

Chapitre 4

Intégration des systèmes différentiels

4.1 Introduction

Soient $[a, b]$ un intervalle fermé de \mathbb{R} et x_0 un point fixé de $[a, b]$. Soient f une application définie et continue sur $[a, b] \times \mathbb{R}^m$ dans \mathbb{R}^m et y_0 un élément de \mathbb{R}^m .

On cherche une application y continue et dérivable de $[a, b]$ dans \mathbb{R}^m telle que

$$\forall x \in [a, b], \quad y'(x) = \frac{dy(x)}{dx} = f(x, y(x)) \quad (4.1)$$

$$\text{et } y(x_0) = y_0. \quad (4.2)$$

On dit que y est solution du système différentiel (4.1) sur $[a, b]$ si y vérifie cette relation $\forall x \in [a, b]$.

(4.1) et (4.2) forment un problème de Cauchy pour le système différentiel (4.1) ; la condition (4.2) s'appelle condition de Cauchy.

Remarque 4.1.1 Afin d'être plus explicite, il est bon de remarquer que le système (4.1) est un système de m équations :

$$\begin{cases} y'_1(x) = f_1(x, y_1(x), \dots, y_m(x)) \\ y'_2(x) = f_2(x, y_1(x), \dots, y_m(x)) \\ \vdots \\ y'_m(x) = f_m(x, y_1(x), \dots, y_m(x)). \end{cases} \quad (4.3)$$

Comme il ne fait intervenir que des dérivées premières de y , ce système est dit du premier ordre.

Un système différentiel d'ordre p de la forme

$$y^{(p)}(x) = \varphi(x, y(x), y'(x), \dots, y^{(p-1)}(x)) \text{ avec } x \in [a, b] \quad (4.4)$$

peut se ramener à un système différentiel du type (4.1). Il suffit de poser :

$$z_1(x) = y(x), z_2(x) = y'(x), \dots, z_p(x) = y^{(p-1)}(x).$$

On obtient alors

$$\begin{cases} z_{t_1}(x) = z_2(x) \\ \vdots \\ z_{t_{p-1}}(x) = z_p(x) \\ z_{t_p}(x) = \varphi(x, z_1(x), \dots, z_p(x)). \end{cases}$$

Cet ensemble de relations est de la forme (4.1). En effet si on pose $z = (z_1, z_2, \dots, z_p)^T$, $F(x, z(x)) = (z_2, z_3, \dots, z_p, \varphi(x, z_1, \dots, z_p))^T$ alors $z : [a, b] \rightarrow \mathbb{R}^p$ et $F : [a, b] \times \mathbb{R}^p \rightarrow \mathbb{R}^p$ et $z'(x) = F(x, z(x))$ pour $x \in [a, b]$.

La condition de Cauchy pour ce problème est :

$$z(x_0) = (z_1(x_0), \dots, z_p(x_0)) = (y(x_0), y'(x_0), \dots, y^{(p-1)}(x_0)).$$

Le théorème de Cauchy-Lipschitz donne l'existence et l'unicité de la solution du problème de Cauchy. x_0 sera fixé à a .

Théorème 4.1.2 (Cauchy-Lipschitz) Si f est définie et continue de $[a, b] \times \mathbb{R}^m$ et si f vérifie une condition de Lipschitz, c'est-à-dire s'il existe un réel $L > 0$ tel que $\forall (x, y)$ et $(x, z) \in [a, b] \times \mathbb{R}^m$ on ait

$$\| f(x, y) - f(x, z) \| \leq L \| y - z \| \quad (4.5)$$

où $\| \cdot \|$ est une norme vectorielle sur \mathbb{R}^m (L est appelée constante de Lipschitz de f), alors le problème de Cauchy (4.1) et (4.2) admet une solution et une seule sur $[a, b]$ et ceci quel que soit $y_0 \in \mathbb{R}^m$.

Démonstration.

La démonstration qui en est donnée reste valable si on remplace \mathbb{R}^m par un espace de Banach. Pour simplifier cette démonstration nous l'effectuerons dans le cas où $m = 1$ et on considérera l'espace de Banach $C_\infty([a, b])$ des fonctions continues définies sur $[a, b]$ à valeurs dans \mathbb{R} , muni de la norme

$$\| g \|_\infty = \max_{x \in [a, b]} | g(x) | \quad \text{si } g \in C_\infty([a, b]).$$

La solution y du problème de Cauchy vérifie

$$y(x) = y_0 + \int_a^x f(t, y(t)) dt \quad \forall x \in [a, b]. \quad (4.6)$$

Commençons par démontrer par l'absurde l'unicité de la solution si elle existe. Supposons qu'il existe une autre solution $z \in C_\infty([a, b])$. Nous avons à partir de (4.5)

$$y(x) - z(x) = \int_a^x \left(f(t, y(t)) - f(t, z(t)) \right) dt.$$

D'où en utilisant la constante de Lipschitz de f (cf. (4.5)) nous obtenons

$$\begin{aligned} | y(x) - z(x) | &\leq L \int_a^x | y(t) - z(t) | dt \\ &\leq L \int_a^x \max_{t \in [a, b]} | y(t) - z(t) | dt. \end{aligned} \quad (4.7)$$

Soit

$$| y(x) - z(x) | \leq L(x - a) \| y - z \|_\infty \quad \forall x \in [a, b].$$

En reportant cette inégalité dans (4.7) on trouve

$$\begin{aligned} |y(x) - z(x)| &\leq L^2 \int_a^x (t-a) \|y - z\|_\infty dt \\ &= L^2 \frac{(x-a)^2}{2} \|y - z\|_\infty \end{aligned}$$

inégalité qui peut être reportée à nouveau dans (4.7).

Montrons par récurrence que

$$|y(x) - z(x)| \leq L^n \frac{(x-a)^n}{n!} \|y - z\|_\infty \quad \forall n \in \mathbb{N}. \quad (4.8)$$

Si on reporte cette inégalité dans (4.7), il vient

$$\begin{aligned} |y(x) - z(x)| &\leq \frac{L^{n+1}}{n!} \|y - z\|_\infty \int_a^x (t-a)^n dt \\ &= L^{n+1} \frac{(x-a)^{n+1}}{(n+1)!} \|y - z\|_\infty. \end{aligned}$$

Puisque (4.8) est vraie $\forall x \in [a, b]$, alors

$$\|y - z\|_\infty \leq L^n \frac{(b-a)^n}{n!} \|y - z\|_\infty.$$

D'où

$$(1 - L^n \frac{(b-a)^n}{n!}) \|y - z\|_\infty \leq 0.$$

Donc quand n tend vers $+\infty$, il en résulte que $\|y - z\|_\infty \leq 0$, ce qui n'est possible que si $\|y - z\|_\infty = 0$, c'est-à-dire que $y = z$. D'où l'unicité de la solution.

Démontrons maintenant l'existence de la solution à l'aide des itérations de Picard.

Considérons la suite des fonctions \hat{y}_k de $C_\infty([a, b])$ définies par

$$\begin{aligned} \hat{y}_0(x) &= y_0 \\ \hat{y}_{k+1}(x) &= y_0 + \int_a^x f(t, \hat{y}_k(t)) dt \quad \text{pour } k = 0, 1, \dots \end{aligned} \quad (4.9)$$

Formons la différence $\hat{y}_{k+1}(x) - \hat{y}_k(x)$ et utilisons la propriété que f est Lipschitzienne.

$$\begin{aligned} |\hat{y}_{k+1}(x) - \hat{y}_k(x)| &= \left| \int_a^x (f(t, \hat{y}_k(t)) - f(t, \hat{y}_{k-1}(t))) dt \right| \\ &\leq \int_a^x |f(t, \hat{y}_k(t)) - f(t, \hat{y}_{k-1}(t))| dt \\ &\leq L \int_a^x |\hat{y}_k(t) - \hat{y}_{k-1}(t)| dt. \end{aligned}$$

On retrouve une expression analogue à (4.7)). Par récurrence on peut donc montrer simplement de la même façon qui nous a permis d'obtenir (4.8)), que nous avons

$$|\hat{y}_{k+1}(x) - \hat{y}_k(x)| \leq L^k \frac{(x-a)^k}{k!} \|\hat{y}_1 - \hat{y}_0\|_\infty \quad \forall x \in [a, b],$$

soit encore

$$\|\hat{y}_{k+1} - \hat{y}_k\|_\infty \leq L^k \frac{(b-a)^k}{k!} \|\hat{y}_1 - \hat{y}_0\|_\infty.$$

Pour un entier $p \geq 1$ nous avons donc

$$\| \hat{y}_{k+p} - \hat{y}_k \|_{\infty} \leq \sum_{i=0}^{p-1} \| \hat{y}_{k+i+1} - \hat{y}_{k+i} \|_{\infty} \leq L^k \frac{(b-a)^k}{k!} \sum_{i=0}^{p-1} L^i \frac{(b-a)^i}{i!} \| \hat{y}_1 - \hat{y}_0 \|_{\infty}$$

en utilisant le fait que $k!i! \leq (k+i)!$ pour $k \geq 1$ et $i \geq 0$.

D'autre part

$$\sum_{i=0}^{p-1} L^i \frac{(b-a)^i}{i!} \leq e^{L(b-a)}.$$

Finalement nous obtenons

$$\| \hat{y}_{k+p} - \hat{y}_k \|_{\infty} \leq L^k \frac{(b-a)^k}{k!} e^{L(b-a)} \| \hat{y}_1 - \hat{y}_0 \|_{\infty}.$$

$L^k \frac{(b-a)^k}{k!}$ est une quantité qui tend vers 0 quand k tend vers l'infini. Donc $\forall \varepsilon > 0$ on peut trouver $K(\varepsilon)$ tel que $\forall k \geq K(\varepsilon)$ et $\forall p \geq 1$ on ait

$$| \| \hat{y}_{k+p} \|_{\infty} - \| \hat{y}_k \|_{\infty} | \leq \| \hat{y}_{k+p} - \hat{y}_k \|_{\infty} \leq \varepsilon,$$

ce qui démontre que la suite $\{\hat{y}_k\}$ est une suite de Cauchy. Comme $C_{\infty}([a, b])$ est un espace de Banach (i.e. complet : toute suite de Cauchy converge dans cet espace), \hat{y}_k converge uniformément vers $\hat{y} \in C_{\infty}([a, b])$ quand k tend vers l'infini.

En passant à la limite dans (4.9) on a

$$\hat{y}(x) = y_0 + \int_a^x f(t, \hat{y}(t)) dt,$$

ce qui démontre que \hat{y} est solution du problème de Cauchy (4.1) et (4.2)). \square

Pratiquement la vérification de la condition de Lipschitz est souvent difficile. Par contre il peut être plus simple de vérifier une condition sur les différentielles de Fréchet.

Définition 4.1.3 Soit E un espace de Banach et U une partie ouverte de E . Soit f une application de U dans lui-même. On dit que f est différentiable au sens de Fréchet au voisinage de x s'il existe une application linéaire $f'(x)$ de E dans lui-même telle que

$$f(x+h) - f(x) = f'(x)h + \Phi(x, h)$$

pour tout h appartenant à un voisinage V de l'origine et pour tout $x+h \in V$, et telle que

$$\lim_{\|h\| \rightarrow 0} \frac{\| \Phi(x, h) \|}{\| h \|} = 0.$$

f' est appelée la différentielle de Fréchet de f en x .

Théorème 4.1.4 Si f est différentiable au sens de Fréchet par rapport à y , si la différentielle est continue et si

$$\left\| \frac{\partial f}{\partial y} \right\| \leq M \quad \forall x \in [a, b] \text{ et } \forall y \in \mathbb{R}^m$$

(où $\| \cdot \|$ est une norme matricielle), alors f vérifie une condition de Lipschitz.

Par conséquent le problème de Cauchy (4.1) et (4.2) admet une solution unique.

Démonstration.

Dans ce cas on peut écrire que $\forall x \in [a, b]$ et $\forall y$ et $z \in \mathbb{R}^m$, il existe $\bar{y} \in \mathbb{R}^m$ tel que

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

d'où la condition de Lipschitz. \square

Remarque fondamentale

Les théorèmes 4.1.2 et 4.1.4 démontrent l'existence et l'unicité d'une **solution globale** sur $[a, b]$. C'est la conséquence de la condition de Lipschitz vérifiée par f , $\forall x \in [a, b]$ et $\forall y$ et $z \in \mathbb{R}^m$.

Si la condition de Lipschitz n'est vérifiée que localement, alors l'existence et l'unicité de la solution ne peuvent être démontrées qu'au voisinage de (x_0, y_0) . On peut perdre l'unicité de la solution si la condition de Lipschitz n'est pas vérifiée comme le montre l'exemple suivant :

$$y' = 3y^{\frac{2}{3}}.$$

Au voisinage de 0 la condition de Lipschitz n'est pas vérifiée. En effet

$$\frac{|y^{\frac{2}{3}} - 0|}{|y - 0|} = \frac{1}{|y|^{\frac{1}{3}}}$$

n'est pas borné quand y tend vers zéro.

La solution générale de cette équation différentielle est

$$y = (x - C)^3$$

et $y = 0$.

Le problème de Cauchy correspondant à $y(0) = 0$ admet une infinité de solutions. En voici quelques-unes :

$$y(x) = 0; y(x) = x^3; y(x) = \begin{cases} 0 & \text{pour } x \leq 0 \\ x^3 & \text{pour } x > 0 \end{cases}; \text{ etc ...}$$

Dans toute la suite de ce chapitre nous supposons que f vérifie toujours une condition de Lipschitz.

