Table des matières

Chapitre 1. Equations différentielles du premier ordre 1
1.1. Concepts fondamentaux 1
1.2. Equations séparables 3
1.3. Equations à coefficients homogènes 5
1.4. Equations exactes 7
1.5. Facteurs d'intégration 13
1.6. Equations linéaires 18
1.7. Familles de courbes orthogonales 19
1.8. Champ des tangentes et solutions approchées 22
1.9. Existence et unicité de la solution 22

Chapitre 2. Equations différentielles linéaires du deuxième ordre 27
2.1. Équations linéaires homogènes 27
2.2. Equations homogènes à coefficients constants 27
2.3. Base de l'espace solution 28
2.4. Solutions indépendantes 30
2.5. Modélisation en mécanique 32
2.6. Equation d'Euler–Cauchy 36

Chapitre 3. Équations différentielles linéaires d'ordre quelconque 41
3.1. Équations homogènes 41
3.2. Equations linéaires homogènes à coefficients constants 47
3.3. Equations linéaires nonhomogènes 51
3.4. Méthode des coefficients indéterminés 52
3.5. Solution particulière par variation des paramètres 56
3.6. Systèmes asservis 61

Chapitre 4. Systèmes d'équations différentielles linéaires 65
4.1. Introduction 65
4.2. Théorème d'existence et d'unicité 67
4.3. Système fondamental 67
4.4. Systèmes linéaires à coefficients constants 70
4.5. Systèmes linéaires nonhomogènes 77

Chapitre 5. Résolution numérique d'équations différentielles 81
5.1. Le problème à valeur initiale 81
5.2. Méthodes explicites à un pas 82
5.3. Méthodes prédicteur-correcteur multipas 88
5.4. Systèmes différentiels raides 98

Chapitre 6. Solutions séries 105

6.1. La méthode 105
6.2. Fondements de la méthode des séries de puissances 106
6.3. Equation et polynômes de Legendre 113
6.4. Orthogonalité des polynômes de Legendre 115
6.5. Série de Fourier–Legendre 118
6.6. Une application: la quadrature gaussienne 120
6.7. Résolution numérique d'équations intégrales de 2e espèce 124

Chapitre 7. Calcul matriciel 129
7.1. Solution LU de $Ax = b$ 129
7.2. La décomposition de Cholesky 136
7.3. Le méthode itérative de Gauss–Seidel 140
7.4. Normes de matrices 142
7.5. Systèmes surdéterminés 145
7.6. Valeurs propres de matrices 147

Chapitre 8. Transformation de Laplace 155
8.1. Définition 155
8.2. Transformées de dérivées et d'intégrales 159
8.3. Déplacements en s et en t 163
8.4. La fonction delta de Dirac 171
8.5. Dérivée et intégrale de la transformée 172
8.6. Équation différentielle de Laguerre 175
8.7. Convolution 177
8.8. Fractions simples 180
8.9. Transformées de fonctions périodiques 180

Chapitre 9. Introduction aux méthodes numériques 183
9.1. Calculs 183
9.2. Résolution des équations nonlinéaires par récurrence 186
9.3. Interpolation et extrapolation 201
9.4. Intégration numérique 208
9.5. Problèmes aux valeurs initiales 214

Chapitre 10. Examens partiels et final 225

Chapitre 11. Formulaire et tables 235
11.1. Facteur d'intégration de $M(x, y)\,dx + N(x, y)\,dy = 0$ 235
11.2. Les polynômes de Legendre $P_n(x)$ sur $[-1, 1]$ 235
11.3. Les polynômes de Laguerre sur $0 \leq x < \infty$ 236
11.4. Développements de Fourier–Legendre 237
11.5. Table d'intégrales 238
11.6. La transformée de Laplace 238

Exercices 241
Exercices pour le chapitre premier 241
Exercices pour le chapitre 2 243
Exercices pour le chapitre 3 244
Exercices pour le chapitre 4 246
Exercices pour le chapitre 5 247

Exercices pour le chapitre 6 248
Exercices pour le chapitre 7 250
Exercices pour le chapitre 8 253
Exercices pour le chapitre 9 255

CHAPITRE 1

Equations différentielles du premier ordre

1.1. Concepts fondamentaux

(a) Une *équation différentielle* contient une ou plusieurs dérivées et un signe "=".

Voici trois équations différentielles où $' := \dfrac{d}{dx}$:

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

Voici une *équation aux dérivées partielles* du 2ème ordre:

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

(b) L'ordre d'une équation différentielle est égal à l'ordre de la dérivée d'ordre le plus élevé.

Les équations (1), (2) et (3) ci-haut sont respectivement d'ordre 1, 2 et 3.

(c) Une *solution explicite* d'une équation différentielle en x sur $]a, b[$ est une fonction d'une variable, $y = g(x)$, décrivant une courbe, telle que l'équation différentielle devient une identité en x sur $]a, b[$ quand on substitue $g(x)$, $g'(x)$, etc. à y, y', etc. dans l'équation différentielle.

On voit que la fonction

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

est une solution explicite de l'équation différentielle

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

En effet, on a

$$\text{1er M} := y' = g'(x) = 2\, e^{2x},$$
$$\text{2ème M} := 2y = 2g(x) = 2\, e^{2x}.$$

Donc,

$$\text{1er M} = \text{2ème M}, \qquad \text{pour tout } x.$$

On a donc une identité en x sur $]-\infty, \infty[$. □

(d) Une *solution implicite* est une courbe définie par une équation de la forme $G(x, y) = 0$.

On remarque qu'une solution implicite contient toujours le signe "=".

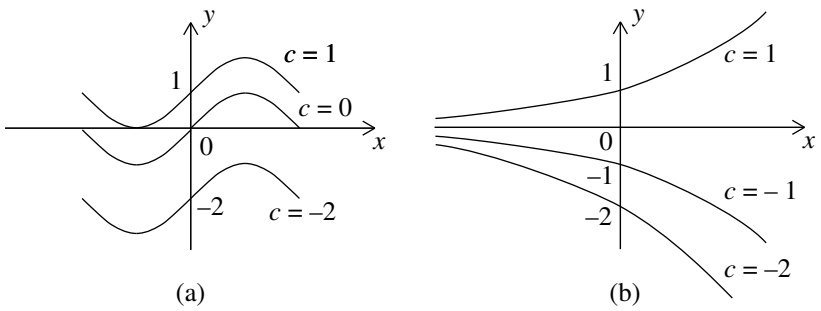

FIGURE 1.1. (a) Deux familles de courbes: (a) $y = \sin x + c$; (b) $y = c\exp(x)$.

On voit que la fonction

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

est une solution implicite de l'équation différentielle

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

En effet, en considérant y comme fonction de x et en dérivant l'équation de la courbe par rapport à x,

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

on a

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

(e) La *solution générale* d'une équation différentielle d'ordre n contient n constantes arbitraires.

La famille de fonctions

$$y = \sin x + c$$

est la solution générale de l'équation différentielle

$$y' = \cos x$$

du 1er ordre. Pour $c = 1$ fixé, on obtient la solution unique,

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

qui passe par le point $(0, 1)$ de $\mathbb{R}^2$. Etant donné un point quelconque (x_0, y_0) du plan il y a une et une seule courbe de la famille donnée qui passe par ce point (V. figure 1.1(a)).

On voit de la même façon que la famille de fonctions

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

est la solution générale de l'équation différentielle

$$y' = y.$$

En fixant $c = -1$, on obtient la solution unique,

$$y = -e^x,$$

qui passe par le point $(0,-1)$ de $\mathbb{R}^2$. Etant donné un point quelconque (x_0,y_0) du plan il y a une et une seule courbe de la famille donnée qui passe par ce point (V. figure 1.1(b)).

1.2. Equations séparables

Considérons une équation différentielle séparable de la forme

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

On sépare l'équation récrite sous forme de différentiels en mettant au 1er membre tous les termes en y et au second membre tous les termes en x:

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

La solution d'une équation séparée s'obtient au moyen d'une intégrale indéfinie (primitive ou antidérivée) de chacun des membres à laquelle on ajoute une constante:

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

c'est-à-dire

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

Ces deux formes implicites de la solution définissent y comme fonction de x ou x comme fonction de y.

Supposons que $y = y(x)$ soit fonction de x et vérifions que (1.3) est bien une solution de (1.1):

$$\frac{d}{dx}(\text{1er M}) = \frac{d}{dx}G\left(y(x)\right) = G'\left(y(x)\right)y'(x) = g(y)y',$$
$$\frac{d}{dx}(\text{2ème M}) = \frac{d}{dx}[F(x)+c] = F'(x) = f(x). \quad \square$$

EXEMPLE 1.1. Résoudre $y' = 1 + y^2$.

RÉSOLUTION. Puisque l'équation différentielle est séparable, on a

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

Alors

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

est la solution générale, puisqu'elle contient une constante arbitaire. $\square$

EXEMPLE 1.2. Résoudre $y' = -2xy$ avec la condition initiale $y(0) = y_0$.

RÉSOLUTION. Puisque l'équation différentielle est séparable, la solution générale est

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

En prenant l'exponentielle de la solution, on a

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

qu'on récrit sous la forme

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

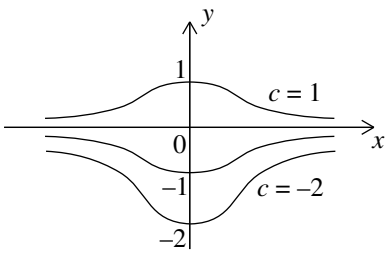

FIGURE 1.2. Trois fonctions cloches.

On remarquera que la constante additive c_1 est devenue une constante multiplicative après l'exponentiation. Dans la figure 1.2 on voit trois fonctions cloches membres de la famille de la solution générale.

Enfin, la solution qui satisfait la condition initiale, est

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

Cette solution est unique. □

EXEMPLE 1.3. D'après la loi du refroidissement de Newton, le taux de changement de la température $T(t)$ d'un corps dans un milieu environnant de température T_0 est proportionnel à la différence $T(t) - T_0$ des températures:

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

On plonge une boule de cuivre dans un grand bassin de liquide dont la température est maintenue à 30 degrés. Si la température initiale de la boule est de 100 degrés et si sa température après 3 min est de 70 degrés, quand sera-t-elle de 31 degrés?

RÉSOLUTION. On a l'équation différentielle séparable:

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

Alors

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

À $t = 0$,

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

À $t = 3$,

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

Si $T(t) = 31$, alors

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

En prenant le logarithme des deux membres, on obtient

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

d'où il vient que

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

1.3. Equations à coefficients homogènes

DÉFINITION 1.1. Une fonction $M(x, y)$ est homogène de degré s simultanément en x et en y si

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

Les équations différentielles à coefficients homogènes du même degré sont séparables de la façon suivante.

THÉORÈME 1.1. *Soit une équation différentielle à coefficients homogènes de degré s,*

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

Alors chacune des substitutions $y = xu$ et $x = yu$ rend l'équation différentielle séparable.

DÉMONSTRATION. Posons

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

dans (1.5). Alors,

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

On sépare cette dernière equation:

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

qui admet la solution générale

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

EXEMPLE 1.4. Résoudre $2xyy' - y^2 + x^2 = 0$.

RÉSOLUTION. On récrit l'équation sous forme différentielle:

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

Puisque les coefficients sont des fonctions homogènes de degré 2 en x et y, posons

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

Alors la dernière équation devient

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

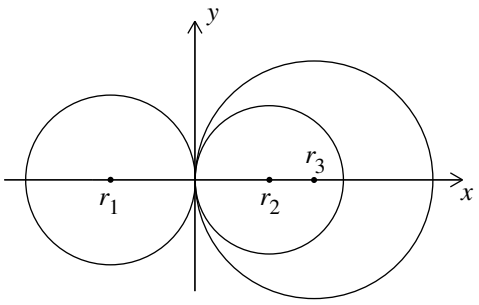

FIGURE 1.3. Familles de cercles de centre $(r, 0)$.

Puisque le 1er membre de cette équation semble difficile à intégrer, recommençons avec la substitution

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

Alors,

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

Cette dernière équation s'intègre facilement:

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

Alors la solution générale est

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

Si l'on pose $c = 2r$ dans la formule précédente, on obtient

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

La solution générale décrit donc une famille de cercle de centre $(r, 0)$ et de rayon $|r|$ (V. figure 1.3). □

EXEMPLE 1.5. Résoudre l'équation différentielle

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

RÉSOLUTION. Récrivons cette équation sous forme différentielle:

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

C'est une équation aux coefficients homogènes de degré zéro où la fonction g au second membre est homogène de degré zéro en x et y. Si l'on pose

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

la dernière équation devient

$$g(u)\,dx - x\,du - u\,dx = 0,$$
$$x\,du = [g(u) - u]\,dx,$$
$$\frac{du}{g(u)-u} = \frac{dx}{x},$$

qui est une équation séparée. On peut donc l'intégrer directement:

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

et substituer $u = y/x$ dans la solution après intégration. □

1.4. Equations exactes

DÉFINITION 1.2. L'équation différentielle du 1er ordre

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

est exacte si le 1er membre est une différentielle totale ou exacte d'une fonction $u(x,y)$:

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

Si (1.6) est exacte, alors

$$du = 0$$

et la solution générale de (1.6) est

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

Si l'on compare les expressions (1.6) et (1.7), on voit que

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

Le théorème suivant donne une condition nécessaire et suffisante pour que l'équation (1.6) soit exacte.

THÉORÈME 1.2. *Soit $M(x,y)$ et $N(x,y)$ deux fonctions continûment dérivables sur un ensemble $\Omega \in \mathbb{R}^2$ connexe et simplement connexe (c'est-à-dire d'un seul morceau et sans trou). Alors l'équation différentielle*

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

est exacte si et seulement si

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}, \qquad \text{sur } \Omega. \tag{1.11}$$

DÉMONSTRATION. **Nécessité:** Supposons que (1.10) est exacte; alors

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

Par conséquent,

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

où l'échange de l'ordre des dérivées en x et y est justifiée par la continuité des deux termes extrêmes.

Suffisance: On suppose (1.11) et l'on construit une fonction $F(x, y)$ telle que

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

Soit $\varphi(x, y) \in C^2(\Omega)$ une fonction telle que

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

On peut prendre, par exemple,

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

Alors,

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

Puisque

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

par la continuité des deux membres, il suit que

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

Si l'on intègre en x, on obtient

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

Prenons

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

Alors

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

On illustre par des exemples une **méthode pratique** pour résoudre des équations exactes.

Exemple 1.6. Trouver la solution générale de l'équation différentielle

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

et la solution qui satifait la condition initiale $y(1) = -1$. Tracer la solution sur $1 \leq x \leq 4$.

RÉSOLUTION. **(a) Résolution analytique par la méthode pratique.**— On vérifie que l'équation est exacte:

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

Il suit que l'équation est exacte. On peut donc l'intégrer. De

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

on a

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

De

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

on a

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

Donc

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

Il est essentiel que $T'(y)$ soit fonction de y seulement, sinon il y a erreur quelque part: ou l'équation n'est pas exacte ou on a une erreur de calcul.

On intègre $T'(y)$:

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

Il n'est pas nécessaire d'ajouter une constante d'intégration ici parce qu'on aura une constante dans $u = c$. On a donc la **surface**

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

Puisque $du = 0$, alors $u(x,y) = c$ et la solution générale implicite, qui contient *une constante arbitraire et un signe* "=", c'est-à-dire une **courbe**, est donc

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

On détermine la constante c au moyen de la condition initiale $y(1) = -1$. On pose donc $x = 1$ et $y = -1$ dans la solution générale et l'on obtient

$$c = -3.$$

Alors, la solution implicite qui satisfait la condition initiale est

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

(b) Résolution par Matlab symbolique.— La solution générale est :

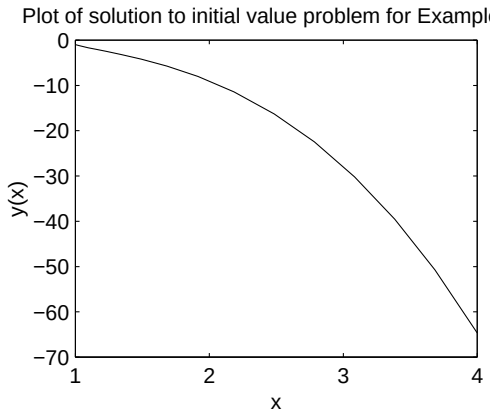

FIGURE 1.4. Graphe de la solution pour l'exemple 1.6.

```
>> 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)]
```

La solution du problème à valeur initiale est la branche inférieure avec $C1 = -3$, comme on peut voir en insérant la condition initiale `'y(1)=-1'`, dans la commande précédente:

```
>> 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 du problème à valeur initiale par Matlab numérique.— On emploie la condition initiale $y(1) = -1$. Le fichier M `exp1_6.m` est

```
function yprime = exp1_6(x,y); %MAT 2731, Exp 1.6.
yprime = -3*x*(x*y-2)/(x^3+2*y);
```

L'appel du solveur `ode23` et de la commande `plot` sont:

```
>> xspan = [1 4]; % solution pour 1<=x<=4
>> y0 = -1; % condition initiale
>> [x,y] = ode23('exp1_6',xspan,y0);%xspan en Matlab 5
>> subplot(2,2,1); plot(x,y);
>> title('Graphe de la solution pour l'exemple 1.6');
>> xlabel('x'); ylabel('y(x)');
>> print Fig.exp1.6
```

□

EXEMPLE 1.7. Trouver la solution générale de l'équation différentielle

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

et la solution qui satifait la condition initiale $y(1) = -1$. Tracer la solution sur $1 \leq x \leq 4$.

RÉSOLUTION. **(a) Résolution analytique par la méthode pratique.**— Remarquer que $N(x,y) = -(x^2y + 2x)$ du fait que le 1er membre de l'équation est $M\,dx + N\,dy$. On vérifie que l'équation est exacte:

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

Il suit que l'équation est exacte. On peut donc l'intégrer. De

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

on a

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

De

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

on a

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

Donc

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

Il est essentiel que $T'(x)$ soit fonction de x seulement, sinon il y a erreur quelque part: ou l'équation n'est pas exacte ou on a une erreur de calcul.

On intègre $T'(x)$:

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

Il n'est pas nécessaire d'ajouter une constante d'intégration ici parce qu'on aura une constante dans $u = c$. On a donc la **surface**

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

Puisque $du = 0$, alors $u(x,y) = c$ et la solution générale implicite, qui contient *une constante arbitraire et un signe* "=", c'est-à-dire une **courbe**, est donc

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

Puisque $x = 1$ et $y = -1$, on obtient

$$c = 10.$$

Alors, la solution implicite qui satisfait la condition initiale est

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

(b) Résolution par Matlab symbolique.— La solution générale est :

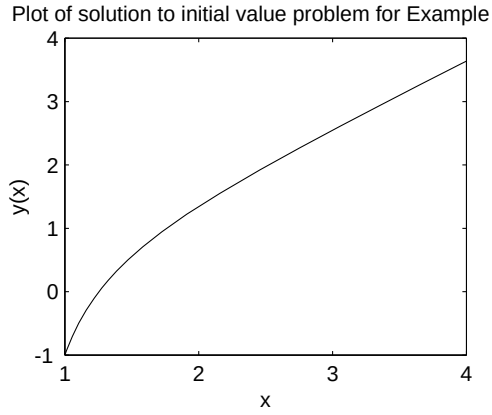

FIGURE 1.5. Graphe de la solution pour l'exemple 1.7.

```
\begin{verbatim}
>> 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]
```

La solution du problème à valeur initiale est la branche inférieure avec $C1 = -5$:

```
>> 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 du problème à valeur initiale par Matlab numérique.— On emploie la condition initiale $y(1) = -1$. Le fichier M `exp1_7.m` est

```
function yprime = exp1_7(x,y); %MAT 2731, Exp 1.7.
yprime = (2*x^3-x*y^2-2*y+3)/(x^2*y+2*x);
```

L'appel du solveur `ode23` et de la commande `plot` sont:

```
>> xspan = [1 4]; % solution pour 1<=x<=4
>> y0 = -1; % condition initiale
>> [x,y] = ode23('exp1_7',xspan,y0);%xspan en Matlab 5
>> subplot(2,2,1); plot(x,y);
>> title('Graphe de la solution pour l'exemple 1.7');
>> xlabel('x'); ylabel('y(x)');
>> print Fig.exp1.7
```

□

L'exemple suivant montre que la méthode flanche si l'on tente d'intégrer une équation qui n'est pas exacte.

EXEMPLE 1.8. Résoudre

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

RÉSOLUTION. On récrit l'équation sous forme standard:

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

L'équation n'est pas exacte parce que

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

Essayons de résoudre l'équation par la méthode proposée:

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

Donc

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

Ceci est impossible parce que T(y) doit être fonction de y seulement. □

EXEMPLE 1.9. Soit l'équation différentielle

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

Choisir a, b, k, l pour que l'équation soit exacte.

RÉSOLUTION.

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

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

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

Donc

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

La solution générale est

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

Dans la suite, on notera les dérivées partielles:

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

1.5. Facteurs d'intégration

Si l'équation différentielle

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

n'est pas exacte, on peut la rendre exacte en la multipliant par un facteur d'intégration $\mu(x, y)$:

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

On récrit cette équation sous la forme

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

Alors

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

L'équation (1.13) sera exacte si

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

En général, il est difficile de résoudre l'équation aux dérivées partielles (1.14).

On considère les deux cas particuliers où μ est une fonction d'une seule variable: $\mu = \mu(x)$ ou $\mu = \mu(y)$.

Cas 1. Si $\mu = \mu(x)$ est une fonction de x seulement, alors $\mu_x = \mu'(x)$ et $\mu_y = 0$. Donc (1.14) devient une équation différentielle:

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

Si le 1er membre de l'expression suivante

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

est une fonction de x seulement, alors (1.15) est séparable:

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

En intégrant cette équation séparée, on obtient le facteur d'intégration

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

Cas 2. De même, si $\mu = \mu(y)$ est une fonction de y seulement, alors $\mu_x = 0$ et $\mu_y = \mu'(y)$. Donc (1.14) devient une équation différentielle:

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

Si le 1er membre de l'expression suivante

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

est une fonction de y seulement, alors (1.18) est séparable:

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

En intégrant cette équation séparée, on obtient le facteur d'intégration

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

Remarquer la présence du signe moins dans (1.20) et son absence dans (1.17).

EXEMPLE 1.10. Résoudre

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

RÉSOLUTION. **(a) Résolution analytique.**— Puisque

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

est une fonction de x seulement, on a le facteur d'intégration

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

Multipliant l'équation différentielle par x^2, on obtient l'équation exacte

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

On résout cette équation par la méthode pratique:

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

Alors

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

Pas de constante d'intégration ici; elle apparaîtra plus loin. On a donc

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

et la solution générale est

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

(b) Résolution par Matlab symbolique.— Matlab ne trouve pas la solution générale de l'équation inexacte:

```
>> 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 ]
```

mais résout l'équation exacte:

```
>> 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)]
```

□

EXEMPLE 1.11. Résoudre

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

RÉSOLUTION. **(a) Résolution analytique.**— Puisque

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

on essaie

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

qui est fonction de y seulement. On a le facteur d'intégration

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

Multipliant l'équation différentielle par y, on obtient l'équation exacte

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

On résout cette équation par la méthode pratique:

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

Alors

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

On a donc

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

et la solution générale est

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

(b) Résolution par Matlab symbolique.— La commande `dsolve` de Matlab symbolique produit une solution générale très complexe pour l'équation inexacte et pour l'équation exacte. Ces solutions ne se simplifient pas au moyen des commandes `simplify` et `simple`.

On reprend donc la méthode pratique et demande à Matlab symbolique de faire les simples manipulations algébriques et analytiques.

```
>> 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') % equation exacte ou non
 test = -x-y-1 % equation non exacte
>> 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 % une fonction de y seulement
>> mu = exp(-int(g,'y')) % facteur d'int\'egration
 mu = y
>> syms MM NN
>> MM = mu*M; NN = mu*N; % multiplication par le facteur d'integration
>> u = int(MM,'x') % solution u; T(y) arbitraire pas inclus
 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) % solution generale u = c.
```

La solution générale est

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

REMARQUE 1.1. Une équation séparée,

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

est exacte. En effet, $M_y = 0$ et $N_x = 0$, ce qui donne le facteur d'intégration

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

Si l'on résout cette équation par la méthode pratique des équations exactes, on a:

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

On obtient donc la même solution que celle obtenue par la méthode (1.3).

REMARQUE 1.2. Le facteur qui transforme une équation séparable en une équation séparée est un facteur d'intégration puisque cette dernière est exacte.

EXEMPLE 1.12. Soit l'équation séparable

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

Montrer que le facteur $\left(1 + y^2\right)^{-1}$ qui sépare l'équation est un facteur d'intégration.

RÉSOLUTION. On a

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

Donc

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

Dans le prochain exemple, on trouve facilement un facteur d'intégration $\mu(x,y)$ qui est une fonction de x et de y.

EXEMPLE 1.13. Soit l'équation séparable

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

Montrer que le facteur

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

qui sépare l'équation, est un facteur d'intégration.

RÉSOLUTION. L'équation différentielle

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

est séparée; donc elle est exacte. $\square$

1.6. Equations linéaires

Considérons l'équation nonhomogène du 1er ordre de la forme

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

Le 1er membre de cette équation est une expression linéaire en y et y'. On dira donc que (1.21) est une équation différentielle *linéaire*. On récrit cette équation sous forme différentielle:

$$M\,dx + N\,dy := f(x)y\,dx + dy = r(x)\,dx. \tag{1.22}$$

Puisque $M_y = f(x)$ et $N_x = 0$, on voit que le 1er membre de cette équation n'est pas une différentielle exacte. Comme

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

est une fonction de x seulement, par (1.17) on a le facteur d'intégration

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

On multiplie l'équation (1.21) par $\mu(x)$ pour rendre le 1er membre exacte:

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

Alors,

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

La solution générale de (1.21) est donc

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

À l'exemple 3.8, on présente une seconde méthode pour résoudre une équation linéaire nonhomogène du 1er ordre

Exemple 1.14. Résoudre l'équation différentielle linéaire

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

Résolution. On récrit l'équation sous forme standard:

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

Le facteur d'intégration est

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

Alors,

$$d(x^2y) = \sinh 3x\,dx, \qquad \text{c'est-à-dire} \quad \frac{d}{dx}(x^2y) = \sinh 3x.$$

On a donc

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

ou

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

EXEMPLE 1.15. Résoudre l'équation différentielle linéaire

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

RÉSOLUTION. On récrit cette équation sous la forme d'une équation linéaire en $x(y)$:

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

Le facteur d'intégration qui rend le 1er membre exact, est

$$\mu(y) = e^{\int \left(\frac{3}{y} - 1\right) dy} = e^{\ln y^3 - y} = y^3\, e^{-y}.$$

On a donc

$$\frac{d}{dy}\left(y^3\, e^{-y} x\right) = -2y^2\, e^{-y}, \qquad \text{c'est-à-dire} \quad d\left(y^3\, e^{-y} x\right) = -2y^2\, e^{-y}\, dy.$$

Alors

$$\begin{aligned} y^3\, e^{-y} x &= -2\int y^2\, e^{-y}\, dy + c \\ &= 2y^2\, e^{-y} - 4\int y e^{-y}\, dy + c \\ &= 2y^2\, e^{-y} + 4y\, e^{-y} - 4\int e^{-y}\, dy + c \\ &= 2y^2\, e^{-y} + 4y\, e^{-y} + 4\, e^{-y} + c. \end{aligned}$$

La solution générale est donc

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

1.7. Familles de courbes orthogonales

Une famille de courbes peut être donnée au moyen d'une équation

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

où le paramètre c est explicite, ou d'une équation

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

implicite en c.

Dans le 1er cas, les courbes de la famille donnée satisfont l'équation différentielle

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

où m est la pente de la courbe au point (x, y). Cette équation différentielle ne contient pas le paramètre c.

Dans le 2ème cas, on a

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

Pour éliminer le paramètre c de cette équation différentielle, on résout l'équation $F(x, y, c) = 0$ pour c en fonction de x et y,

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

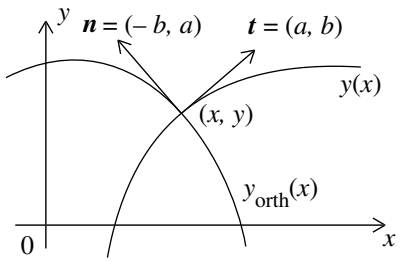

FIGURE 1.6. Deux courbes orthogonales au point (x, y).

et l'on substitue cette fonction dans l'équation différentielle:

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

Soit $\boldsymbol{t} = (a, b)$ la tangente et $\boldsymbol{n} = (-b, a)$ la normale à la courbe donnée $y = y(x)$ au point (x, y) de la courbe. Alors la pente de la tangente est

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

La pente de la courbe $y_{\text{orth}}(x)$ orthogonale à la courbe $y(x)$ en (x, y) est

$$m_1 = -\frac{a}{b} = y'_{\text{orth}}(x) \tag{1.25}$$

(V. figure 1.6). Alors, la famille orthogonale satisfait l'équation différentielle

$$y'_{\text{orth}}(x) = m_1(x).$$

EXEMPLE 1.16. Soit la famille de cercles

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

de centre $(0, c)$ sur l'axe Oy et de rayon $|c|$. Trouver l'équation différentielle de cette famille et celle de la famille orthogonale, résoudre cette dernière équation et tracer quelques courbes des deux familles.

RÉSOLUTION. On obtient l'équation différentielle de la famille donnée en dérivant (1.26) par rapport à x,

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

et en résolvant (1.26) pour c,

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

On substitue cette valeur de c dans l'équation différentielle,

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

L'équation différentielle de la famille orthogonale est alors

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

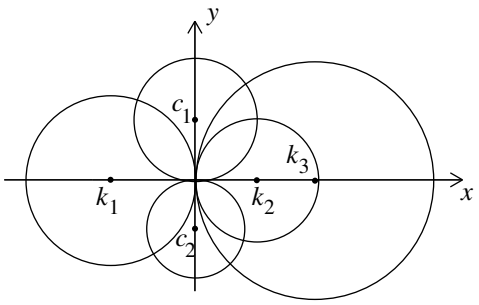

FIGURE 1.7. Quelques courbes des deux familles orthogonales.

On récrit cette équation sous forme différentielle $M\,dx + N\,dy = 0$, tout en supprimant la mention "orth":

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

Puisque $M_y = -2y$ et $N_x = 2y$, cette équation n'est pas exacte, mais

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

est une fonction de x seulement. Donc

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

est un facteur d'intégration. On multiplie l'équation différentielle par $\mu(x)$:

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

et l'on résout par la méthode ordinaire:

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

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

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

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

c'est-à-dire la solution

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

est une famille de cercles. On récrit cette solution d'une façon plus explicite:

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

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

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

La famille orthogonale est donc une famille de cercles de centre $(k, 0)$ sur l'axe Ox et de rayon $|k|$. La figure 1.7 montre quelques courbes des deux familles. □

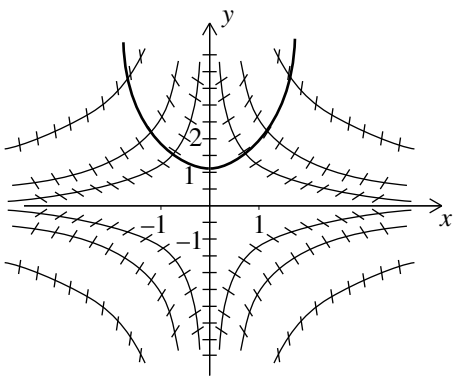

FIGURE 1.8. Champs des tangentes pour l'example 1.17.

1.8. Champ des tangentes et solutions approchées

On peut utiliser les solutions approchées d'une équation différentielle si l'on ne peut trouver la solution exacte ou si la complexité de la formule de cette solution en rend l'utilisation très difficile. Dans ce cas, on a recours soit à une méthode de résolution numérique (V. chapitre 5), ou soit à la méthode du champ des tangentes. La méthode du champ des tangentes nous permet de tracer plusieurs courbes intégrales, sans pour autant résoudre l'équation différentielle.

La méthode du champ des tangentes s'applique aux équations différentielles de la forme générale

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

Il suffit de tracer une courbe intégrale de pente y'. La pente de la courbe qui passe par le point (x_0, y_0) est égale à $f(x_0, y_0)$ en ce point. Ainsi, on peut tracer des petits segments de droite de pente $f(x, y)$ en plusieurs points (x, y) et tracer une courbe intégrale suivant le champ des tangentes.

En pratique, on trace d'abord, des courbes de pentes constantes, $f(x, y) =$ const, appelées *lignes isoclines*, puis on trace le long de chaque ligne isocline $f(x, y) = k$ plusieurs segments de droite de pente k. On obtient ainsi un champ de tangentes. Enfin, on trace une solution approchée de l'équation (1.27).

EXEMPLE 1.17. Tracer le champ des tangentes de l'équation différentielle du 1er ordre

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

et la solution approchée passant par le point $(1, 2)$.

RÉSOLUTION. Les courbes de pente constantes sont les hyperboles équilatérales $xy = k$ ainsi que les axes Ox et Oy tracés sur la fig. 1.8 □

1.9. Existence et unicité de la solution

DÉFINITION 1.3. Une fonction $f(y)$ est lipschitzienne sur $]c, d[$ s'il existe une constante $M > 0$, appelée constante de Lipschitz, telle que

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

On remarque que la condition (1.29) implique l'existence des dérivées premières à gauche et à droite de $f(y)$, mais non pas leur égalité. Géométriquement, la pente de la courbe $f(y)$ reste bornée sur $]c, d[$.

On énonce le théorème d'existence et d'unicité suivant.

THÉORÈME 1.3. *Soit le problème à valeur initiale*

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

Si la fonction $f(x, y)$ est continue et bornée,

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

sur le rectangle

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

et lipschitzienne en y sur R, alors (1.30) *admet une et une seule solution pour tout x satisfaisant*

$$|x - x_0| < \alpha, \quad \text{où} \quad \alpha = \min(a, b/K).$$

Sous les hypothèses du théorème 1.3, la solution du problème (1.30) peut s'obtenir au moyen de la méthode de Picard, c'est-à-dire la suite $y_0, y_1, \ldots, y_n, \ldots$, définie par la récurrence de Picard,

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

converge vers la solution $y(x)$.

On applique le théorème 1.3 à l'exemple suivant.

EXEMPLE 1.18. Résoudre le problème à valeur initiale:

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

et tracer la solution

RÉSOLUTION. **(a) Résolution analytique.**— On écrit l'équation différentielle sous la forme générale $y' = f(x, y)$,

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

Puisque $f(x, y)$ n'est pas continue en $y = 0$, on aura une solution pour $y < 0$ et une autre pour $y > 0$. On sépare l'équation et l'on intègre:

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

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

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

La solution générale est donc une famille de cercles de centre l'origine et de rayon r. On a les deux solutions

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

Puisque $y(0) = -2$, on prend la 2ème solution et l'on détermine la valeur de r:

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

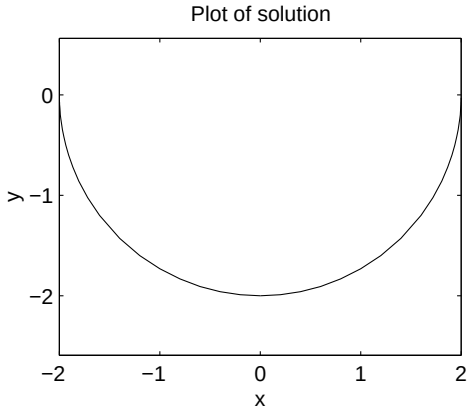

FIGURE 1.9. Graphe de la solution pour l'exemple 1.18.

Donc la solution, qui est unique, est

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

On voit que la pente $y'(x)$ de la solution tend vers $\pm\infty$ lorsque $y \to 0\pm$. Pour avoir une solution continue au voisinage de $y = 0$, on résout pour $x = x(y)$.

(b) Résolution par Matlab symbolique.—

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

(c) Résolution par Matlab numérique.— La résolution numérique de ce problème à valeur initiale est un peu futée parce que la solution générale admet deux branches: $y_{\pm}$. On a besoin d'une fonction fichier M pour implémenter le solveur `ode23`. Le fichier M `halfcircle.m` est

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

Pour obtenir la branche inférieure de la solution générale, on appelle le solveur `ode23` et la commande `plot`:

```
xspan1 = [0 -2]; % span de x = 0 = 0 a x = -2
xspan2 = [0 2]; % span de x = 0 a x = 2
y0 = [0; -2]; % condition initiale
[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('Trace de la solution')
```

La solution numérique est tracée dans fig. 1.9. □

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 5.3, we shall solve this equation numerically.

EXEMPLE 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 $y_0 = y(0)$. 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. □

EXEMPLE 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)
```

□

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.

EXEMPLE 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 $a\le b$,

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

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 solution to the initial value problem. Hence the solution is not unique. □

CHAPITRE 2

Equations différentielles linéaires du deuxième ordre

Dans ce chapitre, on introduit quelques concepts fondamentaux sur les équations différentielles linéaires du 2ème ordre. On résout les équations à coefficients constants et l'équation d'Euler–Cauchy.

Au chapitre suivant, on reprendra ces idées pour les équations linéaires nonhomogènes d'ordre quelconque.

2.1. Équations linéaires homogènes

Soit l'équation différentielle linéaire nonhomogène du 2ème ordre

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

L'équation est linéaire en y, y' et y''. Elle est nonhomogène si le 2ème membre, $r(x)$, est non nul.

On représentera souvent un opérateur différentiel linéaire quelconque par la lettre L:

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

Si le 2ème membre de (2.1) est nul, on a une équation *homogène*:

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

THÉORÈME 2.1. *Les solutions de* (2.2) *forment un espace vectoriel.*

DÉMONSTRATION. Soit y_1 et y_2 deux solutions de (2.2). La linéarité de L implique:

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

2.2. Equations homogènes à coefficients constants

Soit l'équation différentielle linéaire homogène du 2ème ordre à *coefficients constants*:

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

On résout cette équation en supposant que la solution est de la forme exponentielle suivante:

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

Alors,

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

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

Puisque $e^{\lambda x}$ ne s'annule jamais, on obtient l'*équation caractéristique*

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

pour λ et les *valeurs propres*

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

Si $\lambda_1 \neq \lambda_2$, on a deux solutions distinctes

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

et l'équation générale, qui contient deux constantes arbitraires, est

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

EXEMPLE 2.1. Résoudre l'équation différentielle

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

RÉSOLUTION. L'équation caractéristique est

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

Alors $\lambda_1 = -2$ et $\lambda_2 = -3$. La solution générale est donc:

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

2.3. Base de l'espace solution

On étend aux fonctions définies sur $[a, b]$ la notion d'indépendance linéaire de deux vecteurs dans $\mathbb{R}^2$.

DÉFINITION 2.1. Deux fonctions $f_1(x)$ et $f_2(x)$ sont linéairement indépendantes sur $[a, b]$ si l'identité

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

implique

$$c_1 = c_2 = 0.$$

Sinon, elles sont linéairement dépendantes.

Si $f_1(x)$ et $f_2(x)$ sont linéairement dépendantes sur $[a, b]$, il existe deux nombres $(c_1, c_2) \neq (0, 0)$ tels que l'identité (2.8) est satisfaite sur $[a, b]$. Supposant que $c_1 \neq 0$, on a

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

On conclut que si

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

alors f_1 et f_2 sont linéairement indépendantes sur $[a, b]$. On emploiera souvent cette caractérisation d'indépendance linéaire pour deux fonctions.

DÉFINITION 2.2. La solution générale de l'équation homogène (2.2) engendre l'espace vectoriel des solutions de (2.2).

THÉORÈME 2.2. *Soit $y_1(x)$ et $y_2(x)$ deux solutions de (2.2) sur $[a,b]$. Alors, la solution*

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

est une solution générale de (2.2) si et seulement si y_1 et y_2 sont linéairement indépendantes sur $[a,b]$.

DÉMONSTRATION. On donnera la démonstration pour les équations d'ordre n quelconque au chapitre suivant. □

L'exemple suivant illustre l'utilité de la solution générale.

EXEMPLE 2.2. Résoudre le problème aux valeurs initiales:

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

RÉSOLUTION. **(a) Résolution analytique.**— L'équation caractéristique est

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

Alors $\lambda_1 = 1$ et $\lambda_2 = -2$. Les solutions

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

sont linéairement indépendantes car

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

La solution générale est donc

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

On détermine la valeur des constantes au moyen des conditions initiales:

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

On obtient donc le système linéaire

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

Puisque

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

la solution $\boldsymbol{c}$ est unique. On obtient cette solution par la règle de Cramer:

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

La solution du problème aux valeurs initiales est donc

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

Cette solution est unique.

(b) Résolution par Matlab symbolique.—

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

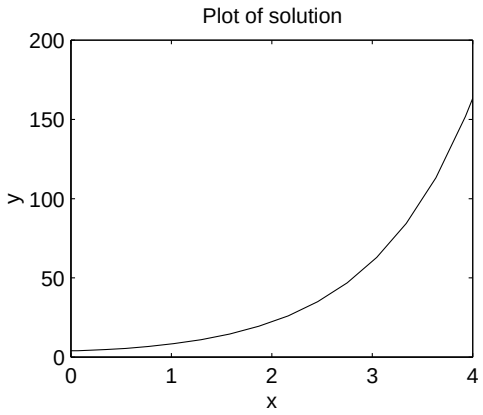

FIGURE 2.1. Graphe de la solution de l'équation linéaire de l'exemple 2.2.

(c) Résolution par Matlab numérique.— On récrit l'équation différentielle du second ordre en un système du premier ordre au moyen des variables

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

Alors,

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

Le fichier M `exp22.m`:

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

On appelle le solveur `ode23` et la commande `plot`:

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

La solution numérique se trouve à la fig. 2.1. □

2.4. Solutions indépendantes

La forme des solutions indépendantes de l'équation homogène

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

dépend de la forme des racines

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

de l'équation caractéristique

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

Il faut considérer trois cas: $\lambda_1 \neq \lambda_2$ réelles, $\lambda_2 = \bar{\lambda}_1$ complexes et $\lambda_1 = \lambda_2$ réelles.

Cas I. Dans le cas de deux valeurs propres réelles et distinctes, $\lambda_1 \neq \lambda_2$, on a vu à la section 2.1 que les deux solutions:

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

sont indépendantes. La solution générale est donc

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

Cas II. Dans le cas de deux valeurs propres complexes, conjuguées l'une de l'autre, on a

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

On emploie l'identité d'Euler,

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

On a alors les deux solutions complexes

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

Puisque $\lambda_1 \neq \lambda_2$, les solutions u_1 et u_2 sont indépendantes. Pour avoir deux solutions réelles indépendantes, on fait le changement de base suivant:

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

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

On voit immédiatement que y_1 et y_2 sont indépendantes. Alors, la solution générale est

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

Cas III. Dans le cas d'une valeur propre réelle double, $\lambda = \lambda_1 = \lambda_2$, (2.11) admet une solution de la forme

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

Pour obtenir une seconde solution indépendante de y_1 par la méthode de la variation des paramètres décrite plus loin, on pose

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

Il est important de remarquer que le paramètre u est fonction de x et que y_1 est solution de (2.11). On substitue y_2 dans (2.11) et l'on additionne les trois équations suivantes:

$$\begin{aligned}
b y_2(x) &= b u(x) y_1(x) \\
a y_2'(x) &= a u(x) y_1'(x) + a y_1(x) u'(x) \\
y_2''(x) &= u(x) y_1''(x) \;\; + 2 y_1'(x) u'(x) + y_1(x) u''(x) \\
L y_2 &= u(x) L y_1 \quad\;\; + [a y_1(x) + 2 y_1'(x)] u'(x) + y_1(x) u''(x).
\end{aligned}$$

Le 1er membre de la somme est nul puisqu'on suppose que y_2 est solution de $Ly = 0$. Le premier terme du 2ème membre est nul puisque y_1 est solution de $Ly = 0$.

Le 2ème terme du 2ème membre est nul parce que

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

du fait que le discriminant $\Delta = a^2 - 4b = 0$. Alors,

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

Il suit que

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

d'où

$$u'(x) = k_1$$

et

$$u(x) = k_1 x + k_2.$$

On a donc

$$y_2(x) = k_1 x\,e^{\lambda x} + k_2\,e^{\lambda x}.$$

Il suffit de prendre $k_2 = 0$ parce que le 2ème terme du 2ème membre est déjà contenu dans l'enveloppe linéaire de y_1. On peut aussi prendre $k_1 = 1$ puique la solution générale contient déjà une constante multipliant y_2.

On voit immédiatement que les solutions

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

sont linéairement indépendantes.

Alors, la solution générale est

$$y(x) = c_1\,e^{\lambda x} + c_2 x\,e^{\lambda x}. \tag{2.21}$$

2.5. Modélisation en mécanique

On considère quelques modèles de la mécanique élémentaire.

EXEMPLE 2.3 (Oscillation libre). Soit une ressort en position verticale pendante fixé à une poutre rigide. Le ressort résiste à l'extension et à la compression et sa constante de Hooke est k. Etudier le problème de l'oscillation libre verticale d'une masse de m kg fixée au bout inférieur du ressort.

RÉSOLUTION. On considère la direction Oy vers le bas comme positive. Soit s_0 m l'étirement du ressort par le poids au repos en position $y = 0$ (V. figure 2.2).

On néglige toute friction. La force due à la gravité est

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

La force de restauration exercée par le ressort est

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

Quand le système est au repos, la résultante est nulle,

$$F_1 + F_2 = 0,$$

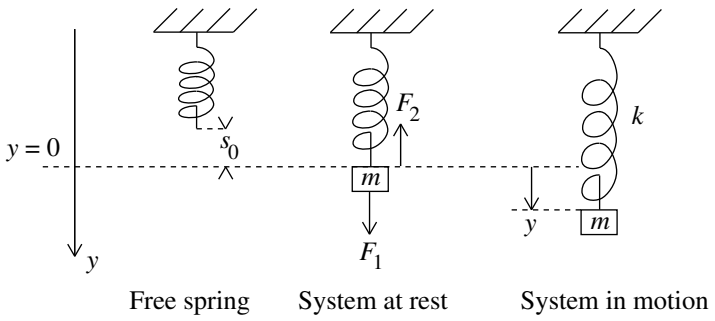

FIGURE 2.2. Système non amorti.

par la 2ème loi de Newton. Puisque le système en mouvement est non amorti, par la même loi, la résultante est

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

Or l'accélération est donnée par $a = y''$. Donc

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

où $\omega/2\pi$ Hz est la fréquence du système. L'équation caractéristique de cette équation différentielle,

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

admet les valeurs propres

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

imaginaires. La solution générale est donc

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

On voit que le système oscille librement sans perte d'énergie. □

EXEMPLE 2.4 (Système amorti). Soit une ressort en position verticale pendante fixé à une poutre rigide. Le ressort résiste à l'extension et à la compression et sa constante de Hooke est k. Etudier le problème du mouvement vertical amorti d'une masse de m kg fixée au bout inférieur du ressort (V. figure 2.3). La constante d'amortissement est c

RÉSOLUTION. On considère la direction Oy vers le bas comme positive. Soit s_0 m l'étirement du ressort par le poids au repos en position $y = 0$ (V. figure 2.2).

La force due à la gravité est

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

La force de restauration exercée par le ressort est

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

Quand le système est au repos, la résultante est nulle,

$$F_1 + F_2 = 0,$$

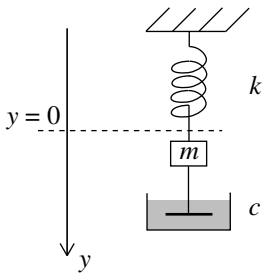

FIGURE 2.3. Système amorti.

par la 2ème loi de Newton. Puisque l'amortissement est dans la direction opposée au mouvement, par la même loi, la résultante est

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

Or l'accélération est donnée par $a = y''$. Donc

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

L'équation caractéristique de cette équation différentielle,

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

admet les valeurs propres

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

On a les trois cas suivants.

Cas I: Sur-amortissement. Si $c^2 > 4mk$, le système est sur-amorti. Les deux valeurs propres sont négatives puisque

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

La solution générale,

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

décroit exponentiellement vers zéro sans oscillation du système.

Cas II: Sous-amortissement. Si $c^2 < 4mk$, le système est sous-amorti. Les deux valeurs propres sont complexes conjuguées l'une de l'autre,

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

La solution générale,

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

décroît exponentiellement vers zéro avec oscillation du système.

Cas III: Amortissement critique. Si $c^2 = 4mk$, le système est critiquement amorti. Les deux valeurs propres sont réelles et égales,

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

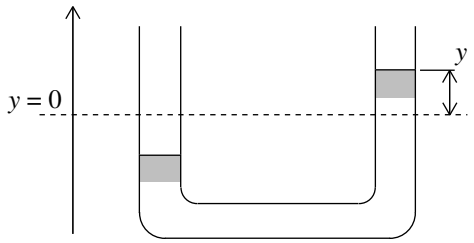

FIGURE 2.4. Mouvement vertical d'un liquide dans un tube en U.

La solution générale,

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

décroît exponentiellement vers zéro avec une augmentation initiale de $y(t)$ si $c_2 > 0$. □

EXEMPLE 2.5 (Oscillation aqueuse dans un tube en forme de U).
Calculer la fréquence du mouvement oscillatoire vertical de 2 L d'eau dans un tube en forme de U de 0.04 m de diamètre.

RÉSOLUTION. On néglige la friction entre le liquide et la paroi du tube. La masse du liquide est $m = 2\,\text{kg}$. Le volume causant la force de restauration est

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

(V. figure 2.4). La masse du volume V est

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

et la force de restauration est

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

Par la 2ème loi de Newton,

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

c'est-à-dire

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

où

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

La fréquence est donc

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

EXEMPLE 2.6 (Oscillation d'un pendule). Calculer la fréquence des oscillations de faible amplitude d'un pendule de masse m kg et de longueur $L = 1$ m.

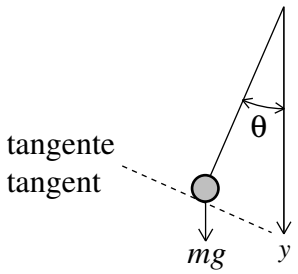

FIGURE 2.5. Pendule en mouvement.

RÉSOLUTION. On néglige la résistance de l'air et la masse de la tige. Soit θ l'angle, en radians, porté par le pendule à partir de la verticale (V. figure 2.5).

La force tangentielle est

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

Puisque la longueur de la tige est fixée, la composante de la force orthogonale au mouvement est nulle. Il suffit donc de considérer la composante tangentielle de la force de restauration due à la gravité. On a donc

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

puisque $\sin\theta \approx \theta$ si θ est suffisamment petit. Alors,

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

La fréquence est donc

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

$\square$

2.6. Equation d'Euler–Cauchy

Considérons l'équation d'Euler–Cauchy homogène

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

A cause de la forme particulière de l'opérateur différentiel à coefficients variables,

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

dont chaque terme est de la forme $a_kx^kD^k$, où a_k est une constante, on peut résoudre (2.22) en posant

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

dans (2.22):

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

On peut diviser par x^m. On a alors l'équation caractéristique

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

Les valeurs propres sont

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

Il faut considérer trois cas: $m_1 \neq m_2$ réelles, $m_2 = \overline{m}_1$ complexes distinctes et $m_1 = m_2$ réelles.

Cas I. Si l'on a deux racines réelles distinctes, la solution générale de (2.22) est

$$y(x) = c_1 x^{m_1} + c_2 x^{m_2}. \tag{2.26}$$

Cas II. Si l'on a deux racines complexes, conjuguées l'une de l'autre,

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

on a deux solutions complexes indépendantes:

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

et

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

Pour $x > 0$, on obtient les deux solutions réelles indépendantes suivantes en additionnant et soustrayant u_1 et u_2, et en divisant la somme et la différence respectivement par 2 et $2i$,

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

La solution génrérale de (2.22) est donc

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

Cas III. Si l'on a deux racines réelles égales,

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

On a une solution de la forme

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

Pour obtenir une 2ème solution indépendante au moyen de la variation des paramètres, on pose

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

dans (2.22) et l'on additionne les 1ers et 2èmes membres respectifs des trois expressions suivantes:

$$\begin{aligned}
b y_2(x) &= b u(x) y_1(x) \\
a x y_2'(x) &= a x u(x) y_1'(x) + a x y_1(x) u'(x) \\
x^2 y_2''(x) &= x^2 u(x) y_1''(x) + 2x^2 y_1'(x) u'(x) + x^2 y_1(x) u''(x) \\
L y_2 &= u(x) L y_1 \qquad + [a x y_1(x) + 2x^2 y_1'(x)] u'(x) + x^2 y_1(x) u''(x).
\end{aligned}$$

Le 1er membre de la somme est nul puisqu'on suppose que y_2 est solution de $Ly = 0$. Le premier terme du 2ème membre est nul puisque y_1 est solution de $Ly = 0$.

Le coefficient de u' est

$$\begin{aligned}
a x y_1(x) + 2x^2 y_1'(x) &= a x x^m + 2m x^2 x^{m-1} = a x^{m+1} + 2m x^{m+1} \\
&= (a + 2m) x^{m+1} = \left(a + 2\,\frac{1-a}{2}\right) x^{m+1} = x^{m+1}.
\end{aligned}$$

On obtient donc

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

On peut diviser par x^{m+1}:

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

Puisque u est absente de cette équation différentielle, on peut en réduire l'ordre en posant

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

Alors on a l'équation séparable

$$x\,\frac{dv}{dx} + v = 0, \qquad \text{c'est-à-dire} \quad \frac{dv}{v} = -\frac{dx}{x},$$

qu'on peut intégrer:

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

La 2ème solution, indépendante de la 1ère, est

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

La solution générale de (2.22) est donc

$$y(x) = c_1 x^m + c_2(\ln x)x^m. \tag{2.28}$$

EXEMPLE 2.7. Trouver la solution générale de l'équation d'Euler–Cauchy

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

RÉSOLUTION. **(a) Résolution analytique.**— En posant $y = x^m$ dans l'équation différentielle, on obtient

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

L'équation caractéristique est

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

Les valeurs propres,

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

sont rélles et distinctes. La solution générale est donc

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

(b) Résolution par Matlab symbolique.—

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

□

EXEMPLE 2.8. Résoudre:

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

RÉSOLUTION. **(a) Résolution analytique.**— La solution générale trouvée à l'exemple 2.7 est

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

Utilisant les conditions initiales, on obtient le système linéaire en c_1 et c_2:

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

qui admet la solution

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

Donc, l'unique solution est

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

(b) Résolution par Matlab symbolique.—

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

(c) Résolution par Matlab numérique.— On récrit l'équation différentielle du second ordre en un système du premier ordre au moyen des variables

$$\begin{aligned} y(1) &= y, \\ y(2) &= y', \end{aligned}$$

avec les conditions initiales en $x = 1$:

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

Alors,

$$\begin{aligned} y_1' &= y_2, \\ y_2' &= 6y_1/x^2. \end{aligned}$$

Le fichier M `euler2.m`:

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

On appelle le solveur `ode23` et la commande `plot`:

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

Le graphe de la solution numérique est à la fig. 2.6. □

EXEMPLE 2.9. Trouver la solution générale de l'équation d'Euler–Cauchy

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

RÉSOLUTION. L'équation caractéristique

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

admet une racine double $m = -3$. La solution générale est donc

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

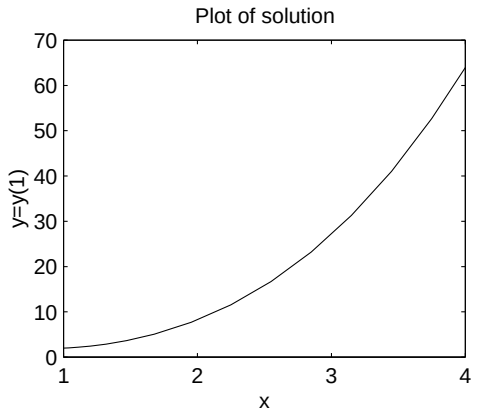

FIGURE 2.6. Graphe de la solution de l'équation linéaire de l'exemple 2.8.

EXEMPLE 2.10. Trouver la solution générale de l'équation d'Euler–Cauchy

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

RÉSOLUTION. L'équation caractéristique

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

admet deux racines complexes conjuguées l'une de l'autre

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

La solution générale est donc

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

On traitera de l'existence et de l'unicité de la solution des problèmes aux valeurs initiales au chapitre suivant.

CHAPITRE 3

Équations différentielles linéaires d'ordre quelconque

3.1. Équations homogènes

Considérons l'équation différentielle linéaire *nonhomogène* d'ordre n,

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

à coefficients variables, $a_0(x), a_1(x), \ldots, a_{n-1}(x)$. Notons L l'opérateur différentiel du 1er membre:

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

Alors, l'équation nonhomogène (3.1) s'écrit

$$Lu = r(x).$$

Si $r \equiv 0$, l'équation (3.1) est dite *homogène*:

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

c'est-à-dire

$$Ly = 0.$$

DÉFINITION 3.1. Une *solution* de (3.1) ou (3.3) sur $]a, b[$ est une fonction $y(x)$ n fois continûment dérivable sur $]a, b[$ qui satisfait identiquement l'équation différentielle.

Le théorème 2.1 se généralise aux équations linéaires homogènes d'ordre n quelconque.

THÉORÈME 3.1. *Les* ***solutions*** *de* (3.3) *forment un espace vectoriel.*

DÉMONSTRATION. Soit $y_1, y_2, \ldots, y_k$, k solutions de $Ly = 0$. La linéarité de L implique:

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

DÉFINITION 3.2. On dit que n **fonctions**, $f_1, f_2, \ldots, f_n$, sont *linéairement dépendantes* sur $]a, b[$ si, et seulement si, il existe n constantes non toutes nulles,

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

telles que

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

Sinon, elles sont *linéairement indépendantes.*

REMARQUE 3.1. Soit $f_1, f_2, \ldots, f_n$, n **fonctions** linéairement dépendantes. Sans perte de généralité, on peut supposer que $k_1 \neq 0$ dans (3.4). Alors f_1 est une combinaison linéaires de $f_2, f_3, \ldots, f_n$.

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

On a le théorème d'existence et d'unicité suivant.

THÉORÈME 3.2. *Si les fonctions $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ sont continues sur $]a,b[$ et $x_0 \in]a,b[$, alors le problème aux valeurs initiales*

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

admet une et une seule solution.

DÉMONSTRATION. On peut démontrer le théorème en réduisant l'équation différentielle d'ordre n à un système de n équations différentielles du 1er ordre. En effet, posons

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

Alors le problème aux valeurs initiales devient

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

Sous forme matricielle, on a

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

On peut démontrer par la méthode de Picard que ce système admet une et une seule solution. La récurrence de Picard est de la forme:

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

DÉFINITION 3.3. Le wronskien de n **fonctions**, $f_1(x), f_2(x), \ldots, f_n(x)$, $n-1$ fois différentiables sur $]a,b[$ est le déterminant d'ordre n:

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

On caractérise la dépendance linéaire de n **solutions** de l'équation linéaire homogène (3.3) au moyen de leur wronskien.

Montrons d'abord un lemme d'Abel.

LEMME 3.1 (Abel). *Soit n **solutions**, $y_1, y_2, \ldots, y_n$, de (3.3) sur $]a,b[$. Alors le wronskien $W(x) = W(y_1, y_2, \ldots, y_n)(x)$ satisfait l'identité suivante:*

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

DÉMONSTRATION. Pour simplifier l'écriture, prenons $n = 3$; le cas général se traite de la même façon. Soit $W(x)$ le wronskien de trois solutions $y_1, y_2, , y_3$. La dérivée du wronskien est de la forme:

$$\begin{aligned}
W'(x) &= \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1'' & y_2'' & y_3'' \end{vmatrix}' \\
&= \begin{vmatrix} y_1' & y_2' & y_3' \\ y_1' & y_2' & y_3' \\ y_1'' & y_2'' & y_3'' \end{vmatrix} + \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1'' & y_2'' & y_3'' \\ y_1'' & y_2'' & y_3'' \end{vmatrix} + \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1''' & y_2''' & y_3''' \end{vmatrix} \\
&= \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1''' & y_2''' & y_3''' \end{vmatrix} \\
&= \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ -a_0y_1 - a_1y_1' - a_2y_1'' & -a_0y_2 - a_1y_2' - a_2y_2'' & -a_0y_3 - a_1y_3' - a_2y_3'' \end{vmatrix},
\end{aligned}$$

puisque les deux premiers déterminants sont nuls et dans le dernier déterminant on emploie le fait que y_k, $k = 1, 2, 3$, est une solution de l'équation homogène (3.3).

Si l'on additionne a_0 fois la 1ère ligne et a_1 fois la 2ème ligne à la 3ème ligne, on obtient

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

C'est une équation différentielle séparable:

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

La solution est

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

c'est-à-dire

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

THÉORÈME 3.3. *Si les coefficients $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ de (3.3) sont continus sur $]a, b[$, alors n **solutions**, $y_1, y_2, \ldots, y_n$, de (3.3) sont linéairement dépendantes si et seulement si leur wronskien s'annule en un point quelconque $x_0 \in]a, b[$:*

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

DÉMONSTRATION. Si les solutions sont linéairement dépendantes, alors par la définition 3.2 il existe n constantes non toutes nulles,

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

telles que

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

Si l'on dérive cette identité $n-1$ fois, on obtient

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

Ce système linéaire homogène en $k_1,k_2,\ldots,k_n$ se récrit sous la forme matricielle:

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

c'est-à-dire

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

Puisque, par hypothèse, la solution $\boldsymbol{k}$ est non nulle, le déterminant du système doit être nul,

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

D'autre part, si le wronskien de n **solutions** s'annule en un point $x_0\in]a,b[$,

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

il s'annule pour tout $x\in]a,b[$ par le lemme 3.1 d'Abel. Alors le déterminant $W(x)$ du système (3.9) est nul pour tout $x\in]a,b[$. Donc ce système admet une solution $\boldsymbol{k}$ non nulle. Par conséquent les solutions $y_1,y_2,\ldots,y_n$ de (3.3) sont linéairement dépendantes. □

REMARQUE 3.2. Le wronskien de n fonctions linéairement dépendantes sur $]a,b[$ est nécesairement nul sur $]a,b[$, comme on peut voir à la 1ère partie de la démonstration du théorème 3.3. Mais pour des fonctions qui *ne sont pas solutions* d'une équation différentielle linéaire homogène, le fait que le wronskien soit nul sur $]a,b[$ n'est pas une condition suffisante pour que ces fonctions soient dépendantes. Par exemple, les fonctions $u_1=x^3$ et $u_2=|x|^3$ sont de classe $C^1[-1,1]$ et sont linéairement independantes sur l'intervalle $[-1,1]$, mais leur wronskien est identiquement nul sur cet intervalle.

COROLLAIRE 3.1. *Si les coefficients $a_0(x),a_1(x),\ldots,a_{n-1}(x)$ de (3.3) sont continus sur $]a,b[$, alors n **solutions**, $y_1,y_2,\ldots,y_n$, de (3.3) sont linéairement indépendantes si et seulement si leur wronskien ne s'annule pas en un seul point $x_0\in]a,b[$.*

COROLLAIRE 3.2. *Soit $f_1(x),f_2(x),\ldots,f_n(x)$ n fonctions de classe C^n sur un intervalle réel I. Si $W(f_1,\ldots,f_n)(x)\neq 0$ sur I, alors il existe une unique équation différentielle homogène d'ordre n (avec le coefficient de $y^{(n)}$ l'unité):*

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

pour laquelle ces fonctions forment un système de n solutions indépendantes.

EXEMPLE 3.1. Montrer que les fonctions

$$y_1 = \cosh x \quad \text{et} \quad y_2 = \sinh x$$

sont linéairement indépendantes.

RÉSOLUTION. Puisque y_1'' et y_2'' sont continues, on peut donc appliquer le corollaire 3.2:

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

pour tout x. Donc y_1 et y_2 sont indépendantes. On voit facilement que y_1 et y_2 sont solutions de l'équation différentielle

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

□

Dans la résolution on a employé l'identité suivante:

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

EXEMPLE 3.2. Utiliser le wronskien pour montrer que les fonctions

$$y_1 = x^m \quad \text{et} \quad y_2 = x^m \ln x$$

sont linéairement indépendantes sur $x > 0$ et construire une équation différentielle du second ordre pour laquelle ces fonctions sont solutions.

RÉSOLUTION. On vérifie que le wronskien de y_1 et y_2 ne s'annule pas sur $x > 0$:

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

Alors, par le corollaire 3.2, y_1 et y_2 sont linéairement indpendantes. Par le même corollaire,

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

Pour évaluer ce déterminant, on multiplie la 2e et la 3e lignes respectivement par x et x^2, on divise la 2e et la 3e colonnes par x^m, on soustrait m fois la 1ère ligne de la seconde ligne et $m(m-1)$ fois la 1ère ligne de la 3e ligne:

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

Enfin, on dévelope ce déterminat suivant la 2e colonne, ce qui donne l'équation d'Euler–Cauchy:

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

DÉFINITION 3.4. On appelle *système fondamental* ou *base* sur $]a,b[$ n solutions, $y_1, y_2, \ldots, y_n$, de (3.3) linéairement indépendantes sur $]a,b[$.

DÉFINITION 3.5. Soit $y_1, y_2, \ldots, y_n$ un système fondamental pour (3.3). On appelle *solution générale* de (3.3) sur $]a,b[$ une solution de la forme

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

où $c_1, c_2, \ldots, c_n$ sont n constantes arbitraires.

THÉORÈME 3.4. *Si les fonctions $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ sont continues sur $]a,b[$, alors l'équation linéaire homogène* (3.3) *admet une solution générale sur $]a,b[$.*

DÉMONSTRATION. Par le théorème 3.2, pour $i = 1, 2, \ldots, n$, le problème aux valeurs initiales (3.5),

$$Ly = 0, \qquad k_i = 1, \qquad k_j = 0 \quad j \neq i,$$

admet une (et une seule) solution $y_i(x)$ telle que

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

Alors le wronskien

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

où I_n est la matrice identité d'ordre n. Les solutions sont donc indépendantes par le corollaire 3.1. $\square$

THÉORÈME 3.5. *Si les fonctions $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ sont continues sur $]a,b[$, alors toute solution du problème aux valeurs initiales* (3.5) *sur $]a,b[$ s'obtient au moyen d'une solution générale.*

DÉMONSTRATION. Soit

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

une solution générale de (3.3). Le système

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

admet une solution unique $\boldsymbol{c}$ puisque le déterminant du système est non nul. $\square$

3.2. Equations linéaires homogènes à coefficients constants

Considérons l'équation différentielle linéaire homogène d'ordre n,

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1 y' + a_0 y = 0, \tag{3.11}$$

à coefficients constants, $a_0, a_1, \ldots, a_{n-1}$. Notons L l'opérateur différentiel du 1er membre:

$$L := D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0 I, \qquad D := {}' = \frac{d}{dx}. \tag{3.12}$$

On pose $y(x) = e^{\lambda x}$ dans (3.11). On obtient alors l'équation caractéristique

$$p(\lambda) := \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0, \tag{3.13}$$

Si les n racines de $p(\lambda) = 0$ sont distinctes, on a n solutions indépendantes:

$$y_1(x) = e^{\lambda_1 x}, \qquad y_2(x) = e^{\lambda_2 x}, \qquad \ldots, \qquad y_n(x) = e^{\lambda_n x}, \tag{3.14}$$

et la solution générale est de la forme

$$y(x) = c_1\, e^{\lambda_1 x} + c_2\, e^{\lambda_2 x} + \cdots + c_n\, e^{\lambda_n x}. \tag{3.15}$$

Si (3.13) admet une racine double, disons, $\lambda_1 = \lambda_2$, on a deux solutions indépendantes de la forme

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

De même, si l'on a une racine triple, disons, $\lambda_1 = \lambda_2 = \lambda_3$, on a trois solutions indépendantes de la forme

$$y_1(x) = e^{\lambda_1 x}, \qquad y_2(x) = x\, e^{\lambda_1 x}, \qquad y_3(x) = x^2\, e^{\lambda_1 x}.$$

On démontre le théorème suivant.

THÉORÈME 3.6. *Soit μ une racine de multiplicité m de l'équation caractéristique* (3.13). *Alors l'équation différentielle* (3.11) *admet m solutions indépendantes de la forme*

$$y_1(x) = e^{\mu x}, \qquad y_2(x) = x\, e^{\mu x}, \qquad \ldots, \qquad y_m(x) = x^{m-1}\, e^{\mu x}. \tag{3.16}$$

DÉMONSTRATION. Écrivons

$$p(D)y = (D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0)y = 0.$$

Puisque, par hypothèse,

$$p(\lambda) = q(\lambda)(\lambda - \mu)^m,$$

et que les coefficients sont constants, l'opérateur différentiel s'écrit sous la forme

$$p(D) = q(D)(D - \mu)^m.$$

On voit par récurrence que les m fonctions (3.16),

$$x^k\, e^{\mu x}, \qquad k = 0, 1, \ldots, m-1,$$

satisfont les équations suivantes:

$$\begin{aligned}(D-\mu)\left(x^k\,e^{\mu x}\right) &= kx^{k-1}\,e^{\mu x}+\mu x^k\,e^{\mu x}-\mu x^k\,e^{\mu x}\\ &= kx^{k-1}\,e^{\mu x},\\ (D-\mu)^2\left(x^k\,e^{\mu x}\right) &= (D-\mu)\left(kx^{k-1}\,e^{\mu x}\right)\\ &= k(k-1)x^{k-2}\,e^{\mu x},\\ &\vdots\\ (D-\mu)^k\left(x^k\,e^{\mu x}\right) &= k!\,e^{\mu x},\\ (D-\mu)^{k+1}\left(x^k\,e^{\mu x}\right) &= k!\,(\mu\,e^{\mu x}-\mu\,e^{\mu x})=0.\end{aligned}$$

Puisque $m\geq k+1$, on a

$$(D-\mu)^m\left(x^k\,e^{\mu x}\right)=0,\qquad k=0,1,\ldots,m-1.$$

Donc, par le lemme 3.2 qui suit, les fonctions (3.16) forment m solutions indépendantes de (3.11). □

LEMME 3.2. *Soit*

$$y_1(x)=e^{\mu x},\qquad y_2(x)=x\,e^{\mu x},\qquad \ldots,\qquad y_m(x)=x^{m-1}\,e^{\mu x},$$

m solutions d'une équation différentielle linéaire homogène. Alors elles sont indépendantes.

DÉMONSTRATION. Par le corollaire 3.1, il suffit de montrer que le wronskien des solutions ne s'annule pas en $x=0$. On a vu dans la démonstration du théorème précédent que

$$(D-\mu)^k\left(x^k\,e^{\mu x}\right)=k!\,e^{\mu x},$$

c'est-à-dire

$$D^k\left(x^k\,e^{\mu x}\right)=k!\,e^{\mu x}+\text{ des termes en } x^l\,e^{\mu x},\qquad l=1,2,\ldots,k-1.$$

Donc

$$D^k\left(x^k\,e^{\mu x}\right)\big|_{x=0}=k!,\qquad D^k\left(x^{k+l}\,e^{\mu x}\right)\big|_{x=0}=0,\quad l\geq 1.$$

Il suit que la matrice M du wronskien est triangulaire inférieure avec $m_{i,i}=(i-1)!$,

$$W(0)=\begin{bmatrix}0! & 0 & 0 & \ldots & 0\\ \times & 1! & 0 & & 0\\ \times & \times & 2! & 0 & \vdots\\ \vdots & & \ddots & \ddots & 0\\ \times & \times & \ldots & \times & (m-1)!\end{bmatrix}\neq 0.\quad\square$$

EXEMPLE 3.3. Trouver la solution générale de

$$(D^4-13D^2+36I)y=0.$$

RÉSOLUTION. Le polynôme caractéristique se met facilement en facteur:

$$\begin{aligned}\lambda^4-13\lambda^2+36 &= (\lambda^2-9)(\lambda^2-4)\\ &= (\lambda+3)(\lambda-3)(\lambda+2)(\lambda-2).\end{aligned}$$

Alors:

$$y(x) = c_1\, e^{-3x} + c_2\, e^{3x} + c_3\, e^{-2x} + c_4\, e^{2x}.$$

Le solveur de polynôme de Matlab.— Pour trouver les zéros du polynôme caractéristique

$$\lambda^4 - 13\lambda^2 + 36$$

avec Matlab, on représente le polynôme par le vecteur de ses coefficients,

$$p = \begin{bmatrix} 1 & 0 & -13 & 0 & 36 \end{bmatrix}$$

et l'on applique la commande `roots` sur 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
```

