Chapitre 2 : méthodes à 1 pas.
Méthodes de Runge-Kutta

Il s'agit de résoudre de façon approchée une équation différentielle scalaire :

$$\begin{cases} y'(t) = f(t, y(t)) & t \in [t_0, t_0+T] = I \quad (1) \\ y(t_0) = y^0 & (2) \end{cases}$$

Nous supposerons ici que f est continue par rapport aux deux variables t et y, et lipschitzienne par rapport à y :

$$(3) \quad \forall t \in I, \ \forall (y,z) \in \mathbb{R}^2, \ |f(t,y) - f(t,z)| \leq L |y-z|$$

D'après le chapitre I, ceci nous permet d'affirmer que le problème (1) (2) admet une solution et une seule.

Comme il a été indiqué au chapitre I, dans la plupart des cas on n'est pas capable de trouver une solution analytique : il faut alors se contenter de la résoudre "numériquement". Pour cela on calcule des valeurs approchées y_n de y en des points t_n de l'intervalle I_0. On se donne donc une subdivision ou partage de l'intervalle I_0 :

(4) $\quad t_0 < t_1 \dots < t_n \dots < t_N = t_0 + T$

et l'on note h_n la longueur du segment $[t_n, t_{n+1}]$:

(5) $\quad h_n = t_{n+1} - t_n \qquad h = \max_{0 \le n \le N-1} h_i$

Souvent, pour des raisons de simplicité, on utilise un partage uniforme, i.e les h_n sont tous égaux à h, si bien que $h = \frac{T}{N}$.

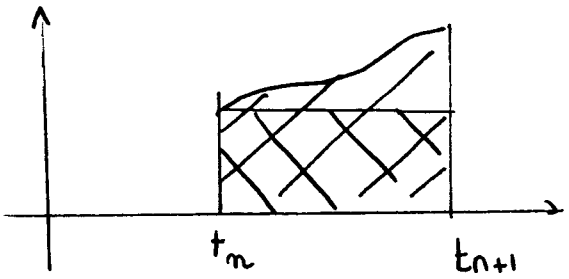
Historiquement, la première méthode numérique est celle du polygone d'Euler que nous avons introduite au chapitre I pour démontrer l'existence et l'unicité locale pour (1)(2). Rappelons-la brièvement : il s'agit d'approcher y en chaque point t_n par sa tangente. En termes d'intégration numérique, qui nous sera utile plus tard, écrivons :

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

et nous approchons cette dernière intégrale en interpolant $f(t, y(t))$ sur l'intervalle $[t_n, t_{n+1}]$ par la constante égale à $f(t_n, y(t_n))$ (formule des rectangles à gauche).

$$\int_{t_n}^{t_{n+1}} f(t, y(t))\,dt \sim h_n f(t_n, y(t_n))$$

Nous écrirons plus loin une estimation plus précise du reste.



Nous remplaçons alors (6) par :

$$y(t_{n+1}) \approx y(t_n) + h_n f(t_n, y(t_n))$$

le schéma approché sera alors

$$(7) \qquad y_{n+1} = y_n + h_n f(t_n, y_n) \qquad 0 \le n \le N-1$$

C'est la méthode d'Euler, ou Euler progressive, ou Euler explicite. Explicite car la donnée de y_0 permet de calculer explicitement y_n, pour $1 \le n \le N-1$ par la relation de récurrence (7). La donnée initiale pour le schéma, y_0, est égale à y^0 lorsqu'on en dispose, ou une approximation de y^0 (ce qui est souvent le cas pratiquement à cause de la précision de la machine).

Si l'on avait estimé l'intégrale en (6) par la formule des rectangles à gauche on aurait obtenu :

$$y(t_{n+1}) \simeq y(t_n) + h_n f(t_{n+1}, y(t_{n+1}))$$

que l'on aurait approché par le schéma

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

C'est la méthode d'Euler <u>implicite</u>, ou <u>rétrograde</u>. Implicite car à chaque <u>pas</u>, pour calculer y_{n+1} à partir de y_n, il faut résoudre une équation non linéaire.

Plus généralement une méthode à un pas est définie par une relation de récurrence ne faisant intervenir que deux valeurs consécutives de la suite y_n :

(9) $$y_{n+1} = y_n + h_n \varphi(t_n, y_n; h_n).$$

où φ est une fonction continue de $\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$ dans $\mathbb{R}$.

Pour la méthode d'Euler on a :

$$\varphi(t, y; h) \equiv f(t, y)$$

φ est indépendante de h.

Pour la méthode d'Euler implicite, c'est plus compliqué.

Supposons que l'équation

$$y + h f(t+h, z) = z$$

admette pour tout y une solution z unique. On la note

$$z = \mathcal{P}(t, y; h).$$

On a alors

$$y_{n+1} = y_n + h_n f(t_{n+1}, \mathcal{P}(t_n, y_n; h_n))$$

et l'on pose :

(10) $$\varphi(t_n, y_n; h_n) = f(t_n + h_n, \mathcal{P}(t_n, y_n; h_n)).$$

Dans ce cas on dit que la méthode à un pas est implicite.
Nous ne considérerons ici en général que des méthodes explicites.

Trois notions sont ici fondamentales : les valeurs y_n données par le schéma (9) sont-elles de bonnes approximations de $y(t_n)$ lorsque le pas h est petit ? c'est la notion de convergence.
Certaines méthodes sont-elles plus précises que d'autres ?
Ensuite ces méthodes sont-elles sensibles, et dans quelle mesure, aux erreurs faites sur la condition initiale (erreurs d'arrondi) : c'est la notion de stabilité.

1. Notions de base.

Définition 1 : L'erreur de troncature locale est la quantité ε_n définie par :

(11) $$\varepsilon_n = y(t_{n+1}) - y(t_n) - h_n \varphi(t_n, y(t_n); h_n)$$

La méthode est dite d'ordre p si p est le plus grand entier tel que $\varepsilon_n = O(h_n^{p+1})$. $0 \le n \le N-1$.
pour toute solution de (1), pour tout t_0, T, et pour tout partage.

Définition 2 : On dit que la méthode est consistante si elle vérifie :

(12) $$\varphi(t, y; 0) \equiv f(t, y).$$

Théorème 1 : une méthode consistante est d'ordre supérieur ou égal à 1.

Démonstration :

On utilise la formule de Taylor dans (11) en supposant y $\mathcal{C}^2$ et φ $\mathcal{C}^1$ par rapport à h.

$$y(t_{n+1}) - y(t_n) = h_n y'(t_n) + O(h_n^2) = h_n f(t_n, y(t_n)) + O(h_n^2)$$

$$\varphi(t_n, y(t_n); h_n) = \varphi(t_n, y(t_n); 0) + O(h_n)$$
$$= f(t_n, y(t_n)) + O(h_n)$$

par hypothèse.

D'où

$$\varepsilon_n = O(h_n^2) + h_n O(h_n) \text{ et}$$

(13) $$|\varepsilon_n| \leq C h_n^2$$

Exemple : la formule d'Euler progressive est évidemment consistante : supposons y $\mathcal{C}^2$ et calculons ε_n :

$$\varepsilon_n = y(t_{n+1}) - y(t_n) - h_n f(t_n, y(t_n))$$

(14) $$\varepsilon_n = \int_{t_n}^{t_{n+1}} f(t, y(t)) - h_n f(t_n, y(t_n))$$

Rappelons la formule d'intégration numérique :

(15) $$\int_{t_n}^{t_{n+1}} g(t)\, dt - h_n g(t_n) = \frac{h_n^2}{2} g'(\xi) \qquad \xi \in]t_n, t_{n+1}[$$

si g est une fonction $\mathcal{C}^1$.

En effet, considérons pour $h \in [0, h_n]$ la fonction

$$G(h) = \int_{t_n}^{t_n+h} g(t)\,dt - h\,g(t_n) - C\frac{h^2}{2}$$

C est une constante, définie par $G(h_n) = 0$.

On a donc $G(0) = G(h_n) = 0$

D'après le théorème de Rolle, G s'annule sur $]0, h_n[$:

$\exists\, h_1 \in]0, h_n[\quad G'(h_1) = 0$. Or

$$G'(h) = g(t_n + h) - g(t_n) - Ch$$

Donc $G'(0) = G'(h_1) = 0$ et G'' s'annule sur $]0, h_1[$

$\exists\, h_2 \in]0, h_1[\quad G''(h_2) = 0$, et

$$G''(h) = g'(t_n + h) - C.$$

Donc $C = g'(t_n + h_2) = g'(\xi)$, $\xi \in]t_n, t_{n+1}[$ et (14) est démontré. Donc

$$\varepsilon_n = \frac{h_n^2}{2}\left[\frac{d}{dt} f(t, y(t))\right](\xi) \qquad \xi \in]t_n, t_{n+1}[.$$
$$= \frac{h_n^2}{2} y''(\xi) \quad \text{si } y \text{ est } \mathcal{C}^2.$$

et la méthode est au moins d'ordre 1.

