cos x - \frac{1}{2}\,e^x\sin x$$

and the general solution is

$$\begin{aligned} y(x) &= y_h(x) + y_p(x) \\ &= c_1 + c_2\,e^x - \frac{1}{2}\,e^x\cos x - \frac{1}{2}\,e^x\sin x. \end{aligned}$$

□

Ex. 3.35. Solve the nonhomogeneous differential equation

$$y'' + 3y' + 2y = \frac{1}{1+e^x}.$$

SOLUTION. The corresponding homogeneous equation, $y'' + 3y' + 2y = 0$, has characteristic equation

$$\lambda^2 + 3\lambda + 2 = (\lambda+1)(\lambda+2) = 0$$

and general solution

$$y_h(x) = c_1\,e^{-x} + c_2\,e^{-2x}.$$

The right-hand side $r(x) = \frac{1}{1+e^x}$ admits infinitely many independent derivatives so we must use Variation of Parameters. The equations for $c_1'(x)$ and $c_1'(x)$ are

$$\begin{aligned} c_1'y_1 + c_2'y_2 &= 0, \\ c_1'y_1' + c_2'y_2' &= \frac{1}{1+e^x}, \end{aligned}$$

or

$$c_1'\,e^{-x} + c_2'\,e^{-2x} = 0, \tag{a}$$

$$-c_1'\,e^{-x} - 2c_2'\,e^{-2x} = \frac{1}{1+e^x}. \tag{b}$$

Thus, (a)+(b) gives

$$-c_2'\,e^{-2x} = \frac{1}{1+e^x}$$

or

$$\begin{aligned} c_2' &= \frac{-e^{2x}}{1+e^x} = \frac{1-e^{2x}-1}{1+e^x} \\ &= \frac{(1+e^x)(1-e^x)-1}{1+e^x} = 1-e^x-\frac{1}{1+e^x}, \end{aligned}$$

so

$$c_2'(x) = 1-e^x-\frac{e^{-x}}{e^{-x}+1},$$

and

$$c_2(x) = x-e^x+\ln\left(e^{-x}+1\right).$$

Then, (a) says

$$c_1' = -c_2'\,e^{-x} = \frac{e^x}{1+e^x} \quad \Rightarrow \quad c_1(x) = \ln\left(e^x+1\right).$$

So the particular solution is

$$\begin{aligned} y_p(x) &= c_1(x)y_1(x)+c_2(x)y_2(x) \\ &= \ln(e^x+1)\,e^{-x}+\left(x-e^x+\ln(e^{-x}+1)\right)e^{-2x} \\ &= e^{-x}\ln(e^x+1)+x\,e^{-2x}-e^{-x}+e^{-2x}\ln(e^{-x}+1). \end{aligned}$$

Since e^{-x} appears in $y_h(x)$, we can delete that term and take

$$y_p(x) = e^{-x}\ln(e^x+1)+x\,e^{-2x}+e^{-2x}\ln(e^{-x}+1).$$

The general solution is

$$y(x) = c_1\,e^{-x}+c_2\,e^{-2x}+e^{-x}\ln(e^x+1)+x\,e^{-2x}+e^{-2x}\ln(e^{-x}+1). \quad \square$$

Ex. 4.3. Solve the system $\boldsymbol{y}' = \begin{bmatrix} -1 & 1 \\ 4 & -1 \end{bmatrix}\boldsymbol{y}$.

SOLUTION. Letting A be the matrix of the system, we have

$$\begin{aligned} \det(A-\lambda I) &= \begin{vmatrix} -1-\lambda & 1 \\ 4 & -1-\lambda \end{vmatrix} \\ &= (-1-\lambda)^2-4 = \lambda^2+2\lambda-3 = (\lambda+3)(\lambda-1) = 0, \end{aligned}$$

so the eigenvalues are $\lambda_1 = -3$ and $\lambda_2 = 1$.
For $\lambda_1 = -3$,

$$(A-\lambda_1 I)\boldsymbol{u}_1 = \begin{bmatrix} -1 & 1 \\ 4 & -1 \end{bmatrix}\boldsymbol{u}_1 = \boldsymbol{0}, \quad \Rightarrow \quad \boldsymbol{u}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

For $\lambda_2 = 1$,

$$(A-\lambda_2 I)\boldsymbol{u}_2 = \begin{bmatrix} -2 & 1 \\ 4 & -2 \end{bmatrix}\boldsymbol{u}_2 = \boldsymbol{0}, \quad \Rightarrow \quad \boldsymbol{u}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

The genral solution is

$$\boldsymbol{y}(x) = c_1\,e^{-3x}\begin{bmatrix} 1 \\ -2 \end{bmatrix}+c_2\,e^{x}\begin{bmatrix} 1 \\ 2 \end{bmatrix}. \quad \square$$

Ex. 5.5. Find the Laplace transform of $f(t) = \cos^2 t = \dfrac{1}{2}\big(1+\cos(2t)\big)$.

SOLUTION.

$$F(s) = \mathcal{L}\{\cos^2 t\} = \mathcal{L}\left\{\frac{1}{2}\big(1+\cos(2t)\big)\right\} = \frac{1}{2}\left(\frac{1}{s} + \frac{s}{s^2+4}\right). \quad \square$$

Ex. 5.12. Find the Laplace transform of $f(t) = u(t-1)t^2$.

SOLUTION.

$$F(s) = \mathcal{L}\{u(t-1)t^2\} = e^{-s}\mathcal{L}\{t^2+2t+1\} = e^{-s}\left(\frac{2}{s^3} + \frac{2}{s^2} + \frac{1}{s}\right),$$

where the transformation $f(t-1) = t^2$ into $f(t) = (t+1)^2 = t^2+2t+1$ has been used. $\square$

Ex. 5.23. Find the inverse Laplace transform of $F(s) = \dfrac{2s+1}{s^2+5s+6}$.

SOLUTION.

$$F(s) = \frac{2s+1}{s^2+5s+6} = \frac{2s+1}{(s+3)(s+2)} = \frac{A}{s+3} + \frac{B}{s+2} = \frac{5}{s+3} - \frac{3}{s+2}$$

since

$$A(s+2) + B(s+3) = 2s+1 \quad \Rightarrow \quad \begin{cases} A+B = 2, & A = 5, \\ 2A+3B = 1, & B = -3. \end{cases}$$

So

$$f(t) = \mathcal{L}^{-1}\left\{\frac{5}{s+3} - \frac{3}{s+2}\right\} = 5\,e^{-3t} - 3\,e^{-2t}. \quad \square$$

Ex. 5.29. Find the inverse Laplace transform of $F(s) = \dfrac{e^{-3s}}{s^2(s-1)}$.

SOLUTION.

$$\begin{aligned} F(s) &= \frac{e^{-3s}}{s^2(s-1)} = e^{-3s}\left[\frac{As+B}{s^2} + \frac{C}{s+1}\right] \\ &= e^{-3s}\left[\frac{-s-1}{s^2} + \frac{1}{s-1}\right] = e^{-3s}\left[\frac{1}{s-1} - \frac{1}{s} - \frac{1}{s^2}\right] \end{aligned}$$

since

$$(As+B)(s-1) + Cs^2 = 1 \quad \Rightarrow \quad \begin{cases} A+C = 0, & B = -1, \\ B-A = 0, & A = -1, \\ -B = 1, & C = 1. \end{cases}$$

Thus

$$\begin{aligned} f(t) &= \mathcal{L}^{-1}\left\{\frac{e^{-3}}{s^2(s-1)}\right\} = u(t-3)g(t-3) \\ &= u(t-3)\left[e^{t-3} - 1 - (t-3)\right] = u(t-3)\left(e^{t-3} - t + 2\right). \end{aligned}$$

$\square$

Ex. 5.36. Find the Laplace transform of $f(t) = t\,e^{-t}\cos t$.