De fait, la commande `roots` construit une matrice compagnon C de p (voir la démonstration du théorème 3.2)

$$C = \begin{bmatrix} 0 & 13 & 0 & -36 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

et emploie l'algorithme QR pour trouver les valeurs propres de C qui, de fait, sont les zéros de 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
```

□

EXEMPLE 3.4. Trouver la solution générale de

$$(D - I)^3 y = 0.$$

RÉSOLUTION. Le polynôme caractéristique $(\lambda - 1)^3$ admet un zéro triple:

$$\lambda_1 = \lambda_2 = \lambda_3 = 1,$$

Alors:

$$y(x) = c_1\, e^x + c_2 x\, e^x + c_3 x^2\, e^x. \quad \square$$

EXEMPLE 3.5. Trouver la solution générale de l'équation d'Euler–Cauchy

$$x^3 y''' - 3x^2 y'' + 6xy' - 6y = 0.$$

RÉSOLUTION. **(a) Résolution analytique.**— En posant

$$y = x^m$$

dans l'équation différentielle, on obtient

$$m(m-1)(m-2)x^m - 3m(m-1)x^m + 6mx^m - 6x^m = 0,$$

d'où l'équation caractéristique:

$$m(m-1)(m-2) - 3m(m-1) + 6m - 6 = 0.$$

Si l'on remarque que $m-1$ est un facteur commun, on obtient

$$\begin{aligned}(m-1)[m(m-2) - 3m + 6] &= (m-1)(m^2 - 5m + 6)\\ &= (m-1)(m-2)(m-3) = 0.\end{aligned}$$

Alors:

$$y(x) = c_1\, x + c_2\, x^2 + c_3\, x^3.$$

(b) Résolution par Matlab symbolique.—

```
dsolve('x^3*D3y-3*x^2*D2y+6*x*Dy-6*y=0','x')
y = C1*x+C2*x^2+C3*x^3
```

□

EXEMPLE 3.6. Soit

$$y_1(x) = e^{x^2}$$

une solution de l'équation différentielle du 2ème ordre:

$$Ly := y'' - 4xy' + (4x^2 - 2)y = 0.$$

Trouver une seconde solution indépendante.

RÉSOLUTION. **(a) Résolution analytique.**— On utilise la méthode de la variation des paramètres. En posant

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

on obtient

$$\begin{aligned}(4x^2 - 2)y_2 &= (4x^2 - 2)uy_1,\\ -4xy_2' &= -4xuy_1' - 4xu'y_1,\\ y_2'' &= uy_1'' + 2u'y_1' + u''y_1.\end{aligned}$$

On additionne respectivement les 1ers et les 2èmes membres:

$$Ly_2 = uLy_1 + (2y_1' - 4xy_1)u' + y_1u''.$$

On a $Ly_1 = 0$ et $Ly_2 = 0$ puisque y_1 est une solution et l'on veut que y_2 soit une solution. Maintenant on remplace y_1 par son expression dans l'équation différentielle en u:

$$e^{x^2}u'' + \left(4x\, e^{x^2} - 4x\, e^{x^2}\right)u' = 0,$$

d'où

$$u'' = 0 \implies u' = k_1 \implies u = k_1x + k_2$$

et

$$y_2(x) = (k_1x + k_2)e^{x^2}.$$

Il suffit de prendre $k_2 = 0$ parce que $k_2 e^{x^2}$ est contenu dans y_1 et $k_1 = 1$ parce que $x\, e^{x^2}$ sera multiplié par une constante arbitraire. Alors la solution générale est de la forme

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

(b) Résolution par Matlab symbolique.—

```
dsolve('D2y-4*x*Dy+(4*x^2-2)*y=0','x')
y = C1*exp(x^2)+C2*exp(x^2)*x
```

□

3.3. Equations linéaires nonhomogènes

Considérons l'équation différentielle linéaire *nonhomogène* d'ordre n,

$$Ly := y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = r(x). \tag{3.17}$$

Soit

$$y_h(x) = c_1 y_1(x) + c_2 y_2(x) + \cdots + c_n y_n(x), \tag{3.18}$$

une solution générale de l'équation homogène

$$Ly = 0.$$

De plus, soit $y_p(x)$ une *solution particulière* de l'équation nonhomogène (3.17). Alors,

$$y_g(x) = y_h(x) + y_p(x)$$

est la solution générale de (3.17). En effet,

$$Ly_g = Ly_h + Ly_p = 0 + r(x).$$

EXEMPLE 3.7. Touver la solution générale $y_g(x)$ de

$$y'' - y = 3\, e^{2x}$$

si

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

est une solution particulière.

RÉSOLUTION. **(a) Résolution analytique.**— Il est facile de voir que e^{2x} est bien une solution particulière. Puisque

$$y'' - y = 0 \implies \lambda^2 - 1 = 0 \implies \lambda = \pm 1,$$

alors

$$y_h(x) = c_1\, e^{x} + c_2\, e^{-x}$$

et

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

(b) Résolution par Matlab symbolique.—

```
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
```

□

Voici une seconde méthode de résolution d'une équation linéaire nonhomogène du 1er ordre,

EXEMPLE 3.8. Trouver la solution générale de l'équation linéaire nonhomogène du 1er ordre,

$$Ly := y' + f(x)y = r(x). \tag{3.19}$$

RÉSOLUTION. L'équation homogène $Ly = 0$ est séparable:

$$\frac{dy}{y} = -f(x)\,dx \implies \ln|y| = -\int f(x)\,dx \implies y_h(x) = e^{-\int f(x)\,dx}.$$

On trouve une solution particulière par variation des paramètres en posant

$$y_p(x) = u(x)y_h(x)$$

dans l'équation nonhomogène $Ly = r(x)$:

$$\begin{aligned} y_p' &= uy_h' + u'y_h \\ f(x)y_p &= uf(x)y_h. \end{aligned}$$

On additionne chacun des deux membres de ces deux expressions:

$$Ly_p = uLy_h + u'y_h = u'y_h = r(x).$$

L'équation différentielle $u'y_h = r$ est séparable:

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

Alors

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

et la solution générale de (3.19) est

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

Dans les deux sections qui suivent, on présente deux méthodes pour trouver une solution particulière, soit la méthode des coefficients indéterminés et la méthode de la variation des paramètres. La première méthode, plus particulière que la seconde, ne requiert pas toujours la solution générale de l'équation homogène, à la différence de la deuxième méthode.

3.4. Méthode des coefficients indéterminés

Considérons l'équation différentielle linéaire nonhomogène d'ordre n,

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1y' + a_0y = r(x), \tag{3.20}$$

à coefficients constants, $a_0, a_1, \ldots, a_{n-1}$.

Si l'espace des dérivées du 2ème membre de (3.20) est de dimension *finie*, on peut employer la méthode des coefficients indéterminés.

Voici des exemples de fonctions $r(x)$ qui admettent un nombre fini de dérivées linéairement indépendantes; de plus, on indique la dimension de l'espace des

dérivées.

$$\begin{aligned}
r(x) &= x^2 + 2x + 1, \qquad r'(x) = 2x + 2, \qquad r''(x) = 2,\\
r^{(k)}(x) &= 0,\ k = 3, 4, \ldots, \qquad \Longrightarrow \text{ dim. } = 3;\\
r(x) &= \cos 2x + \sin 2x, \qquad r'(x) = -2\sin 2x + 2\cos 2x,\\
r''(x) &= -4r(x), \qquad \Longrightarrow \text{ dim. } = 2;\\
r(x) &= x\,e^x, \qquad r'(x) = e^x + x\,e^x,\\
r''(x) &= 2r'(x) - r(x), \qquad \Longrightarrow \text{ dim. } = 2.
\end{aligned}$$

La méthode des coefficients indéterminés consiste à prendre une solution particulière qui est une combinaison linéaire,

$$y_p(x) = c_1 p_1(x) + c_2 p_2(x) + \cdots + c_l p_l(x),$$

des dérivées indépendantes de la fonction $r(x)$ du 2ème membre. On détermine les coefficients c_k en substituant $y_p(x)$ dans (3.20). Si l'on obtient une contradiction, on a fait un mauvais choix ou une erreur.

EXEMPLE 3.9. Touver la solution générale $y_g(x)$ de

$$Ly := y'' + y = 3x^2$$

par la méthode des coefficients indéterminés.

RÉSOLUTION. **(a) Résolution analytique.**— Posons

$$y_p(x) = ax^2 + bx + c$$

dans l'équation différentielle et additionnons chacun des deux membres:

$$\begin{aligned}
y_p &= ax^2 + bx + c\\
y_p'' &= 2a\\
Ly_p &= ax^2 + bx + (2a + c)\\
&= 3x^2.
\end{aligned}$$

En identifiant les coefficients de 1, x et x^2 des deux derniers membres, on obtient:

$$a = 3, \qquad b = 0, \qquad c = -2a = -6.$$

La solution générale de $Ly = 0$ est

$$y_h(x) = A\cos x + B\sin x.$$

Alors la solution générale de $Ly = 3x^2$ est

$$y_g(x) = A\cos x + B\sin x + 3x^2 - 6.$$

(b) Résolution par Matlab symbolique.—

```
dsolve('D2y+y=3*x^2','x')
y = -6+3*x^2+C1*sin(x)+C2*cos(x)
```

□

Remarque importante. Si pour un terme choisi $p_j(x)$, $x^k p_j(x)$ est une solution de l'équation homogène, mais non pas $x^{k+1}p_j(x)$, alors il faut remplacer $p_j(x)$ par $x^{k+1}p_j(x)$.

EXEMPLE 3.10. Trouver la forme d'une solution particulière pour l'équation

$$y'' - 4y' + 4y = 3\,e^{2x} + 32\sin x$$

par coefficients indéterminés.

RÉSOLUTION. Puisque la solution générale de l'équation homogène est

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

une solution particulière est de la forme

$$y_p(x) = ax^2\,e^{2x} + b\cos x + c\sin x. \quad \square$$

EXEMPLE 3.11. Résoudre

$$y''' - y' = 4\,e^{-x} + 3\,e^{2x}, \qquad y(0) = 0, \quad y'(0) = -1, \quad y''(0) = 2$$

et tracer la solution.

RÉSOLUTION. **(a) Résolution analytique.**— L'équation caractéristique est

$$\lambda^3 - \lambda = \lambda(\lambda - 1)(\lambda + 1) = 0.$$

La solution générale de l'équation homogène est

$$y_h(x) = c_1 + c_2\,e^{x} + c_3\,e^{-x}.$$

Puisque e^{-x} apparaît au second membre de l'équation différentielle, on choisit une solution particulière de la forme

$$y_p(x) = ax\,e^{-x} + b\,e^{2x}.$$

Alors,

$$\begin{aligned} y' &= -ax\,e^{-x} + a\,e^{-x} + 2b\,e^{2x}, \\ y'' &= \quad ax\,e^{-x} - 2a\,e^{-x} + 4b\,e^{2x}, \\ y''' &= -ax\,e^{-x} + 3a\,e^{-x} + 8b\,e^{2x}. \end{aligned}$$

Donc,

$$\begin{aligned} y''' - y' &= 2a\,e^{-x} + 6b\,e^{2x} \\ &= 4\,e^{-x} + 3\,e^{2x}, \qquad \text{pour tout} \quad x. \end{aligned}$$

Si l'on identifie les coefficients de e^{-x} et e^{2x}, on a

$$a = 2, \qquad b = \frac{1}{2},$$

ce qui donne la solution particulière de l'équation nonhomogène

$$y_p(x) = 2x\,e^{-x} + \frac{1}{2}\,e^{2x}$$

et la solution générale de l'équation nonhomogène

$$y(x) = c_1 + c_2\,e^{x} + c_3\,e^{-x} + 2x\,e^{-x} + \frac{1}{2}\,e^{2x}.$$

On détermine les constantes arbitraires c_1, c_2 et c_3 au moyen des conditions initiales:

$$\begin{aligned} y(0) &= c_1 + c_2 + c_3 + \frac{1}{2} = \quad 0, \\ y'(0) &= \qquad c_2 - c_3 \; + 3 = -1, \\ y''(0) &= \qquad c_2 + c_3 \; - 2 = \quad 2, \end{aligned}$$

d'où le système algébrique linéaire

$$\begin{aligned} c_1 + c_2 + c_3 &= -\frac{1}{2}, \\ c_2 - c_3 &= -4, \\ c_2 + c_3 &= \quad 4, \end{aligned}$$

qui admet la solution

$$c_1 = -\frac{9}{2}, \qquad c_2 = 0, \qquad c_3 = 4.$$

Donc l'unique solution:

$$y(x) = -\frac{9}{2} + 4\,e^{-x} + 2x\,e^{-x} + \frac{1}{2}\,e^{2x}.$$

(b) Résolution par Matlab symbolique.—

```
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) Résolution par Matlab numérique.— On récrit l'équation différentielle du 3e ordre au moyen des variables

$$\begin{aligned} y(1) &= y, \\ y(2) &= y', \\ y(3) &= y''. \end{aligned}$$

Alors

$$\begin{aligned} y(1)' &= y(2), \\ y(2)' &= y(3), \\ y(3)' &= y(2) + 4 * \exp(-x) + 3 * \exp(2 * x). \end{aligned}$$

Le fichier M `exp39.m`:

```
function yprime = exp39(x,y);
yprime=[y(2); y(3); y(2)+4*exp(-x)+3*exp(2*x)];
```

On appelle le solveur `ode23` et la commande `plot`:

```
xspan = [0 2]; % solution pour 0<=x<=2
y0 = [0;-1;2]; % conditions initiales
[x,y] = ode23('exp39',xspan,y0);
plot(x,y(:,1))
```

Le graphe de la solution numérique est dans la fig. 3.1. □

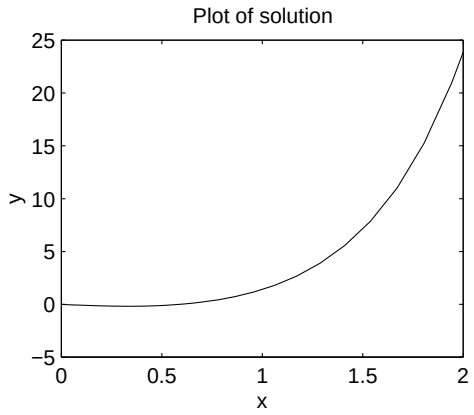

FIGURE 3.1. Graphe de la solution de l'équation linéaire de l'exemple 3.11.

3.5. Solution particulière par variation des paramètres

Considérons l'équation différentielle linéaire *nonhomogène* d'ordre n,

$$Ly := y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = r(x), \tag{3.21}$$

sous forme standard, c'est-à-dire, le coefficient de y^n est égal à 1.

Soit

$$y_h(x) = c_1y_1(x) + c_2y_2(x) + \cdots + c_ny_n(x), \tag{3.22}$$

une solution générale de l'équation homogène

$$Ly = 0.$$

Par simplicité, on dérive la méthode de la variation des paramètres dans le cas $n = 3$; le cas général suit de la même façon.

Suivant une idée de Lagrange, on prend une solution particulière de la forme

$$y_p(x) = c_1(x)y_1(x) + c_2(x)y_2(x) + c_3(x)y_3(x), \tag{3.23}$$

où l'on a fait varier les paramètres c_1, c_2 et c_3 de la solution générale y_h. Ceci nous donne trois degrés de liberté.

On dérive $y_p(x)$:

$$\begin{aligned} y_p'(x) &= \big[c_1'(x)y_1(x) + c_2'(x)y_2(x) + c_3'(x)y_3(x)\big] \\ &\quad + c_1(x)y_1'(x) + c_2(x)y_2'(x) + c_3(x)y_3'(x) \\ &= c_1(x)y_1'(x) + c_2(x)y_2'(x) + c_3(x)y_3'(x). \end{aligned}$$

où l'on a employé un degré de liberté pour supposer que le terme entre crochets est nul:

$$c_1'(x)y_1(x) + c_2'(x)y_2(x) + c_3'(x)y_3(x) = 0. \tag{3.24}$$

On dérive $y_p'(x)$:

$$\begin{aligned} y_p''(x) &= \big[c_1'(x)y_1'(x) + c_2'(x)y_2'(x) + c_3'(x)y_3'(x)\big] \\ &\quad + c_1(x)y_1''(x) + c_2(x)y_2''(x) + c_3(x)y_3''(x) \\ &= c_1(x)y_1''(x) + c_2(x)y_2''(x) + c_3(x)y_3''(x), \end{aligned}$$

où l'on a employé un degré de liberté pour supposer que le terme entre crochets est nul:

$$c_1'(x)y_1'(x) + c_2'(x)y_2'(x) + c_3'(x)y_3'(x) = 0. \tag{3.25}$$

Enfin

$$\begin{aligned} y_p'''(x) = {} & \big[c_1'(x)y_1''(x) + c_2'(x)y_2''(x) + c_3'(x)y_3''(x)\big] \\ & + \big[c_1(x)y_1'''(x) + c_2(x)y_2'''(x) + c_3(x)y_3'''(x)\big]. \end{aligned}$$

Utilisant les expressions obtenues pour y_p, y_p', y_p'' et y_p''', on a

$$\begin{aligned} Ly_p &= c_1'y_1'' + c_2'y_2'' + c_3'y_3'' \\ &\quad + \big[c_1Ly_1 + c_2Ly_2 + c_3Ly_3\big] \\ &= c_1'y_1'' + c_2'y_2'' + c_3'y_3'' \\ &= r(x), \end{aligned}$$

puisque y_1, y_2 et y_3 sont des solutions de $Ly = 0$ et donc le terme entre crochets est nul. De plus, on cherche y_p telle que $Ly_p = r(x)$; ceci utilise le 3e degré de liberté. On a donc

$$c_1'y_1'' + c_2'y_2'' + c_3'y_3'' = r(x). \tag{3.26}$$

On récrit les trois équations (3.24)–(3.26) en $c_1'(x)$, $c_2'(x)$ et $c_3'(x)$ sous forme matricielle

$$\begin{bmatrix} y_1(x) & y_2(x) & y_3(x) \\ y_1'(x) & y_2'(x) & y_3'(x) \\ y_1''(x) & y_2''(x) & y_3''(x) \end{bmatrix} \begin{bmatrix} c_1'(x) \\ c_2'(x) \\ c_3'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ r(x) \end{bmatrix}, \tag{3.27}$$

c'est-à-dire

$$A(x)\boldsymbol{c}'(x) = \begin{bmatrix} 0 \\ 0 \\ r(x) \end{bmatrix}.$$

Puisque y_1, y_2 et y_3 forment un système fondamental, il suit que

$$\det A = W(y_1, y_2, y_3) \neq 0,$$

le wronskien n'étant jamais nul d'après le corollaire 3.1.

On résout le système linéaire pour $\boldsymbol{c}'(x)$ et l'on intègre

$$\boldsymbol{c}(x) = \int \boldsymbol{c}'(x)\,dx.$$

Aucune constante d'intégration n'est requise ici puisque la solution générale en contiendra trois. La solution générale de (3.21) est donc

$$y_g(x) = Ay_1 + By_2 + Cy_3 + c_1(x)y_1 + c_2(x)y_2 + c_3(x)y_3. \tag{3.28}$$

Remarque 3.3. Si le coefficient $a_n(x)$ de $y^{(n)}$ n'est pas égal à 1, il faut diviser le 2ème membre de (3.27) par $a_n(x)$, c'est-à-dire remplacer $r(x)$ par $r(x)/a_n(x)$.

Exemple 3.12. Trouver la solution générale de l'équation différentielle

$$(D^2 + 1)y = \sec x \tan x,$$

par variation des paramètres.

RÉSOLUTION. **(a) Résolution analytique.**— Puisque le 2ème membre n'admet pas un nombre fini de dérivées dépendantes, la méthode des coefficients indéterminés ne fonctionne pas.

On sait que la solution générale de l'équation homogène est

$$y_h(x) = c_1 \cos x + c_2 \sin x.$$

On cherche une solution particulière de l'équation nonhomogène par variation des paramètres:

$$y_p(x) = c_1(x) \cos x + c_2(x) \sin x.$$

On a donc

$$\begin{bmatrix} \cos x & \sin x \\ -\sin x & \cos x \end{bmatrix} \begin{bmatrix} c_1'(x) \\ c_2'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \sec x \tan x \end{bmatrix},$$

c'est-à-dire

$$Q(x)\boldsymbol{c}'(x) = \begin{bmatrix} 0 \\ \sec x \tan x \end{bmatrix}. \tag{3.29}$$

Puisque la matrice Q est orthogonale, c'est-à-dire

$$QQ^T = \begin{bmatrix} \cos x & \sin x \\ -\sin x & \cos x \end{bmatrix} \begin{bmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

il suit que l'inverse Q^{-1} de Q est la transposée Q^T de Q:

$$Q^{-1} = Q^T.$$

On obtient $\boldsymbol{c}'$ en multipliant (3.29) à gauche par Q^T:

$$\boldsymbol{c}' = Q^T Q \boldsymbol{c}' = Q^T \begin{bmatrix} 0 \\ \sec x \tan x \end{bmatrix},$$

c'est-à-dire

$$\begin{bmatrix} c_1'(x) \\ c_2'(x) \end{bmatrix} = \begin{bmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{bmatrix} \begin{bmatrix} 0 \\ \sec x \tan x \end{bmatrix}.$$

Alors

$$c_1' = -\frac{\sin x}{\cos x} \tan x = -\tan^2 x,$$
$$c_2' = \frac{\cos x}{\cos x} \tan x = \tan x = \frac{\sin x}{\cos x},$$

qu'on intègre:

$$c_1 = -\int (\sec^2 x - 1)\, dx = x - \tan x,$$
$$c_2 = -\ln|\cos x| = \ln|\sec x|.$$

La solution particulière est donc

$$y_p(x) = (x - \tan x)\cos x + (\ln|\sec x|)\sin x$$

et la solution générale est

$$\begin{aligned} y(x) &= y_h(x) + y_p(x) \\ &= A\cos x + B\sin x + (x - \tan x)\cos x + (\ln|\sec x|)\sin x. \end{aligned}$$

(b) Résolution par Matlab symbolique.—

```
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)
```

□

EXEMPLE 3.13. Trouver la solution générale de l'équation différentielle

$$y''' - y' = \cosh x,$$

par variation des paramètres.

RÉSOLUTION. L'équation caractéristique:

$$\lambda^3 - \lambda = \lambda(\lambda^2 - 1) = 0 \implies \lambda_1 = 0,\ \lambda_2 = 1,\ \lambda_3 = -1.$$

La solution générale de l'équation homogène est

$$y_h(x) = c_1 + c_2\, e^x + c_3\, e^{-x}.$$

Par variation des paramètres, la solution particulière de l'équation nonhomogène est

$$y_p(x) = c_1(x) + c_2(x)\, e^x + c_3(x)\, e^{-x}.$$

On a donc le système

$$\begin{bmatrix} 1 & e^x & e^{-x} \\ 0 & e^x & -e^{-x} \\ 0 & e^x & e^{-x} \end{bmatrix} \begin{bmatrix} c_1'(x) \\ c_2'(x) \\ c_3'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cosh x \end{bmatrix}.$$

On résout ce système par élimination gaussienne:

$$\begin{bmatrix} 1 & e^x & e^{-x} \\ 0 & e^x & -e^{-x} \\ 0 & 0 & 2e^{-x} \end{bmatrix} \begin{bmatrix} c_1'(x) \\ c_2'(x) \\ c_3'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cosh x \end{bmatrix}.$$

Alors

$$\begin{aligned} c_3' &= \frac{1}{2}\, e^x \cosh x = \frac{1}{2}\, e^x \left(\frac{e^x + e^{-x}}{2} \right) = \frac{1}{4} \left(e^{2x} + 1 \right), \\ c_2' &= e^{-2x} c_3' = \frac{1}{4} \left(1 + e^{-2x} \right), \\ c_1' &= -e^x c_2' - e^{-x} c_3' = -\frac{1}{2} \left(e^x + e^{-x} \right) = -\cosh x. \end{aligned}$$

On intègre:

$$\begin{aligned} c_1 &= -\sinh x \\ c_2 &= \frac{1}{4} \left(x - \frac{1}{2}\, e^{-2x} \right) \\ c_3 &= \frac{1}{4} \left(\frac{1}{2}\, e^{2x} + x \right). \end{aligned}$$

La solution particulière est

$$\begin{aligned} y_p(x) &= -\sinh x + \frac{1}{4} \left(x\, e^x - \frac{1}{2}\, e^{-x} \right) + \frac{1}{4} \left(\frac{1}{2}\, e^x + x\, e^{-x} \right) \\ &= -\sinh x + \frac{1}{4} x \left(e^x + e^{-x} \right) + \frac{1}{8} \left(e^x - e^{-x} \right) \\ &= \frac{1}{2} x \cosh x - \frac{3}{4} \sinh x. \end{aligned}$$

La solution générale de l'équation nonhomogène est donc

$$\begin{aligned} y_g(x) &= A + B'\,e^x + C'\,e^{-x} + \frac{1}{2}\,x\cosh x - \frac{3}{4}\sinh x \\ &= A + B\,e^x + C\,e^{-x} + \frac{1}{2}\,x\cosh x, \end{aligned}$$

puisque la fonction

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

est déjà contenue dans la solution générale y_h de l'équation homogène. Matlab symbolique ne produit pas une solution générale aussi simple. □

Si l'on emploie la méthode des coefficients indéterminés pour résoudre ce problème, il faut prendre une solution particulière de la forme

$$y_p(x) = ax\cosh x + bx\sinh x,$$

puisque $\cosh x$ et $\sinh x$ sont des combinaisons linéaires de e^x et e^{-x} qui sont des solutions de l'équation homogène. En effet, si l'on pose

$$y_p(x) = ax\cosh x + bx\sinh x$$

dans l'équation $y''' - y' = \cosh x$, on obtient

$$\begin{aligned} y_p''' - y_p' &= 2a\cosh x + 2b\sinh x \\ &= \cosh x, \end{aligned}$$

d'où

$$a = \frac{1}{2} \quad \text{et} \quad b = 0.$$

EXEMPLE 3.14. Trouver la solution générale de l'équation différentielle

$$Ly := y'' + 3y' + 2y = \frac{1}{1+e^x}.$$

RÉSOLUTION. Puisque la dimension de l'espace des dérivées du 2e membre est infinie, on emploie la méthode de la variation des paramètres.

On remarque que la commande `dsolve` de Matlab symbolique produit une lonque solution inutilisable. On suit donc la méthode théorique de Lagrange, mais on exécute les simples manipulations d'algèbre et de calcul infinitésimal au moyen de Matlab.

Le polynôme caractéristique de l'équation homogène $Ly = 0$ est

$$\lambda^2 + 3\lambda + 2 = (\lambda+1)(\lambda+2) = 0 \implies \lambda_1 = -1, \quad \lambda_2 = -2.$$

Donc, la solution générale $y_h(x)$ de $Ly = 0$ est

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

Par variation des paramètres, on cherche une solution particulière de l'équation nonhomogène de la forme

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

Les fonctions $c_1(x)$ et $c_2(x)$ sont les intégrales des solutions $c_1'(x)$ et $c_2'(x)$ du système algébrique $A\boldsymbol{c'} = \boldsymbol{b}$,

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

On utilise Matlab symbolique pour résoudre ce petit système.

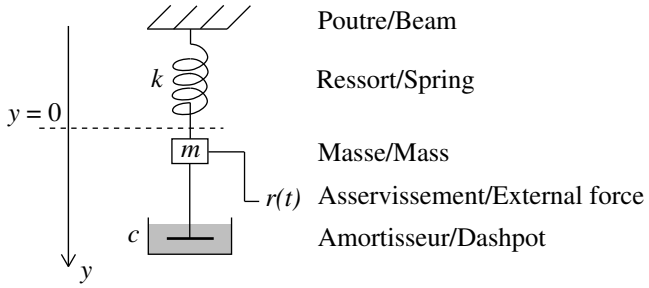

FIGURE 3.2. Système amorti et asservi.

```
>> 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) % c(x) est l'integrale de 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)
```

Puisque $-e^{-x}$ est contenu dans $y_h(x)$, la solution génnérale de l'équation nonhomogène est

$$y(x) = A\,e^{-x} + B\,e^{-2x} + [\ln(1+e^x)]\;e^{-x} + [\ln(1+e^x)]\;e^{-2x}. \quad \square$$

3.6. Systèmes asservis

On présente deux exemples de systèmes mécaniques asservis.

Soit une ressort en position verticale pendante fixé à une poutre rigide. Le ressort résiste à l'extension et à la compression et sa constante de Hooke est k. On étudie le mouvement vertical amorti et asservi d'une masse de m kg fixée au bout inférieur du ressort (V. figure 3.2). La constante d'amortissement est c et la force d'asservissement est $r(t)$.

On se réfère à l'exemple 2.4 pour la dérivation de l'équation différentielle qui gouverne le système non asservi, auquel on ajoute au 2ème membre la fonction d'asservissement:

$$y'' + \frac{c}{m}\,y' + \frac{k}{m}\,y = \frac{1}{m}\,r(t).$$

EXEMPLE 3.15 (Oscillation asservie sans résonnance). Touver la solution du problème asservi

$$Ly := y'' + 9y = 8\sin t, \qquad y(0) = 1, \qquad y'(0) = 1,$$

et tracer la solution sur $[0, 7]$.

RÉSOLUTION. **(a) Résolution analytique.**— La solution générale de $Ly = 0$ est

$$y_h(t) = A\cos 3t + B\sin 3t.$$

Par la méthode des coefficients indéterminés, on prend

$$y_p(t) = a\cos t + b\sin t,$$

qu'on substitue dans $Ly = 8\sin t$. On obtient

$$\begin{aligned} y_p'' + 9y_p &= (-a + 9a)\cos t + (-b + 9b)\sin t \\ &= 8\sin t. \end{aligned}$$

En identifiant les coefficients des deux derniers membres, on a

$$a = 0, \qquad b = 1.$$

La solution générale de $Ly = 8\sin t$ est

$$y(t) = A\cos 3t + B\sin 3t + \sin t.$$

On détermine A et B au moyen des conditions initiales:

$$\begin{aligned} y(0) &= A = 1, \\ y'(t) &= -3A\sin 3t + 3B\cos 3t + \cos t, \\ y'(0) &= 3B + 1 = 1 \implies B = 0. \end{aligned}$$

La solution, qui est unique, est donc

$$y(t) = \cos 3t + \sin t.$$

(b) Résolution par Matlab symbolique.—

```
dsolve('D2y+9*y=8*sin(t)','y(0)=1','Dy(0)=1','t')
y = sin(t)+cos(3*t)
```

(c) Résolution par Matlab numérique.— Pour récrire l'équation différentielle du 2e ordre en système du 1er ordre on utilise les variables

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

Alors,

$$\begin{aligned} y_1' &= y_2, \\ y_2' &= -9y_1 + 8\sin t. \end{aligned}$$

Le fichier M `exp312.m`:

```
function yprime = exp312(t,y);
yprime = [y(2); -9*y(1)+8*sin(t)];
```

On appelle le solveur `ode23` et la commande `plot`:

```
tspan = [0 7]; % solution sur 0<=t<=7
y0 = [1; 1]; % conditions initiales
[x,y] = ode23('exp312',tspan,y0);
plot(x,y(:,1))
```

Le graphe de la solution numérique est dans la fig. 3.3. □

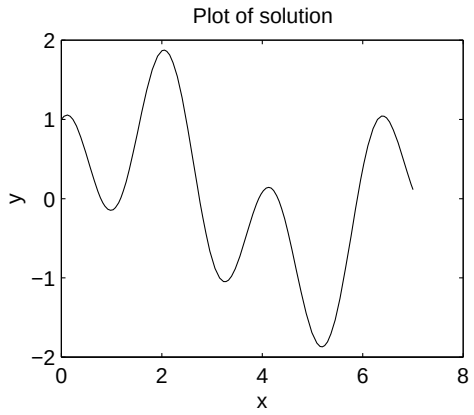

FIGURE 3.3. Graphe de la solution de l'équation linéaire de l'exemple 3.15.

EXEMPLE 3.16 (Oscillation asservie avec résonnance). Touver la solution du problème asservi:

$$Ly := y'' + 9y = 6\sin 3t, \qquad y(0) = 1, \qquad y'(0) = 2,$$

et tracer la solution sur $[0, 7]$.

RÉSOLUTION. **(a) Résolution analytique.**— La solution générale de $Ly = 0$ est

$$y_h(t) = A\cos 3t + B\sin 3t.$$

Puisque le 2ème membre de $Ly = 6\sin 3t$ est contenu dans y_h, pour la méthode des coefficients indéterminés, on prend

$$y_p(t) = at\cos 3t + bt\sin 3t.$$

On obtient alors

$$\begin{aligned} y_p'' + 9y_p &= -6a\sin 3t + 6b\cos 3t \\ &= 6\sin 3t. \end{aligned}$$

En identifiant les coefficients des deux derniers membres, on a

$$a = -1, \qquad b = 0.$$

La solution générale de $Ly = 6\sin 3t$ est

$$y(t) = A\cos 3t + B\sin 3t - t\cos 3t.$$

On détermine A et B au moyen des conditions initiales:

$$\begin{aligned} y(0) &= A = 1, \\ y'(t) &= -3A\sin 3t + 3B\cos 3t - \cos 3t + 3t\sin 3t, \\ y'(0) &= 3B - 1 = 2 \implies B = 1. \end{aligned}$$

La solution, qui est unique, est donc

$$y(t) = \cos 3t + \sin 3t - t\cos 3t.$$

Le terme $-t\cos 3t$, dont l'amplitude croît, provient de la résonnance du système parce que la fréquence de la force d'asservissement et celle du système coïncident.

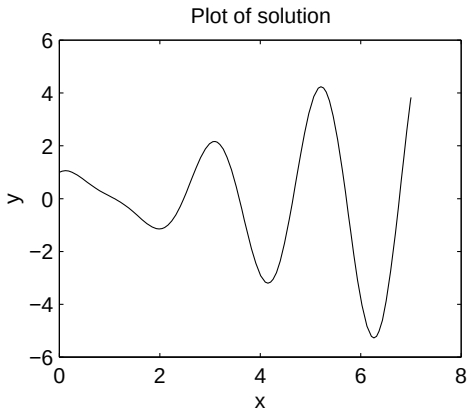

FIGURE 3.4. Graphe de la solution de l'équation linéaire de l'exemple 3.16.

(b) Résolution par Matlab symbolique.—

```
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) Résolution par Matlab numérique.— Pour récrire l'équation différentielle du 2e ordre en système du 1er ordre on utilise les variables

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

Alors,

$$y_1' = y_2,$$
$$y_2' = -9y_1 + 6\sin 3t.$$

Le fichier M `exp313.m`:

```
function yprime = exp313(t,y);
yprime = [y(2); -9*y(1)+6*sin(3*t)];
```

On appelle le solveur `ode23` et la commande `plot`:

```
tspan = [0 7]; % solution sur 0<=t<=7
y0 = [1; 1]; % conditions initiales
[x,y] = ode23('exp313',tspan,y0);
plot(x,y(:,1))
```

Le graphe de la solution numérique est dans la fig. 3.4. □

CHAPITRE 4

Systèmes d'équations différentielles linéaires

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, $r(x)$, become

$$\begin{bmatrix} u_1(x_0) \\ u_2(x_0) \\ \vdots \\ u_{n-1}(x_0) \\ u_n(x_0) \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_{n-1} \\ k_n \end{bmatrix}, \qquad \boldsymbol{g}(x) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ r(x) \end{bmatrix},$$

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 the *companion matrix*.

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.

EXEMPLE 4.1. 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

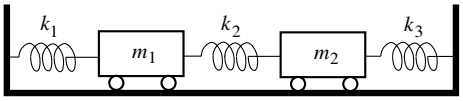

FIGURE 4.1. Mechanical system for Example 4.1.

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:

$$m_1x_1'' = -k_1x_1 + k_2(x_2 - x_1), \qquad m_2x_2'' = -k_2(x_2 - x_1) - k_3x_2. \tag{4.1}$$

We convert each equation in (4.1) to a first-order system of equations by introducing two new variables y_1 and y_2 representing the velocities of each mass:

$$y_1 = x_1', \qquad y_2 = x_2'. \tag{4.2}$$

Using these new dependent variables, we rewrite (4.1) as the following four simultaneous equations in the four unknowns x_1, y_1, x_2 and y_2:

$$\begin{aligned} x_1' &= y_1, \\ y_1' &= \frac{-k_1x_1 + k_2(x_2 - x_1)}{m_1}, \\ x_2' &= y_2, \\ y_2' &= \frac{-k_2(x_2 - x_1) - k_3x_2}{m_2}, \end{aligned} \tag{4.3}$$

which, in matrix form, become

$$\begin{bmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{k_1+k_2}{m_1} & 0 & \frac{k_2}{m_1} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{k_2}{m_2} & 0 & -\frac{k_2+k_3}{m_2} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix}. \tag{4.4}$$

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

$$\boldsymbol{u}' = A\boldsymbol{u}, \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0. \tag{4.5}$$

It is to be noted that the matrix A is not in the form of a companion matrix. □

4.2. Théorème d'existence et d'unicité

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

$$\boldsymbol{y}' = \boldsymbol{f}(x, \boldsymbol{y}), \qquad \boldsymbol{y}(x_0) = \boldsymbol{y}_0, \tag{4.6}$$

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 system of the form

$$\boldsymbol{y}' = A(x)\boldsymbol{y} + \boldsymbol{g}(x), \qquad \boldsymbol{y}(x_0) = \boldsymbol{y}_0, \tag{4.7}$$

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. Système fondamental

It is readily seen that the solutions to the linear homogeneous system

$$\boldsymbol{y}' = A(x)\boldsymbol{y}, \qquad x \in]a, b[, \tag{4.8}$$

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

$$c_1\boldsymbol{y}_1(x) + c_2\boldsymbol{y}_2(x) + \cdots + c_m\boldsymbol{y}_m(x) = \boldsymbol{0}, \qquad \text{for all} \quad x \in]a, b[,$$

implies that

$$c_1 = c_2 = \cdots = c_m = 0.$$

Otherwise, this set of functions is said to be *linearly dependent*.

For general system, 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,

$$\operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn}.$$

LEMME 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:*

$$W(x) = W(x_0)\, e^{-\int_{x_0}^{x} tr\, A(t)\, dt}, \qquad x_0 \in]a, b[. \tag{4.9}$$

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.

COROLLAIRE 4.1. *If n solutions to the homogeneous differential system* (4.8) *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.*

REMARQUE 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 (4.8).

DÉFINITION 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 solution matrix*.

LEMME 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, $Z(x_0) = I$. □

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.

THÉORÈME 4.1. *Let $Y(x)$ be a fundamental solution 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 rely on the uniqueness theorem. To prove the first part, let $Y(x)$ be a solution 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 (4.7) in terms of a fundamental solution.

LEMME 4.3. *Let $Y(x)$ be a fundamental matrix for the system* (4.8)*. Then, $(Y^T)^{-1}(x)$ is a fundamental solution for the adjoint system*

$$\boldsymbol{y}' = -A^T(x)\boldsymbol{y}. \tag{4.10}$$

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 (4.8), 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 (4.10). □

THÉORÈME 4.2 (Solution formula). *Let $Y(x)$ be a fundamental solution matrix of the homogeneous linear system* (4.8) *Then the unique solution to the initial value problem* (4.7) *is*

$$\boldsymbol{y}(x) = Y(x)Y^{-1}(x_0)\boldsymbol{y}_0 + Y(x)\int_{x_0}^{x} Y^{-1}(t)\boldsymbol{g}(t)\,dt. \tag{4.11}$$

PROOF. Multiply both sides of (4.7) 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. Systèmes linéaires à coefficients constants

A solution of a linear system with constant coefficients can be expressed in terms of the eigenvalues and eigenvectors of the coefficient matrix A. Given a linear homogeneous system of the form

$$\boldsymbol{y}' = A\boldsymbol{y}, \tag{4.12}$$

where the $n \times n$ matrix A has real constant entries, we seek solutions of the form

$$\boldsymbol{y}(x) = e^{\lambda x}\boldsymbol{v}, \tag{4.13}$$

where the number λ and the vector $\boldsymbol{v}$ are to be determined. Substituting (4.13) into (4.12) and dividing by $e^{\lambda x}$ we obtain the eigenvalue problem

$$(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}, \qquad \boldsymbol{v} \neq \boldsymbol{0}. \tag{4.14}$$

This equation has a nonzero solution $\boldsymbol{v}$ if and only if

$$\det(A - \lambda I) = 0.$$

The left-hand side of this determinental equation is a polynomial of degree n and the n roots of this equation are called the *eigenvalues* of the matrix A. The corresponding nonzero vectors $\boldsymbol{v}$ 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.

EXEMPLE 4.2. 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

$$(A-I)\boldsymbol{u}=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}u_1\\ u_2\end{bmatrix}=0.$$

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 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 $\boldsymbol{y}' = Y\boldsymbol{x}'$. Hence the given system $\boldsymbol{y}' = A\boldsymbol{y}$ becomes

$$\boldsymbol{x}' = D\boldsymbol{x}.$$

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.

EXEMPLE 4.3. Solve the system of Example 4.1 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,:),'--');
```

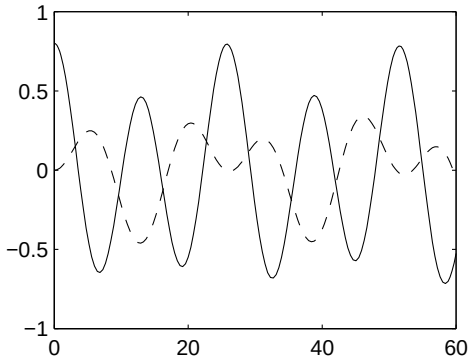

FIGURE 4.2. Graph of solution $x_1(t)$ (solid line) and $x_2(t)$ (dashed line) to Example 4.3.