<u>Remarque</u> : de la même façon la méthode d'Euler rétrograde est d'ordre 1. On peut alors se demander pourquoi utiliser une méthode plus lourde si elle n'est pas plus précise. En fait nous verrons qu'elle a de meilleures propriétés de stabilité.

De la formule (13) on déduit un résultat élémentaire

Lemme 1 : si la méthode est consistante, on a

$$\lim_{h \to 0} \sum_{0}^{N-1} |\varepsilon_n| = 0 \qquad (16)$$

En effet d'après (13)

$$\sum_{0}^{N-1} |\varepsilon_n| \leq C \sum_{0}^{N-1} h_n^2 \leq Ch \sum_{0}^{N-1} h_n$$

or $\sum_{0}^{N-1} h_n = T$

d'où $\underbrace{\sum_{0}^{N-1} |\varepsilon_n|}_{\text{erreur de troncature globale}} \leq CTh$ et tend vers 0 avec h.

De même que la consistance est une condition facile à vérifier sur φ, on démontre :

Théorème 2 : On suppose que f est de classe $\mathcal{C}^p$ et que toutes les dérivées partielles $\varphi, \frac{\partial \varphi}{\partial h_n}, \dots, \frac{\partial^p \varphi}{\partial h_n^p}$ existent et sont continues.

Alors une condition nécessaire et suffisante pour que la méthode à un pas (9) soit d'ordre supérieur ou égal à p s'écrit

pour tout (t,y)

$$\begin{cases} \varphi(t,y;0) = f(t,y) \\ \frac{\partial \varphi}{\partial h_n}(t,y;0) = \frac{1}{2} f^{(1)}(t,y) \\ \vdots \\ \frac{\partial^{p-1} \varphi}{\partial h_n^{p-1}}(t,y;0) = \frac{1}{p} f^{(p-1)}(t,y) \end{cases} \qquad (17)$$

pour tout couple $(t,y) \in I_0 \times \mathbb{R}$

où les $f^{(k)}$ sont définies par récurrence par :

$$(18)\quad \begin{cases} f^{(0)}(t,y) = f(t,y) \\ f^{(1)}(t,y) = \dfrac{\partial f^{(0)}}{\partial t}(t,y) + \dfrac{\partial f^{(0)}}{\partial y}(t,y) f(t,y) \\ \vdots \\ f^{(k)}(t,y) = \dfrac{\partial f^{(k-1)}}{\partial t}(t,y) + \dfrac{\partial f^{(k-1)}}{\partial y}(t,y) f(t,y). \end{cases}$$

Pourquoi cette définition ? On a pour y solution de (1) de classe C^{p+1} :

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

$$y''(t) = \frac{d}{dt}[f(t, y(t))] = \frac{\partial f}{\partial t}(t, y(t)) + y'(t) \frac{\partial f}{\partial y}(t, y(t))$$

$$y''(t) = \frac{\partial f}{\partial t}(t, y(t)) + f(t, y(t)) \frac{\partial f}{\partial y}(t, y(t))$$

$$y''(t) = f^{(1)}(t, y(t))$$

Supposons que $y^{(k)}(t) = f^{(k-1)}(t, y(t))$

alors
$$y^{(k+1)}(t) = \frac{\partial f^{(k-1)}}{\partial t}(t, y(t)) + y'(t) \frac{\partial f^{(k-1)}}{\partial y}(t, y(t))$$
$$= f^{(k)}(t, y(t))$$

On a donc :

$$(19)\quad \forall k \geq 0, \quad y^{(k+1)}(t) = \frac{d^k}{dt^k}[f(t, y(t))] = f^{(k)}(t, y(t)).$$

Démontrons le théorème :

Pour simplifier les notations nous posons $h_n = h$ et

$$\varepsilon(t, h) = y(t+h) - y(t) - h\,\varphi(t, y(t); h)$$

si bien que $\varepsilon_n = \varepsilon(t_n, h_n)$

Par la formule de Taylor on a :

$$y(t+h) = \sum_{0}^{p} \frac{h^k}{k!} y^{(k)}(t) + \frac{h^{p+1}}{(p+1)!} y^{(p+1)}(t+\theta h) \quad 0<\theta<1$$

et par changement d'indices

$$y(t+h) - y(t) = \sum_{k=0}^{p-1} \frac{h^{k+1}}{(k+1)!} y^{(k+1)}(t) + \frac{h^{p+1}}{(p+1)!} y^{(p+1)}(t+\theta h) \quad 0<\theta<1$$

et d'après (19)

$$y(t+h) - y(t) = \sum_{k=0}^{p-1} \frac{h^{k+1}}{(k+1)!} f^{(k)}(t, y(t)) + \frac{h^{p+1}}{(p+1)!} y^{(p+1)}(t+\theta h) \quad 0<\theta<1$$

De même

$$\varphi(t, y(t); h) = \sum_{k=0}^{p-1} \frac{h^k}{k!} \frac{\partial^k \varphi}{\partial h^k}(t, y(t); 0) + \frac{h^p}{p!} \frac{\partial^p \varphi}{\partial h^p}(t, y(t), \sigma h)$$
$$0<\sigma<1.$$

et donc

$$\varepsilon(t,h) = \sum_{k=0}^{p-1} \frac{h^{k+1}}{k!} \Psi_k(t, y(t)) + \frac{h^{p+1}}{p!} \left[\frac{1}{p+1} y^{(p+1)}(t+\theta h) - \frac{\partial^p \varphi}{\partial h^p}(t, y(t), \sigma h) \right]$$
$$0<\theta<1$$
$$0<\sigma<1$$

avec

$$\Psi_k(t,y) = \frac{1}{k+1} f^{(k)}(t,y) - \frac{\partial^k \varphi}{\partial h^k}(t,y,0)$$

Montrons maintenant que (18) est une condition suffisante.

si (18) est vérifié, alors $\Psi_k \equiv 0$ pour $0 \le k \le p-1$ et

$$\varepsilon(t,h) = O(h^{p+1}) \text{ et } \varepsilon_n = O(h_n^{p+1}).$$

Réciproquement montrons que si la méthode est d'ordre p alors les conditions (18) sont vérifiées. Procédons par l'absurde : supposons qu'elles ne sont pas vérifiées

Il existe alors un plus petit entier k tel que

$$k<p, \quad \Psi_k(t_1, y^1) \neq 0 \quad \text{pour un couple } (t_1, y^1) \text{ dans } I_0 \times \mathbb{R}$$

Soit alors $y(t)$ la solution de (1) telle que $y(t_1)=y_1$. On aurait alors

$$\varepsilon(t_1, h) = \frac{h^k}{k!} \Psi_k(t_1, y^1) + O(h^{k+1}).$$

et $\varepsilon(t_1, h) = \varepsilon_1 = O(h^k)$ avec $k<p$, ce qui aboutit à une contradiction.

Définition 3 : La méthode (9) est dite convergente si pour tout (t_0, y^0), pour toute y solution de (1),(2), pour toute y_n solution de (9) satisfaisant la condition initiale $y_0 = y_0(h)$, avec $y_0(h) \to y^0$ lorsque $h \to 0$, on a

$$\lim_{h \to 0} |y_n - y(t_n)| = 0, \quad 1 \leq n \leq N.$$

Remarquons que cette définition permet de démarrer le schéma avec une valeur y_0 qui n'est pas y^0.

Comme nous l'avions fait pour le problème continu, nous allons maintenant définir la stabilité du schéma, c'est à dire la sensibilité par rapport à une erreur sur les données. Considérons donc y_n, solution de (9), et z_n solution d'un schéma perturbé

$$z_{n+1} = z_n + h_n \varphi(t_n, z_n; h_n) + \eta_n \tag{20}$$

Définition 4 : La méthode (9) est dite stable s'il existe des constantes M et H telles que

$$|y_n - z_n| \leq M\left(|y_0 - z_0| + \sum_{n=0}^{N-1} |\eta_n|\right) \quad 0 \leq n \leq N \tag{21}$$

pour tout $h_n < H$.

Théorème 3 : Un schéma stable et consistant est convergent.

Démonstration :

La suite $z_n = y(t_n)$ vérifie (20) avec $\eta_n = \varepsilon_n$. Si le schéma est stable on a donc

$$|y_n - y(t_n)| \leq M\left(|y_0 - y^0| + \sum_{0}^{N-1} |\varepsilon_n|\right) \quad 0 \leq n \leq N. \tag{22}$$

or nous avons montré au lemme 1 que si la méthode est consistante, $\sum_{0}^{N-1} |\varepsilon_n| \to 0$. Donc si $|y_0 - y^0|$ tend vers 0,

$$|y_n - y(t_n)| \to 0 \quad 0 \leq n \leq N.$$

On a même une estimation d'erreur pour un schéma d'ordre p

Théorème 4 : On suppose que f est de classe C^p. Alors si la méthode est stable et d'ordre p, il existe une constante $K > 0$ indépendante de h telle que

$$|Y_n - Y(t_n)| \leq M\left(|y_0 - y^0| + K h^p\right). \tag{22}$$

Démonstration :

Si la méthode est d'ordre p, $|\varepsilon_n| \leq C h_n^{p+1}$, et comme au lemme 1

$$\sum_0^{N-1} |\varepsilon_n| \leq C T h^p$$

En reportant dans (22)

$$|y_n - y(t_n)| \leq M(|y_0 - y^0| + C T h^p).$$

Ceci montre que si l'on peut choisir $y_0 = y^0$, l'erreur sur la solution est $O(h^p)$.

