

Université Paris - Sud

CENTRE D'ORSAY

C3 Analyse Numérique

- :: -

Interpolation, Quadratures, Equations Différentielles.

Cours professé par : M. R. TEMAM

rédigé par : MM. CORNIL et SCHEURER

(année 1970/71)

(Secrétariat de Mathématique 2ème cycle)

Université Paris - Sud

CENTRE D'ORSAY

C3 Analyse Numérique

- :: -

Interpolation, Quadratures, Equations Différentielles.

Cours professé par : M. R. TEMAM

rédigé par : MM. CORNIL et SCHEURER

(année 1970/71)

(Secrétariat de Mathématique 2ème cycle)

Table des matières.

CHAPITRE I - INTERPOLATION, DERIVATION ET INTEGRATION NUMERIQUES.

A - INTERPOLATION	I.1
1. Interpolation polynomiale	I.1
2. Formules d'erreur	I.2
3. Formules de Newton	I.5
B - DERIVATION NUMERIQUE	I.10
C - INTEGRATION NUMERIQUE	I.12

CHAPITRE II - INTEGRATION NUMERIQUE DES EQUATIONS DIFFERENTIELLES

A - THEOREMES D'EXISTENCES	II.1
Problème de Cauchy des équations différentielles ..	II.1
Rappel. Théorème d'Ascoli.....	II.4
Rappel. Théorème du point fixe.....	II.8
B - PROLONGEMENT DES SOLUTIONS	II.11
a. Solutions maximales	II.11
b. Critère de prolongement	II.13
C - RESOLUTION NUMERIQUE (Méthodes de différence finie).....	II.17
1. Les méthodes à un et plusieurs pas	II.17
2. Méthodes à un pas	II.20
3. Méthodes à plusieurs pas. Utilisation des formules de quadrature	II.27
4. Etude générale des méthodes à plusieurs pas. Stabilité	II.32
Méthode d'Adams-Bashforth	II.44
Méthode d'Adams-Moulton	II.45
Méthode de Nyström	II.45
Méthode de Milne Simpson	II.45

A INTERPOLATION

§ 1. Interpolation polynomiale.

Le problème est le suivant : étant donné une fonction numérique f définie continue sur un intervalle $[a, b]$ et des points $x_0 < x_1 \dots < x_n$ de $[a, b]$, trouver un polynôme qui prend les mêmes valeurs que f aux points $x_0 \dots x_n$.

Proposition 1. Soient $x_0 < x_1 \dots < x_n$ des points de $[a, b]$ et $f_0 \dots f_n$, $n+1$ nombres réels, il existe un polynôme unique de degré n , tel que :

$$P_n(x_i) = f_i \quad 0 \leq i \leq n.$$

Démonstration.

(1) Existence et unicité de P_n .

Soit $P_n(x) = \sum_{j=0}^n a_j x^j$ le polynôme. Les inconnues sont les coefficients $a_0 \dots a_n$. Par hypothèse P_n est tel que :

$$P_n(x_0) = f_0 \dots P_n(x_n) = f_n.$$

Ce système d'équations s'écrit sous forme matricielle :

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{pmatrix} \quad \text{soit } A a = f \quad (1.1)$$

Le déterminant de A est un déterminant de Vandermonde : le système (1.1) est donc un système de Cramer car :

$$\det A = \prod_{i < j} (x_i - x_j) \neq 0$$

car tous les points x_i sont distincts. Le système (1.1) admet donc une solution unique.

(2) Forme explicite de P_n .

Introduisons les polynômes élémentaires suivants :

$$* \quad \omega_n(x) = (x-x_0) \dots (x-x_n) = \prod_{i=0}^n (x-x_i)$$

donc

$$\omega_n'(x_k) = \prod_{\substack{i=0 \\ i \neq k}}^n (x_k - x_i)$$

$$* \quad l_k^{(n)}(x) = \frac{\prod_{\substack{i=0 \\ i \neq k}}^n (x-x_i)}{\prod_{\substack{i=0 \\ i \neq k}}^n (x_k - x_i)} = \frac{\omega_n(x)}{(x-x_k)\omega_n'(x_k)}$$

On vérifie que $l_k^{(n)}(x_i) = \begin{cases} 0 & \text{si } i \neq k \\ 1 & \text{si } i = k \end{cases}$

$$d^0 l_k^{(n)}(x) = n.$$

Considérons alors le polynôme

$$P_n(x) = \sum_{k=0}^n f_k \frac{\omega_n(x)}{(x-x_k)\omega_n'(x_k)} = \sum_{k=0}^n f_k l_k^{(n)}(x). \quad (1.2)$$

Ce polynôme est de degré n et d'autre part :

$$P_n(x_i) = \sum_{k=0}^n f_k l_k^{(n)}(x_i) = f_i \quad 0 \leq i \leq n.$$

Il vérifie donc le système (1.1) ; comme ce dernier admet une solution unique, le polynôme (1.2) est bien la solution cherchée.

Définition. Polynôme d'interpolation.

Soit $f : [a, b] \rightarrow \mathbb{R}$ où f est continu. On note $P_n(f ; x)$ et on appelle polynôme d'interpolation de f aux points $x_0 \dots x_n$ de $[a, b]$ l'unique polynôme P_n de degré n tel que :

$$P_n(f ; x_i) = f(x_i) = f_i \quad 0 \leq i \leq n.$$

Remarque. Le polynôme $P_n(f ; x)$ "approche" la fonction, mais il serait faux

d'en déduire que lorsque $n \rightarrow \infty$, P_n converge uniformément vers f dans $[a, b]$.

§ 2. Formules d'erreur.

Proposition 2.1. Soient $f \in \mathcal{C}^{n+1}([a, b])$ et $x_0 \dots x_n$ des points de $[a, b]$.

Pour tout $x \in]a, b[$, il existe ξ_x tel que :

$$\text{Min}(x, x_0 \dots x_n) \leq \xi_x \leq \text{Max}(x, x_0 \dots x_n) \text{ et}$$

$$f(x) - P_n(f; x) = \frac{(x-x_0) \dots (x-x_n)}{(n+1)!} f^{(n+1)}(\xi_x) = \frac{\omega_n(x)}{(n+1)!} f^{(n+1)}(\xi_x) \quad (2.1)$$

Démonstration. Considérons la fonction de t :

$$\phi(t) = f(t) - P_n(f; t) - \frac{(t-x_0) \dots (t-x_n)}{(x-x_0) \dots (x-x_n)} [f(x) - P_n(f; x)] .$$

Alors $\phi \in \mathcal{E}^{n+1}([a, b])$ et s'annule aux points $t = x_i, 0 \leq i \leq n$ et $t = x$.

D'après le lemme ci-après il existe donc $\xi = \xi_x$ tel que :

$$\phi^{(n+1)}(\xi_x) = 0 .$$

Donc il existe ξ_x tel que :

$$\phi^{(n+1)}(\xi_x) = f^{(n+1)}(\xi_x) - \frac{f(x) - P_n(f; x)}{(x-x_0) \dots (x-x_n)} (n+1)! = 0 .$$

(En effet $P_n(f; t)$ étant de degré n $\frac{d^{n+1}}{dt^{n+1}} P_n(f; t) = 0$ et $(t-x_0) \dots (t-x_n)$ étant également de degré n , $\frac{d^{n+1}}{dt^{n+1}} \omega_n(t) = (n+1)! \text{ coef.}[t^{n+1} \text{ dans } \omega_n(t)] = (n+1)!$).

Lemme. Soit $\phi : [a, b] \rightarrow \mathbb{R}$, $\phi \in \mathcal{E}^{n+1}[a, b]$,

(plus précisément f est n fois continument dérivable dans $[a, b]$ et sa dérivée $(n+1)^{\text{e}}$ existe dans $]a, b[$),

tel que $\phi(y_j) = 0$ avec $a < y_0 < \dots < y_{n+1} < b$. Alors il existe ξ ,

$y_0 < \xi < y_{n+1}$ tel que :

$$\phi^{(n+1)}(\xi) = 0 .$$

Démonstration. Elle repose sur le théorème de Rolle

	$\phi = 0$	en	$y_0 \quad y_1$	$y_n \quad y_{n+1}$	soit $n+2$ points
↓ Rolle	$\phi' = 0$	en	$\begin{matrix} y_0 & \nearrow & y_1 \\ & (1) & \end{matrix}$	$\begin{matrix} y_n & \nearrow & y_{n+1} \\ & (1) & \end{matrix}$	soit $n+1$ points
Rolle ↓	$\phi'' = 0$	en	$y_0^{(2)}$	$y_{n-1}^{(2)}$	soit n points
	\vdots					
	$\phi^{(n+1)} = 0$	en	$y_0^{(n+1)}$	= ξ		soit 1 point.

Proposition 2.2. Si f est analytique dans un ouvert Ω du plan complexe, qui contient l'intervalle réel $[a, b]$. Si Γ est une courbe de Jordan régulière contenant $[a, b]$ et contenu dans Ω alors :

$$|f(x) - P_n(f; x)| \leq \frac{\ell\mu}{2\pi\delta^{n+2}} |x-x_0| \dots |x-x_n| \quad (2.2)$$

où ℓ = longueur de Γ , $\mu = \max_{x \in \Gamma} |f(x)|$, $\delta = d([a, b], \Gamma)$.

Démonstration. On utilise la formule de Cauchy :

$$j(x; \Gamma) f^{(n+1)}(\xi) = \frac{(n+1)!}{2\pi i} \int_{\Gamma} \frac{f(z)}{(z-\xi)^{n+2}} dz.$$

D'après la formule (2.1)

$$|f(x) - P_n(f; x)| = \frac{|x-x_0| \dots |x-x_n|}{(n+1)!} |f^{(n+1)}(\xi)|.$$

Comme $j(x; \Gamma) = 1$:

$$|f^{(n+1)}(\xi)| = \frac{(n+1)!}{2\pi} \left| \int_{\Gamma} \frac{f(z)}{(z-\xi)^{n+2}} dz \right| \leq \frac{(n+1)!}{2\pi} \frac{\mu\ell}{\delta^{n+2}}$$

donc

$$\begin{aligned} |f(x) - P_n(f; x)| &\leq \frac{|x-x_0| \dots |x-x_n|}{(n+1)!} \frac{(n+1)!}{2\pi} \frac{\mu\ell}{\delta^{n+2}} \\ &\leq \frac{|x-x_0| \dots |x-x_n|}{2\pi} \frac{\mu\ell}{\delta^{n+2}} \end{aligned}$$

On a défini P_n comme un opérateur dans $\mathcal{C}^{(n+1)}[a, b]$. Définissons l'opérateur

L comme suit

$$L : f \in \mathcal{C}^{n+1}[a, b] \mapsto f - P_n(f) = L(f).$$

On a alors la :

Proposition 2.3. Formule d'erreur de Peano.

$$\text{Pour } f \in \mathcal{C}^{n+1}[a, b] \text{ on a : } L(f)(x) = f(x) - P_n(f; x) = \int_a^b f^{(n+1)}(t) k(x, t) dt \quad (2.3)$$

où

$$k(x, t) = \frac{1}{n!} L_x \left[\frac{(x-t)_+^n}{+} \right]$$

$$L_x \text{ désignant l'opérateur } L \text{ appliqué a : } x \mapsto \frac{(x-t)_+^n}{+} = \begin{cases} (x-t)^n & \text{si } (x-t) \geq 0 \\ 0 & \text{si } (x-t) < 0 \end{cases}.$$

Démonstration. Utilisons la formule de Taylor à reste intégrale :

$$f(x) = f(a) + (x-a)f'(a) + \dots + \frac{(x-a)^n}{n!} f^{(n)}(a) + \frac{1}{n!} \int_a^b f^{(n+1)}(t)(x-t)^n dt,$$

donc $L(f)(x) = f(x) - P_n(f; x) = \frac{1}{n!} \int_a^b f^{(n+1)}(t) L_x(x-t)_+^n dt$

$$= \int_a^b f^{(n+1)}(t) k(x, t) dt$$

Proposition 2.4. Pour $f \in \mathcal{E}^{n+1}[a, b]$ on a :

$$L(f)(x) = \sum_{i=0}^{i=n} \frac{\ell_i^{(n)}(x)}{n!} \int_{x_i}^x (x_i-t)^n f^{(n+1)}(t) dt, \quad x_0 \dots x_n \in [a, b].$$

Démonstration. Par définition de L :

$$k(x, t) = \frac{1}{n!} L_x[(x-t)_+^n] = \frac{1}{n!} (x-t)_+^n - \frac{1}{n!} \sum_{i=0}^n (x_i-t)_+^n \cdot \ell_i^{(n)}(x)$$

donc d'après la formule (2.3) :

$$\begin{aligned} L(f)(x) &= \int_a^b f^{(n+1)}(t) k(x, t) dt = \int_a^b f^{(n+1)}(t) \left[\frac{1}{n!} (x-t)_+^n - \sum_{i=0}^n (x_i-t)_+^n \ell_i^{(n)}(x) \right] dt \\ &= \frac{1}{n!} \int_a^b f^{(n+1)}(t) (x-t)_+^n dt - \frac{1}{n!} \sum_{i=0}^n \int_a^b f^{(n+1)}(t) (x_i-t)_+^n \ell_i^{(n)}(x) dt \\ &= \frac{1}{n!} \int_a^x f^{(n+1)}(t) (x-t)^n dt - \sum_{i=0}^n \frac{\ell_i^{(n)}(x)}{n!} \int_a^{x_i} f^{(n+1)}(t) (x_i-t)^n dt. \end{aligned}$$

D'après la proposition 1 $(x-t)^n$, étant de degré n , est son propre polynôme d'interpolation : $(x-t)^n = P_n((x-t)^n, x) = \sum_{i=0}^n (x_i-t)^n \ell_i^{(n)}(x)$, donc

$$\begin{aligned} L(f)(x) &= \frac{1}{n!} \int_a^x f^{(n+1)}(t) \sum_{i=0}^n (x_i-t)^n \ell_i^{(n)}(x) dt + \sum_{i=0}^n \frac{\ell_i^{(n)}(x)}{n!} \int_{x_i}^a f^{(n+1)}(t) (x_i-t)^n dt \\ &= \sum_{i=0}^n \frac{\ell_i^{(n)}(x)}{n!} \int_{x_i}^x f^{(n+1)}(t) (x_i-t)^n dt \end{aligned}$$

§ 3. Formules de Newton.

Soient $f \in \mathcal{E}^{n+1}[a, b]$ et $a < x_0 < \dots < x_n < b$. Cherchons une représentation du polynôme d'interpolation P_n du type $\sum_{k=0}^n a_k \omega_{k-1} = P_n$.

Une telle représentation existe puisque les $(n+2)$ polynômes élémentaires

$$\omega_{-1}(x) = 1, \quad \omega_0(x) = x - x_0, \quad \dots, \quad \omega_k(x) = \prod_{i=0}^k (x - x_i), \quad \dots, \quad \omega_n(x) = \prod_{i=0}^n (x - x_i)$$

forment

On a alors :
$$a_n = \sum_{k=0}^n \frac{f_k}{\omega'_n(x_k)} = \sum_{k=0}^n \frac{f(x_k)}{\omega'_n(x_k)} .$$

De même a_j ($j < n$) coefficient de x^j dans $P_j(f; x) = \sum_{k=0}^j a_k \omega_{k-1}(x)$ est égal par identification avec $P_j(f; x) = \sum_{k=0}^j \ell_k^{(j)}(x) f_k$ à $\sum_{k=0}^j \frac{f_k}{\omega'_j(x_k)}$.

Ainsi les coefficients a_k de $P_n = \sum_{k=0}^n a_k \omega_{k-1}$ sont donnés par :

$$a_k = \sum_{i=0}^k \frac{f(x_i)}{\omega'_k(x_i)} = \sum_{i=0}^k \frac{f(x_i)}{k \prod_{\substack{\ell=0 \\ \ell \neq i}}^k (x_i - x_\ell)} \quad 0 \leq k \leq n \quad (3.1)$$

Définition. Le coefficient a_k s'appelle la k^{e} différence divisée de f aux points $x_0 \dots x_k$ et on note : $a_k = f[x_0 \dots x_k]$.

Résumons les résultats précédents. On a la :

Proposition 3.1. Soient $f \in \mathcal{C}[a, b]$ et $x_0 \dots x_n$ des points de $[a, b]$ alors le polynôme d'interpolation de f aux points $x_0 \dots x_n$ s'écrit :

$$P_n(f; x) = \sum_{k=0}^n f[x_0, \dots, x_k] \omega_{k-1}(x) .$$

Proposition 3.2. On a la relation de récurrence suivante entre les différences divisées : $f[x_0] = f(x_0) = f_0$, et pour $n \geq 1$

$$f[x_0, \dots, x_n] = \frac{f[x_1, \dots, x_k] - f[x_0, \dots, x_{n-1}]}{x_n - x_0} .$$

Démonstration.

$$\begin{aligned} \sum_{k=0}^n \frac{f_k}{\omega'_n(x_k)} (x_n - x_0) &= a_n (x_n - x_0) = \sum_{k=0}^n \frac{f_k}{\omega'_n(x_k)} (x_n - x_k) + \sum_{k=0}^n \frac{f_k}{\omega'_n(x_k)} (x_k - x_0) \\ &= \sum_{k=0}^{n-1} \frac{f_k}{\omega'_n(x_k)} + \sum_{k=1}^n \frac{f_k}{\omega'_n(x_k)} \\ &= - f[x_0, \dots, x_{n-1}] + f[x_1, \dots, x_n] \end{aligned}$$

d'où
$$a_n = \frac{f[x_1, \dots, x_n] - f[x_0, \dots, x_{n-1}]}{x_n - x_0}$$

Nous allons préciser l'expression de $P_n(f; x) = \sum_{k=0}^n a_k \omega_{k-1}(x)$ dans deux cas particuliers.

* Cas des points équidistants.

Définissons les points x_0, \dots, x_n de l'intervalle $[a, b]$ de la façon suivante : $x_j = x_0 + j h$, $0 \leq j \leq n$, h étant le pas de l'intervalle.