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));
```

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.

DÉFINITION 4.2. Let A be an $n \times n$ matrix. We say that λ is a *deficient eigenvalue* of A if it has 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

$$(A - \lambda I)^s \boldsymbol{u} = \mathbf{0},$$

but

$$(A - \lambda I)^{s-1} \boldsymbol{u} \neq \mathbf{0}.$$

In general, given a matrix A with an eigenvalue λ that has a degree of deficiency r, we construct a set of generalized eigenvectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ such that

$$(A - \lambda I)\boldsymbol{u}_2 = \boldsymbol{u}_1, \quad (A - \lambda I)\boldsymbol{u}_3 = \boldsymbol{u}_2, \quad \ldots, \quad (A - \lambda I)\boldsymbol{u}_r = \boldsymbol{u}_{r-1},$$

and $\boldsymbol{u}_1$ is an eigenvector of A associated with λ. The set of generalized eigenvectors, in turn, generates the following set of linearly independent solutions of (4.12):

$$\boldsymbol{y}_1(x) = e^{\lambda x}\boldsymbol{u}_1, \quad \boldsymbol{y}_2(x) = e^{\lambda x}(x\boldsymbol{u}_1+\boldsymbol{u}_2), \quad \boldsymbol{y}_3(x) = e^{\lambda x}\left(\frac{x^2}{2}\boldsymbol{u}_1 + x\boldsymbol{u}_2 + \boldsymbol{u}_3\right), \quad \ldots.$$

It is a result of linear algebra that any $n \times n$ matrix had n linearly independent generalized eigenvectors.

EXEMPLE 4.4. 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 with a single eigenvector $\boldsymbol{e}_1$:

$$A - I \sim \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

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

$$(A - I)\boldsymbol{e}_2 = \boldsymbol{e}_1.$$

Thus,

$$\boldsymbol{e}_2 = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}$$

and

$$\boldsymbol{y}_2(x) = (x\boldsymbol{e}_1 + \boldsymbol{e}_2)\, e^x = \left(x \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} \right) e^x$$

is a second linearly independent solution.

To construct a second generalized eigenvector, we solve the equation

$$(A - I)\boldsymbol{e}_3 = \boldsymbol{e}_2.$$

Thus,

$$\boldsymbol{e}_3 = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$$

and

$$\boldsymbol{y}_3(x) = \left(\frac{x^2}{2}\boldsymbol{e}_1 + x\boldsymbol{e}_2 + \boldsymbol{e}_2\right) e^x = \left(\frac{x^2}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} \right) e^x$$

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.

EXEMPLE 4.5. 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 with two independent eigenvectors $\boldsymbol{e}_1$:

$$A - 3I \sim \begin{bmatrix} -2 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{e}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

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

$$(A - 3I)\boldsymbol{e}_3 = \alpha\boldsymbol{e}_1 + \beta\boldsymbol{e}_2,$$

where the parameters α and β are to be chosen so that the right-hand side,

$$\boldsymbol{e}_4 = \alpha\boldsymbol{e}_1 + \beta\boldsymbol{e}_2,$$

is in the space spanned by the columns of the matrix $(A - 3I)$. We find that that

$$\alpha + 2\beta = 0 \implies \alpha = 2, \quad \beta = -1.$$

Thus,

$$\boldsymbol{e}_4 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \qquad \boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and

$$\boldsymbol{y}_3(x) = (x\boldsymbol{e}_4 + \boldsymbol{e}_3)\, e^{3x} = \left(x \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) e^{x}$$

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. Systèmes linéaires nonhomogènes

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

$$\boldsymbol{y}' = A\boldsymbol{y} + \boldsymbol{f}(x). \tag{4.15}$$

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. Méthode des coefficients indéterminés. The method of undetermined coefficients can be used when the matrix A in (4.15) is constant and the dimension of the vector space spanned by the derivatives of the right-hand side $\boldsymbol{f}(x)$ of (4.15) 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 derivatives.

Exemple 4.6. Find the general solution of the nonhomogeneous linear system

$$\boldsymbol{y}' = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \boldsymbol{y} + \begin{bmatrix} 4\,e^{-3x} \\ e^{-2x} \end{bmatrix}.$$

Solution. The eigenvalues of the matrix A of the system are

$$\lambda_1 = i, \qquad \lambda_2 = -i,$$

and the corresponding eigenvectors are

$$\boldsymbol{u}_1 = \begin{bmatrix} -i \\ 1 \end{bmatrix}, \qquad \boldsymbol{u}_2 = \begin{bmatrix} i \\ 1 \end{bmatrix}.$$

Hence the general solution of the homogeneous system is

$$\boldsymbol{y}_h(x) = k_1\, e^{ix} \begin{bmatrix} -i \\ 1 \end{bmatrix} + k_2\, e^{-ix} \begin{bmatrix} i \\ 1 \end{bmatrix},$$

where k_1 and k_2 are complex constants. Since the real and imaginary parts of a solution of a real homogeneous linear equation are solutions, setting

$$c_1 = k_1 + k_2, \qquad c_2 = -i(k_1 - k_2),$$

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,

$$\boldsymbol{a} = -(A + 3I)^{-1}(4\boldsymbol{e}_1) = -\frac{1}{5}\begin{bmatrix} 6 \\ 2 \end{bmatrix}$$

$$\boldsymbol{b} = -(A + 2I)^{-1}(\boldsymbol{e}_2) = \frac{1}{5}\begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Finally,

$$\boldsymbol{y}_h(x) = -\begin{bmatrix} \frac{6}{5}\, e^{-3x} - \frac{1}{5}\, e^{-2x} \\ \frac{2}{5}\, e^{-3x} + \frac{2}{5}\, e^{-2x} \end{bmatrix}.$$

The general solution is

$$\boldsymbol{y}(x) = \boldsymbol{y}_h(x) + \boldsymbol{y}_p(x).$$

□

4.5.2. Méthode de la variation des paramètres. The method of variation of parameters can be applied, at least theoretically, to nonhomogeneous systems with nonconstant matrix $A(x)$ and general right-hand side $\boldsymbol{f}(x)$. 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 parameter 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 4.6 by the method of variation of parameters.

CHAPITRE 5

Résolution numérique d'équations différentielles

5.1. Le problème à valeur initiale

Soit le problème à valeur initiale du 1er ordre:

$$y' = f(x,y), \qquad y(x_0) = y_0. \tag{5.1}$$

On suppose que, pour tout $x \in [a,b]$, la fonction $f(x,y)$ du second membre est continue en x et y et lipschitzienne en y,

$$|f(x,z) - f(x,y)| \le M|z-y|, \qquad \text{pour tout} \quad x \in [a,b]. \tag{5.2}$$

Il suit par le théorème 1.3 que le problème (5.1) admet une et une seule solution sur $[a,b]$ si $x_0 \in [a,b]$.

On considère quelques méthodes numériques pour la résolution de (5.1).

On note

- $h > 0$ le *pas* d'intégration
- $x_n = x_0 + nh$ le n-ième nœud
- $y(x_n)$ la *solution exacte* en x_n
- y_n la *solution numérique* en x_n
- $f_n = f(x_n, y_n)$ la valeur numérique de f en (x_n, y_n)

On dit qu'une fonction $g(x)$ est d'*ordre* p quand $x \to x_0$, et l'on écrit $g = O(|x-x_0|^p)$, si

$$|g(x)| < M|x-x_0|^p, \qquad M \quad \text{une constante,}$$

pout tout x près de x_0.

Définition 5.1. On dit qu'une méthode est d'ordre p si l'erreur locale en y_{n+1}^{local} après un pas du problème local

$$y' = f(x,y), \qquad y(x_n) = y_n$$

est d'ordre $p+1$, c'est-à-dire

$$\left|y_{n+1}^{\text{local}} - y^{\text{local}}(x_{n+1})\right| = O(h^{p+1}).$$

Cette notation s'explique par le fait que si l'erreur locale est d'ordre $p+1$, alors l'erreur globale sera d'ordre p. En effet, si N est le nombre de pas d'intégration de x_0 à x_N et $x_N - x_0 = A = \text{const}$, alors l'erreur globale est d'ordre p:

$$Nh^{p+1} = Nh\,h^p = A\,h^p = O(h^p).$$

On dit qu'une méthode aux différences finies pour résoudre (5.1) est *absolument stable* si elle n'amplifie pas les erreurs de calculs causées aux pas précédents. L'*intervalle de stabilité absolue* $(\alpha, 0)$ d'une méthode restreint le pas h tel que

$$\alpha < h\frac{\partial f}{\partial y} < 0,$$

pour assurer la stabilité absolue de la méthode.

5.2. Méthodes explicites à un pas

5.2.1. Méthode d'Euler, d'ordre 1. La méthode la plus simple est celle d'Euler:

$$y_{n+1} = y_n + hf(x_n, y_n). \tag{5.3}$$

La méthode d'Euler est une méthode explicite à un pas d'ordre 1.

5.2.2. Méthode d'Euler améliorée, d'ordre 2. La méthode d'Euler améliorée se compose d'un prédicteur et d'un correcteur:

$$y^{\mathrm{P}}_{n+1} = y^{\mathrm{C}}_n + hf(x_n, y^{\mathrm{C}}_n). \tag{5.4}$$

$$y^{\mathrm{C}}_{n+1} = y^{\mathrm{C}}_n + \frac{1}{2}\, h \left[f(x_n, y^{\mathrm{C}}_n) + f(x_{n+1}, y^{\mathrm{P}}_{n+1}) \right]. \tag{5.5}$$

Cette méthode est d'ordre 2.

Elle s'écrit aussi sous la forme d'une méthode de Runge-Kutta explicite à deux étapes:

$$\begin{aligned} k_1 &= hf(x_n, y_n), \\ k_2 &= hf\left(x_n + h, y_n + k_1\right), \end{aligned}$$

et

$$y_{n+1} = y_n + \frac{1}{2}(k_1 + k_2).$$

On l'appelle alors méthode de Heun d'ordre 2. L'intervalle de stabilité absolue des méthodes de Runge–Kutta explicites à deux étapes est $(-2, 0)$.

5.2.3. Méthodes de Runge–Kutta. La méthode de Runge–Kutta d'ordre 4, dite classique, est une méthode explicite à quatre étapes:

$$\begin{aligned} k_1 &= hf(x_n, y_n), \\ k_2 &= hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right), \\ k_3 &= hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_2\right), \\ k_4 &= hf\left(x_n + h, y_n + k_3\right), \end{aligned}$$

et

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4).$$

On représente la méthode de Runge–Kutta classique dans un tableau de Butcher. Le vecteur c représente l'incrément de x_n, la matrice A les coefficients des accroissements de y_n et le vecteur b les coefficients des k_i.

$$\begin{array}{c|cccc} c & A & & & \\ \hline 0 & 0 & & & \\ 1/2 & 1/2 & 0 & & \\ 1/2 & 0 & 1/2 & 0 & \\ 1 & 0 & 0 & 1 & 0 \\ \hline b^T & 1/6 & 2/6 & 2/6 & 1/6 \end{array} \tag{5.6}$$

Runge–Kutta classique (1/3) d'ordre 4.

L'intervalle de stabilité absolue des méthodes de Runge–Kutta explicite à quatre étapes est $(-2.78, 0)$.

EXEMPLE 5.1. Soit l'équation différentielle

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

Calculer y_4 au moyen de la méthode de Runge–Kutta d'ordre 4 avec $h = 0.1$.

RÉSOLUTION. On présente la solution sous forme de tableau.

n	x_n	y_n	Exacte $y(t_n)$	Erreur $y(t_n) - y_n$
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

□

EXEMPLE 5.2. Employer la méthode de Runge–Kutta d'ordre 4 avec $h = 0.01$ pour résoudre, au dix-milionnième près, le problème à valeur initiale

$$y' = x + \arctan y, \quad y(0) = 0,$$

sur $0 \leq x \leq 1$. Imprimer chaque dixième valeur et tracer la solution numérique.

RÉSOLUTION. **Résolution par Matlab numérique.**— Le fichier M `exp4_2` pour l'exemple 5.2 est

```
function yprime = exp4_2(x,y); %MAT 2731, Exemple 5.2.
yprime = x+atan(y);
```

On applique la méthode de Runge–Kutta d'ordre 4 à l'équation différentielle:

```
clear
h = 0.01; x0= 0; xf= 1; y0 = 0;
n = ceil((xf-x0)/h); % nombre de pas
%
count = 2; every_so_often = 10; % quand imprimer les resultats
x = x0; y = y0; % initialiser x et y
output = [0 x0 y0];
for i=1:n
 k1 = h*exp4_2(x,y);
 k2 = h*exp4_2(x+h/2,y+k1/2);
 k3 = h*exp4_2(x+h/2,y+k2/2);
 k4 = h*exp4_2(x+h,y+k3);
 z  = y + (1/6)*(k1+2*k2+2*k3+k4);
 x = x + h;
   if count > every_so_often
      output = [output; i x z];
      count = count - every_so_often;
```

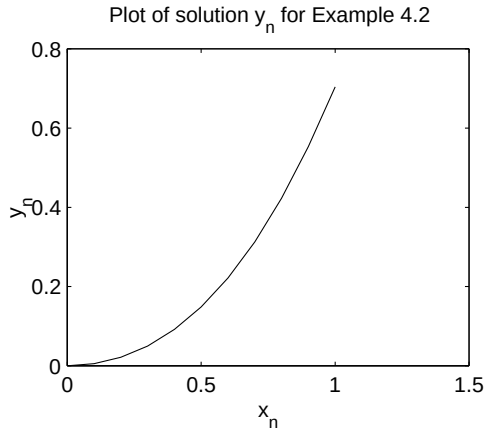

FIGURE 5.1. Graphe de la solution pour l'exemple 5.2.

```
   end
 y = z;
count = count + 1;
end
output
save output % pour imprimer le graphe
```

La commande `output` imprime les valeurs de n, x et 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
```

La commande suivante imprime les résultats.

```
load output;
subplot(2,2,1); plot(output(:,2),output(:,3));
title('Graphe de la solution y_n pour l'example 5.2');
xlabel('x_n'); ylabel('y_n');
```

□

5.2.4. La paire RKF(4,5) à 6 étapes avec estimation de l'erreur locale. La paire de méthodes de Runge–Kutta–Fehlberg avec contrôle de l'erreur locale utilise une méthode d'ordre 4 pour obtenir la valeur numérique y_{n+1} et une méthode d'ordre 5 pour obtenir la valeur auxiliaire $\hat{y}_{n+1}$ pour calculer l'erreur locale au moyen de la différence $y_{n+1} - \hat{y}_{n+1}$. On la présente sous forme de tableau de Butcher. On obtient une estimation de l'erreur au moyen de la dernière ligne. La méthode d'ordre 4 minimise l'erreur locale.

(5.7)

	c	A					
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
y_{n+1}	b^T	$\frac{25}{216}$	0	$\frac{1408}{2565}$	$\frac{2197}{4104}$	$-\frac{1}{5}$	0
$\hat{y}_{n+1}$	$\hat{b}^T$	$\frac{16}{135}$	0	$\frac{6656}{12825}$	$\frac{28561}{56430}$	$-\frac{9}{50}$	$\frac{2}{55}$
$\hat{b}^T - b^T$		$\frac{1}{360}$	0	$-\frac{128}{4275}$	$-\frac{2197}{75240}$	$\frac{1}{50}$	$\frac{2}{55}$

Paire de Runge–Kutta–Fehlberg d'ordre 4 et 5 à 6 étapes

L'intervalle de stabilité absolue de la paire RKF(4,5) est approximativement $(-3.78, 0)$.

5.2.5. La paire DP(5,4)7M à 7 étapes avec estimation de l'erreur locale. La paire de méthodes de Dormand–Prince avec contrôle de l'erreur locale utilise une méthode d'ordre 5 pour obtenir la valeur numérique y_{n+1} et une méthode d'ordre 4 pour obtenir la valeur auxiliaire $\hat{y}_{n+1}$ pour calculer l'erreur locale au moyen de la différence $y_{n+1} - \hat{y}_{n+1}$. On la présente sous forme de tableau de Butcher. On obtient une estimation de l'erreur au moyen de la dernière ligne. La méthode d'ordre 5 minimise l'erreur locale, comme l'indique la lettre M dans son nom.

(5.8)

	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	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	
$\hat{y}_{n+1}$	$\hat{b}^T$	$\frac{5179}{57600}$	0	$\frac{7571}{16695}$	$\frac{393}{640}$	$-\frac{92097}{339200}$	$\frac{187}{2100}$	$\frac{1}{40}$
y_{n+1}	b^T	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
$b^T - \hat{b}^T$		$\frac{71}{57\,600}$	0	$-\frac{71}{16\,695}$	$\frac{71}{1\,920}$	$-\frac{17\,253}{339\,200}$	$\frac{22}{525}$	$-\frac{1}{40}$

Paire de Dormand–Prince DP(5,4)7M à 7 étapes,

Cette méthode à 7 étapes se réduit en pratique à une méthode à 6 étapes parce que $k_1^{n+1} = k_7^n$ puisque la ligne du vecteur b^T est identique à la 7ème ligne. En anglais on dit que c'est une méthode FSAL (First Same As Last).

L'intervalle de stabilité absolue de la paire DP(5,4)7M est approximativement $(-4, 0)$.

5.2.6. La paire ode23 de Matlab à 4 étapes avec estimation de l'erreur locale. The code `ode23` of the Matlab ODE suite consists in 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:

$$\begin{aligned} k1 &= hf(x_n, y_n), \\ k2 &= hf(x_n + (1/2)h, y_n + (1/2)k_1), \\ k3 &= hf(x_n + (3/4)h, y_n + (3/4)k_2), \\ k4 &= hf(x_n + h, y_n + (2/9)k_1 + (1/3)k_2 + (4/9)k_3), \end{aligned}$$

The first three stages produce the solution at the next time step:

$$y_{n+1} = y_n + (2/9)k_1 + (1/3)k_2 + (4/9)k_3,$$

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 $k_1^{[n+1]} = k_4^{[n]}$ (that is, a FSAL method).

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 $x = x_{n+1}$.

Les anciennes versions de Matlab employaient une paire d'ordre 2 et 3 du type Runge–Kutta à 3 étapes et une estimation de l'erreur locale comme une formule

d'ordre faible pour la résolution de systèmes différentiels du premier ordre. La méthode, notée `ode23`, est

$$\begin{aligned} k_1 &= h\, f(x_n, y_n), \\ k_2 &= h\, f(x_n + h, y_n + k_1), \\ k_3 &= h\, f\left(x_n + \frac{1}{2}\, h, y_n + \frac{1}{4}\, k_1 + \frac{1}{4}\, k_2\right); \\ y_{n+1} &= y_n + \frac{1}{6}\,(k_1 + k_2 + 4k_3); \\ \text{Erreur locale} &\approx \frac{1}{3}\,(k_1 - 2k_3 + k_2). \end{aligned}$$

EXEMPLE 5.3. Soit le problème à valeur initiale:

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

Calculer $y(0.1)$ et $y(0.2)$, au dix-millième près, au moyen de la méthode du type Runge–Kutta ode23 à trois étapes de Matlab 4 avec $h = 0.1$ et estimer l'erreur locale.

RÉSOLUTION. La fonction f au second membre de l'équation différentielle est

$$f(x, y) = xy + 1.$$

Avec $n = 0$:

$$\begin{aligned} k_1 &= 0.1 \times 1 = 0.1 \\ k_2 &= 0.1 \times (0.1 \times 1.1 + 1) = 0.111 \\ k_3 &= 0.1 \times (0.05 \times (1 + 0.025 + 0.02775) + 1) = 0.105\,264 \\ y_1 &= 1.105\,342\,5 \end{aligned}$$

L'estimation de l'erreur locale est

$$\text{Erreur locale} \approx 0.001\,417\,5$$

Avec $n = 1$:

$$\begin{aligned} k_1 &= 0.1(0.1y_1 + 1) = 0.111\,053\,425 \\ k_2 &= 0.1(0.2(y_1 + k_1) + 1) = 0.124\,327\,918\,5 \\ k_3 &= 0.117\,462\,817\,538\,13 \\ y_2 &= 1.222\,881\,268\,942\,08 \end{aligned}$$

L'estimation de l'erreur locale est

$$\text{Erreur locale} \approx 0.001\,367\,125$$

Pour employer la commande `ode23` de Matlab numérique pour résoudre ce problème sur $[0, 1]$, on utilise la fonction fichier M `exp4_3.m`:

```
function yprime = exp4_3(x,y)
yprime = x.*y+1
```

et les commandes

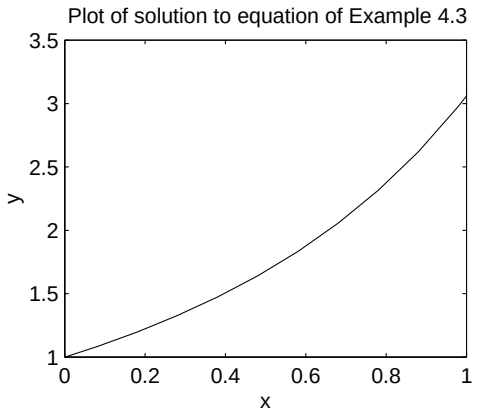

FIGURE 5.2. Graphe de la solution numérique de l'exemple 5.3.

```
>> clear
>> xspan = [0 1]; y0 = 1; % xspan et valeur initiale
>> [x,y] = ode23('exp4_3',xspan,y0);
>> subplot(2,2,1); plot(x,y); xlabel('x'); ylabel('y');
>> title('Graphe de la solution de l'equation de l'exemple 5.3');
print Fig.exp4.3 % imprimer la figure dans le fichier Fig.exp4.3
```

□

Matlab numérique a deux méthodes, chacune formée d'une paire imbriquée de méthodes de Runga–Kutta explicites pour résoudre les équations différentielles non raides:

- `ode23` implémente la paire Runge–Kutta (2,3) explicite de Bogacki et Shampine, appelée BS23. Le solveur utilise un interpolant "libre" d'ordre 3 et l'extrapolation locale. *Extrapolation locale* signifie qu'on avance la solution au moyen de la méthode d'ordre supérieur, en l'occurrence, d'ordre 3.
- `ode45` implémente la paire Runge–Kutta (4,5) explicite de Dormand et Prince, appelée RK5(4)7M, DOPRI5, DP(4,5) ou DP54. Le solveur utilise un interpolant "libre" d'ordre 4 communiquée privément par Dormand et Prince, et l'extrapolation locale.

On trouve les détails sur ces méthodes dans l'article *The MATLAB ODE Suite*, L. F. Shampine et M. W. Reichelt, SIAM Journal on Scientific Computing, **18**(1), 1997.

5.3. Méthodes prédicteur-correcteur multipas

5.3.1. Méthodes multipas générales. Soit le problème à valeur initiale:

$$y' = f(x, y), \qquad y(a) = \eta, \tag{5.9}$$

où f est continue en x et lipschitzienne en y sur $[a, b] \times (-\infty, \infty)$. Alors la solution exacte $y(x)$ existe et est unique sur $[a, b]$.
On cherche une solution approchée $\{y_n\}$ de y aux points $x_n = a + nh$ où h est le

pas et $1 \leq n \leq (b-a)/h$.
Pour ce faire, on considére la *méthode linéaire à k pas*:

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f_{n+j}, \tag{5.10}$$

où $y_n \approx y(x_n)$ et $f_n := f(x_n, y_n)$. On normalise par la condition $\alpha_k = 1$ et l'on stipule que le nombre de pas est bien k par la condition $(\alpha_0, \beta_0) \neq (0,0)$.
On choisit k *valeurs initiales* $y_0, y_1, \ldots, y_{k-1}$. La méthode est *explicite* si $\beta_k = 0$; on obtient alors y_{n+1} directement; elle est *implicite* si $\beta_k \neq 0$; dans ce cas il faut résoudre par la récurrence:

$$y_{n+k}^{[s+1]} = h\beta_k f\left(x_{n+k}, y_{n+k}^{[s]}\right) + g, \qquad y_{n+k}^{[0]} \text{ arbitraire.} \tag{5.11}$$

On a noté

$$g = g(x_n, \ldots, x_{n+k-1}, y_0, \ldots, y_{n+k-1}).$$

La récurrence (5.11) converge quand $s \to \infty$ si la constante de Lipschitz M du second membre de (5.11) par rapport à y_{n+k} satisfait $0 \leq M < 1$. Si la constante de Lipschitz de f par rapport à y est L, alors

$$M := Lh|\beta_k| < 1 \tag{5.12}$$

et il y a convergence si

$$h < \frac{1}{L|\beta_k|}.$$

5.3.2. Méthode d'Adams–Bashford–Moulton à quatre pas et d'ordre 4. A titre d'exemple, parmi toutes les méthodes multipas, on présente la méthode d'Adams–Bashford–Moulton à quatre pas et d'ordre 4.

Le *prédicteur d'Adams–Bashford* et le *correcteur d'Adams–Moulton* d'ordre 4 sont respectivement

$$y_{n+1}^{\mathrm{P}} = y_n^{\mathrm{C}} + \frac{h}{24}\left(55 f_n^{\mathrm{C}} - 59 f_{n-1}^{\mathrm{C}} + 37 f_{n-2}^{\mathrm{C}} - 9 f_{n-3}^{\mathrm{C}}\right) \tag{5.13}$$

et

$$y_{n+1}^{\mathrm{C}} = y_n^{\mathrm{C}} + \frac{h}{24}\left(9 f_{n+1}^{\mathrm{P}} + 19 f_n^{\mathrm{C}} - 5 f_{n-1}^{\mathrm{C}} + f_{n-2}^{\mathrm{C}}\right), \tag{5.14}$$

où

$$f_n^{\mathrm{C}} = f(x_n, y_n^{\mathrm{C}}) \quad \text{et} \quad f_n^{\mathrm{P}} = f(x_n, y_n^{\mathrm{P}}).$$

On démarre avec une méthode de Runge–Kutta ou autrement et l'on contrôle l'erreur locale au moyen de l'estimation suivante:

$$C_5 h^5 y^{(5)}(x_{n+4}) \approx -\frac{19}{270}\left[y_{n+4}^{\mathrm{C}} - y_{n+4}^{\mathrm{P}}\right]. \tag{5.15}$$

On garde en mémoire un nombre de valeurs de y_n et de f_n pour pouvoir allonger le pas si l'erreur est faible devant un seuil donné. Si l'erreur est trop forte devant le seuil donné, on peut réduire le pas de moitié en utilisant les valeurs intermédiaires $y_{n-1/2}$ et $y_{n-3/2}$ calculées au moyen des formules suivantes:

$$y_{n-1/2} = \frac{1}{128}\left(35 y_n + 140 y_{n-1} - 70 y_{n-2} + 28 y_{n-3} - y_{n-4}\right), \tag{5.16}$$

$$y_{n-3/2} = \frac{1}{162}\left(-y_n + 24 y_{n-1} + 54 y_{n-2} - 16 y_{n-3} + 3 y_{n-4}\right). \tag{5.17}$$

En mode *prédicteur–évaluation de f–correcteur–évaluation de f*, noté $PECE$, la méthode d'Adams–Bashford–Moulton admet l'intervalle de stabilité absolue $(-1.25, 0)$, c'est-à-dire, la méthode n'amplifie pas les erreurs antérieures si le pas h est suffisamment petit pour qu'on ait

$$-1.25 < h\frac{\partial f}{\partial y} < 0.$$

EXEMPLE 5.4. Soit le problème à valeur initiale

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

Calculer la solution en $x = 2$ par la méthode d'Adams–Bashford–Moulton d'ordre 4 avec $h = 0.2$. Employer la méthode de Runge–Kutta d'ordre 4 pour obtenir les valeurs de départ. Faire les calculs au cent-millième près. Calculer l'erreur globale au moyen de la solution exacte et estimer l'erreur locale au moyen de (5.15).

RÉSOLUTION. Le calcul de l'erreur globale se fait au moyen de la solution exacte,

$$y(x) = e^x - x - 1.$$

On présente la solution sous forme de tableau: valeurs de départ, valeurs prédites, corrigées et exactes, l'erreur globale $\times 10^6$, EG $\times\ 10^6 = 10^6(y(x_n) - y_n^C)$ et l'estimation de l'erreur locale $\times 10^6$, EEL $\times\ 10^6$ au moyen de (5.15).

n	x_n	Départ y_n^C	Prédite y_n^P	Corrigée y_n^C	Exacte $y(t_n)$	EG $\times 10^6$	EEL $\times 10^6$
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	12
5	1.0		0.718 066	0.718 270	0.718 282	12	14
6	1.2		1.119 855	1.120 106	1.120 117	11	18
7	1.4		1.654 885	1.655 191	1.655 200	9	22
8	1.6		2.352 653	2.353 026	2.353 032	6	26
9	1.8		3.249 190	3.249 646	3.249 647	1	32
10	2.0		4.388 505	4.389 062	4.389 056	−6	39

On voit que la méthode est stable puisque l'erreur ne croît pas. □

EXEMPLE 5.5. Résoudre le problème à valeur initiale

$$y' = \arctan x + \arctan y, \qquad y(0) = 0,$$

au dix-milionnième près au moyen de la méthode d'Adams–Bashforth–Moulton d'ordre 4 sur l'intervalle $[0, 2]$ avec $h = 0.2$. Obtenir les valeurs de départs par Runge–Kutta 4. Utiliser la formule (5.15) pour estimer l'erreur locale à chaque pas.

RÉSOLUTION. **Résolution par Matlab numérique.**— Le fichier M `exp4_5` pour l'exemple 5.5:

```
function yprime = exp4_5(x,y); %MAT 2731, Exemple 5.5.
yprime = atan(x)+atan(y);
```

On emploie les conditions initiales et la méthode de Runge–Kutta d'ordre 4 pour obtenir les 4 valeurs de départ:

```
clear
h = 0.2; x0= 0; xf= 2; y0 = 0;
n = ceil((xf-x0)/h); % nombre de pas
%
count = 2; every_so_often = 1; % quand ecrire les resultats
x = x0; y = y0; % initialiser x et y
output = [0 x0 y0 0];
%RK4
for i=1:3
 k1 = h*exp4_5(x,y);
 k2 = h*exp4_5(x+h/2,y+k1/2);
 k3 = h*exp4_5(x+h/2,y+k2/2);
 k4 = h*exp4_5(x+h,y+k3);
 z  = y + (1/6)*(k1+2*k2+2*k3+k4);
  x = x + h;
   if count > every_so_often
      output = [output; i x z 0];
      count = count - every_so_often;
   end
 y = z;
count = count + 1;
end
% ABM4
for i=4:n
 zp = y + (h/24)*(55*exp4_5(output(i,2),output(i,3))-...
                  59*exp4_5(output(i-1,2),output(i-1,3))+...
                  37*exp4_5(output(i-2,2),output(i-2,3))-...
                   9*exp4_5(output(i-3,2),output(i-3,3)) );
  z = y + (h/24)*( 9*exp4_5(x+h,zp)+...
                  19*exp4_5(output(i,2),output(i,3))-...
                   5*exp4_5(output(i-1,2),output(i-1,3))+...
                     exp4_5(output(i-2,2),output(i-2,3)) );
  x = x + h;
   if count > every_so_often
      errest = -(19/270)*(z-zp);
      output = [output; i x z errest];
      count = count - every_so_often;
   end
 y = z;
count = count + 1;
end
output
save output % pour produire le graphe
```

La commande `output` imprime les valeurs de n, x et y.

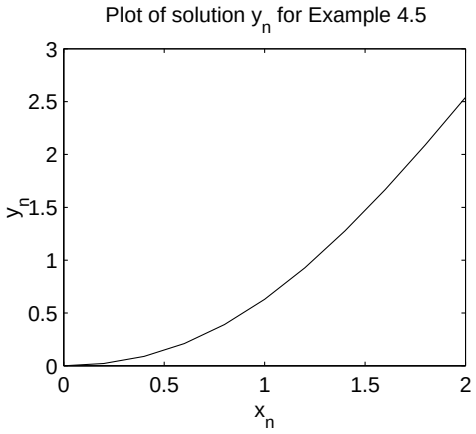

FIGURE 5.3. Graphe de la solution numérique de l'exemple 5.5.

n	x	y	Est. de l'erreur
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

Les commandes suivantes impriment les résultats:

```
load output;
subplot(2,2,1); plot(output(:,2),output(:,3));
title('Graphe de la solution y_n pour l'exemple 5.5');
xlabel('x_n'); ylabel('y_n');
```

□

Dans un 3ème exemple de méthode multipas, on considère la méthode d'Adams–Bashforth–Moulton d'ordre 3 à 3 pas, dont la paire de formules est

$$y_{n+1}^{\mathrm{P}} = y_n^{\mathrm{C}} + \frac{h}{12}\left(23f_n^{\mathrm{C}} - 16f_{n-1}^{\mathrm{C}} + 5f_{n-2}^{\mathrm{C}}\right), \qquad f_k^{\mathrm{C}} = f\left(x_k, y_k^{\mathrm{C}}\right), \tag{5.18}$$

$$y_{n+1}^{\mathrm{C}} = y_n^{\mathrm{C}} + \frac{h}{12}\left(5f_{n+1}^{\mathrm{P}} + 8f_n^{\mathrm{C}} - f_{n-1}^{\mathrm{C}}\right), \qquad f_k^{\mathrm{P}} = f\left(x_k, y_k^{\mathrm{P}}\right). \tag{5.19}$$

et l'estimation de l'erreur locale est

$$\text{Err.} \approx -\frac{1}{10}\left[y_{n+1}^{\mathrm{C}} - y_{n+1}^{\mathrm{P}}\right]. \tag{5.20}$$

EXEMPLE 5.6. Résoudre au millionième près le problème à valeur initiale

$$y' = x + \sin y, \qquad y(0) = 0,$$

au moyen de la méthode d'Adams–Bashforth–Moulton d'ordre 3 sur l'intervalle $[0, 2]$ avec le pas $h = 0.2$. Les valeurs de démarrage proviennent d'une méthode de haute précision. Utiliser la formule (5.20) pour estimer l'erreur locale à chaque pas.

RÉSOLUTION. On présente la solution sous forme de tableau.

n	x_n	Départ y_n^C	Prédite y_n^P	Corrigée y_n^C	$10^5\times$Erreur locale en y_n^C $\approx -(y_n^C - y_n^P) \times 10^4$
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

□

Les méthodes d'Adams–Bashforth–Moulton d'ordres 1 à 5 à pas fixes sont implémentées dans les fonctions fichiers M de Matlab qu'on trouve dans `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
```

On obtient les valeurs initiales dans les fichiers M suivants au moyen d'une méthode de Runge–Kutta.

```
  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
```

On trouve la fonction fichier M `Rkstep` à la sous-section 9.5.3 Le pas du prédicteur d'Adams-Bashforth se fait avec le fichier M:

```
  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);
```

Le pas du correcteur d'Adams-Moulton se fait avec le fichier M:

```
  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);
```

Le pas du prédicteur-correcteur se fait avec le fichier M:

```
  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);
```

5.3.3. Spécification des méthodes multipas. Le premier membre des méthodes d'Adams est de la forme

$$y_{n+1} - y_n.$$

On nomme Adams–Bashforth les méthodes explicites et Adams–Moulton les méthodes implicites. On obtient les méthodes d'Adams dans les tableaux suivants en prenant les paramètres a et b égaux à 0. Le nombre k indique le nombre de pas de la méthode. Le nombre p indique l'ordre de la méthode et la constante C_{p+1} est la constante du premier terme de l'erreur.

Méthodes explicites

$$
\begin{aligned}
k = 1 : & \\
& \alpha_1 = 1, \\
& \alpha_0 = -1, \quad \beta_0 = 1, \\
& p = 1; \qquad C_{p+1} = \tfrac{1}{2}.
\end{aligned}
$$

$$
\begin{aligned}
k = 2 : & \\
& \alpha_2 = 1, \\
& \alpha_1 = -1 - a, \quad \beta_1 = \tfrac{1}{2}(3 - a), \\
& \alpha_0 = a, \qquad \beta_0 = \tfrac{1}{2}(-1 + a), \\
& p = 2; \qquad C_{p+1} = \tfrac{1}{12}(5 + a).
\end{aligned}
$$

La stabilité absolue limite l'ordre à 2.

$$
\begin{aligned}
k = 3 : & \\
& \alpha_3 = 1, \\
& \alpha_2 = -1 - a, \quad \beta_2 = \tfrac{1}{12}(23 - 5a - b), \\
& \alpha_1 = a + b, \qquad \beta_1 = \tfrac{1}{3}(-4 - 2a + 2b), \\
& \alpha_0 = -b, \qquad \beta_0 = \tfrac{1}{12}(5 + a + 5b), \\
& p = 3; \qquad C_{p+1} = \tfrac{1}{24}(9 + a + b).
\end{aligned}
$$

La stabilité absolue limite l'ordre à 3.

$$
\begin{aligned}
k = 4 : & \\
& \alpha_4 = 1, \\
& \alpha_3 = -1 - a, \quad \beta_3 = \tfrac{1}{24}(55 - 9a - b - c), \\
& \alpha_2 = a + b, \qquad \beta_2 = \tfrac{1}{24}(-59 - 19a + 13b - 19c), \\
& \alpha_1 = -b - c, \quad \beta_1 = \tfrac{1}{24}(37 + 5a + 13b - 19c), \\
& \alpha_0 = c, \qquad \beta_0 = \tfrac{1}{24}(-9 - a - b - 9c), \\
& p = 4; \qquad C_{p+1} = \tfrac{1}{720}(251 + 19a + 11b + 19c).
\end{aligned}
$$

La stabilité absolue limite l'ordre à 4.

Méthodes implicites

$$
\begin{aligned}
k = 1 : & \\
& \alpha_1 = 1, \qquad \beta_1 = \tfrac{1}{2}, \\
& \alpha_0 = -1, \quad \beta_0 = \tfrac{1}{2}, \\
& p = 2; \qquad C_{p+1} = -\tfrac{1}{12}.
\end{aligned}
$$

$k=2$:

$$\begin{array}{lll} & \alpha_2 = 1, & \beta_2 = \frac{1}{12}(5+a), \\ & \alpha_1 = -1-a, & \beta_1 = \frac{2}{3}(1-a), \\ & \alpha_0 = a, & \beta_0 = \frac{1}{12}(-1-5a), \\ \text{Si } a \neq -1, & p=3; & C_{p+1} = -\frac{1}{24}(1+a), \\ \text{Si } a = -1, & p=4; & C_{p+1} = -\frac{1}{90}. \end{array}$$

$k=3$:

$$\begin{array}{ll} \alpha_3 = 1, & \beta_3 = \frac{1}{24}(9+a+b), \\ \alpha_2 = -1-a, & \beta_2 = \frac{1}{24}(19-13a-5b), \\ \alpha_1 = a+b, & \beta_1 = \frac{1}{24}(-5-13a+19b), \\ \alpha_0 = -b, & \beta_0 = \frac{1}{24}(1+a+9b), \\ p=4; & C_{p+1} = -\frac{1}{720}(19+11a+19b). \end{array}$$

La stabilité absolue limite l'ordre à 4.

$k=4$:

$$\begin{array}{ll} \alpha_4 = 1, & \beta_4 = \frac{1}{720}(251+19a+11b+19c), \\ \alpha_3 = -1-a, & \beta_3 = \frac{1}{360}(323-173a-37b-53c), \\ \alpha_2 = a+b, & \beta_2 = \frac{1}{30}(-11-19a+19b+11c), \\ \alpha_1 = -b-c, & \beta_1 = \frac{1}{360}(53+37a+173b-323c), \\ \alpha_0 = c, & \beta_0 = \frac{1}{720}(-19-11a-19b-251c). \end{array}$$

Si $27+11a+11b+27c \neq 0$, alors

$$p=5; \qquad C_{p+1} = -\frac{1}{1440}(27+11a+11b+27c).$$

Si $27+11a+11b+27c = 0$, alors

$$p=6; \qquad C_{p+1} = -\frac{1}{15\,120}(74+10a-10b-74c).$$

La stabilité absolue limite l'ordre à 6.

Le solveur `ode113` de Matlab est une implémentation PECE à pas variables d'une modification en différences divisées d'une famille de formules d'Adams–Bashforth–Moulton d'ordres 1 à 12. Il utilise les interpolants naturels "libres" et l'extrapolation locale. On trouve les détails dans *The MATLAB ODE Suite*, L. F. Shampine et M. W. Reichelt, SIAM Journal on Scientific Computing, **18**(1), 1997.

5.4. Systèmes différentiels raides

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.

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

$$A = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}.$$

DÉFINITION 5.2. Suppose the eigenvalues of the Jacobian matrix A of a system of differential equation have negative real parts and are ordered as follows

$$\Re\lambda_n \leq \Re\lambda_{n-1} \leq \cdots \leq \Re\lambda_2 \leq \Re\lambda_1 \leq 0.$$

The stiffness ratio of the system is

$$\frac{|\Re\lambda_n|}{|\Re\lambda_1|}.$$

The following definition of stiffness will suffice for our purpose, but the reader should be aware that the phenomenon of stiffness may be more complex in general.

DÉFINITION 5.3. A system of differential equation

$$\boldsymbol{y}' = \boldsymbol{f}(x, \boldsymbol{y})$$

is said to be stiff if the stiffness ratio of the system is large.

We consider an example of a second-order equation, with one real parameter q, which we first solve analytically.

EXEMPLE 5.7. 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 5.7 with numeric Matlab, we reformulate it into a system of two first-order equations.

EXEMPLE 5.8. 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

$$u_1 = y,$$
$$u_2 = y'$$

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 5.8 is

$$10^q.$$

EXEMPLE 5.9. Use the five Matlab ode solvers to solve the nonstiff differential equations

$$y'' + (10^q + 1)y' + 10^q = 0 \qquad \text{on} \quad [0, 1],$$

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 `exp443.m` is

```
function uprime = exp443(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('exp443',xspan,u0);
>> [x45,u45] = ode45('exp443',xspan,u0);
>> [x113,u113] = ode113('exp443',xspan,u0);
>> [x23s,u23s] = ode23s('exp443',xspan,u0);
>> [x15s,u15s] = ode15s('exp443',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. □

EXEMPLE 5.10. Use the five Matlab ode solvers to solve the stiff differential equations

$$y'' + (10^q + 1)y' + 10^q = 0 \qquad \text{on} \quad [0, 1],$$

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 4.4 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('exp443',xspan,u0);
[x45,u45] = ode45('exp443',xspan,u0);
[x113,u113] = ode113('exp443',xspan,u0);
[x23s,u23s] = ode23s('exp443',xspan,u0);
[x15s,u15s] = ode15s('exp443',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 two solvers for stiff systems.

- The Matlab solver `ode23s` is an implementation of a new modified Rosenbrock (2,3) pair with a "free" interpolant. Local extrapolation is not done. By default, Jacobians are generated numerically.
- 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 Formulas of orders 1 to 5. The natural "free" interpolants are used. 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.

CHAPITRE 6

Solutions séries

6.1. La méthode

On illustre la méthode par un exemple très simple.

EXEMPLE 6.1. Soit

$$y'' + 25y = 0, \quad y(0) = 3, \quad y'(0) = 13.$$

Trouver la solution en série de puissances:

$$y(x) = a_0 + a_1 x + a_2 x^2 + \dots$$

RÉSOLUTION. Dans ce cas simple, on connaît déjà l'unique solution:

$$y(x) = a\cos 5x + b\sin 5x,$$

où les constantes a et b sont déterminées par les conditions initiales:

$$y(0) = a = 3 \implies a = 3,$$
$$y'(0) = 5b = 13 \implies b = \frac{13}{5}.$$

On sait aussi que

$$\cos 5x = 1 - \frac{(5x)^2}{2!} + \frac{(5x)^4}{4!} - \frac{(5x)^6}{6!} + - \cdots$$

et

$$\sin 5x = 5x - \frac{(5x)^3}{3!} + \frac{(5x)^5}{5!} - \frac{(5x)^7}{7!} + - \cdots$$

Pour obtenir la solution en série, on pose

$$y(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

dans

$$y'' + 25y = 0;$$

alors

$$y''(x) = 2a_2 + 3 \times 2a_3 x + 4 \times 3a_4 x^2 + 5 \times 4a_5 x^3 + 6 \times 5a_6 x^4 + \cdots$$
$$25y(x) = 25a_0 + 25a_1 x + 25a_2 x^2 + 25a_3 x^3 + 25a_4 x^4 + \cdots$$

et l'on somme:

$$0 = (2a_2 + 25a_0) + (3 \times 2a_3 + 25a_1)x + (4 \times 3a_4 + 25a_2)x^2 + \cdots, \quad \text{pour tout } x.$$

Puisque nous avons une identité en

$$1,\ x,\ x^2,\ x^3, \dots,$$

il suit que tous les coefficients sont nuls. Donc avec a_0 et a_1 indéterminés, on a

$$\begin{aligned}
2a_2 + 25a_0 = 0 &\implies a_2 = -\frac{5^2}{2!}a_0,\\
3\times 2a_3 + 25a_1 = 0 &\implies a_3 = -\frac{5^2}{3!}a_1,\\
4\times 3a_4 + 25a_2 = 0 &\implies a_4 = \frac{5^4}{4!}a_0,\\
5\times 4a_5 + 25a_3 = 0 &\implies a_5 = \frac{5^4}{5!}a_1,
\end{aligned}$$

etc., d'où l'on obtient le développement

$$\begin{aligned}
y(x) &= a_0\left[1 - \frac{1}{2!}(5x)^2 + \frac{1}{4!}(5x)^4 - \frac{1}{6!}(5x)^6 + \cdots\right]\\
&\quad + \frac{a_1}{5}\left[5x - \frac{1}{3!}(5x)^3 + \frac{1}{5!}(5x)^5 - \cdots\right]\\
&= a_0\cos 5x + \frac{a_1}{5}\sin 5x.
\end{aligned}$$

La condition initiale $y(0) = 3$ détermine a_0:

$$a_0 = 3,$$

et la condition initiale $y'(0) = 13$ détermine a_1:

$$5\,\frac{a_1}{5} = 13, \quad a_1 = 13. \quad \square$$

6.2. Fondements de la méthode des séries de puissances

Il sera avantageux de considérer les séries entières dans le plan complexe. On rappelle qu'un point z du plan complexe $\mathbb{C}$ admet les représentations suivantes:

- *cartésienne* ou *algébrique*:
$$z = x + iy, \quad i^2 = -1,$$
- *trigonométrique*:
$$z = r(\cos\theta + i\sin\theta),$$
- *polaire* ou *eulérienne*:
$$z = r\,e^{i\theta},$$

où

$$r = \sqrt{x^2+y^2}, \quad \theta = \arg z = \arctan\frac{y}{x}.$$

On note $\bar z = x - iy$ le conjugué complexe de z et

$$|z| = \sqrt{x^2+y^2} = \sqrt{z\bar z} = r$$

le module de z (V. figure 6.1).

Exemple 6.2. Prolonger la fonction

$$f(x) = \frac{1}{1-x},$$

au plan complexe et la developper en série de puissances de centre $z_0 = 0$, $z_0 = -1$ et $z_0 = i$.

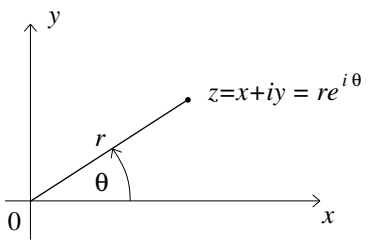

FIGURE 6.1. Un point $z = x + iy = r\,e^{i\theta}$ du plan complexe $\mathbb{C}$.

RÉSOLUTION. On prolonge la fonction d'une variable réelle

$$f(x) = \frac{1}{1-x}, \qquad x \in \mathbb{R} \setminus \{1\},$$

au plan complexe,

$$f(z) = \frac{1}{1-z}, \qquad z = x + iy \in \mathbb{C}.$$

C'est une fonction rationnelle qui admet un pôle simple en $z = 1$. On dit que $z = 1$ est un pôle de $f(z)$ puisque $|f(z)|$ tend vers $+\infty$ quand $z \to 1$. De plus, il s'agit d'un pôle simple puisque $1 - z$ apparaît à la première puissance au dénominateur.

On développe $f(z)$ en série de TAYLOR près de 0:

$$f(z) = f(0) + \frac{1}{1!}f'(0)z + \frac{1}{2!}f''(0)z^2 + \cdots$$

Puisque

$$\begin{aligned}
f(z) &= \frac{1}{(1-z)} && \implies f(0) = 1,\\
f'(z) &= \frac{1!}{(1-z)^2} && \implies f'(0) = 1!,\\
f''(z) &= \frac{2!}{(1-z)^3} && \implies f''(0) = 2!,\\
&\vdots\\
f^{(n)}(z) &= \frac{n!}{(1-z)^{n+1}} && \implies f^{(n)}(0) = n!,
\end{aligned}$$

il suit que

$$f(z) = \frac{1}{1-z} = 1 + z + z^2 + z^3 + \cdots = \sum_{n=0}^{\infty} z^n.$$

La série converge **absolument** pour $|z| \equiv \sqrt{x^2 + y^2} < 1$, c'est-à-dire

$$\sum_{n=0}^{\infty} |z|^n < \infty, \qquad \text{pour tout } |z| < 1,$$

et **uniformément** pour $|z| \leq \rho < 1$, c'est-à-dire

$$\left| \sum_{n=N}^{\infty} z^n \right| < \epsilon, \quad \text{pour tout } N > N_\epsilon, \quad \text{et pour tout } |z| \leq \rho < 1.$$

Le rayon de convergence R de $\sum_{n=0}^{\infty} z^n$ est 1.

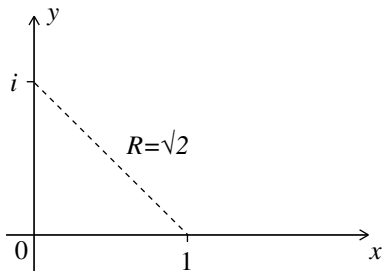

FIGURE 6.2. Distance du centre $a = i$ au pôle $z = 1$.

Maintenant, on développe f au voisinage de $z = -1$:

$$\begin{aligned} f(z) &= \frac{1}{1-z} = \frac{1}{1-(z+1-1)} \\ &= \frac{1}{2-(z+1)} = \frac{1}{2}\,\frac{1}{1-(\frac{z+1}{2})} \\ &= \frac{1}{2}\left\{1 + \frac{z+1}{2} + \left(\frac{z+1}{2}\right)^2 + \left(\frac{z+1}{2}\right)^3 + \left(\frac{z+1}{2}\right)^4 + \cdots\right\}. \end{aligned}$$

La série converge absolument pour

$$\left|\frac{z+1}{2}\right| < 1, \quad |z+1| < 2, \quad |z-(-1)| < 2.$$

Le centre du disque de convergence est $z = -1$ et son rayon est $R = 2$.

Enfin, on développe f au voisinage de $z = i$:

$$\begin{aligned} f(z) &= \frac{1}{1-z} = \frac{1}{1-(z-i+i)} \\ &= \frac{1}{(1-i)-(z-i)} = \frac{1}{1-i}\,\frac{1}{1-(\frac{z-i}{1-i})} \\ &= \frac{1}{1-i}\left\{1 + \frac{z-i}{1-i} + \left(\frac{z-i}{1-i}\right)^2 + \left(\frac{z-i}{1-i}\right)^3 + \left(\frac{z-i}{1-i}\right)^4 + \cdots\right\}. \end{aligned}$$

La série converge absolument pour

$$\left|\frac{z-i}{1-i}\right| < 1, \quad \text{c'est-à-dire} \quad |z-i| < |1-i| = \sqrt{2}.$$

On voit que le centre du disque de convergence est bien $z = i$ et son rayon est $R = \sqrt{2}$ (V. figure 6.2). □

Cet exemple montre que le développement de TAYLOR d'une fonction $f(z)$, en $z = a$ et de rayon de convergence R, n'est plus convergent dès que $|z-a| \geq R$, c'est-à-dire dès que $|z-a|$ est plus grand que la distance de a à la singularité z_1 la plus proche (V. figure 6.3).

On emploiera le résultat suivant.

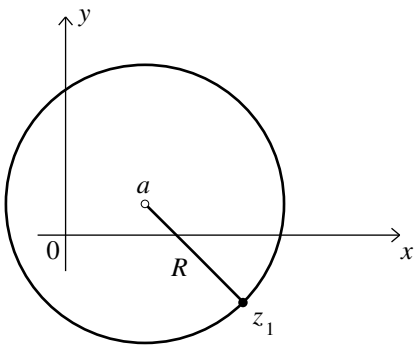

FIGURE 6.3. Distance R du centre a à la singularité z_1 la plus proche.