SOLUTION.

$$\begin{aligned} F(s) &= \mathcal{L}\left\{t\,e^{-t}\cos t\right\} = -\frac{d}{ds}\mathcal{L}\left\{e^{-t}\cos t\right\} \\ &= -\frac{d}{ds}\left[\frac{s+1}{(s+1)^2+1}\right] = -\left[\frac{(1)\left((s+1)^2+1\right)-2(s+1)(s+1)}{\left((s+1)^2+1\right)^2}\right] \\ &= \frac{(s+1)^2-1}{\left((s+1)^2+1\right)^2} = \frac{s^2+2s}{\left(s^2+2s+2\right)^2}. \end{aligned}$$

□

Ex. 5.51. Solve $y'' + 4y' = u(t-1), \quad y(0) = 0, \quad y'(0) = 0$, by Laplace transform.

SOLUTION. Let $Y(s) = \mathcal{L}\{y(t)\}$ and take the Laplace transform of the equation to get

$$\begin{gathered} \mathcal{L}\{y''\} + 4\mathcal{L}\{y'\} = \mathcal{L}\{u(t-1)\}, \\ s^2Y(s) - sy(0) - y'(0) + 4(sY(s) - y(0)) = \frac{e^{-s}}{s}, \\ \left(s^2+4s\right)Y(s) = \frac{e^{-s}}{s} \quad \Rightarrow \quad Y(s) = \frac{e^{-s}}{s^2(s+4)}. \end{gathered}$$

By partial fractions,

$$Y(s) = e^{-s}\left[\frac{As+B}{s^2} + \frac{C}{s+4}\right].$$

Now

$$(As+B)(s+4) + Cs^2 = 1 \quad \Rightarrow \quad \begin{cases} A + C = 0, & B = 1/4, \\ 4A + B = 0, & A = -1/16 \\ 4B = 1, & C = 1/16. \end{cases}$$

So

$$Y(s) = \frac{e^{-s}}{16}\left[\frac{-s+4}{s^2} + \frac{1}{s+4}\right] = \frac{e^{-s}}{16}\left[\frac{4}{s^2} - \frac{1}{s} + \frac{1}{s+4}\right]$$

and the solution to the initial value problem is

$$\begin{aligned} y(t) &= \mathcal{L}^{-1}\{Y(s)\} = \mathcal{L}^{-1}\left\{\frac{e^{-s}}{16}\left[\frac{4}{s^2} - \frac{1}{s} + \frac{1}{s+4}\right]\right\} \\ &= u(t-1)g(t-1) = \frac{1}{16}\,u(t-1)\left[4(t-1) - 1 + e^{-4(t-1)}\right] \\ &= \frac{1}{16}\,u(t-1)\left[4t - 5 + e^{-4(t-1)}\right]. \end{aligned}$$

□

Ex. 6.12. Find the power series solutions of the equation $y'' - xy' - y = 0$.

SOLUTION. Let

$$y(x) = \sum_{m=0}^{\infty} a_m x^m,$$

so

$$y'(x) = \sum_{m=0}^{\infty} m a_m x^{m-1} \quad \text{and} \quad y''(x) = \sum_{m=0}^{\infty} m(m-1) a_m x^{m-2}.$$

Then

$$\begin{aligned} y'' - xy'' - y &= \sum_{m=0}^{\infty} m(m-1)a_m x^{m-2} - x\sum_{m=0}^{\infty} ma_m x^{m-1} - \sum_{m=0}^{\infty} a_m x^m \\ &= \sum_{m=0}^{\infty} (m+1)(m+2)a_{m+2}x^m - \sum_{m=0}^{\infty} ma_m x^m - \sum_{m=0}^{\infty} a_m x^m \\ &= \sum_{m=0}^{\infty} \big[(m+1)(m+2)a_{m+2} - ma_m x^m - a_m\big]x^m \\ &= 0, \qquad \text{for all } x, \end{aligned}$$

so

$$(m+1)(m+2)a_{m+2} - (m+1)a_m = 0, \qquad \text{for all } m,$$

and the recurrence relation is

$$a_{m+2} = \frac{a_m}{m+2}.$$

So for even m,

$$a_2 = \frac{a_0}{2}, \quad a_4 = \frac{a_2}{4} = \frac{a_0}{2\cdot 4}, \quad a_6 = \frac{a_4}{6} = \frac{a_0}{2\cdot 4\cdot 6}, \quad a_8 = \frac{a_6}{8} = \frac{a_0}{2\cdot 4\cdot 6\cdot 8} = \frac{a_0}{2^4 4!}.$$

Thus, the general pattern is

$$a_{2k} = \frac{a_0}{2^k k!}.$$

For odd m, we have

$$a_3 = \frac{a_1}{3}, \quad a_5 = \frac{a_3}{5} = \frac{a_1}{3\cdot 5}, \quad a_7 = \frac{a_5}{7} = \frac{a_1}{5\cdot 5\cdot 7},$$

and the general pattern is

$$a_{2k+1} = \frac{a_1}{1\cdot 3\cdot 5\cdots(2k+1)} = \frac{a_1 2^k k!}{1\cdot 3\cdot 5\cdots(2k+1)2\cdot 4\cdot 6\cdots(2k)} = \frac{2^k k! a_1}{(2k+1)!}.$$

The general solution is

$$\begin{aligned} y(x) &= \sum_{m=0}^{\infty} a_m x^m = \sum_{k=0}^{\infty} a_{2k}x^{2k} + \sum_{k=0}^{\infty} a_{2k+1}x^{2k+1} \\ &= a_0 \sum_{k=0}^{\infty} \frac{x^{2k}}{k!2^k} + a_1 \sum_{k=0}^{\infty} \frac{2^k k!}{(2k+1)!}x^{2k+1}. \end{aligned}$$

□

Ex. 6.31. Evaluate $I = \displaystyle\int_{0.2}^{1.5} e^{-x^2}\,dx$ by the three-point Gaussian Quadrature formula.

SOLUTION. Use the substitution

$$x = \frac{1}{2}\left[0.2(1-t) + 1.5(t+1)\right] = \frac{1}{2}\left(1.3t + 1.7\right) = 0.65t + 0.85$$

and $dx = 0.65\,dt$ to get the limits -1 and 1. Then

$$\begin{aligned}
\int_{0.2}^{1.5} e^{-x^2}\,dx &= \int_{-1}^{1} e^{-(0.65t+0.85)^2}(0.65)\,dt \\
&\approx 0.65\sum_{j=1}^{3} w_j f(t_j) \\
&= 0.65\left[\frac{5}{9}\,e^{-(0.65(-0.774\,596\,669\,2)+0.85)^2} + \frac{8}{9}\,e^{-(0.65t(0)+0.85)^2}\right. \\
&\qquad \left. + \frac{5}{9}\,e^{-(0.65(0.774\,596\,669\,2)+0.85)^2}\right] \\
&= 0.65\left[\frac{5}{9}\,e^{-0.120\,070\,7} + \frac{8}{9}\,e^{-0.7225} + \frac{5}{9}\,e^{-1.831\,929\,2}\right] \\
&= 0.65(0.492\,698\,7 + 0.431\,588\,3 + 0.088\,946, 8) \\
&= 0.658\,602.
\end{aligned}$$

□

Solutions to Exercises from Chapter 7

Ex. 7.11. Sketch the function

$$f(x) = e^{-x} - \tan x$$

and compute a root of the equation $f(x) = 0$ to six decimals by means of Newton's method with $x_0 = 1$.