Il reste maintenant à écrire une condition suffisante de stabilité facile à utiliser. Elle est parallèle à celle écrite au ch1.

Théorème 5 : On suppose que φ satisfait une condition de Lipschitz en y : il existe une constante Λ telle que :

(23) $\forall t \in I_0$, $\forall (y,z) \in \mathbb{R}^2$, $\forall h \in [0,H]$, $|\varphi(t,y;h) - \varphi(t,z;h)| \leq \Lambda |y-z|$.

Alors la méthode (9) est stable.

Démonstration :

Réécrivons (9) et (20)

$$y_{n+1} = y_n + h_n \varphi(t_n, y_n; h_n)$$
$$z_{n+1} = z_n + h_n \varphi(t_n, z_n; h_n) + \eta_n$$

et soustrayons ces deux égalités.

$$y_{n+1} - z_{n+1} = y_n - z_n + h_n [\varphi(t_n, y_n; h_n) - \varphi(t_n, z_n; h_n)] + \eta_n$$

Par l'inégalité triangulaire et en utilisant (23) on obtient :

$$|y_{n+1} - z_{n+1}| \leq (1 + \Lambda h_n) |y_n - z_n| + |\eta_n|.$$

Posons $e_n = |y_n - z_n|$.

On a :

$$e_{n+1} \leq (1 + \Lambda h_n)\, e_n + |\eta_n| \tag{24}$$

Utilisons un lemme technique sur les suites récurrentes

Lemme 2 : Soient e_n et b_n deux suites de nombres positifs vérifiant

$$e_{n+1} \leq (1 + \Lambda h_n)\, e_n + b_n \tag{25}$$

Alors on a pour tout $n \geq 0$

$$e_n \leq e_0 \exp(\Lambda (t_n - t_0)) + \sum_{\ell=0}^{n-1} b_\ell \exp(\Lambda (t_n - t_{\ell+1})) \tag{26}$$

Appliquons ce lemme technique à $e_n = |y_n - z_n|$.

Majorons $t_n - t_{\ell+1}$ par T pour tout n et ℓ, et $t_n - t_0$ par T.

Il vient

$$|y_n - z_n| \leq \left(|y_0 - z_0| + \sum_{\ell=0}^{n-1} |\eta_\ell|\right) e^{\Lambda T}$$

$$|y_n - z_n| \leq \left(|y_0 - z_0| + \sum_{\ell=0}^{N-1} |\eta_\ell|\right) e^{\Lambda T}$$

ce qui correspond à (21) avec $M = e^{\Lambda T}$.

Remarque : Comme nous l'avions remarqué au chapitre 1, dans le cas d'équations raides, c'est-à-dire lorsque la constante de Lipschitz Λ est grande, alors l'estimation d'erreur (22) avec $M = e^{\Lambda T}$ n'est plus très intéressante.

Démonstration du lemme 2 :

Elle se fait par récurrence sur n.

- (26) est évidemment vraie pour $n=0$
- supposons (26) vraie pour n. Alors

$$e_{n+1} \leq e_0 (1+\Lambda h_n) \exp(\Lambda(t_n - t_0)) + (1+\Lambda h_n) \sum_{\ell=0}^{n-1} b_\ell \exp(\Lambda(t_n - t_{\ell+1})) + b_n .$$

Maintenant nous savons que :

$$1 + \Lambda h_n \leq e^{\Lambda h_n} = e^{\Lambda(t_{n+1} - t_n)}$$

D'où

$$(1+\Lambda h_n) e^{\Lambda(t_n - t_0)} \leq e^{\Lambda(t_{n+1} - t_0)}$$

$$(1+\Lambda h_n) e^{\Lambda(t_n - t_{\ell+1})} \leq e^{\Lambda(t_{n+1} - t_{\ell+1})}$$

et $e_{n+1} \leq e_0 e^{\Lambda(t_{n+1} - t_0)} + \underbrace{\sum_{\ell=0}^{n-1} b_\ell e^{\Lambda(t_{n+1} - t_{\ell+1})} + b_n}_{= \sum_{\ell=0}^{n} b_\ell e^{\Lambda(t_{n+1} - t_{\ell+1})}}$

et le lemme est démontré.

2. Schémas de Runge et Kutta

Nous avons vu un schéma, le schéma d'Euler. Il est d'ordre 1, ce qui signifie que pour avoir une grande précision, il faut prendre des pas h_n très petits, et donc beaucoup de points de discrétisation. Nous allons construire des schémas d'ordre supérieurs, d'une façon systématique, à l'aide des méthodes d'intégration numérique. Mais avant cela, regardons comment construire un schéma d'ordre 2. Une idée naturelle consisterait à utiliser un développement de Taylor ..

On a en effet :

$$Y(t_{n+1}) = Y(t_n) + h_n y'(t_n) + \frac{h_n^2}{2} y''(t_n) + O(h_n^3)$$

$$Y(t_{n+1}) = Y(t_n) + h_n f(t_n, y(t_n)) + \frac{h_n^2}{2} f^{(1)}(t_n, y(t_n)) + O(h_n^3)$$

avec les notations (19). Donc si l'on définit un schéma par :

$$Y_{n+1} = Y_n + h_n f(t_n, y_n) + \frac{h_n^2}{2} f^{(1)}(t_n, y_n)$$

il sera d'ordre 2. Malheureusement l'utilisation pratique de ce schéma nécessiterait de dériver la fonction f, ce qui n'est pas toujours possible, car souvent elle n'est pas connue analytiquement, mais par ses valeurs en un certain nombre de points.

construisons le schéma d'Euler modifié.

La construction se fait en deux étapes. Rappelons d'abord que nous devons déterminer une fonction φ telle que

$$y(t_{n+1}) - y(t_n) = h_n \varphi(t_n, y(t_n); h_n) + O(h_n^3).$$

Reprenons la formule

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

Nous approchons l'intégrale par une formule d'intégration numérique, qui doit être aussi d'ordre 2 : c'est la formule du point milieu :

$$\int_{t_n}^{t_{n+1}} f(t, y(t))\, dt = h_n f\left(t_n + \frac{h_n}{2}, y\left(t_n + \frac{h_n}{2}\right)\right) + O(h_n^3)$$

Pour atteindre notre but, il faut maintenant exprimer, de façon approchée, $y(t_n + \frac{h_n}{2})$ en fonction de $y(t_n)$. Ici, nous n'avons plus besoin que de l'ordre 1, puisque nous avons h_n en facteur : la formule des rectangles à gauche :

$$y\left(t_n + \frac{h_n}{2}\right) - y(t_n) = \int_{t_n}^{t_n + \frac{h_n}{2}} y'(t)\, dt = \frac{h_n}{2} f(t_n, y(t_n)) + O(h_n^2)$$

d'où finalement :

$$y(t_{n+1}) - y(t_n) = h_n f\left(t_n + \frac{h_n}{2}, y(t_n) + \frac{h_n}{2} f(t_n, y(t_n)) + O(h_n^2)\right) + O(h_n^3). \tag{27}$$

et le schéma

$$y_{n+1} = y_n + h_n f\left(t_n + \frac{h_n}{2}, y_n + \frac{h_n}{2} f(t_n, y_n)\right) \tag{28}$$

C'est le schéma d'Euler modifié. Il est d'ordre 2. En effet l'erreur de troncature est ici d'après (27)

$$\varepsilon_n = h_n \Big[f\left(t_n + \frac{h_n}{2}, y(t_n) + \frac{h_n}{2} f(t_n, y(t_n)) + O(h_n^2)\right) - f\left(t_n + \frac{h_n}{2}, y(t_n) + \frac{h_n}{2} f(t_n, y(t_n))\right) \Big] + O(h_n^3)$$

et puisque f est lipschitzienne en y,

$$\varepsilon_n = O(h_n^3).$$

Pour construire le schéma d'<u>Euler amélioré</u>, on utilise la formule des trapèzes sur $[t_n, t_{n+1}]$

$$y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} y'(t)\,dt = \frac{h_n}{2}\left[f(t_n, y(t_n)) + f(t_{n+1}, y(t_{n+1}))\right] + O(h_n^3)$$

Puisque l'on veut un schéma explicite, on doit exprimer $y(t_{n+1})$. On utilise alors la formule du rectangle :

$$y(t_{n+1}) - y(t_n) = h_n f(t_n, y(t_n)) + O(h_n^2)$$

et donc

$$y(t_{n+1}) - y(t_n) = \frac{h_n}{2}\left[f(t_n, y(t_n)) + f(t_{n+1}, y(t_n) + h_n f(t_n, y(t_n)) + O(h_n^2))\right]$$

ce qui donne le schéma

$$(2\mathcal{S}) \quad y_{n+1} = y_n + \frac{h_n}{2}\left[f(t_n, y_n) + f(t_n + h_n, y_n + h_n f(t_n, y_n))\right]$$

Il est aussi d'ordre 2.