THÉORÈME 6.1 (Critères de convergence). *L'inverse du rayon de convergence R d'une série entière de centre a,*

$$\sum_{m=0}^{\infty} a_m (z-a)^m, \tag{6.1}$$

est égal à la limite supérieure:

$$\frac{1}{R} = \limsup_{m\to\infty} |a_m|^{1/m}. \tag{6.2}$$

On a aussi le critère:

$$\frac{1}{R} = \lim_{m\to\infty} \left| \frac{a_{m+1}}{a_m} \right|, \tag{6.3}$$

si cette limite existe.

DÉMONSTRATION. Le critère de convergence de CAUCHY affirme que la série

$$\sum_{m=0}^{\infty} c_m$$

converge si

$$\lim_{m\to\infty} |c_m|^{1/m} < 1.$$

Alors la série entière converge si

$$\lim_{m\to\infty} |a_m (z-a)^m|^{1/m} = \lim_{m\to\infty} |a_m|^{1/m} |z-a| < 1.$$

Soit R le maximum de $|z-a|$ tel que l'égalité

$$\lim_{m\to\infty} |a_m|^{1/m} R = 1$$

soit satisfaite. S'il y a plusieurs limites, il faut prendre la limite supérieure. Ceci démontre le critère (6.2). Le second critère découle du critère de D'ALEMBERT qui affirme que la série

$$\sum_{m=0}^{\infty} c_m$$

converge si

$$\lim_{m\to\infty} \frac{|c_{m+1}|}{|c_m|} < 1.$$

Donc, la série de puissance converges si

$$\lim_{m\to\infty} \frac{|a_{m+1}(z-a)^{m+1}|}{|a_m(z-a)^m|} = \lim_{m\to\infty} \frac{|a_{m+1}|}{|a_m|}\,|z-a| < 1.$$

Soit R le maximum de $|z-a|$ tel que l'égalité

$$\lim_{m\to\infty} \frac{|a_{m+1}|}{|a_m|}\,R = 1$$

soit satisfaite. Ceci démontre le critère (6.3). □

EXEMPLE 6.3. Trouver le rayon de convergence de la série

$$\sum_{m=0}^{\infty} \frac{1}{k^m} x^{3m}$$

et de sa 1ère dérivée terme à terme.

RÉSOLUTION. Par le critère (6.2),

$$\frac{1}{R} = \limsup_{m\to\infty} |a_m|^{1/m} = \lim_{m\to\infty} \left|\frac{1}{k^m}\right|^{1/3m} = \frac{1}{|k|^{1/3}}.$$

Donc le rayon de convergence de la série est

$$R = |k|^{1/3}.$$

Pour employer le critère (6.3), on pose

$$w = z^3$$

dans la série qui devient

$$\sum_{m=0}^{\infty} \frac{1}{k^m} w^m.$$

Alors le rayon de convergence, R_1, de la nouvelle série est donné par la formule

$$\frac{1}{R_1} = \lim_{m\to\infty} \left|\frac{k^m}{k^{m+1}}\right| = \left|\frac{1}{k}\right|.$$

Donc le rayon de convergence de la nouvelle séries est $R_1 = |k|$. La série originale converge pour

$$|z^3| = |w| < |k|, \quad \text{c'est-à-dire} \quad |z| < |k|^{1/3}.$$

Le rayon de convergence R' de la série dérivée,

$$\sum_{m=0}^{\infty} \frac{3m}{k^m} x^{3m-1},$$

s'obtient de la même manière:

$$\begin{aligned}
\frac{1}{R'} &= \lim_{m\to\infty} \left|\frac{3m}{k^m}\right|^{1/(3m-1)} \\
&= \lim_{m\to\infty} |3m|^{1/(3m-1)} \lim_{m\to\infty} \left|\frac{1}{k^m}\right|^{(1/m)(m/(3m-1)} \\
&= \lim_{m\to\infty} \left(\frac{1}{|k|}\right)^{1/(3-1/m)} \\
&= \frac{1}{|k|^{1/3}},
\end{aligned}$$

puisque

$$\lim_{m\to\infty} |3m|^{1/(3m-1)} = 1. \quad \square$$

On voit par récurrence que toutes les séries dérivées terme à terme d'une série donnée admettent la même rayon de convergence R.

DÉFINITION 6.1. On dit que la fonction f est *analytique* dans le disque $D(a, R)$, de centre a et de rayon $R > 0$, si elle admet un développement de centre a,

$$f(z) = \sum_{n=0}^{\infty} a_n (z-a)^n,$$

uniformément convergent dans tout disque fermé strictement contenu dans $D(a, R)$.

Le théorème suivant est une conséquence immédiate de la définition précédente.

THÉORÈME 6.2. *Une fonction f analytique dans $D(a, R)$ admet la représentation*

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(z-a)^n$$

uniformément et absolument convergente dans $D(a, R)$. De plus $f(z)$ est indéfiniment dérivable:

$$f^{(k)}(z) = \sum_{n=k}^{\infty} \frac{f^{(n)}(a)}{(n-k)!}(z-a)^{n-k}, \quad k = 0, 1, 2, \ldots,$$

dans $D(a, R)$.

DÉMONSTRATION. Puisque le rayon de convergence de la série dérivée terme à terme est R, le résultat découle du fait que la série dérivée converge uniformément dans tout disque fermé strictement contenu dans $D(a, R)$ et que $f(z)$ est différentiable sur $D(a, R)$. $\square$

On a le théorème général suivant pour les équations différentielles linéaires à coefficients analytiques.

THÉORÈME 6.3 (Existence de solutions en série). *Soit l'équation différentielle*

$$y'' + f(x)y' + g(x)y = r(x),$$

où f, g et r sont des fonctions analytiques au voisinage de a. Si R est le minimum des rayons de convergence des développements en série entière, de centre a, de f, g et r, alors l'equation différentielle admet une solution analytique *de centre a et de rayon de convergence R.*

DÉMONSTRATION. La démonstration se fait par la méthode des majorantes dans le plan complexe $\mathbb{C}$, qui consiste à trouver une série (à coefficients positifs) absolument convergente dans $D(a, R)$,

$$\sum_{n=0}^{\infty} b_n (x-a)^n,$$

dont les coefficients bornent en module ceux de la solution

$$y(x) = \sum_{n=0}^{\infty} a_n (x-a)^n,$$

c'est-à-dire

$$|a_n| \leq b_n. \quad \square$$

On emploiera les théorèmes 6.2 et 6.3 pour obtenir les solutions en série de puissances d'équations différentielles. Dans les trois sections suivantes, on obtiendra la solution analytique de l'équation de LEGENDRE et l'on démontrera les relations d'orthogonalité des polynômes $P_n(x)$ de LEGENDRE. À la dernière section, on obtiendra les formules de quadrature gaussiennes à deux et à trois points.

EXEMPLE 6.4. Résoudre

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

en série de puissances.

RÉSOLUTION. Posons

$$y(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$$

dans l'équation différentielle:

$$\left.\begin{array}{r} y' \\ -xy \\ -1 \end{array}\right\} = \left\{\begin{array}{l} 1a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots \\ \quad - \ a_0 x - \ a_1 x^2 - \ a_2 x^3 - \dots \\ -1 \end{array}\right.$$

La somme du 1er membre est nulle parce qu'on suppose que $y(x)$ est une solution de l'équation différentielle. Alors la somme du second membre est

$$0 = (a_1 - 1) + (2a_2 - a_0)x + (3a_3 - a_1)x^2 + (4a_4 - a_2)x^3 + \dots.$$

Puisque nous avons une identité en x, le coefficient de chaque terme en x^s est nul pour $s = 0, 1, 2, \dots$. De plus, puisque l'équation différentielle est du 1er ordre, un des coefficients sera indéterminé. Donc,

$$\begin{aligned} a_1 - 1 = 0 &\Longrightarrow a_1 = 1, \\ 2a_2 - a_0 = 0 &\Longrightarrow a_2 = \frac{a_0}{2}, \qquad a_0 \text{ arbitraire}, \\ 3a_3 - a_1 = 0 &\Longrightarrow a_3 = \frac{a_1}{3} = \frac{1}{3}, \\ 4a_4 - a_2 = 0 &\Longrightarrow a_4 = \frac{a_2}{4} = \frac{a_0}{8}, \\ 5a_5 - a_3 = 0 &\Longrightarrow a_5 = \frac{a_3}{5} = \frac{1}{15}, \end{aligned}$$

ainsi de suite. La condition initiale $y(0) = 1$ implique que $a_0 = 1$. Donc la solution est

$$y(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} + \frac{x^5}{15} + \dots.$$

Celle-ci coïncide avec les solutions des exemples 1.19 and 1.20. $\square$

6.3. Equation et polynômes de Legendre

On cherche la solution générale de l'équation de LEGENDRE:

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0, \quad -1 < x < 1, \tag{6.4}$$

sous forme de série de puissances de centre $a = 0$. On récrit l'équation sous forme standard $y'' + f(x)y' + g(x)y = r(x)$:

$$y'' - \frac{2x}{1-x^2}y' + \frac{n(n+1)}{1-x^2}y = 0.$$

Les coefficients,

$$f(x) = \frac{2x}{(x-1)(x+1)}, \qquad g(x) = -\frac{n(n+1)}{(x-1)(x+1)},$$

admettent des pôles simples en $x = \pm 1$. Alors, ils admettent les développements en série de puissance de centre $a = 0$ et de rayon de convergence $R = 1$:

$$\begin{aligned}
f(x) &= -\frac{2x}{1-x^2} = -2x[1 + x^2 + x^4 + x^6 + \cdots], \quad -1 < x < 1, \\
g(x) &= \frac{n(n+1)}{1-x^2} = n(n+1)[1 + x^2 + x^4 + x^6 + \cdots], \quad -1 < x < 1, \\
r(x) &= 0, \quad -\infty < x < \infty.
\end{aligned}$$

On voit donc que f et g sont analytiques sur $-1 < x < 1$ et r est analytique partout.

Par le théorème 6.3, l'équation (6.4) admet deux solutions indépendantes et analytiques sur $-1 < x < 1$.

Posons

$$y(x) = \sum_{m=0}^{\infty} a_m x^m \tag{6.5}$$

dans (6.4), avec $k = n(n+1)$:

$$\left.\begin{array}{r} y'' \\ -x^2y'' \\ -2xy' \\ ky \end{array}\right\} = \left\{\begin{array}{rrrrl} 2\times 1a_2 + & 3\times 2a_3x + & 4\times 3a_4x^2 + & 5\times 4a_5x^3 + & \cdots \\ & & -2\times 1a_2x^2 - & 3\times 2a_3x^3 - & \cdots \\ & -\ 2a_1x - & 2\times 2a_2x^2 - & 2\times 3a_3x^3 - & \cdots \\ ka_0 + & ka_1x + & ka_2x^2 + & ka_3x^3 + & \cdots \end{array}\right.$$

La somme de chacun des membres est nulle puisqu'on suppose que (6.5) est une solution:

$$\begin{aligned}
0 = {} & (2\times 1a_2 + ka_0) + (3\times 2a_3 - 2a_1 + ka_1)x \\
& + (4\times 3a_4 - 2\times 1a_2 - 2\times 2a_2 + ka_2)x^2 \\
& + \cdots \\
& + [(s+2)(s+1)a_{s+2} - s(s-1)a_s - 2sa_s + ka_s]x^s \\
& + \cdots, \quad \text{pour tout } x.
\end{aligned}$$

Puisque nous avons une identité en x, chacun des coefficients de x^s, $s = 0, 1, 2, \ldots$, est nul, et puisque l'équation (6.4) est du second ordre, deux des a_m seront indéterminés. On a donc

$$2!a_2 + ka_0 = 0 \implies a_2 = -\frac{n(n+1)}{2!}a_0, \quad a_0 \text{ indéterminé},$$

$$(3\times 2)a_3+(-2+k)a_1=0 \implies a_3=\frac{2-n(n+1)}{3!}a_1, \quad a_1 \text{ indéterminé,}$$

$$(s+2)(s+1)a_{s+2}+[-s(s-1)-2s+n(n+1)]a_s=0$$
$$\implies a_{s+2}=-\frac{(n-s)(n+s+1)}{(s+2)(s+1)}a_s, \quad s=0,1,2,\ldots,$$

d'où

$$a_2=-\frac{n(n+1)}{2!}a_0, \qquad a_3=-\frac{(n-1)(n+2)}{3!}a_1, \tag{6.6}$$

$$a_4=\frac{(n-2)n(n+1)(n+3)}{4!}a_0, \quad a_5=\frac{(n-3)(n-1)(n+2)(n+4)}{5!}a_1, \tag{6.7}$$

etc. On peut donc écrire la solution de la forme:

$$y(x)=a_0y_1(x)+a_1y_2(x), \tag{6.8}$$

où

$$y_1(x)=1-\frac{n(n+1)}{2!}x^2+\frac{(n-2)n(n+1)(n+3)}{4!}x^4-+\cdots,$$
$$y_2(x)=x-\frac{(n-1)(n+2)}{3!}x^3+\frac{(n-3)(n-1)(n+2)(n+4)}{5!}x^5-+\cdots$$

Ces séries convergent pour $|x|<R=1$. On remarque que y_1 est paire et y_2 est impaire. Puisque

$$\frac{y_1(x)}{y_2(x)}\neq \text{ constante},$$

il suit que y_1 et y_2 sont deux solutions indépendantes et (6.8) est la solution générale.

COROLLAIRE 6.1. *Pour n pair, $y_1(x)$ est un polynôme pair,*

$$y_1(x)=k_nP_n(x),$$

et de même, pour n impair, on a le polynôme impair

$$y_2(x)=k_nP_n(x),$$

où $P_n(x)$ est le polynôme de LEGENDRE *de degré n, tel que $P_n(1)=1$.*

Voici les 6 premiers polynômes de LEGENDRE (V. figure 6.4):

$$P_0(x)=1, \qquad P_1(x)=x,$$
$$P_2(x)=\frac{1}{2}\left(3x^2-1\right), \qquad P_3(x)=\frac{1}{2}\left(5x^3-3x\right),$$
$$P_4(x)=\frac{1}{8}\left(35x^4-30x^2+3\right), \qquad P_5(x)=\frac{1}{8}\left(63x^5-70x^3+15x\right).$$

On remarque que les n zéros du polynôme $P_n(x)$, de degré n, sont tous dans l'intervalle $]-1,1[$. De plus, ils sont simples et entrelacent les $n-1$ zéros de $P_{n-1}(x)$; les zeros de fonctions orthogonales jouissent ordinairement de ces deux propriétés.

REMARQUE 6.1. On peut montrer que les séries donnant y_1, resp. y_2, divergent en $x=\pm 1$ si $n\neq 0,2,4,\ldots$, resp. $n\neq 1,3,5,\ldots$

On peut construire les polynômes de Legengre au moyen de Matlab symbolique avec $P_n(1)=1$:

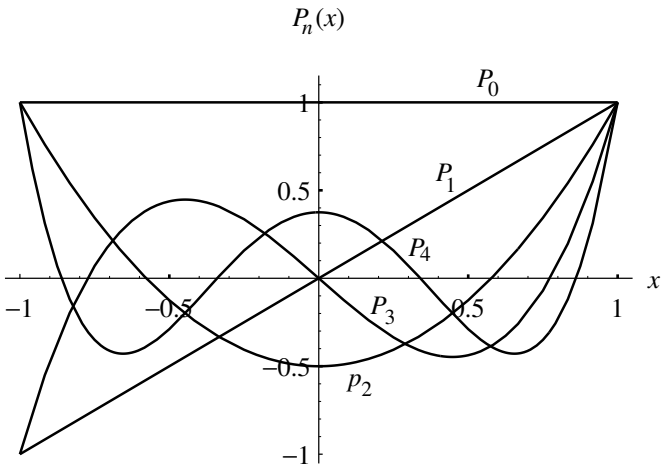

FIGURE 6.4. Les 5 premiers polynômes de LEGENDRE.

```
>> 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
```

ainsi de suite. Avec la boîte symbolique étendue de la version professionnelle de Matlab, on peut obtenir les polynômes de Legendre $P_n(x)$ en recourant au noyau complet de Mable par la commande `orthopoly[L](n,x)`, elle-même référencée par la commande `mhelp orthopoly[L]`.

6.4. Orthogonalité des polynômes de Legendre

THÉORÈME 6.4. *Les polynômes de* LEGENDRE $P_n(x)$ *satisfont la relation d'orthogonalité suivante:*

$$\int_{-1}^{1} P_m(x)P_n(x)\,dx = \begin{cases} 0, & m \neq n, \\ \frac{2}{2n+1}, & m = n. \end{cases} \tag{6.9}$$

DÉMONSTRATION. On donnera deux démonstrations de la 2$^{\text{e}}$ partie. La 1$^{\text{re}}$ partie (c'est-à-dire pour $m \neq n$) découle de l'équation de LEGENDRE:

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0,$$

récrite sous forme de divergence:

$$L_n y := [(1-x^2)y']' + n(n+1)y = 0.$$

Puisque P_m et P_n sont solutions respectivement de $L_m y = 0$ et $L_n y = 0$, on a

$$P_n(x)L_m(P_m) = 0, \quad P_m(x)L_n(P_n) = 0;$$

on intègre ces deux expressions de -1 à 1:

$$\int_{-1}^{1} P_n(x)[(1-x^2)P_m'(x)]'\,dx + m(m+1)\int_{-1}^{1} P_n(x)P_m(x)\,dx = 0,$$
$$\int_{-1}^{1} P_m(x)[(1-x^2)P_n'(x)]'\,dx + n(n+1)\int_{-1}^{1} P_m(x)P_n(x)\,dx = 0,$$

et l'on intègre le 1er terme de chacune de ces expressions par parties:

$$P_n(x)(1-x^2)P_m'(x)\Big|_{-1}^{1} - \int_{-1}^{1} P_n'(x)(1-x^2)P_m'(x)\,dx + m(m+1)\int_{-1}^{1} P_n(x)P_m(x)\,dx = 0,$$
$$P_m(x)(1-x^2)P_n'(x)\Big|_{-1}^{1} - \int_{-1}^{1} P_m'(x)(1-x^2)P_n'(x)\,dx + n(n+1)\int_{-1}^{1} P_m(x)P_n(x)\,dx = 0.$$

Les deux termes intégrés sont nuls et le terme suivant de chacune des équations est identique. Donc, en soustrayant on obtient l'orthogonalité des P_n:

$$[m(m+1)-n(n+1)]\int_{-1}^{1} P_m(x)P_n(x)\,dx = 0$$
$$\Longrightarrow \int_{-1}^{1} P_m(x)P_n(x)\,dx = 0 \quad \text{pour} \quad m \neq n.$$

La 2e partie, $m = n$, suit de la **formule de** Rodrigues:

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left[\left(x^2-1\right)^n\right]. \tag{6.10}$$

En effet,

$$\begin{aligned}
\int_{-1}^{1} P_n^2(x)\,dx &= \frac{1}{2^n}\times\frac{1}{n!}\times\frac{1}{2^n}\times\frac{1}{n!}\int_{-1}^{1}\left[\frac{d^n}{dx^n}\left(x^2-1\right)^n\right]\left[\frac{d^n}{dx^n}\left(x^2-1\right)^n\right]dx \\
&\quad \{\text{et intégrant par parties } n \text{ fois:}\} \\
&= \frac{1}{2^n}\times\frac{1}{n!}\times\frac{1}{2^n}\times\frac{1}{n!}\left[\frac{d^{n-1}}{dx^{n-1}}\left(x^2-1\right)^n\frac{d^n}{dx^n}\left(x^2-1\right)\Big|_{-1}^{1}\right. \\
&\quad \left. + (-1)^1\int_{-1}^{1}\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n\frac{d^{n+1}}{dx^{n+1}}\left(x^2-1\right)^n\,dx\right] \\
&\quad + \cdots \\
&= \frac{1}{2^n}\times\frac{1}{n!}\times\frac{1}{2^n}\times\frac{1}{n!}(-1)^n\int_{-1}^{1}\left(x^2-1\right)^n\frac{d^{2n}}{dx^{2n}}\left(x^2-1\right)^n\,dx \\
&\quad \{\text{et dérivant } 2n \text{ fois:}\} \\
&= \frac{1}{2^n}\times\frac{1}{n!}\times\frac{1}{2^n}\times\frac{1}{n!}(-1)^n(2n)!\int_{-1}^{1} 1\times\left(x^2-1\right)^n\,dx \\
&\quad \{\text{et intégrant de nouveau par parties } n \text{ fois:}\} \\
&= \frac{(-1)^n(2n)!}{2^n n! 2^n n!}\left[\frac{x}{1}\left(x^2-1\right)^n\Big|_{-1}^{1} + \frac{(-1)^1}{1!}2n\int_{-1}^{1} x^2\left(x^2-1\right)^{n-1}\,dx\right]
\end{aligned}$$

$$
\begin{aligned}
&+\cdots\\
&= \frac{(-1)^n(2n)!}{2^n n! 2^n n!}(-1)^n \frac{2n2(n-1)2(n-2)\cdots 2(n-(n-1))}{1\times 3\times 5\times\cdots\times(2n-1)}\int_{-1}^{1} x^{2n}\,dx\\
&= \frac{(-1)^n(-1)^n(2n)!}{2^n n! 2^n n!}\,\frac{2^n n!}{1\times 3\times 5\times\cdots\times(2n-1)}\,\frac{1}{(2n+1)}x^{2n+1}\Big|_{-1}^{1}\\
&= \frac{2}{2n+1}. \quad \square
\end{aligned}
$$

REMARQUE 6.2. La formule de RODRIGUES se démontre par un calcul direct avec $n = 0, 1, 2, 3, \ldots$, ou autrement. On calcule $P_4(x)$ au moyen de la formule de Rodrigues avec la commande `diff` de Matlab symbolique.

```
>> syms x
>> 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
```

On présente une 2e démonstration de la formule:

$$\|P_n\|^2 := \int_{-1}^{1} [P_n(x)]^2\,dx = \frac{2}{2n+1}$$

au moyen de la **fonction génératrice** de $P_n(x)$:

$$\sum_{k=0}^{\infty} P_k(x)t^k = \frac{1}{\sqrt{1-2xt+t^2}}. \tag{6.11}$$

DÉMONSTRATION. Élevons au carré chacun des deux membres:

$$\sum_{k=0}^{\infty} P_k^2(x)t^{2k} + \sum_{j\neq k} P_j(x)P_k(x)t^{j+k} = \frac{1}{1-2xt+t^2},$$

et intégrons:

$$\sum_{k=0}^{\infty}\left[\int_{-1}^{1} P_k^2(x)\,dx\right]t^{2k} + \sum_{j\neq k}\left[\int_{-1}^{1} P_j(x)P_k(x)\,dx\right]t^{j+k} = \int_{-1}^{1}\frac{dx}{1-2xt+t^2}.$$

Comme P_j et P_k sont orthogonaux pour $j \neq k$, le 2e terme du 1er membre est nul et nous obtenons après intégration du 2e membre:

$$
\begin{aligned}
\sum_{k=0}^{\infty}\|P_k\|^2 t^{2k} &= -\frac{1}{2t}\ln(1-2xt+t^2)\Big|_{x=-1}^{x=1}\\
&= -\frac{1}{t}[\ln(1-t) - \ln(1+t)].
\end{aligned}
$$

On multiplie par t:

$$\sum_{k=0}^{\infty}\|P_k\|^2 t^{2k+1} = -\ln(1-t) + \ln(1+t)$$

et l'on dérive par rapport à t:

$$\begin{aligned}\sum_{k=0}^{\infty}(2k+1)\|P_k\|^2t^{2k} &= \frac{1}{1-t}+\frac{1}{1+t}\\ &= \frac{2}{1-t^2}\\ &= 2\left(1+t^2+t^4+t^6+\cdots\right), \quad \text{pour tout } t, |t|<1.\end{aligned}$$

Puisque nous avons une identité en t, on peut identifier les coefficients de t^{2k}:

$$(2k+1)\|P_k\|^2 = 2 \implies \|P_k\|^2 = \frac{2}{2k+1}. \quad \square$$

REMARQUE 6.3. On peut obtenir la fonction génératrice (6.11) en développant le 2$^{\text{e}}$ membre en une série de TAYLOR en puissances de x.

6.5. Série de Fourier–Legendre

On présente des exemples simples de développement en série de FOURIER–LEGENDRE.

EXEMPLE 6.5. Développer le polynôme:

$$p(x) = x^3 - 2x^2 + 4x + 1$$

sur $[-1,1]$ selon les polynômes de LEGENDRE $P_0(x), P_1(x),\ldots$

RÉSOLUTION. On exprime les puissances de x suivant la base des polynômes de LEGENDRE

$$\begin{aligned}P_0(x) = 1 &\implies 1 = P_0(x),\\ P_1(x) = x &\implies x = P_1(x),\\ P_2(x) = \frac{1}{2}(3x^2-1) &\implies x^2 = \frac{2}{3}P_2(x) + \frac{1}{3}P_0(x),\\ P_3(x) = \frac{1}{2}(5x^3-3x) &\implies x^3 = \frac{2}{5}P_3(x) + \frac{3}{5}P_1(x).\end{aligned}$$

On évite ainsi le calcul d'intégrales. Alors

$$\begin{aligned}p(x) &= \frac{2}{5}P_3(x) + \frac{3}{5}P_1(x) - \frac{4}{3}P_2(x) - \frac{2}{3}P_0(x) + 4P_1(x) + P_0(x)\\ &= \frac{2}{5}P_3(x) - \frac{4}{3}P_2(x) + \frac{23}{5}P_1(x) + \frac{1}{3}P_0(x). \quad \square\end{aligned}$$

EXEMPLE 6.6. Développer le polynôme

$$p(x) = 2 + 3x + 5x^2$$

sur $[3,7]$ selon les polynômes de LEGENDRE $P_0(x), P_1(x),\ldots$

RÉSOLUTION. Appliquons $x \in [3,7]$ sur $s \in [-1,1]$ (V. figure 6.5). Posons $s = ax + b$. Alors

$$-1 = 3a + b \quad \text{et} \quad 1 = 7a + b \quad \implies \quad a = \frac{1}{2} \quad \text{et} \quad b = -\frac{5}{2}.$$

On obtient les applications affines suivantes réciproques l'une de l'autre:

$$s = \frac{x-5}{2} \quad \text{et} \quad x = 2s + 5. \tag{6.12}$$

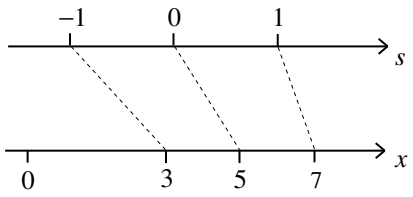

FIGURE 6.5. Application affine de $x \in [3,7]$ sur $s \in [-1,1]$.

Alors

$$\begin{aligned} p(x) &= p(2s+5) \\ &= 2 + 3(2s+5) + 5(2s+5)^2 \\ &= 142 + 106s + 20s^2 \\ &= 142P_0(s) + 106P_1(s) + 20\left[\frac{2}{3}\,P_2(s) + \frac{1}{3}\,P_0(s)\right]. \end{aligned}$$

Donc

$$p(x) = \left(142 + \frac{20}{3}\right) P_0\left(\frac{x-5}{2}\right) + 106P_1\left(\frac{x-5}{2}\right) + \frac{40}{3}P_2\left(\frac{x-5}{2}\right). \quad \square$$

EXEMPLE 6.7. Calculer les trois premiers termes du développement de FOURIER–LEGENDRE de la fonction

$$f(x) = \begin{cases} 0, & -1 < x < 0, \\ x, & 0 < x < 1. \end{cases}$$

RÉSOLUTION. Posons

$$f(x) = \sum_{m=0}^{\infty} a_m P_m(x), \quad -1 < x < 1.$$

Alors,

$$a_m = \frac{2m+1}{2} \int_{-1}^{1} f(x) P_m(x)\, dx.$$

Donc,

$$\begin{aligned} a_0 &= \frac{1}{2}\int_{-1}^{1} f(x)P_0(x)\,dx = \frac{1}{2}\int_0^1 x\,dx = \frac{1}{4}, \\ a_1 &= \frac{3}{2}\int_{-1}^{1} f(x)P_1(x)\,dx = \frac{3}{2}\int_0^1 x^2\,dx = \frac{1}{2}, \\ a_2 &= \frac{5}{2}\int_{-1}^{1} f(x)P_2(x)\,dx = \frac{5}{2}\int_0^1 x\,\frac{1}{2}\,(3x^2-1)\,dx = \frac{5}{16}. \end{aligned}$$

On a donc l'approximation

$$f(x) \approx \frac{1}{4}\,P_0(x) + \frac{1}{2}\,P_1(x) + \frac{5}{16}\,P_2(x). \quad \square$$

EXEMPLE 6.8. Calculer les trois premiers termes du développement de FOURIER–LEGENDRE de la fonction

$$f(x) = e^x, \quad 0 \le x \le 1.$$

RÉSOLUTION. Pour utiliser l'orthogonalité des polynômes de LEGENDRE on transforme le domaine de $f(x)$ de $[0,1]$ à $[-1,1]$ au moyen de la substitution

$$s = 2\left(x - \frac{1}{2}\right), \quad \text{c'est-à-dire} \quad x = \frac{s}{2} + \frac{1}{2}.$$

Alors,

$$f(x) = e^x = e^{(1+s)/2} = \sum_{m=0}^{\infty} a_m P_m(s), \quad -1 \le s \le 1,$$

où

$$a_m = \frac{2m+1}{2} \int_{-1}^{1} e^{(1+s)/2} P_m(s)\, dx.$$

On calcule d'abord trois intégrales par récurrence:

$$\begin{aligned}
I_0 &= \int_{-1}^{1} e^{s/2}\, ds = 2\left(e^{1/2} - e^{-1/2}\right), \\
I_1 &= \int_{-1}^{1} s\, e^{s/2}\, ds = 2s\, e^{s/2}\Big|_{-1}^{1} - 2\int_{-1}^{1} e^{s/2}\, ds \\
&= 2\left(e^{1/2} + e^{-1/2}\right) - 2I_0 \\
&= -2\, e^{1/2} + 6\, e^{-1/2}, \\
I_2 &= \int_{-1}^{1} s^2\, e^{s/2}\, ds = 2s^2\, e^{s/2}\Big|_{-1}^{1} - 4\int_{-1}^{1} s\, e^{s/2}\, ds \\
&= 2\left(e^{1/2} + e^{-1/2}\right) - 4I_1 \\
&= 10\, e^{1/2} - 26\, e^{-1/2}.
\end{aligned}$$

Il suit que

$$\begin{aligned}
a_0 &= \frac{1}{2}\, e^{1/2} I_0 = e - 1 \approx 1.7183, \\
a_1 &= \frac{3}{2}\, e^{1/2} I_1 = -3\, e + 9 \approx 0.8452, \\
a_2 &= \frac{5}{2}\, e^{1/2} \frac{1}{2}(3I_2 - I_0) = 35\, e - 95 \approx 0.1399.
\end{aligned}$$

On a donc l'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. Une application: la quadrature gaussienne

On obtient facilement les formules de quadrature gaussienne à n points au moyen des polynômes de LEGENDRE. Par souci de simplicité, on ne considère que $n = 2$ et $n = 3$. On précise que le nombre n de points se réfère aux n points en lesquels on évalue la fonction à intégrer sur l'intervalle standardisé $[-1,1]$, et non pas au nombre de sous-intervalles en lesquels on a l'habitude de diviser l'intervalle d'intégration $[a,b]$ afin de diminuer l'erreur de la valeur numérique de l'intégrale.

EXEMPLE 6.9. Déterminer les 4 paramètres de la quadrature gaussienne à 2 points:

$$\int_{-1}^{1} f(x)\,dx = af(x_1) + bf(x_2).$$

RÉSOLUTION. Par symétrie, on prévoit que les nœuds sont opposés, $x_1 = -x_2$, et les poids sont égaux, $a = b$. Puisqu'on a 4 paramètres, la formule est exacte pour les polynômes de degré 3 et, par l'exemple 6.5, il suffit de considérer les polynômes $P_0(x), \ldots, P_3(x)$. Comme $P_0(x) = 1$ est orthogonal à $P_n(x)$, $n = 1, 2, \ldots$, on a

$$2 = \int_{-1}^{1} P_0(x)\,dx = aP_0(x_1) + bP_0(x_2) = a + b, \tag{6.13}$$

$$0 = \int_{-1}^{1} 1 \times P_1(x)\,dx = aP_1(x_1) + bP_1(x_2) = ax_1 + bx_2, \tag{6.14}$$

$$0 = \int_{-1}^{1} 1 \times P_2(x)\,dx = aP_2(x_1) + bP_2(x_2), \tag{6.15}$$

$$0 = \int_{-1}^{1} 1 \times P_3(x)\,dx = aP_3(x_1) + bP_3(x_2), \tag{6.16}$$

Pour satisfaire (6.15) on choisit x_1 et x_2 tels que

$$P_2(x_1) = P_2(x_2) = 0,$$

c'est-à-dire:

$$P_2(x) = \frac{1}{2}(3x^2 - 1) = 0 \Rightarrow -x_1 = x_2 = \frac{1}{\sqrt{3}} = 0.577\,350\,27.$$

Alors, par (6.14)

$$a = b.$$

De plus, (6.16) est automatiquement satisfaite car $P_3(x)$ est impair. Enfin, par (6.13),

$$a = b = 1.$$

On a donc la formule de GAUSS à deux points:

$$\int_{-1}^{1} f(x)\,dx = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right). \quad \square \tag{6.17}$$

EXEMPLE 6.10. Déterminer les 6 paramètres de la quadrature gaussienne à 3 points:

$$\int_{-1}^{1} f(x)\,dx = af(x_1) + bf(x_2) + cf(x_3).$$

RÉSOLUTION. Par symétrie, on prévoit que les nœuds extrêmes sont opposés, $x_1 = -x_3$, et $x_2 = 0$, les poids extrêmes sont égaux, $a = c$, et le poids central dépasse les deux autres, $b > a = c$. Puisqu'on a 6 paramètres, la formule est exacte pour les polynômes de degré 5 et, par l'exemple 6.5, il suffit de considérer

la base $P_0(x), \ldots, P_5(x)$:

$$2 = \int_{-1}^{1} P_0(x)\,dx = aP_0(x_1) + bP_0(x_2) + cP_0(x_3), \tag{6.18}$$

$$0 = \int_{-1}^{1} P_1(x)\,dx = aP_1(x_1) + bP_1(x_2) + cP_1(x_3), \tag{6.19}$$

$$0 = \int_{-1}^{1} P_2(x)\,dx = aP_2(x_1) + bP_2(x_2) + cP_2(x_3), \tag{6.20}$$

$$0 = \int_{-1}^{1} P_3(x)\,dx = aP_3(x_1) + bP_3(x_2) + cP_3(x_3), \tag{6.21}$$

$$0 = \int_{-1}^{1} P_4(x)\,dx = aP_4(x_1) + bP_4(x_2) + cP_4(x_3), \tag{6.22}$$

$$0 = \int_{-1}^{1} P_5(x)\,dx = aP_5(x_1) + bP_5(x_2) + cP_5(x_3). \tag{6.23}$$

Pour satisfaire (6.21), prenons pour x_1, x_2, x_3 les 3 zéros de

$$P_3(x) = \frac{1}{2}(5x^3 - 3x) = \frac{1}{2}x(5x^2 - 3),$$

c'est-à-dire

$$-x_1 = x_3 = \sqrt{\frac{3}{5}} = 0.774\,596\,7, \quad x_2 = 0.$$

Alors (6.19) implique

$$-\sqrt{\frac{3}{5}}\,a + \sqrt{\frac{3}{5}}\,c = 0 \Rightarrow a = c;$$

donc (6.23) est satisfaite puisque $P_5(x)$ est impair. De plus, par la substitution $a = c$ dans (6.20), on obtient

$$a\,\frac{1}{2}\left(3 \times \frac{3}{5} - 1\right) + b\left(-\frac{1}{2}\right) + a\,\frac{1}{2}\left(3 \times \frac{3}{5} - 1\right) = 0,$$

c'est-à-dire

$$4a - 5b + 4a = 0 \quad \text{ou} \quad 8a - 5b = 0. \tag{6.24}$$

Maintenant, de (6.18) on déduit

$$2a + b = 2 \quad \text{ou} \quad 10a + 5b = 10. \tag{6.25}$$

Si l'on additionne la seconde expression de (6.24) et de (6.25), on obtient

$$a = \frac{10}{18} = \frac{5}{9} = 0.555;$$

alors

$$b = 2 - \frac{10}{9} = \frac{8}{9} = 0.888.$$

Enfin, on vérifie que (6.22) est satisfaite. Puisque

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3),$$

on a

$$2\times\frac{5\times1}{9\times8}\left(35\times\frac{9}{25}-30\times\frac{3}{5}+3\right)+\frac{8}{9}\times\frac{3}{8}=\frac{2\times5}{9\times8}\left(\frac{315-450+75}{25}\right)+\frac{8}{9}\times\frac{3}{8}$$
$$=\frac{2\times5}{9\times8}\times\frac{(-60)}{25}+\frac{8\times3}{9\times8}$$
$$=\frac{-24+24}{9\times8}=0.$$

On a donc la formule de Gauss à trois points:

$$\int_{-1}^{1}f(x)\,dx=\frac{5}{9}f\left(-\sqrt{\frac{3}{5}}\right)+\frac{8}{9}f(0)+\frac{5}{9}f\left(\sqrt{\frac{3}{5}}\right). \quad \square \tag{6.26}$$

Remarque 6.4. On a normalisé l'intervalle d'intégration des quadratures de Gauss sur $[-1,1]$. Pour intégrer sur $[a,b]$ on emploie le changement de variable déjà utilisée à l'exemple 6.8 :

$$x=\frac{(b-a)t+b+a}{2},\quad dx=\left(\frac{b-a}{2}\right)dt.$$

Alors

$$\int_a^b f(x)\,dx=\frac{b-a}{2}\int_{-1}^{1}f\left(\frac{(b-a)t+b+a}{2}\right)dt.$$

Exemple 6.11. Évaluer

$$I=\int_0^{\pi/2}\sin x\,dx$$

par la formule de Gauss à 2 point appliquée une seule fois sur tout l'intervalle $[0,\pi/2]$ et sur les demi-intervalles $[0,\pi/4]$ et $[\pi/4,\pi/2]$.

Résolution. Posons

$$x=\frac{(\pi/2)t+\pi/2}{2},\qquad dx=\frac{\pi}{4}\,dt.$$

En $t=-1$, $x=0$ et, en $t=1$, $x=\pi/2$. Alors

$$I=\frac{\pi}{4}\int_{-1}^{1}\sin\left(\frac{\pi t+\pi}{4}\right)dt$$
$$\approx\frac{\pi}{4}\left[1.0\times\sin\left(0.105\,66\pi\right)+1.0\times\sin\left(0.394\,34\pi\right)\right]$$
$$=0.998\,47.$$

L'erreur est 1.53×10^{-3}. Sur les demi-intervalles, on a

$$I=\frac{\pi}{8}\int_{-1}^{1}\sin\left(\frac{\pi t+\pi}{8}\right)dt+\frac{\pi}{8}\int_{-1}^{1}\sin\left(\frac{\pi t+3\pi}{8}\right)dt$$
$$\approx\frac{\pi}{8}\left[\sin\frac{\pi}{8}\left(-\frac{1}{\sqrt{3}}+1\right)+\sin\frac{\pi}{8}\left(\frac{1}{\sqrt{3}}+1\right)\right.$$
$$\left.+\sin\frac{\pi}{8}\left(-\frac{1}{\sqrt{3}}+3\right)+\sin\frac{\pi}{8}\left(\frac{1}{\sqrt{3}}+3\right)\right]$$
$$=0.999\,910\,166\,769\,89.$$

L'erreur est 8.983×10^{-5}. Voici une solution obtenue par Matlab. Par soucis de généralité, définissons la fonction fichier M `exp5_10.m`,

```
function f=exp5_10(t)
% evaluate the function f(t)
f=sin(t);
```

Voici le programme de la quadrature gaussienne à deux points:

```
>> 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)' % valeur numerique de l'integrale
 nv1 = 0.9985
>> error1 = 1 - nv1 % l'erreur dans la solution
 error1 = 0.0015
```

On obtient l'autre partie de la même façon. □

REMARQUE 6.5. La quadrature gaussienne est la formule la plus précise pour un nombre de points donné. L'erreur pour la formule à n points est

$$E_n(f) = \frac{2}{(2n+1)!}\left[\frac{2^n(n!)^2}{(2n)!}\right]^2 f^{(2n)}(\xi), \quad -1 < \xi < 1.$$

La formule à n points est donc exacte pour un polynôme de degré $2n-1$.

Matlab's adaptive Simpson's rule `quad` and adaptive Newton–Cotes 8-panel rule `quad8` evaluate the integral of Example 6.11 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}.

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 addresses this problem by discovering where the integrand is ill behaved and shortening the subintervals accordingly.

6.7. Résolution numérique d'équations intégrales de 2e espèce

The theory and application of integral equations is an important subject in applied mathematics, science and engineering. In this section we restrict attention to Fredholm integral equations of the second kind in one variable. The general form of such equation is

$$f(t) = \lambda \int_a^b K(t,s) f(s)\, ds + g(t), \qquad \lambda \neq 0. \tag{6.27}$$

We shall assume that the kernel $K(t,s)$ is continuous on the square $[a,b]\times[a,b]\in\mathbb{R}^2$.

A significant use of Gaussian quadrature formulae is in the numerical solution of Fredholm integral equations of the second kind by the Nyström method. We explain this method.

Let a numerical integration scheme be given:

$$\int_a^b y(s)\,ds \approx \sum_{j=1}^{N} w_j y(s_j), \tag{6.28}$$

where the N numbers $\{w_j\}$ are the weights of the quadrature rule and the N points $\{s_j\}$ are the nodes used by the method. One may use the trapezoidal or Simpson's rules, but for smooth nonsingular problems Gaussian quadrature seems by far superior.

If we apply the numerical integration scheme to the integral equation (6.27), we get

$$f(t) = \lambda \sum_{j=1}^{N} w_j K(t,s_j) f(s_j) + g(t), \tag{6.29}$$

where, for simplicity, we have written $f(t)$ for $f_N(t)$. We evaluate this equation at the quadrature points:

$$f(t_j) = \lambda \sum_{j=1}^{N} w_j K(t_j,s_j) f(s_j) + g(t_j). \tag{6.30}$$

Let f_i be the vector $f(t_i)$, g_i the vector $g(t_i)$, K_{ij} the matrix $K(t_i,s_j)$, and define

$$\widetilde{K}_{ij} = K_{ij} w_j.$$

Then, in matrix notation, the previous equation becomes

$$(I - \lambda \widetilde{K})\boldsymbol{f} = \boldsymbol{g}. \tag{6.31}$$

This is a set of N linear algebraic equations in N unknowns that can be solved by the LU decomposition (see Chapter 7).

Having obtained the solution at the quadrature points $\{t_i\}$, how do we get the solution at some other point t? We do not simply use polynomial interpolation since this destroys the accuracy we worked hard to achieve. Nyström's key observation is to use (6.29) as an interpolatory formula which maintains the accuracy of the solution.

In Example 6.12, we compare the performance of the three-point Simpson rule and three-point Gaussian quadrature, respectively.

Exemple 6.12. Consider the integral equation

$$f(t) = \lambda \int_0^1 e^{ts} f(s)\,ds + g(t), \qquad 0 \le t \le 1, \tag{6.32}$$

with $\lambda = 0.5$ and $f(t) = e^t$. Compare the errors in the numerical solutions at the nodes of Simpson's rule and three-point Gaussian quadrature, respectively.

SOLUTION. Substituting $f(t) = e^t$ in the integral equation, we see that the function $g(t)$ on the right-hand side is

$$g(t) = e^t - \frac{1}{2(t+1)} \left(1 - e^{t+1}\right).$$

This is easily obtained by the symbolic Matlab commands

```
>> clear; syms s t; lambda = 1/2;
>> g = exp(t)-lambda*int(exp((t+1)*s),s,0,1)
 g = exp(t)-1/2/(t+1)*exp(t+1)+1/2/(t+1)
```

Applying Simpson's rule to equation (6.32), with nodes

$$t_1 = 0, \qquad t_2 = 0.5, \qquad t_3 = 1,$$

and solving the resulting algebraic system (6.31), say, by the LU decomposition, we have the error in the solution f_3:

$$\begin{bmatrix} f(0) \\ f(0.5) \\ f(1) \end{bmatrix} - \begin{bmatrix} f_3(0) \\ f_3(0.5) \\ f_3(1) \end{bmatrix} = \begin{bmatrix} -0.0047 \\ -0.0080 \\ -0.0164 \end{bmatrix}$$

Applying the three-point Gaussian quadrature to equation (6.32), with nodes

$$t_1 = \frac{1 - \sqrt{0.6}}{2} \approx 0.112\,701\,67, \qquad t_2 = 0.5, \qquad t_3 = \frac{1 + \sqrt{0.6}}{2} \approx 0.887\,298\,33,$$

and solving the resulting algebraic system (6.31), say, by the LU decomposition, we have the error in the solution f_3:

$$\begin{bmatrix} f(t_1) \\ f(t_2) \\ f(t_3) \end{bmatrix} - \begin{bmatrix} f_3(t_1) \\ f_3(t_2) \\ f_3(t_3) \end{bmatrix} = \begin{bmatrix} 0.2099 \times 10^{-4} \\ 0.3195 \times 10^{-4} \\ 0.6315 \times 10^{-4} \end{bmatrix} \tag{6.33}$$

which is much smaller than with Simpson's rule when using the same number of nodes.

The function M-file `exp5_11.m`:

```
function g=exp5_11(t)
% evaluate right-hand side
global lambda
syms s
g = exp(t)-lambda*int(exp(t*s)*exp(s),s,0,1);
\end{varbatim}
computes the value of the function $g(t)$, and
the following Matlab commands produce these results.
\begin{verbatim}
clear; global lambda
lambda = 1/2; h = 1/2;
snode = [0 1/2 1]; sweight = [1/3 4/3 1/3]; % Simpson's rule
sK = h*exp(snode'*snode)*diag(sweight);
sA = eye(3)-lambda*sK;
sb = double(exp5_11(snode)');
sf3 = sA\sb;
serror = exp(snode)-sf3'
 serror =
```

TABLE 1. Nyström–trapezoidal method in Example 6.13.

N	E_1	Ratio	E_2	Ratio
2	5.35E–03		5.44E–3	
4	1.35E–03	3.9	1.37E–03	4.0
8	3.39E–04	4.0	3.44E–04	4.0
16	8.47E–05	4.0	8.61E–05	4.0

```
    -0.0047   -0.0080   -0.0164
gnode = [(1-sqrt(0.6))/2 1/2 (1+sqrt(0.6))/2]; % Gaussian quadrature
gweight = [5/18 8/18 5/18];
gK = exp(gnode'*gnode)*diag(gweight);
gA = eye(3)-lambda*gK;
gb = double(exp5_11(gnode)');
gf3 = gA\gb;
gerror = exp(gnode)-gf3'
 gerror = 1.0e-04 *
    0.2099    0.3195    0.6315
```

Note that the use of matrices in computing `sK` and `gK` avoids recourse to loops.

Quadratic interpolation can be used to extend the numerical solution to all other $t \in [0,1]$, but it generally results in a much larger error. For example,

$$f(1.0) - P_2 f_3(1.0) = 0.0158,$$

where $P_2 f_3(t)$ denotes the quadratic polynomial interpolating the Nyström solution at the Gaussian quadrature nodes given above. In contrast, the Nyström formula (6.29) gives errors that are consistent in size with those in (6.33). For example,

$$f(1.0) - f_3(1.0) = 8.08 \times 10^{-5}. \quad \square$$

EXEMPLE 6.13. Consider the integral equation of Example 6.12 with $\lambda = 1/50$ and $f(t) = e^t$. Compare the errors in the Nyström–trapezoidal method and Nyström–Gaussian method, respectively.

SOLUTION. In Table 1 we give numerical results when using the trapezoidal rule with n nodes, with $N = 2, 4, 8, 16$. In Table 2 we give results when using n-point Gaussian quadratures for $N = 1, 2, 3, 4, 5$. The following norms are used

$$E_1 = \max_{1 \le i \le N} |f(t_i) - f_N(t_i)|, \qquad E_2 = \max_{0 \le t \le 1} |f(t) - f_N(t)|.$$

For E_2, $f_N(t)$ is obtained using the Nyström interpolation formula (6.29). The results for the trapezoidal rule show clearly the $O(h^2)$ behavior of the error. It is seen that the use of Gaussian quadrature leads to very rapid convergence of f_N to $f(x)$. $\square$

TABLE 2. Nyström–Gaussian method in Example 6.13.

N	E_1	Ratio	E_2	Ratio
1	4.19E–03		9.81E–03	
2	1.22E–04	34	2.18E–04	45
3	1.20E–06	100	1.86E–06	117
4	5.09E–09	200	8.47E–09	220
5	1.74E–11	340	2.39E–11	354

CHAPITRE 7

Calcul matriciel

With the advent of digitized systems in many areas of science and engineering, matrix computation is occupying a central place in modern computer software. In this chapter, we study the solutions of linear systems,

$$Ax = b, \quad A \in \mathbb{R}^{m\times n},$$

and eigenvalue problems,

$$Ax = \lambda x, \quad A \in \mathbb{R}^{n\times n}, \quad x \neq 0,$$

as implemented in softwares, where accuracy, stability and algorithmic complexity are of the utmost importance.

7.1. Solution LU de $Ax = b$

The solution of a linear system

$$Ax = b, \quad A \in \mathbb{R}^{n\times n},$$

with partial pivoting on rows is obtained by the LU decomposition of A,

$$A = LU,$$

where L is a row-permutation of a lower triangular matrix M with $m_{ii} = 1$ and $|m_{ij}| \leq 1$, for $i > j$, and U is an upper triangular matrix. Thus the system becomes

$$LUx = b.$$

The solution is obtained in two steps. First,

$$Ly = b$$

is solved for y by forward substitution and, second,

$$Ux = y$$

is solved for x by backward substitution. The following example illustrates the above steps.

EXEMPLE 7.1. Solve the system $Ax = b$,

$$\begin{bmatrix} 3 & 9 & 6 \\ 18 & 48 & 39 \\ 9 & -27 & 42 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 23 \\ 136 \\ 45 \end{bmatrix}$$

by the LU decomposition **with pivoting**.

SOLUTION. Since $a_{21} = 18$ is the largest pivot in absolute value in the first column of A,

$$|18| > |3|, \quad |18| > |9|,$$

we interchange the second and first rows of A,

$$P_1 A = \begin{bmatrix} 18 & 48 & 39 \\ 3 & 9 & 6 \\ 9 & -27 & 42 \end{bmatrix}, \quad P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We now apply a Gaussian transformation on $P_1 A$ to put zeros under 18 in the first column,

$$\begin{bmatrix} 1 & 0 & 0 \\ -1/6 & 1 & 0 \\ -1/2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 18 & 48 & 39 \\ 3 & 9 & 6 \\ 9 & -27 & 42 \end{bmatrix} = \begin{bmatrix} 18 & 48 & 39 \\ 0 & 1 & -1/2 \\ 0 & -51 & 45/2 \end{bmatrix},$$

with multipliers $-1/6$ and $-1/2$. Thus

$$M_1 P_1 A = A_1.$$

Considering the 2×2 submatrix

$$\begin{bmatrix} 1 & -1/2 \\ -51 & 45/2 \end{bmatrix},$$

we see that -51 is the pivot in the first column since

$$|-51| > |1|.$$

Hence we interchange the third and second row,

$$P_2 A_1 = \begin{bmatrix} 18 & 48 & 39 \\ 0 & -51 & 45/2 \\ 0 & 1 & -1/2 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

To zero the $(3,2)$ element we apply a Gaussian transformation M_2 on $P_2 A_1$,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/51 & 1 \end{bmatrix} \begin{bmatrix} 18 & 48 & 39 \\ 0 & -51 & 45/2 \\ 0 & 1 & -1/2 \end{bmatrix} = \begin{bmatrix} 18 & 48 & 39 \\ 0 & -51 & 22.5 \\ 0 & 0 & -0.0588 \end{bmatrix} = U,$$

where $1/51$ is the multiplier. Thus

$$M_2 P_2 A_1 = U.$$

Therefore

$$M_2 P_2 M_1 P_1 A = U,$$

and

$$A = P_1^{-1} M_1^{-1} P_2^{-1} M_2^{-1} U.$$

The inverse of a Gaussian transformation is easily written:

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ -b & 0 & 1 \end{bmatrix} \quad \Longrightarrow \quad M_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & 0 & 1 \end{bmatrix},$$

$$M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -c & 1 \end{bmatrix} \quad \Longrightarrow \quad M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix},$$

once the multipliers $-a$, $-b$, $-c$ are known. Moreover the product $M_1^{-1}M_2^{-1}$ can be easily written:

$$M_1^{-1}M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{bmatrix}.$$

It is easily seen that a permutation P, which consists of the identity matrix I with permuted rows, is an orthogonal matrix. Hence,

$$P^{-1} = P^T.$$

Therefore, if

$$L = P_1^T M_1^{-1} P_2^T M_2^{-1},$$

then

$$\begin{aligned} L &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1/6 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/51 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1/6 & 1 & 0 \\ 1 & 0 & 0 \\ 1/2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/51 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/6 & -1/51 & 1 \\ 1 & 0 & 0 \\ 1/2 & 1 & 0 \end{bmatrix} \end{aligned}$$

which is the row-permutation of a lower triangular matrix, that is, it becomes lower triangular if the second and third rows are interchanged, and then the new third row is interchanged with the first row, namely, P_2P_1L is lower triangular.