SOLUTION. We use the `newton1_11` M-file

```
function f = newton1_11(x); % Exercise 1.11.
f = x - (exp(-x) - tan(x))/(-exp(-x) - sec(x)^2);
```

We iterate Newton's method and monitor convergence to six decimal places.

```
>> xc = input('Enter starting value:'); format long;
Enter starting value:1
>> xc = newton1_11(xc)
xc = 0.68642146135728
>> xc = newton1_11(xc)
xc = 0.54113009740473
>> xc = newton1_11(xc)
xc = 0.53141608691193
>> xc = newton1_11(xc)
xc = 0.53139085681581
>> xc = newton1_11(xc)
xc = 0.53139085665216
```

All the digits in the last value of xc are exact. Note the convergence of order 2. Hence the root is $xc = 0.531391$ to six decimals.

We plot the two functions and their difference. The x-coordinate of the point of intersection of the two functions is the root of their difference.

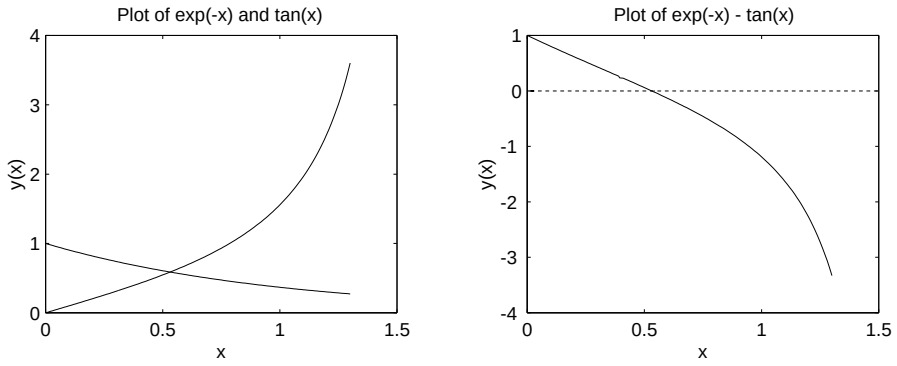

FIGURE 12.1. Graph of two functions and their difference for Exercise 8.11.

```
x=0:0.01:1.3;
subplot(2,2,1); plot(x,exp(-x),x,tan(x));
title('Plot of exp(-x) and tan(x)'); xlabel('x'); ylabel('y(x)');
subplot(2,2,2); plot(x,exp(-x)-tan(x),x,0);
title('Plot of exp(-x)-tan(x)'); xlabel('x'); ylabel('y(x)');
print -deps Fig9_2
```

□

Ex. 7.21. Compute a root of the equation $f(x) = x - \tan x$ with the secant method with starting values $x_0 = 1$ and $x_1 = 0.5$. Find the order of convergence to the root.

SOLUTION. Since

$$f(0) = f'(0) = f''(0) = 0, \qquad f'''(0) \neq 0,$$

$x = 0$ is a triple root.

```
x0 = 1; x1 = 0.5;  % starting values
x = zeros(20,1);
x(1) = x0; x(2) = x1;
for n = 3:20
  x(n) = x(n-1) -(x(n-1)-x(n-2)) ...
        /(x(n-1)-tan(x(n-1))-x(n-2)+tan(x(n-2)))*(x(n-1)-tan(x(n-1)));
end
dx = abs(diff(x));
p = 1;  % checking convergence of order 1
dxr = dx(2:19)./(dx(1:18).^p);
table = [[0:19]' x [0; dx] [0; 0; dxr]]
table =
   n          x_n      x_n - x_{n-1}     |x_n - x_{n-1}|
                                           /|x_{n-1} - x_{n-2}|

   0   1.00000000000000
   1   0.50000000000000   0.50000000000000
   2   0.45470356524435   0.04529643475565   0.09059286951131
```

3	0.32718945543123	0.12751410981312	2.81510256824784
4	0.25638399918811	0.07080545624312	0.55527546204022
5	0.19284144711319	0.06354255207491	0.89742451283310
6	0.14671560243705	0.04612584467614	0.72590481763723
7	0.11082587909404	0.03588972334302	0.77808273420254
8	0.08381567002072	0.02701020907332	0.75258894628889
9	0.06330169146740	0.02051397855331	0.75948980985777
10	0.04780894321090	0.01549274825651	0.75522884145761
11	0.03609714636358	0.01171179684732	0.75595347277403
12	0.02725293456160	0.00884421180198	0.75515413367179
13	0.02057409713542	0.00667883742618	0.75516479882196
14	0.01553163187404	0.00504246526138	0.75499146627099
15	0.01172476374403	0.00380686813002	0.75496169684658
16	0.00885088980844	0.00287387393559	0.75491817353192
17	0.00668139206035	0.00216949774809	0.75490358892216
18	0.00504365698583	0.00163773507452	0.75489134568691
19	0.00380735389990	0.00123630308593	0.75488588182657

Therefore, $x_{19} = 0.0038$ is an approximation to the triple root $x = 0$. Since the ratio

$$\frac{|x_n - x_{n-1}|}{|x_{n-1} - x_{n-2}|} \to 0.75 \approx \text{ constant}$$

as n grows, we conclude that the method converges to order 1.

Convergence is slow to a triple root. In general, the secant method may not converge at all to a multiple root. □

Solutions to Exercises for Chapter 8

Ex. 8.4. The three points

$$(0.1, 1.0100502), \qquad (0.2, 1.04081077), \qquad (0.4, 1.1735109)$$

lie on the graph of a certain function $f(x)$. Use these points to estimate $f(0.3)$.

Solution. We have

$$f[0.1, 0.2] = \frac{1.04081077 - 1.0100502}{0.1} = 0.307606,$$

$$f[0.2, 0.4] = \frac{1.1735109 - 1.04081077}{0.2} = 0.663501$$

and

$$f[0.1, 0.2, 0, 4] = \frac{0.663501 - 0.307606}{0.3} = 1.18632.$$

Therefore,

$$p_2(x) = 1.0100502 + (x - 0.1) \times 0.307606 + (x - 0.1)(x - 0.2) \times 1.18632$$

and

$$p_2(0.3) = 1.0953. \quad \square$$

Ex. 8.12. Approximate $f(0.65)$ using the data in Exercise 2.10

x	0.0	0.2	0.4	0.6	0.8
$f(x)$	1.00000	1.22140	1.49182	1.82212	2.22554

and Gregory–Newton's backward interpolating polynomial of degree four.

SOLUTION. We construct the difference table.

```
f = [1 1.2214 1.49182 1.82212 2.22554];
ddt = [f' [0 diff(f)]' [0 0 diff(f,2)]' [0 0 0 diff(f,3)]' ...
      [0 0 0 0 diff(f,4)]']
```

The backward difference table is

n	x_n	f_n	∇f_n	$\nabla^2 f_n$	$\nabla^3 f_n$	$\nabla^4 f_n$
0	0.0	1.0000				
			0.22140			
1	0.2	1.2214		0.04902		
			0.27042		0.01086	
2	0.4	1.4918		0.05998		0.00238
			0.33030		0.01324	
3	0.6	1.8221		0.07312		
			0.40342			
4	0.8	2.2255				

```
s = (0.65-0.80)/0.2  % the variable s
s =  -0.7500
format long
p4 = ddt(5,1) + s*ddt(5,2) + s*(s+1)*ddt(5,3)/2 ....
    + s*(s+1)*(s+2)*ddt(5,4)/6 + s*(s+1)*(s+2)*(s+3)*ddt(5,5)/24
p4 = 1.91555051757812
```

□

Solutions to Exercises for Chapter 10

The M-file `exr5_25` for Exercises 10.3, 10.8, 10.13 and 10.25 is

```
function yprime = exr5_25(x,y); % Exercises 10.3, 10.8, 10.13 and 10.25.
yprime = x+cos(y);
```

Ex. 10.3. Use Euler's method with $h = 0.1$ to obtain a four-decimal approximation for the initial value problem

$$y' = x + \cos y, \quad y(0) = 0$$

on $0 \le x \le 1$ and plot the numerical solution.