La résolution numérique des problèmes de Cauchy consiste à subdiviser l'intervalle $[a, b]$ en n sous-intervalles $[x_i, x_{i+1}]$ tels que

$$x_0 = a < x_1 < x_2 < \dots < x_{n-1} < x_n = b.$$

On pose $h_i = x_{i+1} - x_i$. On approche la suite $\{y(x_i)\}$, valeur de la solution en x_i , par la suite $\{y_i\}$ des valeurs approchées données par une méthode d'intégration numérique exposée ci-après. Ces méthodes se partagent en deux classes :

- les méthodes à un pas (ou à pas séparés) : y_{i+1} est calculé uniquement en fonction de y_i .
- les méthodes à plusieurs pas (ou à pas liés) : y_{i+1} est calculé en fonction de $y_i, y_{i-1}, \dots, y_{i-k}$.

Chacune de ces classes se divise elle-même en deux sous-classes : celle des méthodes explicites et celle des méthodes implicites, suivant que y_{i+1} est donné explicitement comme solution d'un système d'équations linéaires par rapport aux y_i précédents ou que y_{i+1} est donné implicitement comme solution d'un système d'équations non-linéaires.

4.2 Les méthodes à un pas

Ces méthodes (explicites) s'écrivent sous la forme :

$$y_{i+1} = y_i + h_i \Phi(x_i, y_i, h_i) \quad (4.10)$$

avec $y_0 = y(x_0) = y(a)$.

Φ est une application continue de $[a, b] \times \mathbb{R}^m \times [0, h^*]$ dans \mathbb{R}^m , où h^* est tel que $0 \leq h^* \leq b - a$.

Bien évidemment y_i ne peut être une bonne approximation de $y(x_i)$ sans que Φ satisfasse un certain nombre de conditions théoriques que nous allons passer en revue.

Nous supposons que $m = 1$ pour simplifier les démonstrations bien que le fait d'avoir $m > 1$ ne présente pas de difficulté majeure.

4.2.1 La consistance

Cette notion exprime le fait que la fonction incrément Φ doit être liée à l'équation différentielle.

Définition 4.2.1 *La méthode (4.10)) est consistante avec l'équation différentielle (4.1) si pour toute solution $y(\cdot)$ de l'équation différentielle (4.1) on a*

$$\lim_{h \rightarrow 0} \sum_{j=0}^{n-1} |y(x_{j+1}) - y(x_j) - h_j \Phi(x_j, y(x_j), h_j)| = 0 \quad (4.11)$$

avec $h = \max_{0 \leq j \leq n-1} h_j$.

La quantité $y(x_{j+1}) - y(x_j) - h_j \Phi(x_j, y(x_j), h_j)$ qui est appelée erreur de consistance en x_j , représente dans un certain sens l'erreur que l'on fait au j ème pas en remplaçant l'équation différentielle (4.1) par la méthode (4.10)).

La relation (4.11) représente la somme de ces erreurs en valeur absolue qui devient négligeable dès que h est assez petit.

Le théorème suivant donne une condition nécessaire et suffisante de consistance qui est bien plus pratique que la définition 4.2.1.

Théorème 4.2.2 *Une condition nécessaire et suffisante pour que la méthode (4.10)) soit consistante avec l'équation différentielle (4.1) est que*

$$\Phi(x, y(x), 0) = f(x, y(x)) \quad \forall x \in [a, b] \text{ et } \forall y \in \mathbb{R}.$$

Démonstration.

Posons $\varepsilon_j = y(x_{j+1}) - y(x_j) - h_j \Phi(x_j, y(x_j), h_j)$.

D'après la formule des accroissements finis, il existe $c_j \in [x_j, x_{j+1}]$ tel que

$$\varepsilon_j = h_j [f(c_j, y(c_j)) - \Phi(x_j, y(x_j), h_j)].$$

Posons

$$\begin{cases} \alpha_j = h_j [f(c_j, y(c_j)) - \Phi(c_j, y(c_j), 0)], \\ \beta_j = \Phi(c_j, y(c_j), 0) - \Phi(x_j, y(x_j), h_j). \end{cases}$$

Donc $\varepsilon_j = \alpha_j + h_j \beta_j$.

La fonction $x \rightarrow |f(x, y(x)) - \Phi(x, y(x), 0)|$ est continue, donc intégrable au sens de Riemann, c'est-à-dire que

$$\lim_{h \rightarrow 0} \sum_{j=0}^{n-1} |\alpha_j| = \int_a^b |f(t, y(t)) - \Phi(t, y(t), 0)| dt$$

D'autre part

$$|\beta_j| \leq \beta(h) = \max_{\substack{|x-x'| \leq h \\ 0 \leq h' \leq h}} |\Phi(x, y(x), 0) - \Phi(x', y(x'), h')|.$$

La fonction $(x, h) \rightarrow \Phi(x, y(x), h)$ est continue sur le compact $[a, b] \times [0, h^*]$. Elle est donc uniformément continue et par conséquent $\lim_{h \rightarrow 0} \beta(h) = 0$. D'où

$$\sum_{j=0}^{n-1} |h_j \beta_j| \leq \sum_{j=0}^{n-1} h_j \beta(h) = (b-a)\beta(h),$$

ce qui entraîne que

$$\lim_{h \rightarrow 0} \sum_{j=0}^{n-1} |h_j \beta_j| = 0.$$

Nous obtenons donc, en utilisant le fait que $\varepsilon_j = \alpha_j + h_j \beta_j$,

$$\lim_{h \rightarrow 0} \sum_{j=0}^{n-1} |\varepsilon_j| = \int_a^b |f(t, y(t)) - \Phi(t, y(t), 0)| dt.$$

La définition 4.2.1 de la consistance de la méthode (4.10) nous dit que pour toute solution de l'équation différentielle (4.1) on a $\lim_{h \rightarrow 0} \sum_{j=0}^{n-1} |\varepsilon_j| = 0$.

Par conséquent $\forall x \in [a, b]$ on a $f(x, y(x)) = \Phi(x, y(x), 0)$.

D'après le théorème 4.1.2, pour tout $(x^*, y^*) \in [a, b] \times \mathbb{R}$, il existe une solution de l'équation (4.1) qui vérifie $y(x^*) = y^*$; par conséquent la consistance entraîne que $f(x^*, y^*) = \Phi(x^*, y^*, 0)$. \square

4.2.2 La stabilité théorique

Définition 4.2.3 Soient $\{y_i\}$, $\{z_i\}$ et $\{\tilde{\varepsilon}_i\}$ pour $i = 0, 1, \dots, n$ les suites vérifiant

$$\begin{cases} y_{i+1} = y_i + h_i \Phi(x_i, y_i, h_i) \\ y_0 \in \mathbb{R} \text{ quelconque donné.} \end{cases}$$

$$\begin{cases} z_{i+1} = z_i + h_i \Phi(x_i, z_i, h_i) + \tilde{\varepsilon}_i \\ z_0 \in \mathbb{R} \text{ quelconque donné.} \end{cases}$$

La méthode (4.10) est dite stable s'il existe une constante M indépendante de h telle que

$$\max_{0 \leq i \leq n} |z_i - y_i| \leq M (|z_0 - y_0| + \sum_{j=0}^{n-1} |\tilde{\varepsilon}_j|).$$

Théorème 4.2.4 Une condition suffisante pour que la méthode (4.10) soit stable est que Φ vérifie une condition de Lipschitz, c'est-à-dire $\forall x \in [a, b]$, $\forall y$ et $z \in \mathbb{R}$, $\forall h \in [0, h^*]$, on a $|\Phi(x, y, h) - \Phi(x, z, h)| \leq M^* |y - z|$ où M^* est une constante indépendante de h .

Démonstration.

$$\begin{aligned} |y_{i+1} - z_{i+1}| &\leq |y_i - z_i| + h_i |\Phi(x_i, y_i, h_i) - \Phi(x_i, z_i, h_i)| + |\tilde{\varepsilon}_i| \\ &\leq (1 + h_i M^*) |y_i - z_i| + |\tilde{\varepsilon}_i|. \end{aligned}$$

On en déduit par récurrence que

$$|y_{i+1} - z_{i+1}| \leq \prod_{j=0}^i (1 + h_j M^*) |y_0 - z_0| + \sum_{j=0}^i \prod_{k=j+1}^i (1 + h_k M^*) |\tilde{\varepsilon}_j|.$$

En utilisant le fait que $1 + K \leq e^K$ pour tout K (la courbe représentative de l'exponentielle est toujours au-dessus de sa tangente au point (0,1)), on a

$$\prod_{k=j+1}^i (1 + h_k M^*) \leq \prod_{k=j+1}^i e^{h_k M^*} = e^{M^* \sum_{k=j+1}^i h_k} = e^{M^* (x_{i+1} - x_{j+1})}.$$

D'où

$$\begin{aligned} |y_{i+1} - z_{i+1}| &\leq e^{M^*(x_{i+1}-x_0)} |y_0 - z_0| + \sum_{j=0}^i |\tilde{\varepsilon}_j| e^{M^*(x_{i+1}-x_{j+1})} \\ &\leq e^{M^*(b-a)} (|y_0 - z_0| + \sum_{j=0}^i |\tilde{\varepsilon}_j|). \end{aligned}$$

Il suffit de prendre $M = e^{M^*(b-a)}$ pour que la propriété soit démontrée. \square

4.2.3 La convergence

Définition 4.2.5 La méthode (4.10) est convergente si

$$\lim_{h \rightarrow 0} \max_{0 \leq j \leq n} |y_j - y(x_j)| = 0$$

avec $h = \max_{0 \leq j \leq n-1} h_j$.

Théorème 4.2.6 Si la méthode (4.10) est stable et consistante, alors elle est convergente.

Démonstration.

Posons $\varepsilon_i = y(x_{i+1}) - y(x_i) - h_i \Phi(x_i, y(x_i), h_i)$ et $y(x_0) = y_0$.

Si la méthode est consistante, alors d'après la définition 4.2.1 nous avons $\lim_{h \rightarrow 0} \sum_{j=0}^{n-1} |\varepsilon_j| = 0$.

Puisque la méthode est stable, alors, d'après la définition 4.2.3 (en considérant dans cette définition que $z_i = y(x_i)$ et $z_{i+1} = y(x_{i+1})$), il existe une constante M indépendante de h telle que

$$\max_{0 \leq i \leq n} |y(x_i) - y_i| \leq M \sum_{i=0}^{n-1} |\varepsilon_i|.$$

D'où la convergence en faisant tendre h vers 0. \square

4.2.4 L'ordre

Définition 4.2.7 On dit que la méthode (4.10) est d'ordre p ($p > 0$) s'il existe un réel K ne dépendant que de $y(\cdot)$ et de Φ tel que

$$\sum_{i=0}^{n-1} |y(x_{i+1}) - y(x_i) - h_i \Phi(x_i, y(x_i), h_i)| \leq K h^p$$

pour toute solution $y(\cdot) \in C^{p+1}([a, b])$ de l'équation différentielle (4.1).

Théorème 4.2.8 Si la méthode (4.10) est stable et d'ordre p et si $f \in C^p([a, b] \times \mathbb{R})$, alors $\forall i \in \mathbb{N} \quad |y(x_i) - y_i| \leq M K h^p$.

Démonstration.

Posons $\varepsilon_i = y(x_{i+1}) - y(x_i) - h_i \Phi(x_i, y(x_i), h_i)$.

Comme dans le théorème 4.2.6, la stabilité donne

$$|y(x_i) - y_i| \leq M \sum_{j=0}^{n-1} |\varepsilon_j|.$$

D'autre part l'ordre p donne

$$\sum_{j=0}^{n-1} |\varepsilon_j| \leq K h^p.$$

D'où le résultat. \square

Nous allons démontrer une propriété qui permettra l'obtention pratique de l'ordre.

Tout d'abord remarquons que l'on a

$$f^{(1)}(x, y(x)) = \frac{d}{dx} f(x, y(x)) = \frac{\partial}{\partial x} f(x, y(x)) + f(x, y(x)) \frac{\partial}{\partial y} f(x, y(x)).$$

Plus généralement

$$\begin{aligned} f^{(k)}(x, y(x)) &= \frac{d}{dx} f^{(k-1)}(x, y(x)) \\ &= \frac{\partial}{\partial x} f^{(k-1)}(x, y(x)) + f(x, y(x)) \frac{\partial}{\partial y} f^{(k-1)}(x, y(x)). \end{aligned}$$

Si y est solution de $y'(x) = f(x, y(x))$ et si $f(x, y(x))$ est suffisamment dérivable, alors

$$\forall k \geq 0, y^{(k+1)}(x) = f^{(k)}(x, y(x)) = \frac{d^k}{dx^k} f(x, y(x)),$$

où $f^{(0)}(x, y(x)) = f(x, y(x))$.

Théorème 4.2.9 *On suppose que f est une fonction p fois continûment différentiable dans $[a, b] \times \mathbb{R}$ et que les fonctions $\Phi, \frac{\partial \Phi}{\partial h}, \dots, \frac{\partial^p \Phi}{\partial h^p}$ existent et sont continues dans $[a, b] \times \mathbb{R} \times [0, h^*]$.*

Dans ce cas une condition nécessaire et suffisante pour que la méthode soit d'ordre p est que, quel que soit $(x, y) \in [a, b] \times \mathbb{R}$,

$$\frac{\partial^j}{\partial h^j} \Phi(x, y(x), h) \big|_{h=0} = \frac{1}{j+1} f^{(j)}(x, y(x)) \text{ pour } j = 0, \dots, p-1. \quad (4.12)$$

Démonstration.