The system

$$Ly = b$$

is solved by forward substitution

$$\begin{bmatrix} 1/6 & -1/51 & 1 \\ 1 & 0 & 0 \\ 1/2 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 23 \\ 136 \\ 45 \end{bmatrix},$$

$$\begin{aligned} &y_1 = 136, \\ &y_2 = 45 - 136/2 = -23, \\ &y_3 = 23 - 136/6 - 23/51 = -0.1176. \end{aligned}$$

Finally, the system

$$Ux = y$$

is solved by backward substitution

$$\begin{bmatrix} 18 & 48 & 39 \\ 0 & -51 & 22.5 \\ 0 & 0 & -0.0588 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 136 \\ -23 \\ -0.1176 \end{bmatrix},$$

$$\begin{aligned} &x_3 = 0.1176/0.0588 = 2, \\ &x_2 = (-23 - 22.5 \times 2)/(-51) = 1.3333, \\ &x_1 = (136 - 48 \times 1.3333 - 39 \times 2)/18 = -0.3333. \end{aligned}$$

□

The following Matlab session does exactly that.

```
>> A = [3 9 6; 18 48 39; 9 -27 42]

A =
     3     9     6
    18    48    39
     9   -27    42

>> [L,U] = lu(A)

L =
    0.1667   -0.0196    1.0000
    1.0000         0         0
    0.5000    1.0000         0

U =
   18.0000   48.0000   39.0000
         0  -51.0000   22.5000
         0         0   -0.0588

>> b = [23; 136; 45]

b =
    23
   136
    45

>> y = L\b   % forward substitution

y =
  136.0000
  -23.0000
   -0.1176

>> x = U\y   % backward substitution

x =
   -0.3333
    1.3333
    2.0000

>> z = A\b   % Matlab left-inverse to solve Az = b by the LU decomposition

z =
   -0.3333
    1.3333
    2.0000
```

EXEMPLE 7.2. Given

$$A = \begin{bmatrix} 3 & 2 & 0 \\ 12 & 13 & 6 \\ -3 & 8 & 9 \end{bmatrix}, \qquad b = \begin{bmatrix} 14 \\ 40 \\ -28 \end{bmatrix},$$

find the LU decomposition of A **without pivoting** and solve

$$Ax = b.$$

SOLUTION. For $M_1 A = A_1$, we have

$$\begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 0 \\ 12 & 13 & 6 \\ -3 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 0 \\ 0 & 5 & 6 \\ 0 & 10 & 9 \end{bmatrix}.$$

For $M_2 A_1 = U$, we have

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 0 \\ 0 & 5 & 6 \\ 0 & 10 & 9 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 0 \\ 0 & 5 & 6 \\ 0 & 0 & -3 \end{bmatrix} = U,$$

that is

$$M_2 M_1 A = U, \quad A = M_1^{-1} M_2^{-1} U = LU.$$

Thus

$$L = M_1^{-1} M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix}.$$

Forward substitution is used to obtain y from $Ly = b$,

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 40 \\ -28 \end{bmatrix};$$

thus

$$\begin{aligned} y_1 &= 14, \\ y_2 &= 40 - 56 = -16, \\ y_3 &= -28 + 14 + 32 = 18. \end{aligned}$$

Finally, backward substitution is used to obtain x from $Ux = y$,

$$\begin{bmatrix} 3 & 2 & 0 \\ 0 & 5 & 6 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ -16 \\ 18 \end{bmatrix};$$

thus

$$\begin{aligned} x_3 &= -6, \\ x_2 &= (-16 + 36)/5 = 4, \\ x_1 &= (14 - 8)/3 = 2. \end{aligned}$$

□

We note that, without pivoting, $|l_{ij}|$, $i > j$, may be larger than 1.

The LU decomposition without partial pivoting is an unstable procedure which may lead to large errors in the solution. In practice, partial pivoting is usually stable. However, in some cases, one needs to resort to complete pivoting on rows and columns to ensure stability, or to use the stable QR decomposition.

Sometimes it is useful to scale the rows or columns of the matrix of a linear system before solving it. This may alter the choice of the pivots. In practice, one has to consider the meaning and physical dimensions of the unknown variables to decide upon the type of scaling or balancing of the matrix. Softwares provide some of these options. Scaling in the l_∞-norm is used in the following example.

EXEMPLE 7.3. Scale each equation in the l_∞-norm, so that the largest coefficient of each row on the left-hand side is equal to 1 in absolute value, and solve the following system:

$$\begin{array}{rcrcr} 30.00x_1 & + & 591400x_2 & = & 591700 \\ 5.29x_1 & - & 6.130x_2 & = & 46.70 \end{array}$$

by the LU decomposition with pivoting with four-digit arithmetic.

SOLUTION. Dividing the first equation by

$$s_1 = \max\{|30.00|, |591400|\} = 591400$$

and the second equation by

$$s_2 = \max\{|5.291|, |6.130|\} = 6.130,$$

we find that

$$\frac{|a_{11}|}{s_1} = \frac{30.00}{591400} = 0.5073 \times 10^{-4}, \qquad \frac{|a_{21}|}{s_2} = \frac{5.291}{6.130} = 0.8631.$$

Hence the scaled pivot is in the second equation. Note that the scaling is done only for comprison purposes, so the division to determine the scaled pivots produces no round-off error in solving the system. Thus the LU decomposition applied to the interchanged system

$$\begin{array}{rcrcr} 5.29x_1 & - & 6.130x_2 & = & 46.70 \\ 30.00x_1 & + & 591400x_2 & = & 591700 \end{array}$$

produces the correct results:

$$x_1 = 10.00, \qquad x_2 = 1.000.$$

Note that the LU decomposition with four-digit arithmetic applied to the non-interchanged system produces the erroneous results $x_1 \approx -10.00$ and $x_2 \approx 1.001$. □

The following Matlab function M-files are found in `ftp://ftp.cs.cornell.edu/pub/cv`. The forward substitution algorithm solves a lower triangular system:

```
  function x = LTriSol(L,b)
%
% Pre:
%   L   n-by-n nonsingular lower triangular matrix
%   b   n-by-1
%
```

```
% Post:
%   x   Lx = b

   n = length(b);
   x = zeros(n,1);
   for j=1:n-1
      x(j) = b(j)/L(j,j);
      b(j+1:n) = b(j+1:n) - L(j+1:n,j)*x(j);
   end
   x(n) = b(n)/L(n,n);
```

The backward substitution algorithm solves a upper triangular system:

```
  function x = UTriSol(U,b)
%
% Pre:
%   U   n-by-n nonsingular upper triangular matrix
%   b   n-by-1
%
% Post:
%   x   Lx = b

   n = length(b);
   x = zeros(n,1);
   for j=n:-1:2
      x(j) = b(j)/U(j,j);
      b(1:j-1) = b(1:j-1) - x(j)*U(1:j-1,j);
   end
   x(1) = b(1)/U(1,1);
```

The LU decomposition without pivoting is performed by the following function.

```
  function [L,U] = GE(A);
%
% Pre:
%   A      n-by-n
%
% Post:
%   L      n-by-n unit lower triangular with |L(i,j)|<=1.
%   U      n-by-n upper triangular.
%          A = LU

    [n,n] = size(A);
    for k=1:n-1
       A(k+1:n,k) = A(k+1:n,k)/A(k,k);
       A(k+1:n,k+1:n) = A(k+1:n,k+1:n) - A(k+1:n,k)*A(k,k+1:n);
    end
        L = eye(n,n) + tril(A,-1);
    U = triu(A);
```

The LU decomposition with pivoting is performed by the following function.

```
  function [L,U,piv] = GEpiv(A);
```

```
%
% Pre:
%   A       n-by-n
%
% Post:
%   L       n-by-n unit lower triangular with |L(i,j)|<=1.
%   U       n-by-n upper triangular
%   piv     integer n-vector that is a permutation of 1:n.
%
%           A(piv,:) = LU

    [n,n] = size(A);
        piv = 1:n;
    for k=1:n-1
       [maxv,r] = max(abs(A(k:n,k)));
       q = r+k-1;
           piv([k q]) = piv([q k]);
           A([k q],:) = A([q k],:);
       if A(k,k) ~= 0
          A(k+1:n,k) = A(k+1:n,k)/A(k,k);
          A(k+1:n,k+1:n) = A(k+1:n,k+1:n) - A(k+1:n,k)*A(k,k+1:n);
       end
    end
        L = eye(n,n) + tril(A,-1);
    U = triu(A);
```

7.2. La décomposition de Cholesky

The important class of positive definite symmetric matrices admits the Cholesky decomposition

$$A = GG^T$$

where G is lower triangular.

DÉFINITION 7.1. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is said to be positive definite if

$$x^T Ax > 0, \quad \text{for all } x \neq 0, \quad x \in \mathbb{R}^n.$$

In that case we write $A > 0$.

A symmetric matrix A is positive definite if and only if all its eigenvalues λ,

$$Ax = \lambda x, \quad x \neq 0,$$

are positive, $\lambda > 0$.

A matrix A is positive definite if and only if all its principal minors are positive. For example,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} > 0$$

if and only if

$$\det a_{11} = a_{11} > 0, \quad \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} > 0, \quad \det A > 0.$$

If $A > 0$, then $a_{ii} > 0, i = 1, 2, \ldots, n$.

An $n \times n$ matrix A is *diagonally dominant* if

$$|a_{ii}| > |a_{i1}| + |a_{i2}| + \cdots + |a_{i,i-1}| + |a_{i,i+1}| + \cdots + |a_{in}|, \qquad i = 1, 2, \ldots, n.$$

A diagonally dominant matrix with positive diagonal entries is positive definite.

THÉORÈME 7.1. *If A is positive definite, the Cholesky decomposition*

$$A = GG^T$$

does not require any pivoting, and hence $Ax = b$ can be solved by the Cholesky decomposition without pivoting,

$$Gy = b, \quad G^T x = y.$$

EXEMPLE 7.4. Let

$$A = \begin{bmatrix} 4 & 6 & 8 \\ 6 & 34 & 52 \\ 8 & 52 & 129 \end{bmatrix}, \qquad b = \begin{bmatrix} 0 \\ -160 \\ -452 \end{bmatrix}.$$

Find the Cholesky decomposition of A and use this decomposition to solve the system

$$Ax = b.$$

SOLUTION. The Cholesky decomposition is obtained (without pivoting) by solving the following system for g_{ij}:

$$\begin{bmatrix} g_{11} & 0 & 0 \\ g_{21} & g_{22} & 0 \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \begin{bmatrix} g_{11} & g_{21} & g_{31} \\ 0 & g_{22} & g_{32} \\ 0 & 0 & g_{33} \end{bmatrix} = \begin{bmatrix} 4 & 6 & 8 \\ 6 & 34 & 52 \\ 8 & 52 & 129 \end{bmatrix}.$$

$$\begin{aligned} g_{11}^2 = 4 \qquad &\Longrightarrow \ g_{11} = 2 > 0, \\ g_{11}g_{21} = 6 \qquad &\Longrightarrow \ g_{21} = 3, \\ g_{11}g_{31} = 8 \qquad &\Longrightarrow \ g_{31} = 4, \\ g_{21}^2 + g_{22}^2 = 34 \qquad &\Longrightarrow \ g_{22} = 5 > 0, \\ g_{21}g_{31} + g_{22}g_{32} = 52 \qquad &\Longrightarrow \ g_{32} = 8, \\ g_{31}^2 + g_{32}^2 + g_{33}^2 = 129 \qquad &\Longrightarrow \ g_{33} = 7 > 0. \end{aligned}$$

Hence

$$G = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 5 & 0 \\ 4 & 8 & 7 \end{bmatrix},$$

$$\det A = \det G \det G^T = (\det G)^2 = (2 \times 5 \times 7)^2 > 0.$$

Solving $Gy = b$ by forward substitution,

$$\begin{bmatrix} 2 & 0 & 0 \\ 3 & 5 & 0 \\ 4 & 8 & 7 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -160 \\ -452 \end{bmatrix},$$

we have

$$\begin{aligned} y_1 &= 0, \\ y_2 &= -32, \\ y_3 &= (-452 + 256)/7 = -28. \end{aligned}$$

Solving $G^T x = y$ by backward substitution,

$$\begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 8 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -32 \\ -28 \end{bmatrix},$$

we have

$$\begin{aligned} x_3 &= -4, \\ x_2 &= (-32 + 32)/5 = 0, \\ x_1 &= (0 - 3 \times 0 + 16)/2 = 8. \end{aligned}$$

□

The numeric Matlab command `chol` find the Cholesky decomposition $R^T R$ of the symmetric matrix A as follows.

```
>>  A = [4 6 8;6 34 52;8 52 129];
>> R = chol(A)
R =
     2     3     4
     0     5     8
     0     0     7
```

The following Matlab function M-files are found in `ftp://ftp.cs.cornell.edu/pub/cv`. They are introduced here to illustrate the different levels of matrix-vector multiplications.

The simplest "scalar" Cholesky decomposition is obtained by the following function.

```
  function G = CholScalar(A);
%
% Pre:  A is a symmetric and positive definite matrix.
% Post: G is lower triangular and A = G*G'.

   [n,n] = size(A);
   G = zeros(n,n);
   for i=1:n
      % Compute G(i,1:i)
      for j=1:i
             s = A(j,i);
                 for k=1:j-1
                    s = s - G(j,k)*G(i,k);
                 end
                 if j<i
                    G(i,j) = s/G(j,j);
             else
                G(i,i) = sqrt(s);
```

```
                end
        end
   end
```

The dot product of two vectors returns a scalar, $c = x^T y$. Noticing that the k-loop in `CholScalar` oversees an inner product between subrows of G, we obtain the following level-1 dot product implementation.

```
  function G = CholDot(A);
%
% Pre:  A is a symmetric and positive definite matrix.
% Post: G is lower triangular and A = G*G'.

   [n,n] = size(A);
   G = zeros(n,n);
   for i=1:n
      % Compute G(i,1:i)
      for j=1:i
             if j==1
                s = A(j,i);
             else
                    s = A(j,i) - G(j,1:j-1)*G(i,1:j-1)';
                end
                if j<i
                   G(i,j) = s/G(j,j);
             else
                G(i,i) = sqrt(s);
                end
          end
   end
```

An update of the form

$$\text{vector} \leftarrow \text{vector} + \text{vector} \cdot \text{scalar}$$

is called a *saxpy* operation, which stands for "scalar a times x plus y", that is $y = Ax + y$.A column-orientation version that features the saxpy operation is the following implementation.

```
  function G = CholSax(A);
%
% Pre:  A is a symmetric and positive definite matrix.
% Post: G is lower triangular and A = G*G'.

   [n,n] = size(A);
   G = zeros(n,n);
   s = zeros(n,1);
   for j=1:n
      s(j:n) = A(j:n,j);
          for k=1:j-1
             s(j:n) = s(j:n) - G(j:n,k)*G(j,k);
          end
          G(j:n,j) = s(j:n)/sqrt(s(j));
```

```
   end
```

An update of the form

$$\text{vector} \leftarrow \text{vector} + \text{matrix} \times \text{vector}$$

is called a *gaxpy* operation, which stands for "general a times x plus y" (general saxpy), that is $y = Ax + y$. A version that features level-2 gaxpy operation is the following implementation.

```
  function G = CholGax(A);
%
% Pre:  A is a symmetric and positive definite matrix.
% Post: G is lower triangular and A = G*G'.


   [n,n] = size(A);
   G = zeros(n,n);
   s = zeros(n,1);
   for j=1:n
      if j==1
         s(j:n) = A(j:n,j);
          else
             s(j:n) = A(j:n,j) - G(j:n,1:j-1)*G(j,1:j-1)';
          end
          G(j:n,j)  = s(j:n)/sqrt(s(j));
   end
```

There is also a recursive implementation which computes the Cholesky factor row by row, just like `ChoScalar`

```
  function G = CholRecur(A);
%
% Pre:  A is a symmetric and positive definite matrix.
% Post: G is lower triangular and A = G*G'.

   [n,n] = size(A);
   if n==1
      G = sqrt(A);
   else
      G(1:n-1,1:n-1) = CholRecur(A(1:n-1,1:n-1));
          G(n,1:n-1)     = LTriSol(G(1:n-1,1:n-1),A(1:n-1,n))';
          G(n,n)         = sqrt(A(n,n) - G(n,1:n-1)*G(n,1:n-1)');
   end
```

There is even a high performance level-3 implementation of the Cholesky decomposition `CholBlock`

7.3. Le méthode itérative de Gauss–Seidel

One can solve linear systems by iterative methods, especially when dealing with very large systems. One such method is Gauss–Seidel's method which uses the latest values for the variables. This method is best explained by means of an example.

EXEMPLE 7.5. Apply two iterations of Gauss–Seidel's iterative scheme to the system

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & + & x_3 & = & 14, \\ x_1 & + & 5x_2 & - & x_3 & = & 10, \\ x_1 & + & x_2 & + & 8x_3 & = & 20, \end{array} \qquad \text{with} \qquad \begin{array}{rcl} x_1^{(0)} & = & 1 \\ x_2^{(0)} & = & 1 \\ x_3^{(0)} & = & 1 \end{array}$$

SOLUTION. Since the system is diagonally dominant, Gauss–Seidel's iterative scheme will converge. This scheme is

$$\begin{array}{rcrcrcrcl} x_1^{(n+1)} & = & \frac{1}{4}(14 & & & - & 2x_2^{(n)} & - & x_3^{(n)}), \\ x_2^{(n+1)} & = & \frac{1}{5}(10 & - & x_1^{(n+1)} & & & + & x_3^{(n)}), \\ x_3^{(n+1)} & = & \frac{1}{8}(20 & - & x_1^{(n+1)} & - & x_2^{(n+1)} & &), \end{array} \qquad \begin{array}{rcl} x_1^{(0)} & = & 1 \\ x_2^{(0)} & = & 1 \\ x_3^{(0)} & = & 1 \end{array}$$

For $n = 0$ we have

$$\begin{aligned} x_1^{(1)} &= \frac{1}{4}(14 - 2 - 1) = \frac{11}{4} = 2.75 \\ x_2^{(1)} &= \frac{1}{5}(10 - 2.75 + 1) = 1.65 \\ x_3^{(1)} &= \frac{1}{8}(20 - 2.75 - 1.65) = 1.95. \end{aligned}$$

For $n = 1$:

$$\begin{aligned} x_1^{(2)} &= \frac{1}{4}(14 - 2 \times 1.65 - 1.95) = 2.1875 \\ x_2^{(2)} &= \frac{1}{5}(10 - 2.1875 + 1.95) = 1.9525 \\ x_3^{(2)} &= \frac{1}{8}(20 - 2.1875 - 1.9525) = 1.9825 \end{aligned}$$

□

The Gauss–Seidel method to solve the system $A\boldsymbol{x} = \boldsymbol{b}$ is given by the following iterative scheme:

$$\boldsymbol{x}^{(m+1)} = D^{-1}\left(\boldsymbol{b} - L\boldsymbol{x}^{(m+1)} - U\boldsymbol{x}^{(m)}\right), \qquad \text{with properly chosen } \boldsymbol{x}^{(0)},$$

where the matrix A has been split as the sum of three matrices,

$$A = D + L + U,$$

with D diagonal, L strictly lower triangular, and U strictly upper triangular.

This Gauss-Seidel algorithm is programmed in Matlab as follows:

```
A = [7 1 -1;1 11 1;-1 1 9]; b = [3 0 -17]';
D = diag(A); L = tril(A,-1); U = triu(A,1);
m = size(b,1); % number of rows of b
x = ones(m,1); % starting value
y = zeros(m,1); % temporary storage
k = 5; % number of iterations
for j = 1:k
uy = U*x(:,j);
for i = 1:m
 y(i) = (1/D(i))*(b(i)-L(i,:)*y-uy(i));
end
x = [x,y];
```

```
end
>> x
x =
    1.0000    0.4286    0.1861    0.1380    0.1357    0.1356
    1.0000   -0.1299    0.1492    0.1588    0.1596    0.1596
    1.0000   -1.8268   -1.8848   -1.8912   -1.8915   -1.8916
```

It is important to rearrange the coefficient matrix of a given linear system in as much a diagonally dominant matrix as possible since this may assure or improve the convergence of the Gauss–Seidel iterative scheme.

EXEMPLE 7.6. Rearrange the system

$$\begin{array}{rcrcrcr} 2x_1 & + & 10x_2 & - & x_3 & = & -32 \\ -x_1 & + & 2x_2 & - & 15x_3 & = & 17 \\ 10x_1 & - & x_2 & + & 2x_3 & = & 35 \end{array}$$

such that Gauss–Seidel's scheme converges.

SOLUTION. By placing the last equation first, the system will be diagonally dominant,

$$\begin{array}{rcrcrcr} 10x_1 & - & x_2 & + & 2x_3 & = & 35 \\ 2x_1 & + & 10x_2 & - & x_3 & = & -32 \\ -x_1 & + & 2x_2 & - & 15x_3 & = & 17 \end{array}$$

□

7.4. Normes de matrices

In matrix computations, norms are used to quantify results, like error estimates and to study the convergence of iterative schemes.

Given a matrix $A \in \mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$, and a vector norm $\|x\|$ for $x \in \mathbb{R}^n$, a *subordinate matrix norm*, $\|A\|$, is defined by the supremum

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|.$$

There are three important vector norms in scientific computation, the l_1-*norm* of x,

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| = |x_1| + |x_2| + \cdots + |x_n|,$$

the *Euclidean norm*, or l_2-*norm*, of x,

$$\|x\|_2 = \left[\sum_{i=1}^{n} |x_i|^2\right]^{1/2} = \left[|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2\right]^{1/2},$$

and the *supremum norm*, or l_∞-*norm*, of x,

$$\|x\|_\infty = \sup_{i=1,2,\ldots,n} |x_i| = \sup\{|x_1|, |x_2|, \ldots, |x_n|\}.$$

It can be shown that the corresponding matrix norms are given by the following formulae.

The l_1-*norm*, or column "sum" norm, of A is

$$\|A\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{n} |a_{ij}| \qquad \text{(largest column in the } l_1 \text{ vector norm)},$$

the l_∞-*norm*, or a row "sum" norm, of A is

$$\|A\|_\infty = \max_{i=1,2,\ldots,n} \sum_{j=1}^{n} |a_{ij}| \qquad \text{(largest row in the } l_1 \text{ vector norm)},$$

and the l_2-*norm* of A is

$$\|A\|_2 = \max_{i=1,2,\ldots,n} \{\sigma_i^{1/2}\} \qquad \text{(largest singular value of } A),$$

where the $\sigma_i \geq 0$ are the eigenvalues of $A^T A$. The singular values of a matrix are considered in Subsection 7.6.5.

An important non-subordinate matrix norm is the *Frobenius norm*, or *matrix Euclidean norm*,

$$\|A\|_F = \left[\sum_{j=1}^{n}\sum_{i=1}^{n} |a_{ij}|^2\right]^{1/2}.$$

EXEMPLE 7.7. Compute the l_1, l_2 and l_∞-norms of the vector

$$v = \begin{bmatrix} 1 & -2 & 2 \end{bmatrix}.$$

SOLUTION. We have

$$\begin{aligned} \|v\|_1 &= 1 + 2 + 2 = 5, \\ \|v\|_2 &= \sqrt{1^2 + 2^2 + 2^2} = 3, \\ \|v\|_\infty &= \max\{1, 2, 2\} = 2. \end{aligned}$$

Numeric Matlab obtains these norms as follows.

```
>> v = [1 -2 2]
v =     1    -2     2
>> n1= norm(v,1)
n1 =     5
>> n2 = norm(v)
n2 =     3
>> ninf = norm(v,inf)
ninf =     2
```

□

EXEMPLE 7.8. Compute the l_1, l_∞ and Frobenius norms of the matrix

$$A = \begin{bmatrix} 1 & -2 & 2 \\ -3 & 10 & 4 \\ -1 & 6 & 2 \end{bmatrix}.$$

Obtain the l_2 norm of A by means of Matlab.

SOLUTION. We have

$$\begin{aligned} \|A\|_1 &= \max\{1 + 3 + 1, 2 + 10 + 6, 2 + 4 + 2\} = 18, \\ \|A\|_\infty &= \max\{1 + 2 + 2, 3 + 10 + 4, 1 + 6 + 2\} = 17, \\ \|A\|_{\text{Frobenius}} &= \sqrt{1 + 4 + 4 + 9 + 100 + 16 + 1 + 36 + 4} = 13.2288, \end{aligned}$$

Numeric Matlab obtains these norms as follows.

```
>> A=[1 -2 2;-3 10 4;-1 6 2]
A =     1    -2     2
       -3    10     4
       -1     6     2
>> N1 = norm(A,1)
N1 =    18
>> N2 = norm(A)
N2 =   12.9392
>> Ninf = norm(A,inf)
Ninf =    17
>> Nfro = norm(A,'fro')
Nfro =   13.2288
```

□

DÉFINITION 7.2 (Condition number). The *condition number* of a matrix $A \in \mathbb{R}^{n\times n}$ is the number

$$\kappa(A) = \|A\|\,\|A^{-1}\|.$$

Note that $\kappa(A) \geq 1$ if $\|I\| = 1$.

The condition number of A appears in an upper bound for the relative error in the solution to the system

$$Ax = b.$$

In fact, let $\hat{x}$ be the exact solution to the perturbed system

$$(A + \Delta A)\hat{x} = b + \delta b,$$

where all experimental and numerical round-off errors are lumped into ΔA and δb. Then we have the bound

$$\frac{\|\hat{x} - x\|}{\|x\|} \leq \kappa(A)\left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right]. \tag{7.1}$$

We say that a system $Ax = b$ is well conditioned if $\kappa(A)$ is small; otherwise it is ill conditioned.

EXEMPLE 7.9. Study the ill condition of the following system

$$\begin{bmatrix} 1.0001 & 1 \\ 1 & 1.0001 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2.0001 \\ 2.0001 \end{bmatrix}$$

with exact and some approximate solutions

$$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \hat{x} = \begin{bmatrix} 2.0000 \\ 0.0001 \end{bmatrix},$$

respectively.

SOLUTION. The approximate solution has a very small residual (to 4 decimals), $r = b - A\hat{x}$,

$$\begin{aligned} r &= \begin{bmatrix} 2.0001 \\ 2.0001 \end{bmatrix} - \begin{bmatrix} 1.0001 & 1 \\ 1 & 1.0001 \end{bmatrix}\begin{bmatrix} 2.0000 \\ 0.0001 \end{bmatrix} \\ &= \begin{bmatrix} 2.0001 \\ 2.0001 \end{bmatrix} - \begin{bmatrix} 2.0003 \\ 2.0001 \end{bmatrix} = \begin{bmatrix} -0.0002 \\ 0.0000 \end{bmatrix}. \end{aligned}$$

However, the relative error in $\hat{x}$ is

$$\frac{\|\hat{x} - x\|_1}{\|x\|_1} = \frac{(1.0000 + 0.9999)}{1 + 1} \approx 1,$$

that is 100%. This is explained by the fact that the system is very ill conditioned. In fact,

$$A^{-1} = \frac{1}{0.0002} \begin{bmatrix} 1.0001 & -1.0000 \\ -1.0000 & 1.0001 \end{bmatrix} = \begin{bmatrix} 5000.5 & -5000.0 \\ -5000.0 & 5000.5 \end{bmatrix},$$

and

$$\kappa_1(A) = (1.0001 + 1.0000)(5000.5 + 5000.0) = 20\,002.$$

□

The l_1-norm of the matrix A of the previous example and its l_1 condition number are obtained by the following numeric Matlab commands:

```
>> A = [1.0001 1;1 1.0001];
>> N1 = norm(A,1)
N1 = 2.0001
>> K1 = cond(A,1)
K1 = 2.0001e+04
```

7.5. Systèmes surdéterminés

In curve fitting we are given N points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N),$$

and want to determine a function $f(x)$ such that

$$f(x_i) \approx y_i, \qquad i = 1, 2, \ldots, N.$$

For properly chosen functions $\varphi_0(x), \varphi_1(x), \ldots, \varphi_n(x)$, we put

$$f(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_n\varphi_n(x),$$

and minimize the quadratic form

$$Q(a_0, a_1, \ldots, a_n) = \sum_{i=1}^{N} (f(x_i) - y_i)^2.$$

If the functions $\varphi_j(x)$ are "linearly independent", the quadratic form is nondegenerate and the minimum is attained for values of $a_0, a_1, \ldots, a_n$, such that

$$\frac{\partial Q}{\partial a_j} = 0, \qquad j = 0, 1, 2, \ldots, n.$$

Writing the quadratic form Q explicitly,

$$Q = \sum_{i=1}^{N} (a_0\varphi_0(x_i) + \cdots + a_n\varphi_n(x_i) - y_i)^2,$$

and equating the partial derivatives of Q with respect to a_j to zero, we have

$$\frac{\partial Q}{\partial a_j} = 2\sum_{i=1}^{N} (a_0\varphi_0(x_i) + \cdots + a_n\varphi_n(x_i) - y_i)\varphi_j(x_i) = 0.$$

This is an $(n+1)\times(n+1)$ symmetric linear algebraic system

$$\begin{bmatrix} \sum\varphi_0(x_i)\varphi_0(x_i) & \sum\varphi_1(x_i)\varphi_0(x_i) & \cdots & \sum\varphi_n(x_i)\varphi_0(x_i) \\ \vdots & & & \vdots \\ \sum\varphi_0(x_i)\varphi_n(x_i) & \sum\varphi_1(x_i)\varphi_n(x_i) & \cdots & \sum\varphi_n(x_i)\varphi_n(x_i) \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum\varphi_0(x_i)y_i \\ \vdots \\ \sum\varphi_n(x_i)y_i \end{bmatrix}, \tag{7.2}$$

where all sums are over i from 1 to N. Setting the $N\times(n+1)$ matrix A, and the $(n+1)$ vector y as

$$A = \begin{bmatrix} \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_n(x_1) \\ \vdots & & & \vdots \\ \varphi_0(x_N) & \varphi_1(x_N) & \cdots & \varphi_n(x_N) \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix},$$

we see that the previous square system can be written in the form

$$A^TA \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = A^Ty.$$

These equations are called the *normal equations*.

In the case of linear regression, we have

$$\varphi_0(x) = 1, \qquad \varphi_1(x) = x,$$

and the normal equations are

$$\begin{bmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_iy_i \end{bmatrix}.$$

This is the least-square fit by a straight line.

In the case of quadratic regression, we have

$$\varphi_0(x) = 1, \qquad \varphi_1(x) = x, \qquad \varphi_2(x) = x^2,$$

and the normal equations are

$$\begin{bmatrix} N & \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 \\ \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 & \sum_{i=1}^N x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_iy_i \\ \sum_{i=1}^N x_i^2y_i \end{bmatrix}.$$

This is the least-square fit by a parabola.

EXEMPLE 7.10. Using the method of least squares, fit a parabola

$$f(x) = a_0 + a_1x + a_2x^2$$

to the following data

i	1	2	3	4	5
x_i	0	1	2	4	6
y_i	3	1	0	1	4

SOLUTION. **(a) The analytic solution.**— The normal equations are

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 4 & 6 \\ 0 & 1 & 4 & 16 & 36 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 4 & 6 \\ 0 & 1 & 4 & 16 & 36 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \\ 4 \end{bmatrix}, \tag{7.3}$$

that is

$$\begin{bmatrix} 5 & 13 & 57 \\ 13 & 57 & 289 \\ 57 & 289 & 1569 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 9 \\ 29 \\ 161 \end{bmatrix},$$

or

$$Na = b.$$

Using the Cholesky decomposition $N = GG^T$, we have

$$G = \begin{bmatrix} 2.2361 & 0 & 0 \\ 5.8138 & 4.8166 & 0 \\ 25.4921 & 29.2320 & 8.0430 \end{bmatrix}.$$

The solution a is obtained by forward and backward substitutions with $Gw = b$ and $G^T a = w$,

$$\begin{aligned} a_0 &= 2.8252 \\ a_1 &= -0.0490 \\ a_2 &= 0.3774. \end{aligned}$$

(b) The Matlab numeric solution.—

```
>> x = [0 1 2 4 6]';
>> A = [x.^0 x x.^2];
>> y = [3 1 0 1 4]';
>> a = (A'*A\(A'*y))'
a =   2.8252   -2.0490    0.3774
```

The result is plotted in Fig. 7.1 □

7.6. Valeurs propres de matrices

An *eigenvalue*, or *characteristic value*, of a matrix $A \in \mathbb{R}^{n\times n}$, or $\mathbb{C}^{n\times n}$, is a real or complex number such that the vector equation

$$Ax = \lambda x, \quad x \in \mathbb{R}^n \text{ or } \mathbb{C}^n, \tag{7.4}$$

has a nontrivial solution, $x \neq 0$, called an *eigenvector*. We rewrite (7.4) in the form

$$(A - \lambda I)x = 0, \tag{7.5}$$

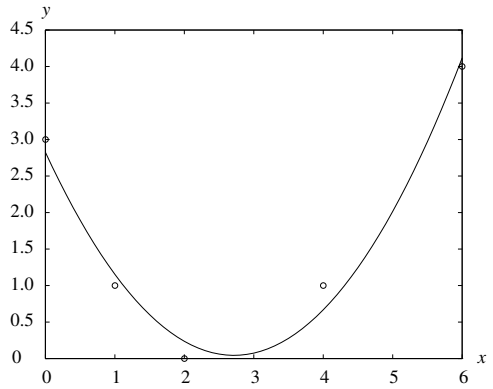

FIGURE 7.1. Quadratic least-square approximation in Example 7.10.

where I is the $n \times n$ unit matrix. This equation has a nonzero solution x if and only if the *characteristic determinant* is zero,

$$\det(A - \lambda I) = 0, \tag{7.6}$$

that is, λ is a zero of the characteristic polynomial of A.

7.6.1. Les disques de Gershgorin. The inclusion theorem of Gershgorin states that each eigenvalue of A lies in a Gershgorin disk.

THÉORÈME 7.2 (Gershgorin Theorem). *Let λ be an eigenvalue of an arbitrary $n \times n$ matrix $A = (a_{ij})$. Then for some i, $1 \le i \le n$, we have*

$$|a_{ii} - \lambda| \le |a_{i1}| + |a_{i2}| + \cdots + |a_{i,i-1}| + |a_{i,i+1}| + \cdots + |a_{in}|. \tag{7.7}$$

PROOF. Let x be an eigenvector corresponding to the eigenvalue λ. Then

$$Ax = \lambda x \quad \text{or} \quad (A - \lambda I)x = 0. \tag{7.8}$$

Let x_i be a component of x that is largest in absolute value. Then we have $|x_j/x_i| \le 1$ for $j = 1, 2, \ldots, n$. The vector equation (7.8) is equivalent to a system of n equations for the n components of the vectors on both sides, and the ith of these n equations is

$$a_{i1}x_1 + \cdots + a_{i,i-1}x_{i-1} + (a_{ii} - \lambda)x_i + a_{i,i+1}x_{i+1} + \cdots + a_{in}x_n = 0.$$

Division by x_i and reshuffling terms gives

$$a_{ii} - \lambda = -a_{i1}\frac{x_1}{x_i} - \cdots - a_{i,i-1}\frac{x_{i-1}}{x_i} - a_{i,i+1}\frac{x_{i+1}}{x_i} - \cdots - a_{in}\frac{x_n}{x_i}.$$

Taking absolute values on both sides of this equation, applying the triangle inequality $|a+b| \le |a|+|b|$ (where a and b are any complex numbers), and observing that because of the choice of i,

$$\left|\frac{x_1}{x_i}\right| \le 1, \quad \ldots, \quad \left|\frac{x_n}{x_j}\right| \le 1,$$

we obtain (7.7), and the theorem is proved. □

EXEMPLE 7.11. Using Gershgorin Theorem, determine and sketch the Gershgorin disks D_k that contain the eigenvalues of the matrix

$$A = \begin{bmatrix} -3 & 0.5i & -i \\ 1-i & 1+i & 0 \\ 0.1i & 1 & -i \end{bmatrix}.$$

SOLUTION. The centres, c_i, and radii, r_i, of the disks are

$$\begin{aligned} c_1 &= -3, & r_1 &= |0.5i| + |-i| = 1.5 \\ c_2 &= 1+i, & r_2 &= |1-i| + |0| &= \sqrt{2} \\ c_3 &= -i, & r_3 &= |0.1i| + 1 &= 1.1 \end{aligned}$$

as shown in Fig. 7.2. □

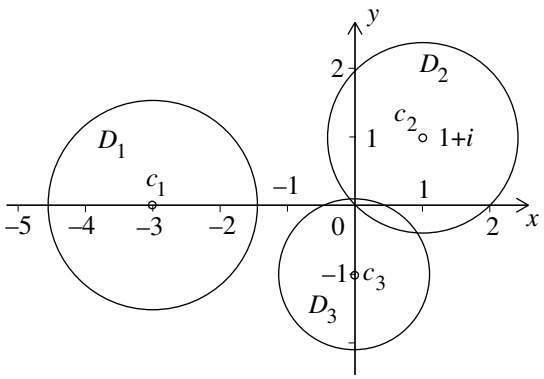

FIGURE 7.2. Gershgorin disks for Example 7.11.

7.6.2. La méthode de la puissance. The *power method* may be used to determine the eigenvalue of largest modulus of a matrix A and the corresponding eigenvector. The method is derived as follows.

For simplicity we assume that A admits n linearly independent eigenvectors $z_1, z_2, \ldots, z_n$ corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, ordered such that

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|.$$

Then any vector x can be represented in the form

$$x = a_1 z_1 + a_2 z_2 + \cdots + a_n z_n.$$

Applying A^k to x, we have

$$\begin{aligned} A^k x &= a_1 \lambda_1^k z_1 + a_2 \lambda_2^k z_2 + \cdots + a_n \lambda_n^k z_n \\ &= \lambda_1^k \left[a_1 z_1 + a_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k z_2 + \cdots + a_n \left(\frac{\lambda_n}{\lambda_1} \right)^k z_n \right] \\ &\to \lambda_1^k a_1 z_1 = y \quad \text{as} \quad k \to \infty. \end{aligned}$$

Thus $Ay = \lambda_1 y$. In practice, successive vectors are scaled to avoid overflows.

$$\begin{aligned} Ax^{(0)} &= x^{(1)}, & u^{(1)} &= \frac{x^{(1)}}{\|x^{(1)}\|_\infty}, \\ Au^{(1)} &= x^{(2)}, & u^{(2)} &= \frac{x^{(2)}}{\|x^{(2)}\|_\infty}, \\ &\vdots \\ Au^{(n)} &= x^{(n+1)} \\ &= \lambda_1 u^{(n)}. \end{aligned}$$

EXEMPLE 7.12. Using the power method, find the largest eigenvalue and the corresponding eigenvector of the matrix

$$\begin{bmatrix} 3 & 2 \\ 2 & 5 \end{bmatrix}.$$

SOLUTION. Letting $x^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, we have

$$\begin{aligned} \begin{bmatrix} 3 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} 5 \\ 7 \end{bmatrix} = x^{(1)}, & u^{(1)} &= \begin{bmatrix} 5/7 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} 3 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 5/7 \\ 1 \end{bmatrix} &= \begin{bmatrix} 4.14 \\ 6.43 \end{bmatrix} = x^{(2)}, & u^{(2)} &= \begin{bmatrix} 0.644 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} 3 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 0.644 \\ 1 \end{bmatrix} &= \begin{bmatrix} 3.933 \\ 6.288 \end{bmatrix} = x^{(3)}, & u^{(3)} &= \begin{bmatrix} 0.6254 \\ 1 \end{bmatrix}. \end{aligned}$$

Hence

$$\lambda_1 \approx 6.288, \qquad z_1 \approx \begin{bmatrix} 0.6254 \\ 1 \end{bmatrix}.$$

□

Numeric Matlab has the command `eig` to find the eigenvalues and eigenvectors of a numeric matrix. For example

```
>> A = [3 2;2 5];
>> [X,D] = eig(A)
X =
    0.8507    0.5257
   -0.5257    0.8507
D =
    1.7639         0
         0    6.2361
```

where the columns of the matrix X are the eigenvectors of A and the diagonal elements of the diagonal matrix D are the eigenvalues of A. The numeric command `eig` uses the QR algorithm with shifts to be described next.

7.6.3. La méthode de la puissance inverse. A more versatile method to determine any eigenvalue of a matrix $A \in \mathbb{R}^{n\times n}$, or $\mathbb{C}^{n\times n}$, is the *inverse power method*. It is derived as follows, under the simplifying assumption that A has n linearly independent eigenvectors $z_1, \ldots, z_n$, and λ is near λ_1.

We have

$$(A-\lambda I)x^{(1)} = x^{(0)} = a_1 z_1 + \cdots + a_n z_n,$$
$$x^{(1)} = a_1 \frac{1}{\lambda_1 - \lambda} z_1 + a_2 \frac{1}{\lambda_2 - \lambda} z_2 + \cdots + a_n \frac{1}{\lambda_n - \lambda} z_n.$$

Similarly, by recurrence,

$$x^{(k)} = a_1 \frac{1}{(\lambda_1 - \lambda)^k} \left[z_1 + a_2 \left(\frac{\lambda_1 - \lambda}{\lambda_2 - \lambda} \right)^k z_2 + \cdots + a_n \left(\frac{\lambda_1 - \lambda}{\lambda_n - \lambda} \right)^k z_n \right]$$
$$\to a_1 \frac{1}{(\lambda_1 - \lambda)^k} z_1, \qquad \text{as} \quad k \to \infty,$$

since

$$\left| \frac{\lambda_1 - \lambda}{\lambda_j - \lambda} \right| < 1, \qquad j \neq 1.$$

Thus, the sequence $x^{(k)}$ converges in the direction of z_1. In practice the vectors $x^{(k)}$ are normalized and the system

$$(A-\lambda I)x^{(k+1)} = x^{(k)}$$

is solved by the LU decomposition. The algorithm is as follows.

Choose $x^{(0)}$
For $k = 1, 2, 3, \ldots,$ do
Solve
$(A-\lambda I)y^{(k)} = x^{(k-1)}$ by the LU decomposition with partial pivoting.
$x^{(k)} = y^{(k)}/\|y^{(k)}\|_\infty$
Stop if $\|(A-\lambda I)x^{(k)}\|_\infty < c\epsilon\|A\|_\infty$, where c is a constant of order unity and ϵ is the machine epsilon.

7.6.4. La méthode QR. A very powerful method to solve ill-conditioned and overdetermined system

$$Ax = b, \qquad A \in \mathbb{R}^{m\times n}, \qquad m \geq n,$$

is the QR decomposition,

$$A = QR,$$

where Q is orthogonal, or unitary, and R is upper triangular. In this case,

$$\|Ax - b\|_2 = \|QRx - b\|_2 = \|Rx - Q^T b\|_2.$$

If A has full rank, that is, rank of A is equal to n, we can write

$$R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}, \qquad Q^T b = \begin{bmatrix} c \\ d \end{bmatrix},$$

where $R_1 \in \mathbb{R}^{n\times n}$, $0 \in \mathbb{R}^{(m-n)\times n}$, $c \in \mathbb{R}^n$, $d \in \mathbb{R}^{m-n}$, and R_1 is upper triangular and non singular.

Then the least-square solution is

$$x = R_1^{-1} c$$

obtained by solving

$$R_1 x = c$$

by backward substitution and the residual

$$\rho = \min_{x\in\mathbb{R}^n} \|Ax - b\|_2 = \|d\|_2.$$

In the QR decomposition, the matrix A is transformed into an upper-triangular matrix by the successive application of $n-1$ Householder reflections, the kth one zeroing the elements below the diagonal element in the k column. For example, to zero the elements x_2, x_3, ..., x_n in the vector $x \in \mathbb{R}^n$, one applies the Householder reflection

$$P = I - 2\,\frac{vv^T}{v^T v},$$

with

$$v = x + \operatorname{sgn}(x_1)\, e_1, \qquad \text{where } e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

In this case,

$$P \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} -\|x\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The matrix P is orthogonal and is equal to its own inverse, that is, it satisfies the relations

$$P^T = P = P^{-1}.$$

To minimize the number of floating point operations and memory allocation, the scalar

$$s = 2/v^T v$$

is first computed and then

$$Px = x - s(v^T x)v$$

is computed taking the special structure of the matrix P into account. To keep P in memory, only the number s and the vector v need be stored.

Softwares systematically use the QR decomposition to solve overdetermined systems. So does the Matlab left-division command \ with an overdetermined or singular system.

The QR algorithm, which uses a sequence of QR decompositions

$$\begin{aligned} A &= Q_1 R_1 \\ A_1 &= R_1 Q_1 = Q_2 R_2 \\ A_2 &= R_2 Q_2 = Q_3 R_3 \\ &\vdots \end{aligned}$$

yields the eigenvalues of A, since A_n converges to an upper or quasi-upper triangular matrix with the real eigenvalues on the diagonal and complex eigenvalues in 2×2 diagonal blocks. Combined with simple shifts, double shifts, and other shifts, convergence is very fast.

The numeric Matlab command `qr` produces the QR decomposition of a matrix:

```
>> A = [1 2 3; 4 5 6; 7 8 9];
>> [Q,R] = qr(A)
Q =
   -0.1231    0.9045    0.4082
```

```
   -0.4924    0.3015   -0.8165
   -0.8616   -0.3015    0.4082
R =
   -8.1240   -9.6011  -11.0782
         0    0.9045    1.8091
         0         0   -0.0000
```

The diagonal elements of R are the eigenvalues of A. It is seen that the matrix A is singular since the diagonal element $r_{33} = 0$.

For large matrices, of order $n \geq 100$, one seldom wants all the eigenvalues. To find selective eigenvalues, one may use Lanczos method.

The Jacobi method to find the eigenvalues of a symmetric matrix is being revived since it is parallelizable for parallel computers.

7.6.5. La décomposition en valeurs singulières. The *singular value decomposition* is a very powerful tool in matrix computation. It is more expensive in time than the previous methods. Any matrix $A \in \mathbb{R}^{m\times n}$, say, with $m \geq n$, can be factored in the form

$$A = U\Sigma V^T,$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix, whose diagonal elements σ_i ordered in decreasing order

$$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n \geq 0,$$

are the *singular values* of A. If $A \in \mathbb{R}^{n\times n}$ is a square matrix, it is seen that

$$\|A\|_2 = \sigma_1, \qquad \|A^{-1}\|_2 = \sigma_n.$$

The same decomposition holds for complex matrices $A \in \mathbb{C}^{m\times n}$. In this case U and V are unitary and the transpose V^T is replaced by the Hermitian transpose

$$V^* = \overline{V}^T.$$

The rank of a matrix A is the number of nonzero singular values of A. If A is symmetric, $A^T = A$, Hermitian symmetric $A^H = A$ or, more generally, normal, $AA^* = A^*A$, the the moduli of the eigenvalues of A are the singular values of A.

The numeric Matlab command `svd` produces the singular value of a matrix:

```
A = [1 2 3; 4 5 6; 7 8 9];
[U,S,V] = svd(A)
U =
    0.2148    0.8872   -0.4082
    0.5206    0.2496    0.8165
    0.8263   -0.3879   -0.4082
S =
   16.8481         0         0
         0    1.0684         0
         0         0    0.0000


V =

    0.4797   -0.7767    0.4082
    0.5724   -0.0757   -0.8165
```

```
    0.6651    0.6253    0.4082
```

The diagonal elements of the matrix S are the singular values of A. The l_2 norm of A is $\|A\|_2 = \sigma_1 = 16.8481$. Since $\sigma_3 = 0$, the matrix A is singular.

CHAPITRE 8

Transformation de Laplace

8.1. Définition

DÉFINITION 8.1. Soit $f(x)$ une fonction définie sur $[0, +\infty[$. La transformée de Laplace $F(s)$ de la fonction $f(t)$ est définie par l'intégrale

$$\mathcal{L}(f)(s) := F(s) = \int_0^\infty e^{-st} f(t)\, dt, \tag{8.1}$$

sous la condition que l'intégrale existe pour $s > \gamma$. Dans ce cas on dit que $f(t)$ est transformable et qu'elle est l'*originale* de $F(s)$.

On voit que la fonction exponentielle

$$f(t) = e^{t^2}$$

n'admet pas de transformée de Laplace puisque l'intégrale (8.1) n'existe pas quelque soit le choix de $s > 0$.

On illustre la définition par quelques exemples utiles.

EXEMPLE 8.1. Trouver la transformée de Laplace de $f(t) = 1$.

RÉSOLUTION. **(a) Résolution analytique.**—

$$\begin{aligned}\mathcal{L}(1)(s) &= \int_0^\infty e^{-st}\, dt, \qquad s > 0,\\ &= -\frac{1}{s} e^{-st}\Big|_0^\infty = -\frac{1}{s}(0 - 1)\\ &= \frac{1}{s}.\end{aligned}$$

(b) Résolution par Matlab symbolique.—

```
>> f = sym('Heaviside(t)');
>> F = laplace(f)
F = 1/s
```

La fonction `Heaviside` est une fonction de Maple. L'assistance pour les fonctions de Maple s'obtient par la commande `mhelp`. □

EXEMPLE 8.2. Montrer que

$$\mathcal{L}\left(e^{at}\right)(s) = \frac{1}{s - a}. \tag{8.2}$$

RÉSOLUTION. **(a) Résolution analytique.**—

$$\begin{aligned}\mathcal{L}\left(e^{at}\right)(s) &= \int_0^\infty e^{-st}\, e^{at}\, dt \\ &= \int_0^\infty e^{-(s-a)t}\, dt \qquad (s > a) \\ &= -\frac{1}{s-a}\left[e^{-(s-a)t}\right]_0^\infty \\ &= \frac{1}{s-a}.\end{aligned}$$

(b) Résolution par Matlab symbolique.—

```
>> syms a t;
>> f = exp(a*t);
>> F = laplace(f)
F = 1/(s-a)
```

□

THÉORÈME 8.1. *La transformation de Laplace,*

$$\mathcal{L} : f(t) \;\mapsto\; F(s),$$

est linéaire.

DÉMONSTRATION.

$$\begin{aligned}\mathcal{L}(af + bg) &= \int_0^\infty e^{-st}\big[af(t) + bg(t)\big]\, dt \\ &= a\int_0^\infty e^{-st} f(t)\, dt + b\int_0^\infty e^{-st} g(t)\, dt \\ &= a\mathcal{L}(f)(s) + b\mathcal{L}(g)(s). \quad \square\end{aligned}$$

EXEMPLE 8.3. Trouver la transformée de Laplace de $f(t) = \cosh at$.

RÉSOLUTION. **(a) Résolution analytique.**— Puisque

$$\cosh at = \frac{1}{2}\left(e^{at} + e^{-at}\right),$$

on obtient

$$\begin{aligned}\mathcal{L}(\cosh at)(s) &= \frac{1}{2}\left[\mathcal{L}\left(e^{at}\right) + \mathcal{L}\left(e^{-at}\right)\right] \\ &= \frac{1}{2}\left[\frac{1}{s-a} + \frac{1}{s+a}\right] \\ &= \frac{s}{s^2 - a^2}.\end{aligned}$$

(b) Résolution par Matlab symbolique.—

```
>> syms a t;
>> f = cosh(a*t);
>> F = laplace(f)
F = s/(s^2-a^2)
```

□

EXEMPLE 8.4. Trouver la transformée de Laplace de $f(t) = \sinh at$.

RÉSOLUTION. **(a) Résolution analytique.**— Puisque

$$\sinh at = \frac{1}{2}\left(e^{at} - e^{-at}\right),$$

on obtient

$$\begin{aligned}\mathcal{L}(\sinh at)(s) &= \frac{1}{2}\left[\mathcal{L}\left(e^{at}\right) - \mathcal{L}\left(e^{-at}\right)\right] \\ &= \frac{1}{2}\left[\frac{1}{s-a} - \frac{1}{s+a}\right] \\ &= \frac{a}{s^2 - a^2}.\end{aligned}$$

(b) Résolution par Matlab symbolique.—

```
>> syms a t;
>> f = sinh(a*t);
>> F = laplace(f)
F = a/(s^2-a^2)
```

□

REMARQUE 8.1. On voit que $\mathcal{L}(\cosh at)(s)$ est une fonction paire de a et $\mathcal{L}(\sinh at)(s)$ est une fonction impaire de a.

EXEMPLE 8.5. Trouver la transformée de Laplace de $f(t) = t^n$.

RÉSOLUTION. On procède par induction mathématique. Supposons que

$$\mathcal{L}(t^{n-1})(s) = \frac{(n-1)!}{s^n}.$$

Cette formule est vraie pour $n = 1$,

$$\mathcal{L}(1)(s) = \frac{0!}{s^1} = \frac{1}{s}.$$

Si $s > 0$, il suit par intégration par parties:

$$\begin{aligned}\mathcal{L}(t^n)(s) &= \int_0^\infty e^{-st} t^n \, dt \\ &= -\frac{1}{s}\left[t^n e^{-st}\Big|_0^\infty + \frac{n}{s}\int_0^\infty e^{-st} t^{n-1}\, dt\right. \\ &= \frac{n}{s}\,\mathcal{L}(t^{n-1})(s).\end{aligned}$$

Alors par récurrence:

$$\begin{aligned}\mathcal{L}(t^n)(s) &= \frac{n}{s}\,\frac{(n-1)!}{s^n} \\ &= \frac{n!}{s^{n+1}}, \qquad s > 0. \quad \square\end{aligned}$$

On trouve la transformée de Laplace, par exemple, de t^5, par les commandes de Matlab symbolique:

```
>> syms t
>> f = t^5;
>> F = laplace(f)
F = 120/s^6
```

ou

```
>> F = laplace(sym('t^5'))
F = 120/s^6
```

ou

```
>> F = laplace(sym('t')^5)
F = 120/s^6
```

EXEMPLE 8.6. Trouver la transformée de Laplace de $\cos\omega t$ et de $\sin\omega t$.

RÉSOLUTION. **(a) Résolution analytique.**— On emploie l'identité d'Euler:

$$e^{i\omega t} = \cos\omega t + i\sin\omega t, \qquad i = \sqrt{-1}.$$

Alors

$$\begin{aligned}
\mathcal{L}\left(e^{i\omega t}\right)(s) &= \int_0^\infty e^{-st}\, e^{i\omega t}\, dt \qquad (s>0)\\
&= \int_0^\infty e^{-(s-i\omega)t}\, dt\\
&= -\frac{1}{s-i\omega}\left[e^{-(s-i\omega)t}\Big|_0^\infty\right.\\
&= -\frac{1}{s-i\omega}\left[e^{-st}e^{i\omega t}\big|_{t\to\infty} - 1\right]\\
&= \frac{1}{s-i\omega} = \frac{1}{s-i\omega}\,\frac{s+i\omega}{s+i\omega}\\
&= \frac{s+i\omega}{s^2+\omega^2}.
\end{aligned}$$

Par la linéarité de $\mathcal{L}$, on obtient

$$\begin{aligned}
\mathcal{L}\left(e^{i\omega t}\right)(s) &= \mathcal{L}(\cos\omega t + i\sin\omega t)\\
&= \mathcal{L}(\cos\omega t) + i\mathcal{L}(\sin\omega t)\\
&= \frac{s}{s^2+\omega^2} + i\frac{\omega}{s^2+\omega^2}
\end{aligned}$$

On a donc

$$\mathcal{L}(\cos\omega t) = \frac{s}{s^2+\omega^2}, \tag{8.3}$$

qui est paire en ω, et

$$\mathcal{L}(\sin\omega t) = \frac{\omega}{s^2+\omega^2}, \tag{8.4}$$

qui est impaire en ω.

(b) Résolution par Matlab symbolique.—

```
>> 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)
```

□

On supposera, dans la suite, que la transformée de Laplace des fonctions considérées dans ce chapitre existe et est différentiable et intégrable sous des conditions additionnelles. Les hypothèses de base sont énoncées dans la définition et le théorème suivants. La formule générale de la transformation de Laplace réciproque, qui n'est pas introduite dans ce chapitre, requiert aussi les résultats suivants.

DÉFINITION 8.2. On dit que la fonction $f(t)$ est de type exponentiel d'ordre γ s'il existe des constantes γ, $M > 0$ et $T > 0$ telles que

$$|f(t)| \leq M\, e^{\gamma t}, \qquad \text{pour tout } t > T. \tag{8.5}$$

On appelle *abscisse de convergence* de $f(t)$ la borne inférieure $\gamma_0 \leq \gamma$ telle que (8.5) est valide.

THÉORÈME 8.2. *Soit $f(t)$ une fonction continue par morceaux sur l'intervalle $[0, \infty)$. Si γ_0 est l'abscisse de convergence de $f(t)$, alors l'intégrale*

$$\int_0^\infty e^{-st} f(t)\, dt$$

est absolument et uniformément convergente pour tout $s > \gamma_0$.

PROOF. On démontre la convergence absolue:

$$\begin{aligned}\left|\int_0^\infty e^{-st} f(t)\, dt\right| &\leq \int_0^\infty M\, e^{-(s-\gamma_0)t}\, dt \\ &= -\frac{M}{s-\gamma_0}\, e^{-(s-\gamma_0)t}\Big|_0^\infty = \frac{M}{s-\gamma_0}. \quad \square\end{aligned}$$

8.2. Transformées de dérivées et d'intégrales

En vue des applications aux équations différentielles, il faut savoir transformer la dérivée d'une fonction.