SOLUTION. **The Matlab numeric solution.**— Euler's method applied to the given differential equation:

```
clear
h = 0.1; x0= 0; xf= 1; y0 = 0;
n = ceil((xf-x0)/h); % number of steps
%
count = 2; print_time = 1; % when to write to output
x = x0; y = y0; % initialize x and y
output1 = [0 x0 y0];
for i=1:n
 z = y + h*exr5_25(x,y);
 x = x + h;
   if count > print_time
```

```
      output1 = [output1; i x z];
      count = count - print_time;
   end
 y = z;
count = count + 1;
end
output1
save output1 %for printing the graph
```

The command `output1` prints the values of n, x, and y.

```
         n                  x                  y
                  0                  0                  0
   1.00000000000000   0.10000000000000   0.10000000000000
   2.00000000000000   0.20000000000000   0.20950041652780
   3.00000000000000   0.30000000000000   0.32731391010682
   4.00000000000000   0.40000000000000   0.45200484393704
   5.00000000000000   0.50000000000000   0.58196216946658
   6.00000000000000   0.60000000000000   0.71550074191996
   7.00000000000000   0.70000000000000   0.85097722706339
   8.00000000000000   0.80000000000000   0.98690209299587
   9.00000000000000   0.90000000000000   1.12202980842386
  10.00000000000000   1.00000000000000   1.25541526027779
```

□

Ex. 10.8. Use the improved Euler method with $h = 0.1$ to obtain a four-decimal approximation for the initial value problem

$$y' = x + \cos y, \quad y(0) = 0$$

on $0 \leq x \leq 1$ and plot the numerical solution.

SOLUTION. **The Matlab numeric solution.**— The improved Euler method applied to the given differential equation:

```
clear
h = 0.1; x0= 0; xf= 1; y0 = 0;
n = ceil((xf-x0)/h); % number of steps
%
count = 2; print_time = 1; % when to write to output
x = x0; y = y0; % initialize x and y
output2 = [0 x0 y0];
for i=1:n
 zp = y + h*exr5_25(x,y); % Euler's method
 z  = y + (1/2)*h*(exr5_25(x,y)+exr5_25(x+h,zp));
 x = x + h;
   if count > print_time
      output2 = [output2; i x z];
      count = count - print_time;
   end
 y = z;
count = count + 1;
```

```
end
output2
save output2 %for printing the graph
```

The command `output2` prints the values of n, x, and y.

```
         n                  x                  y
                0                  0                  0
   1.00000000000000   0.10000000000000   0.10475020826390
   2.00000000000000   0.20000000000000   0.21833345972227
   3.00000000000000   0.30000000000000   0.33935117091202
   4.00000000000000   0.40000000000000   0.46622105817179
   5.00000000000000   0.50000000000000   0.59727677538612
   6.00000000000000   0.60000000000000   0.73088021271199
   7.00000000000000   0.70000000000000   0.86552867523997
   8.00000000000000   0.80000000000000   0.99994084307400
   9.00000000000000   0.90000000000000   1.13311147003613
  10.00000000000000   1.00000000000000   1.26433264384505
```

□

Ex. 10.13. Use the Runge–Kutta method of order 4 with $h = 0.1$ to obtain a six-decimal approximation for the initial value problem

$$y' = x + \cos y, \quad y(0) = 0$$

on $0 \le x \le 1$ and plot the numerical solution.

SOLUTION. **The Matlab numeric solution.**— The Runge–Kutta method of order 4 applied to the given differential equation:

```
clear
h = 0.1; x0= 0; xf= 1; y0 = 0;
n = ceil((xf-x0)/h); % number of steps
%
count = 2; print_time = 1; % when to write to output
x = x0; y = y0; % initialize x and y
output3 = [0 x0 y0];
for i=1:n
 k1 = h*exr5_25(x,y);
 k2 = h*exr5_25(x+h/2,y+k1/2);
 k3 = h*exr5_25(x+h/2,y+k2/2);
 k4 = h*exr5_25(x+h,y+k3);
 z  = y + (1/6)*(k1+2*k2+2*k3+k4);
 x = x + h;
   if count > print_time
      output3 = [output3; i x z];
      count = count - print_time;
   end
 y = z;
count = count + 1;
end
output3
```

```
save output3 % for printing the graph
```

The command `output3` prints the values of n, x, and y.

```
         n                  x                  y
                0                  0                  0
   1.00000000000000   0.10000000000000   0.10482097362427
   2.00000000000000   0.20000000000000   0.21847505355285
   3.00000000000000   0.30000000000000   0.33956414151249
   4.00000000000000   0.40000000000000   0.46650622608728
   5.00000000000000   0.50000000000000   0.59763447559658
   6.00000000000000   0.60000000000000   0.73130914485224
   7.00000000000000   0.70000000000000   0.86602471267959
   8.00000000000000   0.80000000000000   1.00049620051241
   9.00000000000000   0.90000000000000   1.13371450064800
  10.00000000000000   1.00000000000000   1.26496830711844
```

□

Ex. 10.25. Use the Adams–Bashforth–Moulton four-step predictor-corrector method with $h = 0.1$ to obtain a six-decimal approximation for the initial value problem

$$y' = x + \cos y, \quad y(0) = 0$$

on $0 \le x \le 1$, estimate the local error at $x = 0.5$, and plot the numerical solution.

Solution. **The Matlab numeric solution.**— The initial conditions and the Runge–Kutta method of order 4 are used to obtain the four starting values for the ABM four-step method.

```
clear
h = 0.1; x0= 0; xf= 1; y0 = 0;
n = ceil((xf-x0)/h); % number of steps
%
count = 2; print_time = 1; % when to write to output
x = x0; y = y0; % initialize x and y
output4 = [0 x0 y0 0];
%RK4
for i=1:3
 k1 = h*exr5_25(x,y);
 k2 = h*exr5_25(x+h/2,y+k1/2);
 k3 = h*exr5_25(x+h/2,y+k2/2);
 k4 = h*exr5_25(x+h,y+k3);
 z  = y + (1/6)*(k1+2*k2+2*k3+k4);
  x = x + h;
   if count > print_time
      output4 = [output4; i x z 0];
      count = count - print_time;
   end
 y = z;
count = count + 1;
end
```

```
% ABM4
for i=4:n
 zp = y + (h/24)*(55*exr5_25(output4(i,2),output4(i,3))-...
                  59*exr5_25(output4(i-1,2),output4(i-1,3))+...
                  37*exr5_25(output4(i-2,2),output4(i-2,3))-...
                   9*exr5_25(output4(i-3,2),output4(i-3,3)) );
  z = y + (h/24)*( 9*exr5_25(x+h,zp)+...
                  19*exr5_25(output4(i,2),output4(i,3))-...
                   5*exr5_25(output4(i-1,2),output4(i-1,3))+...
                     exr5_25(output4(i-2,2),output4(i-2,3)) );
  x = x + h;
   if count > print_time
      errest = -(19/270)*(z-zp);
      output4 = [output4; i x z errest];
      count = count - print_time;
   end
 y = z;
count = count + 1;
end
output4
save output4 %for printing the grap
```