Proposition 3.3. Soient $f \in \mathcal{C}[a, b]$, $x_j = x_0 + j h$, $0 < j < n$ des points de

$[a, b]$, alors $P_n(f; x) = \sum_{k=0}^n a_k \omega_{k-1}(x)$ avec

$$a_k = \sum_{i=0}^k \frac{(-1)^{k-i}}{h^k k!} \binom{k}{i} f_i.$$

Démonstration. D'après la formule (3.2)

$$a_k = \sum_{i=0}^k \frac{f_i}{\omega_k(x_i)} = \sum_{i=0}^k \frac{f_i}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_k)}.$$

Mais $x_j = x_0 + j h$, $0 \leq j \leq n$. Donc en remplaçant dans l'expression

ci-dessus

$$a_k = \sum_{i=0}^k \frac{(-1)^{k-i}}{h^k i! (k-i)!} f_i = \sum_{i=0}^k \frac{(-1)^{k-i}}{h^k k!} \binom{k}{i} f_i$$

* Cas des différences progressives.

Considérons une suite infinie $(y) = y_0, y_1, \dots, y_k, \dots$. On définit dans l'espace vectoriel des suites, l'opérateur linéaire ∇ de différence progressive par : $(\nabla y)_k = \nabla \{y_k\}_{k \in \mathbb{N}} = \{y_{k+1} - y_k\}_{k \in \mathbb{N}}$. On peut aussi définir les puissances ∇^n de cet opérateur. En effet on a le :

Lemme. $(\nabla^n y)_k = \sum_{j=0}^n [(-1)^{n-j} \binom{n}{j} y_{k+j}]$.

Démonstration. Raisonnons par récurrence.

* Par définition de ∇y_k la proposition est vraie pour $n = 1$.

* Admettons le résultat pour n :

$$(\nabla^n y)_k = \sum_{j=0}^n [(-1)^{n-j} \binom{n}{j} y_{k+j}]$$

* Démontrons le pour $n+1$:

$$\begin{aligned}
 (\nabla^{n+1} y)_k &= (\nabla^n y)_{k+1} - (\nabla^n y)_k \\
 &= \sum_{j=0}^n (-1)^{n-j} \binom{n}{j} y_{k+1+j} - \sum_{j=0}^n (-1)^{n-j} \binom{n}{j} y_{k+j} \\
 &= \sum_{j=1}^{n+1} (-1)^{n+1-j} \binom{n}{j+1} y_{k+j} - \sum_{j=0}^n (-1)^{n-j} \binom{n}{j} y_{k+j} \\
 &= \sum_{j=1}^n (-1)^{n+1-j} y_{k+j} \left(\binom{n}{j-1} + \binom{n}{j} \right) + (-1)^0 \binom{n}{n} + (-1)^{k+1} \binom{n}{0} \\
 &= \sum_{j=1}^n (-1)^{n+1-j} y_{k+j} \binom{n+1}{j} + (-1)^0 \binom{n+1}{n+1} + (-1)^{n+1} \binom{n+1}{0} \\
 &= \sum_{j=0}^{n+1} (-1)^{n+1-j} y_{k+j} \binom{n+1}{j}
 \end{aligned}$$

Exprimons alors le polynôme $P_n(f; x)$ à l'aide des différences progressives.

Proposition 3.4. Soit $f \in \mathcal{C}[a, b]$, x_0, \dots, x_n des points de $[a, b]$ tels que

$x_j = x_0 + j h$, $0 \leq j \leq n$ et posons $s = \frac{x-x_0}{h}$. Alors

$$P_n(f; x) = \sum_{k=0}^n \binom{s}{k} \nabla^k f_0.$$

Démonstration. D'après la proposition 3.3. on a vu que :

$$a_k = \sum_{i=0}^k \frac{(-1)^{k-i}}{h^k k!} \binom{k}{i} f_i$$

ce qui s'écrit encore d'après le lemme ci-dessus :

$$a_k = \frac{1}{h^k k!} (\nabla^k f)_j = \frac{1}{h^k k!} \nabla^k f_0.$$

On en déduit donc que :

$$\begin{aligned}
 P_n(f; x) &= \sum_{k=0}^n a_k \omega_{k-1}(x) = \sum_{k=0}^n a_k (x-x_0) \dots (x-x_j) \dots (x-x_{k-1}) \\
 &= \sum_{k=0}^n \frac{1}{h^k k!} \nabla^k f_0 (x-x_0) \dots (x-x_j) \dots (x-x_{k-1}) \\
 &= \sum_{k=0}^n \frac{1}{h^k k!} \nabla^k f_0 (x-x_0) \dots (x-x_0 - j h) \dots (x-x_0 - (k-1)h)
 \end{aligned}$$

car $x_j = x_0 + j h$, $0 \leq j \leq n$. Effectuons le changement de variable $s = \frac{x-x_0}{h}$:

$$\begin{aligned}
 P_n(f; x) &= \sum_{k=0}^n \frac{1}{h^k k!} \nabla^k f_0 s(s-1)h \dots (s-(k-1))h \\
 &= \sum_{k=0}^n \frac{s(s-1) \dots (s-(k-1))}{k!} \nabla^k f_0 = \sum_{k=0}^n \binom{s}{k} \nabla^k f_0
 \end{aligned}$$

B DERIVATION NUMERIQUE

1^e approximation.

$$f'(x) \approx \frac{f(x+\Delta x) - f(x)}{\Delta x} .$$

C'est l'approximation la plus simple, mais elle est toujours très mauvaise (perte de chiffres significatifs).

2^e approximation.

Cette approximation, valable lorsqu'on connaît f en plusieurs points x_0, \dots, x_n "voisins" de x , est plus raffinée. Elle consiste à approcher la dérivée de f au point x par la dérivée du polynôme d'interpolation de f aux points x_0, \dots, x_n :

$$f'(x) \approx P'_n(f; x).$$

Proposition 4. Soit f $(n+1)$ fois continument différentiable sur $[a, b]$.

Soit $Sp(x_0, \dots, x_n, x)$ l'intervalle engendré par x_0, \dots, x_n, x . Alors il

existe ξ_x tel que :

$$(i) \quad f'(x_j) - P'_n(f; x_j) = \frac{1}{(n+1)!} \omega'_n(x_j) f^{(n+1)}(\xi_x)$$

$$\text{et on a (ii) } f'(x_j) - P'_n(f; x_j) = \sum_{i=0}^n \left(\frac{\omega_i^{(x)}(x_j)}{n!} \right)' \int_{x_i}^{x_j} (x_i - t)^n f^{(n+1)}(t) dt.$$

Démonstration. (i) Sous l'hypothèse ci-dessus, on a vu que (prop. 2.1) il existe

ξ_x tel que

$$(2.1) \quad \text{Min}(x_0, \dots, x_n, x) \ll \xi_x \ll \text{Max}(x_0, \dots, x_n, x) \quad \text{et} \quad f(x) - P_n(f; x) = \frac{\omega_n^{(x)}}{n!} f^{(n+1)}(\xi_x).$$

Le problème est de savoir si la fonction $x \rightarrow f^{(n+1)}(\xi_x)$ est continument dérivable. Posons :

$$g(x) = f^{(n+1)}(\xi_x) = \frac{f(x) - P_n(f; x)}{\omega_n(x)} (n+1)! \quad \text{si } x \neq x_j$$

$$g(x_j) = \frac{f'(x_j) - P'_n(f; x_j)}{\omega'_n(x_j)} (n+1)!, \quad 0 \leq j \leq n.$$

On vérifie que g est continument dérivable sur $]a, b[$. $\frac{d}{dx} [g(x)]_{x=x_j}$ a donc un sens. On peut donc dériver la formule d'erreur (2.1) et faire $x = x_j$. On

obtient :

$$\left[\frac{d}{dx} (f(x) - P_n(f; x)) \right]_{x=x_j} = \frac{\omega'_n(x_j)}{(n+1)!} f^{(n+1)}(\xi_x) + \frac{\omega_n(x_j) g'(x_j)}{(n+1)!}$$

or par définition $\omega_n(x_j) = 0$. Donc :

$$f'(x_j) - P'_n(f; x_j) = \frac{\omega'_n(x_j)}{(n+1)!} f^{(n+1)}(\xi_x)$$

■

Il est plus rapide d'appliquer le théorème de Rolle au polynôme

$$[f'(x_j) - P'_n(f; x_j)] \prod_{i=1}^n (u - x_i) - \prod_{i \neq j} (x_j - x_i) [f(u) - P_n(f, u)]$$

qui possède x_j comme racine double et les x_i comme racine simple.

(ii) Sous l'hypothèse ci-dessus, on a vu que (prop. 2.4) :

$$f(x) - P_n(f; x) = \sum_{i=0}^n \frac{\ell_i^{(n)}(x)}{n!} \int_{x_i}^x (x_i - t)^n f^{(n+1)}(t) dt.$$

Dérivons cette égalité et faisons $x = x_j$; on obtient (la dérivation étant

licite) :

$$f'(x_j) - P'_n(f; x_j) = \sum_{i=0}^n \frac{(\ell_i^{(n)}(x_j))'}{n!} \int_{x_i}^{x_j} (x_i - t)^n f^{(n+1)}(t) dt$$

$$+ \underbrace{\sum_{i=0}^n \ell_i^{(n)}(x_j) (x_i - x_j)^n f^{(n+1)}(x_j)}_{\text{terme nul}}.$$

D'où :

$$f'(x_j) - P'_n(f; x_j) = \sum_{i=0}^n \frac{(\ell_i^{(n)}(x_j))'}{n!} \int_{x_i}^{x_j} (x_i - t)^n f^{(n+1)}(t) dt$$

3^e approximation.

On a vu (prop. 3.4) qu'on pouvait écrire :

$$P_n(f; x) = \sum_{k=0}^n \binom{s}{k} \nabla^k f_0 \quad \text{avec} \quad s = \frac{x-x_0}{h} .$$

Par suite :

$$P'_n(f; x) = \frac{1}{h} \frac{d}{ds} \sum_{k=0}^n \binom{s}{k} \nabla^k f_0 .$$

Donc :
$$P'_n(f; x_0) = \sum_{k=0}^n \frac{1}{h} \frac{(-1)^{k-1}}{k} \nabla^k f_0 .$$

Par conséquent, on peut écrire :

$$f'(x_0) \approx P'_n(f; x_0) = \sum_{k=0}^n \frac{(-1)^{k-1}}{k h} \nabla^k f_0$$

C- INTEGRATION NUMERIQUE

Soit $x \mapsto \mu(x)$ une fonction "poids" que l'on prendra dans $L^1([a,b])$. On veut calculer de manière approchée $\int_a^b f(x)\mu(x)dx$, connaissant f en certains points x_0, \dots, x_n de $[a,b]$.

Définition 5.1. On appelle formule de quadrature tout formule d'approximation

du genre :

$$\int_a^b f(x)\mu(x)dx \approx \sum_{j=0}^n A_j^{(n)} f_j$$

où $f_j = f(x_j)$, $0 \leq j \leq n$.

Ex : Formule des trapèzes.

$$x \mapsto \mu(x) = 1$$

$$x_0 = a$$

$$x_n = b$$

$$x_j = x_0 + j h, \quad 0 \leq j \leq n$$

$$A_0^{(n)} = A_n^{(n)} = \frac{h}{2}$$

$$A_j^{(n)} = h, \quad 1 \leq j \leq n-1$$

$$\int_a^b f(x)dx = h \left[\frac{f_0}{2} + f_1 + \dots + f_{n-1} + \frac{f_n}{2} \right] .$$

① Une 1^e méthode consiste à approcher la fonction par un polynôme, puis à calculer l'intégrale de ce polynôme multiplié par $\mu(x)$.

Définition 5.2. On appelle formule de quadrature du type interpolation, toute approximation du type :

$$\int_a^b f(x)\mu(x)dx \approx \int_a^b P_n(f; x)\mu(x)dx$$

où $P_n(f; x)$ est le polynôme d'interpolation de f aux points x_0, \dots, x_n de

$[a, b]$: $P_n(f; x) = \sum_{j=0}^n l_j^{(n)}(x) f(x_j)$. On a donc explicitement :

$$\int_a^b f(x)\mu(x)dx \approx \sum_{j=0}^n f(x_j) \int_a^b l_j^{(n)}(x)\mu(x)dx .$$

On pose

$$C_j^{(n)} = \int_a^b l_j^{(n)}(x)\mu(x)dx .$$

Dans les applications les $C_j^{(n)}$ doivent être connus explicitement. On va donc

préciser leur valeur. On a vu que :

$$l_j^{(n)} = \frac{\omega_n(x)}{(x-x_j)\omega_n'(x_j)} .$$

$$\begin{aligned} \text{Donc : } C_j^{(n)} &= \int_a^b l_j^{(n)}(x)\mu(x)dx = \int_a^b \frac{\omega_n(x)}{(x-x_j)\omega_n'(x_j)} \mu(x) dx \\ &= \int_a^b \frac{(x-x_0)\dots(x-x_k)\dots(x-x_n)}{(x-x_j)(x_j-x_0)\dots(x_j-x_{j-1})(x_j-x_{j+1})\dots(x_j-x_n)} \mu(x)dx . \end{aligned}$$

* Supposons que les points x_0, \dots, x_n soient équidistants. Autrement dit :

$$x_0 = a$$

$$x_n = b , \quad x_j = x_0 + jh , \quad 0 \leq j \leq n . \text{ Posons } s = \frac{x-x_0}{h} .$$

$$\begin{aligned} \text{Alors } l_j^{(n)} &= \frac{(x-x_0)\dots(x-x_{j-1})(x-x_{j+1})\dots(x-x_n)}{(x_j-x_0)\dots(x_j-x_{j-1})(x_j-x_{j+1})\dots(x_j-x_n)} = \frac{(x-x_0)\dots(x-x_{j-1})(x-x_{j+1})\dots(x-x_n)}{jh(j-1)h\dots h(-h)(-2h)\dots(-(n-j)h)} \\ &= \frac{(x-x_0)\dots(x-x_{j-1})(x-x_{j+1})\dots(x-x_n)}{j! h^n (-1)^{n-j} (n-j)!} = \frac{s(s-1)\dots(s-j+1)(s-j-1)\dots(s-n)}{j! (-1)^{n-j} (n-j)!} \end{aligned}$$

$$\begin{aligned}
\text{Donc } C_j^{(n)} &= \int_0^n \frac{s(s-1)\dots(s-j+1)(s-j-1)\dots(s-n)}{j!(-1)^{n-j}(n-j)!} \mu(sh + x_0) ds \\
&= (-1)^{n-j} h \int_0^n \prod_{\substack{i=0 \\ i \neq j}}^n (s-i) \frac{\mu(sh + x_0)}{j!(n-j)!} ds .
\end{aligned}$$

Dans le cas où $x \mapsto \mu(x) = 1$, on obtient la formule de Newton-Cotes

$$\boxed{C_j^{(n)} = \frac{1}{h} C_j^{(n)} = \frac{(-1)^{n-j}}{j!(n-j)!} \int_0^n \prod_{\substack{i=0 \\ i \neq j}}^n (s-i) ds}$$

Définition 5.3. On dit qu'une formule de quadrature est exacte sur un espace de fonction \mathcal{E} si :

$$\int_a^b f(x)\mu(x)dx = \sum_{j=c}^n A_j^{(n)} f(x_j) \quad \forall f \in \mathcal{E} .$$

Proposition 5.4. Une formule de quadrature est exacte sur \mathcal{P}_n (espace des polynômes de degré $\leq n$) si et seulement si elle est du type interpolation.

Démonstration.

* Supposons la formule du type interpolation. Alors :

$$\int_a^b f(x)\mu(x)dx \approx \sum_{j=0}^n \left(\int_a^b \ell_j^{(n)}(x)\mu(x)dx \right) f(x_j) .$$

Mais par hypothèse $\mathcal{E} = \mathcal{P}_n$ donc $f = P_n(f) \quad \forall f \in \mathcal{P}_n$. Donc : $\forall f \in \mathcal{P}_n$

$$\begin{aligned}
\int_a^b f(x)\mu(x)dx &= \int_a^b P_n(f; x)\mu(x)dx \\
&= \int_a^b \sum_{j=0}^n \ell_j^{(n)}(x)\mu(x)dx f(x_j) \\
&= \sum_{j=0}^n \left(\int_a^b \ell_j^{(n)}(x)\mu(x)dx \right) f(x_j) \\
&= \sum_{j=0}^n A_j^{(n)} f(x_j) .
\end{aligned}$$

La formule est donc exacte sur \mathcal{P}_n .

* Inversement, supposons la formule de quadrature exacte sur \mathcal{P}_n . Elle

l'est donc en particulier pour les polynômes élémentaires $\ell_i^{(n)}$, $0 \leq i \leq n$,

d'où :

$$\int_a^b \ell_i^{(n)}(x) \mu(x) dx = \sum_{j=0}^n A_j^{(n)} \ell_i^{(n)}(x_j) = A_i^{(n)} \quad \forall i, 0 \leq i \leq n,$$

$$\begin{aligned} \text{donc } \sum_{j=0}^n A_j^{(n)} f(x_j) &= \sum_{j=0}^n \left(\int_a^b \ell_j^{(n)}(x) \mu(x) dx \right) f(x_j) \\ &= \int_a^b \left(\sum_{j=0}^n \ell_j^{(n)}(x) f(x_j) \right) \mu(x) dx = \int_a^b P_n(f; x) \mu(x) dx \end{aligned}$$

Proposition 5.5. Formule d'erreur.