THÉORÈME 8.3.

$$\mathcal{L}(f')(s) = s\mathcal{L}(f) - f(0). \tag{8.6}$$

DÉMONSTRATION. On intègre par partie:

$$\begin{aligned}\mathcal{L}(f')(s) &= \int_0^\infty e^{-st} f'(t)\, dt \\ &= e^{-st} f(t)\Big|_0^\infty - (-s)\int_0^\infty e^{-st} f(t)\, dt \\ &= s\mathcal{L}(f)(s) - f(0). \quad \square\end{aligned}$$

REMARQUE 8.2. On montre par récurrence les formules suivantes:

$$\mathcal{L}(f'')(s) = s^2\mathcal{L}(f) - sf(0) - f'(0), \tag{8.7}$$

$$\mathcal{L}(f''')(s) = s^3\mathcal{L}(f) - s^2 f(0) - sf'(0) - f''(0). \tag{8.8}$$

En effet,

$$\begin{aligned}\mathcal{L}(f'')(s) &= s\mathcal{L}(f')(s) - f'(0)\\ &= s[s\mathcal{L}(f)(s) - f(0)] - f'(0)\\ &= s^2\mathcal{L}(f) - sf(0) - f'(0)\end{aligned}$$

et

$$\begin{aligned}\mathcal{L}(f''')(s) &= s\mathcal{L}(f'')(s) - f''(0)\\ &= s[s^2\mathcal{L}(f) - sf(0) - f'(0)] - f''(0)\\ &= s^3\mathcal{L}(f) - s^2f(0) - sf'(0) - f''(0). \quad \square\end{aligned}$$

On démontre le théorème général suivant par récurrence.

THÉORÈME 8.4. *Soit $f(t), f'(t), \ldots, f^{(n-1)}(t)$ continues pour $t \geq 0$ et $f^{(n)}(t)$ transformable pour $s \geq \gamma$. Alors*

$$\mathcal{L}\left(f^{(n)}\right)(s) = s^n\mathcal{L}(f) - s^{n-1}f(0) - s^{n-2}f'(0) - \cdots - f^{(n-1)}(0). \tag{8.9}$$

DÉMONSTRATION. La démonstration est par récurrence comme à la remarque 8.2 pour $n = 2$ et $n = 3$. $\square$

EXEMPLE 8.7. Résoudre le problème aux valeurs initiales:

$$y'' + 4y' + 3y = 0, \qquad y(0) = 3, \qquad y'(0) = 1.$$

Tracer la solution.

RÉSOLUTION. **(a) Résolution analytique.**— On note

$$\mathcal{L}(y)(s) = Y(s)$$

et l'on transforme l'équation différentielle:

$$\begin{aligned}\mathcal{L}(y'') + 4\mathcal{L}(y') + 3\mathcal{L}(y) &= s^2Y(s) - sy(0) - y'(0) + 4[sY(s) - y(0)] + 3Y(s)\\ &= 0.\end{aligned}$$

On obtient donc

$$(s^2 + 4s + 3)Y(s) - (s + 4)y(0) - y'(0) = 0,$$

qu'on récrit en remplaçant $y(0)$ et $y'(0)$ par leurs valeurs:

$$\begin{aligned}(s^2 + 4s + 3)Y(s) &= (s + 4)y(0) + y'(0)\\ &= 3(s + 4) + 1 = 3s + 13.\end{aligned}$$

On isole l'inconnue $Y(s)$ et l'on développe le 2ème membre en fractions simples:

$$\begin{aligned}Y(s) &= \frac{3s + 13}{s^2 + 4s + 3}\\ &= \frac{3s + 13}{(s + 1)(s + 3)}\\ &= \frac{A}{s + 1} + \frac{B}{s + 3}.\end{aligned}$$

Pour évaluer A et B, on chasse les dénominateurs et l'on récrit les deux derniers membres sous la forme

$$\begin{aligned} 3s+13 &= (s+3)A + (s+1)B \\ &= (A+B)s + (3A+B). \end{aligned}$$

Des 1er et 3ème membres on obtient le système linéaire:

$$\begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 3 \\ 13 \end{bmatrix} \implies \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}.$$

Dans ce cas simple, on peut évaluer A et B tout simplement en posant $s = -1$ puis $s = -3$ dans l'identité

$$3s+13 = (s+3)A + (s+1)B.$$

Alors

$$-3+13 = 2A \implies A = 5, \qquad -9+13 = -2B \implies B = -2.$$

On a donc

$$Y(s) = \frac{5}{s+1} - \frac{2}{s+3}.$$

On trouve l'originale au moyen de la tranformation de Laplace réciproque qui, en l'occurrence, est donnée de la formule (8.2):

$$\begin{aligned} y(t) = \mathcal{L}^{-1}(Y) &= 5\mathcal{L}^{-1}\left(\frac{1}{s+1}\right) - 2\mathcal{L}^{-1}\left(\frac{1}{s+3}\right) \\ &= 5\,e^{-t} - 2\,e^{-3t}. \end{aligned}$$

(b) Résolution par Matlab symbolique.— On utilise l'expression pour $Y(s)$:

```
>> 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) Résolution par Matlab numérique.— La fonction fichier M `exp77.m`:

```
function yprime = exp77(t,y);
yprime = [y(2); -3*y(1)-4*y(2)];
```

Le solveur `ode45` de Matlab numérique produit la 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('Graphe de la solution')
```

La commande `subplot` produit la fig. 8.1 dont le lettrage est suffisamment gros après réduction. □

Remarque 8.3. On remarque que le polynôme caractéristique de l'équation différentielle homogène multiplie l'inconnue $Y(s)$ de l'équation transformée.

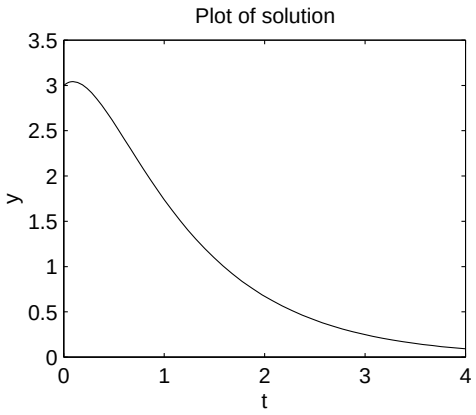

FIGURE 8.1. Graphe de la solution de l'équation différentielle de l'exemple 8.7.

REMARQUE 8.4. La méthode de résolution d'une équation différentielle au moyen de la transformation de Laplace prend en main les valeurs initiales. Ce procédé équivaut à la methode des coefficients indéterminés ou de la variation des paramètres.

Puisque l'intégration est la réciproque de la dérivation et que la transformée de la dérivée de $f(t)$ est essentiellement la transformée de $f(t)$ multipliée par s on prévoit que la transformée de l'intégrale indéfinie de $f(t)$ sera la transformée de $f(t)$ divisée par s car la division est l'inverse de la multiplication.

THÉORÈME 8.5. *Soit $f(t)$ transformable pour $s \geq \gamma$. Alors*

$$\mathcal{L}\left\{\int_0^t f(\tau)\,d\tau\right\} = \frac{1}{s}\mathcal{L}(f), \tag{8.10}$$

ou, en employant la transformation inverse de Laplace,

$$\mathcal{L}^{-1}\left\{\frac{1}{s}F(s)\right\} = \int_0^t f(\tau)\,d\tau. \tag{8.11}$$

DÉMONSTRATION. Soit

$$g(t) = \int_0^t f(\tau)\,d\tau.$$

Alors,

$$\mathcal{L}\big(f(t)\big) = \mathcal{L}\big(g'(t)\big) = s\mathcal{L}\big(g(t)\big) - g(0).$$

Puisque $g(0) = 0$, on a $\mathcal{L}(f) = s\mathcal{L}(g)$, d'où (8.10). □

EXEMPLE 8.8. Soit

$$\mathcal{L}(f) = \frac{1}{s(s^2+\omega^2)}.$$

Trouver $f(t)$.

RÉSOLUTION. **(a) Résolution analytique.**— Puisque

$$\mathcal{L}^{-1}\left(\frac{1}{s^2+\omega^2}\right) = \frac{1}{\omega}\sin\omega t,$$

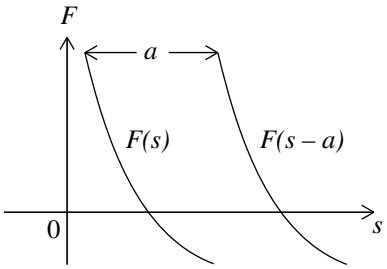

FIGURE 8.2. Déplacement $F(s-a)$ de $F(s)$ pour $a>0$.

par (8.11) on a

$$\mathcal{L}^{-1}\left\{\frac{1}{s}\left(\frac{1}{s^2+\omega^2}\right)\right\}=\frac{1}{\omega}\int_0^t \sin\omega\tau\,d\tau=\frac{1}{\omega^2}\,(1-\cos\omega t).$$

(b) Résolution par Matlab symbolique.—

```
>> syms s omega t
>> F = 1/(s*(s^2+omega^2));
>> f = ilaplace(F)
f = 1/omega^2-1/omega^2*cos(omega*t)
```

□

8.3. Déplacements en s et en t

Dans la pratique, on a besoin de l'originale de $F(s-a)$ et de la transformée de $u(t-a)f(t-a)$ où $u(t)$ est la fonction d'Heaviside,

$$u(t)=\begin{cases}0, & \text{si } x<0,\\ 1, & \text{si } x>0.\end{cases} \tag{8.12}$$

THÉORÈME 8.6. *Soit*

$$\mathcal{L}(f)(s)=F(s),\qquad s>\gamma.$$

Alors

$$\mathcal{L}\big(e^{at}f(t)\big)(s)=F(s-a),\qquad s-a>\gamma. \tag{8.13}$$

DÉMONSTRATION. (V. figure 8.2)

$$\begin{aligned}F(s-a)&=\int_0^\infty e^{-(s-a)t}f(t)\,dt\\ &=\int_0^\infty e^{-st}\left[e^{at}f(t)\right]\,dt\\ &=\mathcal{L}\big(e^{at}f(t)\big)(s).\quad\square\end{aligned}$$

EXEMPLE 8.9. Appliquer le théorème 8.6 aux trois fonctions simples t^n, $\cos\omega t$ et $\sin\omega t$.

RÉSOLUTION. **(a) Résolution analytique.**— On présente les résultats sous forme de tableau.

$f(t)$	$F(s)$	$e^{at}f(t)$	$F(s-a)$
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) Résolution par Matlab symbolique.— Matlab symbolique produit la 2ème et la 3ème fonctions:

```
>> 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)
```

□

EXEMPLE 8.10. Trouver la solution du système amorti:

$$y'' + 2y' + 5y = 0, \qquad y(0) = 2, \qquad y'(0) = -4,$$

au moyen de la transformation de Laplace.

RÉSOLUTION. Posons

$$\mathcal{L}(y)(s) = Y(s).$$

Alors,

$$s^2Y(s) - sy(0) - y'(0) + 2[sY(s) - y(0)] + 5Y(s) = 0.$$

On regroupe $Y(s)$ au 1er membre:

$$\begin{aligned}(s^2+2s+5)Y(s) &= sy(0) + y'(0) + 2y(0)\\ &= 2s - 4 + 4\\ &= 2s.\end{aligned}$$

On isole $Y(s)$ et l'on réarrange le 2ème membre:

$$\begin{aligned}Y(s) &= \frac{2s}{s^2+2s+1+4}\\ &= \frac{2(s+1)-2}{(s+1)^2+2^2}\\ &= \frac{2(s+1)}{(s+1)^2+2^2} - \frac{2}{(s+1)^2+2^2}.\end{aligned}$$

On a donc la solution

$$y(t) = 2\,e^{-t}\cos 2t - e^{-t}\sin 2t. \quad \square$$

DÉFINITION 8.3. La translatée $u_a(t) = u(t-a)$ de la fonction d'Heaviside $u(t)$, dite échelon unité, est la fonction (V. figure 8.3):

$$u_a(t) := u(t-a) = \begin{cases} 0, & \text{si } t < a,\\ 1, & \text{si } t > a, \end{cases} \qquad a \geq 0. \tag{8.14}$$

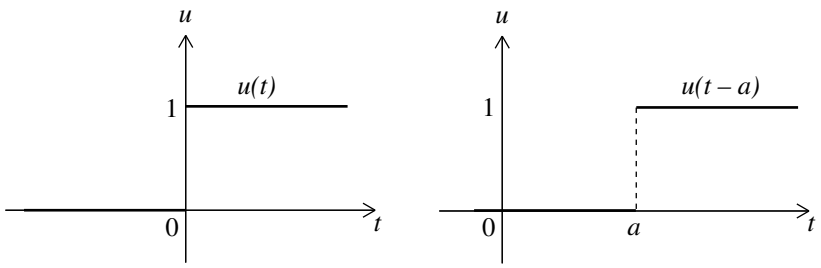

FIGURE 8.3. La fonction $u(t)$ d'Heaviside et sa translatée $u(t-a)$, $a>0$.

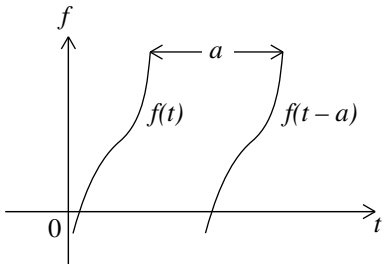

FIGURE 8.4. Déplacement $f(t-a)$ de $f(t)$ pour $a>0$.

Selon les auteurs, on note $u(t)$ aussi par $\alpha(t)$ ou $H(t)$. En Matlab symbolique, la fonction de Heaviside s'obtient par les commandes:

```
>> sym('Heaviside(t)')
>> u = sym('Heaviside(t)')
u = Heaviside(t)
```

L'assistance pour les fonctions de Maple s'obtient par la commande `mhelp`.

THÉORÈME 8.7. *Soit*

$$\mathcal{L}(f)(s) = F(s).$$

Alors

$$\mathcal{L}^{-1}\big(e^{-as}F(s)\big) = u(t-a)f(t-a), \tag{8.15}$$

c'est-à-dire

$$\mathcal{L}\big(u(t-a)f(t-a)\big)(s) = e^{-as}F(s). \tag{8.16}$$

DÉMONSTRATION. (V. figure 8.4)

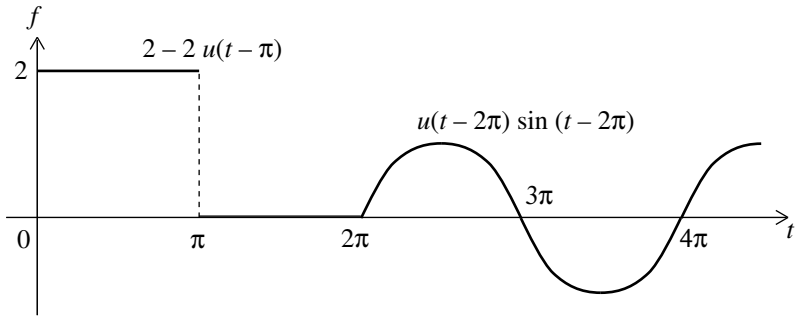

FIGURE 8.5. La fonction $f(t)$ de l'exemple 8.11.

$$\begin{aligned}
e^{-as}F(s) &= e^{-as}\int_0^\infty e^{-s\tau}f(\tau)\,d\tau \\
&= \int_0^\infty e^{-s(\tau+a)}f(\tau)\,d\tau \\
&\qquad \text{(On pose } \tau + a = t,\ d\tau = dt\text{)} \\
&= \int_a^\infty e^{-st}f(t-a)\,dt \\
&= \int_0^a e^{-st}0f(t-a)\,dt + \int_a^\infty e^{-st}1f(t-a)\,dt \\
&= \int_0^\infty e^{-st}u(t-a)f(t-a)\,dt \\
&= \mathcal{L}\big(u(t-a)f(t-a)\big)(s). \quad \square
\end{aligned}$$

On voit que

$$\mathcal{L}\big(u(t-a)\big) = \frac{e^{-as}}{s}, \qquad s > 0. \tag{8.17}$$

Cette formule découle directement de la définition:

$$\begin{aligned}
\mathcal{L}\big(u(t-a)\big) &= \int_0^\infty e^{-st}u(t-a)\,dt \\
&= \int_0^a e^{-st}\,0\,dt + \int_a^\infty e^{-st}\,1\,dt \\
&= -\frac{1}{s}\,e^{-st}\Big|_a^\infty = \frac{e^{-as}}{s}.
\end{aligned}$$

EXEMPLE 8.11. Soit

$$f(t) = \begin{cases} 2, & \text{si } 0 < t < \pi, \\ 0, & \text{si } \pi < t < 2\pi, \\ \sin t, & \text{si } 2\pi < t, \end{cases}$$

(V. figure 8.5). Trouver $F(s)$.

RÉSOLUTION. On récrit $f(t)$ au moyen de la fonction d'Heaviside et de la 2π-périodicité de $\sin t$:

$$f(t) = 2 - 2u(t-\pi)1 + u(t-2\pi)\sin(t-2\pi).$$

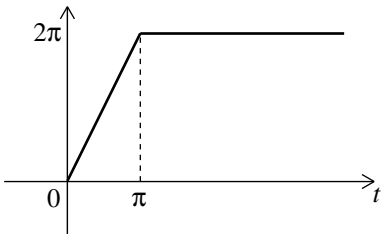

FIGURE 8.6. La fonction $f(t)$ de l'exemple 8.12.

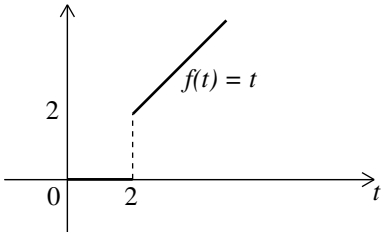

FIGURE 8.7. La fonction $f(t)$ de l'exemple 8.13.

Alors,

$$\begin{aligned} F(s) &= 2\mathcal{L}(1) - 2\mathcal{L}\big(u(t-\pi)1(t-\pi)\big) + \mathcal{L}\big(u(t-2\pi)\sin(t-2\pi)\big) \\ &= \frac{2}{s} - e^{-\pi s}\frac{2}{s} + e^{-2\pi s}\frac{1}{s^2+1}. \quad \square \end{aligned}$$

EXEMPLE 8.12. Soit

$$f(t) = \begin{cases} 2t, & \text{si } 0 < t < \pi, \\ 2\pi, & \text{si } \pi < t, \end{cases}$$

(V. figure 8.6). Trouver $F(s)$.

RÉSOLUTION. On récrit $f(t)$ au moyen de la fonction d'Heaviside:

$$\begin{aligned} f(t) &= 2t - u(t-\pi)(2t) + u(t-\pi)2\pi \\ &= 2t - 2u(t-\pi)(t-\pi). \end{aligned}$$

Alors,

$$F(s) = \frac{2\times 1!}{s^2} - 2\,e^{-\pi s}\,\frac{1}{s^2}. \quad \square$$

EXEMPLE 8.13. Soit

$$f(t) = \begin{cases} 0, & \text{si } 0 \le t < 2, \\ t, & \text{si } 2 < t, \end{cases}$$

(V. figure 8.7). Trouver $F(s)$.

RÉSOLUTION. On récrit $f(t)$ au moyen de la fonction d'Heaviside:

$$\begin{aligned} f(t) &= u(t-2)t \\ &= u(t-2)(t-2) + u(t-2)2. \end{aligned}$$

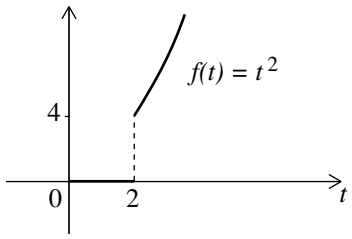

FIGURE 8.8. La fonction $f(t)$ de l'exemple 8.14.

Alors,

$$F(s) = e^{-2s}\,\frac{1!}{s^2} + 2\,e^{-2s}\,\frac{0!}{s} = e^{-2s}\left[\frac{1}{s^2} + \frac{2}{s}\right]. \quad \square$$

EXEMPLE 8.14. Soit

$$f(t) = \begin{cases} 0, & \text{si } 0 \le t < 2, \\ t^2, & \text{si } 2 < t, \end{cases}$$

(V. figure 8.8). Trouver $F(s)$.

RÉSOLUTION. **(a) Résolution analytique.**— On récrit $f(t)$ au moyen de la fonction d'Heaviside:

$$\begin{aligned} f(t) &= u(t-2)t^2 \\ &= u(t-2)\big[(t-2)+2\big]^2 \\ &= u(t-2)\left[(t-2)^2 + 4(t-2) + 4\right]. \end{aligned}$$

Alors,

$$F(s) = e^{-2s}\left[\frac{2!}{s^3} + \frac{4}{s^2} + \frac{4}{s}\right].$$

(b) Résolution par Matlab symbolique.—

```
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
```

$\square$

EXEMPLE 8.15. Soit

$$F(s) = e^{-\pi s}\,\frac{s}{s^2+4}.$$

Trouver $f(t)$.

RÉSOLUTION. **(a) Résolution analytique.**— On voit que

$$\begin{aligned} \mathcal{L}^{-1}F(s) &= u(t-\pi)\cos\big(2(t-\pi)\big) \\ &= \begin{cases} 0, & \text{si } 0 \le t < \pi, \\ \cos\big(2(t-\pi)\big) = \cos 2t, & \text{si } \pi < t. \end{cases} \end{aligned}$$

On a tracé $f(t)$ dans la figure 8.9.

(b) Résolution par Matlab symbolique.—

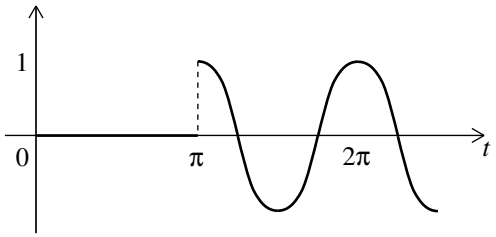

FIGURE 8.9. La fonction $f(t)$ de l'exemple 8.15.

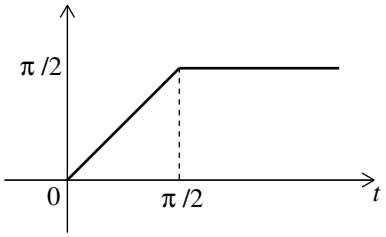

FIGURE 8.10. La fonction $g(t)$ de l'exemple 8.16.

```
>> syms s;
>> F = exp(-pi*s)*s/(s^2+4);
>> f = ilaplace(F)
f = Heaviside(t-pi)*cos(2*t)
```

□

EXEMPLE 8.16. Résoudre le problème aux valeurs initiales:

$$y'' + 4y = g(t) = \begin{cases} t, & \text{si } 0 \leq t < \pi/2, \\ \pi/2, & \text{si } \pi/2 < t, \end{cases}$$
$$y(0) = 0, \qquad y'(0) = 0,$$

au moyen de la transformation de Laplace.

RÉSOLUTION. Posons

$$\mathcal{L}(y) = Y(s) \quad \text{et} \quad \mathcal{L}(g) = G(s).$$

Alors,

$$\begin{aligned} \mathcal{L}(y'' + 4y) &= s^2 Y(s) - sy(0) - y'(0) + 4Y(s) \\ &= (s^2 + 4)Y(s) \\ &= G(s), \end{aligned}$$

où l'on a utilisé les valeurs données de $y(0)$ et $y'(0)$. On a donc

$$Y(s) = \frac{G(s)}{s^2 + 4}.$$

On récrit $g(t)$ (V. figure 8.10) au moyen de la fonction d'Heaviside:

$$\begin{aligned} g(t) &= t - u(t-\pi/2)t + u(t-\pi/2)\frac{\pi}{2} \\ &= t - u(t-\pi/2)(t-\pi/2), \end{aligned}$$

d'où

$$G(s) = \frac{1}{s^2} - e^{-(\pi/2)s}\,\frac{1}{s^2} = \Big[1 - e^{-(\pi/2)s}\Big]\,\frac{1}{s^2}.$$

Donc

$$Y(s) = \Big[1 - e^{-(\pi/2)s}\Big]\,\frac{1}{(s^2+4)s^2}.$$

On décompose la fonction rationnelle au 2ème terme en fractions simples:

$$\frac{1}{(s^2+4)s^2} = \frac{A}{s} + \frac{B}{s^2} + \frac{Cs+D}{s^2+4}.$$

On chasse les dénominateurs:

$$\begin{aligned} 1 &= (s^2+4)sA + (s^2+4)B + s^2(Cs+D) \\ &= (A+C)s^3 + (B+D)s^2 + 4As + 4B. \end{aligned}$$

On identifie les coefficients:

$$\begin{aligned} 4A = 0 &\implies A = 0, \\ 4B = 1 &\implies B = \frac{1}{4}, \\ B + D = 0 &\implies D = -\frac{1}{4}, \\ A + C = 0 &\implies C = 0, \end{aligned}$$

d'où

$$\frac{1}{(s^2+4)s^2} = \frac{1}{4}\,\frac{1}{s^2} - \frac{1}{4}\,\frac{1}{s^2+4}.$$

Donc

$$Y(s) = \frac{1}{4}\,\frac{1}{s^2} - \frac{1}{8}\,\frac{2}{s^2+2^2} - \frac{1}{4}\,e^{-(\pi/2)s}\frac{1}{s^2} + \frac{1}{8}\,e^{-(\pi/2)s}\frac{2}{s^2+2^2}$$

et, par la transformation inverse de Laplace,

$$y(t) = \frac{1}{4}\,t - \frac{1}{8}\sin 2t - \frac{1}{4}\,u(t-\pi/2)(t-\pi/2) + \frac{1}{8}\,u(t-\pi/2)\sin\big(2[t-\pi/2]\big). \quad \square$$

On peut aussi trouver l'originale de la transformée

$$Y(s) = \frac{1}{2}\,\Big[1 - e^{-(\pi/2)s}\Big]\,\frac{2}{(s^2+4)s^2}$$

de l'exemple 8.16 par deux integrations au moyen de la formule (8.11) du théorème 8.5:

$$\begin{aligned} \mathcal{L}^{-1}\left(\frac{1}{s}\,\frac{2}{s^2+2^2}\right) &= \int_0^t \sin 2\tau\,d\tau = \frac{1}{2} - \frac{1}{2}\,\cos(2t), \\ \mathcal{L}^{-1}\left(\frac{1}{s}\,\left[\frac{1}{s}\frac{2}{s^2+2^2}\right]\right) &= \frac{1}{2}\int_0^t (1-\cos 2\tau)\,d\tau = \frac{t}{2} - \frac{1}{4}\,\sin(2t). \end{aligned}$$

On obtient alors l'originale $y(t)$ par la formule (8.15) du théorème 8.7.

8.4. La fonction delta de Dirac

Soit la fonction

$$f_k(t;a) = \begin{cases} 1/k, & \text{si } a \le t \le a+k, \\ 0, & \text{sinon.} \end{cases} \tag{8.18}$$

On voit que l'intégrale de $f_k(t;a)$ est 1:

$$I_k = \int_0^\infty f_k(t;a)\,dt = \int_a^{a+k} \frac{1}{k}\,dt = 1. \tag{8.19}$$

On note

$$\delta(t-a)$$

la limite de $f_k(t;a)$ quand $k \to 0$, Cette limite s'appelle *fonction delta de Dirac*, ou plus précisément, masse ou mesure de Dirac.

On peut représenter $f_k(t;a)$ au moyen de la différence de deux fonctions d'Heaviside:

$$f_k(t;a) = \frac{1}{k}\,[u(t-a) - u(t-(a+k))].$$

De (8.17) on obtient

$$\mathcal{L}\big(f_k(t;a)\big) = \frac{1}{ks}\left[e^{-as} - e^{-(a+k)s}\right] = e^{-as}\,\frac{1-e^{-ks}}{ks}. \tag{8.20}$$

Le quotient au dernier membre tend vers 1 quand $k \to 0$ comme on peut voir par la règle de l'Hôpital. On a donc

$$\mathcal{L}\big(\delta(t-a)\big) = e^{-as}. \tag{8.21}$$

Matlab symbolique produit la transformée de Laplace de la fonction symbolique $\delta(t)$:

```
>> f = sym('Dirac(t)');
>> F = laplace(f)
F = 1
```

Exemple 8.17. Résoudre le système amorti

$$y'' + 3y' + 2y = \delta(t-a), \qquad y(0) = 0, \qquad y'(0) = 0,$$

au repos pour $0 \le t < a$ et percuté à l'instant $t = a$.

Résolution. Par (8.21), la transformée de l'équation différentielle est

$$s^2Y + 3sY + 2Y = e^{-as}.$$

On isole $Y(s)$:

$$Y(s) = F(s)\,e^{-as},$$

où

$$F(s) = \frac{1}{(s+1)(s+2)} = \frac{1}{s+1} - \frac{1}{s+2}.$$

Alors,

$$f(t) = \mathcal{L}^{-1}(F) = e^{-t} - e^{-2t}.$$

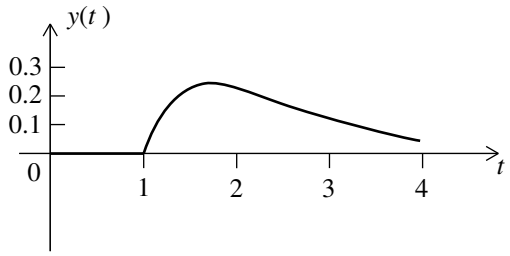

FIGURE 8.11. La solution $y(t)$ de l'exemple 8.17.

Donc par (8.15),

$$\begin{aligned} y(t) &= \mathcal{L}^{-1}\left(e^{-as}F(s)\right) \\ &= f(t-a)u(t-a) \\ &= \begin{cases} 0, & \text{si } 0 \le t < a, \\ e^{-(t-a)} - e^{-2(t-a)}, & \text{si } t > a. \end{cases} \end{aligned}$$

On illustre la solution dans la figure 8.11 pour $a = 1$. □

8.5. Dérivée et intégrale de la transformée

On démontre les formules suivantes.

THÉORÈME 8.8. *Si $F(s) = \mathcal{L}\{f(t)\}(s)$, alors*

$$\mathcal{L}\{tf(t)\}(s) = -F'(s), \tag{8.22}$$

ou, en se servant de la transformation de Laplace inverse,

$$\mathcal{L}^{-1}\{F'(s)\} = -tf(t). \tag{8.23}$$

De plus, si la limite

$$\lim_{t\to 0+} \frac{f(t)}{t}$$

existe, alors

$$\mathcal{L}\left\{\frac{f(t)}{t}\right\}(s) = \int_s^\infty F(\tilde{s})\,d\tilde{s}, \tag{8.24}$$

ou, en se servant de la transformation de Laplace inverse,

$$\mathcal{L}^{-1}\left\{\int_s^\infty F(\tilde{s})\,d\tilde{s}\right\} = \frac{1}{t}\,f(t). \tag{8.25}$$

DÉMONSTRATION. Soit

$$F(s) = \int_0^\infty e^{-st} f(t)\,dt.$$

Par le théorème 8.2, (8.22) suit par dérivation:

$$\begin{aligned} F'(s) &= -\int_0^\infty e^{-st}[tf(t)]\,dt \\ &= -\mathcal{L}\{tf(t)\}(s). \end{aligned}$$

D'autre part, par le théorème 8.2, (8.24) suit par intégration:

$$\begin{aligned}\int_s^\infty F(\tilde{s})\,d\tilde{s} &= \int_s^\infty \int_0^\infty e^{-\tilde{s}t} f(t)\,dt\,d\tilde{s}\\ &= \int_0^\infty f(t)\left[\int_s^\infty e^{-\tilde{s}t} d\tilde{s}\right] dt\\ &= \int_0^\infty f(t)\left[-\frac{1}{t}\,e^{-\tilde{s}t}\right]\Big|_{\tilde{s}=s}^{\tilde{s}=\infty} dt\\ &= \int_0^\infty e^{-st}\left[\frac{1}{t}\,f(t)\right] dt\\ &= \mathcal{L}\left\{\frac{1}{t}\,f(t)\right\}. \quad \square\end{aligned}$$

On généralise la formule (8.22) dans le théorème suivant.

THÉORÈME 8.9. *Si $t^n f(t)$ est transformable, alors*

$$\mathcal{L}\{t^n f(t)\}(s) = (-1)^n F^{(n)}(s), \tag{8.26}$$

ou, réciproquement,

$$\mathcal{L}^{-1}\{F^{(n)}(s)\} = (-1)^n t^n f(t). \tag{8.27}$$

EXEMPLE 8.18. Obtenir l'originale de $\dfrac{1}{(s+1)^2}$ par (8.23).

RÉSOLUTION. Posons

$$\frac{1}{(s+1)^2} = -\frac{d}{ds}\left(\frac{1}{s+1}\right) =: -F'(s).$$

Alors par (8.23)

$$\begin{aligned}\mathcal{L}^{-1}\{-F'(s)\} = tf(t) &= t\,\mathcal{L}^{-1}\left\{\frac{1}{s+1}\right\}\\ &= t\,e^{-t}. \quad \square\end{aligned}$$

EXEMPLE 8.19. Etant donné $f(t)$ trouver $F(s)$ par (8.22).

	$f(t)$	$F(s)$
(8.28)	$\dfrac{1}{2\beta^3}\,[\sin\beta t - \beta t\cos\beta t]$	$\dfrac{1}{(s^2+\beta^2)^2}$,
(8.29)	$\dfrac{t}{2\beta}\sin\beta t$	$\dfrac{s}{(s^2+\beta^2)^2}$,
(8.30)	$\dfrac{1}{2\beta}\,[\sin\beta t + \beta t\cos\beta t]$	$\dfrac{s^2}{(s^2+\beta^2)^2}$.

RÉSOLUTION. **(a) Résolution analytique.**— On applique (8.22) au 1er membre de (8.29):

$$\begin{aligned}\mathcal{L}\big(t\sin\beta t\big)(s) &= -\frac{d}{ds}\left[\frac{\beta}{s^2+\beta^2}\right]\\ &= \frac{2\beta s}{(s^2+\beta^2)^2},\end{aligned}$$

d'où, après division par 2β, on obtient le 2ème membre de (8.29).

De même, par (8.22), on obtient

$$\begin{aligned}\mathcal{L}\big(t\cos\beta t\big)(s) &= -\frac{d}{ds}\left[\frac{s}{s^2+\beta^2}\right]\\ &= -\frac{s^2+\beta^2-2s^2}{(s^2+\beta^2)^2}\\ &= \frac{s^2-\beta^2}{(s^2+\beta^2)^2}.\end{aligned}$$

Alors

$$\begin{aligned}\mathcal{L}\left(t\cos\beta t\pm\frac{1}{\beta}\,\sin\beta t\right)(s) &= \frac{s^2-\beta^2}{(s^2+\beta^2)^2}\pm\frac{1}{s^2+\beta^2}\\ &= \frac{s^2-\beta^2\pm(s^2+\beta^2)}{(s^2+\beta^2)^2}.\end{aligned}$$

Si l'on prend le signe + et l'on divise par deux, on obtient (8.30). Si l'on prend le signe − et l'on divise par $-2\beta^2$, on obtient (8.28).

(b) Résolution par Matlab symbolique.—

```
>> 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
```

□

Exemple 8.20. Trouver

$$\mathcal{L}^{-1}\left\{\ln\left(1+\frac{\omega^2}{s^2}\right)\right\}(t).$$

RÉSOLUTION. **(a) Résolution analytique.**— On a

$$\begin{aligned}
-\frac{d}{ds}\ln\left(1+\frac{\omega^2}{s^2}\right) &= -\frac{d}{ds}\ln\left(\frac{s^2+\omega^2}{s^2}\right)\\
&= -\frac{s^2}{s^2+\omega^2}\,\frac{2s^3-2s(s^2+\omega^2)}{s^4}\\
&= \frac{2\omega^2}{s(s^2+\omega^2)}\\
&= 2\,\frac{(\omega^2+s^2)-s^2}{s(s^2+\omega^2)}\\
&= \frac{2}{s}-2\,\frac{s}{s^2+\omega^2}\\
&=: F(s).
\end{aligned}$$

Alors

$$f(t)=\mathcal{L}^{-1}(F)=2-2\cos\omega t.$$

Puisque

$$\frac{f(t)}{t}=2\omega\,\frac{1-\cos\omega t}{\omega t}\to 0 \qquad \text{quand} \quad t\to 0,$$

par (8.25) on a

$$\begin{aligned}
\mathcal{L}^{-1}\left\{\ln\left(1+\frac{\omega^2}{s^2}\right)\right\} &= \mathcal{L}^{-1}\left\{\int_s^\infty F(\tilde{s})\,d\tilde{s}\right\}\\
&= \frac{1}{t}\,f(t)\\
&= \frac{2}{t}\,(1-\cos\omega t).
\end{aligned}$$

(b) Résolution par Matlab symbolique.—

```
>> syms omega t s
>> F = log(1+(omega^2/s^2));
>> f = ilaplace(F,s,t)
f = 2/t-2/t*cos(omega*t)
```

□

8.6. Équation différentielle de Laguerre

On peut résoudre certaines équations différentielles à coefficients variables de la forme $at+b$ au moyen de la transformation de Laplace. En effet, par (8.22), (8.6) et (8.7), on obtient

$$\begin{aligned}
\mathcal{L}\big(ty'(t)\big) &= -\frac{d}{ds}[sY(s)-y(0)] \qquad (8.31)\\
&= -Y(s)-sY'(s),\\
\mathcal{L}\big(ty''(t)\big) &= -\frac{d}{ds}[s^2Y(s)-sy(0)-y'(0)] \qquad (8.32)\\
&= -2sY(s)-s^2Y'(s)+y(0).
\end{aligned}$$

EXEMPLE 8.21. Trouver les solutions polynomiales $L_n(t)$ de l'équation de Laguerre:

$$ty'' + (1-t)y' + ny = 0, \qquad n = 0, 1, \ldots. \tag{8.33}$$

RÉSOLUTION. La transformée de Laplace de l'équation (8.33) est

$$\begin{aligned} &-2sY(s) - s^2Y'(s) + y(0) + sY(s) - y(0) + Y(s) + sY'(s) + nY(s) \\ &\qquad = (s - s^2)Y'(s) + (n + 1 - s)Y(s) = 0. \end{aligned}$$

Cette équation est séparable:

$$\begin{aligned} \frac{dY}{Y} &= \frac{n+1-s}{(s-1)s}\, ds \\ &= \left(\frac{n}{s-1} - \frac{n+1}{s} \right) ds, \end{aligned}$$

d'où sa solution

$$\begin{aligned} \ln|Y(s)| &= n \ln|s-1| - (n+1)\ln s \\ &= \ln \left| \frac{(s-1)^n}{s^{n+1}} \right|, \end{aligned}$$

c'est-à-dire

$$Y(s) = \frac{(s-1)^n}{s^{n+1}}.$$

Posons

$$L_n(t) = \mathcal{L}^{-1}(Y)(t),$$

où, par exception, la lettre majuscule L dans L_n désigne une fonction de t. En effet, on note $L_n(t)$ le polynôme de Laguerre de degré n. On montre que

$$L_0(t) = 1, \qquad L_n(t) = \frac{e^t}{n!} \frac{d^n}{dt^n} \left(t^n e^{-t} \right), \qquad n = 1, 2, \ldots.$$

On voit bien que $L_n(t)$ est un polynôme de degré n puisque les exponentielles s'annulent après la différentiation. Puisque, par le théorème 8.4,

$$\mathcal{L}\big(f^{(n)}\big)(s) = s^n F(s) - s^{n-1} f(0) - s^{n-2} f'(0) - \cdots - f^{(n-1)}(0),$$

on a

$$\mathcal{L} \left\{ \left(t^n e^{-t} \right)^{(n)} \right\} (s) = s^n \frac{n!}{(s+1)^{n+1}},$$

et, par conséquent,

$$\begin{aligned} &\mathcal{L} \left\{ \frac{e^t}{n!} \left(t^n e^{-t} \right)^{(n)} \right\} \\ &\qquad = \frac{n!}{n!} \frac{(s-1)^n}{s^{n+1}} \\ &\qquad = Y(s) \qquad\qquad = \mathcal{L}(L_n). \quad \square \end{aligned}$$

Les 4 premiers polynômes de Laguerre sont (V. figure 8.12):

$$\begin{aligned} L_0(x) &= 1, & L_1(x) &= 1 - x, \\ L_2(x) &= 1 - 2x + \frac{1}{2}x^2, & L_3(x) &= 1 - 3x + \frac{3}{2}x^2 - \frac{1}{6}x^3. \end{aligned}$$

On peut obtenir les $L_n(x)$ par la récurrence:

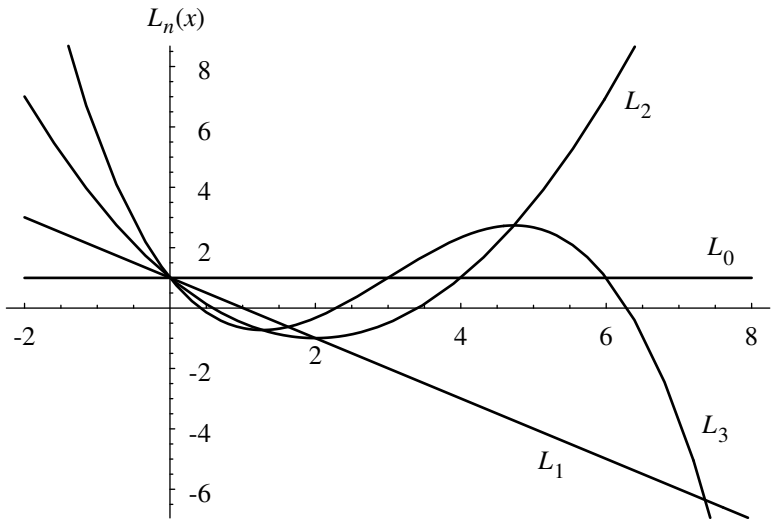

FIGURE 8.12. Les 4 premiers polynômes de LAGUERRE.

$$(n+1)L_{n+1}(x) = (2n+1-x)L_n(x) - nL_{n-1}(x).$$

Les polynômes de Laguerre satisfont les relations d'orthogonalité suivantes avec le poids $p(x) = e^{-x}$:

$$\int_0^\infty e^{-x} L_m(x) L_n(x)\, dx = \begin{cases} 0, & m \neq n, \\ 1, & m = n. \end{cases}$$

Matlab produit les polynômes de Laguerre:

```
>>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
```

ainsi de suite. La commande `simple` de Matlab symbolique a pour but, nonorthodoxe en mathématiques, de simplifier une expression afin d'en réduire le nombre de caractères.

8.7. Convolution

L'originale du produit de deux transformées est la convolution des deux originales.

DÉFINITION 8.4. La convolution de $f(t)$ et de $g(t)$, notée $(f * g)(t)$, est la fonction

$$h(t) = \int_0^t f(\tau)g(t-\tau)\,d\tau. \tag{8.34}$$

On dit "f convoluée avec g".

On vérifie que la convolution est commutative:

$$\begin{aligned}(f*g)(t) &= \int_0^t f(\tau)g(t-\tau)\,d\tau \\ &\qquad \text{(on pose } t-\tau=\sigma,\ d\tau=-d\sigma) \\ &= -\int_t^0 f(t-\sigma)g(\sigma)\,d\sigma \\ &= \int_0^t g(\sigma)f(t-\sigma)\,d\sigma \\ &= (g*f)(t).\end{aligned}$$

THÉORÈME 8.10. *Soit*

$$F(s)=\mathcal{L}(f),\quad G(s)=\mathcal{L}(g),\quad H(s)=F(s)G(s),\quad h(t)=\mathcal{L}^{-1}(H).$$

Alors

$$h(t) = (f*g)(t) = \mathcal{L}^{-1}\big(F(s)G(s)\big). \tag{8.35}$$

DÉMONSTRATION. Par définition et par (8.16), on a

$$\begin{aligned}e^{-s\tau}G(s) &= \mathcal{L}\big(g(t-\tau)u(t-\tau)\big) \\ &= \int_0^\infty e^{-st}g(t-\tau)u(t-\tau)\,dt \\ &= \int_\tau^\infty e^{-st}g(t-\tau)\,dt.\end{aligned}$$

Par ceci et par la définition de $F(s)$ on a

$$\begin{aligned}F(s)G(s) &= \int_0^\infty e^{-s\tau}f(\tau)G(s)\,d\tau \\ &= \int_0^\infty f(\tau)\left[\int_\tau^\infty e^{-st}g(t-\tau)\,dt\right]d\tau, \qquad (s>\gamma) \\ &= \int_0^\infty e^{-st}\left[\int_0^t f(\tau)g(t-\tau)\,d\tau\right]dt \\ &= \mathcal{L}\big[(f*g)(t)\big](s) \\ &= \mathcal{L}(h)(s).\end{aligned}$$

On illustre dans la figure 8.13 le domaine d'intégration dans le plan $t\tau$ utilisé dans la démonstration du théorème 8.10. □

EXEMPLE 8.22. Caluler $(1*1)(t)$.

RÉSOLUTION.

$$(1*1)(t) = \int_0^t 1\times 1\,d\tau = t. \quad \square$$

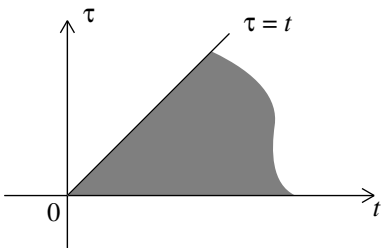

FIGURE 8.13. Région d'intégration du plan $t\tau$ utilisée dans la démonstration du théorème 8.10.

EXEMPLE 8.23. Caluler $e^t * e^t$.

RÉSOLUTION.

$$\begin{aligned} e^t * e^t &= \int_0^t e^\tau \, e^{t-\tau} \, d\tau \\ &= \int_0^t e^t \, d\tau = t \, e^t. \quad \square \end{aligned}$$

EXEMPLE 8.24. Caluler l'originale de

$$\frac{1}{(s-a)(s-b)}, \qquad a \neq b,$$

au moyen de la convolution.

RÉSOLUTION.

$$\begin{aligned} \mathcal{L}^{-1}\left[\frac{1}{(s-a)(s-b)}\right] &= e^{at} * e^{bt} \\ &= \int_0^t e^{a\tau} \, e^{b(t-\tau)} \, d\tau \\ &= e^{bt} \int_0^t e^{(a-b)\tau} \, d\tau \\ &= e^{bt} \frac{1}{a-b} \, e^{(a-b)\tau} \Big|_0^t \\ &= \frac{e^{bt}}{a-b} \left[e^{(a-b)t} - 1\right] \\ &= \frac{e^{at} - e^{bt}}{a-b}. \quad \square \end{aligned}$$

On peut résoudre certaines équations intégrales au moyen de la transformation de Laplace.

EXEMPLE 8.25. Résoudre l'équation intégrale

$$y(t) = t + \int_0^t y(\tau) \sin(t-\tau) \, d\tau. \tag{8.36}$$

RÉSOLUTION. Le dernier terme de (8.36) est une convolution. Alors

$$y(t) = t + y * \sin t.$$

Donc

$$Y(s) = \frac{1}{s^2} + Y(s)\,\frac{1}{s^2+1},$$

d'où

$$Y(s) = \frac{s^2+1}{s^4} = \frac{1}{s^2} + \frac{1}{s^4}.$$

Enfin,

$$y(t) = t + \frac{1}{6}\,t^3. \quad \square$$

8.8. Fractions simples

La décomposition d'une fonction rationnelle en fractions simples a été étudiée dans le cours de calcul différentiel et intégral.

Il suffit de mentionner que si $p(\lambda)$ est le polynôme caractéristique d'une équation différentielle $Ly = r(t)$ à coefficients constants, la mise en facteur de $p(\lambda)$ requise pour trouver les zéros de p et par conséquent les solutions indépendantes de $Ly = 0$, est aussi requise pour la réduction de $1/p(s)$ en fractions simples avec l'emploi de la transformation de Laplace.

Le phénomène de résonnance correspond à des racines multiples.

Avec la boîte symbolique étendue de la version professionnelle de Matlab, on peut obtenir le développement en fractions simples en recourant au noyau complet de Mable par la commande `convert`, elle-même référencée par la commande `mhelp convert[parfrac]`.

8.9. Transformées de fonctions périodiques

DÉFINITION 8.5. Une fonction $f(t)$ définie pour tout $t > 0$ est périodique de période p, $p > 0$, si

$$f(t+p) = f(t), \qquad \text{pour tout } t > 0. \tag{8.37}$$

THÉORÈME 8.11. *Soit $f(t)$ une fonction p-périodique. Alors*

$$\mathcal{L}(f)(s) = \frac{1}{1-e^{-ps}} \int_0^p e^{-st} f(t)\,dt, \qquad s > 0. \tag{8.38}$$

DÉMONSTRATION. On exploite la périodicité de f:

$$\begin{aligned}\mathcal{L}(f)(s) &= \int_0^\infty e^{-st} f(t)\,dt \\ &= \int_0^p e^{-st} f(t)\,dt + \int_p^{2p} e^{-st} f(t)\,dt + \int_{2p}^{3p} e^{-st} f(t)\,dt + \dots.\end{aligned}$$

On substitue

$$t = \tau + p, \qquad t = \tau + 2p, \qquad \dots,$$

respectivement dans la 2ème, la 3ème intégrales, etc. Les nouvelles limites d'intégration sont alors 0 et p. Donc

$$\begin{aligned}\mathcal{L}(f)(s) &= \int_0^p e^{-st} f(t)\,dt + \int_0^p e^{-s(t+p)} f(t)\,dt + \int_0^p e^{-s(t+2p)} f(t)\,dt + \dots \\ &= \left(1 + e^{-sp} + e^{-2sp} + \dots\right) \int_0^p e^{-st} f(t)\,dt \\ &= \frac{1}{1-e^{-ps}} \int_0^p e^{-st} f(t)\,dt. \quad \square\end{aligned}$$

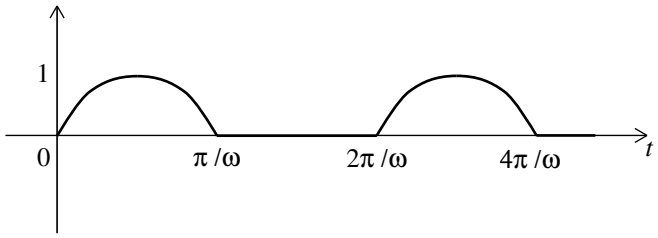

FIGURE 8.14. Semi-rectification de l'onde de l'exemple 8.26.

EXEMPLE 8.26. Trouver la transformée de Laplace de la semi-rectification de l'onde

$$f(t) = \sin \omega t$$

(V. figure 8.14).

(A) RÉSOLUTION ANALYTIQUE. L'onde semi-rectifiée de période $p = 2\pi/\omega$ est

$$f(t) = \begin{cases} \sin \omega t, & \text{si } 0 < t < \pi/\omega, \\ 0, & \text{si } \pi/\omega < t < 2\pi/\omega. \end{cases}$$

Par (8.38)

$$\mathcal{L}(f)(s) = \frac{1}{1 - e^{-2\pi s/\omega}} \int_0^{\pi/\omega} e^{-st} \sin \omega t \, dt.$$

On remarque que cette intégrale est la partie imaginaire de l'intégrale

$$\begin{aligned} \int_0^{\pi/\omega} e^{(-s+i\omega)t} \, dt &= \frac{1}{-s+i\omega} \, e^{(-s+i\omega)t} \Big|_0^{\pi/\omega} \\ &= \frac{-s-i\omega}{s^2+\omega^2} \left(-e^{-s\pi/\omega} - 1 \right). \end{aligned}$$

Alors, par la formule

$$1 - e^{-2\pi s/\omega} = \left(1 + e^{-\pi s/\omega} \right) \left(1 - e^{-\pi s/\omega} \right),$$

on obtient

$$\begin{aligned} \mathcal{L}(f)(s) &= \frac{\omega \left(1 + e^{-\pi s/\omega} \right)}{(s^2+\omega^2) \left(1 - e^{-2\pi s/\omega} \right)} \\ &= \frac{\omega}{(s^2+\omega^2) \left(1 - e^{-\pi s/\omega} \right)}. \end{aligned}$$

(b) Résolution par Matlab symbolique.—

```
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)
```

□

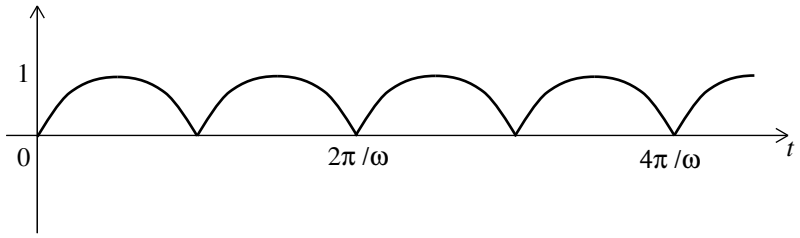

FIGURE 8.15. Rectification de l'onde de l'exemple 8.27.

EXEMPLE 8.27. Trouver la transformée de Laplace de la rectification totale de l'onde

$$f(t) = \sin \omega t$$

(V. figure 8.15).

RÉSOLUTION. L'onde rectifiée de période $p = 2\pi/\omega$ est

$$f(t) = |\sin \omega t| = \begin{cases} \sin \omega t, & \text{si } 0 < t < \pi\omega, \\ -\sin \omega t, & \text{si } \pi < t < 2\pi\omega. \end{cases}$$

Par la méthode utilisée à l'exemple 8.26, on obtient

$$\mathcal{L}(f)(s) = \frac{\omega}{s^2 + \omega^2} \coth \frac{\pi s}{2\omega}. \quad \square$$

CHAPITRE 9

Introduction aux méthodes numériques

9.1. Calculs

9.1.1. Définitions. 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 **absolute error** in $\tilde{a}$ and 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 **truncation error** occurs when (for instance) 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 9.1.2). The term "truncation error" is also used for the error committed when an infinite series is truncated after a finite number of terms.

(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

$$0.1203 \times 10^2, \quad 0.1230 \times 10^{-2} \quad 0.1000 \times 10^3.$$

is 4, 3, and 1, respectively.

REMARQUE 9.1. 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:

$$0.001203, \quad 12300.04$$

9.1.2. Nombres arrondis ou tronqués. 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

$$0.b_1 b_2 \dots b_{k-1}b_k \times 10^N.$$

EXEMPLE 9.1. 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.

9.1.3. Cancellation dans les calculs. 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 9.2 illustrates these points.

EXEMPLE 9.2. Solve the quadratic equation

$$x^2 - 1634x + 2 = 0.$$

SOLUTION. The usual formula yields

$$x = 817 \pm \sqrt{667\,487}.$$

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 9.2, it is seen that a numerically stable formula for solving the quadratic equation

$$ax^2 + bx + c = 0, \qquad a \neq 0,$$

is

$$x_1 = \frac{1}{2a}\left[-b - \operatorname{sgn}(b)\sqrt{b^2 - 4ac}\right], \qquad x_2 = \frac{c}{ax_1},$$

where the signum function is

$$\operatorname{sgn}(x) = \begin{cases} +1, & \text{if} \quad x \geq 0, \\ -1, & \text{if} \quad x < 0. \end{cases}$$

EXEMPLE 9.3. 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

$$4.805 - 12.7525 < x - y < 4.815 - 12.7515 \Leftrightarrow -7.9475 < x - y < -7.9365. \quad \square$$

EXEMPLE 9.4. Find the absolute and relative errors 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 absolute and relative errors 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 absolute and relative errors 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 absolute and relative errors 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 absolute and relative errors in 355/113 rounded to three digits are

$$\epsilon = 0.267 \times 10^{-6} \quad \text{and} \quad \epsilon_r = 0.849 \times 10^{-7}. \quad \square$$

9.2. Résolution des équations nonlinéaires par récurrence

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

$$f(x) = 0. \tag{9.1}$$

A **root** of the equation $f(x) = 0$ or a **zero** of $f(x)$ is a number α such that $f(\alpha) = 0$

The following results from elementary calculus are needed to justify the methods of solution presented here.