The command `output4` prints the values of n, x, and y.

n	x	y	Error estimate
0	0	0	0
1.00000000000000	0.10000000000000	0.10482097362427	0
2.00000000000000	0.20000000000000	0.21847505355285	0
3.00000000000000	0.30000000000000	0.33956414151249	0
4.00000000000000	0.40000000000000	0.46650952510670	-0.00000234408483
5.00000000000000	0.50000000000000	0.59764142006542	-0.00000292485029
6.00000000000000	0.60000000000000	0.73131943222018	-0.00000304450366
7.00000000000000	0.70000000000000	0.86603741396612	-0.00000269077058
8.00000000000000	0.80000000000000	1.00050998975914	-0.00000195879670
9.00000000000000	0.90000000000000	1.13372798977088	-0.00000104794662
10.00000000000000	1.00000000000000	1.26498035231682	-0.00000017019624

□

The numerical solutions for Exercises 10.3, 10.8, 10.13 and 10.25 are plotted by the commands:

```
load output1; load output2; load output3; load output4;
subplot(2,2,1); plot(output1(:,2),output1(:,3));
title('Plot of solution y_n for Exercise 10.3');
xlabel('x_n'); ylabel('y_n');
subplot(2,2,2); plot(output2(:,2),output2(:,3));
title('Plot of solution y_n for Exercise 10.8');
xlabel('x_n'); ylabel('y_n');
subplot(2,2,3); plot(output3(:,2),output3(:,3));
```

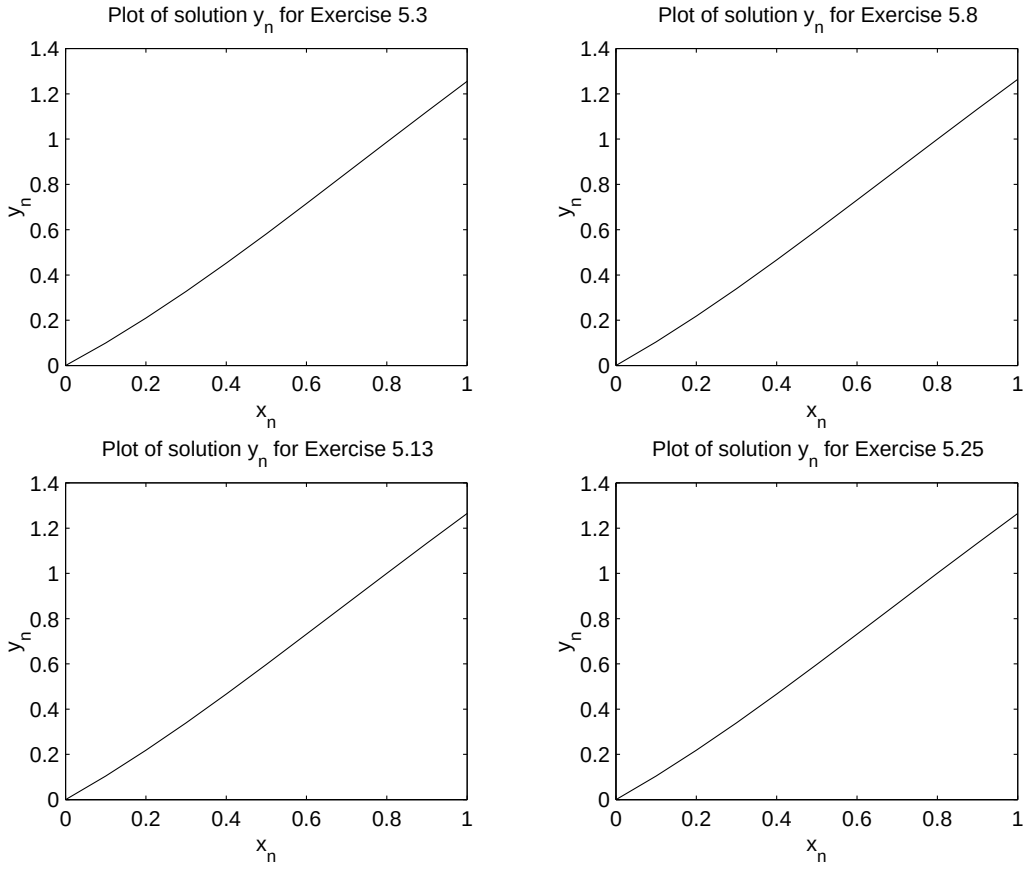

FIGURE 12.2. Graph of numerical solutions of Exercises 10.3 (Euler), 10.8 (improved Euler), 10.13 (RK4) and 10.25 (ABM4).

```
title('Plot of solution y_n for Exercise 10.13');
xlabel('x_n'); ylabel('y_n');
subplot(2,2,4); plot(output4(:,2),output4(:,3));
title('Plot of solution y_n for Exercise 10.25');
xlabel('x_n'); ylabel('y_n');
print -deps Fig9_3
```

Part 4

Formulas and Tables

CHAPTER 13

Formulas and Tables

13.1. Integrating Factor of $M(x,y)\,dx + N(x,y)\,dy = 0$

Consider the first-order homogeneous differential equation

$$M(x,y)\,dx + N(x,y)\,dy = 0. \tag{13.1}$$

If

$$\frac{1}{N}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = f(x)$$

is a function of x only, then

$$\mu(x) = e^{\int f(x)\,dx}$$

is an integrating factor of (13.1).

If

$$\frac{1}{M}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = g(y)$$

is a function of y only, then

$$\mu(y) = e^{-\int g(y)\,dy}$$

is an integrating factor of (13.1).

13.2. Solution of First-Order Linear Differential Equations

The solution of the first order-linear differential equation

$$y' + f(x)y = r(x)$$

is given by the formula

$$y(x) = e^{-\int f(x)\,dx}\left[\int e^{\int f(x)\,dx} r(x)\,dx + c\right].$$

13.3. Laguerre Polynomials on $0 \le x < \infty$

Laguerre polynomials on $0 \le x < \infty$ are defined by the expression

$$L_n(x) = \frac{e^x}{n!}\frac{d^n(x^n e^{-x})}{dx^n}, \quad n = 0, 1, \ldots$$

The first four Laguerre polynomials are (see figure 13.1)

$$L_0(x) = 1, \qquad L_1(x) = 1 - x,$$

$$L_2(x) = 1 - 2x + \frac{1}{2}x^2, \qquad L_3(x) = 1 - 3x + \frac{3}{2}x^2 - \frac{1}{6}x^3.$$

The $L_n(x)$ can be obtained by the three-point recurrence formula

$$(n+1)L_{n+1}(x) = (2n + 1 - x)L_n(x) - nL_{n-1}(x).$$

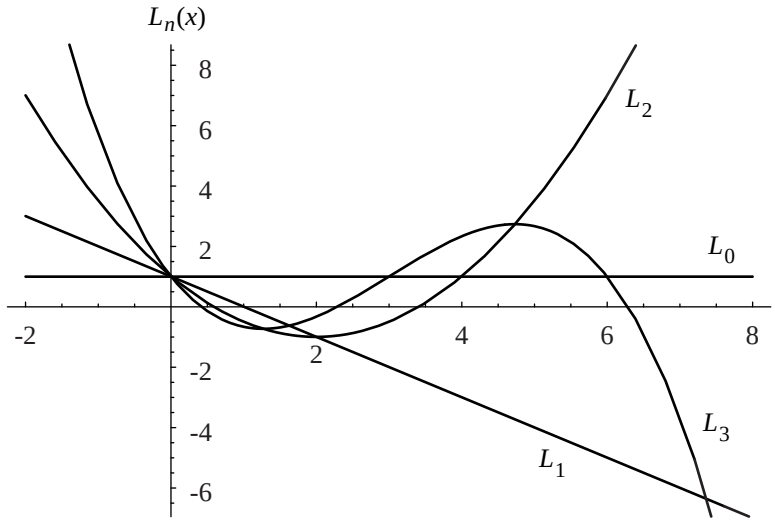

FIGURE 13.1. Plot of the first four Laguerre polynomials.