Soit $f \in \mathcal{C}^{n+1}[a, b]$.

$$\text{Si } e_f = \int_a^b f(x) \mu(x) dx - \int_a^b P_n(f; x) \mu(x) dx$$

$$\text{alors } |e_f| \leq \frac{M_{n+1}}{(n+1)!} \int_a^b |\omega_n(x)| |\mu(x)| dx$$

$$\text{où } M_{n+1} = t \in [a, b].$$

Démonstration. On a vu que (prop. 2.1), il existait ξ_x tel que :

$$\text{Min}(x_0, \dots, x_n, x) \leq \xi_x \leq \text{Max}(x_0, \dots, x_n, x)$$

$$\text{et : } f(x) - P_n(f; x) = \frac{\omega_n(x)}{(n+1)!} f^{(n+1)}(\xi_x).$$

Par conséquent :

$$e_f = \int_a^b f(x) \mu(x) dx - \int_a^b P_n(f; x) \mu(x) dx = \int_a^b [f(x) - P_n(f; x)] \mu(x) dx$$

$$\text{d'où } |e_f| = \left| \int_a^b [f(x) - P_n(f; x)] \mu(x) dx \right| \leq \int_a^b |f(x) - P_n(f; x)| |\mu(x)| dx$$

$$\text{soit } |e_f| \leq \int_a^b \frac{|\omega_n(x)|}{(n+1)!} |f^{(n+1)}(\xi_x)| |\mu(x)| dx$$

$$\leq \sup_{t \in [a, b]} |f^{(n+1)}(t)| \int_a^b \frac{|\omega_n(x)|}{(n+1)!} |\mu(x)| dx$$

$$\text{d'où } |e_f| \leq \frac{M_{n+1}}{(n+1)!} \int_a^b |\omega_n(x)| |\mu(x)| dx .$$

Remarque. * Soit f une fonction continue quelconque. L'erreur ne tend pas vers 0, pour $n \rightarrow \infty$. D'autre part à partir de $n = 5, 6$ les $C_j^{(n)}$ sont de signes variables, ce qui introduit des instabilités numériques et limite considérablement les possibilités de la formule de Newton-Cotes.

* Utilisation pratique de Newton-Cotes.

- (1) On subdivise l'intervalle $[a, b]$ en $x_0 = a, x_1, \dots, x_n = b$.
- (2) Sur chaque intervalle on utilise une formule de quadrature du type interpolation avec n "petit" ($n \leq 4$).

On retrouve comme cas particulier la formule de Simpson (cf. Ex.)

② Une 2^e méthode consiste à chercher s'il existe un choix judicieux des points x_0, \dots, x_n (x_0, \dots, x_n , $(n+1)$ points de $[a, b]$) tel qu'une formule de quadrature basée sur x_0, \dots, x_n soit exacte sur un espace de polynômes \mathcal{P}_n , avec $n' > n$. Les propositions 5.6 et 5.7 répondent complètement à la question dans le cas d'un poids $\mu > 0$. La réponse est alors liée aux polynômes orthogonaux relatifs à μ .

Proposition 5.6. Une formule de quadrature est exacte sur \mathcal{P}_{2n+1} si et seulement si : (i) elle est exacte sur $\mathcal{P}_n \iff$ elle est du type interpolation

(ii) le polynôme $\omega_n(x) = (x-x_0)\dots(x-x_n)$ vérifie :

$$\int_a^b x^p \omega_n(x) \mu(x) dx = 0 \quad , \quad 0 \leq p \leq n .$$

[autrement dit $\omega_n(x)$ est orthogonal pour le produit scalaire de poids μ à $1, x, \dots, x^n$].

Démonstration.

* Supposons la formule exacte sur \mathcal{P}_{2n+1} . Elle l'est donc à fortiori sur

\mathcal{P}_n . D'autre part :

$$\int_a^b x^p \omega_n(x) \mu(x) dx = \sum_{j=0}^n A_j^{(n)} x_j^p \omega_n(x_j) = 0 \quad \text{car} \quad \omega_n(x_j) = 0 .$$

* Inversement supposons (i) et (ii) satisfaits. Soit $P \in \mathcal{P}_{2n+1}$

$$P = \omega_n Q + R \quad \text{avec} \quad \begin{cases} d^0 R \leq n \\ d^0 Q \leq n \end{cases} \quad (\text{division euclidienne})$$

$$\text{donc} \quad \int_a^b P(x) \mu(x) dx = \int_a^b Q(x) \omega_n(x) \mu(x) dx + \int_a^b R(x) \mu(x) dx .$$

$$\text{D'après (ii)} \quad \int_a^b Q(x) \omega_n(x) \mu(x) dx = 0 .$$

$$\text{Donc} \quad \int_a^b P(x) \mu(x) dx = \int_a^b R(x) \mu(x) dx .$$

$$\text{D'après (i)} \quad \int_a^b R(x) \mu(x) dx = \sum_{j=0}^n A_j^{(n)} R(x_j) .$$

$$\text{Donc} \quad \int_a^b P(x) \mu(x) dx = \sum_{j=0}^n A_j^{(n)} R(x_j) .$$

Il reste à chercher si il existe des polynômes ω_n vérifiant (i) et (ii).

Proposition 5.7. Si $\mu(x) > 0$, il existe un polynôme $\omega_n(x) = (x-x_0)\dots(x-x_n)$

tel que (i) et (ii) soient vérifiés et ce polynôme est unique.

Démonstration. * Unicité

Soient ω_n et ω_n^* deux polynômes répondant à la question. Posons

$\omega_n - \omega_n^* = S$. On a $d^0 S \leq n$, donc d'après (ii) :

$$\int_a^b x^p S(x) \mu(x) dx = 0 \quad \text{avec } p \leq n.$$

Posons $S(x) = \sum_{p=0}^n \alpha_p x^p$. Alors

$$\sum_{p=0}^n \alpha_p \int_a^b x^p S(x) \mu(x) dx = \int_a^b \left(\sum_{p=0}^n \alpha_p x^p \right) S(x) \mu(x) dx = \int_a^b [S(x)]^2 \mu(x) dx.$$

Mais $\sum_{p=0}^n \alpha_p \int_a^b x^p S(x) \mu(x) dx = 0$ car $p \in n$. Donc

$$\int_a^b [S(x)]^2 \mu(x) dx = 0$$

ce qui implique $S \equiv 0$ car $\mu > 0$.

* Existence

Soit $\omega_n \in \mathcal{P}_{2n+1}$; $\int_a^b x^p \omega_n(x) \mu(x) dx = 0$ par hypothèse.

Définissons sur \mathcal{P}_{2n+1} un produit scalaire de poids μ par :

$$(f, g)_\mu = \int_a^b f(x) g(x) \mu(x) dx,$$

alors $(f, f)_\mu = \int_a^b f^2(x) \mu(x) dx$.

Par conséquent $\int_a^b x^p \omega_n(x) \mu(x) dx = 0 \iff (\omega_n, x^p)_\mu = 0 \iff \omega_n \perp x^p, p \leq n$.

La direction de ω_n est donc déterminée par le fait que ω_n est orthogonal

à l'hyperplan engendré par $\{x^0, x^1, \dots, x^n\}$. D'autre part, par construction

le coefficient du terme de degré n dans ω_n est 1. Cependant les racines

x_0, \dots, x_n de ω_n pourraient être complexes. Il n'en est rien :

Lemme. ω_n a $(n+1)$ racine réelles dans $]a, b[$.

Soient ξ_1, \dots, ξ_m les racines de ω_n appartenant à $]a, b[$, de multipli-

ciétés impaires ; on suppose que ces racines sont distinctes.

On a $m \geq n+1$ et supposons $m \leq n$.

Soit $(x-\xi_1)\dots(x-\xi_m)$ polynôme dont le degré est inférieur ou égal à n .

Donc :
$$\int_a^b (x-\xi_1)\dots(x-\xi_m) \omega_n(x) dx = 0 .$$

$(x-\xi_1)\dots(x-\xi_m) \omega_n(x)$ s'annule donc en $\xi_1 \dots \xi_m$ et est de signe constant sur $]a,b[$. Il est donc identiquement nul ce qui est contradictoire avec l'hypothèse.

Il y a donc exactement $(n+1)$ racines de ω_n appartenant à $]a,b[$ et ce sont les seules racines de ω_n .

Remarque. Les polynômes $\omega_0, \dots, \omega_n$ forment les polynômes orthogonaux pour le poids μ et les points x_0, \dots, x_n sont les racines du polynôme orthogonal aux monômes $1, x, \dots, x^n$.

③ La 3^e méthode qui va être décrite, utilise l'accélération de convergence.

(Méthode de Romberg).

Il s'agit toujours de calculer $T = \int_a^b f(x)\mu(x)dx$ et le problème est le suivant : ayant calculé des valeurs approchées T_0^0, \dots, T_0^k de T on veut chercher à exploiter ces $(k+1)$ calculs approchés pour obtenir une approximation de T encore meilleure.

Une façon de résoudre ce problème est d'utiliser la technique d'accélération de convergence, présentée dans un cadre général.

* Partons d'une suite (T_0^k) , (les T_0^k étant des valeurs approchées de T)

qui converge vers T .

Construisons le tableau triangulaire suivant, selon la formule :

$$\begin{array}{ccc} T_{m-1}^k & & \\ & \searrow & \\ T_{m-1}^{k+1} & \longrightarrow & T_m^k \end{array}$$

$$T_m^k = \frac{4^m T_{m-1}^{k+1} - T_{m-1}^k}{4^m - 1}$$

$$\begin{array}{ccccccc} T_0^0 & & & & & & \\ T_0^1 & T_1^0 & & & & & \\ T_0^2 & T_1^1 & T_2^0 & & & & \\ \vdots & \vdots & \vdots & \ddots & & & \\ T_0^k & T_1^{k-1} & \dots & & T_{k-1}^0 & & \\ \vdots & \vdots & \dots & & T_{k-1}^1 & T_k^0 & \\ & & \dots & & \dots & \dots & \end{array}$$

On va montrer que si la suite (T_0^k) converge, il en est de même des autres suites "colonnes" (T_i^k) [i indice de la colonne] et également des suites constituées par les termes d'une même diagonale.

* Ce résultat, non limité aux quadratures est tout à fait général. Sous certaines conditions que nous ne développerons pas ici, les suites ainsi construites peuvent converger vers T beaucoup plus rapidement que la suite initiale (T_0^k) . Il en est par exemple ainsi, pour une quadrature, lorsque T_0^k est la valeur obtenue par une formule des trapèzes à $2^k + 1$ points équidistants.

Cherchons tout d'abord une relation entre T_m^k, T_0^k, T_0^{k-m} .

On vérifie aisément sur le schéma triangulaire que T_m^k est une combinaison

linéaire de T_0^k, \dots, T_0^{k+m} à coefficients indépendants de k . Posons donc :

$$T_m^k = \sum_{i=0}^m C_{m, m-i} T_0^{k+i} .$$

Changeons d'indice dans cette formule, on obtient :

$$\text{d'une part} \quad T_{m-1}^{k+1} = \sum_{i=0}^{m-1} C_{m-1, m-1-i} T_0^{k+1+i}$$

$$\text{d'autre part} \quad T_{m-1}^k = \sum_{i=0}^{m-1} C_{m-1, m-1-i} T_0^{k+i}$$

c'est-à-dire les 2 éléments à partir desquels on obtient T_m^k .

$$\text{Ecrivons alors que} \quad T_m^k = \frac{4^m T_{m-1}^{k+1} - T_{m-1}^k}{4^m - 1} \iff (4^m - 1) T_m^k = 4^m T_{m-1}^{k+1} - T_{m-1}^k$$

$$\text{On a :} \quad (4^m - 1) \sum_{i=0}^m C_{m, m-i} T_0^{k+i} = 4^m \sum_{i=0}^{m-1} C_{m-1, m-1-i} T_0^{k+1+i} - \sum_{i=0}^{m-1} C_{m-1, m-1-i} T_0^{k+i} ,$$

d'où si $m \geq 0$, $i = 0, \dots, m$, la relation de récurrence suivante entre les

coefficients $C_{\alpha, \beta}$:

$$(4^m - 1) C_{m, m-i} = 4^m C_{m-1, m-i} - C_{m-1, m-1-i} .$$

Pour englober les cas limites, on pose : $C_{m-1, 0} = C_{m-1, -1} = 0$.

Formons maintenant le polynôme associé aux $C_{m, k}$:

$$\Pi_m(X) = \sum_{k=0}^m C_{m, k} X^k .$$

On a la :

Proposition 6.1. Le polynôme $\Pi_m(X) = \sum_{k=0}^m C_{m, k} X^k$ vérifie la relation de récur-

$$\text{rence :} \quad \Pi_m(X) = \frac{4^m - X}{4^m - 1} \Pi_{m-1}(X)$$

$$\text{et on a :} \quad \Pi_m(X) = \frac{(1 - \frac{X}{4^m}) \dots (1 - \frac{X}{4})}{(1 - \frac{1}{4^m}) \dots (1 - \frac{1}{4})} \Pi_0(X)$$

$$\text{avec} \quad \Pi_0(X) = C_{0,0} = 1 .$$

Considérons le polynôme $\Pi_{m-1}(X) = \sum_{k=0}^{m-1} C_{m-1,k} X^k$. On a :

$$\begin{aligned} (4^m - X)\Pi_{m-1}(X) &= 4^m \sum_{k=0}^{m-1} C_{m-1,k} X^k - \sum_{k=0}^{m-1} C_{m-1,k} X^{k+1} \\ &= 4^m \sum_{j=0}^{m-1} C_{m-1,m-1-j} X^{m-1-j} - \sum_{j=0}^{m-1} C_{m-1,m-1-j} X^{m-j} \end{aligned}$$

(on a fait le changement d'indice $j = m-1-k$).

Le coefficient de X^{m-j} dans $(4^m - X)\Pi_{m-1}(X)$ soit $\frac{4^m C_{m-1,m-j} - C_{m-1,m-1-j}}{4^m - 1} = C_{m,m-j}$

est donc égal au coefficient de X^{m-j} dans $\Pi_m(X)$.

On a donc :

$$\Pi_m(X) = \frac{(1 - \frac{X}{4^m})\Pi_{m-1}(X)}{(1 - \frac{1}{4^m})} = \frac{(4^m - X)}{(4^m - 1)} \Pi_{m-1}(X).$$

En "remontant" cette formule de récurrence, on obtient finalement en posant

$$\Pi_0(X) = C_{00} = 1$$

$$\Pi_m(X) = \frac{(1 - \frac{X}{4^m}) \dots (1 - \frac{X}{4})}{(1 - \frac{1}{4^m}) \dots (1 - \frac{1}{4})} \Pi_0(X).$$

Explicitons maintenant le coefficient $C_{m,k}$ de X^k dans $\Pi_m(X)$. On a le

résultat suivant

Proposition 6.2. Propriétés des coefficients $C_{m,k}$

- (i) $\sum_{k=0}^m C_{m,k} = \Pi_m(1) = 1$
- (ii) $(-1)^k C_{m,k} > 0$
- (iii) $\sum_{k=0}^m |C_{m,k}| = \sum_{k=0}^m (-1)^k C_{m,k} = \Pi_m(-1)$

$$= \frac{(1 + \frac{1}{4}) \dots (1 + \frac{1}{4^m})}{(1 - \frac{1}{4}) \dots (1 - \frac{1}{4^m})} \ll \frac{\prod_{\ell=0}^{\infty} (1 + \frac{1}{4^\ell})}{\prod_{\ell=0}^{\infty} (1 - \frac{1}{4^\ell})} \ll C_*$$

(On montre que $C_* < 2$).

$$(iv) \quad C_{m,k} = \frac{C_{m-1,k-1}}{1-4^k} = \left\{ \frac{1}{(1-4)^k \dots (1-4)} \right\} C_{m-k,0}.$$

Démonstration. Les 3 premières propriétés sont immédiates. Montrons la 4^e.

$$\text{On a} \quad \Pi_m(4z) = \frac{(1 - \frac{z}{4^{m+1}})(1 - \frac{z}{4^{m-2}}) \dots (1 - \frac{z}{4})(1-z)}{(1 - \frac{1}{4^m}) \dots (1 - \frac{1}{4})}$$

Remarquons que : $1 - z = 1 - \frac{z}{4^m} + \frac{z}{4^m} - z$. Donc :

$$\begin{aligned} \Pi_m(4z) &= \Pi_m(z) + \frac{(1 - \frac{z}{4^{m-1}}) \dots (1 - \frac{z}{4})(\frac{1}{4^m} - 1)}{(1 - \frac{1}{4^m}) \dots (1 - \frac{1}{4})} z \\ &= \Pi_m(z) - z \Pi_{m-1}(z) \end{aligned}$$

$$\text{d'où} \quad \sum_{k=0}^m C_{m,k} 4^k z^k = \sum_{k=0}^m C_{m,k} z^k - \sum_{k=0}^{m-1} C_{m-1,k} z^{k+1}$$

donc $(4^k - 1)C_{m,k} = -C_{m-1,k-1}$. Il en résulte que :

$$C_{m,k} = \frac{C_{m-1,k-1}}{1-4^k} = \frac{C_{m-2,k-2}}{(1-4^k)(1-4^{k-1})} = \left\{ \frac{1}{(1-4^k) \dots (1-4)} \right\} C_{m-k,0}$$

en remontant la récurrence.

Etudions maintenant la convergence de $\Pi_m(z)$ pour $z \in \mathbb{C}$.

Proposition 6.3. $\Pi_m(z)$ converge uniformément sur tout compact de \mathbb{C} vers $\Pi(z)$.

Démonstration. $\log \left| 1 - \frac{z}{4^{m'}} \right| \leq \frac{|z|}{4^{m'}}$. Par conséquent $\prod_{m'=0}^{\infty} (1 - \frac{z}{4^{m'}})$ est un pro-

duit uniformément et normalement convergent sur tout compact de \mathbb{C} . Il en est

donc de même pour :

$$\frac{\prod_{m'=0}^{\infty} (1 - \frac{z}{4^{m'}})}{\prod_{m'=0}^{\infty} (1 - \frac{1}{4^{m'}})} \Pi_0(X) = \lim_{m \rightarrow \infty} \Pi_m(z).$$

Remarque. Comme les fonctions Π_m sont analytiques dans le plan, la fonction

Π est aussi analytique et les suites dérivées k^e de Π_m , soient $\Pi_m^{(k)}(z)$,

convergent uniformément vers la dérivée correspondante $\Pi^{(k)}(z)$ sur tout compact de \mathbb{C} . Résumons, on a :

Corollaire 6.3. $\Pi(z) = \lim_{m \rightarrow \infty} \Pi_m(z)$ est analytique.
 $z \in K \subset \mathbb{C}$ K compact de \mathbb{C}

Les $\Pi_m^{(k)}$ convergent uniformément vers $\Pi^{(k)}$ sur tout compact de \mathbb{C} .

Proposition 6.4. Quand $m \rightarrow \infty$, $C_{m,i} \rightarrow C_i$.