Ces deux schémas s'écrivent sous la forme

$$\begin{cases} k_1 = f(t_n, y_n) \\ k_2 = f(t_n + a h_n, y_n + a h_n k_1) \qquad 0 < a \leq 1. \\ y_{n+1} = y_n + h_n (b_1 k_1 + b_2 k_2) \end{cases}$$

Plus généralement on définit une méthode de Runge-Kutta à R niveaux par la façon suivante :

$$y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} f(t, y(t))\,dt \sim h_n \sum_1^R b_i f(t_n + h_n \tau_i, y(t_n + h_n \tau_i))$$

$0 \leq \tau_i \leq 1.$

puis $y(t_n + h_n \tau_i) = y(t_n) + \int_{t_n}^{t_n + h_n \tau_i} f(t, y(t))\,dt$

$$= y(t_n) + h_n \sum_{j=1}^{R} a_{ij} f(t_n + h_n \tau_j, y(t_n + h_n \tau_j)).$$

posons $y_{n,j} \sim y(t_n + h_n \tau_j)$

$y_n \sim y(t_n)$

$$1 \leq i \leq R \qquad y_{n,i} = y_n + h_n \sum_{j=1}^{R} a_{ij} f(t_n + h_n \tau_j, y_{n,j})$$

$$y_{n+1} = y_n + h_n \sum_1^R b_i f(t_n + h_n \tau_i, y_{n,i}).$$

autre formulation :

$$k_i = f(t_n + h_n c_i, y_{n,i})$$

$$\begin{cases} 1 \le i \le R \quad k_{ni} = f(t_n + h_n c_i, y_n + h_n \sum_{j=1}^{R} a_{ij} k_j) \\ y_{n+1} = y_n + h_n \sum_{i=1}^{R} b_i k_{ni} \end{cases}$$

$$K_i(t_n, y_n, h_n)$$

Traditionnellement on l'écrit de la façon suivante.

$$k_1 = f(t_n + h_n c_1, y_n + h_n \sum_{1}^{R} a_{1j} k_j)$$

$$\vdots$$

$$k_R = f(t_n + h_n c_R, y_n + h_n \sum_{1}^{R} a_{Rj} k_j)$$

$$y_{n+1} = y_n + h_n (b_1 k_1 - + b_R k_R)$$

c_1	a_{11} ——	a_{1R}
↓	‖	‖
c_R	a_{R1} ——	a_{RR}
	b_1 ——	b_R .

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1R} \\ \vdots & & \vdots \\ a_{R1} & \cdots & a_{RR} \end{pmatrix} \qquad c = \begin{pmatrix} c_1 \\ \vdots \\ c_R \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

les c_i sont rangés dans l'ordre croissant.

Si a est triangulaire inférieure avec des zéros sur la diagonale. Ceci définit bien un schéma : Runge - Kutta explicite. Il s'écrit sous la forme méthode à un pas avec

$$\begin{cases} \varphi(t, y; h) = \sum_{1}^{R} b_i k_i(t, y; h) \\ k_i(t, y; h) = f(t + c_i h, y + h \sum_{1}^{i-1} a_{ij} k_j) \end{cases} \quad \text{(convention } a_{10} = 0\text{)}$$

Sinon : on a un schéma implicite. Ceci définit bien un schéma si on sait résoudre pour tout t, y, h :

$$k_i(t, y; h) = f(t + h c_i, y + h \sum_{1}^{R} a_{ij} k_j)$$

Posons $K = \begin{pmatrix} k_1 \\ \vdots \\ k_R \end{pmatrix}$ $F(K) = \begin{cases} F_1(K) \\ \vdots \\ F_R(K) \end{cases}$ avec $F_i(K) = f(t + h c_i, y + h \sum_{1}^{R} a_{ij} k_j)$

Il existe k tel que $F(K) = K \iff$ F est strictement contractante dans $\mathbb{R}^R$ pour une norme de Banach.

$$|F_i(k) - F_i(\tilde{k})| \le Lh \left| \sum_{j=1}^{R} a_{ij} (k_j - \tilde{k}_j) \right| = Lh \left| [A(k - \tilde{k})]_j \right|.$$

donc pour toute norme de $\mathbb{R}^R$ subordonnée.

$$\| F(k) - F(\tilde{k}) \| \le Lh \|A\| \, \| k - \tilde{k} \|.$$

Théorème : Si $\forall n$, $Lh_n \|A\| < 1$ pour une norme subordonnée, alors le schéma de Runge - Kutta implicite est bien défini.

Et on a encore un schéma à un pas.

Exemples

La méthode d'Euler modifié s'écrit avec ces notations :

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

et la méthode d'Euler amélioré :

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

et la méthode de Runge-Kutta à 2 niveaux (2g) :

0	0	0
á	a	0
	$1-\frac{1}{2a}$	$\frac{1}{2a}$

Exemples :

Schéma de Heun : Ordre 3.

$$k_1 = f(t_n, y_n)$$

$$k_2 = f\left(t_n + \frac{h_n}{3}, y_n + \frac{h_n}{3} k_1\right)$$

$$k_3 = f\left(t_n + \frac{2h_n}{3}, y_n + \frac{2h_n}{3} k_2\right)$$

$$y_{n+1} = y_n + \frac{h_n}{4}(k_1 + 3k_3)$$

0	0	0	0
$\frac{1}{3}$	$\frac{1}{3}$	0	0
$\frac{2}{3}$	0	$\frac{2}{3}$	0
	$\frac{1}{4}$	0	$\frac{3}{4}$

Schéma de Runge-Kutta d'ordre 4.

$$k_1 = f(t_n, y_n)$$
$$k_2 = f(t_n + \frac{h_n}{2}, y_n + \frac{h_n}{2} k_1)$$
$$k_3 = f(t_n + \frac{h_n}{2}, y_n + \frac{h_n}{2} k_2)$$
$$k_4 = f(t_n + h_n, y_n + h_n k_3)$$
$$y_{n+1} = y_n + \frac{h_n}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

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

Etudions la stabilité.

Théorème : • Un schéma de Runge-Kutta explicite est stable

• si $h_n \leq h^*$ avec $Lh^* \|A\|_\infty < 1$, alors le schéma de Runge-Kutta implicite est stable

Dém : Rappelons qu'il suffit de montrer que φ est lipschitzienne en y, où

$$\varphi(t, y, h) = \sum b_i k_i(t, y, h).$$

Il suffit donc de montrer que les k_i sont lipschitz en y.

$$1 \leq i \leq R \quad |k_i(y) - k_i(z)| \leq L\left[|y - z| + h \sum_1^R |a_{ij}|\, |k_j(y) - k_j(z)|\right]$$

• cas explicite . $|k_1(y) - k_1(z)| \leq L|y - z|$

si pour $j \leq i-1$ $|k_j(y) - k_j(z)| \leq \Lambda_j |y - z|$

alors $|k_i(y) - k_i(z)| \leq L\left(1 + h \sum_1^{i-1} \Lambda_j |a_{ij}|\right) |y - z|$.

$$\Lambda_j = L\left(1 + h \sum_1^{i-1} \Lambda_j |a_{ij}|\right)$$

• cas implicite :

$$\forall i \quad |k_i(y) - k_i(z)| \leq L|y - z| + Lh \|k(y) - k(z)\|_\infty \sum_1^R |a_{ij}|.$$

$$\|k(y) - k(z)\|_\infty \leq L|y - z| + Lh \|A\|_\infty \|k(y) - k(z)\|_\infty.$$

donc si $Lh\|A\|_\infty < 1$.

$$\|k(y) - k(z)\|_\infty \leq \frac{L}{1 - Lh\|A\|_\infty} |y - z|.$$

et φ est lipsch.

Théorème 7 : la méthode de Runge-Kutta (31)(32) est ordre ≥ 1 si et seulement si :

(33) $\quad {}^t b.e = 1 \qquad e = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

elle est d'ordre ≥ 2 si et seulement si

(34) $\quad {}^t b.e = 1 \qquad {}^t b\, c = {}^t b A e = \frac{1}{2}$

Démonstration : on utilise la caractérisation du théorème 2.

. La méthode est d'ordre ≥ 1 ssi $\varphi(t,y;0) = f(t,y)$

or
$$\varphi(t,y;0) = \sum_{j=1}^{R} b_j f(t,y)$$

donc $\varphi(t,y;0) = f(t,y) \Leftrightarrow \sum_{j=1}^{R} b_j = 1 \Leftrightarrow {}^t b e = 1$.