On pose encore $\varepsilon_i = y(x_{i+1}) - y(x_i) - h_i \Phi(x_i, y(x_i), h_i)$ et

$$\Psi_k(x, y(x)) = \frac{1}{(k+1)!} f^{(k)}(x, y(x)) - \frac{1}{k!} \frac{\partial^k \Phi}{\partial h^k} \Phi(x, y(x), h) \big|_{h=0}.$$

La formule de Taylor donne

$$\varepsilon_i = \sum_{k=0}^{p-1} h_i^{k+1} \Psi_k(x_i, y(x_i)) + O(h_i^{p+1}).$$

Les conditions (4.12) sont équivalentes à $\Psi_k(x, y(x)) \equiv 0, \forall k < p$.

Donc si elles sont vérifiées, on a

$$|\varepsilon_i| \leq C h_i^{p+1} \leq C h_i h^p, \text{ et } \sum_{i=0}^{n-1} |\varepsilon_i| \leq C h^p \sum_{i=0}^{n-1} h_i = C(b-a)h^p,$$

ce qui montre que les conditions (4.12) sont suffisantes.

Montrons à présent qu'elles sont aussi nécessaires. Supposons que les conditions (4.12) ne soient pas vérifiées. Il existe donc un entier $k < p$ tel que $\Psi_k(x, y) \not\equiv 0$. On peut se placer dans le cas où $h_i = h$ est constant. Nous avons par conséquent

$$\varepsilon_i = h^{k+1} \Psi_k(x_i, y(x_i)) + O(h^{k+2})$$

$$\text{et } \sum_{i=0}^{n-1} |\varepsilon_i| = h^k \sum_{i=0}^{n-1} h |\Psi_k(x_i, y(x_i))| + (b-a) O(h^{k+1}).$$

Si la méthode est d'ordre p , on a

$$0 = \lim_{h \rightarrow 0} \sum_{i=0}^{n-1} \frac{|\varepsilon_i|}{h^k} = \lim_{h \rightarrow 0} \sum_{i=0}^{n-1} h |\Psi_k(x_i, y(x_i))| = \int_a^b |\Psi_k(t, y(t))| dt,$$

ce qui entraîne que $\Psi_k(t, y(t)) = 0$ quel que soit $t \in [a, b]$ et pour toute solution $y(\cdot)$ de $y'(x) = f(x, y(x))$.

On en déduit donc, comme dans le théorème 4.2.2, que $\Psi_k(x, y(x)) \equiv 0$, ce qui est en contradiction avec l'hypothèse. Donc les conditions (4.12) sont vérifiées. \square

Remarque 4.2.10 *On retiendra, d'après la démonstration précédente, que pour une méthode d'ordre p on a*

$$\varepsilon_i = O(h_i^{p+1}).$$

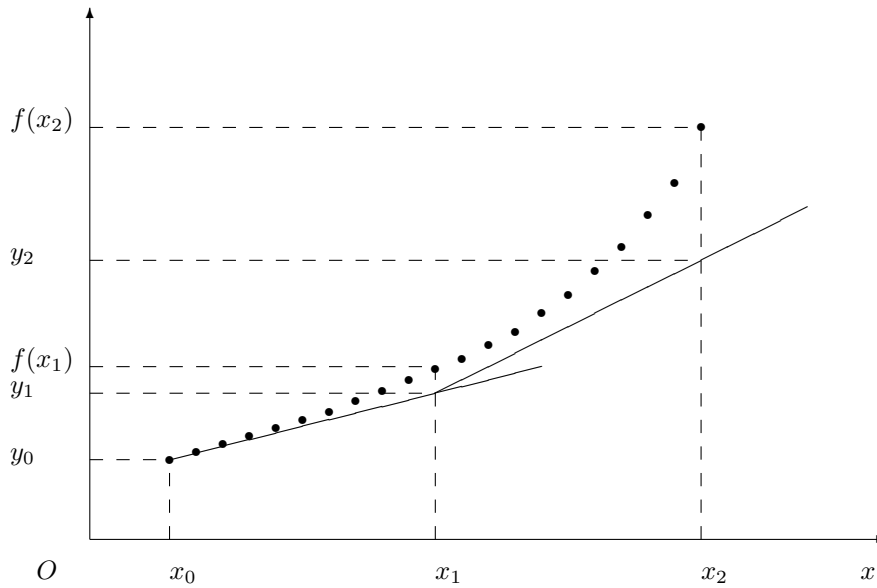
4.2.5 La méthode d'Euler

La méthode d'Euler est la méthode à un pas la plus simple qui vérifie les trois premières notions théoriques de consistance, de stabilité et donc de convergence. Elle s'écrit

$$\begin{cases} y_{k+1} = y_k + h_k f(x_k, y_k) \\ y_0 \in R \text{ quelconque donné,} \end{cases}$$

avec $\Phi(x, y, h) = f(x, y)$. Elle vérifie donc immédiatement la notion de consistance. Comme f est lipschitzienne, Φ l'est également par voie de conséquence. Donc la méthode d'Euler est convergente. A l'aide du théorème 4.2.9, on voit que cette méthode est d'ordre 1, puisque $\Phi(x, y, 0) = f(x, y)$.

On peut donner une interprétation géométrique de cette méthode. En effet le point y_1 est situé sur la tangente à la courbe intégrale solution de l'équation différentielle munie de la condition initiale $y(x_0) = y_0$. Le point y_2 est situé sur la tangente à la courbe intégrale solution de l'équation différentielle munie de la valeur y_1 comme condition initiale en x_1 .



4.2.6 Les méthodes de Runge et Kutta

On peut justifier les méthodes de Runge et Kutta en les associant à des formules de quadrature.

Soient r réels, distincts ou non, $\theta_1, \theta_2, \dots, \theta_r$ (en pratique $0 \leq \theta_i \leq 1$). On considère les formules de quadrature numérique suivantes :

$$\int_0^{\theta_i} \Psi(t) dt \simeq \sum_{j=1}^r a_{ij} \Psi(\theta_j) \text{ pour } i = 1, 2, \dots, r \quad (4.13)$$

et

$$\int_0^1 \Psi(t) dt \simeq \sum_{j=1}^r c_j \Psi(\theta_j). \quad (4.14)$$

Si on prend les abscisses intermédiaires $x_{k,i} = x_k + \theta_i h_k$ et si la fonction y est solution de $y'(x) = f(x, y(x))$, alors on obtient en utilisant (4.13)

$$\begin{aligned} y(x_{k,i}) &= y(x_k) + \int_{x_k}^{x_{k,i}} f(t, y(t)) dt \\ &\simeq y(x_k) + h_k \sum_{j=1}^r a_{ij} f(x_{k,j}, y(x_{k,j})) \text{ pour } i = 1, 2, \dots, r. \end{aligned} \quad (4.15)$$

Avec (4.14) on obtient

$$y(x_{k+1}) \simeq y(x_k) + h_k \sum_{i=1}^r c_i f(x_{k,i}, y(x_{k,i})). \quad (4.16)$$

Dans les méthodes de Runge-Kutta on remplace les deux égalités approchées (4.15) et (4.16) (pour avoir des égalités il y manque le terme d'erreur de quadrature) par des égalités dans lesquelles $y(x_{k,j})$ est remplacé par $y_{k,j}$, ce qui donne

$$y_{k,i} = y_k + h_k \sum_{j=1}^r a_{ij} f(x_{k,j}, y_{k,j}) \text{ pour } i = 1, \dots, r, \quad (4.17)$$

$$y_{k+1} = y_k + h_k \sum_{i=1}^r c_i f(x_{k,i}, y_{k,i}), \quad (4.18)$$

avec $x_{k,i} = x_k + \theta_i h_k$ pour $i = 1, \dots, r$.

Ce schéma est encore équivalent à l'écriture d'une méthode à un pas (relation (4.10)),

$$y_{k+1} = y_k + h_k \Phi(x_k, y_k, h_k), \quad (4.19)$$

où la fonction Φ est définie par les relations suivantes :

$$\hat{y}_i = y + h \sum_{j=1}^r a_{ij} f(x + \theta_j h, \hat{y}_j) \text{ pour } i = 1, \dots, r, \quad (4.20)$$

$$\Phi(x, y, h) = \sum_{i=1}^r c_i f(x + \theta_i h, \hat{y}_i). \quad (4.21)$$

Bien évidemment une méthode de Runge-Kutta est entièrement définie dès que l'on connaît r , les coefficients $a_{i,j}$, c_j et θ_j . Il est commode de représenter une méthode de Runge-Kutta par le tableau suivant :

θ_1	a_{11}	a_{12}	\cdots	a_{1r}
θ_2	a_{21}	a_{22}	\cdots	a_{2r}
\vdots	\vdots	\vdots	\cdots	\vdots
θ_r	a_{r1}	a_{r2}	\cdots	a_{rr}
	c_1	c_2	\cdots	c_r

(4.22)

On remarquera que l'on a aussi

$$\begin{pmatrix} y_{k,1} \\ \vdots \\ y_{k,r} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} y_k \\ \vdots \\ y_k \end{pmatrix} + h_k \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ \vdots & \vdots & \cdots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rr} \\ c_1 & c_2 & \cdots & c_r \end{pmatrix} \begin{pmatrix} f(x_{k,1}, y_{k,1}) \\ \vdots \\ f(x_{k,r}, y_{k,r}) \end{pmatrix}.$$

Par exemple on retrouve la méthode d'Euler avec le tableau

$$\begin{array}{|c|c|} \hline 0 & 0 \\ \hline & 1 \\ \hline \end{array} \quad \text{et} \quad \begin{cases} y_{n,1} = y_n \\ y_{n+1} = y_n + h_n f(x_n, y_{n,1}). \end{cases}$$

4.2.7 Consistance, stabilité et ordre des méthodes de Runge-Kutta

Les relations (4.19), (4.20) et (4.21) ont une forme équivalente qui est souvent plus commode pour la programmation des méthodes de Runge-Kutta. On pose $k_{\ell,i} = f(x_{\ell,i}, y_{\ell,i})$. On obtient

$$k_{\ell,i} = f(x_{\ell,i}, y_{\ell} + h_{\ell} \sum_{j=1}^r a_{i,j} k_{\ell,j}) \quad \text{pour } i = 1, \dots, r \quad (4.23)$$

et

$$y_{\ell+1} = y_{\ell} + h_{\ell} \sum_{i=1}^r c_i k_{\ell,i}. \quad (4.24)$$

On résout le système (4.23) de r équations à r inconnues $k_{\ell,i}$ pour $i = 1, \dots, r$. On en déduit $y_{\ell+1}$ par (4.24).

Pour réobtenir le système (4.17) et (4.18), il suffit de poser

$$y_{\ell,i} = y_{\ell} + h_{\ell} \sum_{j=1}^r a_{i,j} k_{\ell,j}.$$

Une condition nécessaire et suffisante pour qu'une méthode de Runge-Kutta soit consistante est que

$$\sum_{i=1}^r c_i = 1.$$

En effet $\Phi(x, y, 0) = \sum_{i=1}^r c_i f(x, y_i)$ où $y_i = y$ d'après (4.20). D'où le résultat.

Notons A la matrice $r \times r$ du tableau (4.22) qui a pour coefficients les $a_{i,j}$ pour i et $j = 1, \dots, r$, θ la matrice diagonale $r \times r$ dont les éléments diagonaux sont les θ_i , c le vecteur de \mathbb{R}^r de composantes c_1, c_2, \dots, c_r .

On voit que si A est strictement triangulaire inférieure (i.e. $a_{i,j} = 0$ si $i \leq j$) la résolution du système (4.17) (ou sa forme équivalente (4.23)) est immédiate, puisque chaque $y_{k,i}$ n'est fonction que des $y_{k,j}$ pour $j = 1, \dots, i-1$. La méthode de Runge-Kutta est alors dite explicite.

Si A est triangulaire inférieure (i.e. $a_{i,j} = 0$ si $i < j$), la méthode de Runge-Kutta est dite semi-implicite, car on doit résoudre à chaque étape r fois une équation non-linéaire à une inconnue.

Dans le cas général d'une matrice A carrée $r \times r$, on doit résoudre un système de r équations à r inconnues. On parle alors de méthode de Runge-Kutta implicite.

Nous noterons $|A|$ la matrice $r \times r$ de coefficients $|a_{i,j}|$ pour i et $j = 1, \dots, r$ et par $\rho(|A|)$ son rayon spectral.

Théorème 4.2.11 *Si $h_i L \rho(|A|) < 1$, où L est la constante de Lipschitz de f , alors le schéma de Runge-Kutta (4.17), (4.18) admet une solution unique. De plus si $h^* L \rho(|A|) < 1$, alors le schéma de Runge-Kutta est stable pour tout h tel que $0 < h \leq h^*$.*

Si la fonction $f(x, y)$ est k fois continûment différentiable sur $[a, b] \times \mathbb{R}$, alors la fonction $\Phi(x, y, h)$ est aussi k fois continûment différentiable sur $[a, b] \times \mathbb{R} \times [0, h^]$.*

Démonstration.

Commençons par prouver l'existence.

Le système non linéaire (4.20) de r équations à r inconnues peut s'écrire

$$y_i = F_i(y_1, \dots, y_r) \text{ pour } i = 1, \dots, r.$$

Soit Y le vecteur de \mathbb{R}^r de composantes y_1, \dots, y_r et $F(Y)$ le vecteur de \mathbb{R}^r de composantes $F_1(Y), \dots, F_r(Y)$. De (4.20) on déduit que

$$|F_i(Y) - F_i(Z)| \leq h \sum_{j=1}^r |a_{i,j}| L |y_j - z_j|.$$

On notera $|\cdot|$, où \cdot est un vecteur de \mathbb{R}^r , le vecteur de \mathbb{R}^r ayant pour composantes la valeur absolue des composantes du vecteur. Alors

$$|F(Y) - F(Z)| \leq h L A |y - z|,$$

où le signe \leq doit être compris comme une inégalité relative à chacune des composantes du vecteur du membre de gauche avec celui du membre de droite (i.e. $|Y| \leq |Z| \iff |y_i| \leq |z_i|$ pour $i = 1, \dots, r$).