Démonstration. On a vu que $\Pi_m(z) = \sum_{i=0}^m C_{m,i} z^i$ converge uniformément sur tout compact vers $\Pi(z) = \sum_{i=0}^{\infty} C_i z^i$. Fixons i et prenons $m \geq i$. Alors :

$$\lim_{m \rightarrow \infty} C_{m,i} = \lim_{m \rightarrow \infty} \frac{1}{i!} \frac{d^i}{dz^i} \Pi_m(0)$$

soit
$$C_i = \frac{1}{i!} \frac{d^i}{dz^i} \Pi(0)$$

Corollaire 6.4.1. Quand $m \rightarrow \infty$, $C_{m,m-k} \rightarrow 0$.

Corollaire 6.4.2. Quand $m \rightarrow \infty$, $C_{m,k} \rightarrow C_k = C_0 \left\{ \frac{1}{(1-4^k) \dots (1-4)} \right\}$.

Démonstration. C'est une conséquence immédiate du corollaire 6.4.1. et de (iv) dans la proposition 6.2.

Corollaire 6.4.3. $\Pi(z) = \sum_{k=0}^{\infty} C_k z^k = C_0 \sum_{k=0}^{\infty} \frac{z^k}{(1-4^k) \dots (1-4)}$

et
$$\Pi(1) = \lim_{m \rightarrow \infty} \Pi_m(1) = 1.$$

Démonstration. Conséquence directe du corollaire 6.4.2.

Nous allons maintenant montrer deux propositions sur la convergence des T_m^k .

Proposition 6.5. (Convergence des colonnes)

Si la limite lorsque $k \rightarrow \infty$ de T_0^k existe (soit T cette limite) alors :

$$\forall m \quad \lim_{k \rightarrow \infty} T_m^k = T.$$

Démonstration. On a vu que :

$$T_m^k = \sum_{i=0}^m C_{m,m-i} T_0^{k+i}, \quad T = \sum_{i=0}^m C_{m,m-i} T$$

$$\text{Formons } T_m^k - T = \sum_{i=0}^m C_{m,m-i} (T_0^{k+i} - T).$$

D'après les propositions précédentes (analyticité et $C_{m,m-i} \rightarrow 0$)

$$\lim_{n \rightarrow \infty} T_m^k - T = 0$$

Proposition 6.6. (Convergence des termes diagonaux)

Si la limite de T_0^k existe et égale T quand $k \rightarrow \infty$, alors, pour tout

$$k \text{ fixé : } \lim_{m \rightarrow \infty} T_m^k = T.$$

Démonstration. Formons $T_m^k - T = \sum_{i=0}^m C_{m,m-i} (T_0^{k+i} - T)$

$$= \sum_{i=0}^m C_{m,j} (T_0^{k+m-j} - T)$$

(on a posé $m - i = j$)

$$\forall \varepsilon > 0, \exists i_\varepsilon \text{ t.q. } i \gg i_\varepsilon \Rightarrow |T_0^i - T| < \varepsilon \quad \text{par hypothèse}$$

$$\text{ou } \forall \varepsilon > 0, \exists i_\varepsilon \text{ t.q. } i \gg i_\varepsilon \Rightarrow |T_0^{k+i} - T| < \varepsilon.$$

Pour $m \gg i_\varepsilon$, découpons la somme $\sum_{i=0}^m$ en deux :

$$T_m^k - T = \sum_{i=0}^{i_\varepsilon} C_{m,m-i} (T_0^{k+i} - T) + \sum_{i=i_\varepsilon}^m C_{m,m-i} (T_0^{k+i} - T).$$

$$\text{Donc } |T_m^k - T| \leq \left| \sum_{i=0}^{i_\varepsilon} C_{m,m-i} (T_0^{k+i} - T) \right| + \sum_{i=i_\varepsilon}^m |C_{m,m-i}| \varepsilon$$

$$\leq \left| \sum_{i=0}^{i_\varepsilon} C_{m,m-i} (T_0^{k+i} - T) \right| + C^* \varepsilon$$

Mais $\lim_{m \rightarrow \infty} C_{m,m-i} = 0 \quad \forall i$. Faisons tendre m vers l'infini dans la dernière

inégalité :

$$\limsup_{m \rightarrow \infty} |T_m^k - T| \leq \varepsilon C_* .$$

Comme $\varepsilon > 0$ est arbitraire, on a :

$$0 \leq \liminf_{m \rightarrow \infty} |T_m^k - T| \leq \limsup_{m \rightarrow \infty} |T_m^k - T| = 0$$

d'où $\lim_{m \rightarrow \infty} T_m^k = T$.

A THEOREMES D'EXISTENCES

Problème de Cauchy des équations différentielles.

α) Cas de \mathbb{R}^N .

Soit $\Omega \subset \mathbb{R} \times \mathbb{R}^N$ et $f : \Omega \rightarrow \mathbb{R}^N$.

On cherche un intervalle ouvert I et une fonction $y : I \subset \mathbb{R} \rightarrow \mathbb{R}^N$ vérifiant :

$$(i) \quad y'(t) = f(t, y(t))$$

$$(ii) \quad y(t_0) = y_0.$$

Plus précisément supposons f continue de Ω dans \mathbb{R}^N . Soit $(t_0, y_0) \in \Omega$.

Peut-on trouver un intervalle ouvert I contenant t_0 tel qu'il existe y

appartenant à $\mathcal{C}^1(I, \mathbb{R}^N)$ vérifiant :

$$(i) \quad y'(t) = f(t, y(t)) \quad \forall (t, y(t)) \in \Omega, \quad t \in I$$

$$(ii) \quad y(t_0) = y_0.$$

Ce problème est appelé le problème aux valeurs initiales ou problème de Cauchy

pour l'équation différentielle $y'(t) = f(t, y(t))$.

β) Cas d'un espace de Banach.

Plus généralement, soit E un espace de Banach et soit la fonction continue $f : \Omega \rightarrow E$ où Ω est un ouvert de $\mathbb{R} \times E$.

Peut-on trouver un ouvert I contenant t_0 tel qu'il existe y appartenant

à $\mathcal{C}^1(I, E)$ vérifiant :

$$(i) \quad y'(t) = f(t, y(t)) \quad \forall (t, y(t)) \in \Omega, \quad t \in I$$

$$(ii) \quad y(t_0) = y_0.$$

Si y existe on dit que y est solution sur I du problème de Cauchy.

Suivant l'hypothèse de départ, le problème de Cauchy admet une réponse plus ou moins précise. Nous allons nous placer dans les 2 cas suivants :

$$\begin{array}{l} \text{Hypothèses 1} \\ \text{Hypothèses 2} \end{array} \quad \begin{array}{l} H_1 \\ H_2 \end{array} \quad \left\{ \begin{array}{l} E \text{ est de dimension finie} \\ f \text{ est continue sur } \Omega \\ E \text{ de dimension finie ou infinie} \\ f \text{ est continue sur } \Omega \\ f \text{ est localement lipschitzienne en} \\ y \text{ sur } \Omega \end{array} \right.$$

Définition 1. Fonction lipschitzienne.

On dit qu'une fonction $f : \Omega \rightarrow E$ continue est lipschitzienne en y pour la constante k si :

$$\begin{array}{l} \forall (t,y) \in \Omega \\ \forall (t,y_*) \in \Omega \end{array} \quad \|f(t,y) - f(t,y_*)\| \leq k \|y - y_*\| .$$

Définition 2. Fonction localement lipschitzienne.

On dit qu'une fonction $f : \Omega \rightarrow E$ continue est localement lipschitzienne en y si pour tout point de Ω il existe un voisinage v de ce point dans lequel f est lipschitzienne en y .

Remarquons que :

Proposition. Si $f : \Omega \rightarrow E$ est continument différentiable en y , f est localement lipschitzienne en y dans Ω .

Démonstration. Soit $(t_0, y_0) \in \Omega$. Par hypothèse la différentielle f'_y existe et est continue. Donc il existe un voisinage fermé \mathcal{V} de (t_0, y_0) , contenu dans Ω , dans lequel :

$$\|f'_y\| \leq K .$$

Donc d'après le théorème des accroissements finis :

$$\begin{aligned} \forall (t,y) \in \mathcal{V} \\ \forall (t,y_*) \in \mathcal{V} \end{aligned} \quad \begin{aligned} \|f(t,y_*) - f(t,y)\| &\leq \|f'_y\| \|y-y_*\| \\ &\leq K \|y-y_*\| \end{aligned}$$

On va maintenant montrer 2 résultats correspondant respectivement aux hypothèses H_1 et H_2 .

Théorème 1. (Théorème d'existence ou théorème de Cauchy-Arzela).

Sous l'hypothèse H_1 le problème de Cauchy possède au moins une solution y définie sur un intervalle I contenant t_0 .

Démonstration. On va seulement démontrer l'existence sur un intervalle $[t_0, t_0+h]$, le résultat complet étant facile à obtenir.

L'idée est de construire une solution approchée, puis de passer à la limite. La solution approchée sera constituée par une suite de fonctions continues, linéaires par morceaux sur $[t_0, t_0+h]$.

Commençons par déterminer un voisinage commode de $(t_0, y_0) = (t_0, y(t_0))$ dans lequel, on le verra, la solution est définie. La fonction f étant continue, il existe un voisinage \mathcal{V} fermé de (t_0, y_0) tel que :

$$\|f(t,y)\| \leq M \quad (t,y) \in \mathcal{V}.$$

Choisissons ensuite $h > 0$ assez petit de sorte que :

$$\mathcal{V}_h \subset \mathcal{V} \text{ avec } \mathcal{V}_h = \{(t,y) \mid |t-t_0| \leq h, \|y-y_0\| \leq Mh\}.$$

On a donc : $\|f(t,y)\| \leq M \quad \forall (t,y) \in \mathcal{V}_h$.

Subdivisons $[t_0, t_0+h]$ en N intervalles de largeur $\frac{h}{N}$, h étant fixé :

$$t_0 < t_1 < \dots < t_N = t_0+h, \quad t_i = t_0 + \frac{i}{N} h.$$

Considérons alors une solution approchée, continue, linéaire sur chaque intervalle $[t_i, t_{i+1}]$ définie de la façon suivante :

$$y_N(t_0) = y_0$$

$$y_N'(t) = \frac{y_N(t_{i+1}) - y_N(t_i)}{t_{i+1} - t_i} \quad t \in]t_i, t_{i+1}[$$

Une telle définition a un sens, si l'on montre que l'on peut définir y_N de proche en proche sur chaque intervalle $[t_i, t_{i+1}]$, c'est-à-dire si l'on montre que :

$$(t_j, y_N(t_j)) \in \mathcal{V}_h, \quad 0 \leq j \leq i \implies (t, y_N(t)) \in \mathcal{V}_h, \quad t \in [t_i, t_{i+1}]$$

Supposons que : $(t_j, y_N(t_j)) \in \mathcal{V}_h, \quad 0 \leq j \leq i$.

$$\begin{aligned} \text{On a alors : } \|y_N(t_{j+1}) - y_N(t_j)\| &= \|f(t_j, y_N(t_j))\| \cdot |t_{j+1} - t_j| \\ &\leq M |t_{j+1} - t_j| = M \frac{h}{N} \end{aligned}$$

et ceci pour tout j tel que : $0 \leq j \leq i$.

Donc d'après l'hypothèse :

$$\|y_N(t_{i+1}) - y_0\| \leq \sum_{j=0}^i \|y_N(t_{j+1}) - y_N(t_j)\| \leq \sum_{j=0}^i M |t_{j+1} - t_j| = M N \frac{h}{N} = M h$$

Comme $|t_{i+1} - t_0| \leq h$ par construction, on en déduit que :

$$(t_{i+1}, y_N(t_{i+1})) \in \mathcal{V}_h$$

La définition de y_N a donc bien un sens.

Nous allons maintenant montrer que la suite $\{y_N\}$ ainsi définie tend uniformément sur $[t_0, t_0+h]$ vers une fonction y . Pour cela nous allons utiliser le théorème d'Ascoli.

Rappel. Théorème d'Ascoli.

On suppose que B est un espace de Banach et E un espace métrique compact. Pour qu'une partie H de l'espace de Banach $\mathcal{C}^0(E, B)$ soit relati-

vement compacte il faut et il suffit que H soit équicontinue et que pour tout $x \in E$ l'ensemble $H(x)$ de tous les $f(x)$ tels que $f \in H$ soit relativement compact dans B .

Dans le cas présent : $E = I = [t_0, t_0+h]$ qui est bien compact

$B = E$ qui est un espace de Banach par hypothèse.

* Montrons tout d'abord que $H = \{y_N, N \in \mathbb{N}\} \subset \mathcal{C}^0(I, E)$ est équicontinu sur $[t_0, t_0+h]$.

Par définition $y_N'(t)$ existe p.p et pour $t \in]t_i, t_{i+1}[$:

$$\|y_N'(t)\| = \|f(t_i, y_N(t_i))\| \leq M.$$

Soient t, t_* quelconques dans $[t_0, t_0+h]$, alors :

$$y_N(t) - y_N(t_*) = \int_{t_*}^t y_N'(s) ds \quad (1)$$

$$(1) \Rightarrow \|y_N(t) - y_N(t_*)\| = \left\| \int_{t_*}^t y_N'(s) ds \right\| \leq \sup_{t \in]t, t_*[} \|y_N'(t)\| \cdot |t - t_*|$$

ainsi $\forall t \in [t_0, t_0+h], \|y_N(t) - y_N(t_*)\| \leq M|t - t_*|$

$$\forall t_* \in [t_0, t_0+h].$$

* Montrons que $H(t) = \{y_N(t) | y_N \in H\} \forall t \in [t_0, t_0+h]$ est relativement compact.

E étant de dimension finie, il suffit, pour prouver que $H(t)$ est relativement compact, de montrer que $H(t)$ est borné.

Soit t quelconque dans $[t_0, t_0+h]$. Pour tout $y_N \in H$, nous avons :

$$y_N(t) = \int_{t_0}^t y_N'(s) ds + y_0$$

donc :

$$\begin{aligned} \|y_N(t)\| &\leq \int_{t_0}^t \|y_N'(s)\| ds + \|y_0\| \\ &\leq \sup_{s \in]t_0, t[} \|y_N'(s)\| \cdot |t - t_0| + \|y_0\| \\ &\leq M h + \|y_0\| < \infty. \end{aligned}$$

Ainsi : $\forall t \in [t_0, t_0+h]$ les éléments $y_N(t)$ de $H(t)$ sont bornés par une quantité finie. Donc $H(t)$ est borné pour tout t de $[t_0, t_0+h]$.

* Nous pouvons alors appliquer le théorème d'Ascoli et en déduire que

$H = \{y_N ; N \in \mathbb{N}\}$ est relativement compact. Autrement dit il existe une fonction y de $\mathcal{C}^0(I, E)$ tel que :

$$y_{\nu} \rightarrow y \text{ uniformément sur } I = [t_0, t_0+h] ,$$

où $\{y_{\nu}\}$ est une suite extraite de $\{y_N\}$. En effet rappelons que :

$H \subset \mathcal{C}^1(I, E)$ relativement compact \Leftrightarrow toute suite de point de H possède une valeur d'adhérence dans $\mathcal{C}^1(I, E)$.

Dans ce qui suit on notera, pour simplifier, $\{y_N\}$ la suite extraite $\{y_{\nu}\}$.

On sait que par hypothèse : $y_N(t_0) = y_0$

$$\|y_N(t) - y_0\| \leq M h \quad \forall t \in [t_0, t_0+h] .$$

Par passage à la limite on en déduit donc :

$$y(t_0) = y_0$$

$$\|y(t) - y_0\| \leq M h , \quad \forall t \in [t_0, t_0+h]$$

ce qui est équivalent à : $(t, y(t)) \in \mathcal{V}_h \subset \Omega , \forall t \in [t_0, t_0+h]$.

La limite y de la suite $\{y_N\}$ ne "sort" donc pas du voisinage \mathcal{V}_h dans lequel on a défini les fonctions y_N .

* Montrons enfin que y satisfait (i) et (ii) c'est-à-dire est solution du problème de Cauchy. On sait que les conditions (i) et (ii) sont équivalentes à :

$$y(t) = y_0 + \int_{t_0}^t f(s, y(s)) ds , \quad \forall t \in [t_0, t_0+h] .$$

Or si $t \in]t_i, t_{i+1}[$, par construction :

$$y_N(t) - y_N(t_i) = (t-t_i) f(t_i, y_N(t_i))$$

$$y_N(t_i) - y_N(t_{i-1}) = (t_i - t_{i-1}) f(t_{i-1}, y_N(t_{i-1}))$$

.....

$$y_N(t_1) - y_0 = (t_1 - t_0) f(t_0, y_N(t_0)) .$$

Sommons ces égalités, on obtient :

$$y_N(t) - y_0 = (t-t_i) f(t_i, y_N(t_i)) + \sum_{j=1}^i (t_j - t_{j-1}) f(t_{j-1}, y_N(t_{j-1})) .$$

Posons
$$I_{1,N} = (t-t_i) f(t_i, y_N(t_i)) + \sum_{j=1}^i (t_j - t_{j-1}) f(t_{j-1}, y(t_{j-1}))$$

$$I_{2,N} = (t-t_i) [f(t_i, y_N(t_i)) - f(t_i, y(t_i))] + \sum_{j=1}^i (t_j - t_{j-1}) [f(t_{j-1}, y_N(t_{j-1})) - f(t_{j-1}, y(t_{j-1}))] .$$

On vérifie que :
$$y_N(t) = y_0 + I_{1,N} + I_{2,N} \quad (1) .$$

Il est clair que, f étant continue, $I_{1,N}$ est une somme de Riemann pour l'intégrale :

$$\int_{t_0}^t f(s, y(s)) ds$$

et pour la subdivision $t_0 < t_1 < \dots < t_i < t$. Donc :

$$\lim_{N \rightarrow \infty} I_{1,N} = \int_{t_0}^t f(s, y(s)) ds .$$

Par ailleurs, f étant continu et y_N convergeant uniformément vers y sur $[t_0, t_0+h]$ il en résulte que :

$$f(t, y_N(t)) \rightarrow f(t, y(t)) \text{ uniformément sur } [t_0, t_0+h] .$$

Autrement dit :

$$\forall \varepsilon > 0, \exists N_\varepsilon \text{ t.q. } N \geq N_\varepsilon \implies \|f(s, y_N(s)) - f(s, y(s))\| \leq \varepsilon, \forall s \in [t_0, t_0+h] .$$

Par conséquent, comme t_j pour $1 \leq j \leq i$ appartient à $[t_0, t_0+h]$ pour

$N \geq N_\varepsilon$ on a :

$$\|I_{2,N}\| \leq |t-t_i| \cdot \|f(t_i, y_N(t_i)) - f(t_i, y(t_i))\| \\ + \sum_{j=1}^i |t_j - t_{j-1}| \cdot \|f(t_{j-1}, y_N(t_{j-1})) - f(t_{j-1}, y(t_{j-1}))\|$$

d'où $\|I_{2,N}\| \leq \varepsilon [|t-t_i| + |t_i-t_{i-1}| + \dots + |t_1-t_0|] = \varepsilon |t-t_0|$.