. puis la méthode est d'ordre ≥ 2 ssi on a de plus $\frac{\partial \varphi}{\partial h}(t,y;0) = \frac{1}{2} f^{(1)}$

$= \frac{1}{2}(f_t + f f_y)$

$$\frac{\partial k_i}{\partial h}(t,y;0) = c_i \frac{\partial f}{\partial t}(t,y) + \frac{\partial f}{\partial y}(t,y) \sum_{j=1}^{R} a_{ij} k_j(t,y;0)$$

$$= c_i \frac{\partial f}{\partial t}(t,y) + f(t,y) \frac{\partial f}{\partial y}(t,y) \sum_{j=1}^{R} a_{ij}.$$

d'où

$$\frac{\partial \varphi}{\partial h}(t,y;0) = \left(\sum_{1}^{R} b_j c_j \right) \frac{\partial f}{\partial t}(t,y) + \left(\sum_{j=1}^{R} \sum_{i=1}^{R} a_{ji} b_j \right) f(t,y) \frac{\partial f}{\partial y}(t,y)$$

La méthode est alors d'ordre ≥ 2 si et seulement si

$$\sum_{1}^{R} b_j c_j = {}^t b.c = \frac{1}{2}$$

$$\sum_{j=1}^{R} \sum_{i=1}^{R} a_{ji} b_j = {}^t b.Ae = \frac{1}{2}.$$

De la même façon on peut demontrer le

Théorème 8 : On suppose que $Ae = Te$. La méthode de Runge-Kutta est d'ordre ≥ 3 si et seulement si

$$(35) \qquad {}^tb\,Te = \frac{1}{2} \qquad {}^tb\,Te^2 = \frac{1}{3} \qquad {}^tb\,ATe = \frac{1}{6}$$

elle est d'ordre ≥ 4 si en outre on a

$$(36) \qquad {}^tb\,T^3e = \frac{1}{4} \qquad {}^tb\,AT^2e = \frac{1}{12} \qquad {}^tb\,A^2Te = \frac{1}{24} \qquad {}^tb\,TATe = \frac{1}{8}.$$

où l'on a posé $T = \begin{pmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_k \end{pmatrix}$.

3. Notion de stabilité absolue.

Considérons pour commencer un exemple. Prenons $t_0 = 0$ pour simplifier et considérons l'équation différentielle

$$(37) \qquad y' = -\lambda(y - g(t)) + g'(t). \qquad \lambda \in \mathbb{R}.$$

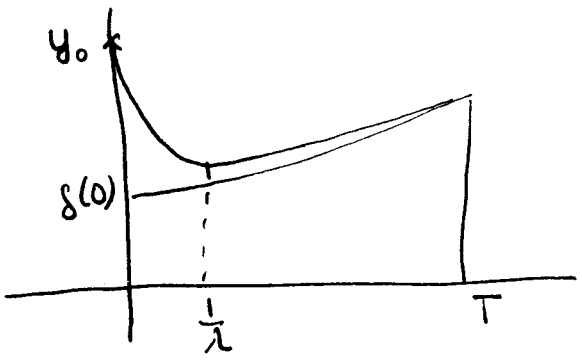
avec la condition de Cauchy

$$(38) \qquad y(0) = y^0$$

la solution de (37), (38) s'écrit

$$y(t) = (y_0 - g(0))\exp -\lambda t + g(t).$$

si $y_0 \neq g(0)$ et λ est très grand, on a un phénomène de couche limite



dans la phase initiale d'une durée de l'ordre de $\frac{1}{\lambda}$, $y(t)$ varie très rapidement. Ensuite le terme exponentiel devient négligeable : on a une zone régulière.

Appliquons le schéma d'Euler,

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

Soit ici :

$$y_{n+1} = y_n (1 - \lambda h_n) + [\lambda g(t_n) + g'(t_n)] h_n$$

l'erreur $e_n = y(t_n) - y_n$ vérifie

$$e_{n+1} = (1 - \lambda h_n) e_n + \varepsilon_n$$

et nous avons vu que

$$\varepsilon_n = \frac{h_n^2}{2} y''(\xi_n) \quad \xi_n \in]t_n, t_{n+1}[.$$

On peut alors calculer e_n explicitement

$$e_n = \sum_{\ell=0}^{n-1} \left(\prod_{k=\ell+1}^{n-1} (1 - \lambda h_k) \right) \varepsilon_\ell$$

Dans la couche limite on a

$$y''(t) \sim \lambda^2 \exp -\lambda t \sim \lambda^2 \quad \text{puisque } t < \frac{1}{\lambda}.$$

donc $\varepsilon_\ell \sim \frac{(h_\ell \lambda)^2}{2}$

et pour que ε_ℓ soit petit, il faut que λh_ℓ soit petit ; ce qui limite nettement le pas ; et augmente notablement le nombre de points : on a affaire

à un système raide. On ne peut pas se permettre par contre de choisir λh_ℓ petit partout. Dans la zone régulière on aimerait choisir $\lambda h_\ell \gg 1$, mais alors c'est le terme $\prod_{\ell+1}^{n-1}(1-\lambda h_k)$ qui devient très grand : la méthode d'Euler s'applique mal à des systèmes raides.

Considérons maintenant le schéma d'Euler rétrograde $y_{n+1} = y_n + h_n f(t_{n+1}, y_{n+1})$. Appliqué à notre équation il s'écrit

$$(1+\lambda h_n) y_{n+1} = y_n + h_n(-\lambda g(t_{n+1}) + g'(t_{n+1}))$$

et l'erreur vérifie

$$e_{n+1} = \frac{1}{1+\lambda h_n}(e_n + \varepsilon_n)$$

et $\quad \varepsilon_n = -\frac{h_n^2}{2} y''(\xi_n) \quad \xi_n \in]t_n, t_{n+1}[$.

On obtient maintenant :

$$e_n = \sum_{\ell=0}^{n-1} \left(\prod_{k=\ell}^{n-1} \frac{1}{1+\lambda h_k} \right) \varepsilon_\ell .$$

Dans la couche limite, il faudra encore assurer $\lambda h_\ell \ll 1$, mais ici on pourra prend $\lambda h_k \gg 1$, dans la zone régulière. Ceci nous amène à la notion de stabilité absolue.

Considérons l'équation différentielle linéaire

(39) $\quad y' = -\lambda y \qquad \lambda \in \mathbb{R}_+$

la solution générale est $y = e^{-\lambda t}$

si bien que

$$(40) \qquad y(t_{n+1}) = e^{-\lambda h_n}\, y(t_n)$$

et la suite des $y(t_n)$ est décroissante.

Soit maintenant un schéma de Runge-Kutta pour cette équation. Il s'écrit :

$$(41) \quad \begin{cases} k_i = -\lambda \left(y_n + h_n \sum_{j=1}^{i-1} a_{ij}\, k_j \right) \\ y_{n+1} = y_n + h_n \sum_{j=1}^{s} b_j\, k_j \end{cases}$$

Montrons par récurrence que $h_n k_i$ est un polynôme de degré i en λh_n :

$$h_n k_i = P_i(\lambda h_n)\, y_n .$$

- c'est vrai pour $i = 1$ car $k_1 = -\lambda y_n$
- si c'est vrai pour $j = 1, \dots, i-1$

$$h_n k_i = \left[-\lambda h_n - \lambda h_n \sum_{j=1}^{i-1} a_{ij}\, P_j(\lambda h_n) \right] y_n$$

et donc

$$h_n k_i = P_i(\lambda h_n)\, y_n$$

avec

$$P_i(z) = -z \left(1 + \sum_{j=1}^{i-1} a_{ij}\, P_j(z) \right)$$

puis

$$y_{n+1} = y_n + \sum_{j=1}^{s} b_j\, P_i(\lambda h_n)\, y_n$$

et donc

$$(41) \qquad y_{n+1} = Q(\lambda h_n)\, y_n .$$

puisque la solution exacte est exponentiellement décroissante, il est raisonnable de demander que

$$\left|\frac{y_{n+1}}{y_n}\right| \leq 1$$

Definition 5 : On appelle domaine de stabilité absolue l'ensemble des $z \in \mathbb{R}_+$ tels que

$$|Q(z\ .)| \leq 1.$$

Nous ne démontrerons pas le résultat suivant :

Théorème 7 : Tous les schémas de Runge-Kutta définis par (30), consistants et de même ordre R, ont le même intervalle de stabilité absolue.

le tableau suivant donne les domaines d'absolue stabilité suivant l'ordre de la méthode.

ordre	intervalle d'absolue stabilité
1	$[0,2]$
2	$[0,2]$
3	$[0, 2.51]$
4	$[0, 2.78]$

ce qui montre que si accroître l'ordre accroit la précision, cela n'accroit que peu la stabilité absolue. Comme l'a montré l'exemple précédent, il faut alors se tourner vers d'autres méthodes : implicites (matrice A pleine) ou semi-implicite ($a_{ij} = 0$ pour $j > i$).