THÉORÈME 9.1 (Intermediate value theorem). *Let $a < b$ and $f(x)$ be a continuous function on $[a, b]$. If A is a number strictly between $f(a)$ and $f(b)$, then there exists a number c such that $a < c < b$ and $f(c) = A$.*

COROLLAIRE 9.1. *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 $A = 0$. $\square$

THÉORÈME 9.2 (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*

$$f(\alpha) \leq f(x) \leq f(\beta).$$

THÉORÈME 9.3 (Mean value theorem). *Let $a < b$ and $f(x)$ be a continuous function on $[a, b]$ which is differentiable on $]a, b[$. There exists a number c such that $a < c < b$ and*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

THÉORÈME 9.4 (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*

$$\int_a^b f(x)\, g(x)\, dx = f(c) \int_a^b g(x)\, dx.$$

9.2.1. Récurrence pour point fixe. To find a root of equation (9.1), we rewrite

$$f(x) = 0,$$

in an equivalent form

$$x = g(x), \tag{9.2}$$

for instance, $g(x) = x - f(x)$.

If for a given initial value x_0, the sequence x_0, x_1, ..., defined by the recurrence

$$x_{n+1} = g(x_n), \qquad n = 0, 1, \ldots, \tag{9.3}$$

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 (9.3) as $n \to \infty$. The number p is called a **fixed point** for $g(x)$. We say that (9.1) and (9.2) are **equivalent** (on a given interval) if any root of (9.1) is a fixed point for (9.2) and vice-versa.

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.

DÉFINITION 9.1. 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 $g(x)$ satisfies

$$|g'(p)| < 1, \quad |g'(p)| > 1, \quad \text{or} \quad |g'(p)| = 1,$$

respectively.

THÉORÈME 9.5. *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 $|g'(x)| \leq K$ 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 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 the mean value theorem 9.3, 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 9.3, there exists a number c between p and q (and hence in $[a, b]$) such that

$$|p - q| = |g(p) - g(q)| = |g'(c)|\,|p - q| \leq K|p - q| < |p - q|,$$

which is a contradiction. Thus $p = q$ and the attractive fixed point in $[a, b]$ is unique. We now prove convergence. By the intermediate value theorem 9.1, 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,

$$|g(x) - g(y)| \leq K|x - y|.$$

In particular,

$$|x_{n+1} - p| = |g(x_n) - g(p)| \leq K|x_n - p|.$$

Repeating this procedure $n + 1$ times, we have

$$|x_{n+1} - p| \leq K^{n+1}|x_0 - p| \to 0, \quad \text{as} \quad n \to \infty$$

since $0 < K < 1$. Thus the sequence $\{x_n\}$ converges to p. □

EXEMPLE 9.5. 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 the find a fixed point for

$$g(x) = (9 - x^3)/9.$$

By Corollary 9.1, we know that $f(x)$ has a root p between 0 and 1 since

$$f(0)f(1) = -9 < 0.$$

Condition (3) of Theorem 9.5 is satisfied with $K = 1/3$ since

$$|g'(x)| = | - x^2/3| \leq 1/3$$

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_5.m` is

```
function x1 = exp8_5(x0); %MAT 2331, Example 8.5.
x1 = (9-x0^2)/9;
```

The exact solution:

$$0.914\,907\,841\,533\,66,$$

is obtained by means of some 30 iterations. The following iterative procedure solves the problem.

```
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_5(x(i,2));
x(i+1,:) = [i xt (xt-xexact), (xt-xexact)/x(i,4)];
end
```

The iterates, their errors and the ratios of successive errors are listed in Table 8.1. One sees that the ratios of successive errors are decreasing; therefore the order of convergence, defined in Subsection 9.2.3, is greater than one, but smaller than two since the number of correct digits does not double from one iterate to the next. □

In Example 9.6 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. In fact, besides fixed points, an iterative scheme may have cycles which are defined in Definition 9.2.

TABLE 1. Results of Example 9.5.

n	x_n	abs. err. ϵ_n	$\epsilon_n/\epsilon_{n-1}$
0	0.50000000000000	−0.41490784153366	1.00000000000000
1	0.98611111111111	0.07120326957745	0.07120326957745
2	0.89345451579409	−0.02145332573957	−0.30129691890395
3	0.92075445888550	0.00584661735184	−0.01940483617658
4	0.91326607850598	−0.00164176302768	0.08460586900804
5	0.91536510274262	0.00045726120896	0.00540460389243

DÉFINITION 9.2. 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

$$x_1 = g(x_0), \quad x_2 = g^2(x_1), \quad \ldots, x_{k-1} = g^{k-1}(x_0), \quad x_0 = g^k(x_0).$$

The *multiplier* of a k cycle is

$$(g^k)'(x_j) = g'(x_{k-1}) \cdots g'(x_0),$$

where

$$g^{s+1}(x) = g\big(g^s(x)\big).$$

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.

EXEMPLE 9.6. 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 solutions 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.

TABLE 2. Results of Example 9.6.

	$g_1(x)$	$g_2(x)$	$g_3(x)$	$g_4(x)$
n	$10+x-4x^2-x^3$	$\sqrt{(10/x)-4x}$	$0.5\sqrt{10-x^3}$	$\frac{x^3+4x^2+10}{2x^2+8x}$
0	1.5	1.5	1.5	1.5
1	-0.8750	0.816	1.286953	1.356060
2	6.732421875	2.996	1.402540	1.366437
3	-4.6972001×10^2	$0.00 - 2.94\,i$	1.345458	1.365077
4	1.0275×10^8	$2.75 - 2.75\,i$	1.375170	1.365249
5	-1.08×10^{24}	$1.81 - 3.53\,i$	1.360094	1.365227
6	1.3×10^{72}	$2.38 - 3.43\,i$	1.367846	1.365230

Six iterations are performed with the following four rearrangements $x = g_j(x)$, $j = 1, 2, 3, 4$, of the given equation $f(x) = 0$. The derivative of $g_j'(x)$ is evaluated at the real root $x_\infty \approx 1.365$.

$$
\begin{aligned}
x = g_1(x) &=: 10 + x - 4x^2 - x^3, && g_1'(x_\infty) \approx -15.51,\\
x = g_2(x) &=: \sqrt{(10/x) - 4x}, && g_2'(x_\infty) \approx -3.42,\\
x = g_3(x) &=: 0.5\sqrt{10 - x^3}, && g_3'(x_\infty) \approx -0.51,\\
x = g_4(x) &=: \frac{x^3 + 4x^2 - 10}{3x^2 + 8x}, && g_4'(x_\infty) \approx -0.13.
\end{aligned}
$$

The Matlab function M-file `exp8_5.m` is

```
function y = exp8_6(x); %MAT 2331, Example 8.6.
y = [10+x(1)-4*x(1)^2-x(1)^3; sqrt((10/x(2))-4*x(2));
 sqrt(10-x(3)^3)/2; (10+4*x(4)^2+x(4)^3)/(8*x(4)+2*x(4)^2)]';
```

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=exp8_6(x(i,2:5));
x(i+1,:) = [i xt];
end
```

The results are summarized in Table 8.2. We see from the table that x_∞ is an attractive fixed point of $g_3(x)$ and $g_4(x)$. Moreover, $g_4(x_n)$ converges more quickly to the root 1.367 than $g_3(x_n)$. 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

$$g_2'(z_+)g'(z_-) = 0.19790433047378$$

which is smaller than one. 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 $-\infty$. □

REMARQUE 9.2. 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 could be accelerated by different acceleration processes.

9.2.2. Critères d'arrêt. 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.

9.2.3. Ordre et taux de convergence d'une méthode itérative. 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

$$\begin{aligned} x_{n+1} = g(x_n) = g(p + \epsilon_n) &= g(p) + g'(p)\epsilon_n + \frac{g''(p)}{2!}\,\epsilon_n^2 + \dots \\ &= p + g'(p)\epsilon_n + \frac{g''(p)}{2!}\,\epsilon_n^2 + \dots . \end{aligned}$$

Hence,

$$\epsilon_{n+1} = x_{n+1} - p = g'(p)\epsilon_n + \frac{g''(p)}{2}\,\epsilon_n^2 + \dots .$$

DÉFINITION 9.3. The **order of the iterative method** 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 converge p.

The iterative schemes $g_3(x)$ and $g_4(x)$ used in Example 9.6 converge to first order.

Note that for a second-order iterative scheme, we have

$$\frac{\epsilon_{n+1}}{\epsilon_n^2} \approx \frac{g''(p)}{2} = \text{constant}.$$

9.2.4. Méthode de Newton. Let x_n be an approximation to the 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. 9.1. 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**,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{9.4}$$

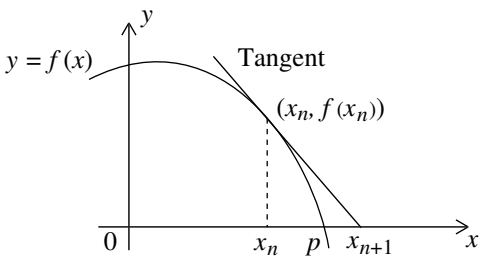

FIGURE 9.1. The nth step of Newton's method.

Note that Newton's method is a fixed point method since it can be rewritten in the form

$$x_{n+1} = g(x_n), \qquad \text{where} \quad g(x) = x - \frac{f(x)}{f'(x)}.$$

EXEMPLE 9.7. 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

$$f(x) = x^2 - 2 = 0.$$

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 following results:

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

Therefore,

$$\sqrt{2} \approx 1.414214.$$

Note that the number of zeros in the errors doubles as it is the case with methods of second order. □

EXEMPLE 9.8. Use six iterations of Newton's method to approximate a root $p \in [1, 2]$ of the function

$$f(x) = x^3 + 4\,x^2 - 10 = 0$$

given in Example 9.6.

SOLUTION. In this case, Newton's method becomes

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{-10 + 4x_n^2 + x_n^3}{8x_n + 3x_n^2} = \frac{2(5 + 2x_n^2 + x_n^3)}{8x_n + 3x_n^2}.$$

We take $x_0 = 1.5$. The results are listed in Table 8.3. □

TABLE 3. Results of Example 9.8.

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}

THÉORÈME 9.6. *Let p be a simple root $f(x) = 0$, 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 $f(p) = 0$ and $f'(p) \neq 0$, we have

$$g'(p) = 0.$$

Therefore, Newton's method is of order two. □

REMARQUE 9.3. Taking the second derivative of $g(x)$ in Newton's method, we have

$$\begin{aligned} g''(x) &= \frac{[f'(x)f''(x) + f(x)f'''(x)][f'(x)]^2 - 2f(x)f''(x)f'(x)f''(x)}{[f'(x)]^4} \\ &= \frac{[f'(x)f''(x) + f(x)f'''(x)]f'(x) - 2f(x)[f''(x)]^2}{[f'(x)]^3}. \end{aligned}$$

If $f'''(p)$ exists, we obtain

$$g''(p) = \frac{f''(p)}{f'(p)}.$$

Thus, the successive errors satisfy the approximate relation

$$\epsilon_{n+1} \approx \frac{f''(p)}{f'(p)}\, \epsilon_n^2,$$

which explains the doubling of the number of leading zeros in the error of Newton's method near a simple root of $f(x) = 0$.

EXEMPLE 9.9. Use six iterations of the ordinary and modified Newton's methods

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad x_{n+1} = x_n - 2\,\frac{f(x_n)}{f'(x_n)}$$

TABLE 4. Results of Example 9.9.

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	0
6	0.972	0.512	1	0

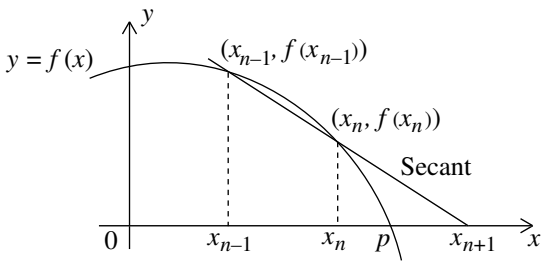

FIGURE 9.2. The nth step of the secant method.

to approximate the double root of the function

$$f(x) = (x-1)^2(x-2).$$

SOLUTION. In this case, the two methods are

$$g_1(x) = x - \frac{(x-1)(x-2)}{2(x-2)+(x-1)}, \qquad g_2(x) = x - \frac{2(x-1)(x-2)}{2(x-2)+(x-1)}.$$

We take $x_0 = 0$. The results are listed in Table 8.4. 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 p the method

$$x_{n+1} = x_n - p\,\frac{f(x_n)}{f'(x_n)}$$

has 2nd-order convergence. □

In general, Newton's method may converge to the desired root, to another root, or to an attractive cycle, especially in the complex plane.

9.2.5. La méthode de la sécante. 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 $(x_n, f(x_n))$ and $(x_{n-1}, f(x_{n-1}))$. 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. 9.2,

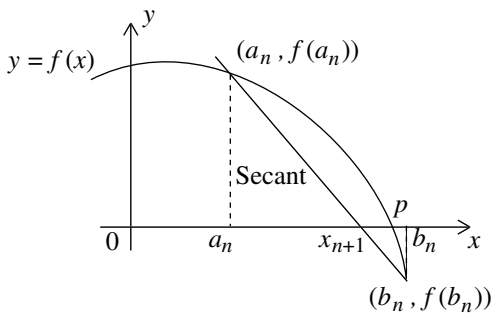

FIGURE 9.3. The nth step of the method of false position.

$$0 = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} (x_{n+1} - x_n) + f(x_n).$$

Solving for x_{n+1}, we obtain the **secant method**:

$$x_{n+1} = x_n - \frac{(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})} f(x_n).$$

The secant method algorithm is as follows:

(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 $f(x_n) - f(x_{n-1}) \neq 0$. 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 9.2.2).

This method is generally slower than Newton's method. However, it does not require the derivative of $f(x)$. In general applications of the Newton's method, the derivative of the function $f(x)$ is approximated numerically by the slope of a secant to the curve.

9.2.6. La méthode de la position fausse. 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. 9.3, 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

$$0 = f(a_n) + \frac{f(b_n) - f(a_n)}{b_n - a_n} (x_{n+1} - a_n),$$

TABLE 5. Results of Example 9.10.

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	−	−

which leads to the method of false position:

$$x_{n+1} = \frac{a_n\, f(b_n) - b_n\, f(a_n)}{f(b_n) - f(a_n)}.$$

Given that $f(x)$ is continuous on $[a, b]$ and that $f(a)\, f(b) < 0$, the algorithm for the method of false position is:

(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) Stop if $f(x_{n+1}) = 0$.
(4) If $f(x_{n+1}) \neq 0$ and
   (a) $f(x_{n+1})$ and $f(a_n)$ are of opposite signs, set $a_{n+1} = a_n$ and $b_{n+1} = x_{n+1}$;
   (b) $f(x_{n+1})$ and $f(a_n)$ are of the same sign, set $a_{n+1} = x_{n+1}$ and $b_{n+1} = b_n$.
(5) Repeat (2), (3), and (4) until the selected stopping criterion is satisfied (see Subsection 9.2.2).

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 to the root is one-sided, convergence can be accelerated by replacing the value of $f(x)$ at the stagnant end position with $f(x)/2$.

EXEMPLE 9.10. Find an approximation to $\sqrt{2}$ using the method of false position. Stop iterating when $|x_{n+1} - x_n| < 10^{-3}$.

SOLUTION. This problem is equivalent to 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 8.5. Therefore, $\sqrt{2} \approx 1.414141$. □

9.2.7. La méthode de la bisection. The bisection method constructs a sequence of intervals of decreasing length which contain a root p of $f(x) = 0$. If

$$f(a)\, f(b) < 0 \quad \text{and} \quad f \quad \text{is continuous,}$$

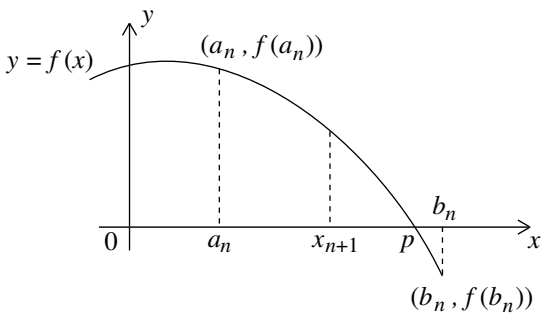

FIGURE 9.4. The nth step of the bisection method.

then $f(x)$ has a root between a and b. The root is either between

$$a \quad \text{and} \quad \frac{a+b}{2}, \quad \text{if} \quad f(a)\, f\left(\frac{a+b}{2}\right) < 0,$$

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. 9.4.

The algorithm of the **bisection method** is as follows. Given that $f(x)$ is continuous on $[a, b]$ and that $f(a)\, f(b) < 0$, we

(1) choose $a_0 = a$ and $b_0 = b$ such that $f(a_0) f(b_0) < 0$.
(2) For $n = 0, 1, 2, \ldots$, compute

$$x_{n+1} = \frac{a_n + b_n}{2}.$$

(3) Stop if $f(x_{n+1}) = 0$.
(4) If $f(x_{n+1}) \neq 0$ and
   (a) $f(x_{n+1})$ and $f(a_n)$ are of opposite signs, set $a_{n+1} = a_n$ and $b_{n+1} = x_{n+1}$.
   (b) $f(x_{n+1})$ and $f(a_n)$ are of the same sign, set $a_{n+1} = x_{n+1}$ and $b_{n+1} = b_n$.
(5) Repeat (2), (3) and (4) until the selected stopping criterion is satisfied (see Subsection 9.2.2).

The rate of convergence of the bisection method is low but the method always converges.

This 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.
%
% 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;
```

EXEMPLE 9.11. 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 8.6. The answer is $\sqrt{2} \approx 1.414063$ with an accuracy of 10^{-2}. Note that a root lies in the interval $[1.414063, 1.421875]$. □

EXEMPLE 9.12. 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.

TABLE 6. Results of Example 9.11.

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. Results of Example 9.12.

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	−	−

SOLUTION. Since

$$f(1) = -5 < 0 \quad \text{and} \quad f(2) = 14 > 0,$$

then $f(x)$ has a root 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 8.7.

After eight iterations, we find that the root is approximately 1.36328125. Since the exact root is between 1.363281250 and 1.367187500, the error is bounded by

$$1.367187500 - 1.363281250 = 0.00390625. \quad \square$$

EXEMPLE 9.13. Find the number of iterations needed in Example 9.12 to have an absolute error less than 10^{-4}.

SOLUTION. Since the root 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. □

9.2.8. Une méthode globale de Newton-bisection. The many difficulties that can occur with the Newton 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 $[a, b]$ 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 equal 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 with 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;
```

9.2.9. La fonction `fzero` de Matlab. 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`.

9.3. Interpolation et extrapolation

Quite often, experimental results provide only a few values of an unknown function $f(x)$, say,

$$(x_0, f_0), \qquad (x_1, f_1), \qquad (x_2, f_2), \qquad \ldots, \qquad (x_n, f_n), \tag{9.5}$$

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 a polynomial $p_n(x)$ of degree n whose graph passes through the points (9.5). This polynomial will be used to estimate $f(x)$ (we hope that $f(x) \approx p_n(x)$). If the points x_i are distinct, this polynomial is unique, Clearly, the following polynomial, called the **Lagrange interpolation polynomial** of $f(x)$,

$$p_n(x) = \sum_{i=0}^{n} f_i \frac{\prod_{j\neq i}(x - x_j)}{\prod_{j\neq i}(x_i - x_j)} \tag{9.6}$$

is of degree n and interpolates $f(x)$ at the points (9.5).

PROPOSITION 9.1. *Let $p_n(x)$ be as in (9.6), where $f_i = f(x_i)$. Suppose that $f(x)$ has a continuous derivative of order $(n+1)$ in the interval $[a, b]$ which contains the points $x_0, \ldots, x_n$. Then*

$$f(x) - p_n(x) = \frac{f^{(n+1)}(c(x))}{(n+1)!}(x - x_0)(x - x_1)\cdots(x - x_n)$$

where $c(x) \in [a, b]$. In particular, if

$$m_{n+1} = \min_{a\leq x\leq b} |f^{(n+1)}(x)| \quad \textit{and} \quad M_{n+1} = \max_{a\leq x\leq b} |f^{(n+1)}(x)|,$$

then

$$\begin{aligned}\frac{m_{n+1}}{(n+1)!}|(x - x_0)(x - x_1)\cdots(x - x_n)| &\leq |f(x) - p_n(x)| \\ &\leq \frac{M_{n+1}}{(n+1)!}|(x - x_0)(x - x_1)\cdots(x - x_n)|\end{aligned}$$

for $a \leq x \leq b$.

PROOF. The proof is omitted. □

From a computational point of view, (9.6) 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 its degree to improve the interpolation.

9.3.1. Interpolation linéare. Let $x_0 \neq x_1$ and consider the two data points: (x_0, f_0) and (x_1, f_1). Then

$$p_1(x) = f_0 + (x - x_0) f[x_0, x_1]$$

where

$$f[x_0, x_1] = \frac{f_1 - f_0}{x_1 - x_0}$$

is the **first divided difference**.

EXEMPLE 9.14. Consider a function $f(x)$ which passes through the points $(2.2, 6.2)$ and $(2.5, 6.7)$. Find a linear interpolation to $f(x)$ and use it to approximate $f(2.35)$.

SOLUTION. Since we have

$$f[2.2, 2.5] = \frac{6.7 - 6.2}{2.5 - 2.2} = 1.6667,$$

then

$$p_1(x) = 6.2 + (x - 2.2) \times 1.6667 = 2.5333 + 1.6667\, x.$$

In particular, $p_1(2.35) = 6.45$. □

EXEMPLE 9.15. 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

$$\cos\frac{\pi}{8} = \frac{1}{2}\sqrt{\sqrt{2}+2}$$

keeping in mind that $\cos(\pi/4) = \sqrt{2}/2$.) 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. □

9.3.2. Interpolation du second ordre. 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

$$p_2(x) = f_0 + (x - x_0)\, f[x_0, x_1] + (x - x_0)\,(x - x_1)\, f[x_0, x_1, x_2]$$

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.

EXEMPLE 9.16. Interpolate a given function $f(x)$ through the three points

$$(2.2, 6.2), \qquad (2.5, 6.7), \qquad (2.7, 6.5),$$

and approximate $f(2.35)$ by means of this interpolation.

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, $p_2(2.35) = 6.57$. □

EXEMPLE 9.17. Construct the quadratic interpolation polynomial 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 9.15 that

$$\cos \frac{\pi}{8} = \frac{1}{2}\sqrt{\sqrt{2} + 2}$$

Hence, from the three data points

$$(0, 1), \qquad (\pi/8, \cos \pi/8), \qquad (\pi/4, \sqrt{2}/2),$$

we obtain the divided differences

$$f[0, \pi/8] = \frac{4}{\pi}\left(\sqrt{\sqrt{2} + 2} - 2\right), \qquad f[\pi/8, \pi/4] = \frac{4}{\pi}\left(\sqrt{2} - \sqrt{\sqrt{2} + 2}\right),$$

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,

$$p_2(x) = 1 + x\,\frac{4}{\pi}\left(\sqrt{\sqrt{2} + 2} - 2\right) + x(x - \pi/8)\frac{16}{\pi^2}\left(\sqrt{2} - 2\sqrt{\sqrt{2} + 2}\right).$$

Evaluating this polynomial at $x = 0.2$, we obtain

$$p_2(0.2) = 0.97881.$$

The absolute error is 0.00189. □

EXEMPLE 9.18. 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

$$p_2(0.3) = 1.0953. \quad \square$$

9.3.3. Interpolation par différences divisées. Given the data points

$$(x_0, f_0), \qquad (x_1, f_1), \qquad \ldots, \qquad (x_n, f_n).$$

where $x_i \neq x_j$ for $i \neq j$, we have

$$p_n(x) = f_0 + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2] + \cdots + (x - x_0)(x - x_1) \cdots (x - x_{n-1}) f[x_0, x_1, \ldots, x_n] \tag{9.7}$$

where

$$f[x_j, x_{j+1}, \ldots, x_k] = \frac{f[x_{j+1}, \ldots, x_k] - f[x_j, x_{j+1}, \ldots, x_{k-1}]}{x_k - x_j}.$$

To derive this formula, we note that

$$p_2(x) = p_1(x) + \text{“something”}$$

and in general

$$p_n(x) = p_{n-1}(x) + \text{“something”},$$

where

$$p_{n-1}(x_i) = f_i, \qquad \text{if} \qquad i < n.$$

The term “something” is 0 if $x = x_i$, $i < n$, and $p_n(x_n) = f_n$. (This derivation is omitted.)

EXEMPLE 9.19. Construct the cubic interpolation through the points $(1.0, 2.4)$, $(1.3, 2.2)$, $(1.5, 2.3)$ and $(1.7, 2.4)$ on the graph of a certain function $f(x)$ and approximate $f(1.4)$.

SOLUTION. Newton's divided difference table is

x_i	$f(x_i)$	$f[x_i, x_j]$	$f[x_i, x_j, x_k]$	$f[x_i, x_j, x_k, x_l]$
1.0	$\boxed{2.4}$			
		$\boxed{-0.66667}$		
1.3	2.2		$\boxed{2.33340}$	
		0.500000		$\boxed{-3.3334}$
1.5	2.3		0.00000	
		0.500000		
1.7	2.4			

Therefore,

$$p_3(x) = 2.4 + (x - 1.0)\,(-0.66667) + (x - 1.0)\,(x - 1.3) \cdot 2.33340 \\ + (x - 1.0)\,(x - 1.3)\,(x - 1.5)\,(-3.3334).$$

The approximation to $f(1.4)$ is

$$p_3(1.4) = 2.2400. \quad \square$$

9.3.4. La méthode d'interpolation par différence-avant de Gregory–Newton. We rewrite (9.7) in the special case where the points x_i are equidistant,

$$x_i = x_0 + i\,h.$$

The **first forward difference** of $f(x)$ at x_j is

$$\Delta f_j := f_{j+1} - f_j.$$

The **second forward difference** of $f(x)$ at x_j is

$$\Delta^2 f_j := \Delta f_{j+1} - \Delta f_j.$$

The k**th 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 points, (9.7) becomes

$$p_n(r) = f_0 + \sum_{k=1}^{n} \frac{r\,(r-1)\,\cdots\,(r-k+1)}{k!}\,\Delta^k f_0 \tag{9.8}$$

$$= \sum_{k=0}^{n} \binom{r}{k}\,\Delta^k f_0,$$

where

$$\binom{r}{k} = \begin{cases} \frac{r\,(r-1)\,\cdots\,(r-k+1)}{k!} & \text{if} \quad k > 0, \\ 1 & \text{if} \quad k = 0. \end{cases}$$

EXEMPLE 9.20. Suppose that we are given the following data:

x	1988	1989	1990	1991	1992	1993
y	35000	36000	36500	37000	37800	39000

Extrapolate the value of y in 1994.

SOLUTION. The 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

$$p_5(x) = 35000 + r\,1000 + \frac{r(r-1)}{2}(-500) + \frac{r(r-1)(r-2)}{6}(500) + \frac{r(r-1)(r-2)(r-3)}{24}(-200) + \frac{r(r-1)(r-2)(r-3)(r-4)}{120}\cdot 0.$$

In particular,

$$p_5(1994) = 40500. \quad \square$$

EXEMPLE 9.21. Use the following 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 difference table is

i	x_i	y_i	$\Delta^2 y_i$	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
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

$$p_4(x) = 0.7651977 + r\,(-0.145112) + \frac{r(r-1)}{2}(-0.0195721) + \frac{r(r-1)(r-2)}{6}(0.0106723) + \frac{r(r-1)(r-2)(r-3)}{24}(0.0003548)$$

and

$$p_4(1.5) = 0.511819. \quad \square$$

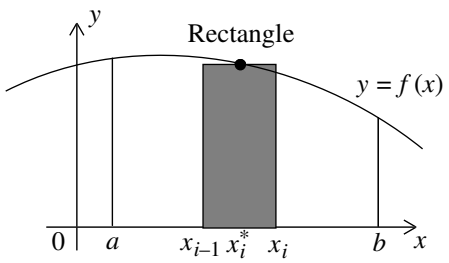

FIGURE 9.5. The ith panel of the rectangular rule.

9.4. Intégration numérique

To approximate the value of the definite integral

$$\int_a^b f(x)\,dx$$

where the function $f(x)$ is continuous on $[a, b]$ and $a < b$, we subdivide the interval $[a, b]$ into n sub-intervals of equal length $h = (b - a)/n$. The function $f(x)$ is approximated on each of these sub-intervals by an interpolation polynomial and the polynomials are integrated.

We shall consider three methods:

(i) For the **rectangular rule**, $f(x)$ is interpolated on each sub-interval by a constant (not necessarily the same constant from one sub-interval to another), and the integral of $f(x)$ over a sub-interval is estimated by the area of a rectangle.

(ii) For the **trapezoidal rule**, $f(x)$ is interpolated on each sub-interval by a polynomial of degree one (not necessarily the same polynomial from one sub-interval to another), and the integral of $f(x)$ over a sub-interval is estimated by the area of a trapezoid.

(iii) For **Simpson's rule**, $f(x)$ is interpolated on each sub-interval by a polynomial of degree two (not necessarily the same polynomial from one sub-interval to another), and the integral of $f(x)$ over a sub-interval is estimated by the area of under a parabola.

9.4.1. La méthode du rectangle. We subdivide the interval $[a, b]$ into n sub-intervals 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 sub-interval $[x_{i-1}, x_i]$, 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. 9.5 Thus, on the sub-interval $[x_{i-1}, x_i]$, the integral of $f(x)$ is approximated by $f(x_i^*)h$. This gives the **rectangular rule**:

$$\int_a^b f(x)\,dx \approx h\,[f(x_1^*) + \cdots + f(x_n^*)]\,. \tag{9.9}$$

PROPOSITION 9.2 (Error bound for the rectangular rule). *Let $f(x)$ have a continuous first derivative on the interval $[a, b]$ and let*

$$M = \max_{a \le x \le b} |f'(x)|.$$

The absolute value of the global error, ϵ_R, in the approximation to the integral $\int_a^b f(x)\,dx$ by the rectangular rule is of order $O(h)$:

$$|\epsilon_R| \le M\,\frac{(b-a)^2}{4n} = \frac{1}{4}M\,(b-a)h = O(h).$$

PROOF. On $[x_{i-1}, x_i]$ we have

$$\int_{x_{i-1}}^{x_i} f(x)\,dx - f(x_i^*)\,h = \int_{x_{i-1}}^{x_i} [f(x) - f(x_i^*)]\,dx = \int_{x_{i-1}}^{x_i} f'(c(x))\,(x - x_i^*)\,dx$$

by the mean value theorem. Since

$$\big|f'\big(c(x)\big)\big| \le M.$$

it follows that

$$\begin{aligned}
\left|\int_{x_{i-1}}^{x_i} f(x)\,dx - f(x_i^*)\,h\right| &= \left|\int_{x_{i-1}}^{x_i} f'(c(x))\,(x - x_i^*)\,dx\right| \\
&\le \int_{x_{i-1}}^{x_i} |f'(c(x))\,(x - x_i^*)|\,dx \\
&\le \int_{x_{i-1}}^{x_i} M\,|x - x_i^*|\,dx \\
&= M\int_{x_{i-1}}^{x_i^*} (x_i^* - x)\,dx + M\int_{x_i^*}^{x_i} (x - x_i^*)\,dx \\
&= M\frac{(x_i^* - x_{i-1})^2}{2} + M\frac{(x_i - x_i^*)^2}{2} = M\frac{h^2}{4}
\end{aligned}$$

because

$$x_i^* - x_{i-1} = x_i - x_i^* = \frac{h}{2}.$$

Thus

$$\begin{aligned}
|\epsilon_R| &= \left|\sum_{i=1}^{n}\left[\int_{x_{i-1}}^{x_i} f(x)\,dx - f(x_i^*)\,h\right]\right| \le \sum_{i=1}^{n}\left|\int_{x_{i-1}}^{x_i} f(x)\,dx - f(x_i^*)\,h\right| \\
&\le \sum_{i=1}^{n} M\frac{h^2}{4} = nM\frac{h^2}{4} = \frac{1}{4}M(b-a)h \\
&= O(h),
\end{aligned}$$

since $nh = b - a$. □

EXEMPLE 9.22. Use the rectangular rule to approximate the integral

$$I = \int_0^1 e^{x^2}\,dx$$

with step size h such that the error is bounded by 10^{-4}.

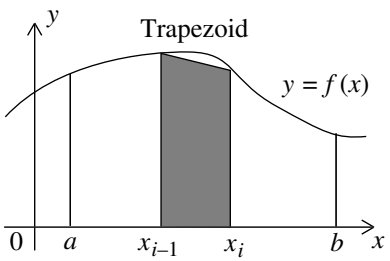

FIGURE 9.6. The ith panel of the trapezoidal rule.

SOLUTION. We have

$$f(x) = e^{x^2} \quad \text{and} \quad f'(x) = 2\,x\,e^{x^2}.$$

Thus

$$0 \le f'(x) \le 2\,e \quad \text{for } x \text{ in } [0,1].$$

The absolute value of the error is smaller than or equal to $displaystyle\frac{2\,e}{4}\,(1-0)^2 h$. It thus suffices to choose h such that

$$\frac{1}{2}\,eh < 10^{-4}, \quad \text{that is} \quad h < 7.357\,589 \times 10^{-5}.$$

To have $n = 1/h \ge 13591.4091$, we take $n = 13\,592$ and

$$h = \frac{1}{13\,592}.$$

Thus we have the approximation

$$I \approx \frac{1}{13592}\left[e^{(0.5/13592)^2} + e^{(1.5/13592)^2} + \cdots + e^{(13590.5/13592)^2} + e^{(13591.5/13592)^2}\right]$$
$$\approx 1.46265 \quad \square$$

9.4.2. La méthode du trapèze. We divide the interval $[a,b]$ into n sub-intervals of equal length $h = (b-a)/n$. The sub-intervals have end-points

$$x_0 = a, \quad x_1 = a + h, \quad \ldots, \quad x_i = a + ih, \quad \ldots, \quad x_n = b.$$

On the sub-interval $[x_{i-1}, x_i]$, 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. 9.6 Thus,

$$\int_{x_{i-1}}^{x_1} f(x)\,dx \approx \frac{h}{2}\,[f(x_{i-1}) + f(x_i)].$$

Summing over all the intervals, we obtain the **trapezoidal rule**:

$$\int_a^b f(x)\,dx \approx \frac{h}{2}\sum_{i=1}^{n}\big[f(x_{i-1}) + f(x_i)\big] \tag{9.10}$$
$$= \frac{h}{2}\big[f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-2}) + 2f(x_{n-1}) + f(x_n)\big].$$

PROPOSITION 9.3 (Error bound for the trapezoidal rule). *Let $f(x)$ have a continuous second derivative on $[a,b]$. Let ϵ_T be the error made when approximating the integral $\int_a^b f(x)\,dx$ by means of the trapezoidal rule. If*

$$M = \max_{a\le x\le b} |f''(x)|,$$

then

$$|\epsilon_T| \le M\,\frac{(b-a)^3}{12\,n^2} = \frac{1}{12}\,M(b-a)h^2 = O(h^2).$$

PROOF. On the sub-interval $[x_{i-1}, x_i]$, we interpolate the function $f(x)$ by a polynomial of first order through the points $(x_{i-1}, f(x_{i-1}))$ and $(x_i, f(x_i))$,

$$p_1(x) = f(x_{i-1}) + \frac{1}{h}\big[f(x_i) - f(x_{i-1})\big](x - x_{i-1}),$$

and notice that

$$\int_{x_{i-1}}^{x_i} p_1(x)\,dx = \frac{1}{h}\big[f(x_i) + f(x_{i-1})\big].$$

Thus, we have

$$\begin{aligned}
&\int_{x_{i-1}}^{x_i} f(x)\,dx - \frac{h}{2}\,[f(x_{i-1}) + f(x_i)] \\
&\qquad = \int_{x_{i-1}}^{x_i} [f(x) - p_1(x)]\,dx \qquad \text{(and by Proposition 9.1)} \\
&\qquad = \int_{x_{i-1}}^{x_i} \frac{f''(c(x))}{2}\,(x - x_{i-1})(x - x_i)\,dx \qquad \text{(and by Theorem 9.4)} \\
&\qquad = \frac{f''(c_i)}{2}\int_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_i)\,dx \\
&\qquad = -\frac{f''(c_i)}{12}\,h^3
\end{aligned}$$

We note that $(x - x_{i-1})(x - x_i) \le 0$ for all $x \in [x_{i-1}, x_i]$. Hence, the absolute value of the local error, $\epsilon_{T,i}$, on the sub-interval $[x_{i-1}, x_i]$ is of order $O(h^3)$,

$$\begin{aligned}
|\epsilon_{T,i}| &= \left|\int_{x_{i-1}}^{x_i} f(x)\,dx - \frac{h}{2}\,[f(x_{i-1}) + f(x_i)]\right| \\
&\le \left|-\frac{f''(c_i)}{12}\,h^3\right| \\
&\le M\frac{h^3}{12}.
\end{aligned}$$

Finally, the global error, ϵ_T, of the trapezoidal rule on the interval $[a,b]$ is of order $O(h^2)$

$$|\epsilon_T| = \left|\sum_{i=1}^{n} \epsilon_{T,i}\right| \le \sum_{i=1}^{n} M\frac{h^3}{12} = nM\frac{h^3}{12} = \frac{1}{12}\,M(b-a)h^2 = O(h^2),$$

since $nh = b - a$. □

EXEMPLE 9.23. Use the trapezoidal rule to approximate the integral

$$I = \int_0^1 e^{x^2}\,dx$$

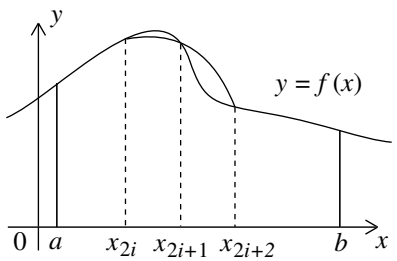

FIGURE 9.7. A pair of panels for Simpson's rule.

with step size h such that the error is bounded by 10^{-4}. Compare with Example 9.22.

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 $n = 117 > 1/h = 116.6$ (compared to 13 592 for the rectangular rule). The approximate value of I is

$$\begin{aligned} I &\approx \frac{1}{117 \cdot 2}\Big[e^{(0/117)^2} + 2\,e^{(1/117)^2} + 2\,e^{(2/117)^2} + \cdots \\ &\qquad\qquad + 2\,e^{(115/117)^2} + 2\,e^{(116/117)^2} + e^{(117/117)^2}\Big] \\ &= 1.46268. \quad \square \end{aligned}$$

9.4.3. La méthode de Simpson. We subdivide the interval $[a, b]$ into an even number, $2n$, of sub-intervals of equal length, $h = (b-a)/(2n)$, with endpoints

$$x_0 = a, \quad x_1 = a + h, \quad \ldots, \quad x_i = a + i\,h, \quad \ldots, \quad x_{2n} = b.$$

On the sub-interval $[x_{2i}, x_{2i+2}]$, the function $f(x)$ can be 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. 9.7 This polynomial, in its Lagrange form (9.6), is

$$\begin{aligned} p_2(x) = f(x_{2i})\,\frac{(x - x_{2i+1})(x - x_{2i+2})}{(x_{2i} - x_{2i+1})(x_{2i} - x_{2i+2})} &+ f(x_{2i+1})\,\frac{(x - x_{2i})(x - x_{2i+2})}{(x_{2i+1} - x_{2i})(x_{2i+1} - x_{2i+2})} \\ &+ f(x_{2i+2})\,\frac{(x - x_{2i})(x - x_{2i+1})}{(x_{2i+2} - x_{2i})(x_{2i+2} - x_{2i+1})}. \end{aligned}$$

Since $x_{2i+1} = x_{2i} + h$ and $x_{2i+2} = x_{2i} + 2h$, we have

$$\begin{aligned}\int_{x_{2i}}^{x_{2i+2}} f(x)\,dx &\approx \int_{x_{2i}}^{x_{2i+2}} p_2(x)\,dx \\ &= \frac{h}{3}\big[f(x_{2i}) + 4f(x_{2i}+h) + f(x_{2i}+2h)\big].\end{aligned}$$

Summing over all intervals, we obtain

$$\begin{aligned}\int_a^b f(x)\,dx &= \sum_{i=0}^{n-1} \int_{x_{2i}}^{x_{2i+2}} f(x)\,dx \\ &\approx \sum_{i=0}^{n-1} \frac{h}{3}\big[f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})\big],\end{aligned}$$

which is **Simpson's rule**:

$$\begin{aligned}\int_a^b f(x)\,dx \approx \frac{h}{3}\big[&f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots \\ &+ 2f(x_{2n-2}) + 4f(x_{2n-1}) + f(x_{2n})\big],\end{aligned} \tag{9.11}$$

with $h = (b-a)/(2n)$.

PROPOSITION 9.4 (Error bound for Simpson's rule). *Suppose that the function $f(x)$ has a continuous fourth-order derivative on the interval $[a,b]$. Let ϵ_S be the error made when approximating the integral $\int_a^b f(x)\,dx$ by means of Simpson's rule. If*

$$M = \max_{a\le x\le b} |f^{(4)}(x)|,$$

then

$$|\epsilon_S| \le M\,\frac{(b-a)^5}{180(2n)^4} = \frac{1}{180}\,M(b-a)h^4 = O(h^4).$$

PROOF. The proof is omitted. □

EXEMPLE 9.24. Use Simpson's rule to approximate the integral

$$I = \int_0^1 e^{x^2}\,dx$$

with stepsize h such that the error is bounded by 10^{-4}. Compare with Examples 9.22 and 9.23.

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 value of the 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 have $2n > 1/h = 10.4$ we take $n = 6$ and $h = 1/12$. The approximation is

$$I \approx \frac{1}{12 \cdot 3}\left[e^{(0/12)^2} + 4\,e^{(1/12)^2} + 2\,e^{(2/12)^2} + \cdots + 2\,e^{(10/12)^2} + 4\,e^{(11/12)^2} + e^{(12/12)^2}\right]$$
$$= 1.46267.$$

We obtain a value which is similar to those found in Examples 9.22 and 9.23. 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 rectangular and trapezoidal rules. □

EXEMPLE 9.25. Use Simpson's rule to approximate the integral

$$I = \int_0^2 \sqrt{1 + \cos^2 x}\, dx$$

with an accuracy of 0.0001.

SOLUTION. We must determine the step size h such that the error will be bounded by 0.0001. For

$$f(x) = \sqrt{1 + \cos^2 x},$$

we have

$$f^{(4)}(x) = \frac{-3\cos^4(x)}{(1+\cos^2(x))^{3/2}} + \frac{4\cos^2(x)}{\sqrt{1+\cos^2(x)}} - \frac{18\cos^4(x)\sin^2(x)}{(1+\cos^2(x))^{5/2}}$$
$$+ \frac{22\cos^2(x)\sin^2(x)}{(1+\cos^2(x))^{3/2}} - \frac{4\sin^2(x)}{\sqrt{1+\cos^2(x)}} - \frac{15\cos^4(x)\,\sin^4(x)}{(1+\cos^2(x))^{7/2}}$$
$$+ \frac{18\cos^2(x)\,\sin^4(x)}{(1+\cos^2(x))^{5/2}} - \frac{3\sin^4(x)}{(1+\cos^2(x))^{3/2}}.$$

Since every denominator is greater than one, we have

$$|f^{(4)}(x)| \le 3 + 4 + 18 + 22 + 4 + 15 + 18 + 3 = 87.$$

The absolute value of the error is thus less than or equal to $\frac{87}{180}(2-0)h^4$. Thus h must satisfy the inequality

$$\frac{87}{180}\,2h^4 < 10^{-4}, \quad \text{that is} \quad h < 0.100\,851\,140.$$

To have $2n > 2/h = 2 \times 9.9$ we take $n = 10$ and $h = 1/20$. The approximation is

$$I \approx \frac{1}{20 \cdot 3}\left[\sqrt{1+\cos^2(0/20)} + 4\sqrt{1+\cos^2(1/20)} + 2\sqrt{1+\cos^2(2/20)} + \ldots\right.$$
$$\left. + 2\sqrt{1+\cos^2(18/20)} + 4\sqrt{1+\cos^2(19/20)} + \sqrt{1+\cos^2(20/20)}\right]$$
$$= 2.48332. \quad \square$$

9.5. Problèmes aux valeurs initiales

In this section, we shall consider first-order initial value problems of the form

$$\frac{d\,y(t)}{d\,t} = f(t, y(t)), \qquad t_0 \le t \le t_f \tag{9.12}$$
$$y(t_0) = y_0$$

where $f(t, y)$ is a real-valued function of two variables.

We want to find an **approximation** to the solution $y(t)$ of (9.12) for $t_0 \le t \le t_f$. That is, we choose N points $t_0 < t_1 < t_2 < \ldots < t_N = t_f$ and we construct approximations y_i to $y(t_i)$, $i = 0, 1, \ldots, N$.

It is important to know whether or not a *small perturbation* of (9.12) will 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 (9.12). Truncation errors which occur when computing $f(t, y)$ can be identified with perturbations of (9.12) and of the initial condition. The following theorem gives sufficient conditions for an initial value problem to be **well-posed**; that is, there exists a unique solution and any small perturbation of the problem (9.12) leads to a correspondingly small change in the solution.

THÉORÈME 9.7. *Let*

$$D = \{(t, y) : t_0 \le t \le t_f \text{ and } -\infty < y < \infty\}.$$

If $f(t, y)$ is continuous on D and satisfies the ***Lipschitz condition***

$$|f(t, y_1) - f(t, y_2)| \le L|y_1 - y_2|$$

for all (t, y_1) and (t, y_2) in D, where L is a constant, then the initial value problem (9.12) is well-posed.

9.5.1. La méthode d'Euler. From this point on, we shall assume that (9.12) is well-posed. Moreover, we shall suppose that $f(t, y)$ has mixed partial derivatives of arbitrary order.

We choose the N points $t_j = t_0 + jh$ where $h = (t_f - t_0)/N$. From Taylor's Theorem we get

$$y(t_{j+1}) = y(t_j) + y'(t_j)\,(t_{j+1} - t_j) + \frac{y''(\xi_j)}{2}\,(t_{j+1} - t_j)^2$$

for ξ_j between x_j and x_{j+1}; $j = 0, 1, \ldots, N-1$. Since $y'(t_j) = f(t_j, y(t_j))$ and $t_{j+1} - t_j = h$, it follows that

$$y(t_{j+1}) = y(t_j) + f\big(t_j, y(t_j)\big)h + \frac{y''(\xi_j)}{2}\,h^2.$$

We obtain Euler's method by deleting the term of order $O(h^2)$. The term

$$\frac{y''(\xi_j)}{2}\,h^2$$

is called the **local discretization error**.

The algorithm for **Euler's method** is as follows.

(1) Choose N and set $h = (t_f - t_0)/N$.
(2) Given y_0, for $j = 0, 1, \ldots, N$, iterate the scheme

$$y_{j+1} = y_j + hf(t_0 + jh, y_j). \tag{9.13}$$

(3) Use y_j as an approximation to $y(t_j)$.

EXEMPLE 9.26. Use Euler's method with $N = 5$ to approximate the solution to the initial value problem

$$y'(t) = 0.2ty, \qquad \text{on } 1 \le t \le 1.5, \quad \text{with } y(1) = 1. \tag{9.14}$$

TABLE 8. Numerical results of Example 9.26.

t_j	y_j	$y(t_j)$	Absolute error	Relative error
1.00	1.0000	1.0000	0.0000	0.00
1.10	1.0200	1.0212	0.0012	0.12
1.20	1.0424	1.0450	0.0025	0.24
1.30	1.0675	1.0714	0.0040	0.37
1.40	1.0952	1.1008	0.0055	0.50
1.50	1.1259	1.1331	0.0073	0.64

SOLUTION. We have

$$t_0 = 1, \quad t_f = t_5 = 1.5, \quad y_0 = 1, \quad f(t, y) = 0.2ty.$$

Hence

$$h = (t_5 - t_0)/5 = 0.1, \quad t_j = t_0 + hj = 1 + 0.1j, \quad y_j \approx y(t_j),$$

where y_j is given by the iteration

$$y_{j+1} = y_j + 0.2(1 + 0.1j)y_j, \quad \text{with } y_0 = 1,$$

for $j = 0, 1, \ldots, 4$. The numerical results are listed in Table 8. Note that the differential equation in (9.14) is separable. The particular solution of (9.14) is $y(t) = e^{(0.1t^2 - 0.1)}$. This formula has been used to compute the exact values $y(x_j)$ in the previous table. □

The next example illustrates the limitations of Euler's method. In the next subsections, we shall see more accurate methods than Euler's method.

EXEMPLE 9.27. Use Euler's method with $N = 5$ to approximate the solution to the initial value problem

$$y'(t) = 2ty, \qquad \text{on } 1 \le t \le 1.5, \quad \text{with } y(1) = 1. \tag{9.15}$$

SOLUTION. As in the previous example, we have $t_0 = 1$, $t_f = t_5 = 1.5$ and $y_0 = 1$. However, $f(t, y) = 2ty$. Hence

$$h = (t_5 - t_0)/5 = 0.1, \qquad t_j = t_0 + hj = 1 + 0.1j,$$

and the approximation y_j to $y(t_j)$ is given by

$$y_0 = 1,$$
$$y_{j+1} = y_j + 2.0(1 + 0.1j)y_j,$$

for $j = 0, 1, 2, 3, 4$. The numerical results are listed in Table 9. Upon investigation of the relative errors, we must conclude that our approximations are not very good. □

It is generally **incorrect** to say that by taking h sufficiently small one can obtain any desired level of precision, that is, get y_j as close as one wants to $y(t_j)$.

Let y_j be the computed value for $y(t_j)$ in (9.13). Set

$$e_j = y(t_j) - y_j, \quad \text{for} \quad j = 0, 1, \ldots, N.$$

If

$$|e_0| < \delta_0$$

TABLE 9. Numerical results of Example 9.27.

t_j	y_j	$y(t_j)$	Absolute error	Relative error
1.00	1.0000	1.0000	0.0000	0.00
1.10	1.0200	1.2337	0.0337	2.73
1.20	1.4640	1.5527	0.0887	5.71
1.30	1.8154	1.9937	0.1784	8.95
1.40	2.2874	2.6117	0.3244	12.42
1.50	2.9278	3.4904	0.5625	16.12

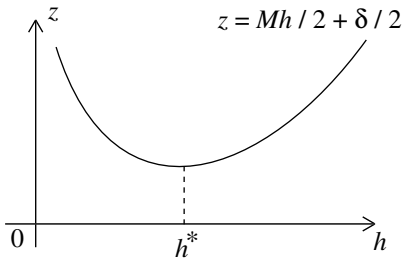

FIGURE 9.8. Round-off error curve as a function of stepsize h.

and the precision in the computations is bounded by δ, then it can be shown that

$$|e_j| \le \frac{1}{L}\left(\frac{Mh}{2} + \frac{\delta}{h}\right)\left(e^{L(t_j - t_0)} - 1\right) + \delta_0\, e^{L(t_j - t_0)},$$

where L is the Lipschitz constant found in Theorem 9.7,

$$M = \max_{t_0 \le t \le t_f} |y''(t)|,$$

and $h = (t_f - t_0)/N$.

We remark that the expression

$$z(h) = \frac{Mh}{2} + \frac{\delta}{h}$$

tends to infinity as h goes to zero, as shown in Fig. 9.8.

DÉFINITION 9.4. The **local truncation error** of a method of the form

$$y_{j+1} = y_j + h\,\phi(t_j, y_j), \tag{9.16}$$

is given by the expression

$$\tau_{j+1} = \frac{1}{h}\big[y(t_{j+1}) - y(t_j)\big] - \phi(t_j, y(t_j)) \qquad \text{for} \quad j = 0, 1, 2, \ldots, N-1.$$

The method (9.16) is of **order** k if $|\tau_j| \le M\,h^k$ for some constant M and for all j.

EXEMPLE 9.28. The local truncation error of Euler's method is

$$\tau_{j+1} = \frac{1}{h}\big[y(t_{j+1}) - y(t_j)\big] - f\big(t_j, y(t_j)\big) = \frac{h}{2}\,y''(\xi_j)$$

for some ξ_j between t_j and t_{j+1}. If

$$M = \max_{t_0 \le t \le t_f} |y''(t)|,$$

then $|\tau_j| \le \frac{h}{2} M$ for all j. Hence, Euler's method is of order one.

9.5.2. Méthode de Runge–Kutta d'ordre 2. We now develop methods of order greater than one,which, in general, are more precise than Euler's method.

By Taylor's Theorem, we have

$$y(t_{j+1}) = y(t_j) + y'(t_j)(t_{j+1} - t_j) + \frac{1}{2} y''(t_j)(t_{j+1} - t_j)^2 + \frac{1}{6} y'''(\xi_j)(t_{j+1} - t_j)^3 \tag{9.17}$$

for some ξ_j between x_j and x_{j+1} and $j = 0, 1, \ldots, N-1$. From the differential equation

$$y'(t) = f\big(t, y(t)\big),$$

and its first total derivative with respect to t, we obtain expressions for $y'(t_j)$ and $y''(t_j)$,

$$\begin{aligned} y'(t_j) &= f\big(t_j, y(t_j)\big), \\ y''(t_j) &= \frac{d}{dt} f\big(t, y(t)\big)\Big|_{t=t_j} \\ &= f_t\big(t_j, y(t_j)\big) + f_y\big(t_j, y(t_j)\big)\, f\big(t_j, y(t_j)\big). \end{aligned}$$

Therefore, putting $h = t_{j+1} - t_j$ and substituting these expressions in (9.17), we have

$$y(t_{j+1}) = y(t_j) + f\big(t_j, y(t_j)\big)\, h + \frac{1}{2}\big[f_t\big(t_j, y(t_j)\big) + f_y\big(t_j, y(t_j)\big) f\big(t_j, y(t_j)\big)\big] h^2 + \frac{1}{6} y'''(\xi_j) h^3$$

for $j = 0, 1, \ldots, N-1$.

Our goal is to replace the expression

$$f\big(t_j, y(t_j)\big) + \frac{1}{2}\big[\big(f_t\big(t_j, y(t_j)\big) + f_y\big(t_j, y(t_j)\big)\, f\big(t_j, y(t_j)\big)\big] + O(h^2) \tag{9.18}$$

by an expression of the form

$$a f\big(t_j, y(t_j)\big) + b f\big(t_j + \alpha h, y(t_j) + \beta h f(t_j, y(t_j))\, h\big) + O(h^2). \tag{9.19}$$

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

$$f\big(t_j + \alpha h, y(t_j) + \beta h f(t_j, y(t_j))\big) = f\big(t_j, y(t_j)\big) + \alpha h f_t\big(t_j, y(t_j)\big) + \beta h f\big(t_j, y(t_j)\big) f_y\big(t_j, y(t_j)\big) + O(h^2).$$

In order for the expression (9.17) and (9.18) to be equal to order h, we must have $a + b = 1$, $\alpha b = 1/2$ and $\beta b = 1/2$. Standard second-order Runge–Kutta methods are:

(1) For the **mid-point method**, $\alpha = \beta = 1/2$, $a = 0$ and $b = 1$.
(2) For the **modified Euler method**, $\alpha = \beta = 1$ and $a = b = 1/2$.
(3) For **Heun's method**, $\alpha = \beta = 2/3$, $b = 3/4$ and $a = 1/4$.

TABLE 10. Numerical results of Example 9.29.

t_j	y_j^{C}	$y(t_j)$	Absolute error	Relative error
1.00	1.0000	1.0000	0.0000	0.00
1.10	1.2320	1.2337	0.0017	0.14
1.20	1.5479	1.5527	0.0048	0.31
1.30	1.9832	1.9937	0.0106	0.53
1.40	2.5908	2.6117	0.0209	0.80
1.50	3.4509	3.4904	0.0344	1.13

The algorithm for the two-stage **second-order Runge–Kutta methods** is as follows.

(1) Choose N and set $h = (t_f - t_0)/N$.
(2) For given y_0 and $j = 0, 1, \ldots, N$, iterate the formula

$$y_{j+1} = y_j + h\big[af(t_j, y_j) + bf(t_j + \alpha h, y_j + \beta f(t_j, y_j)h)\big],$$

where $t_j = t_0 + jh$.
(3) Use y_j as an approximation to $y(t_j)$.

EXEMPLE 9.29. Use the modified Euler method with $N = 5$ to approximate the solution to the initial value problem

$$y'(t) = 2ty, \qquad \text{on } 1 \le t \le 1.5, \quad \text{with } y(1) = 1,$$

of Example 9.27.

SOLUTION. We have

$$h = (t_5 - t_0)/5 = 0.1, \qquad t_j = t_0 + hj = 1 + 0.1j.$$

The approximation y_j to $y(t_j)$ is given by the predictor-corrector scheme

$$\begin{aligned} y_0^{\mathrm{C}} &= 1 \\ y_{j+1}^{\mathrm{P}} &= y_j^{\mathrm{C}} + 0.2\, t_j\, y_j \\ y_{j+1}^{\mathrm{C}} &= y_j^{\mathrm{C}} + 0.1\left(t_j\, y_j^{\mathrm{C}} + t_{j+1}\, y_{j+1}^{\mathrm{P}}\right) \end{aligned}$$

for $j = 0, 1, \ldots, 4$. The numerical results are listed in Table 10. These results are much better than those listed in Table 9 for Euler's method. □

9.5.3. Méthode de Runge–Kutta d'ordre 4. The fourth-order Runge–Kutta method (also known as the classic Runge–Kutta method) is the very popular among the explicit one-step methods to approximate the solution to an initial value problem.

By Taylor's Theorem, we have

$$\begin{aligned} y(t_{j+1}) = y(t_j) + y'(t_j)(t_{j+1} - t_j) + \frac{y''(t_j)}{2!}(t_{j+1} - t_j)^2 + \frac{y^{(3)}(t_j)}{3!}(t_{j+1} - t_j)^3 \\ + \frac{y^{(4)}(t_j)}{4!}(t_{j+1} - t_j)^4 + \frac{y^{(5)}(\xi_j)}{5!}(t_{j+1} - t_j)^5 \end{aligned}$$

for some ξ_j between x_j and x_{j+1} and $j = 0, 1, \ldots, N-1$. To obtain the fourth-order Runge–Kutta method, we 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

$$y'(t_j)(t_{j+1}-t_j)+\frac{y''(t_j)}{2!}\,(t_{j+1}-t_j)^2+\frac{y^{(3)}(t_j)}{3!}\,(t_{j+1}-t_j)^3 + \frac{y^{(4)}(t_j)}{4!}\,(t_{j+1}-t_j)^4+O(h^5)$$

is equal to

$$h(ak_1+bk_2+ck_3+dk_4)+O(h^5),$$

where

$$\begin{aligned} k_1 &= f(t_j, y_j),\\ k_2 &= f(t_j+\alpha_1\,h, y_j+\beta_1\,h\,k_1),\\ k_3 &= f(t_j+\alpha_2\,h, y_j+\beta_2\,h\,k_2),\\ k_4 &= f(t_j+\alpha_3\,h, y_j+\beta_3\,h\,k_3). \end{aligned}$$

This follows from the relations

$$\begin{aligned} t_{j+1}-t_j &= h,\\ y'(t_j) &= f(t_j, y(t_j)),\\ y''(t_j) &= \frac{d}{d\,t}\,f(t, y(t))|_{t=t_j}\\ &= f_t(t_j, y(t_j)) + f_y(t_j, y(t_j))\,f(t_j, y(t_j)),\ldots, \end{aligned}$$

and Taylor's Theorem for functions of two variables. The lengthy computations are omitted.

We only give the "classic" four-stage **fourth-order Runge–Kutta method**, whose algorithm can be summarized as follows.

(1) Choose N and set $h=(t_f-t_0)/N$.
(2) Given y_0, for $j=0,1,\ldots,N$ iterate the scheme

$$y_{j+1}=y_j+\frac{h}{6}(k_1+2k_2+2k_3+k_4),$$

where

$$\begin{aligned} k_1 &= f(t_j, y_j),\\ k_2 &= f(t_j+h/2, y_j+h\,k_1/2),\\ k_3 &= f(t_j+h/2, y_j+h\,k_2/2),\\ k_4 &= f(t_{j+1}, y_j+h\,k_3), \end{aligned}$$

and $t_j=t_0+jh$.
(3) Use y_j as an approximation to $y(t_j)$.

The next example shows that the fourth-order Runge–Kutta method yields better results for (9.15) than the previous methods.

EXEMPLE 9.30. Use the fourth-order Runge–Kutta method with $N=5$ to approximate the solution to the initial value problem of Example 9.27.

SOLUTION. We have

$$h=(t_5-t_0)/5=0.1, \qquad t_j=1.0+0.1j, \qquad \text{for} \quad j=0,1,\ldots,5.$$

TABLE 11. Numerical results for Example 9.30.

t_j	y_j	$y(t_j)$	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

With the starting value $y_0 = 1.0$, the approximation y_j to $y(t_j)$ is given by the scheme

$$y_{j+1} = y_j + \frac{0.1}{6}(k_1 + 2\,k_2 + 2\,k_3 + k_4)$$

where

$$\begin{aligned} k_1 &= 2(1.0 + 0.1j)y_j, \\ k_2 &= 2(1.05 + 0.1j)(y_j + k_1/20), \\ k_3 &= 2(1.05 + 0.1j)(y_j + k_2/20), \\ k_4 &= 2(1.0 + 0.1(j+1)(y_j + k_3/10), \end{aligned}$$

and $j = 0, 1, 2, 3, 4$. The numerical results are listed in Table 11. These results are much better than all those previously obtained. □

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);
```

9.5.4. Méthodes à pas variables. 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 $y(t)$ does not vary rapidly and smaller when $y(t)$ changes rapidly.