Ainsi : $\forall \varepsilon > 0, \exists N_\varepsilon$ t.q. $N \geq N_\varepsilon \Rightarrow \|I_{2,N}\| \leq |t-t_0| \varepsilon$

par conséquent $\lim_{N \rightarrow \infty} I_{2,N} = 0$.

Faisons alors tendre N vers l'infini dans (1). Par passage à la limite il en résulte que :

$$y(t) = y_0 + \int_{t_0}^t f(s, y(s)) ds, \quad \forall t \in [t_0, t_0+h]$$

ce qui signifie que : $y \in \mathcal{C}^1(I, E)$ est solution du problème de Cauchy. Le théorème 1 est ainsi démontré.

Théorème 2. (Théorème d'existence et d'unicité ou théorème de Cauchy-Lipschitz).

Sous l'hypothèse (H_2) , il existe un intervalle ouvert I contenant t_0 , tel que le problème de Cauchy possède une solution unique sur I .

On va utiliser la méthode des approximations successives et le théorème du point fixe.

Rappel. Théorème du point fixe.

Soit \mathcal{B} une boule fermée d'un espace de Banach et $T : \mathcal{B} \rightarrow \mathcal{B}$, une contraction stricte ($\|Ty - Tz\| \leq C\|y-z\|$ avec $C < 1$).

Alors il existe un y unique de \mathcal{B} tel que : $Ty = y$.

Nous allons définir une application T vérifiant les hypothèses du théorème.

* Pour cela déterminons tout d'abord un voisinage \mathcal{V}_h convenable. On a vu que \mathcal{V}_h est déterminé par la donnée du nombre h . Il existe un voisinage \mathcal{V}

de (t_0, y_0) tel que :

$$\|f(t, y)\| \leq M \quad \text{pour } (t, y) \in \mathcal{V}.$$

Donc d'après le théorème des accroissements finis, il existe un nombre

$k = \left\| \frac{f'_y}{y} \right\|$ tel que :

$$\|f(t, y) - f(t, y_*)\| \leq k \|y - y_*\| \quad (t, y) \in \mathcal{V}, (t, y_*) \in \mathcal{V}.$$

Définissons alors \mathcal{V}_h comme :

$$\mathcal{V}_h = \{(t, y) \in \mathcal{V} \mid |t - t_0| \leq h, \|y - y_0\| \leq Mh\} \quad \text{avec } h < \frac{1}{k}.$$

Posons d'autre part $I =]t_0 - h, t_0 + h[$.

* Définissons maintenant T . On sait que :

$$\left. \begin{array}{l} y \in \mathcal{C}^1(I, E) \\ \{(t, y(t)) \mid t \in I\} \in \Omega \\ y(t) = y_0 + \int_{t_0}^t f(s, y(s)) ds \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} y \text{ est solution du problème} \\ \text{de Cauchy} \end{array} \right.$$

Posons donc : $T : y \mapsto Ty = z$ avec $z(t) = y_0 + \int_{t_0}^t f(s, y(s)) ds$.

T satisfait aux hypothèses du théorème du point fixe.

$$\textcircled{1} \quad T : \mathcal{B} \rightarrow \mathcal{B} \quad \text{où } \mathcal{B} = \mathcal{C}_h^1(\bar{I}, E) = \{y \in \mathcal{C}^1(\bar{I}, E) \mid \|y(t) - y_0\| \leq Mh, t \in \bar{I}\}.$$

Soit $y \in \mathcal{C}_h^1(\bar{I}, E)$. Appliquons T à y : $Ty = z$. $z \in \mathcal{C}^1(\bar{I}, E)$ par construction.

De plus :

$$\|z(t) - y_0\| = \left\| \int_{t_0}^t f(s, y(s)) ds \right\| \leq \int_{t_0}^t \|f(s, y(s))\| ds \leq M(t - t_0)$$

donc, comme $t \in [t_0 - h, t_0 + h]$: $\|z(t) - y_0\| \leq Mh$.

Par conséquent $(s, y(s)) \in \mathcal{V}_h$ et $z \in \mathcal{C}_h^1(\bar{I}, E)$.

(2) T est contractante.

Soient $Ty = z$, $Tx = v$ avec $(t, y) \in \mathcal{V}_h$, $(t, x) \in \mathcal{V}_h$, où \mathcal{V}_h est le voisinage défini précédemment et où t appartient à \bar{I} .

Formons :

$$z(t) - v(t) = \int_{t_0}^t (f(s, y(s)) - f(s, x(s))) ds .$$

On a alors :

$$\begin{aligned} \|z(t) - v(t)\| &= \left\| \int_{t_0}^t (f(s, y(s)) - f(s, x(s))) ds \right\| \\ &\leq \int_{t_0}^t \|f(s, y(s)) - f(s, x(s))\| ds . \end{aligned}$$

Soit d'après l'inégalité des accroissements finis :

$$\begin{aligned} \|z(t) - v(t)\| &\leq k \int_{t_0}^t \|y(s) - x(s)\| ds \quad (\text{rappelons que } k = \|f'_y\|) \\ &\leq k (t - t_0) \|y - x\| \mathcal{C}^1(\bar{I}, E) \end{aligned}$$

d'où comme $t \in [t_0 - h, t_0 + h]$:

$$\|z(t) - v(t)\| < k h \|y - x\| \mathcal{C}^1(\bar{I}, E)$$

Mais cette inégalité est vraie pour tout t de $[t_0 - h, t_0 + h]$.

On peut donc écrire :

$$\|z - v\|_{\mathcal{C}^1(\bar{I}, E)} \leq k h \|y - x\|_{\mathcal{C}^1(\bar{I}, E)}$$

Soit encore par définition de z et v :

$$\|Ty - Tx\|_{\mathcal{C}^1(\bar{I}, E)} \leq k h \|y - x\|_{\mathcal{C}^1(\bar{I}, E)}$$

Mais par hypothèse (construction de \mathcal{V}_h) $kh < 1$; T est alors bien contractante.

On peut alors appliquer le théorème du point fixe et conclure qu'il existe un point y unique de $\mathcal{B} = \mathcal{C}_h^1(\bar{I}, E)$ tel que $Ty = y$, autrement dit que :

$$y \in \mathcal{C}^1(\bar{I}, E) , \quad \{(t, y(t)) \mid t \in \bar{I}\} \subset \mathcal{V}_h \subset \Omega$$

et

$$y(t) = y_0 + \int_{t_0}^t f(s, y(s)) ds$$

par définition de T .

Corollaire 2 . Supposons les hypothèses H_2 réalisées :

soit y (resp. z) la solution du problème de Cauchy sur l'intervalle ouvert J contenant t_0 (resp. l'intervalle ouvert K contenant t_0).

Alors $y = z$ sur $J \cap K$.

Démonstration. Considérons $\mathcal{U} = \{t \in J \cap K \mid y(t) = z(t)\}$.

$$\forall t_1 \in J,]h \text{ t.q. } \begin{cases} y' = f(t, y) \\ v(t_1) = y(t_1) \end{cases} \text{ a une solution unique}$$

sur $]t_1-h, t_1+h[$ d'après le théorème de Cauchy-Lipschitz. On a donc $y = z$

sur $]t_1-h, t_1+h[$ et :

$$\mathcal{U} \supset]t_1-h, t_1+h[$$

J est ouvert, \mathcal{U} est fermé et non vide, donc par connexité $\mathcal{U} = J \cap K$.

Les théorèmes de Cauchy-Arzela et Cauchy-Lipschitz donnent une solution définie sur un petit intervalle $[t_0, t_0+h]$ ou $]t_0-h, t_0+h[$.

Le problème qui se pose est d'examiner si on peut prolonger les solutions, dont on a vu l'existence (et l'unicité) au delà de ces petits intervalles et si oui sous quelles conditions.

B PROLONGEMENT DES SOLUTIONS

a) Solutions maximales.

Nous allons définir une relation d'ordre simple sur l'ensemble des solutions du problème de Cauchy.

On note $y < z$ lorsque : $y : I \rightarrow E$ et $z : J \rightarrow E$ solution du problème

de Cauchy respectivement définies sur I et J sont telles que $I \subset J$.

Cette relation d'ordre est en fait la relation d'inclusion sur les graphes des solutions.

Définition 2.1. On appelle solution maximale, tout élément maximal de l'ensemble des solutions ainsi ordonné.

Lemme 2.2. L'ensemble des solutions ordonné par α est inductif.

Démonstration. Soit $\{y_\alpha, I_\alpha\}_{\alpha \in \Lambda}$ une famille totalement ordonnée de solutions.

On a donc : $\forall \alpha \in \Lambda$ soit $y_\alpha < y_\beta$ soit $y_\alpha > y_\beta$.

$$\forall \beta \in \Lambda$$

Soit $I = \bigcup_{\alpha \in \Lambda} I_\alpha$. I est un intervalle ouvert et $t_0 \in I$.

Soit $y : I \rightarrow E$ t.q. $\forall t \in I$, $y(t) = y_{\alpha_0}(t)$ où α_0 est un élément de Λ tel que $t \in I_{\alpha_0}$.

Ces définitions sont comparables car :

$$y \in \mathcal{E}^1(I, E), \{(t, y(t)) \mid t \in I\} \subset \Omega$$

$$(y_\alpha, I_\alpha) < (y, I) \quad \forall \alpha.$$

Donc y est solution sur I du problème de Cauchy et la famille $\{y_\alpha, I_\alpha\}_{\alpha \in \Lambda}$ admet une borne supérieure (y, I) .

Proposition 2.3. Sous les hypothèses H_1 ou H_2 le problème de Cauchy admet des solutions maximales.

Sous l'hypothèse H_2 il existe une solution maximale unique.

Démonstration. L'existence résulte du Lemme de Zorn et du Lemme 2.2.. Il reste à démontrer l'unicité dans le cas de l'hypothèse H_2 .

Soient (y, I) et (z, J) deux solutions maximales. D'après le corollaire 2, on a $y = z$ sur $J \cap I = K$. Supposons que $I \neq J$. Soit w la solution du problème de Cauchy sur $K = I \cap J$. Alors :

$$(w, K) \underset{\neq}{>} (y, I) \quad \text{et} \quad (w, K) \underset{\neq}{>} (z, J)$$

ce qui est contradictoire avec le fait que (y, I) et (z, J) soient des solutions maximales. Donc $I = J$ et $y = z$ partout.

Résumons le théorème de Cauchy-Lipschitz et la proposition 2.3., on a le théorème fondamental suivant :

Théorème 2.4. Sous l'hypothèse H_2 le problème de Cauchy admet une solution maximale unique.

b) Critère de prolongement.

Théorème 2.5. Sous l'une des hypothèses H_1 ou H_2 , soit y une solution du problème de Cauchy sur $]a, b[$. Alors y est prolongeable au delà de b (resp. en deçà de a) si et seulement si l'une des valeurs d'adhérence de $\{t, y(t)\}$ quand $t \rightarrow b-0$ (resp. $t \rightarrow a+0$) est dans Ω .

Autrement dit, le seul cas où le prolongement n'est pas possible est celui où toutes les valeurs d'adhérence sont sur $\partial\Omega$.

Corollaire 2.6. Sous l'une des hypothèses H_1 ou H_2 et si $\Omega = \mathbb{R} \times E$ la solution y du problème de Cauchy sur $]a, b[$ est prolongeable au delà de b (resp. en deçà de a) si et seulement si :

$$\underline{\lim}_{t \rightarrow b-0} \|y(t)\| < +\infty \quad (\text{resp.} \quad \underline{\lim}_{t \rightarrow a+0} \|y(t)\| < +\infty) .$$

Autrement dit, le seul cas où y n'est pas prolongeable est celui où

$$\lim_{t \rightarrow b-0} \|y(t)\| = \infty \quad (\text{resp. } \lim_{t \rightarrow a+0} \|y(t)\| = \infty) .$$

Démonstration. (Théorème 2.5. et Corollaire 2.6.).

- Supposons y prolongeable.

Il existe donc $z : J \rightarrow E$, $J \supset]a, b[$ tel que :

$$y = z \quad \text{sur }]a, b[.$$

Pour $t \rightarrow b-0$, $y(t) = z(t) \rightarrow z(b)$ et donc $(b, z(b)) \in \Omega$.

- Inversement supposons que (b, β) , valeur d'adhérence pour $t \rightarrow b-0$, soit dans Ω .

On va montrer essentiellement que : $\beta = \lim_{t \rightarrow b-0} y(t)$.

Soit $h > 0$. h définit le voisinage \mathcal{V}_h tel que :

$$\mathcal{V}_h = \{(t, y(t)) \mid |t-b| \leq h, \|y-\beta\| \leq Mh\} .$$

On sait que, f étant continue, il existe un voisinage \mathcal{V} tel que :

$$\mathcal{V} = \{(t, \xi) \mid \|f(t, \xi)\| \leq M\} .$$

Or on a $\mathcal{V}_h \subset \mathcal{V}$ et par conséquent : $\forall (t, \xi) \in \mathcal{V}_h \quad \|f(t, \xi)\| \leq M$.

Définissons de plus : $\mathcal{W}_h = \mathcal{V}_{\frac{h}{3}} = \{(t, y(t)) \mid |t-b| \leq \frac{h}{3}, \|y-\beta\| \leq M \frac{h}{3}\}$.

\mathcal{W}_h rencontre le graphe de y :

$$\exists b_1 \quad \text{t.q.} \quad 0 \leq b-b_1 \leq \frac{h}{3} \quad \|y(b_1) - \beta\| \leq M \frac{h}{3} .$$

* Montrons que $\|y(t) - y(b_1)\| \leq \frac{2}{3} M h$ pour $t \in [b_1, b[$.

Pour $t = b_1$, c'est vérifié.

Cherchons alors l'intervalle maximum $[b_1, b_2]$ sur lequel l'inégalité a lieu.

Supposons $b_2 < b$, alors pour $t \in [b_1, b_2]$:

$$\|y(t) - \beta\| \leq \|y(t) - y(b_1)\| + \|y(b_1) - \beta\| \leq \frac{2}{3} M h + \frac{1}{3} M h = M h .$$

Or $0 \leq t - b_1 \leq b_1 - b \leq M \frac{h}{3}$. Donc : $(t, y(t)) \in \mathcal{U}_h$ pour $t \in [b_1, b_2]$ et par

conséquent : $y(b_2) - y(b_1) = \int_{b_1}^{b_2} f(s, y(s)) ds$. Donc

$$\|y(b_2) - y(b_1)\| \leq (b_1 - b_2)M \leq \frac{2}{3} M h.$$

Par continuité de y il en résulte :

$$\|y(t) - y(b_1)\| \leq \frac{2}{3} M h \quad \text{pour } t \in [b_1, b_2 + \delta[\text{ , } \delta > 0$$

ce qui est contradictoire avec la définition de b_2 . Donc $b_2 = b$ et on a

alors :

$$\|y(t) - y(b_1)\| \leq \frac{2}{3} M h \quad \text{pour } t \in [b_1, b[.$$

* Montrons que $y(t)$ est une suite de Cauchy pour $t \rightarrow b-0$. On a vu que

pour $t \in [b_1, b[$:

$$\|y(t) - \beta\| \leq \|y(t) - y(b_1)\| + \|y(b_1) - \beta\| \leq M h$$

de sorte que : $(t, y(t)) \in \mathcal{U}_h$ pour $t \in [b_1, b[$.

Alors si $t \in [b_1, b[$, $t' \in [b_1, b[$:

$$\|y(t) - y(t')\| \leq \int_{t'}^t \|f(s, y(s))\| ds \leq |t - t'| M$$

$y(t)$ est donc bien une suite de Cauchy pour $t \rightarrow b-0$. y étant à valeurs

dans E qui est complet par hypothèse, il en résulte que $y(t)$ a une limite

quand $t \rightarrow b-0$. Ce ne peut être que β .

* Montrons enfin que y est prolongeable au delà de b .

Traduisons l'existence de solution au voisinage de b :

$$\begin{aligned} \exists \delta > 0, \exists z \text{ définie sur } [b, b + \delta[\text{ t.q. } & z'(t) = f(t, z(t)) \\ & z(b) = \beta \end{aligned}$$

Définissons alors \tilde{y} comme suit :

$$\tilde{y} = \begin{cases} y & \text{sur }]a, b[\\ z & \text{sur } [b, b + \delta[\end{cases}$$

On a alors : $\tilde{y}(t) = y_0 + \int_{t_0}^t f(s, y(s)) ds$ pour $t \in]a, b+\delta[$, et donc $y \in \mathcal{C}^1(]a, b+\delta[, E)$ et \tilde{y} est solution du problème de Cauchy sur $]a, b+\delta[$; \tilde{y} est aussi un prolongement strict de y .

Cas particulier.

Supposons en plus de l'hypothèse H_1 que :

a) $\Omega = \mathbb{R} \times E$

β) f est lipschitzienne en y sur tout l'espace $\mathbb{R} \times E$:

$$\|f(t, y) - f(t, y_*)\| < L \|y - y_*\| \quad \forall t \in \mathbb{R}, \quad \forall y, y_* \in E.$$

On a alors la :

Proposition 2.7. La solution du problème de Cauchy :

$$y'(t) = f(t, y(t)), \quad y(t_0) = y_0 \text{ est définie sur tout } \mathbb{R}.$$

Démonstration.