Méthodes multipas linéaires

Ch 3.

On considère de nouveau le problème :

$$(1) \quad \begin{cases} y' = f(t,y) & t \in [0,T] \\ y(0) = y^0 \end{cases}$$

où f est définie continue sur $I_0 \times \mathbb{R}$, lipschitzienne en y de constante L. On se donne ici un partage équidistant de l'intervalle $[0,T]$: $t_0 = 0 < t_1 < \dots < t_N = T$, $t_{j+1} - t_j = h$, si bien que $Nh = T$

Une méthode à q pas consiste à approcher $y(t_n)$ par y_n, où y_{n+q} dépend de $y_{n+q-1}, \dots, y_n$. Plus précisément une méthode multipas linéaire s'écrit :

$$(2) \quad \sum_{j=0}^{q} \alpha_j y_{n+j} = h \sum_{j=0}^{q} \beta_j f(t_{n+j}, y_{n+j}) \qquad 0 \le n \le N-q$$

où les α_j et β_j sont constants. On suppose que α_0 et β_0 ne sont pas simultanément nuls, et que $\alpha_q = 1$.

La résolution du schéma (2) nécessite la connaissance des q premières valeurs de la suite. y_0 sera choisi proche de y^0, et $y_1, \dots, y_{q-1}$ seront calculés à l'aide de schémas à un pas par exemple.

On dit que la méthode est explicite si $\beta_q = 0$, et implicite si $\beta_q \neq 0$. Pour une méthode explicite on obtient directement y_{n+q} à partir des y_{n+j}, $j = 0, \dots, q-1$. Dans le cas d'une méthode implicite on a à résoudre à chaque étape l'équation :

$$y_{n+q} = h \beta_q f(t_{n+q}, y_{n+q}) + g$$

où g est une fonction connue des y_{n+j}, $j = 0, \dots, q-1$. Si f est linéaire, on en déduit y_{n+q} si $\alpha_q - h \beta_q \frac{\partial f}{\partial y} \neq 0$. Si f n'est pas linéaire, le théorème du point fixe assure l'existence d'une solution pourvu que

$$h |\beta_q| L < 1.$$

y_{n+q} peut alors être calculé par une méthode itérative :

$$y_{n+q}^{[s]} = h \beta_q f\left(t_{n+q}, y_{n+q}^{[s-1]}\right) + g.$$

$y_{n+q}^{[0]}$ arbitraire.

A priori, la convergence peut être lente. Comme pour les méthodes de Runge-Kutta, les méthodes implicites ont de meilleures propriétés de stabilité que les méthodes explicites. Dans le cas de problèmes raides, elles sont donc plus intéressantes. Sur le plan pratique, on limitera le coût du calcul par la méthode itérative en choisissant

bien $y_{n+q}^{[0]}$: par une méthode explicite. On pourra alors se limiter à une ou deux itérations. C'est la méthode des prédicteurs - correcteurs sur laquelle nous reviendrons.

Donnons d'abord deux exemples

Ex 1 : la méthode du point milieu.

le schéma s'écrit :

$$y_{n+1} - y_{n-1} - 2h f(t_n, y_n) = 0 \tag{3}$$

Il est obtenu en écrivant la formule du point milieu pour l'équation intégrale sur $[t_{n-1}, t_{n+1}]$:

$$y(t_{n+1}) - y(t_{n-1}) = \int_{t_{n-1}}^{t_{n+1}} f(t, y(t))\, dt \sim 2h f(t_n, y(t_n)).$$

Ex 2 : La méthode de Simpson.

C'est un schéma a deux pas, implicite :

$$y_{n+2} - y_n = \frac{h}{3}\left(f(t_{n+2}, y_{n+2}) + 4 f(t_{n+1}, y_{n+1}) + f(t_n, y_n)\right) \tag{4}$$

il est obtenu en écrivant la formule d'intégration numérique de Simpson sur $[t_n, t_{n+2}]$

$$y(t_{n+2}) - y(t_n) = \int_{t_n}^{t_{n+2}} f(t, y(t))\, dt$$

$$\sim \frac{2h}{6}\left[f(t_n, y(t_n)) + 4 f(t_{n+1}, y(t_{n+1})) + f(t_{n+2}, y(t_{n+2}))\right]$$

Pour simplifier l'écriture, on introduit les premier et second polynômes caractéristiques de la méthode, polynômes de degré $\leq q$ définis par :

$$(5) \qquad \rho(\zeta) = \sum_{j=0}^{q} \alpha_j \zeta^j \qquad \sigma(\zeta) = \sum_{j=0}^{q} \beta_j \zeta^j$$

La méthode est entièrement déterminée par le pas h, ρ et σ. On pasiera de méthode (ρ, σ).

1. Convergence et stabilité.

A la méthode (2) on associe l'opérateur $\mathcal{L}$ qui représente l'application du schéma à la solution de (1) :

$$(6) \qquad \mathcal{L}(y(t); h) = \sum_{j=0}^{q} \alpha_j y(t+jh) - h \sum_{j=0}^{q} \beta_j f(t+jh, y(t+jh))$$

Définition 1 : (i) l'erreur de troncature locale de la méthode au temps t_n est définie par :

$$(7) \qquad \varepsilon_n = \mathcal{L}(y(t_n); h) \qquad 0 \leq n \leq N-q$$

où y est la solution de (1)

(ii) la méthode est dite d'ordre p si p est le plus grand entier tel que $\mathcal{L}(y(t), h) = O(h^{p+1})$ pour toute solution y de (1) $(p+1)$ fois continuement dérivable.

(iii) la méthode est dite consistante si elle est d'ordre ≥ 1.

D'après la définition la suite z_n définie par $z_n = y(t_n)$ est solution du schéma perturbé :

$$(8) \qquad \sum_0^q \alpha_j z_{n+j} = h \sum_0^q \beta_j f(t_{n+j}, z_{n+j}) + \varepsilon_n \qquad 0 \leq n \leq N-q$$

Comme pour les méthodes de Runge-Kutta, on peut donner une condition nécessaire et suffisante pour que la méthode (ρ, σ) soit d'ordre $\geq p$. Pour cela notons

$$(9) \quad \left| \begin{array}{l} c_0 = \sum_{j=0}^q \alpha_j = \rho(1) \\ c_1 = \sum_{j=0}^q j\alpha_j - \sum_{j=0}^q \beta_j = \rho'(1) - \sigma(1) \\ c_\ell = \sum_{j=0}^q \frac{1}{\ell!} j^\ell \alpha_j - \sum_{j=0}^q \frac{1}{(\ell-1)!} j^{\ell-1} \beta_j . \end{array} \right.$$

<u>Théorème 1</u> : La méthode (ρ, σ) est d'ordre $\geq p$ si et seulement si

$$(10) \qquad c_\ell \equiv 0 \qquad 0 \leq \ell \leq p .$$

Si y est de classe $\mathcal{C}^{p+2}$ on a alors

$$\mathcal{L}(y(t), h) = c_{p+1} h^{p+1} y^{(p+1)}(t) + O(h^{p+2}) \quad \text{où } c_{p+1} \neq 0$$

exo ordre ≥ 1 $\Longleftrightarrow$ $\begin{cases} e(1)=0 \\ e'(1)-\sigma(1)=0 \end{cases}$

$$\mathcal{L} = \sum_0^q \alpha_j\, y(t+jh) - h\sum_0^q \beta_j\, y'(t+jh)$$

$$y(t+jh) = \sum_0^p \frac{(jh)^k}{k!}\, y^{(k)}(t) + \frac{(jh)^{p+1}}{(p+1)!}\, y^{(p+1)}(t+\theta_j jh) \qquad 0<\theta_j<j$$

$$h\, y'(t+jh) = h\sum_0^{p-1} \frac{(jh)^k}{k!}\, y^{(k+1)}(t) + \frac{(jh)^p}{p!}\, h\, y^{(p+1)}(t+\eta_j h) \qquad 0<\eta_j<j$$

$$k+1 = \ell$$

$$\sum_{k=0}^{p} \qquad \sum_1^p \frac{j^{\ell-1} h^\ell}{(\ell-1)!}\, y^{(\ell)}(t) + \frac{h^{p+1} j^p}{p!}\, y^{(p+1)}(t+\eta_j h)$$

$$\sum y(t)\left(\sum_0^q \alpha_j\right)$$

$$+ \sum_1^p \left[\sum_0^q \frac{(jh)^k}{k!}\alpha_j\, y^{(k)}(t) - \sum_0^q \frac{j^{k-1} h^k}{(k-1)!}\, y^{(k)}(t)\,\beta_j \right]$$

$$+ \sum_0^q \alpha_j \frac{(jh)^{p+1}}{(p+1)!}\, y^{(p+1)}(t+\theta_j h) - \sum_0^q \beta_j \frac{j^p h^{p+1}}{p!}\, y^{(p+1)}(t+\eta_j h)$$

$$y(t)\left(\sum_0^q \alpha_j\right) + \sum_1^p \frac{h^k}{k!}\, y^{(k)}(t)\left(\sum_0^q \left(j^k \alpha_j - k j^{k-1}\beta_j\right)\right)$$