Si on note $F^{\ell+1}(Y) = F(F^\ell(Y))$, on obtient par récurrence

$$\begin{aligned} |F_i^\ell(Y) - F_i^\ell(Z)| &= |F_i(F^{\ell-1}(Y)) - F_i(F^{\ell-1}(Z))| \\ &\leq h \sum_{j=1}^r |a_{i,j}| L |F_j^{\ell-1}(Y) - F_j^{\ell-1}(Z)| \end{aligned}$$

car $F_i(F^{\ell-1}(Y)) = y_i + h \sum_{j=1}^r a_{i,j} f(x + \theta_j h, F_j^{\ell-1}(Y))$.

Par conséquent on a

$$|F^\ell(Y) - F^\ell(Z)| \leq h L |A| |F^{\ell-1}(Y) - F^{\ell-1}(Z)|$$

et par récurrence on obtient (à cause de la positivité de toutes les composantes de $|A|$ et des vecteurs manipulés)

$$|F^\ell(Y) - F^\ell(Z)| \leq h^\ell L^\ell |A|^\ell |Y - Z|.$$

D'où $\forall \ell \geq 0, \|F^\ell(Y) - F^\ell(Z)\|_\infty \leq h^\ell L^\ell |A|^\ell \|Y - Z\|_\infty$.

L'hypothèse faite que $h L \rho(|A|) < 1$ montre qu'il existe ℓ tel que $h^\ell L^\ell |A|^\ell < 1$ (puisque sous cette condition $h^\ell L^\ell |A|^\ell$ tend vers 0 quand ℓ tend vers l'infini). Donc l'application qui à $Y \in \mathbb{R}^r$ fait correspondre $F^\ell(Y) \in \mathbb{R}^r$ est contractante, de constante strictement inférieure à 1. \mathbb{R}^r est supposé muni de la norme $\|\cdot\|_\infty$. Dans ce cas l'application qui à Y fait correspondre $F(Y)$ a un point fixe et un seul Y (cf. chapitre 3).

Puisque $Y = F(Y)$ équivaut à

$$y_i = y_i + h \sum_{j=1}^r a_{i,j} f(x + \theta_j h, y_j) \text{ pour } i = 1, \dots, r,$$

ces équations et l'équation $\Phi(x, y, h) = \sum_{i=1}^r c_i f(x + \theta_i h, y_i)$ définissent de façon unique la fonction $\Phi(x, y, h)$.

En effet il est clair que si f est k fois continûment différentiable, l'application qui à (x, y, h, Y) fait correspondre $F(Y)$ est k fois continûment différentiable. On en déduit que l'application qui à (x, y, h) fait correspondre Y , solution de $Y = F(Y)$, est k fois continûment différentiable, et par voie de conséquence la fonction $\Phi(x, y, h)$ est également k fois continûment différentiable.

Démontrons à présent la stabilité.

Si Φ vérifie une condition de Lipschitz, alors le théorème 4.2.4 nous assure de la stabilité de la méthode.

Considérons les deux systèmes de relations

$$\begin{cases} y_i = y + h \sum_{j=1}^r a_{i,j} f(x + \theta_j h, y_j) \text{ pour } i = 1, \dots, r, \\ z_i = z + h \sum_{j=1}^r a_{i,j} f(x + \theta_j h, z_j) \text{ pour } i = 1, \dots, r. \end{cases}$$

On obtient donc

$$|y_i - z_i| \leq |y - z| + h \sum_{j=1}^r |a_{i,j}| L |y_j - z_j|,$$

soit encore

$$|Y - Z| \leq |y - z| e + h L |A| |Y - Z| e + h^* L |A| |Y - Z|,$$

où e est le vecteur de \mathbb{R}^r de composantes toutes égales à 1.

Par récurrence il est simple de voir que l'on obtient

$$|Y - Z| \leq |y - z| \left(\sum_{j=0}^{\ell} (h^*)^j L^j |A|^j \right) e + (h^*)^{\ell+1} L^{\ell+1} |A|^{\ell+1} |Y - Z|.$$

Puisque $h^* L \rho(|A|) < 1$, alors $(h^*)^\ell L^\ell |A|^\ell \rightarrow 0$ quand $\ell \rightarrow \infty$ et la norme donne $(I - h^* L |A|)^{-1}$.

Par conséquent

$$|Y - Z| \leq |y - z| (I - h^* L |A|)^{-1} e,$$

d'où

$$\|Y - Z\|_\infty \leq \|y - z\| \| (I - h^* L |A|)^{-1} \|_\infty.$$

De plus

$$\Phi(x, y, h) - \Phi(x, z, h) = \sum_{i=1}^r c_i (f(x + \theta_i h, y) - f(x + \theta_i h, z)).$$

On a donc

$$|\Phi(x, y, h) - \Phi(x, z, h)| = \sum_{i=1}^r |c_i| L \| (I - h^* L |A|)^{-1} \|_\infty |y - z|.$$

Il suffit de prendre

$$M^* = L \sum_{i=1}^r |c_i| \| (I - h^* L |A|)^{-1} \|_\infty$$

pour que le théorème 4.2.4 soit satisfait. \square

Remarque 4.2.12 Dans le cas d'une méthode de Runge-Kutta explicite, la matrice A est strictement triangulaire inférieure et donc $\rho(|A|) = 0$. Par conséquent le théorème précédent est toujours vérifié.

Il reste à examiner le problème le plus délicat qui est celui de l'ordre d'une méthode de Runge-Kutta.

Afin de condenser l'écriture des relations nous adoptons la notation suivante : le produit $x \cdot y = z$ entre des vecteurs de \mathbb{R}^r donne un vecteur de \mathbb{R}^r ; on a $z_i = x_i y_i$ pour $i = 1, \dots, r$. Le produit $x \cdot x$ est aussi noté $x^{\cdot 2}$. Plus généralement $x^{\cdot k} = (x^{\cdot(k-1)}) \cdot x$.

Nous rappelons que e est le vecteur de \mathbb{R}^r de composantes toutes égales à 1.

Tableau des conditions nécessaires et suffisantes pour que la méthode de Runge-Kutta soit d'ordre p

p	Conditions $A'(p)$
1	$c^T e = 1$
2	$c^T \theta e = c^T A e = \frac{1}{2}$
3	$c^T \theta^2 e = c^T \theta A e = c^T (A e)^{\cdot 2} = \frac{1}{3}$ $c^T A \theta e = c^T A^2 e = \frac{1}{6}$
4	$c^T \theta^3 e = c^T \theta^2 A e = c^T \theta (A e)^{\cdot 2} = c^T (A e)^{\cdot 3} = \frac{1}{8}$ $c^T \theta A \theta e = c^T \theta A^2 e = c^T ((A e)^{\cdot} (A \theta e)) = c^T ((A e)^{\cdot} (A^2 e)) = \frac{1}{8}$ $c^T A \theta^2 e = c^T A \theta A e = c^T A (A e)^{\cdot 2} = \frac{1}{12}$ $c^T A^2 \theta e = c^T A^3 e = \frac{1}{24}$
5	$c^T \theta^4 e = c^T \theta^3 A e = c^T \theta^2 (A e)^{\cdot 2} = c^T \theta (A e)^{\cdot 3} e = c^T (A e)^{\cdot 4} = \frac{1}{5}$ $c^T \theta^2 A \theta e = c^T \theta^2 A^2 e = c^T \theta ((A e)^{\cdot} (A \theta e)) = c^T \theta ((A e)^{\cdot} (A^2 e))$ $= c^T ((A e)^{\cdot 2} (A \theta e)) = c^T ((A e)^{\cdot 2} (A^2 e)) = \frac{1}{10}$ $c^T \theta A \theta^2 e = c^T \theta A \theta A e = c^T \theta A (A e)^{\cdot 2} = c^T ((A e)^{\cdot} (A \theta^2 e)) =$ $c^T ((A e)^{\cdot} (A \theta A e)) = c^T ((A e)^{\cdot} (A (A e)^{\cdot 2})) = \frac{1}{15}$ $c^T A \theta^3 e = c^T A \theta^2 A e = c^T A \theta (A e)^{\cdot 2} = c^T A (A e)^{\cdot 3} = c^T (A \theta e)^{\cdot 2} =$ $c^T ((A \theta e)^{\cdot} (A^2 e)) = c^T (A^2 e)^{\cdot 2} = \frac{1}{20}$ $c^T \theta A^2 \theta e = c^T \theta A^3 e = c^T ((A e)^{\cdot} (A^2 \theta e)) = c^T ((A e)^{\cdot} (A^3 e)) = \frac{1}{30}$ $c^T A \theta A \theta e = c^T A \theta A^2 e = c^T A ((A e)^{\cdot} (A \theta e)) = c^T A ((A e)^{\cdot} (A^2 e)) = \frac{1}{40}$ $c^T A^2 \theta^2 e = c^T A^2 \theta A e = c^T A^2 (A e)^{\cdot 2} = \frac{1}{60}$ $c^T A^3 \theta e = c^T A^4 e = \frac{1}{120}$

On dira que la condition $A(p)$ est vérifiée si les conditions $A'(j)$ sont vérifiées pour tout j tel que $1 \leq j \leq p$.

Théorème 4.2.13 *Une condition nécessaire et suffisante pour qu'une méthode de Runge-Kutta soit d'ordre p , pour toute fonction f suffisamment différentiable, est qu'elle vérifie la condition $A(p)$.*

Démonstration.

On utilise des relations du type (4.23), c'est-à-dire que, pour x et y fixés, on pose

$$k_i(h) = f(x + \theta_i h, y_i),$$

ce qui donne

$$\begin{cases} k_i(h) = f(x + \theta_i h, y + h \sum_{j=1}^r a_{i,j} k_j(h)) \text{ pour } i = 1, \dots, r \\ \Phi(x, y, h) = \sum_{i=1}^r c_i k_i(h). \end{cases}$$

On obtient par conséquent

$$\begin{aligned}
 k_i(0) &= f(x, y), \\
 k'_i(0) &= \theta_i f'_x(x, y) + \sum_{j=1}^r a_{i,j} k_j(0) f'_y(x, y), \\
 k''_i(0) &= \theta_i^2 f''_{x^2}(x, y) + 2 \sum_{j=1}^r \theta_i a_{i,j} k_j(0) f''_{xy}(x, y) \\
 &\quad + \left(\sum_{j=1}^r a_{i,j} k_j(0) \right)^2 f''_{y^2}(x, y) + 2 \sum_{j=1}^r a_{i,j} k'_j(0) f'_y(x, y), \\
 &\text{etc, } \dots
 \end{aligned}$$

ce qui nous donne

$$\begin{aligned}
 \Phi(x, y, 0) &= \sum_{i=1}^r c_i k_i(0) = c^T e f(x, y). \\
 \Phi'_h(x, y, 0) &= \sum_{i=1}^r c_i k'_i(0) \\
 &= \sum_{i=1}^r c_i \theta_i f'_x(x, y) + \sum_{i=1}^r c_i \sum_{j=1}^r a_{i,j} k_j(0) f'_y(x, y) \\
 &= c^T \theta e f'_x(x, y) + c^T A e f(x, y) f'_y(x, y).
 \end{aligned}$$

De la même façon on aurait

$$\begin{aligned}
 \Phi''_{h^2}(x, y, 0) &= c^T \theta^2 e f''_{x^2}(x, y) + 2 c^T \theta A e f''_{xy}(x, y) f(x, y) \\
 &\quad + c^T (A e)^2 f''_{y^2}(x, y) (f(x, y))^2 + 2 c^T A \theta e f'_x(x, y) f'_y(x, y) \\
 &\quad + 2 c^T A^2 e (f'_y(x, y))^2 f(x, y), \\
 &\text{etc, } \dots
 \end{aligned}$$

D'autre part

$$\begin{aligned}
 f^{(0)}(x, y) &= f(x, y) \\
 f^{(1)}(x, y) &= f'_x(x, y) + f(x, y) f'_y(x, y) \\
 f^{(2)}(x, y) &= f''_{x^2}(x, y) + 2 f(x, y) f''_{xy}(x, y) \\
 &\quad + (f(x, y))^2 f''_{y^2}(x, y) + f'_x(x, y) f'_y(x, y) + f(x, y) (f'_y(x, y))^2.
 \end{aligned}$$

La condition $A'(p)$ est obtenue par une fonction $(p-1)$ fois continûment différentiable en utilisant la relation (4.12) :

$$\frac{\partial^{p-1}}{\partial h^{p-1}} \Phi(x, y, h) \big|_{h=0} = \frac{1}{p} f^{(p-1)}(x, y).$$

Il suffit alors d'identifier les coefficients des différentielles de même nature. D'où le tableau proposé. \square

On ne peut que constater l'extrême lourdeur des conditions nécessaires et suffisantes. Leur nombre croît très rapidement avec p . **Pour $p = 5$ il y a 58 conditions en grande majorité non-linéaires.** Dès que p est supérieur ou égal à 3, les conditions nécessaires et suffisantes deviennent inutilisables. On préfère alors les remplacer par des conditions suffisantes plus commodes d'emploi.

On ajoute une condition simplificatrice

$$A e = \theta e \tag{4.25}$$

qui permet de supprimer de nombreuses conditions du tableau. Pour $p = 5$ on passe alors à 17 conditions.