Montrons que si y est solution sur $[t_0, t_1]$ alors $y(t)$ est borné pour $t \rightarrow t_1$, ce qui entrainera que la solution est prolongeable au delà (et en deçà) de t_1 . On a :

$$y'(t) = f(t, y(t)) \implies \|y'(t)\| = \|f(t, y(t))\| \quad (1)$$

f étant lipschitzienne en y :

$$\|f(t, y) - f(t, 0)\| \leq L \|y\| \text{ d'où}$$

$$\|f(t, y)\| \leq L \|y\| + \|f(t, 0)\| \text{ soit d'après (1) :}$$

$$\|y'(t)\| \leq L \|y(t)\| + \|f(t, 0)\|.$$

Posons $z(t) = \|y'(t)\|$. On en déduit que :

$$z'(t) \leq \|f(t, 0)\| + L z(t)$$

$$\text{soit : } z'(t) - L z(t) \leq \|f(t,0)\| .$$

Multiplions les deux membres de cette inégalité par e^{-Lt}

$$e^{-Lt}(z'(t) - L z(t)) \leq \|f(t,0)\| e^{-Lt}$$

$$\text{soit : } \frac{d}{dt}[z(t) e^{-Lt}] \leq \|f(t,0)\| e^{-Lt}$$

intégrons cette inégalité entre t_0 et s :

$$z(s) e^{-Ls} - z(t_0) e^{-Lt_0} \leq \int_{t_0}^s \|f(t,0)\| e^{-Lt} dt ,$$

$$\text{soit } z(s) e^{-Ls} \leq \|y_0\| e^{-Lt_0} + \int_{t_0}^s \|f(t,0)\| e^{-Lt} dt$$

$$\begin{aligned} \text{car } z(t_0) &= \|y(t_0)\| \\ &= \|y_0\| . \end{aligned}$$

Finalement, on a en multipliant les 2 membres par $e^{Ls} > 0$

$$z(s) \leq \|y_0\| e^{L(s-t_0)} + \int_{t_0}^s \|f(t,s)\| e^{L(s-t)} dt .$$

Faisons tendre s vers t_1-0 . On a, à la limite :

$$\begin{aligned} \lim_{s \rightarrow t_1-0} z(s) &= \|y(t_1-0)\| \leq \|y_0\| e^{L(t_1-t_0)} + \int_{t_0}^{t_1-0} \|f(t,s)\| e^{L(t_1-t_0)} dt \\ &\leq \|y_0\| e^{L(t_1-t_0)} + M \int_{t_0}^{t_1-0} e^{L(t_1-t_0)} dt \end{aligned}$$

quantité qui est finie donc $y(t)$ est borné quand $t \rightarrow t_1$.

C RESOLUTION NUMERIQUE

(Méthodes de différence finie)

1 - Les méthodes à un et plusieurs pas.

1.1) Méthodes à un pas.

Nous nous plaçons dans le cas scalaire ($E = \mathbb{R}^N$) du problème de Cauchy

$$(1,1) \quad y'(t) = f(t, y(t))$$

$$(1,2) \quad y(t_0) = y_0 \text{ donné .}$$

On suppose qu'il y a existence et unicité de y dans $[a, b]$. On introduit un

pas $h = \frac{b-a}{N}$ et l'on va d'abord "intégrer" (1,1) dans l'intervalle $(a, a+h)$

en remplaçant la courbe par sa tangente au point a , etc... Posons :

$$(1,3) \quad \begin{cases} t_j = a + jh & 0 \leq j \leq N \\ t_0 = a & t_N = b \end{cases}$$

Désignons par :

$$(1,4) \quad y_n = \text{approximation de } y(t_n).$$

Alors la méthode précédente conduit au schéma suivant (d'Euler) :

$$(1,5) \quad y_{n+1} = y_n + h f(t_n, y_n), \quad n \geq 0 ;$$

$y(t_0) = y_0$ étant connu, les équations (1,5) définissent de proche en proche y_1, y_2, \dots ; de façon générale y_{n+1} est déterminé à partir de la connaissance de y_n par (1,5) ; c'est ce que l'on appelle une méthode à un pas.

Il reste maintenant à savoir si y_n est vraiment une approximation de $y(t_n)$ et si $|y_n - y(t_n)| \rightarrow 0$ lorsque $h \rightarrow 0$, $t_N = a + Nh$ étant fixé.

Naturellement on peut espérer améliorer la précision par une modification convenable de la méthode. C'est l'objet du n° 2.

1.2) Méthodes à plusieurs pas.

Intégrons l'équation (1,1) sur l'intervalle (t_{n-p}, t_{n+1}) , p entier fixé (ceci n'a de sens que si $n \geq p$). Il vient :

$$(1,6) \quad y(t_{n+1}) - y(t_{n-p}) = \int_{t_{n-p}}^{t_{n+1}} f(t, y(t)) dt .$$

On peut ensuite calculer l'intégrale de droite en utilisant une formule de quadrature [cf. Chap. I § C] ; cela conduira par exemple à :

$$(1,7) \quad y(t_{n+1}) - y(t_{n-p}) \approx \sum_{j=n-p}^{n+1} A_j f(t_j, y(t_j)) ,$$

$$\text{ou (1,8)} \quad y(t_{n+1}) - y(t_{n-p}) \approx \sum_{j=n-p}^{n+0} B_j f(t_j, y(t_j)) .$$

L'équation (1,7) (resp. (1,8)) conduit aux équations aux différences :

$$(1,9) \quad y_{n+1} - y_{n-p} = \sum_{j=n-p}^{n+1} A_j f(t_j, y_j)$$

resp.

$$(1,10) \quad y_{n+1} - y_{n-p} = \sum_{j=n-p}^{n+0} B_j f(t_j, y_j) .$$

La détermination de y_{n+1} à l'aide de (1,9) ou (1,10) suppose connus

$u_n, u_{n-1}, \dots, u_{n+p}$, soit $p+1$ valeurs antérieures. C'est ce que l'on ap-

pelle méthode à (p+1) pas. Une différence essentielle apparaît entre (1,9)

et (1,10) ; la formule (1,10) donne explicitement y_{n+1} au moyen de

y_n, \dots, y_{n-p} : c'est une méthode explicite.

Par contre, la formule (1,9) ne donne y_{n+1} qu'implicitement ou après la résolution d'une équation (dont il faut montrer la possibilité). C'est une méthode implicite.

Remarque 1.1. Dans chacune des méthodes (1,9) (1,10) à (p+1) pas, apparaît la difficulté du démarrage. Seul y_0 étant donné on ne peut calculer y_1 au moyen de (1,9) ou (1,10) ; ces formules ne sont applicables qu'une fois y_0, \dots, y_p en notre possession. Il faut donc utiliser d'autres méthodes pour calculer y_1, \dots, y_p .

Remarque 1.2. Tout ce qui vient d'être dit s'adapte sans difficulté au cas des systèmes d'équations différentielles.

2 - Méthodes à un pas.2.1) Méthode d'Euler modifiée.

Reprenons (1,5) ; on peut l'obtenir comme suit : partant de la relation

$$y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} f(t, y(t)) dt.$$

On utilise la formule de la moyenne :

$$(2,1) \quad y(t_{n+1}) - y(t_n) = h f(\bar{t}_n, y(\bar{t}_n)), \quad t_n < \bar{t}_n < t_{n+1}.$$

La méthode d'Euler consiste à remplacer \bar{t}_n par t_n d'où (1,5). Si on rem-

place \bar{t}_n par $t_n + \frac{h}{2}$ et $y(t_n + \frac{h}{2})$ par sa valeur approchée calculée en

(1,5) à savoir :

$$y_n + \frac{1}{2} h f(t_n, y_n)$$

il vient

$$(2,2) \quad y_{n+1} - y_n = h f\left(t_n + \frac{h}{2}, y_n + \frac{1}{2} h f(t_n, y_n)\right).$$

On ajoute à ce système (explicite) la condition initiale

$$(2,3) \quad y_0 \text{ connu.}$$

2.2) Méthode générale à un pas.

Soit $\Phi(t, y ; h)$ une fonction donnée pour $t \in [a, b]$, $y \in \mathbb{R}$, $h > 0$

et que nous supposerons toujours au moins continue en ces 3 variables. Alors

les schémas (1,5) et (2,2) entrent tous deux dans le schéma général suivant :

$$(2,4) \quad y_{n+1} = y_n + h \Phi(t_n, y_n ; h).$$

Dans le cas d'Euler on a :

$$\Phi(t, y ; h) = f(t, y) \quad (\text{donc indépendant de } h!);$$

dans le cas d'Euler modifié (2,2) on a :

$$\Phi(t, y ; h) = f\left(t + \frac{1}{2} h, y + \frac{h}{2} f(t, y)\right).$$

La condition minimum à supposer sur Φ est :

$$(2,7) \quad \Phi(t, y ; 0) = f(t, y) .$$

En effet il faut imposer que $y_n \rightarrow y(t_n)$ et $\frac{y_{n+1} - y_n}{h} \rightarrow y'(t_n)$ lorsque $h \rightarrow 0$, $t_n = a + N h$ fixé ; d'après (2,4) et puisque $y'(t_n) = f(t_n, y(t_n))$ il faut donc que :

$$\Phi(t_n, y(t_n) ; 0) = f(t_n, y(t_n))$$

et ceci quelle que soit la donnée initiale y_0 , d'où (2,7) .

Notre problème est maintenant le suivant : comment choisir $\Phi(t, y ; h)$.

Nous allons examiner ce problème à partir de l'étude de la convergence et de l'estimation de l'erreur.

2.3) Estimation de l'erreur.

On introduit l'erreur de discrétisation

$$(2,8) \quad e_n = y_n - y(t_n) .$$

Nous allons démontrer la

Proposition 2.1. On suppose que les conditions suivantes ont lieu :

$$(i) \quad \Phi \text{ satisfait à (2,7), } \Phi(t, y ; 0) = f(t, y)$$

$$(ii) \quad \left\{ \begin{array}{l} |\Phi(t, y ; h) - \Phi(t, y^* ; h)| \leq L_1 |y - y^*| \\ \text{pour } t \in [a, b], y, y^* \in \mathbb{R} \text{ (ou dans un compact convenable)} \\ \text{et pour } 0 \ll h \leq h_0 \text{ (} \ll b-a \text{)} ; \end{array} \right.$$

$$(iii) \quad \left\{ \begin{array}{l} |\Phi(t, y(t) ; h) - \frac{y(t+h) - y(t)}{h}| = \\ |\Phi(t, y(t) ; h) - \frac{1}{h} \int_t^{t+h} f(\xi, y(\xi)) d\xi| \leq L_2 h^p \\ \text{pour toute solution } y \text{ régulière de } y' = f(t, y(t)). \end{array} \right.$$

Conclusion. Il existe une constante K telle que

$$(2,9) \quad |e_n| \leq K h^p .$$

Démonstration.

$$1/ \text{ De (2,4) et de la formule : } y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} f(t, y(t)) dt$$

on déduit par soustraction :

$$\begin{aligned} e_{n+1} &= e_n + h[\Phi(t_n, y_n ; h) - \frac{1}{h} \int_{t_n}^{t_{n+1}} f(t, y(t)) dt] \\ &= e_n + h[\Phi(t_n, y_n ; h) - \Phi(t_n, y(t_n) ; h)] \\ &\quad + h[\Phi(t, y(t_n) ; h) - \frac{1}{h} \int_{t_n}^{t_{n+1}} f(t, y(t)) dt] . \end{aligned}$$

Utilisant (ii) pour le 1^{er} crochet et (iii) pour le 2^e, on en déduit que :

$$(2,10) \quad |e_{n+1}| \leq |e_n| + h L_1 |e_n| + L_2 h^{p+1} .$$

2/ Pour simplifier l'écriture posons

$$|e_n| = \varepsilon_n .$$

Alors (2,10) donne

$$(2,11) \quad \varepsilon_{n+1} \leq a \varepsilon_n + b$$

avec (2,12) $a = (1+h L_1)$, $b = L_2 h^{p+1}$.

(2,11) est une inégalité récurrente linéaire. On en déduit donc :

$$(2,13) \quad \varepsilon_n \leq a^n \varepsilon_0 + b \frac{a^n - 1}{a - 1} .$$

Mais $\varepsilon_0 = 0$ et

$$\frac{a^n - 1}{a - 1} = \frac{1}{h L_1} [e^{n \log(1+h L_1)} - 1] \leq \frac{1}{h L_1} [e^{n h L_1} - 1] ;$$

or $n h = t_n - a$, donc :

$$(2,14) \quad \varepsilon_n \leq \frac{L_2}{L_1} (e^{(t_n - a) L_1} - 1)$$

Corollaire 2.1. Sous les hypothèses (i), (ii), (iii), la méthode (2,4) est

convergente et l'erreur de discrétisation est $O(h^p)$.

Remarque 2.1. Supposons que l'on ait un renseignement plus précis que (iii) :

$$(2,16) \quad \Phi(t, y(t); h) - \frac{y(t+h) - y(t)}{h} = h^p \varphi(x) + O(h^{p+1})$$

φ étant continue non nulle dans $[a, b]$. (p est alors l'ordre exact ; la conclusion du corollaire 2.1. est en effet que l'ordre est au moins p).

On peut alors démontrer ceci : soit $V(t)$ la solution de l'équation

$$(2,17) \quad V'(t) = \frac{\partial f}{\partial y}(t, y(t)) V(t) + \varphi(x)$$

avec

$$(2,18) \quad V(a) = 0 \quad .$$

Alors

$$(2,19) \quad e_n = h^p V(t_n) + O(h^{p+1}) \quad .$$

Naturellement V dépend de l'inconnue y . Mais on peut aussi calculer V en utilisant les approximations de y .

2.4) Exemples : retour aux méthodes d'Euler.

1/ Dans le cas (2,5)

$$\begin{aligned} \Phi(t, y(t); h) - \frac{y(t+h) - y(t)}{h} &= f(t, y(t)) - \frac{y(t+h) - y(t)}{h} \\ &= y'(t) - \frac{y(t+h) - y(t)}{h} = -\frac{h}{2} y''(t) + O(h^2) \end{aligned}$$

d'où d'après le corollaire 2.1. la convergence de la méthode d'Euler (avec l'erreur $O(h)$).

2/ Dans le cas (2,6)

$$\begin{aligned} \Phi(t, y(t); h) &= f\left(t + \frac{1}{2}h, y(t) + \frac{h}{2}f(t, y(t))\right) \\ &= f(t, y(t)) + \frac{h}{2} \frac{\partial f}{\partial t}(t, y(t)) + \frac{h}{2} f(t, y(t)) \frac{\partial f}{\partial y}(t, y(t)) + O(h^2) \\ &= y'(t) + \frac{h}{2} y''(t) + O(h^2) \quad . \end{aligned}$$

D'où
$$\Phi(t, y(t); h) - \frac{y(t+h) - y(t)}{h} = O(h^2).$$

Il y a donc effectivement amélioration par rapport à 1/.

2.5) Construction de $\Phi(t, y; h)$ par utilisation de la série de Taylor.

Un moyen systématique de construire des $\Phi(t, y; h)$ satisfaisant à (iii).

Proposition 2.1. est évidemment de développer $\frac{y(t+h) - y(t)}{h}$ en série de Taylor,

en y remplaçant ensuite y' par $f(t, y(t))$ etc... .

On a :
$$\frac{y(t+h) - y(t)}{h} = y'(t) + \frac{h}{2} y''(t) + \dots + \frac{h^{p-1}}{(p-1)!} y^{(p)}(t) + O(h^p).$$

Mais $y'(t) = f(t, y(t))$,

d'où

$$(2,20) \quad y''(t) = \frac{\partial f}{\partial t}(t, y(t)) + y'(t) \frac{\partial f}{\partial y}(t, y(t)) = \frac{\partial f}{\partial k}(t, y(t)) + f(t, y) \frac{\partial f}{\partial y}(t, y)$$

et ainsi de suite. Si l'on pose :

$$(2,21) \quad \begin{aligned} y^{(k)}(t) &= \mathcal{F}_k(t, y(t)) \\ y^{(k+1)}(t) &= \frac{d \mathcal{F}_k}{dt} = \frac{\partial \mathcal{F}_k}{\partial t}(t, y(t)) + \frac{\partial \mathcal{F}_k}{\partial y}(t, y(t)) \cdot y'(t) \\ &= \frac{\partial \mathcal{F}_k}{\partial t}(t, y(t)) + \frac{\partial \mathcal{F}_k}{\partial y}(t, y(t)) \cdot f(t, y(t)) \\ &= \mathcal{F}_{k+1}(t, y(t)). \end{aligned}$$

On peut alors prendre

$$(2,22) \quad \Phi(t, y; h) = f(t, y) + \frac{h}{2} \mathcal{F}_1(t, y) + \dots + \frac{h^{p-1}}{(p-1)!} \mathcal{F}_{p-1}(t, y).$$

Avec le choix (2,22) la formule (2,4) donne une méthode convergente avec erreur de discrétisation en $O(h^p)$.

2.6) Construction de $\Phi(t, y; h)$ par utilisation indirecte de la série de Taylor ; méthode de Runge-Katta.

Le point 2/ de 2.4) donne (à propos de la méthode d'Euler améliorée) un

usage différent (et moins direct) de la méthode de Taylor ; cette possibilité a été systématiquement étudiée par Runge-Katta. Considérons ici un cas simple mais typique.

Par analogie avec (2,6) on cherche $\Phi(t, y ; h)$ sous la forme :

$$(2,23) \quad \Phi(t, y ; h) = a_1 f(t, y) + a_2 f(t + p_1 h, y + p_2 h f(t, y)),$$

les constantes a_i , p_i étant à choisir "au mieux".

La condition (i), proposition 2.1. donne :

$$(2,24) \quad a_1 + a_2 = 1 .$$

Développons, en remplaçant y par $y(t)$ et utilisant $y'(t) = f(t, y(t))$:

$$\Phi(t, y(t) ; h) = f(t, y(t)) + a_2 h [p_1 \frac{\partial f}{\partial t}(t, y) + p_2 f(t, y) \frac{\partial f}{\partial y}(t, y)] + o(h^2) ,$$

identifiant à

$$\begin{aligned} \frac{y(t+h) - y(t)}{h} &= y'(t) + \frac{h}{2} y''(t) + o(h^2) \\ &= f(t, y(t)) + \frac{h}{2} \left[\frac{\partial f}{\partial t}(t, y) + f \frac{\partial f}{\partial y}(t, y) \right] + o(h^2) . \end{aligned}$$

On voit que (2,23) définit une méthode convergente, en $o(h^2)$, si, outre (2,24)

on a :

$$a_2 p_1 = \frac{1}{2} , \quad a_2 p_2 = \frac{1}{2} .$$

Si donc l'on pose : $a_2 = \alpha \neq 0$, il vient :

$$a_1 = 1 - \alpha , \quad p_1 = \frac{1}{2\alpha} , \quad p_2 = \frac{1}{2\alpha} , \text{ d'où}$$

$$(2,25) \quad \Phi(t, y ; h) = (1-\alpha)f(t, y) + \alpha f\left(t + \frac{h}{2\alpha}, y + \frac{h}{2\alpha} f(t, y)\right) .$$

Pour ce choix de Φ , la méthode (2,24) est dite de Runge-Katta (simplifiée).