Démonstration :

supposons d'abord y de classe $\mathcal{C}^{p+1}$, et effectuons un développement de Taylor au temps t à l'ordre $p+1$:

$$\sum_{j=0}^{q}\alpha_j\, y(t+jh) = \Big(\sum_{j=0}^{q}\alpha_j\Big)\, y(t) + h\sum_{j=0}^{q}(j\,\alpha_j)\, y'(t) + \dots$$

$$+\frac{h^p}{p!}\Big(\sum_{j=0}^{q} j^p\alpha_j\Big)\, y^{(p)}(t) + \frac{h^{p+1}}{(p+1)!}\sum_{j=0}^{q} j^{p+1}\alpha_j\, y^{(p+1)}(t+\theta_j h)$$

$$0<\theta_j<j\,.$$

de même :

$$\sum_{j=0}^{q}\beta_j\, y'(t+jh) = \Big(\sum_{j=0}^{q}\beta_j\Big)\, y'(t) + h\Big(\sum_{j=0}^{q} j\,\beta_j\Big)\, y''(t) + \dots$$

$$+\frac{h^{p-1}}{(p-1)!}\Big(\sum_{j=0}^{q} j^{p-1}\beta_j\Big) y^{(p)}(t) + \frac{h^p}{p!}\sum_{j=0}^{q} j^p\beta_j\, y^{(p+1)}(t+\eta_j h)$$

$$0<\eta_j<j$$

d'où

$$\mathcal{L}(y(t),h) = \sum_{\ell=0}^{p} c_\ell\, h^\ell\, y^{(\ell)}(t) + h^{p+1}\Big[\frac{1}{(p+1)!}\sum_{j=0}^{q} j^{p+1}\alpha_j\, y^{(p+1)}(t+\theta_j h)$$

$$-\frac{1}{p!}\sum_{j=0}^{q} j^p\beta_j\, y^{(p+1)}(t+\eta_j h)\Big]$$

donc $\mathcal{L}(y(t),h) = O(h^{p+1}) \iff c_0 = c_1 = \dots = c_p = 0$.

Si y est de classe $\mathcal{C}^{p+2}$, on peut alors écrire

$$\mathcal{L}(y(t),h) = c_{p+1}\, h^{p+1}\, y^{(p+1)}(t) + O(h^{p+2})$$

Remarquons que la méthode est exactement d'ordre p si $C_{p+1} \neq 0$.

D'autre part le théorème 1 s'écrit pour $p = 1$:

Corollaire 1 : La méthode (ρ, σ) est consistante si et seulement si

(11) $\qquad \rho(1) = 0 \qquad \rho'(1) - \sigma(1) = 0.$

Corollaire 2 : Si la méthode est consistante on a :

$$\lim_{h \to 0} \sum_{0}^{N-q} |\varepsilon_n| = 0.$$

La démonstration est analogue à celle faite dans le cas d'une méthode à 1 pas :

$$|\varepsilon_n| \leq C h^2 \qquad 0 \leq n \leq N-q$$

d'où

$$\sum_{0}^{N-q} \varepsilon_n \leq C h T$$

Définition 2 : On dit que la méthode (ρ, σ) est zéro-stable s'il existe une constante $M > 0$ indépendante de h telle que

(12) $$|y_n - z_n| \leq M \left\{ \max_{0 \leq \nu \leq q-1} |y_\nu - z_\nu| + \sum_{\ell=0}^{n-q} |\eta_\ell| \right\} \quad q \leq n \leq N$$

où z_n est solution du schéma perturbé

$$\sum_{j=0}^{q} \alpha_j z_{n+j} = h \sum_{j=0}^{q} \beta_j f(t_{n+j}, z_{n+j}) + \eta_n \qquad 0 \leq n \leq N-q.$$

et l'on a de nouveau le résultat

Théorème 2 : une méthode consistante et zero-stable est convergente,

i.e $\lim_{\substack{h\to 0 \\ nh=t}} y_n = y(t)$

pour tout t dans $[0,T]$, pour toute solution y_n du schéma satisfaisant les conditions initiales $y_k = y_k(h)$ pour lesquelles $\lim_{h\to 0} y_k(h) = y^0$, $k=0,1,\dots,q-1$

La démonstration est la même qu'au chapitre 2 : on choisit $z_n = y(t_n)$ dans (12) et on applique le corollaire 2.

La zéro-stabilité s'interprète par des conditions algébriques sur le polynôme ρ :

Définition 3 : On dit que le polynôme ρ est stable si ses racines ξ_i vérifient les propriétés :

(13)
(i) $|\xi_i| \leq 1$, $1 \leq i \leq q$
(ii) les racines ξ_i de module 1 sont simples.

et l'on a alors

Théorème 4 : La méthode (ρ, σ) est zéro-stable si et seulement si le polynôme ρ est stable.

Nous ne démontrerons ici que la condition nécessaire. Si la méthode est zéro stable, choisissons $f \equiv 0$, et $z_n \equiv 0$. La condition de stabilité s'écrit alors

$$|y_n| \leq M \max_{0 \leq \nu \leq q-1} |y_\nu| \tag{14}$$

c'est-à-dire que la suite y_n est bornée indépendamment des données initiales. Et le schéma s'écrit

$$\sum_{j=0}^{q} \alpha_j y_{n+j} = 0 \tag{15}$$

Pour résoudre cette équation aux différences (équivalent discret d'une équation différentielle linéaire d'ordre q) nous écrivons un système en posant :

$$Y_n = \begin{pmatrix} y_{n+q-1} \\ \vdots \\ y_n \end{pmatrix}$$

et (15) se réécrit

$$Y_{n+1} = A Y_n \qquad A = \begin{pmatrix} -\alpha_{q-1} & -\alpha_{q-2} & \cdots & -\alpha_0 \\ 1 & 0 & & 0 \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}$$

on en déduit que

$$Y_n = A^n Y_0 . \qquad \text{où } Y_0 = \begin{pmatrix} y_{q-1} \\ \vdots \\ y_0 \end{pmatrix}.$$

donc (14) est équivalent à dire que

$$\|Y_n\| \leq M \|Y_0\|. \qquad q \leq n \leq N$$

où pour un vecteur $Z = \begin{pmatrix} z_1 \\ \vdots \\ z_q \end{pmatrix}$ $\|Z\| = \max |z_q|$.

Pour calculer A^n, nous mettons A sous forme de Jordan. Faisons une remarque préliminaire : pour toute valeur propre de A : ζ,

$$\dim \ker (A - \zeta I) = 1$$

en effet $A - \zeta I = \begin{pmatrix} -a_{q-1} - \zeta & -a_{q-2} & \cdots & -a_0 \\ 1 & -\zeta & 0 & 0 \\ 0 & 1 & -\zeta & 0 \\ & 0 & \ddots & \vdots \\ & & 1 & -\zeta \end{pmatrix}$

le déterminant constitué des $q-1$ dernières lignes et $q-1$ premières colonnes est égal à 1, donc le rang de $A - \zeta I$ est $q-1$ et la dimension de son noyau est 1. Donc la dimension de l'espace propre associé à ζ n'est égale à sa multiplicité que si ζ est valeur propre simple.

Mettons maintenant A sous forme de Jordan

$$A = P J P^{-1} \qquad J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_N \end{pmatrix}$$

$J_k = [\zeta_k]$ si ζ_k valeur propre simple.

$J_k = \zeta_k I + N_k = \begin{bmatrix} \zeta_k & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \zeta_k \end{bmatrix}$ J_k dimension ν_k

$N_k^{\nu_k} \equiv 0$

si ν_k = multiplicité $\zeta_k > 1$.

Donc $J^n = \begin{pmatrix} J_1^n & & \\ & \ddots & \\ & & J_m^n \end{pmatrix}$.

$J_k = [\zeta_k^n]$ si $\nu_k = 1$

$J_k^n = \sum_0^{\nu_k - 1} C_n^j \zeta_k^{n-j} N_k^j = \begin{pmatrix} \zeta_k^n & n\zeta_k^{n-1} & \cdots & C_n^{\nu_k-1}\zeta_k^{n-\nu_k+1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & n\zeta_k^{n-1} \\ & & & \zeta_k^n \end{pmatrix}$ si $\nu_k > 1$

et $Y_n = P J^n P^{-1} Y_0$.

chaque composante de Y_n sera donc combinaison linéaire de termes du type

ζ_k^n $\quad \nu_k = 1$

$\zeta_k^n, n\zeta_k^{n-1}, \frac{n(n-1)}{2}\zeta_k^{n-2} \ldots, \frac{n(n-1)\cdots(n-\nu_k+1)}{(\nu_k-1)!}\zeta_k^{n-\nu_k+1}$

pour $\nu_k > 1$.