Conditions suffisantes pour $p = 3$

$$Ae = \theta e; c^T e = 1; c^T Ae = \frac{1}{2}; c^T \theta^2 e = \frac{1}{3}; c^T A\theta e = \frac{1}{6}.$$

Conditions suffisantes pour $p = 4$

En plus des conditions suffisantes pour $p = 3$, on a

$$c^T \theta^3 e = \frac{1}{4}; c^T \theta A\theta e = \frac{1}{8}; c^T A\theta^2 e = \frac{1}{12}; c^T A^2 \theta e = \frac{1}{24}.$$

On peut rajouter dans ce cas une autre condition simplificatrice

$$c^T A = c^T (I - \theta). \quad (4.26)$$

Conditions suffisantes pour $p = 4$.

On utilise les deux précédentes conditions simplificatrices (4.25) et (4.26). On peut encore ajouter une autre condition simplificatrice

$$C(2A\theta e - \theta^2 e) = 0,$$

où C est la matrice diagonale qui contient les coefficients c_i sur la diagonale.

Exemples de méthodes de Runge-Kutta

i) Méthodes d'ordre 2

0	0	0
α	α	0
	$1 - \frac{1}{2\alpha}$	$\frac{1}{2\alpha}$

où α est un paramètre non nul.

Pour $\alpha = \frac{1}{2}$ la méthode s'appelle méthode de la tangente améliorée ou méthode d'Euler améliorée.

Pour $\alpha = 1$ elle s'appelle la méthode de Heun.

ii) Méthode d'ordre 4

0	0	0	0	0
$\frac{1}{2}$	$\frac{1}{2}$	0	0	0
$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
1	0	0	1	0
	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

Quelques exemples de méthodes de Runge-Kutta implicites

Ces méthodes sont vivement conseillées dans le cas de problèmes mal conditionnés. On dit qu'un problème est bien conditionné si les méthodes numériques usuelles fournissent une approximation satisfaisante de sa solution pour un coût raisonnable. Les autres problèmes sont dits mal conditionnés.

i) Méthode d'ordre 1

1	1
	1

$$y_{n+1} = y_n + h_n f(x_{n+1}, y_{n+1}),$$

qui est la méthode d'Euler régressive (ou rétrograde) d'ordre un.

ii) θ -schéma

θ	θ
	1

Si $\theta \neq \frac{1}{2}$, son ordre est 1. Si $\theta = \frac{1}{2}$, son ordre est 2; c'est la méthode du point milieu.

iii)

α	α	0
$1 - \alpha$	$1 - 2\alpha$	α
	$\frac{1}{2}$	$\frac{1}{2}$

Si $\alpha = \frac{1}{2} + \frac{1}{2\sqrt{3}}$, l'ordre de cette méthode est égal à 3.

4.2.8 Contrôle du pas

On désire minimiser le coût de la résolution du problème différentiel pour une précision donnée. On veut également détecter les points où la méthode donne une approximation médiocre de la solution du problème différentiel, afin de choisir au voisinage de ces points un pas h_i suffisamment petit qui permettra d'atteindre la précision souhaitée, à moins que l'on constate l'impossibilité d'obtenir cette précision.

La propagation des erreurs est contrôlée au fur et à mesure de la détermination des y_i . On diminue le pas si l'erreur est trop importante ou on l'augmente si la précision souhaitée est supérieure à l'erreur. On suppose qu'on utilise comme méthode principale une méthode d'ordre p . On veut calculer une approximation $\tilde{\varepsilon}_k$ de l'erreur de consistance

$$\varepsilon_k = y(x_{k+1}) - y(x_k) - h_k \Phi(x_k, y(x_k), h_k).$$

On se sert à cette fin d'une méthode auxiliaire d'ordre $p+1$ au moins qui utilise une fonction $\Phi^*(x, y(x), h)$. On notera ε_k^* son erreur de consistance.

$$\varepsilon_k^* = y(x_{k+1}) - y(x_k) - h_k \Phi^*(x_k, y(x_k), h_k).$$

Puisque la méthode est d'ordre $p+1$ au moins, on a

$$\varepsilon_k^* = O(h_k^{p+2}).$$

D'autre part

$$\begin{aligned} \varepsilon_k - \varepsilon_k^* &= h_k (\Phi^*(x_k, y(x_k), h_k) - \Phi(x_k, y(x_k), h_k)) \\ &= h_k (\Phi^*(x_k, y_k, h_k) - \Phi(x_k, y_k, h_k)) \\ &+ h_k \left(\frac{\partial \Phi^*}{\partial y}(x_k, \bar{y}_k, h_k) - \frac{\partial \Phi}{\partial y}(x_k, \bar{y}_k, h_k) \right) (y(x_k) - y_k) \end{aligned}$$

en utilisant un développement au premier ordre dans lequel \bar{y}_k est un point compris entre y_k et $y(x_k)$.

Puisque $y(x_k) - y_k = O(h_k^p)$ pour une méthode stable et d'ordre p , on obtient

$$\begin{aligned} \varepsilon_k &= h_k (\Phi^*(x_k, y_k, h_k) - \Phi(x_k, y_k, h_k)) \\ &+ h_k \left(\frac{\partial \Phi^*}{\partial y}(x_k, \bar{y}_k, 0) - \frac{\partial \Phi}{\partial y}(x_k, \bar{y}_k, 0) \right) (y(x_k) - y_k) + h_k^2 O(h_k^p). \end{aligned}$$

Or $\frac{\partial \Phi^*}{\partial y}(x_k, \bar{y}_k, 0) - \frac{\partial \Phi}{\partial y}(x_k, \bar{y}_k, 0) = \frac{\partial}{\partial y} f(x, \bar{y}_k) - \frac{\partial}{\partial y} f(x, \bar{y}_k) = 0$, puisque les deux méthodes sont au moins d'ordre 1 (ce qui est assuré dès que les méthodes sont consistantes).

Par conséquent on peut écrire que

$$\varepsilon_k = \tilde{\varepsilon}_k + h_k^2 O(h_k^p) \quad (4.27)$$

où $\tilde{\varepsilon}_k = h_k (\Phi^*(x_k, y_k, h_k) - \Phi(x_k, y_k, h_k))$.

La relation (4.27) montre que $\tilde{\varepsilon}_k$ est une bonne approximation de ε_k .

Présentons maintenant différentes techniques de contrôle du pas que nous ne justifierons pas théoriquement.

1ere technique

On se donne un paramètre μ (destiné à tendre vers zéro) ; on calcule à chaque pas $\tilde{\varepsilon}_k$ et une approximation $\tilde{z}_0(x_k)$ de $z_0(x_k)$ qui est la solution du problème différentiel

$$\begin{cases} z_0'(x) = \frac{\partial}{\partial y} f(x, y(x)) z_0(x) \\ z_0(a) = 1. \end{cases}$$

On choisit le pas h_k le plus grand possible tel que

$$\frac{|\tilde{\varepsilon}_k|}{\tilde{z}_0(x_k)} \leq \mu$$

et

$$h_k \leq 10 (\mu \tilde{z}_0(x_k))^{\frac{1}{p+1}}.$$

La deuxième contrainte assure que $h_k^2 O(h_k^p)$ est négligeable dans (4.27) ; on aura alors $y(x_k) - y_k = O(\mu^{1-\frac{1}{p+1}})$.

2eme technique

On choisit h_k le plus grand possible tel que

$$\frac{|\tilde{\varepsilon}_k|}{h_k \tilde{z}_0(x_k)} \leq \nu$$

et

$$h_k \leq 10 \nu \tilde{z}_0(x_k).$$

L'erreur sera alors en $O(\nu)$.

Choix pratiques dans les deux techniques

i) Le choix le plus courant est $\tilde{z}_0 \equiv 1$.

ii) $\tilde{z}_0(x_n) = |y_n|$. On rend donc constant $|\frac{\varepsilon_n}{y_n}|$ (à condition que $y_n \neq 0$).

Le calcul de $\tilde{\varepsilon}_k$ exige l'utilisation de deux méthodes de Runge-Kutta. Afin d'en diminuer le coût on choisit généralement des méthodes emboîtées.

Définition 4.2.14 Un couple $RK_{p,p'}$ de méthodes de Runge-Kutta emboîtées est formé à partir du tableau d'une méthode d'ordre p complété par une ligne et une colonne supplémentaire afin d'obtenir le tableau d'une méthode d'ordre p' .

θ_1	a_{11}	a_{12}	\cdots	a_{1r}	0
θ_2	a_{21}	a_{22}	\cdots	a_{2r}	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
θ_r	a_{r1}	a_{r2}	\cdots	a_{rr}	0
1	c_1	c_2	\cdots	c_r	0
	c'_1	c'_2	\cdots	c'_r	c'_{r+1}

Les termes ajoutés sont ceux extérieurs au domaine délimité par les doubles traits.

La méthode d'ordre p est utilisée pour le calcul de la solution approchée. La méthode d'ordre p' sert à estimer l'erreur de consistance pour le contrôle du pas.

On peut constater que l'utilisation de la méthode $RK_{p,p'}$ n'exige pas plus d'évaluation de la fonction f que la méthode de Runge-Kutta d'ordre p . En effet $f(x_{k+1}, y_{k+1})$ servira pour le pas suivant.

Donnons quelques exemples de méthodes emboîtées.

Méthode RK_{12}

0	0	0
1	1	0
	$\frac{1}{2}$	$\frac{1}{2}$

Méthode RK_{24}

0	0	0	0	0
$\frac{1}{2}$	$\frac{1}{2}$	0	0	0
$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
1	0	0	1	0
	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

Méthode RK_{34}

0	0	0	0	0	0
$\frac{2}{7}$	$\frac{2}{7}$	0	0	0	0
$\frac{4}{7}$	$-\frac{8}{35}$	$\frac{4}{5}$	0	0	0
$\frac{6}{7}$	$\frac{29}{42}$	$-\frac{2}{3}$	$\frac{5}{6}$	0	0
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{5}{12}$	$\frac{1}{4}$	0
	$\frac{11}{96}$	$\frac{7}{24}$	$\frac{35}{96}$	$\frac{7}{48}$	$\frac{1}{12}$

4.3 Les méthodes à pas liés

4.3.1 Introduction

Les méthodes de Runge-Kutta sont très utilisées dans la pratique. On dispose de méthodes ayant un ordre élevé. Ces méthodes peuvent être mises en oeuvre seules, car elles ne nécessitent que la connaissance de y_0 . Le contrôle de l'erreur et donc du pas est relativement simple.

Par contre ce sont des méthodes qui, à ordre identique, sont plus coûteuses que les méthodes à pas liés dont nous allons étudier les propriétés.

Leur forme générale est la suivante :

$$\sum_{j=0}^{k_n} \alpha_{n+k_n-1,j,k_n} y_{n+j} = h_{n+k_n-1} \sum_{j=0}^{k_n} \beta_{n+k_n-1,j,k_n} f_{n+j} \text{ avec } \alpha_{k_n} \neq 0, \quad (4.28)$$

où $x_{i+1} = x_i + h_i$, $\forall i$, $f_i = f(x_i, y_i)$ et y_0 est la condition initiale donnée.

La formule (4.28) est destinée à nous fournir y_{n+k_n} qui dépend de $y_n, y_{n+1}, \dots, y_{n+k_n-1}$ et de h_{n+k_n-1} . On obtiendra directement y_{n+k_n} si $\beta_{n+k_n-1,k_n,k_n} = 0$. Dans ce cas la méthode (4.28) sera dite explicite. Si $\beta_{n+k_n-1,k_n,k_n} \neq 0$, on obtiendra y_{n+k_n} comme solution d'une équation non linéaire en général.

$$\begin{aligned} y_{n+k_n} &= h_{n+k_n-1} \frac{\beta_{n+k_n-1,k_n,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} f(x_{n+k_n}, y_{n+k_n}) \\ &+ h_{n+k_n-1} \sum_{j=0}^{k_n-1} \frac{\beta_{n+k_n-1,j,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} f_{n+j} - \sum_{j=0}^{k_n-1} \frac{\alpha_{n+k_n-1,j,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} y_{n+j}. \end{aligned} \quad (4.29)$$

Dans ce cas la méthode sera dite implicite.

On sait parfaitement (cf. chapitre 3) qu'une équation (4.29) ne nous fournira y_{n+k_n} que sous certaines conditions.

Théorème 4.3.1 Si $\beta_{n+k_n-1,k_n,k_n} \neq 0$, alors l'équation (4.29) a une solution unique si

$$h_{n+k_n-1} < \frac{1}{L} \frac{|\alpha_{n+k_n-1,k_n,k_n}|}{|\beta_{n+k_n-1,k_n,k_n}|},$$

où L est la constante de Lipschitz de f .