The $L_n(x)$ are solutions of the differential equation

$$xy'' + (1-x)y' + ny = 0$$

and satisfy the orthogonality relations with weight $p(x) = e^{-x}$

$$\int_0^\infty e^{-x} L_m(x) L_n(x)\, dx = \begin{cases} 0, & m \neq n, \\ 1, & m = n. \end{cases}$$

13.4. Legendre Polynomials $P_n(x)$ on $[-1, 1]$

1. The Legendre differential equation is

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0, \quad -1 \leq x \leq 1.$$

2. The solution $y(x) = P_n(x)$ is given by the series

$$P_n(x) = \frac{1}{2^n} \sum_{m=0}^{[n/2]} (-1)^m \begin{pmatrix} n \\ m \end{pmatrix} \begin{pmatrix} 2n-2m \\ n \end{pmatrix} x^{n-2m},$$

where $[n/2]$ denotes the greatest integer smaller than or equal to $n/2$.

3. The three-point recurrence relation is

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x).$$

4. The standardization is

$$P_n(1) = 1.$$

5. The square of the norm of $P_n(x)$ is

$$\int_{-1}^{1} [P_n(x)]^2\, dx = \frac{2}{2n+1}.$$

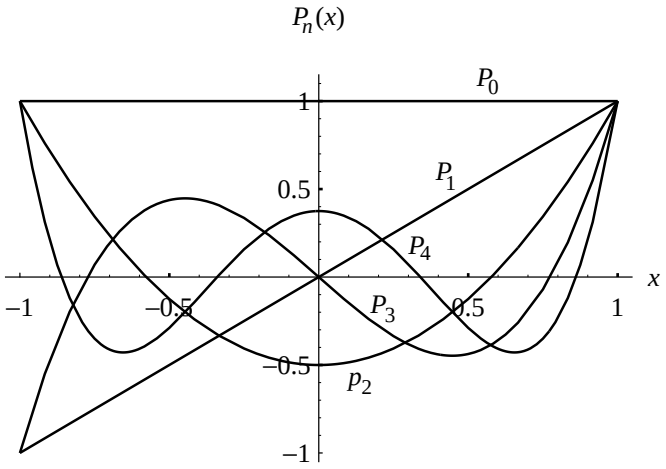

FIGURE 13.2. Plot of the first five Legendre polynomials.

6. Rodrigues's formula is

$$P_n(x) = \frac{(-1)^n}{2^n n!} \frac{d^n}{dx^n} \left[\left(1 - x^2\right)^n \right].$$

7. The generating function is

$$\frac{1}{\sqrt{1 - 2xt + t^2}} = \sum_{n=0}^{\infty} P_n(x) t^n, \quad -1 < x < 1, \ |t| < 1.$$

8. The $P_n(x)$ satisfy the inequality

$$|P_n(x)| \leq 1, \quad -1 \leq x \leq 1.$$

9. The first six Legendre polynomials are:

$$P_0(x) = 1, \qquad P_1(x) = x,$$
$$P_2(x) = \frac{1}{2}\left(3x^2 - 1\right), \qquad P_3(x) = \frac{1}{2}\left(5x^3 - 3x\right),$$
$$P_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right), \qquad P_5(x) = \frac{1}{8}\left(63x^5 - 70x^3 + 15x\right).$$

The graphs of the first five $P_n(x)$ are shown in Fig. 13.2.

13.5. Fourier–Legendre Series Expansion

The Fourier-Legendre series expansion of a function $f(x)$ on $[-1, 1]$ is

$$f(x) = \sum_{n=0}^{\infty} a_n P_n(x), \quad -1 \leq x \leq 1,$$

where

$$a_n = \frac{2n+1}{2} \int_{-1}^{1} f(x) P_n(x)\, dx, \qquad n = 0, 1, 2, \ldots$$

This expansion follows from the orthogonality relations

$$\int_{-1}^{1} P_m(x)P_n(x)\,dx = \begin{cases} 0, & m \neq n, \\ \frac{2}{2n+1}, & m = n. \end{cases}$$

13.6. Table of Integrals

TABLE 13.1. Table of integrals.

1. $\int \tan u\,du = \ln|\sec u| + c$
2. $\int \cot u\,du = \ln|\sin u| + c$
3. $\int \sec u\,du = \ln|\sec u + \tan u| + c$
4. $\int \csc u\,du = \ln|\csc u - \cot u| + c$
5. $\int \tanh u\,du = \ln\cosh u + c$
6. $\int \coth u\,du = \ln\sinh u + c$
7. $\displaystyle\int \frac{du}{\sqrt{a^2-u^2}} = \arcsin\frac{u}{a} + c$
8. $\displaystyle\int \frac{du}{\sqrt{a^2+u^2}} = \ln\left(u+\sqrt{u^2+a^2}\right) + c = \operatorname{arcsinh}\frac{u}{a} + c$
9. $\displaystyle\int \frac{du}{\sqrt{u^2-a^2}} = \ln\left(u+\sqrt{u^2-a^2}\right) + c = \operatorname{arccosh}\frac{u}{a} + c$
10. $\displaystyle\int \frac{du}{a^2+u^2} = \frac{1}{a}\arctan\frac{u}{a} + c$
11. $\displaystyle\int \frac{du}{u^2-a^2} = \frac{1}{2a}\ln\left|\frac{u-a}{u+a}\right| + c$
12. $\displaystyle\int \frac{du}{a^2-u^2} = \frac{1}{2a}\ln\left|\frac{u+a}{u-a}\right| + c$
13. $\displaystyle\int \frac{du}{u(a+bu)} = \frac{1}{a}\ln\left|\frac{u}{a+bu}\right| + c$
14. $\displaystyle\int \frac{du}{u^2(a+bu)} = -\frac{1}{au} + \frac{b}{a^2}\ln\left|\frac{a+bu}{u}\right| + c$
15. $\displaystyle\int \frac{du}{u(a+bu)^2} = \frac{1}{a(a+bu)} - \frac{1}{a^2}\ln\left|\frac{a+bu}{u}\right| + c$
16. $\displaystyle\int x^n \ln ax\,dx = \frac{x^{n+1}}{n+1}\ln ax - \frac{x^{n+1}}{(n+1)^2} + c$

13.7. Table of Laplace Transforms

$$\mathcal{L}\{f(t)\} = \int_0^\infty e^{-st} f(t)\,dt = F(s)$$

TABLE 13.2. Table of Laplace transforms.