Exemple 2.1. Si l'on prend $\alpha = 1$, on retrouve (2,6).

Remarque 2.2. Il est maintenant facile de poser systématiquement le problème.

On introduit

$$k_0 = f(t, y)$$

$$k_1 = f(t + \lambda_1 h, y + \mu_{10} h k_0)$$

$$k_2 = f(t + \lambda_2 h, y + \mu_{20} h k_0 + \mu_{21} h k_1)$$

.....

$$k_p = f(t + \lambda_p h, y + \mu_{p0} h k_0 + \dots + \mu_{p,p-1} h k_{p-1})$$

et on cherche à déterminer les a_i, λ_i, μ_i de façon que

$$(2,26) \quad \Phi(t, y ; h) = a_0 k_0 + a_1 k_1 + \dots + a_p k_p$$

de façon que l'ordre de l'erreur soit le plus élevé possible.

Exemple 2.2. On trouvera par exemple

$$(2,27) \quad \Phi(t, y ; h) = \frac{1}{6} [k_0 + 2k_1 + 2k_2 + k_3]$$

avec

$$k_1 = f(t + \frac{1}{2} h, y + \frac{1}{2} h k_0)$$

$$k_2 = f(t + \frac{1}{2} h, y + \frac{1}{2} h k_1)$$

$$k_3 = f(t + h, y + h k_2) .$$

On trouve que cette méthode est d'ordre 4 .

Remarque 2.3. On peut utiliser des processus d'accélération de convergence ;

cf. chapitre I § C (3) .

Remarque 2.4. Dans les équations différentielles intervenant dans les applications, on sait souvent que la solution est régulière dans un intervalle (approximativement connu) puis irrégulière dans un autre intervalle. Il est alors peu naturel d'utiliser le même pas h tout au long du calcul. On introduit alors des pas variables, t_{n+1} étant défini à partir de t_n par :

$$t_{n+1} = t_n + \theta(t_n) h$$

la fonction $t \mapsto \theta(t)$ étant continue et telle que $\theta(t) \geq \theta_0 > 0$. On remplace alors (2,4) par

$$y_{n+1} = y_n + \theta(t_n) h \Phi(t_n, y_n; \theta(t_n) h).$$

3 - Méthodes à plusieurs pas. Utilisation des formules de quadrature.

3.1) Formule de Newton.

1/ On a rappelé au chapitre I § A, la notion de polynôme d'interpolation, les x_j étant des points distincts croissants sur $[a, b]$; $0 \leq j \leq q$, c'est le polynôme P tel que :

$$(3,1) \quad \begin{cases} P(x_j) = f(x_j) = f_j & (f \text{ donnée}) \text{ pour } 0 \leq j \leq q \\ d^0 P \leq q \end{cases}$$

Lorsque les x_j sont uniformément répartis sur $[a, b]$

$$(3,2) \quad x_j = a + j h, \quad h = \frac{b-a}{N}.$$

On a vu qu'on pouvait donner une expression simple de P en fonctions des différences progressives de la suite $\{f(x_j)\}$ (Cf. Chapitre I, § A proposition 3.4.).

On peut donner une expression analogue de P en utilisant cette fois les différences régressives de la suite $\{f(x_j)\}$.

2/ Opérateur $\bar{\nabla}$.

Soit $\{f_j\}$ une suite donnée. On pose :

$$(3,3) \quad \bar{\nabla} f_j = f_j - f_{j-1};$$

$\bar{\nabla}$ est l'opérateur de différence régressive. On définit ses puissances successives $\bar{\nabla}^2, \bar{\nabla}^3, \dots$; par définition $\bar{\nabla}^0 = \text{identité}$: $\bar{\nabla}^0 f_j = f_j$.

On trouve (par récurrence comme au chapitre I) :

$$\bar{\nabla}^q f_j = \sum_{m=0}^q (-1)^m \binom{q}{m} f_{j-m} .$$

On vérifiera que :

$$f_{j-q} = \sum_{m=0}^q (-1)^m \binom{q}{m}^{-m} f_j .$$

Remarque 3.1. Si f_j n'est pas défini pour tout j , mais seulement pour $j \geq 0$ (où $0 \leq j \leq N$), alors $\bar{\nabla} f_j$ n'est défini que pour $j \geq 1$, $\bar{\nabla}^2 f_j$ pour $j \geq 2$ etc... .

3/ Formule de Newton.

Une notation d'abord ; q et m étant entier, on sait que :

$$\binom{q}{m} = \frac{q(q-1)\dots(q-m+1)}{1.2\dots m} .$$

Cette définition s'étend au cas où $q \in \mathbb{R}$ (ou $\in \mathbb{C}$) quelconque. Ceci posé :

Proposition 3.1. Le polynôme $P(x)$ qui vérifie :

$$(3,6) \quad \begin{cases} d^0 P \leq q \\ P(x_j) = z_j , \quad p-q \leq j \leq p \quad (x_j = a + jh) \end{cases}$$

s'exprime par la formule :

$$P(x) = \sum_{m=0}^q (-1)^m \binom{-s}{m}^{-m} z_p$$

$$\text{où : (3,8) } s = \frac{x-x_p}{h} .$$

Démonstration.

Se reportant à la définition de $\binom{q}{m}$ pour $q \in \mathbb{R}$ et avec (3,8), on voit que $\binom{-s}{m}$ est un polynôme en x de degré m et donc que le 2^e membre de (3,7)

est un polynôme de degré $\leq q$.

Puis pour $0 \leq r \leq q$:

$$P(x_{p-r}) = \sum_{m=0}^q (-1)^m \binom{r}{m} \bar{\nabla}^m z_p .$$

Mais $\binom{r}{m} = 0$ si $r < m$; donc

$$P(x_{p-r}) = \sum_{m=0}^r (-1)^m \binom{r}{m} \bar{\nabla}^m z_p ,$$

et d'après (3,5) ceci vaut z_{p-r} .

3.2) Utilisation de la formule de quadrature

Soit $y(t)$ solution de (1,1) :

$$(3,9) \quad y'(t) = f(t, y(t)).$$

On opère comme pour (1,6) en généralisant un peu. On déduit de (3,9) pour

$$t, t+k \in [a, b]$$

$$(3,10) \quad y(t+k) - y(t) = \int_t^{t+k} f(\xi, y(\xi)) d\xi .$$

On calcule approximativement le deuxième membre de (3,10) en remplaçant

$f(\xi, y(\xi))$ par le polynôme $P(f ; \xi) = P(\xi)$ tel que :

$$(3,11) \quad \begin{cases} d^0 P \leq q \\ P(t_j) = f(t_j, y_j) \quad p-q \leq j \leq p \end{cases}$$

y_j ayant déjà été calculé ou que l'on est éventuellement en train de calculer si $j = p$. On pose :

$$(3,12) \quad f_j = f(t_j, y_j) .$$

Alors d'après la proposition 3.1.

$$(3,13) \quad P(x) = \sum_{m=0}^q (-1)^m \binom{-s}{m} \bar{\nabla}^m f_p , \quad s = \frac{x-x_p}{h} .$$

On a alors le choix de t , $t+k$ et q .

Laisant q arbitraire pour l'instant, les choix classiques sont :

$$(3,14) \quad \left\{ \begin{array}{l} t = t_p, \quad t+k = t_{p+1} ; \text{ méthode d'Adams-Bashforth} ; \\ t = t_{p-1}, \quad t+k = t_p ; \text{ méthode d'Adams-Moulton} ; \\ t = t_{p-1}, \quad t+k = t_{p+1} ; \text{ méthode de Nyström} ; \\ t = t_{p-2}, \quad t+k = t_p ; \text{ méthode de Milne-Simpson} ; \end{array} \right.$$

→ La méthode d'Adams-Bashforth donne :

$$(3,15) \quad y_{p+1} - y_p = h \sum_{m=0}^q \gamma_m \bar{\nabla}^m f_p$$

où

$$(3,16) \quad \gamma_m = (-1)^m \int_0^1 \binom{-s}{m} ds .$$

En effet avec le choix indiqué dans (3,14), la formule (3,10) donne :

$$\gamma_m = \frac{1}{h} \int_{t_p}^{t_{p+1}} (-1)^m \binom{-\frac{\xi-t_p}{h}}{m} d\xi$$

d'où (3,16) en posant $\xi - t_p = hs$.

→ La méthode d'Adams-Moulton donne :

$$(3,17) \quad y_p - y_{p-1} = h \sum_{m=0}^q \gamma_m^* \bar{\nabla}^m f_p$$

$$\text{où } (3,18) : \quad \gamma_m^* = (-1)^m \int_{-1}^0 \binom{-s}{m} ds .$$

→ La méthode Nyström donne :

$$(3,19) \quad y_{p+1} - y_{p-1} = h \sum_{m=0}^q \chi_m \nabla^m f_p .$$

où

$$(3,20) \quad \chi_m = (-1)^m \int_{-1}^{+1} \binom{-s}{m} ds .$$

→ La méthode Milne-Simpson donne :

$$(3,21) \quad y_p - y_{p-2} = h \sum_{m=0}^q \chi_m^* \bar{\nabla}^m f_p$$

où

$$(3,22) \quad \chi_m^* = (-1)^m \int_{-2}^0 \binom{-s}{m} ds .$$

Les problèmes sont maintenant les suivants :

a) un problème technique : trouver des procédés de calcul commode (par récurrence) des coefficients $\gamma_m, \gamma_m^* \dots$; nous allons donner un exemple au point suivant.

b) un problème de comparaison des méthodes, leur intérêt (éventuel), le choix de q, \dots etc. Ceci fait l'objet du n° 4 ci-dessous.

3.3) Fonctions génératrices.

Pour établir des relations de récurrence (éventuelles) entre les γ_m , on associe à γ_m la fonction

$$(3,23) \quad \Gamma(t) = \sum_{m=0}^{\infty} \gamma_m t^m \quad t \in \mathbb{C} ,$$

(sous réserve de vérification de convergence).

Il s'agit là d'une méthode générale souvent utile.

Remplaçant γ_m par sa valeur (3,16) dans (3,23) il vient :

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

on peut vérifier qu'il est loisible, pour $|t| < 1$ de commuter la série et l'intégrale, d'où :

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

d'où

$$(3,24) \quad \Gamma(t) = \frac{-t}{(1-t)\log(1-t)} \quad |t| < 1 .$$

Naturellement, un développement en série de (3,24) reconduirait purement et simplement à (3,23) ; mais écrivons (3,24) sous la forme

$$(3,24 \text{ bis}) \quad -\frac{1}{t} \log(1-t) \Gamma(t) = \frac{1}{1-t}$$

et utilisons les développements :

$$\frac{1}{1-t} = \sum_{n \geq 0} t^n, \\ -\frac{1}{t} \log(1-t) = \sum_{n \geq 0} \frac{t^n}{n+1}$$

portant dans (3,24 bis) on obtient l'identité :

$$(3,25) \quad \left(\sum_{n \geq 0} \frac{t^n}{n+1} \right) \left(\sum_{m \geq 0} \gamma_m t^m \right) = \sum_{n \geq 0} t^n,$$

d'où les relations (par identification) :

$$(3,26) \quad \gamma_m + \frac{1}{2} \gamma_{m-1} + \frac{1}{3} \gamma_{m-2} + \dots + \frac{1}{m+1} \gamma_0 = 1, \quad m = 0, 1, 2, \dots$$

d'où la possibilité de calculer les γ_m par récurrence.

Remarque 3.2. Soient Γ^* , χ , χ^* les fonctions génératrices de γ_m^* , χ_m , χ_m^*

définies par (3,18), (3,20), (3,22) ; on trouve :

$$\Gamma^*(t) = -t \frac{1}{\log(1-t)}, \quad |t| < 1 \\ \chi(t) = -t \frac{2-t}{1-t} \frac{1}{\log(1-t)}, \quad |t| < 1 \\ \chi^*(t) = -t (2-t) \frac{1}{\log(1-t)}, \quad |t| < 1.$$

4. - Etude générale des méthodes à plusieurs pas. Stabilité.

4.1) Formulation générale.

Tous les schémas introduits au n° 3 entrent dans le type suivant :

$$(4,1) \quad \alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \dots + \alpha_0 y_n = h [\beta_k f_{n+k} + \beta_{k-1} f_{n+k-1} + \dots + \beta_0 f_n]$$

où les α_i , β_i sont des nombres réels donnés,

$$f_j = f(t_j, y_j) \quad t_j = a + jh, \quad h = \frac{b-a}{N}.$$

On supposera :

$$(4,2) \quad \alpha_k \neq 0$$

et (4,3) $|\alpha_0| + |\beta_0| \neq 0$.

(De sorte que y_n intervient vraiment dans (4,1)).

Le schéma (4,1) est le schéma général à k pas.

Remarque 4.1. Comme on a déjà signalé (remarque 1.1.) il y a une difficulté au démarrage puisque l'on ne peut espérer obtenir y_{n+k} par (4,1) sans connaître y_{n+k-1}, \dots, y_n . Nous reviendrons sur ce point.

. Méthode explicite. Si $\beta_k = 0$, (4,1) est explicite cf. déjà (1,2) n° 1 :

. Méthode implicite. Si $\beta_k \neq 0$ (4,1) est implicite ; on obtient (éventuellement) y_{n+k} par la résolution de :

$$(4,4) \quad \alpha_k y_{n+k} - h \beta_k f(t_{n+k}, y_{n+k}) = \text{donné} = e.$$

Posons un instant $y_{n+k} = y$. $\frac{h \beta_k}{\alpha_k} f(t_{n+k}, y) = F(y)$.

Alors (4,4) s'écrit :

$$(4,5) \quad y = F(y) + d = G(y), \quad d = \frac{e}{\alpha_k}.$$

Utilisant le fait que f est L -lipschitzienne, on a :

$$|G(y) - G(y^*)| \leq h \left| \frac{\beta_k}{\alpha_k} \right| L |y - y^*|$$

d'où (cf. cours d'Analyse) :

$$(4,6) \quad \begin{cases} \text{si } \beta_k \neq 0, \text{ la formule (4,1) définit } y_{n+k} \text{ de} \\ \text{façon unique si } h \left| \frac{\beta_k}{\alpha_k} \right| L > 1. \end{cases}$$

Il faudra donc choisir h assez petit pour que cette condition ait lieu.

Les problèmes sont maintenant les suivants : comment choisir "au mieux"

les coefficients α_j, β_j dans (4,1) ?

4.2) Définition de la convergence.

Introduisons les polynômes (4,7)
$$\begin{cases} \alpha(\lambda) = \alpha_0 + \alpha_1 \lambda + \dots + \alpha_k \lambda^k \\ \beta(\lambda) = \beta_0 + \beta_1 \lambda + \dots + \beta_k \lambda^k \end{cases}$$

et désignons par (α, β) le schéma (4,1).

A toute fonction $f(x,y) = f$ vérifiant (1,3) on associe :

a) le problème de Cauchy $y'(t) = f(t, y(t)), y(t_0) = y_0$.

b) le problème "approché" (4,1) [avec (4,6) si $\beta_k \neq 0$] avec les données

initiales.

$$(4,8) \quad y_j = \eta_j(h) \quad 0 \leq j \leq k-1$$

les η_j satisfaisant à la condition (minimum) :

$$(4,9) \quad \eta_j(h) \rightarrow y_0 \text{ lorsque } h \rightarrow 0, \quad 0 \leq j \leq k-1 .$$

N.B. On ne prend pas nécessairement y_0 coïncidant avec la donnée de Cauchy.

Définition 4.1. Le schéma (α, β) est dit convergent si, sous la conditions précédentes (et pour toute f), on a :

$$(4,10) \quad y_n - y(t_n) \rightarrow 0 \text{ lorsque } h \rightarrow 0, \quad t_n = a + nh \text{ fixé } \leq b .$$

∑ N.B. La convergence, dans (4,10) n'est pas uniforme en f .

Notre but est maintenant de donner des conditions nécessaires et suffisantes sur les polynômes α, β pour qu'il y ait convergence.

On va commencer par l'étude (importante en elle-même) des équations aux différences.

4.3.) Equations linéaires aux différences. Leur stabilité.

Sur l'espace S des suites $y = \{y_n\}$, $n \geq 0$, on considère l'opérateur

linéaire Λ_α défini par :

$$(\Lambda_\alpha y)_n = \alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \dots + \alpha_0 y_n, \quad \alpha_k \neq 0$$

$$\Lambda_\alpha y = \{(\Lambda_\alpha y)_n\}.$$

Définition 4.2. L'opérateur Λ_α -et, par abus de langage, le polynôme α -

est dit stable, si toute suite $\{y_n\}$ solution de :

$$(4,11) \quad \Lambda_\alpha y = 0 \text{ est } \underline{\text{bornée}}.$$

Montrons là

Proposition 4.1. La condition nécessaire et suffisante de stabilité de l'opé-

rateur Λ_α est que les deux conditions suivantes aient lieu :

(i) les racines du polynôme α sont dans le disque $|\lambda| < 1$

(ii) les racines éventuellement situées sur ce cercle $|\lambda| = 1$ sont

simples.

Démonstration.

1/ Les solutions de (4,11) forment un sous-espace de S de dimension k .

Soit en effet $E^j = \{E_n^j\}$, $0 \leq j \leq k-1$ la solution de :

$$(4,12) \quad \begin{cases} \Lambda_\alpha E^j = 0 \\ E_m^j = \delta_m^j, \quad 0 \leq m \leq k-1 \end{cases}$$

Alors y solution quelconque de (4,11) s'exprime par

$$y = \sum_{j=0}^{k-1} y_j E^j \text{ d'où le résultat.}$$

2/ Solutions particulières.