Démonstration.

y_{n+k_n} sera déterminée par la méthode des approximations successives appliquée à (4.29) :

$$\begin{aligned} y_{n+k_n}^{(p+1)} &= h_{n+k_n-1} \frac{\beta_{n+k_n-1,k_n,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} f(x_{n+k_n}, y_{n+k_n}^{(p)}) \\ &+ h_{n+k_n-1} \sum_{j=0}^{k_n-1} \frac{\beta_{n+k_n-1,j,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} f_{n+j} - \sum_{j=0}^{k_n-1} \frac{\alpha_{n+k_n-1,j,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} y_{n+j}. \end{aligned}$$

On forme $|y_{n+k_n}^{(p+2)} - y_{n+k_n}^{(p+1)}|$ et on utilise le fait que f vérifie une condition de Lipschitz.

$$\begin{aligned} |y_{n+k_n}^{(p+2)} - y_{n+k_n}^{(p+1)}| &= |h_{n+k_n-1} \frac{\beta_{n+k_n-1,k_n,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} \times \\ &\quad (f(x_{n+k_n}, y_{n+k_n}^{(p+1)}) - f(x_{n+k_n}, y_{n+k_n}^{(p)}))| \\ &\leq h_{n+k_n-1} \frac{\beta_{n+k_n-1,k_n,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} L |y_{n+k_n}^{(p+1)} - y_{n+k_n}^{(p)}|. \end{aligned}$$

Si la fonction $h_{n+k_n-1} \frac{\beta_{n+k_n-1,k_n,k_n}}{\alpha_{n+k_n-1,k_n,k_n}} f(x_{n+k_n}, y)$ est une contraction par rapport à la seconde variable, le théorème du point fixe nous indique que les itérations précédentes convergent vers l'unique solution de (4.29) qui est y_{n+k_n} . Pour avoir une contraction, il suffit que $h_{n+k_n-1} L \frac{|\beta_{n+k_n-1,k_n,k_n}|}{|\alpha_{n+k_n-1,k_n,k_n}|} < 1$. \square

Avant de passer à la détermination pratique des α_{n+k_n-1,j,k_n} et β_{n+k_n-1,j,k_n} , nous allons passer en revue un certain nombre de conditions théoriques que doivent satisfaire

les méthodes (4.28). Ces conditions concernent à nouveau la consistance, la stabilité, la convergence. Nous examinerons également la notion d'ordre d'une méthode. Ces conditions introduisent un certain nombre de relations dans lesquelles interviennent les coefficients α_{n+k_n-1,j,k_n} et β_{n+k_n-1,j,k_n} . La méthode (4.28) est liée à la théorie des équations aux différences.

4.3.2 Les équations aux différences

Les équations aux différences qui interviennent dans les méthodes (4.28) sont linéaires, homogènes. Si $h_{n+k_n-1} = h, \forall n$, elles sont aussi à coefficients constants.

Soit y une fonction telle que $y : \mathbb{N} \rightarrow \mathbb{R}$. On appelle équation aux différences d'ordre k une relation de la forme

$$F(y(n), y(n+1), \dots, y(n+k); n) = 0,$$

où F est une fonction donnée. Si F est linéaire par rapport à $y(n), y(n+1), \dots, y(n+k)$, l'équation aux différences est dite linéaire. Si l'équation linéaire est de la forme $\sum_{i=0}^k a_i y(n+i) = g(n)$, où les a_i sont des constantes réelles indépendantes de n , on dira que l'équation est à coefficients constants; elle sera dite homogène si $g(n) = 0, \forall n$.

La théorie des équations aux différences linéaires, homogènes, à coefficients constants et d'ordre k est très semblable à celle des équations différentielles linéaires, homogènes, à coefficients constants et d'ordre k .

Théorème 4.3.2 Si y_1, y_2, \dots, y_p sont solutions de

$$\sum_{i=0}^k a_i y(n+i) = 0, \quad (4.30)$$

alors toute combinaison linéaire de y_1, y_2, \dots, y_p est également solution de (4.30).

Démonstration.

Nous avons $\sum_{i=0}^k a_i y_j(n+i) = 0$ pour $j = 1, \dots, p$.

$$\forall c_1, c_2, \dots, c_p, \sum_{i=0}^k a_i \sum_{j=1}^p c_j y_j(n+i) = \sum_{j=1}^p c_j \sum_{i=0}^k a_i y_j(n+i) = 0,$$

ce qui montre bien que $\sum_{j=1}^p c_j y_j$ est solution de (4.30). \square

Théorème 4.3.3 Si y_1, \dots, y_k sont solution de (4.30) et si le déterminant de la matrice D est non nul, où

$$D = \begin{pmatrix} y_1(0) & y_1(1) & \cdots & y_1(k-1) \\ y_2(0) & y_2(1) & \cdots & y_2(k-1) \\ \vdots & \vdots & \cdots & \vdots \\ y_k(0) & y_k(1) & \cdots & y_k(k-1) \end{pmatrix}, \quad (4.31)$$

alors la solution générale de (4.30) s'écrit $y = \sum_{j=1}^k c_j y_j$.

Démonstration.

Il est clair que, grâce à la relation (4.30), chaque solution y_j est complètement déterminée dès que sont données les valeurs initiales $y_j(0), \dots, y_j(k-1)$ qui peuvent toujours être choisies telles que $\det D$ soit non nul. Le théorème 4.3.2 montre que $y = \sum_{j=1}^k c_j y_j$ est solution de (4.30).

Nous allons montrer que toute solution de (4.30) est de cette forme. Si Φ est une solution de (4.30), elle est aussi complètement déterminée par la connaissance de $\Phi(0), \dots, \Phi(k-1)$.

Or, parmi l'ensemble des fonctions $y = \sum_{j=1}^k c_j y_j$, il en existe une unique qui a les mêmes valeurs initiales. En effet cela revient à trouver la solution du système linéaire $Dc = \varphi$, où c est un vecteur de \mathbb{R}^k de composantes c_1, \dots, c_k et φ un vecteur de \mathbb{R}^k de composantes $\Phi(0), \dots, \Phi(k-1)$. Puisque D est inversible, la solution c est unique lorsque φ est connu. Par conséquent Φ est identique à ce y , ce qui démontre complètement le théorème. \square

Cherchons à présent la forme de la solution générale de (4.30). Nous avons le théorème suivant :

Théorème 4.3.4 *Soit l'équation aux différences*

$$\sum_{i=0}^k a_i y(n+i) = 0, \quad (4.32)$$

où les a_i sont des constantes réelles indépendantes de n avec $a_0 \neq 0$ et $a_k \neq 0$.

La solution générale de cette équation est donnée par

$$y(n) = \sum_{j=1}^p A_j(n) \lambda_j^n + \sum_{j=p+1}^q (B_j(n) \cos b_j n + C_j(n) \sin b_j n) \lambda_j^n,$$

où les quantités

$$\begin{array}{ll} \lambda_j & \text{pour } j = 1, \dots, p \\ \text{et } \lambda_j (\cos b_j + \sin b_j) & \text{pour } j = p+1, \dots, q \quad (i = \sqrt{-1}) \end{array}$$

sont les racines de l'équation caractéristique de (4.32) qui est

$$\sum_{i=0}^k a_i \lambda^i = 0, \quad (4.33)$$

et où A_j , B_j et C_j sont des polynômes en n de degré égal à la multiplicité de la racine correspondante de (4.33) moins un.

Notons d_j le degré de A_j plus un, pour $j = 1, \dots, p$, ainsi que le plus grand degré de B_j et de C_j plus un pour $j = p+1, \dots, q$. Dans ce cas on a

$$\sum_{j=1}^p d_j + 2 \sum_{j=p+1}^q d_j = k.$$

Démonstration.

Nous allons chercher une solution de la forme

$$y(n) = \lambda^n \text{ avec } \lambda \neq 0.$$

L'équation (4.32) nous donne alors

$$\lambda^n \sum_{j=0}^k a_j \lambda^j = 0$$

et donc $\sum_{j=0}^k a_j \lambda^j = 0$ puisque $\lambda \neq 0$.

Cette dernière équation qui est celle donnée par (4.33) est appelée équation caractéristique de l'équation aux différences (4.32). Elle nous montre que, pour que λ^n soit solution de (4.32), il faut que λ soit racine de (4.33).

Commençons par examiner le cas le plus simple, celui où toutes les racines sont simples. Soient $\lambda_1, \dots, \lambda_k$ ces racines. Alors D , donnée par (4.31), a un déterminant qui est non nul, puisque c'est un déterminant de Vandermonde.

Le théorème 4.3.3 nous fournit dans ce cas la solution générale de (4.32) sous la forme

$$y(n) = \sum_{j=1}^k c_j \lambda_j^n. \quad (4.34)$$

L'équation aux différences (4.32) est à coefficients réels, donc aussi son équation caractéristique (4.33). Par conséquent, si une racine λ_j est imaginaire, $\bar{\lambda}_j$ est aussi racine de (4.33). Une telle racine peut s'écrire $\rho(\cos \omega \pm i \sin \omega)$. Dans la solution générale (4.34) la combinaison des deux termes λ_j^n et $\bar{\lambda}_j^n$ pourra être remplacée par la combinaison des deux termes $\rho^n \cos \omega n$ et $\rho^n \sin \omega n$.

Abordons maintenant le cas où l'équation caractéristique admet des racines multiples. Il n'existe plus k solutions de la forme λ^n telles que $\det D \neq 0$. Si $\hat{\lambda}$ est une racine de multiplicité r de l'équation (4.33), on a bien sûr

$$P(\hat{\lambda}) = \sum_{j=0}^k a_j \hat{\lambda}^{n+j} = 0.$$

Donc $\hat{\lambda}^n$ est solution de (4.32). Mais on a aussi

$$P_1(\hat{\lambda}) = \hat{\lambda} P'(\hat{\lambda}) = \sum_{j=0}^k a_j (n+j) \hat{\lambda}^{n+j} = 0,$$

puisque $P'(\hat{\lambda}) = 0$, et donc $n \hat{\lambda}^n$ est solution de (4.32). De même

$$P_\ell(\hat{\lambda}) = \hat{\lambda} P'_{\ell-1}(\hat{\lambda}) = \sum_{j=0}^k a_j (n+j)^\ell \hat{\lambda}^{n+j} = 0,$$

pour $\ell = 1, 2, \dots, r-1$, puisque $\hat{\lambda} P'_{\ell-1}(\hat{\lambda})$ est une combinaison linéaire des $P'(\hat{\lambda})$, $P''(\hat{\lambda})$, \dots , $P^{(\ell-1)}(\hat{\lambda})$. Par conséquent $n^\ell \lambda^n$ est solution de (4.32) pour $\ell = 0, \dots, r-1$. Au total nous obtiendrons bien k solutions, car la somme des multiplicités des racines de (4.33) est naturellement égale à k .

Si $\lambda_1, \dots, \lambda_p$ sont les racines distinctes de l'équation caractéristique (4.33) et si r_1, \dots, r_p sont leurs multiplicités respectives, il suffit à présent de montrer que $\det D \neq 0$ avec

$$D = \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{k-1} \\ 0 & \lambda_1 & \cdots & (k-1) \lambda_1^{k-1} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \lambda_1 & \cdots & (k-1)^{r_1-1} \lambda_1^{k-1} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \lambda_p & \cdots & \lambda_p^{k-1} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \lambda_p & \cdots & (k-1)^{r_p-1} \lambda_p^{k-1} \end{pmatrix}$$

□

Maintenant si nous sommes en présence d'une équation aux différences avec second membre

$$\sum_{i=0}^k a_i y(n+i) = g(n), \quad (4.35)$$

la solution générale d'une telle équation est la somme d'une solution particulière $y^*(n)$ et de la solution générale de l'équation homogène (4.32), puisque, si on prend

$$y(n) = y^*(n) + \phi(n),$$

(4.35) entraîne que

$$\sum_{i=0}^k a_i (y^*(n+i) + \phi(n+i)) = g(n).$$

En tenant compte du fait que $y^*(n)$ est solution de (4.35), il reste

$$\sum_{i=0}^k a_i \phi(n+i) = 0$$

et donc ϕ est solution générale de l'équation homogène correspondante.

Définition 4.3.5 *On dit qu'une équation aux différences linéaire à coefficients constants est stable si, lorsque n tend vers $+\infty$, toute solution de l'équation homogène reste bornée.*

Théorème 4.3.6 *Une équation aux différences linéaire à coefficients constants est stable si et seulement si son équation caractéristique satisfait la condition suivante (condition de stabilité) : l'équation caractéristique a toutes ses racines de module plus petit que 1 et les racines de module 1 sont simples.*

Démonstration.

Le théorème 4.3.4 nous montre que la solution de l'équation homogène est une combinaison linéaire des solutions $u_{ij}(n) = n^i \lambda_j^n$ pour $i = 0, \dots, r_j - 1$ et $j = 1, \dots, p$, où r_j est la multiplicité de la racine λ_j de l'équation caractéristique qui a p racines distinctes $\lambda_1, \dots, \lambda_p$. Si l'équation caractéristique a la propriété de stabilité, alors les $u_{ij}(n)$ restent bornés lorsque n tend vers $+\infty$, donc aussi toute solution de l'équation homogène. Réciproquement, si toute solution de l'équation homogène reste bornée, la propriété est vraie pour tous les $u_{ij}(n)$. Pour qu'il en soit ainsi, il faut

- ou bien que $|\lambda_j| < 1$,
- ou bien que $|\lambda_j| = 1$ et $r_j = 1$ (sinon pour $r_j > 1$, on a $|u_{ij}(n)| = n^i$, pour $i = 1, \dots, r_j - 1$, qui tendrait vers l'infini).

D'où le résultat. \square

4.3.3 La consistance

Les notions de consistance, de stabilité théorique, de convergence et d'ordre qui ont été introduites lors de l'étude des méthodes à un pas, peuvent être à nouveau définies et utilisées dans le cas de méthodes à pas liés.

Dans ce qui suit on considère le cas où

$$\begin{aligned} k_n &= k = \text{constante}, \\ h_{n+k_n-1} &= h = \text{constante}. \end{aligned}$$

La relation (4.28) s'écrit donc

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i f(x_{n+i}, y_{n+i}), \quad \forall n \in \mathbb{N}.$$

Définition 4.3.7 Une méthode à pas liés est dite consistante avec l'équation différentielle si

$$\lim_{h \rightarrow 0} \max_n \left| \frac{1}{h} \sum_{i=0}^k \alpha_i y(x_{n+i}) - \sum_{i=0}^k \beta_i f(x_{n+i}, y(x_{n+i})) \right| = 0, \quad (4.36)$$

pour toute solution $y \in C_\infty([a, b])$ de l'équation différentielle.