One frequently used method which controls the step size is the **Runge–Kutta–Fehlberg (4,5) method**. This method is made of a nested pair of fourth- and fifth-order Runge–Kutta methods.

Let y_0 be the initial condition. Suppose that the approximation y_j to $y(t_j)$ has been computed and satisfies $|y(t_j) - y_j| < \epsilon$ where ϵ is the desired precision. Let $h > 0$.

(1) Compute two approximations for y_{j+1}: one using the fourth-order method

$$y_{j+1} = y_j + \left(\frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5\right), \tag{9.20}$$

and the second using the fifth-order method,

$$\widehat{y}_{j+1} = y_j + \left(\frac{16}{135}k_1 + \frac{6656}{12825}k_3 + \frac{28561}{56430}k_4 - \frac{9}{50}k_5 + \frac{2}{55}k_6\right), \tag{9.21}$$

where

$$k_1 = f(t_j, y_j),$$
$$k_2 = f(t_j + h/4, y_j + h\,k_1/4),$$
$$k_3 = f(t_j + 3h/8, y_j + 3h\,k_1/32 + 9h\,k_2/32),$$
$$k_4 = f(t_j + 12h/13, y_j + 1932h\,k_1/2197 - 7200h\,k_2/2197 + 7296h\,k_3/2197),$$
$$k_5 = f(t_j + h, y_j + 439h\,k_1/216 - 8h\,k_2 + 3680h\,k_3/513 + 845h\,k_4/4104),$$
$$k_6 = f(t_j + h/2, y_j - 8h\,k_1/27 + 2h\,k_2 + 3544h\,k_3/2565 + 1859h\,k_4/4104 - 11h\,k_5/40).$$

(2) If $|\widehat{y}_{j+1} - y_{j+1}| < \epsilon h$, accept y_{j+1} as the approximation to $y(t_{j+1})$. Replace h by qh where

$$q = \left[\epsilon h/(2|\widehat{y}_{j+1} - y_{j+1}|)\right]^{0.25}$$

and go back to step (1) to compute an approximation for y_{j+2}.

(3) If $|\widehat{y}_{j+1} - y_{j+1}| \geq \epsilon h$, replace h by qh where

$$q = \left[\epsilon h/(2|\widehat{y}_{j+1} - y_{j+1}|)\right]^{0.25}$$

and go back to step (1) to compute the next approximation for y_{j+1}.

One can show that the local truncation error for (9.20) is approximately

$$|\widehat{y}_{j+1} - y_{j+1}|/h.$$

At step (2), one requires that this error be smaller than ϵh in order to get $|y(t_j) - y_j| < \epsilon$ for all j (and in particular $|y(t_f) - y_f| < \epsilon$). 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 (9.20) and (9.21).

CHAPITRE 10

Examens partiels et final

Test 1 **MAT 2731** **le 6 octobre 1995**
Durée: 80 min Prof.: Rémi Vaillancourt
Place: MST 333

Instructions:
(a) Examen partiel à livre fermé. Seules les calculatrices approuvées par la Faculté sont autorisées.
(b) Répondre sur le questionnaire.
(c) Les 6 questions ont toutes la même valeur.
(d) Justifiez vos réponses. Une réponse non justifiée ne sera pas corrigée.

L'équation différentielle homogène du 1er ordre, $M(x,y)\,dx + N(x,y)\,dy = 0$, admet un facteur d'intégration $\mu(x)$ ou $\mu(y)$ suivant que:

$$\frac{1}{N}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = f(x) \Longrightarrow \mu(x) = e^{\int f(x)\,dx},$$

$$\frac{1}{M}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = g(y) \Longrightarrow \mu(y) = e^{-\int g(y)\,dy}.$$

$$\begin{aligned}\int (1+2x^2)\,e^{x^2}\,dx &= \int e^{x^2}\,dx + \int x(2x\,e^{x^2})\,dx = \int e^{x^2}\,dx + x\,e^{x^2} - \int e^{x^2}\,dx \\ &= x\,e^{x^2}\end{aligned}$$

Qu. 1. *Résoudre*: $y' = x^3\,e^{-y}, \qquad y(0) = 2.$

Qu. 2. *Trouver un facteur d'intégration, rendre l'équation différentielle exacte et résoudre le problème à valeur initiale.*

$$(1 + 2x^2 + 4xy)\,dx + 2\,dy = 0, \qquad y(0) = 1.$$

Qu. 3. *Résoudre le problème aux valeurs initiales.*

$$y'' + 6y' + 9y = 0, \qquad y(0) = -2, \quad y'(0) = 14.$$

Qu. 4. *Résoudre le problème aux valeurs initiales.*

$$x^2y'' - 4x\,y' + 4y = 0, \qquad y(1) = 4, \quad y'(1) = 13.$$

Qu. 5. *Trouver la solution générale.*

$$y'' - y' - 2y = x + e^x.$$

Qu. 6. *Trouver la solution générale.*

$$y'' - 2y' + y = 12x^{-3}\, e^x.$$

Test 2 **MAT 2731** **le 3 novembre 1995**
Durée: 80 min Prof.: Rémi Vaillancourt
Place: MST 333

Instructions:
(a) Examen partiel à livre fermé. Seules les calculatrices approuvées par la Faculté sont autorisées.
(b) Répondre sur le questionnaire.
(c) Les 6 questions ont toutes la même valeur.
(d) Justifiez vos réponses. Une réponse non justifiée ne sera pas corrigée.

Qu. 1. *Résoudre le problème aux valeurs initiales*:

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

Qu. 2. *Trouver la solution générale de l'équation différentielle*:

$$x^3 y''' + x^2 y'' - 2xy' + 2y = x^{-2}.$$

Qu. 3. *Trouver le rayon de convergence des séries*:

$$(a) \quad \sum_{m=0}^{\infty} \frac{x^m}{3^m}, \qquad\qquad (b) \quad \sum_{m=0}^{\infty} m^{2m} x^m.$$

Qu. 4. Les 6 premiers polynômes de LEGENDRE:

$$P_0(x) = 1, \qquad P_1(x) = x,$$
$$P_2(x) = \frac{1}{2}\left(3x^2 - 1\right), \qquad P_3(x) = \frac{1}{2}\left(5x^3 - 3x\right),$$
$$P_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right), \qquad P_5(x) = \frac{1}{8}\left(63x^5 - 70x^3 + 15x\right).$$

Développer le polynôme $p(x)$ suivant selon les polynômes de LEGENDRE:

$$p(x) = x^4 + x^2 + 1, \qquad -1 \le x \le 1.$$

Qu. 5. Soit la méthode de Runge–Kutta d'ordre 4 explicite à quatre étapes:

$$\begin{aligned} k_1 &= hf(x_n, y_n), \\ k_2 &= hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right), \\ k_3 &= hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_2\right), \\ k_4 &= hf\left(x_n + h, y_n + k_3\right), \end{aligned}$$

et

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4).$$

Calculer y_1 au millionième près (6 décimales) par cette méthode avec $h = 0.2$ pour l'équation différentielle:

$$y' = x + y, \qquad \text{avec} \qquad x_0 = 0, \quad y_0 = 0.021\,400.$$

Qu. 6. Soit la quadrature de Gauss à trois points:

$$\int_{-1}^{1} f(x)\,dx = \frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right).$$

Évaluer, au cent-millième près (5 décimales) l'intégrale:

$$I = \int_{-1}^{1} e^{-x^2}\,dx,$$

par la formule de Gauss à 3 points.

Test 3 **MAT 2731 A** **le 24 novembre 1995**
MST 333; 80 min Prof.: Rémi Vaillancourt

(a) Examen partiel à livre fermé. Calculatrices approuvées seulement.
(b) Répondre sur le questionnaire. Table à la fin du questionnaire.
(c) Les 6 questions ont toutes la même valeur.
(d) Justifiez vos réponses. Une réponse non justifiée ne sera pas corrigée.

Qu. 1. *Résoudre, au dix-millième près ou en arithmétique rationnelle, le système* $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 3.8 \\ 4x_1 + 3x_2 - x_3 &= -5.7 \\ 5x_1 + x_2 + 3x_3 &= 2.8 \end{aligned}$$

(a) *soit par élimination gaussienne AVEC PIVOTAGE sur les lignes,*

OU (b) *soit par la décomposition* $A = LU$ *avec* $l_{ii} = 1$ *(Doolittle).* (Remarque: on résout $L\boldsymbol{y} = \boldsymbol{b}$ par substitution avant et $U\boldsymbol{x} = \boldsymbol{y}$ par substitution arrière.)

Qu. 2. *Faire, au dix-millième près, deux itérations de la méthode de Gauss–Seidel,*

$$\boldsymbol{x}^{(m+1)} = D^{-1}\left(\boldsymbol{b} - L\boldsymbol{x}^{(m+1)} - U\boldsymbol{x}^{(m)}\right), \quad \textit{où} \quad A = D + L + U, \quad A\boldsymbol{x} = \boldsymbol{b},$$

(D diagonale, L triangulaire strictement inférieure et U triangulaire strictement supérieure) sur le système suivant avec $\boldsymbol{x}^{(0)}$ *donné, ayant soin de réarranger les lignes pour assurer la convergence*:

$$\begin{aligned} 2x_1 + 10x_2 - x_3 &= -32 \qquad x_1^{(0)} = 1 \\ -x_1 + 2x_2 + 15x_3 &= 17 \qquad x_2^{(0)} = 1 \\ 10x_1 - x_2 + 2x_3 &= 58 \qquad x_3^{(0)} = 1 \end{aligned}$$

Qu. 3. *Résoudre le système suivant par la méthode de Cholelsky:*

$$GG^T\boldsymbol{x} = \boldsymbol{b}, \quad \text{où} \quad A = GG^T, \quad A\boldsymbol{x} = \boldsymbol{b}, \quad G \text{ triangulaire inférieure.}$$

(Remarque: on résout $G\boldsymbol{y} = \boldsymbol{b}$ par substitution avant et $G^T\boldsymbol{x} = \boldsymbol{y}$ par substitution arrière.)

$$\begin{aligned} 9x_1 + 6x_2 + 12x_3 &= 174 \\ 6x_1 + 13x_2 + 11x_3 &= 236 \\ 12x_1 + 11x_2 + 26x_3 &= 308 \end{aligned}$$

Qu. 4. (a) *Soit* $F(s) = \dfrac{1}{s(s+1)}$. *Trouver* $f(t) = \mathcal{L}^{-1}(F)$.

(b) *Soit* $f(t) = 2t^3\, e^{-t/2}$. *Trouver* $F(s) = \mathcal{L}(f)$.

(c) *Soit* $f(t) = t \sin \omega t$. *Trouver* $F(s) = \mathcal{L}(f)$.

Qu. 5. (a) *Tracer la fonction* $f(t) = \begin{cases} 1, & si\ 0 \le x < 1, \\ 2t^2, & si\ x > 1. \end{cases}$

(b) *Exprimer* $f(t)$ *au moyen de la fonction d'Heaviside* $u(t-a) = \begin{cases} 0, & si\ x < a, \\ 1, & si\ x > a. \end{cases}$

(c) *Trouver* $F(s) = \mathcal{L}(f)$.

Qu. 6. *Résoudre au moyen de la transformation de Laplace:*

$$y'' + 2y' + y = 3t\,e^{-t}, \qquad y(0) = 4, \qquad y'(0) = 2.$$

Examen final **MAT 2731 A** **le 8 décembre 1995**
GYM E; 3h Prof.: Rémi Vaillancourt

(a) Examen à livre fermé. Calculatrices approuvées seulement.
(b) Répondre sur le questionnaire. **Table à la fin du questionnaire**.
(c) Les 10 questions ont toutes la même valeur.
(d) Justifiez vos réponses. Une réponse non justifiée ne sera pas corrigée.

Qu. 1. *Soit*: $(3x\,e^y + 2y)\,dx + (x^2\,e^y + x)\,dy = 0$.

(a) *Montrer que l'équation est exacte ou non.*
(b) *Si non exacte, trouver un facteur d'intégration.*
(c) *Trouver la solution générale de l'équation.*

Qu. 2. *Résoudre le problème à valeur initiale*:

$$y' + 3x^2 y = x\,e^{-x^3}, \qquad y(0) = -1.$$

Qu. 3. *Trouver la solution générale de l'équation nonhomogène*:

$$y''' - 3y'' + 3y' - y = x^{1/2}\,e^x.$$

Qu. 4. (a) *Trouver le rayon de convergence de la série* $\displaystyle\sum_{m=0}^{\infty} \frac{(-1)^m}{k^m}\,x^{2m}$.

(b) *Représenter le polynôme:*

$$p(x) = 70x^4 + 30x^3 - 30x^2 - 15x + 5$$

au moyen d'une combinaison linéaire de polynômes de Legendre.

Qu. 5. Etant donné l'équation différentielle $y' = f(x, y), \quad y(x_0) = y_0$, le prédicteur d'Adams–Bashford et le correcteur d'Adams–Moulton d'ordre 4 sont respectivement

(10.1)

$$y_{n+1}^P = y_n^C + \frac{h}{24}\left(55f_n^C - 59f_{n-1}^C + 37f_{n-2}^C - 9f_{n-3}^C\right), \quad \text{où } f_k^C = f\left(x_k, y_k^C\right),$$

(10.2)

$$y_{n+1}^C = y_n^C + \frac{h}{24}\left(9f_{n+1}^P + 19f_n^C - 5f_{n-1}^C + f_{n-2}^C\right), \qquad \text{où } f_k^P = f\left(x_k, y_k^P\right).$$

Pour l'équation différentielle

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

compléter, au cent-millième près, les **quatre** cases vides de gauche de la table suivante au moyen de (10.1) et (10.2) et les **quatre** cases vides de droite au moyen de la solution exacte.

n	x_n	Départ y_n^C	Prédite y_n^P	Corrigée y_n^C	Exacte $y(t_n)$	Erreur: $10^6\times$ $(y(t_n)-y_n^C)$
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				
4	0.8		0.425 361		0.425 541	12
5	1.0			0.718 270	0.718 282	
6	1.2		1.119 855	1.120 106	1.120 117	11
7	1.4		1.654 885	1.655 191		
8	1.6		2.352 653	2.353 026	2.353 032	6
9	1.8			3.249 646	3.249 647	1
10	2.0		4.388 505		4.389 056	

Qu. 6. *Résoudre, au dix-millième près ou en arithmétique rationnelle, le système*

$$\begin{aligned} x_1 + 2x_2 - 3x_3 &= -4 \\ 2x_1 + 13x_2 - 9x_3 &= -11 \\ -3x_1 - 9x_2 + 35x_3 &= 38 \end{aligned}$$

c'est-à-dire $A\boldsymbol{x} = \boldsymbol{b}$, par l'UNE des trois méthodes suivantes:

(a) *soit par l'élimination gaussienne AVEC PIVOTAGE sur les lignes,*

OU

(b) *soit par la décomposition* $A = LU$ *avec* $l_{ii} = 1$ *(Doolittle),*

OU

(c) *soit par la méthode de Cholesky* $A = GG^T$.

(Remarque: Dans le cas (b) (resp. (c)) on résout $L\boldsymbol{y} = \boldsymbol{b}$ (resp. $G\boldsymbol{y} = \boldsymbol{b}$) par la substitution avant et $U\boldsymbol{x} = \boldsymbol{y}$ (resp. $G^T\boldsymbol{x} = \boldsymbol{y}$) par la substitution arrière.)

Qu. 7. (a) Soit A une matrice d'ordre 3. Le théorème de Gershgorin affirme que chacun des disques dans le plan:

$$|a_{11} - \lambda| \le |a_{12}| + |a_{13}|, \quad |a_{22} - \lambda| \le |a_{21}| + |a_{23}|, \quad |a_{33} - \lambda| \le |a_{31}| + |a_{32}|,$$

contient une valeur propre de A.

Tracer, à l'échelle, les disques de Gershgorin pour la matrice

$$A = \begin{bmatrix} -1+2i & i & 1 \\ -0.5 & 0 & 0.5 \\ 0.8+0.6i & 0 & 2-i \end{bmatrix}.$$

(b) La norme ℓ_1 d'une matrice B d'ordre 2 est

$$\|B\|_1 = \max\{|b_{11}| + |b_{21}|, |b_{12}| + |b_{22}|\} \qquad \text{(colonne maximum)},$$

et le conditionnement de B dans la norme ℓ_1 est

$$\kappa_1(B) = \|B\|_1 \|B^{-1}\|_1.$$

Soit les matrices

$$B = \begin{bmatrix} 600 & 800 \\ 30\,000 & 40\,002 \end{bmatrix}, \qquad B^{-1} \approx \begin{bmatrix} 100 & -2.0 \\ -75 & 1.5 \end{bmatrix}.$$

Calculer la norme ℓ_1 de B et de B^{-1} et le conditionnement de B.

Qu. 8. (a) *Trouver la transformée de Laplace de* $f(t) = 3t\cosh 3t$.

(b) *Soit la convolution* $f(t) = u(t-\pi) * \sin t$ *où* $u(t)$ *est la fonction d'Heaviside. Calculer* $f(t)$ *et trouver* $F(s)$.

(c) *Soit* $F(s) = \dfrac{8}{(s+4)^2}$. *Trouver* $f(t) = \mathcal{L}^{-1}(F)$.

Qu. 9. (a) *Tracer la fonction* 2π*-périodique*

$$f(t) = \pi - t, \quad 0 < t < 2\pi, \qquad f(t+2\pi) = f(t) \quad \textit{pour tout } t,$$

sur l'intervalle $-2\pi < t < 4\pi$ *et trouver la transformée de Laplace* $F(s)$ *de* $f(t)$.

Qu. 10. *Tracer la fonction* $r(t)$ *et résoudre le problèmes aux valeurs initiales au moyen de la transformation de Laplace:*

$$y'' + 9y = r(t) =: \begin{cases} 8\sin t, & si \quad 0 < t < \pi, \\ 0, & si \quad \pi < t, \end{cases} \qquad y(0) = 0, \quad y'(0) = 4.$$

CHAPITRE 11

Formulaire et tables

11.1. Facteur d'intégration de $M(x,y)\,dx + N(x,y)\,dy = 0$

Soit l'équation différentielle homogène du 1er ordre

$$M(x,y)\,dx + N(x,y)\,dy = 0. \tag{11.1}$$

Si

$$\frac{1}{N}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = f(x)$$

est une fonction de x seulement, alors

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

est un facteur d'intégration de (11.1).

Si

$$\frac{1}{M}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = g(y)$$

est une fonction de y seulement, alors

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

est un facteur d'intégration de (11.1).

11.2. Les polynômes de Legendre $P_n(x)$ sur $[-1,1]$

1. L'équation différentielle de Legendre:
$$(1-x^2)y'' - 2xy' + n(n+1)y = 0, \quad -1 \le x \le 1.$$
2. La solution $y(x) = P_n(x)$ explicite:
$$P_n(x) = \frac{1}{2^n}\sum_{m=0}^{[n/2]} (-1)^m \begin{pmatrix} n \\ m \end{pmatrix} \begin{pmatrix} 2n-2m \\ n \end{pmatrix} x^{n-2m},$$
où $[n/2]$ désigne le plus grand entier au plus égal à $n/2$.
3. La relation de récurrence:
$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x).$$
4. La standardisation:
$$P_n(1) = 1.$$
5. La norme des $P_n(x)$:
$$\int_{-1}^{1} [P_n(x)]^2\,dx = \frac{2}{2n+1}.$$

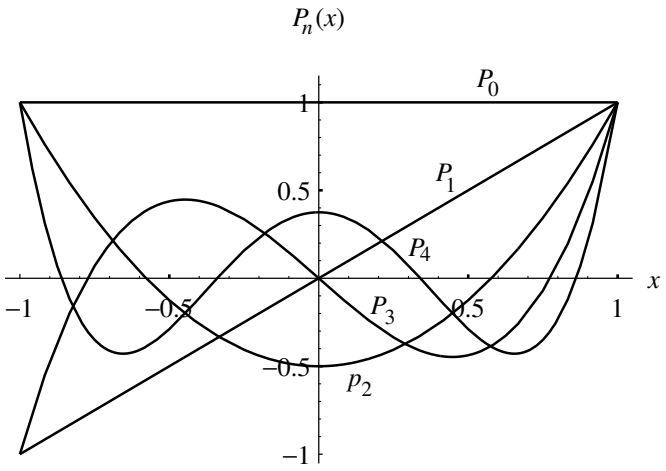

FIGURE 11.1. Les 5 premiers polynômes de LEGENDRE.

6. La formule de RODRIGUES:

$$P_n(x) = \frac{(-1)^n}{2^n n!} \frac{d^n}{dx^n} \left[\left(1 - x^2\right)^n \right].$$

7. La fonction génératrice:

$$\frac{1}{\sqrt{1 - 2xt + t^2}} = \sum_{n=0}^{\infty} P_n(x) t^n, \quad -1 < x < 1, \ |t| < 1.$$

8. L'inégalité sur les $P_n(x)$:

$$|P_n(x)| \leq 1, \quad -1 \leq x \leq 1.$$

9. Les 6 premiers polynômes de LEGENDRE (V. figure 11.1):

$$P_0(x) = 1, \qquad P_1(x) = x,$$
$$P_2(x) = \frac{1}{2}\left(3x^2 - 1\right), \qquad P_3(x) = \frac{1}{2}\left(5x^3 - 3x\right),$$
$$P_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right), \qquad P_5(x) = \frac{1}{8}\left(63x^5 - 70x^3 + 15x\right).$$

11.3. Les polynômes de Laguerre sur $0 \leq x < \infty$

On définit les polynômes de LAGUERRE sur $0 \leq x < \infty$ par l'expression:

$$L_n(x) = \frac{e^x}{n!} \frac{d^n(x^n e^{-x})}{dx^n}, \quad n = 0, 1, \ldots$$

Les 4 premiers sont (V. figure 11.2):

$$L_0(x) = 1, \qquad L_1(x) = 1 - x,$$
$$L_2(x) = 1 - 2x + \frac{1}{2}x^2, \qquad L_3(x) = 1 - 3x + \frac{3}{2}x^2 - \frac{1}{6}x^3.$$

On peut obtenir les $L_n(x)$ par la récurrence:

$$(n+1)L_{n+1}(x) = (2n + 1 - x)L_n(x) - nL_{n-1}(x).$$

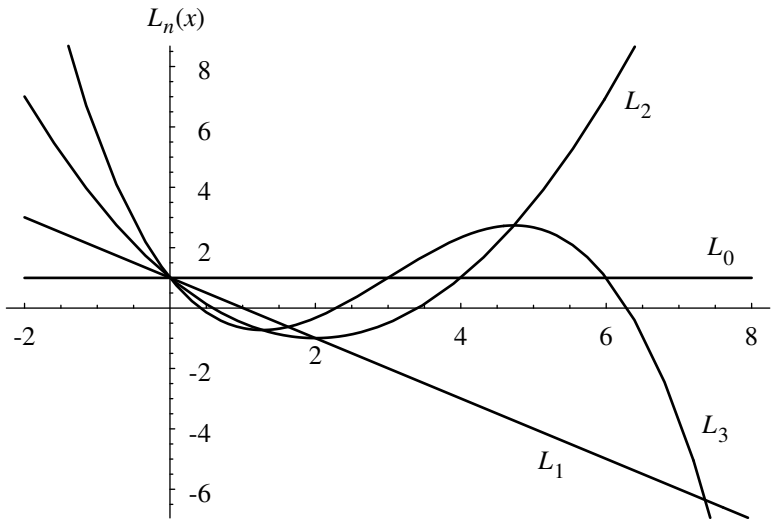

FIGURE 11.2. Les 4 premiers polynômes de LAGUERRE.

Les $L_n(x)$ sont solutions de l'équation différentielle

$$xy'' + (1-x)y' + ny = 0$$

et satisfont les relations d'orthogonalité avec le poids $p(x) = e^{-x}$:

$$\int_0^\infty e^{-x} L_m(x) L_n(x)\, dx = \begin{cases} 0, & m \neq n, \\ 1, & m = n. \end{cases}$$

11.4. Développements de Fourier–Legendre

Le développement de FOURIER–LEGENDRE de $f(\varphi)$ suivant les $P_n(\cos\varphi)$:

$$f(\varphi) \sim \sum_{n=0}^{\infty} a_n P_n(\cos\varphi), \quad 0 \leq \varphi \leq \pi,$$

où

$$\begin{aligned} a_n &= \frac{2n+1}{2} \int_0^\pi f(\varphi) P_n(\cos\varphi) \sin\varphi\, d\varphi \\ &= \frac{2n+1}{2} \int_{-1}^1 f(\arccos w) P_n(w)\, dw, \quad n = 0, 1, 2, \ldots \end{aligned}$$

Le développement de FOURIER–LEGENDRE découle des relations d'orthogonalité suivantes:

$$\int_{-1}^1 P_m(x) P_n(x)\, dx = \begin{cases} 0, & m \neq n, \\ \frac{2}{2n+1}, & m = n. \end{cases}$$

TABLE 1. Table d'intégrales.

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. $\displaystyle\int \frac{du}{\sqrt{a^2 - u^2}} = \ln\left(u + \sqrt{u^2 - a^2}\right) + c = \operatorname{arccosh} \frac{u}{a} + c$
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. $\displaystyle\int x^n \ln ax \, dx = \frac{x^{n+1}}{n + 1} \ln ax - \frac{x^{n+1}}{(n + 1)^2} + c$

11.5. Table d'intégrales

11.6. La transformée de Laplace

$$\mathcal{L}\{f(t)\} = \int_0^\infty e^{-st} f(t) \, dt = F(s)$$

TABLE 2. Table de transformées de Laplace.

	$F(s) = \mathcal{L}\{f(t)\}$	$f(t)$
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.	$\frac{1}{s}\, e^{-cs},\ c>0$	$h(t-c) := \begin{cases} 0, & 0 \le t < c \\ 1, & t \ge c \end{cases}$
4.	$e^{-cs}F(s),\ c>0$	$f(t-c)h(t-c)$
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.	$\frac{1}{1-e^{-ps}} \int_0^p e^{-st} f(t)\, dt$	$f(t+p) = f(t),$ pour tout t

Exercices

Exercices pour le chapitre premier

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. $(x^2 + y^2)\,dx - 2xy\,dy = 0, \quad y(1) = 1$.

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. $(2x\cos y + 3x^2 y)\,dx + (x^3 - x^2\sin y - y)\,dy = 0, \quad y(0) = 2.$

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

1.29. $(1 - xy)y' + y^2 + 3xy^3 = 0.$

1.30. $(2xy^2 - 3y^3)\,dx + (7 - 3xy^2)\,dy = 0.$

1.31. $(2x^2 y + 2y + 5)\,dx + (2x^3 + 2x)\,dy = 0.$

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. $y' + 3x^2 y = x^2, \quad y(0) = 2.$

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 members of the orthogonal trajectories on the same set of axes.

1.41. $x^2 + y^2/4 = c^2.$

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. $9yy' + x = 0$.

Exercices pour le chapitre 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. $y'' + y' + \frac{1}{4}y = 0, \quad y(2) = 1, \quad y'(2) = 1$.

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. $x^2y'' + \dfrac{7}{2}\,xy' - \dfrac{3}{2}\,y = 0, \quad y(4) = 1, \quad y'(4) = 0.$

Exercices pour le chapitre 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. $2x^2y'' + xy' - 3y = x^{-3}, \quad y(1) = 0, \quad y'(1) = 2.$

3.39. $2x^2y'' + xy' - 3y = 2x^{-3}, \quad y(1) = 0, \quad y'(1) = 3.$

Exercices pour le chapitre 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. $A = \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}, \quad \boldsymbol{y}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$

Exercices pour le chapitre 5

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.

5.1. $y' = e^{-y} - y + 1, \quad y(0) = 1.$

5.2. $y' = x + \sin y, \quad y(0) = 0.$

5.3. $y' = x + \cos y, \quad y(0) = 0.$

5.4. $y' = x^2 + y^2, \quad y(0) = 1.$

5.5. $y' = 1 + y^2, \quad y(0) = 0.$

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.

5.6. $y' = e^{-y} - y + 1, \quad y(0) = 1.$

5.7. $y' = x + \sin y, \quad y(0) = 0.$

5.8. $y' = x + \cos y, \quad y(0) = 0.$

5.9. $y' = x^2 + y^2, \quad y(0) = 1.$

5.10. $y' = 1 + y^2, \quad y(0) = 0.$

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.

5.11. $y' = x^2 + y^2, \quad y(0) = 1.$

5.12. $y' = x + \sin y, \quad y(0) = 0.$

5.13. $y' = x + \cos y, \quad y(0) = 0.$

5.14. $y' = e^{-y}, \quad y(0) = 0.$

5.15. $y' = y^2 + 2y - x, \quad y(0) = 0.$

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.

5.16. $y' = x + \sin y, \quad y(0) = 0.$

5.17. $y' = x + \cos y, \quad y(0) = 0.$

5.18. $y' = y^2 - y + 1, \quad y(0) = 0.$

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.

5.19. $y' = x + \sin y, \quad y(0) = 0.$

5.20. $y' = x + \cos y, \quad y(0) = 0.$

5.21. $y' = x + 2\cos y, \quad y(0) = 0.$

5.22. $y' = y^2 - y + 1, \quad y(0) = 0.$

Exercices pour le chapitre 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.

6.8. $y'' - 3y' + 2y = 0.$

6.9. $(1 - x^2)y'' - 2xy' + 2y = 0.$

6.10. $y'' + x^2 y' + xy = 0.$

6.11. $y'' - xy' - y = 0.$

6.12. $(x^2 - 1)y'' + 4xy' + 2y = 0.$

6.13. $(1 - x)y'' - y' + xy = 0.$

6.14. $y'' - 4xy' + (4x^2 - 2)y = 0.$

6.15. $y'' - 2(x - 1)y' + 2y = 0.$

6.16. 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 $x = \cos\theta$.

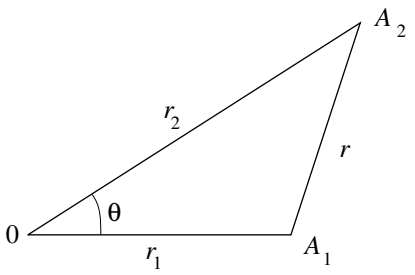

FIGURE 11.3. Distance r from point A_1 to point A_2.

6.17. Derive Rodrigues' formula (6.10).

6.18. Derive the generating function (6.11).

6.19. Let A_1 and A_2 be two points in space (see Fig. 11.3). By means of (6.9) derive the formula

$$\frac{1}{r} = \frac{1}{\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta}} = \frac{1}{r_2} \sum_{m=0}^{\infty} P_m(\cos\theta) \left(\frac{r_1}{r_2}\right)^m,$$

which is important in potential theory.

6.20. Derive Bonnet recurrence formula,

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x), \quad n = 1, 2, \ldots \tag{11.2}$$

(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.)

6.21. Compare the value of $P_4(0.7)$ obtained by means of the three-point recurrence formula (11.2) with the value obtained by evaluating $P_4(x)$ directly at $x = 0.7$.

6.22. For nonnegative integers m and n, with $m \leq n$, 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

$$P_0(x), \quad P_1(x), \quad \ldots$$

6.23. $p(x) = 5x^3 + 4x^2 + 3x + 2, \qquad -1 \leq x \leq 1.$

6.24. $p(x) = 10x^3 + 4x^2 + 6x + 1, \qquad -1 \leq x \leq 1.$

6.25. $p(x) = x^3 - 2x^2 + 4x + 1, \qquad -1 \leq x \leq 2.$

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.

6.26. $f(x) = e^x, \qquad -1 < x < 1.$

6.27. $f(x) = e^{2x}, \qquad -1 < x < 1.$

6.28. $f(x) = \begin{cases} 0 & -1 < x < 0, \\ 1 & 0 < x < 1. \end{cases}$

6.29. 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.

6.30. Evaluate

$$I = \int_{0.2}^{1.5} e^{-x^2}\, dx,$$

by the three-point Gaussian quadrature formula.

6.31. Evaluate

$$I = \int_{0.3}^{1.7} e^{-x^2}\, dx,$$

by the three-point Gaussian quadrature formula.

6.32. Derive the four-point Gaussian quadrature formula.

Exercices pour le chapitre 7

Solve the following system by the LU decomposition **without pivoting**.

7.1.

$$\begin{array}{rcrcrcr} 2x_1 & + & 2x_2 & + & 2x_3 & = & 4 \\ -x_1 & + & 2x_2 & - & 3x_3 & = & 32 \\ 3x_1 & & & - & 4x_3 & = & 17 \end{array}$$

7.2.

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 5 \\ x_1 & + & 2x_2 & + & 2x_3 & = & 6 \\ x_1 & + & 2x_2 & + & 3x_3 & = & 8 \end{array}$$

7.3.

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & -5 \\ x_1 & + & 2x_2 & + & 2x_3 & = & 6 \\ x_1 & + & 2x_2 & + & 3x_3 & = & -8 \end{array}$$

Solve the following system by the LU decomposition **with pivoting**.

7.4.

$$\begin{array}{rcrcrcr} 2x_1 & - & x_2 & + & 5x_3 & = & 4 \\ -6x_1 & + & 3x_2 & - & 9x_3 & = & -6 \\ 4x_1 & - & 3x_2 & & & = & -2 \end{array}$$

7.5.

$$\begin{array}{rcrcrcr} 2x_1 & - & x_2 & + & 5x_3 & = & -4 \\ -6x_1 & + & 3x_2 & - & 9x_3 & = & -6 \\ 4x_1 & - & 3x_2 & & & = & 2 \end{array}$$

7.6.

$$\begin{array}{rcrcrcr} 3x_1 & + & 9x_2 & + & 6x_3 & = & 23 \\ 18x_1 & + & 48x_2 & - & 39x_3 & = & 136 \\ 9x_1 & - & 27x_2 & + & 42x_3 & = & 45 \end{array}$$

7.7. Scale each equation in the l_∞-norm, so that the largest coefficient of each row on the left-hand side be equal to 1 in absolute value, and solve the following system by the LU decomposition with pivoting.

$$\begin{array}{rcrcrcr} x_1 & - & x_2 & + & 2x_3 & = & 3.8 \\ 4x_1 & + & 3x_2 & - & x_3 & = & -5.7 \\ 5x_1 & + & x_2 & + & 3x_3 & = & 2.8 \end{array}$$

7.8. Find the inverse of the Gaussian transformation

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -a & 1 & 0 & 0 \\ -b & 0 & 1 & 0 \\ -c & 0 & 0 & 1 \end{bmatrix}.$$

7.9. Find the product of the three Gaussian transformations

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ a & 1 & 0 & 0 \\ b & 0 & 1 & 0 \\ c & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & d & 1 & 0 \\ 0 & e & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & f & 1 \end{bmatrix}.$$

7.10. Find the Cholesky decomposition of

$$A = \begin{bmatrix} 1 & -1 & -2 \\ -1 & 5 & -4 \\ -2 & -4 & 22 \end{bmatrix}, \quad B = \begin{bmatrix} 9 & 9 & 9 & 0 \\ 9 & 13 & 13 & -2 \\ 9 & 13 & 14 & -3 \\ 0 & -2 & -3 & 18 \end{bmatrix}.$$

Solve by the Cholesky decomposition.

7.11.

$$\begin{bmatrix} 16 & -4 & 4 \\ -4 & 10 & -1 \\ 4 & -1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -12 \\ 3 \\ 1 \end{bmatrix}.$$

7.12.

$$\begin{bmatrix} 16 & -4 & 4 \\ -4 & 10 & -1 \\ 4 & -1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 3 \\ -1 \end{bmatrix}.$$

7.13.

$$\begin{bmatrix} 4 & 10 & 8 \\ 10 & 26 & 26 \\ 8 & 26 & 61 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 44 \\ 128 \\ 214 \end{bmatrix}.$$

Do three iterations of Gauss–Seidel's scheme on the properly permuted given systems with given initial values $x^{(0)}$.

7.14.

$$\begin{array}{rcrcrcr} 6x_1 & + & x_2 & - & x_3 & = & 3 \\ -x_1 & + & x_2 & + & 7x_3 & = & -17 \\ x_1 & + & 5x_2 & + & x_3 & = & 0 \end{array} \quad \text{with} \quad \begin{array}{rcl} x_1^{(0)} & = & 1 \\ x_2^{(0)} & = & 1 \\ x_3^{(0)} & = & 1 \end{array}$$

7.15.

$$\begin{array}{rcrcrcrcrcl} 6x_1 & + & x_2 & - & x_3 & = & 3 & & x_1^{(0)} & = & 1 \\ -x_1 & + & x_2 & + & 7x_3 & = & -17 & \text{with} & x_2^{(0)} & = & -1 \\ x_1 & + & 5x_2 & + & x_3 & = & 0 & & x_3^{(0)} & = & 1 \end{array}$$

7.16.

$$\begin{array}{rcrcrcrcrcl} 3x_1 & + & 3x_2 & + & 7x_3 & = & 4 & & x_1^{(0)} & = & 1 \\ 3x_1 & - & x_2 & + & x_3 & = & 1 & \text{with} & x_2^{(0)} & = & 11 \\ 3x_1 & + & 6x_2 & + & 2x_3 & = & 0 & & x_3^{(0)} & = & 1 \end{array}$$

7.17. Using least squares, fit a straight line to (s, F):

$$(0.9, 10), \quad (0.5, 5), \quad (1.6, 15), \quad (2.1, 20),$$

where s is the elongation of an elastic spring under a force F, and estimate from it the spring modulus $k = F/s$. ($F = ks$ is called Hooke's law).

7.18. Using least squares, fit a parabola to the data

$$(-1, 2), \quad (0, 0), \quad (1, 1), \quad (2, 2).$$

7.19. Using the least squares, fit $f(x) = a_0 + a_1 \cos x$ to the data

$$(0, 3.7), \quad (1, 3.0), \quad (2, 2.4), \quad (3, 1.8).$$

Note: x in radian measures.

7.20. Using least-squares, approximate the data

x_i	-1	-0.5	0	0.25	0.5	0.75	1
y_i	e^{-1}	$e^{-1/2}$	1	$e^{1/4}$	$e^{1/2}$	$e^{3/4}$	e

by means of

$$f(x) = a_0 P_0(x) + a_1 P_1(x) + a_2 P_2(x),$$

where P_0, P_1 and P_2 are the Legendre polynomials of degree 0, 1 and 2 respectively. Plot $f(x)$ and $g(x) = e^x$ on the same graph.

Using Theorem 7.2, determine and sketch Gershgorin's disks that contain the eigenvalues of the following matrices.

7.21.

$$\begin{bmatrix} -i & 0.1 + 0.1i & 0.5i \\ 0.3i & 2 & 0.3 \\ 0.2 & 0.3 + 0.4i & i \end{bmatrix}.$$

7.22.

$$\begin{bmatrix} -2 & 1/2 & i/2 \\ 1/2 & 0 & i/2 \\ -i/2 & -i/2 & 2 \end{bmatrix}.$$

7.23. Find the l_1-norm of the matrix in exercise 21.

7.24. Find the l_∞-norm of the matrix in exercise 22.

Do three iterations of the power method to find the largest eigenvalue, in absolute value, and the corresponding eigenvector of the following matrices.

7.25. $\begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}$ with $x^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

7.26. $\begin{bmatrix} 3 & 2 & 3 \\ 2 & 6 & 6 \\ 3 & 6 & 3 \end{bmatrix}$ with $x^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Exercices pour le chapitre 8

Find the Laplace transforms of the given functions.

8.1. $f(t) = -3t + 2$.

8.2. $f(t) = t^2 + at + b$.

8.3. $f(t) = \cos(\omega t + \theta)$.

8.4. $f(t) = \sin(\omega t + \theta)$.

8.5. $f(t) = \cos^2 t$.

8.6. $f(t) = \sin^2 t$.

8.7. $f(t) = -3t + 2$.

8.8. $f(t) = 2\,e^{-2t} \sin t$.

8.9. $f(t) = e^{-2t} \cosh t$.

8.10. $f(t) = \left(1 + 2e^{-t}\right)^2$.

8.11. $f(t) = u(t-1)(t-1)$.

8.12. $f(t) = u(t-1)t^2$.

8.13. $f(t) = u(t-1)\cosh t$.

8.14. $f(t) = u(t-\pi/2)\sin t$.

Find the inverse Laplace transform of the given functions.

8.15. $F(s) = \dfrac{4(s+1)}{s^2 - 16}$.

8.16. $F(s) = \dfrac{2s}{s^2 + 3}$.

8.17. $F(s) = \dfrac{2}{s^2 + 3}$.

8.18. $F(s) = \dfrac{4}{s^2 - 9}$.

8.19. $F(s) = \dfrac{4s}{s^2 - 9}$.

8.20. $F(s) = \dfrac{3s - 5}{s^2 + 4}$.

8.21. $F(s) = \dfrac{1}{s^2 + s - 20}$.

8.22. $F(s) = \dfrac{1}{(s-2)(s^2 + 4s + 3)}$.

8.23. $F(s) = \dfrac{2s + 1}{s^2 + 5s + 6}$.

8.24. $F(s) = \dfrac{s^2 - 5}{s^3 + s^2 + 9s + 9}$.

8.25. $F(s) = \dfrac{3s^2 + 8s + 3}{(s^2 + 1)(s^2 + 9)}$.

8.26. $F(s) = \dfrac{s - 1}{s^2(s^2 + 1)}$.

8.27. $F(s) = \dfrac{1}{s^4 - 9}$.

8.28. $F(s) = \dfrac{(1 + e^{-2s})^2}{s + 2}$.

8.29. $F(s) = \dfrac{e^{-3s}}{s^2(s - 1)}$.

8.30. $F(s) = \dfrac{\pi}{2} - \arctan \dfrac{s}{2}$.

8.31. $F(s) = \ln \dfrac{s^2 + 1}{s^2 + 4}$.

Find the Laplace transform of the given functions.

8.32. $f(t) = \begin{cases} t, & 0 \le t < 1, \\ 1, & t \ge 1. \end{cases}$

8.33. $f(t) = \begin{cases} 2t + 3, & 0 \le t < 2, \\ 0, & t \ge 2. \end{cases}$

8.34. $f(t) = t \sin 3t$.

8.35. $f(t) = t \cos 4t$.

8.36. $f(t) = e^{-t}\, t \cos t$.

8.37. $f(t) = \displaystyle\int_0^t \tau e^{t-\tau}\, d\tau$.

8.38. $f(t) = 1 * e^{-2t}$.

8.39. $f(t) = e^{-t} * e^{t} \cos t$.

8.40. $f(t) = \dfrac{e^t - e^{-t}}{t}$.

Use Laplace transforms to solve the given initial value problems and plot the solution.

8.41. $y'' - 6y' + 13y = 0, \quad y(0) = 0, \quad y'(0) = -3$.

8.42. $y'' + y = \sin 3t, \quad y(0) = 0, \quad y'(0) = 0$.

8.43. $y'' + y = \sin t, \quad y(0) = 0, \quad y'(0) = 0$.

8.44. $y'' + y = t, \quad y(0) = 0, \quad y'(0) = 0$.

8.45. $y'' + 5y' + 6y = 3\, e^{-2t}, \quad y(0) = 0, \quad y'(0) = 1$.

8.46. $y'' + 2y' + 5y = 4t, \quad y(0) = 0, \quad y'(0) = 0$.

8.47. $y'' - 4y' + 4y = t^3\, e^{2t}, \quad y(0) = 0, \quad y'(0) = 0$.

8.48. $y'' + 4y = \begin{cases} 1, & 0 \le t < 1 \\ 0, & t \ge 1 \end{cases}, \quad y(0) = 0, \quad y'(0) = -1.$

8.49. $y'' - 5y' + 6y = \begin{cases} t, & 0 \le t < 1 \\ 0, & t \ge 1 \end{cases}, \quad y(0) = 0, \quad y'(0) = 1.$

8.50. $y'' + 4y' + 3y = \begin{cases} 4\,e^{1-t}, & 0 \le t < 1 \\ 4, & t \ge 1 \end{cases}, \quad y(0) = 0, \quad y'(0) = 0.$

8.51. $y'' + 4y' = u(t-1), \quad y(0) = 0, \quad y'(0) = 0.$

8.52. $y'' + 3y' + 2y = 1 - u(t-1), \quad y(0) = 0, \quad y'(0) = 1.$

8.53. $y'' - y = \sin t + \delta(t - \pi/2), \qquad y(0) = 3.5, \quad y'(0) = -3.5.$

8.54. $y'' + 5y' + 6y = u(t-1) + \delta(t-2), \qquad y(0) = 0, \quad y'(0) = 1.$

Using Laplace transforms solve the given integral equations and plot the solutions.

8.55. $y(t) = 1 + \displaystyle\int_0^t y(\tau)\, d\tau.$

8.56. $y(t) = \sin t + \displaystyle\int_0^t y(\tau) \sin(t-\tau)\, d\tau.$

8.57. $y(t) = \cos 3t + 2\displaystyle\int_0^t y(\tau) \cos 3(t-\tau)\, d\tau.$

8.58. $y(t) = t + e^t + \displaystyle\int_0^t y(\tau) \cosh(t-\tau)\, d\tau.$

8.59. $y(t) = t\,e^t + 2\,e^t \displaystyle\int_0^t e^{-\tau} y(\tau)\, d\tau.$

Sketch the following 2π-periodic functions over three periods and find their Laplace transforms.

8.60. $f(t) = \pi - t, \qquad 0 < t < 2\pi.$

8.61. $f(t) = 4\pi^2 - t^2, \qquad 0 < t < 2\pi.$

8.62. $f(t) = e^{-t}, \qquad 0 < t < 2\pi.$

8.63. $f(t) = \begin{cases} t, & \text{if } 0 < t < \pi, \\ \pi - t, & \text{if } \pi < t < 2\pi. \end{cases}$

8.64. $f(t) = \begin{cases} 0, & \text{if } 0 < t < \pi, \\ t - \pi, & \text{if } \pi < t < 2\pi. \end{cases}$

Exercices pour le chapitre 9

9.1. *Soit l'équation* / Consider the equation

$$f(x) = x^2 - 2x - 3 = 0.$$

Montrer que la récurrence / Show that the fixed point iteration

$$x_{n+1} = \sqrt{2x_n + 3}$$

converge dans l'intervalle $[2, 4]$. / converges in the interval $[2, 4]$.

9.2. Soit la récurrence $x_{n+1} = \sqrt{2x_n + 3}$ de l'excercice 9.1. *Compléter le tableau* / Complete the table :

n	x_n	Δx_n	$\Delta^2 x_n$
1	$x_1 = 4.000$		
2	$x_2 =$		
3	$x_3 =$		

Accélérer la convergence par Aitken. / Accelerate convergence by Aitken.

$$\hat{x} = x_1 - \frac{(\Delta x_1)^2}{\Delta^2 x_1} =$$

9.3. Sketch the function $f(x) = x - \tan x$ and compute a root of the equation $f(x) = 0$ to six decimals by means of Newton's method with $x_0 = 1$.

9.4. 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 $x_0 = 1$.

9.5. Sketch the function $f(x) = x\,e^x$ and compute a root of the equation $f(x) = 0.5$ to six decimals by means of Newton's method with $x_0 = 1$.

9.6. Interpolate the data

$$(-1, 2), \quad (0, 0), \quad (1, -1), \quad (2, 4),$$

by means of Gregory–Newton's interpolation polynomial of degree three. Plot the data and the interpolation polynomial on the same graph.

9.7. Interpolate the data

$$(-1, 3), \quad (0, 1), \quad (1, 0), \quad (2, 5),$$

by means of Gregory–Newton's interpolation polynomial of degree three. Plot the data and the interpolation polynomial on the same graph.

9.8. Interpolate the data

$$(-1, 2), \quad (0, 0), \quad (1.5, -1), \quad (2, 4),$$

by means of Newton's divided difference interpolation polynomial of degree three. Plot the data and the interpolation polynomial on the same graph.

9.9. Interpolate the data

$$(-1, 1), \quad (0, -1), \quad (1.5, -2), \quad (2, 3),$$

by means of Newton's divided difference interpolation polynomial of degree three. Plot the data and the interpolation polynomial on the same graph.

9.10. *Compléter la table de différences divisées,*
Complete the 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			

Écrire le polynôme de degré 3 qui interpole f en $x_0 = 3.2$, $x_1 = 2.7$, $x_2 = 1.0$ et $x_3 = 4.8$.
Write the interpolating polynomial of degree 3 that fits the data at all points from $x_0 = 3.2$ to $x_3 = 4.8$.

9.11. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+x^2}$ by the trapezoidal rule with $n = 10$.

9.12. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+x^2}$ by Simpson's rule with $2n = 10$.

9.13. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+2x^2}$ by the trapezoidal rule with $n = 10$.

9.14. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+2x^2}$ by Simpson's rule with $2n = 10$.

9.15. *Déterminer les valeurs de h et n pour approcher*
Determine the values of h and n to approximate

$$\int_1^3 \ln x \, dx$$

au 10^{-3} par les méthodes composées suivantes.
to 10^{-3} by the following composite rules:
Trapèzes / Trapezoidal,
Simpson / Simpson's.
Point milieu / Midpoint.

9.16. Same as previous question with

$$\int_0^2 \frac{1}{x+4} \, dx$$

au 10^{-5}. / to 10^{-5}.