La suite $\{\lambda^n\}$, $\lambda \in \mathbb{C}$ est solution de (4,11) si et seule-

ment si λ est racine d'ordre $r+1$ de α (faire le calcul de $(\Lambda_\alpha y)_n$),

$y_n = n^r \lambda^n$) et alors $\{\lambda^n\}, \{n \lambda^n\}, \dots, \{n^r \lambda^n\}$ sont solutions linéairement indépendantes.

Soient alors λ_j les racines de $\alpha(\lambda) = 0$, λ_j étant d'ordre m_j , $0 \leq j \leq v$, $\sum_{j=1}^v m_j = k$. Les suites

$$(4,13) \quad \{n^{r_j} \lambda_j^n\}, \quad 0 \leq r_j \leq m_j, \quad 1 \leq j \leq v$$

sont linéairement indépendantes, en nombre k , et donc forment une base de l'espace des solutions de (4,11).

3/ La condition nécessaire et suffisante de stabilité est alors que toute suite (4,13) soit bornée, d'où la proposition.

Remarque 4.3. Ecrivons explicitement la solution de l'équation non homogène :

$$(4,14) \quad \alpha_k y_{n+k} + \dots + \alpha_0 y_n = d_{n+k}, \quad n \geq 0$$

$$(4,15) \quad y_j = \eta_j, \quad 0 \leq j \leq k.$$

Introduisons la solution élémentaire $E = \{E_n\}$, $n \in \mathbb{Z}$, avec :

$$E_n = E_n^{k-1} \quad \text{si } n \geq 0$$

$$E_n = 0 \quad \text{si } n < 0.$$

$$\begin{aligned} \text{Alors } \sum_{j=0}^k \alpha_j E_{n+j} &= 0 & \text{si } n \geq 0 \\ &= \alpha_k & \text{si } n = -1 \\ &= 0 & \text{si } n < -1 \quad [n+j < k]. \end{aligned}$$

Donc :

$$\sum_{j=0}^k \alpha_j E_{n+j} = \begin{cases} 0 & \text{si } n \neq -1 \\ \alpha_k & \text{si } n = -1 \end{cases}$$

Définissons alors $E \otimes d$ par :

$$(4,17) \quad (E \otimes d)_n = \sum_{m=0}^{\infty} E_{n-m-1} d_{m+k} \quad \text{si } n > 0 .$$

Alors, comme $E_{n-m-1} = 0$ dès que $n-m-1 < k-1$ i.e. $m > n-k$

$$(E \otimes d)_n = \sum_{m=0}^{n-k} E_{n-m-1} d_{m+k} \quad \text{si } n \geq k$$

$$= 0 \quad \text{si } n < k .$$

La solution de (4,14), (4,15) est alors fournie par :

$$(4,18) \quad y = \sum_{j=0}^{k-1} \eta_j E^j + \frac{1}{\alpha_k} E \otimes d .$$

En effet $(E \otimes d)_n = 0$ si $0 \leq n \leq k-1$ est la seule chose à vérifier est que pour $n \geq 0$

$$\xi_n = \sum_{j=0}^k \alpha_j (E \otimes d)_{n+j} \quad \text{est égal à } \alpha_k d_{n+k}$$

or

$$\xi_n = \sum_{m=0}^{\infty} d_{m+k} \left(\sum_{j=0}^k \alpha_j E_{n+j-m-1} \right),$$

et d'après (4,16)

$$\sum_{j=0}^k \alpha_j E_{n+j-m-1} = 0 ,$$

pour tout $n \neq m$ et vaut α_k si $n = m$ d'où (4,18).

Proposition 4.2. Si le polynôme α est stable, il existe une constante K_α

telle que, y_n étant la solution de (4,14), (4,15), on a :

$$(4,19) \quad |y_n| \leq K_\alpha \left(\sum_{j=0}^{k-1} |\eta_j| + \sum_{m=0}^{n-k} |d_{m+k}| \right), \quad (n \geq k) .$$

Démonstration On peut en effet expliciter (4,18) sous la forme

$$y_n = \sum_{j=0}^{k-1} \eta_j E_n^j + \frac{1}{\alpha_k} \sum_{m=0}^{n-k} E_{n-m-1}^{k-1} d_{m+k}$$

d'où (4,19) grâce au fait que $\{E^j\}$, $0 \leq j \leq k-1$, est bornée.

4.4) Conditions nécessaires de convergence.

Proposition 4.3. Pour que le schéma (α, β) soit convergent, il faut que le polynôme α soit stable.

Démonstration.

Si le polynôme α n'est pas stable il existe une suite $Z = \{Z_n\}_n$ vérifiant

$$* \Lambda_{\alpha} Z = 0$$

$$* \lim_{n \rightarrow \infty} Z_n = \infty .$$

Considérons un point x et des valeurs de h telles que :

$$h_{\lambda} = \frac{x}{\lambda} \quad \lambda \in \mathbb{N} .$$

Considérons la suite $y_n = \frac{Z_n}{\sqrt{|Z_{\lambda}|}}$ lorsque h tend vers zéro

$$y_i \rightarrow 0 \quad \forall : i = 0, 1, \dots, k-1 .$$

Mais $y_{\lambda} = \sqrt{|Z_{\lambda}|}$ tend vers l'infini.

La valeur de la "solution approchée" au point x ne tend donc pas vers zéro.

Si on applique ceci au problème de Cauchy

$$\begin{cases} y' = 0 \\ y(0) = 0 \end{cases}$$

on trouve une contradiction. La méthode n'étant pas convergente.

Démonstration. (suite)

1/ Considérons l'équation particulière

$$y' = 0, \quad y(0) = 1$$

de solution $y(x) = 1$. Le schéma (α, β) se réduit à :

$$(4,22) \quad \sum_{j=0}^k \alpha_j y_{n+j} = 0; \quad y_j = \eta_j(h) \quad 0 \leq j \leq k-1$$

avec $\eta_j(h) \rightarrow 1$.

Prenons en particulier

$$\eta_j(h) = 1, \quad 0 \leq j \leq k-1.$$

La solution y_n ne dépend pas de h et la condition de convergence se traduit par $y_n \rightarrow 1$ lorsque $n \rightarrow \infty$; passant à la limite en n dans (4,22),

il vient :

$$\sum_{j=0}^k \alpha_j = \alpha(1) = 0.$$

2/ D'après la proposition 4.3., le polynôme α doit être stable.

Donc la racine 1 est nécessairement simple (proposition 4.1.), donc :

$$\alpha'(1) \neq 0.$$

3/ Considérons maintenant l'équation particulière :

$$y' = 1, \quad y(0) = 0$$

de solution $y(t) = t$. Le schéma (α, β) se réduit à :

$$(4,23) \quad \sum_{j=0}^k \alpha_j y_{n+j} = h (\beta_k + \beta_{k-1} + \dots + \beta_0) = h \beta(1).$$

Choisissons

$$(4,22) \quad \eta_j(h) = \frac{\beta(1)}{\alpha'(1)} j h \quad (\text{loisible d'après 2/}).$$

Alors la solution de (4,23), (4,24) est :

$$y_n = n h \frac{\beta(1)}{\alpha'(1)}$$

et $y(t_n) = n h ;$

la convergence de y_n vers $y(t_n)$ équivaut donc à $\frac{\beta(1)}{\alpha'(1)} = 1$, d'où la proposition.

On est maintenant en mesure de démontrer le premier résultat essentiel de la théorie des schémas (α, β) . C'est l'objet du 4.5).

4.5) Condition nécessaire et suffisante de convergence.

Théorème 4.1. La condition nécessaire et suffisante pour que le schéma (α, β) soit convergent est que les deux conditions suivantes aient lieu :

(i) le polynôme α est stable (cf. proposition 4.1.)

(ii) $\alpha(1) = 0$, $\alpha'(1) = \beta(1) (\neq 0)$.

Démonstration.

1/ La nécessité de (i) et (ii) résulte des propositions 4.3. et 4.4.

2/ Suffisance des conditions.

Formons une inéquation aux différences satisfaites par $e_n = y_n - y(t_n)$

[comparer à la démonstration de la proposition 2.1.]. Posons :

$$\alpha_k y(t_{n+k}) + \dots + \alpha_0 y(t_n) - h[\beta_k f(t_{n+k}, y(t_{n+k}))] + \dots \\ \dots + \beta_0 f(t_n, y(t_n)) = \zeta_n .$$

Alors

$$(4,25) \quad \alpha_k e_{n+k} + \dots + \alpha_0 e_n - h \sum_{j=0}^k \beta_j [f_{n+j} - f(t_{n+j}, y(t_{n+j}))] = -\zeta_n$$

$$\begin{aligned} \text{Mais } f_{n+j} - f(t_{n+j}, y(t_{n+j})) &= f(t_{n+j}, y_{n+j}) - f(t_{n+j}, y(t_{n+j})) \\ &= g_{n+j} e_{n+j}, \end{aligned}$$

avec

$$(4,26) \quad |g_{n+j}| \leq C_1 \quad (\text{constante}).$$

Par ailleurs

$$\begin{aligned} y(t_{n+j}) &= y(t_n + j h) = y(t_n) + j h y'(t_n) + o(j h) \\ &[\frac{o(h)}{h} \rightarrow 0 \text{ si } h \rightarrow 0], \end{aligned}$$

$$y'(t_{n+j}) = y'(t_n + j h) = y'(t_n) + o(j h)$$

d'où en tenant compte de (ii)

$$(4,27) \quad \zeta_n = o(h)$$

On écrit alors (4,25) sous la forme :

$$(4,28) \quad \begin{cases} \alpha_k e_{n+k} + \dots + \alpha_0 e_n = d_{n+k}, & n \geq 0 \\ d_{n+k} = h \sum_{j=0}^k \beta_j g_{n+j} e_{n+j} - \zeta_n, \\ e_j = \eta_j(h) - y(t_j), & 0 \leq j \leq k-1. \end{cases}$$

Comme $\eta_j(h) \rightarrow y(t_j)$ lorsque $h \rightarrow 0$, on peut poser :

$$(4,29) \quad \sum_{0 \leq j \leq k-1} |e_j| = \delta(h), \quad \delta(h) \rightarrow 0, \text{ si } h \rightarrow 0.$$

Comme le polynôme α est stable, la proposition 4.2. donne :

$$|e_n| \leq K_\alpha (\delta(h) + \sum_{m=0}^{n-k} |d_{m+k}|).$$

$$\text{Or } |d_{m+k}| \leq C_2 h \sum_{0 \leq j \leq k} |e_{m+j}| + o(h)$$

$$\text{d'où } \sum_{m=0}^{n-k} |d_{m+k}| \leq C_3 h \sum_{p=0}^n |e_p| + o(1)$$

$[o(1) \rightarrow 0 \text{ si } h \rightarrow 0]$. Donc :

$$(4,30) \quad |e_n| \leq K_\alpha (\delta(h) + o(1) + c_3 h \sum_{p=0}^{n-1} |e_p|) .$$

Choisissons h assez petit pour que

$$1 - c_3 h \geq \frac{1}{2} . \text{ Alors (4,30) donne}$$

$$|e_n| \leq 2 K_\alpha (\delta(h) + o(1) + c_3 h \sum_{p=0}^{n-1} |e_p|)$$

et finalement posant :

$$(4,31) \quad |e_n| = \varepsilon_n .$$

On a :

$$(4,32) \quad \begin{cases} \varepsilon_n \leq a(h) + b h \sum_{p=0}^{n-1} \varepsilon_p \\ a(h) \rightarrow 0 \text{ si } h \rightarrow 0, \quad b = \text{constante.} \end{cases}$$

Mais on déduit par récurrence à partir de (4,32) que :

$$(4,33) \quad \varepsilon_n \leq (a(h) + b h \varepsilon_0) (1 + b h)^{n-1} .$$

Mais $(1 + b h)^n = \exp(n \log(1 + b h))$ est borné (car $n h$ fixé) d'où

$\varepsilon_n \rightarrow 0$, ce qui démontre le théorème.

Remarque 4.4. Stabilité du schéma.

L'inégalité (4,19) de la proposition 4.2. montre que si le polynôme α est stable, alors de petites variations des données initiales du second membre entraînent de petites variations sur la solution de (4,14), (4,15).

Le schéma (α, β) est dit stable si de petites variations des données initiales (4,8) entraînent de petites variations pour la solution du schéma (4,1) [avec (4,8)]. Plus précisément si :

$$|\eta_j(h)| \leq M, \quad \exists M_1 \text{ t.q. } |y_n| \leq M_1 \quad (h \rightarrow 0, n h \text{ fixé} \\ M_1 \text{ constante}).$$

Vérifions brièvement que la stabilité du schéma (α, β) équivaut à la

stabilité du polynôme α .

Ecrivons (4,1) de la façon suivante ; notons que [analogue au calcul fait à propos du théorème 4.1. en utilisant 0 au lieu de $y(t_{n+j})$]

$$f_{n+j} = f(t_{n+j}, y_{n+j}) = f(t_{n+j}, y_{n+j}) - f(t_{n+j}, 0) + f(t_{n+j}, 0)$$

entraîne

$$|f_{n+j}| \leq L |y_{n+j}| + C, \quad C = \sup |f(t, 0)|$$

de sorte que (4,1) donne :

$$\sum_{j=0}^k \alpha_j y_{n+j} = \varepsilon_{n+k},$$

$$|\varepsilon_{n+k}| \leq C_1 h \sum_{j=0}^k |y_{n+j}| + C_2 h.$$

On déduit de là, comme au théorème 4.1. et avec la proposition 4.2.

l'assertion ci-dessus.

Remarque 4.5. Consistance du schéma (α, β) .

Le schéma (4,1) (où (α, β) est dit consistant si lorsqu'on remplace dans (4,1) y_n par $y(t_n)$ la différence des deux membres est $O(h)$; plus précisément :

$$(4,34) \quad \begin{cases} \alpha_k y(t_{n+k}) + \alpha_{k-1} y(t_{n+k-1}) + \dots + \alpha_0 y(t_n) \\ - h[\beta_k f(t_{n+k}, y(t_{n+k}))] + \dots + \beta_0 f(t_n, y(t_n)) = o(h). \end{cases}$$

On vérifie sans peine (exercice) que la consistance du schéma (α, β) équivaut à la condition (ii) du théorème 4.1.

Remarque 4.6. Autre énoncé (équivalent) du théorème 4.1.

Avec la terminologie introduite dans les remarques 4.4. et 4.5. on peut énoncer le théorème 4.1. sous la forme équivalente suivante :

Théorème 4.1. bis. Le schéma (α, β) est convergent si et seulement si il est stable et consistant.

Remarque 4.7. Tout ce qui a été dit s'étend au cas des systèmes. Il suffit de remplacer les modules par des normes. Tout ceci vaut encore pour des équations différentielles dans des espaces de Banach.

Remarque 4.8. Dans la définition 4.1. de la convergence, nous avons utilisé la convergence ponctuelle. Or (4,33) entraîne la convergence uniforme.

Pour $\varepsilon > 0$, on aura $|y_n - y(t_n)| = \varepsilon$, pour h assez petit et tout n tel que $a + nh \leq b$.

4.6) Exemples.

Voyons quels sont les polynômes α et β dans les exemples introduits au n° 3.

Méthode d'Adams-Bashforth

Voici comment trouver aisément les polynômes α et β liés à (3,15) (et même chose pour (3,17) etc...). Si à la suite $\{f_n\}_{n \in \mathbb{Z}}$, on associe la fonction $f = \sum_{n \in \mathbb{Z}} f_n \lambda^{-n}$, l'opérateur de translation $\{f_n\} \rightarrow \{f_{n+1}\}$ se traduit sur les fonctions correspondantes par la multiplication par λ .

L'opérateur $\bar{\nabla}$ correspond alors à la multiplication par $(1 - \frac{1}{\lambda})$, et $\bar{\nabla}^m$ à $(1 - \frac{1}{\lambda})^m$.

Dans (3,15) nous posons (pour nous ramener aux notations du n° 4.4.1.) : $p+1 = n+k$; alors $p-q = n$ d'où $q = k-1$ et l'on trouve

$$(4,35) \quad \begin{cases} \alpha(\lambda) = \lambda^k - \lambda^{k-1} \\ \beta(\lambda) = \lambda^{k-1} \left(\sum_{m=0}^{k-1} \gamma_m \left(1 - \frac{1}{\lambda}\right)^m \right) \end{cases}$$

Méthode d'Adams-Moulton [cf. (3,11), (3,18)].

$$(4,36) \quad \begin{cases} \alpha(\lambda) = \lambda^k - \lambda^{k-1} \\ \beta(\lambda) = \lambda^k \sum_{m=0}^k \gamma_m^* \left(1 - \frac{1}{\lambda}\right)^m \end{cases}$$

Méthode de Nyström [cf. (3,19), (3,20)].

$$(4,37) \quad \begin{cases} \alpha(\lambda) = \lambda^k - \lambda^{k-2} \\ \beta(\lambda) = \lambda^{k-1} \sum_{m=0}^{k-1} \chi_m \left(1 - \frac{1}{\lambda}\right)^m \end{cases}$$

Méthode de Milne Simpson [cf. (3,21), (3,22)].

$$(4,38) \quad \begin{cases} \alpha(\lambda) = \lambda^k - \lambda^{k-2} \\ \beta(\lambda) = \lambda^k \sum_{m=0}^k \chi_m^* \left(1 - \frac{1}{\lambda}\right)^m \end{cases}$$

Grâce au théorème 4.1. on vérifie sans peine la convergence de ces méthodes.

Remarque 4.9. La méthode (α, β) est explicite (resp. implicite) si $d^0\beta < d^0\alpha$ (resp. $d^0\beta = d^0\alpha$).

Remarque 4.10. La question est maintenant évidemment : comment choisir entre ces diverses méthodes (α, β) ?

Des considérations "techniques" permettent de montrer que les méthodes d'Adams-Bashforth et Adams-Moulton ne sont pas optimales. Les méthodes de Nyström et Milne-Simpson ne sont éventuellement optimales que si $k = 2$. Alors $\alpha(\lambda) = \lambda^2 - 1$ et dans ce cas le polynôme β optimal est $\beta(\lambda) = \frac{1}{3}(\lambda^2 + 4\lambda + 1)$.

Cette méthode, appelée méthode de Milne est convergente d'ordre 4.