Nous allons démontrer un résultat pratique pour vérifier la consistance. Nous associons à la méthode à pas liés les deux polynômes suivants

$$\begin{aligned} \alpha(t) &= \sum_{i=0}^k \alpha_i t^i, \\ \beta(t) &= \sum_{i=0}^k \beta_i t^i. \end{aligned}$$

Théorème 4.3.8 Une condition nécessaire et suffisante pour que la méthode à pas liés soit consistante avec l'équation différentielle est que

$$\begin{cases} \alpha(1) &= 0, \\ \alpha'(1) &= \beta(1). \end{cases} \quad (4.37)$$

Démonstration.

Montrons que la condition est nécessaire.

Si la méthode est consistante, les conditions sont satisfaites pour n'importe quelle équation différentielle. En particulier, si on prend $y'(x) = 0$ avec $y(0) = 1$, dont la solution est $y(x) = 1$, la définition de la consistance s'écrit

$$\lim_{h \rightarrow 0} \frac{1}{h} \sum_{i=0}^k \alpha_i = 0.$$

Il est donc nécessaire d'avoir $\sum_{i=0}^k \alpha_i = \alpha(1) = 0$.

Si on prend maintenant l'équation différentielle $y'(x) = 1$ avec $y(0) = 0$, dont la solution est $y(x) = x$, la définition de la consistance s'écrit

$$\sum_{i=0}^k i \alpha_i - \sum_{i=0}^k \beta_i = \alpha'(1) - \beta(1) = 0.$$

D'où la nécessité d'avoir $\alpha'(1) = \beta(1)$.

La condition est aussi suffisante. Si les relations (4.37) sont satisfaites, alors la méthode est consistante pour toute équation différentielle. On peut écrire les développements de Taylor suivants :

$$\begin{aligned} y(x_{n+i}) &= y(x_n) + i h y'(x_n) + O(h^2), \\ y'(x_{n+i}) &= y'(x_n) + O(h). \end{aligned}$$

La condition de consistance (4.36) devient

$$\lim_{h \rightarrow 0} \max_n \left| \frac{y(x_n)}{h} \sum_{i=0}^k \alpha_i + y'(x_n) \sum_{i=0}^k i \alpha_i + O(h) - y'(x_n) \sum_{i=0}^k \beta_i + O(h) \right| = 0,$$

c'est-à-dire

$$\lim_{h \rightarrow 0} \max_n \left| \frac{y(x_n)}{h} \sum_{i=0}^k \alpha_i + y'(x_n) \sum_{i=0}^k (i \alpha_i - \beta_i) + O(h) \right| = 0.$$

Cette dernière relation est bien vérifiée lorsque les relations (4.37) sont satisfaites. \square

4.3.4 La stabilité théorique

Définition 4.3.9 Soient y_n et z_n , pour $n = 0, \dots, N$, les solutions respectives de

$$\begin{cases} \sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i f(x_{n+i}, y_{n+i}) \\ y_0, \dots, y_{k-1} \text{ sont des réels quelconques donnés,} \end{cases}$$

et de

$$\begin{cases} \sum_{i=0}^k \alpha_i z_{n+i} = h \left(\sum_{i=0}^k \beta_i f(x_{n+i}, z_{n+i}) + \tilde{\varepsilon}_n \right) \\ z_0, \dots, z_{k-1} \text{ sont des réels quelconques donnés,} \end{cases}$$

avec $x_0 = a$ et $x_{n+1} = x_n + h$.

La méthode à pas liés est dite stable s'il existe deux constantes M_1 et M_2 , indépendantes de h , telles que

$$\max_n |y_n - z_n| \leq M_1 \max_{0 \leq i \leq k-1} |y_i - z_i| + M_2 \max_n |\tilde{\varepsilon}_n|.$$

On peut démontrer le résultat suivant :

Théorème 4.3.10 Une condition nécessaire et suffisante pour qu'une méthode à pas liés soit stable est que le polynôme α soit stable, c'est-à-dire que toutes les racines de ce polynôme soient de module inférieur ou égal à un et que les racines de module un soient des racines simples.

Lorsque la méthode est consistante, 1 est racine de α , puisque $\alpha(1) = 0$. 1 doit être simple pour que la méthode soit stable.

4.3.5 La convergence

Définition 4.3.11 La méthode à pas liés est dite convergente si

$$\lim_{h \rightarrow 0} \max_n |y_n - y(x_n)| = 0,$$

lorsque $\lim_{h \rightarrow 0} y_i = y_0$ pour $i = 0, \dots, k-1$.

La solution approchée doit donc converger vers la solution de l'équation différentielle en tout point de l'intervalle d'intégration et cela lorsque les k valeurs initiales y_0, \dots, y_{k-1} tendent vers la condition initiale de l'équation différentielle (4.1).

Nous avons le résultat suivant :

Théorème 4.3.12 Une condition nécessaire et suffisante pour qu'une méthode à pas liés soit convergente est qu'elle soit stable et consistante.

Démonstration.

Commençons par montrer la condition suffisante.

Posons

$$\sum_{i=0}^k \alpha_i y_{n+i} - h \sum_{i=0}^k \beta_i f(x_{n+i}, y_{n+i}) = h\varepsilon_n,$$

avec $y(a) = y_0$. Si la méthode est consistante, la définition de cette notion donne

$$\lim_{h \rightarrow 0} \max_n |\varepsilon_n| = 0.$$

Si la méthode est stable, il existe, d'après la définition de cette notion, deux constantes M_1 et M_2 indépendantes de h telles que

$$\max_n |y_n - y(x_n)| \leq M_1 \max_{0 \leq i \leq k-1} |y_i - y(x_i)| + M_2 \max_n |\varepsilon_n|.$$

Le second membre tend donc vers 0 quand h tend vers 0 à cause de l'hypothèse sur la limite de y_i . Ceci démontre la convergence.

La condition est également nécessaire. Tout d'abord la convergence entraîne la stabilité. En effet si on prend l'équation différentielle $y'(x) = 0$, avec $y(0) = 0$, qui a pour solution $y(x) = 0$, la méthode à pas liés donne

$$\sum_{i=0}^k \alpha_i y_{n+i} = 0.$$

Comme la méthode est convergente, on a $\lim_{h \rightarrow 0} y_n = 0$ et par conséquent d'après le théorème 4.3.6 le polynôme $\alpha(t)$ est stable.

Ensuite la convergence entraîne la consistance. Si on prend le problème de Cauchy $y'(x) = 0$ avec $y(0) = 1$ dont la solution est $y(x) = 1$, on a

$$\sum_{i=0}^k \alpha_i y_{n+i} = 0$$

avec $\lim_{h \rightarrow 0} y_n = 1$. Donc $\sum_{i=0}^k \alpha_i = \alpha(1) = 0$.

Si on prend maintenant le problème de Cauchy $y'(x) = 1$ avec $y(0) = 0$ dont la solution est $y(x) = x$, on a

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i = h\beta(1). \quad (4.38)$$

Comme la méthode converge, on a

$$\lim_{h \rightarrow 0} |y_n - y(x_n)| = \lim_{h \rightarrow 0} |y_n - nh| = 0.$$

Donc $y_i = ih + h\varepsilon_i$ avec $\lim_{h \rightarrow 0} \varepsilon_i = 0$, et par conséquent (4.38) s'écrit encore

$$h \left(\sum_{i=0}^k i \alpha_i - \beta(1) \right) = h \sum_{i=0}^k \alpha_i \varepsilon_i,$$

ce qui entraîne que $\alpha'(1) = \beta(1)$. \square

4.3.6 L'ordre

Définition 4.3.13 La méthode à pas liés est dite d'ordre p si

$$\max_n \left| \frac{1}{h} \sum_{i=0}^k \alpha_i y(x_{n+i}) - \sum_{i=0}^k \beta_i f(x_{n+i}, y(x_{n+i})) \right| = O(h^p).$$

Le théorème suivant donne le comportement de l'erreur en fonction de l'ordre de la méthode.

Théorème 4.3.14 Si la méthode est stable et d'ordre p et si $\max_{0 \leq i \leq k-1} |y_i - y(x_i)| = O(h^p)$, alors

$$\max_n |y_n - y(x_n)| = O(h^p).$$

Démonstration.

Posons

$$\varepsilon_n = \frac{1}{h} \sum_{i=0}^k \alpha_i y(x_{n+i}) - \sum_{i=0}^k \beta_i y'(x_{n+i}).$$

Si la méthode est d'ordre p , la définition 4.3.13 entraîne que

$$\max_n |\varepsilon_n| = O(h^p).$$

Comme la méthode est stable, il existe deux constantes M_1 et M_2 indépendantes de h telles que

$$\max_n |y_n - y(x_n)| \leq M_1 \max_{0 \leq i \leq k-1} |y_i - y(x_i)| + M_2 \max_n |\varepsilon_n|.$$

Donc si $\max_{0 \leq i \leq k-1} |y_i - y(x_i)| = O(h^p)$, le théorème est démontré. \square

Ce théorème est important, car il nous montre qu'il suffit que les quantités y_i , pour $i = 1, \dots, k-1$, soient calculées en utilisant une méthode d'ordre p au moins pour avoir

$$\max_n |y_n - y(x_n)| = O(h^p).$$

Remarque 4.3.15 Si y est suffisamment dérivable, on peut écrire le développement de Taylor au voisinage de $h = 0$ de la relation donnée dans la définition 4.3.13. Ceci nous fournit

$$\frac{1}{h} \sum_{i=0}^k \alpha_i y(x_{n+i}) - \sum_{i=0}^k \beta_i y'(x_{n+i}) = \sum_{j=0}^p C_j h^{j-1} y^{(j)}(x_n) + O(h^p),$$

avec

$$\begin{cases} C_0 &= \sum_{i=0}^k \alpha_i, \\ C_1 &= \sum_{i=0}^k (i\alpha_i - \beta_i), \\ \dots & \dots \\ C_q &= \frac{1}{q!} \sum_{i=0}^k i^q \alpha_i - \frac{1}{(q-1)!} \sum_{i=0}^k i^{q-1} \beta_i, \quad \text{pour } q = 2, \dots, p. \end{cases}$$

La méthode est d'ordre p si $C_0 = C_1 = \dots = C_p = 0$.

On remarque que $C_0 = 0$ donne $\alpha(1) = 0$ et que $C_1 = 0$ donne $\alpha'(1) = \beta(1)$.

Par conséquent on a le résultat suivant :

Théorème 4.3.16 Une condition nécessaire et suffisante pour que la méthode soit consistante est qu'elle soit d'ordre un au moins.

4.3.7 Les méthodes d'Adams

Nous allons maintenant étudier l'obtention de méthodes à pas liés qui vérifient les notions théoriques de consistance, de stabilité théorique et par conséquent de convergence, et qui aient de plus un ordre p le plus élevé possible. Par conséquent, lorsque k sera fixé, les polynômes α et β devront vérifier les propriétés relatives à la stabilité de α et à l'ordre p le plus élevé possible ($p \geq 1$).

En remarquant que, si y est un polynôme de degré p , son développement de Taylor est fini et se termine avec les termes en h^p . Par conséquent il faut choisir les polynômes α et β de telle sorte que α soit stable et que la méthode à pas liés soit exacte pour des polynômes de degré le plus élevé possible. Les méthodes d'Adams sont construites selon ce principe. On commence par écrire

$$y(x_q) - y(x_p) = \int_{x_j}^{x_q} f(t, y(t)) dt \quad \text{avec } x_q \geq x_j \quad \text{et } x_j, x_q \in [a, b].$$

Naturellement on ne sait pas calculer l'intégrale figurant dans le membre de droite, puisque y est une fonction inconnue. On va donc remplacer f par son polynôme d'interpolation qui

prend les valeurs f_{n+i} en x_{n+i} pour $i = 0, \dots, k$. Donc l'intégrale est calculée par une formule de quadrature de type interpolation. On obtient

$$y_q - y_j = h \sum_{i=0}^k \beta_i f_{n+i} \quad \text{avec } j \leq q.$$

Puisque $\alpha(t) = t^q - t^j$, on remarque que $\alpha(1) = 0$ et que ce polynôme est stable.

Pour avoir la consistance, les coefficients β_i doivent vérifier $\alpha'(1) = \beta(1)$, c'est-à-dire $q - j = \sum_{i=0}^k \beta_i$.

Les différentes méthodes que l'on peut alors obtenir se différencient par la position de x_j et de x_q par rapport aux points x_n, \dots, x_{n+k} .

- Si $x_j = x_{n+k}$ et $x_q = x_{n+k+1}$, on a les méthodes d'Adams-Bashforth.
- Si $x_j = x_{n+k-1}$ et $x_q = x_{n+k}$, on a les méthodes d'Adams-Moulton.
- Si $x_j = x_{n+k-1}$ et $x_q = x_{n+k+1}$, on a les méthodes de Nyström.
- Si $x_j = x_{n+k-2}$ et $x_q = x_{n+k}$, on a les méthodes de Milne-Simpson.

Il reste à déterminer les coefficients β_i . Ce calcul peut s'effectuer de plusieurs façons.

- On peut reprendre les relations données dans la remarque 4.3.15.
- On peut écrire que la méthode est exacte pour des polynômes de degré le plus élevé possible. On obtient un système linéaire dont les inconnues sont les β_i .

Exemple :

Soit la formule

$$y_{n+2} - y_{n+1} = h\beta_0 f_n + h\beta_1 f_{n+1}. \quad (4.39)$$

On commence par prendre $y'(x) = 0$ avec $y(0) = 1$, dont la solution est $y(x) = 1$. L'équation (4.39) donne $0 = 0$.