	$F(s) = \mathcal{L}\{f(t)\}$	$f(t)$
1.	$F(s-a)$	$e^{at} f(t)$
2.	$F(as+b)$	$\frac{1}{a} e^{-bt/a} f\left(\frac{t}{a}\right)$
3.	$\frac{1}{s} e^{-cs},\ c>0$	$u(t-c) := \begin{cases} 0, & 0 \le t < c \\ 1, & t \ge c \end{cases}$
4.	$e^{-cs}F(s),\ c>0$	$f(t-c)u(t-c)$
5.	$F_1(s)F_2(s)$	$\int_0^t f_1(\tau) f_2(t-\tau)\, d\tau$
6.	$\frac{1}{s}$	1
7.	$\frac{1}{s^{n+1}}$	$\frac{t^n}{n!}$
8.	$\frac{1}{s^{a+1}}$	$\frac{t^a}{\Gamma(a+1)}$
9.	$\frac{1}{\sqrt{s}}$	$\frac{1}{\sqrt{\pi t}}$
10.	$\frac{1}{s+a}$	e^{-at}
11.	$\frac{1}{(s+a)^{n+1}}$	$\frac{t^n e^{-at}}{n!}$
12.	$\frac{k}{s^2+k^2}$	$\sin kt$
13.	$\frac{s}{s^2+k^2}$	$\cos kt$
14.	$\frac{k}{s^2-k^2}$	$\sinh kt$
15.	$\frac{s}{s^2-k^2}$	$\cosh kt$
16.	$\frac{2k^3}{(s^2+k^2)^2}$	$\sin kt - kt\cos kt$
17.	$\frac{2ks}{(s^2+k^2)^2}$	$t\sin kt$
18.	$\frac{1}{1-e^{-ps}} \int_0^p e^{-st} f(t)\, dt$	$f(t+p) = f(t),$ for all t
19.	e^{-as}	$\delta(t-a)$

Index

absolute error, 155
absolutely stable method for ODE, 232
absolutely stable multistep method, 241
Adams–Bashforth multistep method, 241
Adams–Bashforth–Moulton method
  four-step, 244
  three-step, 243
Adams–Moulton multistep method, 241
Aitken's process, 175
amplitude, 40
analytic function, 136

backward differentiation formula, 253
BDF (backward differentiation formula, 253
Bernoulli equation, 22
bisection method, 159
Bonnet recurrence formula, 273
Butcher tableau, 223

centered formula for $f''(x)$, 199
centred formula for $f'(x)$, 198
clamped boundary, 195
clamped spline, 195
classic Runge–Kutta method, 224
composite integration rule
  midpoint, 207
  Simpson's, 210
  trapezoidal, 208
consistent method for ODE, 230
convergence criterion of Cauchy, 135
convergence criterion of d'Alembert, 135
convergence of series
  uniform, 133
convergence of series s
  absolute, 133
convergent method for ODE, 230
corrector, 242
cubic spline, 195

divided difference
  kth, 188
  first, 186
divided difference table, 188
Dormand–Prince pair
  seven-stage, 236
DP(5,4)7M, 236

error, 155
Euler's method, 218
exact solution of ODE, 217
existence of analytic solution, 137
explicit multistep method, 240, 250
extreme value theorem, 158

first forward difference, 189
first-order initial value problem, 217
fixed point, 162
  attractive, 162
  indifferent, 162
  repulsive, 162
floating point number, 155
forward difference
  kth, 189
  second, 189
Fourier-Legendre series, 145
free boundary, 195
function of order p, 217

Gaussian Quadrature, 148, 215
  three-point, 148, 215
  two-point, 148, 215
generating function
  for $P_n(x)$, 144
global Newton-bisection method, 173

Hermite interpolating polynomial, 192
Heun's method
  of order 2, 223
Horner's method, 177

implicit multistep method, 251
improved Euler's method, 221
intermediate value theorem, 158
interpolating polynomial
  Gregory–Newton
    backward-difference, 192
    forward-difference , 190
  Müller's method, 179
  Newton divided difference, 185

parabola method, 179
interval of absolute stability, 231

Lagrange basis, 183
Lagrange interpolating polynomial, 183
Legendre equation, 139
Legendre polynomial $P_n(x)$, 304
Lipschitz condition, 217
local approximation, 242
local error of method for ODE, 230
local extrapolation, 236
local truncation error, 218, 219, 230

MATLAB
`fzero` function, 175
ode113, 252
ode15s, 260
ode23, 234
ode23s, 260
ode23t, 260
ode23tb, 260
matrix
companion, 50
mean value theorem, 158
for integral, 158
for sum, 158
method of false position, 172
method of order p, 231
midpoint rule, 204
multistep method, 240

natural boundary, 195
natural spline, 195
NDF (numerical differentiation formula, 254
Newton's method, 167
modified, 170
Newton–Raphson method, 167
numerical differentiation formula, 254
numerical solution of ODE, 217

order of an iterative method, 167
orthogonality relation
for $P_n(x)$, 142
of $L_n(x)$, 122

PECE mode, 242
PECLE mode, 242
period, 40
phenomenon of stiffness, 253
polynial
Legendre $P_n(x)$, 145
polynomial
Legendre $P_n(x)$, 141
predictor, 242

radius of convergence of a series s, 134
rate of convergence, 167
Ratio Test, 135
rational function, 132
region of absolute stability, 231
regula falsi, 172
relaltive error, 155
Richardson's extrapolation, 202
RKF(4,5), 238
RKV(5,6), 239
Rodrigues' formula
for $P_n(x)$, 142
Root Test, 135
roundoff error, 155, 199, 220
Runge–Kutta method
four-stage, 224
fourth-order, 224
second-order, 223
third-order, 223
Runge–Kutta–Fehlberg pair
six-stage, 238
Runge–Kutta–Verner pair
eight-stage, 239

secant method, 171
signum function sign, 157
Simpson's rule, 205
stability function, 232
stiff system, 252
in an interval, 253
stiffness ratio, 253
stopping criterion, 166

three-point formula for $f'(x)$, 198
trapezoidal rule, 204
truncation error, 155
truncation error of a method, 220
two-point formula for $f'(x)$, 197