Si on prend $y'(x) = 1$ avec $y(0) = 0$, dont la solution est $y(x) = x$, on a

$$x_{n+2} - x_{n+1} = h(\beta_0 + \beta_1) = h.$$

Enfin si on prend $y'(x) = x$ avec $y(0) = 0$, dont la solution est $y(x) = x^2/2$, on a

$$\frac{x_{n+2}^2}{2} - \frac{x_{n+1}^2}{2} = h(\beta_0 x_n + \beta_1 x_{n+1}).$$

On peut prendre $x_n = 0$; alors la relation précédente devient $\beta_1 = 3/2$ et par conséquent $\beta_0 = -1/2$.

- Les coefficients β_i peuvent être calculés d'une autre façon en remplaçant $f(t, y(t))$ par son polynôme d'interpolation de Newton aux abscisses x_n, \dots, x_{n+k} , puis à l'intégrer entre x_j et x_q . La détermination effective des β_i s'effectue ensuite de façon récursive en utilisant les fonctions génératrices.

Cette méthode sera illustrée par la détermination de quelques formules d'Adams-Bashforth.

Comme les abscisses d'interpolation sont équidistantes, nous utiliserons le polynôme de Newton (cf. la relation (5.21) de la section 5.5 du chapitre 5). Dans cette relation on remplace n par k , 0 par n et on effectue le changement de variable $s = (x - x_{n+k})/h$. En utilisant l'opérateur ∇^i défini par

$$\begin{aligned} \nabla^0 f_{n+k} &= f_{n+k}, \\ \nabla^i f_{n+k} &= \nabla^{i-1} f_{n+k} - \nabla^{i-1} f_{n+k-1} \quad \forall i \in \mathbb{N}, i \geq 1, \end{aligned}$$

on obtient

$$P_k(s) = \sum_{i=0}^k (-1)^i \binom{-s}{i} \nabla^i f_{n+k},$$

avec la convention que

$$\binom{a}{i} = \frac{a(a-1)\dots(a-i+1)}{i!}.$$

On a $P_k(-i) = f_i$ pour $i = 0, \dots, k$.

Les formules d'Adams-Bashforth donnent donc

$$\begin{aligned} y_{n+k+1} - y_{n+k} &= \int_{x_{n+k}}^{x_{n+k+1}} P_k(s) ds \\ &= h \sum_{i=0}^k \left(\int_0^1 (-1)^i \binom{-s}{i} ds \right) \nabla^i f_{n+k} \\ &= h \sum_{i=0}^k \gamma_i \nabla^i f_{n+k}, \end{aligned}$$

avec $\gamma_i = (-1)^i \int_0^1 \binom{-s}{i} ds$.

On appelle fonction génératrice de la suite $\{\gamma_i\}_{i \in \mathbb{N}}$ la fonction Γ définie par la série formelle

$$\Gamma(t) = \sum_{i=0}^{\infty} \gamma_i t^i, \quad \forall t \in \mathbb{C}.$$

On suppose que $|t| < \min(R, 1)$ où R est le rayon de convergence de cette série. Par conséquent

$$\Gamma(t) = \sum_{i=0}^{\infty} (-t)^i \int_0^1 \binom{-s}{i} ds.$$

Puisque cette série converge uniformément, on peut commuter l'intégration et la sommation. Donc

$$\begin{aligned} \Gamma(t) &= \int_0^1 \sum_{i=0}^{\infty} (-t)^i \binom{-s}{i} ds = \int_0^1 (1-t)^{-s} ds \\ &= \int_0^1 e^{-s \ln(1-t)} ds = -\frac{t}{(1-t) \ln(1-t)}. \end{aligned}$$

Nous écrivons à partir de cette relation

$$-\frac{1}{t} \ln(1-t) \Gamma(t) = \frac{1}{1-t} = \sum_{i=0}^{\infty} t^i.$$

Or

$$-\frac{1}{t} \ln(1-t) = \sum_{i=0}^{\infty} \frac{t^i}{i+1}.$$

D'où l'identité

$$\left(\sum_{i=0}^{\infty} \frac{t^i}{i+1} \right) \left(\sum_{j=0}^{\infty} \gamma_j t^j \right) = \sum_{i=0}^{\infty} t^i.$$

Il suffit d'identifier les coefficients des termes de même degré en t pour obtenir

$$\begin{aligned} \gamma_0 &= 1, \\ \gamma_1 + \frac{1}{2} \gamma_0 &= 1, \\ \dots &\dots \\ \gamma_i + \frac{1}{2} \gamma_{i-1} + \frac{1}{3} \gamma_{i-2} + \dots + \frac{1}{i+1} \gamma_0 &= 1, \end{aligned}$$

ce qui permet de calculer les coefficients γ_i par récurrence.

i	0	1	2	3	4	5	...
γ_i	1	$\frac{1}{2}$	$\frac{5}{12}$	$\frac{3}{8}$	$\frac{251}{720}$	$\frac{95}{288}$	

Ceci permet d'obtenir les méthodes à pas liés d'Adams-Bashforth suivantes :

$$\begin{aligned}
y_{n+k+1} &= y_{n+k} + h f_{n+k} && \text{d'ordre 1 : c'est la méthode d'Euler,} \\
y_{n+k+1} &= y_{n+k} + \frac{h}{2}(3f_{n+k} - f_{n+k-1}) && \text{d'ordre 2,} \\
y_{n+k+1} &= y_{n+k} + \frac{h}{12}(23f_{n+k} - 16f_{n+k-1} + 5f_{n+k-2}) && \text{d'ordre 3,} \\
y_{n+k+1} &= y_{n+k} + \frac{h}{24}(55f_{n+k} - 59f_{n+k-1} + 37f_{n+k-2} - 9f_{n+k-3}) && \text{d'ordre 4.}
\end{aligned}$$

La méthode des fonctions génératrices appliquée aux méthodes d'Adams-Moulton donne

$$y_{n+k} = y_{n+k-1} + h \sum_{i=0}^k \gamma_i^* \nabla^i f_{n+k},$$

avec $\gamma_i^* = (-1)^i \int_{-1}^0 \binom{-s}{i} ds$.

La fonction génératrice de ces méthodes est

$$\Gamma^*(t) = -\frac{t}{\ln(1-t)}.$$

Pour les méthodes de Nyström on a

$$y_{n+k+1} = y_{n+k-1} + h \sum_{i=0}^k \chi_i \nabla^i f_{n+k},$$

avec $\chi_i = (-1)^i \int_{-1}^1 \binom{-s}{i} ds$.

La fonction génératrice est ici

$$\chi(t) = -\frac{(2-t)t}{(1-t)\ln(1-t)}.$$

Avec les méthodes de Milne-Simpson on a

$$y_{n+k} = y_{n+k-2} + h \sum_{i=0}^k \chi_i^* \nabla^i f_{n+k},$$

avec $\chi_i^* = (-1)^i \int_{-2}^0 \binom{-s}{i} ds$ et

$$\chi^*(t) = -(2-t) \frac{t}{\ln(1-t)}.$$

Donnons pour finir quelques exemples.

La méthode de Nyström d'ordre 2 est

$$y_{n+k+1} = y_{n+k-1} + 2h f_{n+k}.$$

La méthode de Milne-Simpson d'ordre 4 est

$$y_{n+k} = y_{n+k-2} + \frac{h}{3}(f_{n+k} + 4f_{n+k-1} + f_{n+k-2}),$$

(elle est équivalente à la méthode d'intégration de Simpson pour une intégrale définie).

La méthode d'Adams-Moulton d'ordre 4 est

$$y_{n+k} = y_{n+k-1} + \frac{h}{24}(9f_{n+k} + 19f_{n+k-1} - 5f_{n+k-2} + f_{n+k-3}).$$

4.3.8 Les méthodes de prédiction-correction

Si la méthode à pas liés est implicite ($\beta_k \neq 0$), alors y_{n+k} est la solution exacte de l'équation implicite. Le calcul de y_{n+k} s'effectue alors par itérations (cf. chapitre 3). Il se pose donc deux problèmes :

- choix de la valeur initiale des itérations,
- nombre d'itérations avant arrêt.

En pratique ces deux problèmes sont résolus à l'aide d'une méthode de prédiction-correction.

- la valeur initiale des itérations est fournie par une méthode à pas liés explicite : le prédicteur (on prédit la valeur de y_{n+k}),
- on effectue ensuite une seule itération à l'aide de la méthode à pas liés implicite : le correcteur (on corrige la valeur de y_{n+k} donnée par le prédicteur).

On a donc le schéma

$$\begin{aligned} \alpha_k^* y_{n+k}^* + \sum_{i=r}^{k-1} \alpha_i^* y_{n+i} &= h \sum_{i=r}^{k-1} \beta_i^* f_{n+i}, \quad \text{prédicteur,} \\ \sum_{i=0}^k \alpha_i y_{n+i} &= h(\beta_k f(x_{n+k}, y_{n+k}^*) + \sum_{i=0}^{k-1} \beta_i f_{n+i}), \quad \text{correcteur.} \end{aligned}$$

On remarquera que les deux méthodes n'utilisent pas forcément le même nombre de pas antérieurs (k pour la méthode implicite ; $k-r$ pour la méthode explicite).

Nous allons supposer que le prédicteur est d'ordre q , que le correcteur est d'ordre p et nous allons étudier les conditions pour que $\max_n |y_n - y(x_n)| = O(h^p)$ en effectuant seulement une itération pour le correcteur. On a le résultat suivant :

Théorème 4.3.17 *Si le prédicteur est d'ordre q et si le correcteur est d'ordre p , alors une condition suffisante pour que*

$$\max_n |y_n - y(x_n)| = O(h^p)$$

en effectuant seulement une itération avec le correcteur est que $q = p - 1$ (si les quantités y_i , pour $i = 0, \dots, k-1$, ont été calculées en utilisant une méthode d'ordre p au moins).

Démonstration.

Si le prédicteur est d'ordre q , alors

$$|y_{n+k}^* - y(x_{n+k})| = O(h^q).$$

Si le correcteur est d'ordre p , alors

$$\begin{aligned} \sum_{i=0}^k \alpha_i (y_{n+i} - y(x_{n+i})) &= h(\beta_k f(x_{n+k}, y_{n+k}^*) - \beta_k f(x_{n+k}, y(x_{n+k}))) + \\ &\quad h \sum_{i=0}^{k-1} \beta_i (f(x_{n+i}, y_{n+i}) - f(x_{n+i}, y(x_{n+i}))) + \\ &\quad O(h^{p+1}). \end{aligned}$$

On a donc en utilisant le fait que f satisfait une condition de Lipschitz de constante L

$$\begin{aligned} |\alpha_k| |y_{n+k} - y(x_{n+k})| &\leq \sum_{i=0}^{k-1} |\alpha_i| |y_{n+i} - y(x_{n+i})| + \\ &\quad h |\beta_k| |f(x_{n+k}, y_{n+k}^*) - f(x_{n+k}, y(x_{n+k}))| + \end{aligned}$$

$$\begin{aligned}
& h \sum_{i=0}^{k-1} |\beta_i| |f(x_{n+i}, y_{n+i}) - f(x_{n+i}, y(x_{n+i}))| + \\
& O(h^{p+1}) \\
\leq & \sum_{i=0}^{k-1} O(h^p) + hL |\beta_k| O(h^q) + \\
& hL \sum_{i=0}^{k-1} |\beta_i| O(h^p) + O(h^{p+1}) \\
= & O(h^p) + O(h^{q+1}) + O(h^{p+1}) + O(h^{p+1}).
\end{aligned}$$

Par conséquent pour que $\max_n |y_n - y(x_n)| = O(h^p)$, il suffit que $q = p - 1$. \square

On peut affiner les résultats de ce théorème en étudiant le comportement asymptotique de l'erreur. On peut montrer que

- si $q > p - 1$, le comportement asymptotique de l'erreur est le même si on corrige une fois ou si on corrige une infinité de fois,
- si $q = p - 1$, l'ordre de l'erreur n'est pas changé en corrigeant une fois (c'est notre théorème 4.3.17),
- si $q < p - 1$, l'ordre de l'erreur de discrétisation diminue si l'on ne corrige qu'une seule fois.

4.3.9 Démarrage des méthodes à pas liés

Dans une méthode à pas liés ou dans une méthode de prédiction-corrrection, le calcul de y_{n+k} nécessite la connaissance de y_n, \dots, y_{n+k-1} . Pour $n = 0$, il faut donc connaître y_0, \dots, y_{k-1} , alors que l'on ne dispose que de $y(a) = y_0$.

On calculera les valeurs initiales manquantes y_1, \dots, y_{k-1} , à l'aide d'une méthode à pas séparés (méthode de Runge-Kutta par exemple) dont l'ordre devra être au moins égal à l'ordre de la méthode à pas liés (cf. théorème 4.3.14). En fait il est inutile de choisir une méthode à pas séparés d'ordre plus élevé. On prendra donc deux méthodes de même ordre.

On remarque qu'au cours d'une intégration par une méthode à pas liés il est difficile de changer de pas ; ces méthodes sont en effet construites sur des abscisses x_n équidistantes. Si l'on désire changer le pas au cours de l'intégration, il sera donc nécessaire de redémarrer les calculs avec la méthode à pas séparés sur k pas avec la nouvelle valeur de pas choisie.

Dans les méthodes de prédiction-corrrection le contrôle de la propagation des erreurs de discrétisation est facile à réaliser ; en effet si prédicteur et correcteur ont le même ordre, alors la quantité $|y_{n+k}^* - y_{n+k}|$ est une bonne approximation de l'erreur $|y_{n+k} - y(x_{n+k})|$.

Enfin le grand avantage des méthodes à pas liés sur les méthodes de Runge-Kutta est de ne pas nécessiter (sauf au démarrage) d'évaluations de f en des points intermédiaires. Le volume des calculs en est réduit d'autant, ainsi que le temps de calcul.