well-posed problem, 217

zero-stable method for ODE, 230

Integration

$$\int uv'\,dx = uv - \int u'v\,dx$$

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + c \qquad (n \neq -1)$$

$$\int \frac{1}{x}\,dx = \ln|x| + c$$

$$\int e^{ax}\,dx = \frac{1}{a}e^{ax} + c$$

$$\int \sin x\,dx = -\cos x + c$$

$$\int \cos x\,dx = \sin x + c$$

$$\int \tan x\,dx = -\ln|\cos x| + c$$

$$\int \cot x\,dx = \ln|\sin x| + c$$

$$\int \sec x\,dx = \ln|\sec x + \tan x| + c$$

$$\int \csc x\,dx = \ln|\csc x - \cot x| + c$$

$$\int \frac{dx}{x^2+a^2} = \frac{1}{a}\arctan\frac{x}{a} + c$$

$$\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + c$$

$$\int \frac{dx}{\sqrt{x^2+a^2}} = \sinh^{-1}\frac{x}{a} + c$$

$$\int \frac{dx}{\sqrt{x^2-a^2}} = \cosh^{-1}\frac{x}{a} + c$$

$$\int \sin^2 x\,dx = \tfrac{1}{2}x - \tfrac{1}{4}\sin 2x + c$$

$$\int \cos^2 x\,dx = \tfrac{1}{2}x + \tfrac{1}{4}\sin 2x + c$$

$$\int \tan^2 x\,dx = \tan x - x + c$$

$$\int \cot^2 x\,dx = -\cot x - x + c$$

$$\int \ln x\,dx = x\ln x - x + c$$

$$\int e^{ax}\sin bx\,dx = \frac{e^{ax}}{a^2+b^2}(a\sin bx - b\cos bx) + c$$

$$\int e^{ax}\cos bx\,dx = \frac{e^{ax}}{a^2+b^2}(a\cos bx + b\sin bx) + c$$

Laplace Transform: General Formulas

Formula	Name, Comments
$F(s) = \mathcal{L}\{f(t)\} = \int_0^\infty e^{-st}f(t)\,dt$ $f(t) = \mathcal{L}^{-1}\{F(s)\}$	Definition of Transform Inverse Transform
$\mathcal{L}\{af(t) + bg(t)\} = a\mathcal{L}\{f(t)\} + b\mathcal{L}\{g(t)\}$	Linearity
$\mathcal{L}\{e^{at}f(t)\} = F(s-a)$ $\mathcal{L}^{-1}\{F(s-a)\} = e^{at}f(t)$	s-Shifting (First Shifting Theorem)
$\mathcal{L}(f') = s\mathcal{L}(f) - f(0)$ $\mathcal{L}(f'') = s^2\mathcal{L}(f) - sf(0) - f'(0)$ $\mathcal{L}(f^{(n)}) = s^n\mathcal{L}(f) - s^{n-1}f(0) - \cdots - f^{(n-1)}(0)$ $\mathcal{L}\left\{\int_0^t f(\tau)\,d\tau\right\} = \frac{1}{s}\mathcal{L}(f)$	Differentiation of Function Integration of Function
$\mathcal{L}\{f(t-a)u(t-a)\} = e^{-as}F(s)$ $\mathcal{L}^{-1}\{e^{-as}F(s)\} = f(t-a)u(t-a)$	t-Shifting (Second Shifting Theorem)
$\mathcal{L}\{tf(t)\} = -F'(s)$ $\mathcal{L}\left\{\frac{f(t)}{t}\right\} = \int_s^\infty F(\tilde{s})\,d\tilde{s}$	Differentiation of Transform Integration of Transform
$(f*g)(t) = \int_0^t f(\tau)g(t-\tau)\,d\tau = \int_0^t f(t-\tau)g(\tau)\,d\tau$ $\mathcal{L}(f*g) = \mathcal{L}(f)\mathcal{L}(g)$	Convolution
$\mathcal{L}(f) = \frac{1}{1-e^{-ps}}\int_0^p e^{-st}f(t)\,dt$	f Periodic with Period p

$$\sin x \sin y = \tfrac{1}{2}[-\cos(x+y) + \cos(x-y)]$$

$$\cos x \cos y = \tfrac{1}{2}[\cos(x+y) + \cos(x-y)]$$

$$\sin x \cos y = \tfrac{1}{2}[\sin(x+y) + \sin(x-y)]$$

$F(s) = \mathcal{L}\{f(t)\}$	$f(t)$
$1/s$	1
$1/s^2$	t
$1/s^n \quad (n = 1, 2, \cdots)$	$t^{n-1}/(n-1)!$
$1/\sqrt{s}$	$1/\sqrt{\pi t}$
$1/s^{3/2}$	$2\sqrt{t/\pi}$
$1/s^a \quad (a > 0)$	$t^{a-1}/\Gamma(a)$
$\frac{1}{s-a}$	e^{at}
$\frac{1}{(s-a)^2}$	te^{at}
$\frac{1}{(s-a)^n} \quad (n = 1, 2, \cdots)$	$\frac{1}{(n-1)!}t^{n-1}e^{at}$
$\frac{1}{(s-a)^k} \quad (k > 0)$	$\frac{1}{\Gamma(k)}t^{k-1}e^{at}$
$\frac{1}{(s-a)(s-b)} \quad (a \neq b)$	$\frac{1}{(a-b)}(e^{at} - e^{bt})$
$\frac{s}{(s-a)(s-b)} \quad (a \neq b)$	$\frac{1}{(a-b)}(ae^{at} - be^{bt})$
$\frac{1}{s^2+\omega^2}$	$\frac{1}{\omega}\sin \omega t$
$\frac{s}{s^2+\omega^2}$	$\cos \omega t$
$\frac{1}{s^2-a^2}$	$\frac{1}{a}\sinh at$
$\frac{s}{s^2-a^2}$	$\cosh at$
$\frac{1}{(s-a)^2+\omega^2}$	$\frac{1}{\omega}e^{at}\sin \omega t$
$\frac{s-a}{(s-a)^2+\omega^2}$	$e^{at}\cos \omega t$
$\frac{1}{s(s^2+\omega^2)}$	$\frac{1}{\omega^2}(1 - \cos \omega t)$
$\frac{1}{s^2(s^2+\omega^2)}$	$\frac{1}{\omega^3}(\omega t - \sin \omega t)$
$\frac{1}{(s^2+\omega^2)^2}$	$\frac{1}{2\omega^3}(\sin \omega t - \omega t \cos \omega t)$
$\frac{s}{(s^2+\omega^2)^2}$	$\frac{t}{2\omega}\sin \omega t$
$\frac{s^2}{(s^2+\omega^2)^2}$	$\frac{1}{2\omega}(\sin \omega t + \omega t \cos \omega t)$
$\frac{s}{(s^2+a^2)(s^2+b^2)} \quad (a^2 \neq b^2)$	$\frac{1}{b^2-a^2}(\cos at - \cos bt)$
$\frac{1}{s^4+4k^4}$	$\frac{1}{4k^3}(\sin kt \cos kt - \cos kt \sinh kt)$
$\frac{s}{s^4+4k^4}$	$\frac{1}{2k^2}\sin kt \sinh kt$
$\frac{1}{s^4-k^4}$	$\frac{1}{2k^3}(\sinh kt - \sin kt)$
$\frac{s}{s^4-k^4}$	$\frac{1}{2k^2}(\cosh kt - \cos kt)$
$\sqrt{s-a} - \sqrt{s-b}$	$\frac{1}{2\sqrt{\pi t^3}}(e^{bt} - e^{at})$
$\frac{1}{\sqrt{s+a}\sqrt{s+b}}$	$e^{-(a+b)t/2}I_0\left(\frac{a-b}{2}t\right)$
$\frac{1}{\sqrt{s^2+a^2}}$	$J_0(at)$
$\frac{s}{(s-a)^{3/2}}$	$\frac{1}{\sqrt{\pi t}}e^{at}(1 + 2at)$
$\frac{1}{(s^2-a^2)^k} \quad (k > 0)$	$\frac{\sqrt{\pi}}{\Gamma(k)}\left(\frac{t}{2a}\right)^{k-1/2}I_{k-1/2}(at)$
e^{-as}/s	$u(t-a)$
e^{-as}	$\delta(t-a)$
$\frac{1}{s}e^{-k/s}$	$J_0(2\sqrt{kt})$
$\frac{1}{\sqrt{s}}e^{-k/s}$	$\frac{1}{\sqrt{\pi t}}\cos 2\sqrt{kt}$
$\frac{1}{s^{3/2}}e^{k/s}$	$\frac{1}{\sqrt{\pi k}}\sinh 2\sqrt{kt}$
$e^{-k\sqrt{s}} \quad (k > 0)$	$\frac{k}{2\sqrt{\pi t^3}}e^{-k^2/4t}$
$\frac{1}{s}\ln s$	$-\ln t - \gamma \quad (\gamma \approx 0.5772)$
$\ln\frac{s-a}{s-b}$	$\frac{1}{t}(e^{bt} - e^{at})$

INTEGRATING FACTORS

$$\frac{1}{N}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = f(x) \quad \rightarrow \quad \exp\left(\int f(x)\,dx\right)$$

$$\frac{1}{M}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = g(y) \quad \rightarrow \quad \exp\left(-\int g(y)\,dy\right)$$

FOR $M(x, y)\,dx + N(x, y)\,dy = 0.$