Y_n ne peut être borné pour tout n que si tous ces termes sont bornés ce qui suppose

$|\zeta_k| \leq 1$ si $\nu_k = 1$

$|\zeta_k| < 1$ si $\nu_k > 1$.

Donc si le schéma est stable, les valeurs propres de A sont toutes de module ≤ 1, et celles de module 1 sont simples. Pour terminer notre démonstration, il nous reste à prouver que $\rho(\zeta)$ est le polynôme caractéristique de A. Pour cela développons $\det(A-\lambda I)$ par rapport à la première ligne. Il vient

$$\det(A-\lambda I) = (-\lambda)^{q-1}(-\alpha_{q-1}-\lambda) + (-\lambda)^{q-2}\alpha_{q-2} \dots + (-1)^q \alpha_0$$
$$= (-1)^q \rho(\lambda).$$

■

Pour un nombre de pas q donné, quel ordre de convergence peut-on espérer ? Une méthode à q pas dépend de $2q+1$ coefficients α_j et β_j (α_q est fixé à 1), qui se réduisent à $2q$ si la méthode est explicite. Pour atteindre l'ordre p, on écrit que les $(p+1)$ ièrs termes dans le développement de Taylor de l'erreur de troncature s'annulent : on a donc $p+1$ équations à $2q+1$ ou $2q$ inconnues. Pour pouvoir résoudre ce système, il faut donc que $p \leq 2q$ ou $2q-1$. Si l'on veut de plus que la méthode soit zéro stable, on a le

Théorème 5 : Il n'existe pas de méthode à q pas, zéro stable, d'ordre $p \geq q+1$ si q est impair, $q+2$ si q est pair.

Une méthode à q pas, zéro-stable, d'ordre $q+2$ est dite optimale.

Exemple : la méthode de Simpson est optimale : $q=2$ $p=4$.

2. Les méthodes multipas usuelles.

Les méthodes d'Adams utilisent la formule

$$y(t_{n+q}) - y(t_{n+q-1}) = \int_{t_{n+q-1}}^{t_{n+q}} y'(t)\,dt = \int_{t_{n+q-1}}^{t_{n+q}} f(t, y(t))\,dt.$$

et approchent l'intégrale par l'intégrale du polynôme d'interpolation de Lagrange.

* Méthodes d'Adams-Bashforth

Il s'agit de méthodes explicites. Supposons connus $y_n, \dots, y_{n+q-1}$. Posons

(16) $$f_k = f(t_k, y_k).$$

Notons $P_{n,q}$ le polynôme de degré $\leq q-1$ qui interpole $y'(t)$ aux points $t_n, \dots, t_{n+q-1}$.

$$P_{n,q}(t_{n+i}) = f_{n+i} \qquad i=0,\dots,q-1. \tag{17}$$

Le schéma d'Adams-Bashforth à q pas s'écrit :

$$y_{n+q} - y_{n+q-1} = \int_{t_{n+q-1}}^{t_{n+q}} P_{n,q}(t)\,dt \tag{18}$$

D'après la formule de Newton-régressive on a :

$$P_{n,q}(t) = \sum_{j=0}^{q-1} \binom{s+j-1}{j} \bar{\nabla}^j f_{n+q-1} \tag{19}$$

où $\bar{\nabla} f_k = f_k - f_{k-1}$

$\bar{\nabla}^2 f_k = \bar{\nabla}(\bar{\nabla} f_k) = f_k - 2f_{k-1} + f_{k-2}$, etc...

et $s = \dfrac{t - t_{n+q-1}}{h}$.

on a alors :

$$\int_{t_{n+q-1}}^{t_{n+q}} P_{n+q}(t)\,dt = \sum_{j=0}^{q-1} \nabla^j f_{n+q-1} \int_{t_{n+q-1}}^{t_{n+q}} \binom{s+j-1}{j} dt$$

$$s = \frac{t - t_{n+q-1}}{h}.$$

et

$$(20) \qquad y_{n+q} - y_{n+q-1} = h \sum_{j=0}^{q-1} \gamma_j \nabla^j f_{n+q-1}$$

$$(21) \qquad \gamma_j = \int_0^1 \binom{s+j-1}{j} ds.$$

Les γ_j peuvent être calculés par récurrence en notant que :

$$(22) \quad \begin{cases} \gamma_0 = 1 \\ \forall i, \quad \dfrac{\gamma_0}{i+1} + \dfrac{\gamma_1}{i} + \dfrac{\gamma_2}{i-1} + \cdots + \dfrac{\gamma_{i-1}}{2} + \gamma_i = 1 \end{cases}$$

La méthode se met sous la forme (2)

$$(23) \qquad y_{n+q} - y_{n+q-1} = h \sum_{j=0}^{q-1} \beta_{j,q} f_{n+j}$$

avec

$$(24) \qquad \beta_{q-1-i,q} = (-1)^i \sum_{j=i}^{q-1} \binom{j}{i} \gamma_j$$

(en utilisant la formule $\nabla^i f_k = \sum_{i=0}^{i} (-1)^i \binom{i}{i} f_{k-i}$)

* $q = 1$. $\gamma_0 = 1$

$\beta_{0,1} = 1$

$$y_{n+1} - y_n = h f_n$$

c'est la méthode d'Euler

on peut écrire pour les $\beta_{j,q}$ le tableau suivant

	$\beta_{0,q}$	$\beta_{1,q}$	$\beta_{2,q}$	$\beta_{3,q}$.
$q=1$	1			
$q=2$	$-\frac{1}{2}$	$\frac{3}{2}$		
$q=3$	$\frac{5}{12}$	$-\frac{4}{3}$	$\frac{23}{12}$	
$q=4$	$-\frac{3}{8}$	$\frac{37}{24}$	$-\frac{59}{24}$	$\frac{55}{24}$.

Théorème 6: une méthode d'Adams-Bashforth à q pas est d'ordre $\geq q$.

Démonstration: il suffit d'appliquer le résultat d'erreur sur le polynôme d'interpolation de Lagrange.

on a

$$\varepsilon_n = y(t_{n+q}) - y(t_{n+q-1}) - \int_{t_{n+q-1}}^{t_{n+q}} \tilde{P}_{n,q}(t)\, dt$$

avec $$\tilde{P}_{n,q}(t) = \sum_{j=0}^{q-1} \binom{s+j-1}{j} \nabla^j \tilde{f}_{n+q-1}$$

$$\tilde{f}_k = f(t_k, y(t_{n+k})).$$

et $$y'(t) = \tilde{P}_{n,q}(t) + h^q \binom{s+q-1}{q} y^{(q+2)}(\xi_t).$$

d'où

$$\varepsilon_n = \int_{t_{n+q-1}}^{t_{n+q}} [y'(t) - \tilde{P}_{n,q}(t)]\, dt$$

$$= h^q \int_{t_{n+q-1}}^{t_{n+q}} \binom{s+q-1}{q} y^{(q+2)}(\xi_t)\, dt$$

et $$|\varepsilon_n| \leq C h^{q+1}.$$ ∎

On peut en fait montrer que ces méthodes sont exactement d'ordre q.

* les méthodes d'Adams-Moulton à q pas.

on interpole maintenant $y'(t)$ par le polynôme $Q_{n,q}$ de degré $\leq q$, aux points $t_n, \dots, t_{n+q}$:

$$Q_{n,q}(t_{n+k}) = f_{n+k} \qquad k = 0, \dots, q. \tag{25}$$

le polynôme $Q_{n,q}$ s'écrit alors

$$Q_{n,q}(t) = \sum_{j=0}^{q} f_{n+i} \; L_{n,i,q}(t) \tag{26}$$

$$L^*_{n,i,q}(t) = \prod_{\substack{j=0 \\ j \neq i}}^{q} \frac{t - t_{n+j}}{t_{n+i} - t_{n+j}}$$

et le schéma d'Adams-Moulton est :

$$y_{n+q} - y_{n+q-1} = h \sum_{j=0}^{q} \beta^*_{j,q} \, f_{n+j} \,. \tag{27}$$

$$h \, \beta^*_{i,q} = \int_{t_{n+q-1}}^{t_{n+q}} L^*_{n,i,q}(t) \, dt$$

$$\beta^*_{i,q} = \int_0^1 \prod_{\substack{j=0 \\ j \neq i}}^{q} \frac{q - 1 - j + s}{i - j} \, ds. \tag{28}$$

C'est un schéma implicite. Nous avons vu qu'il étant résoluble en y_{n+q} si et seulement si

$$h \, |\beta_q| \, L < 1$$

où L est la constante de Lipschitz associée à L.

Nous nous placerons désormais toujours dans ce cadre.

utilisons de nouveau la formule de Newton régressive.

$$(29)\quad Q_{n,q}(t) = \sum_{j=0}^{q} \binom{s+j-1}{j} \nabla^j f_{n+q}.$$

$$s = \frac{t - t_{n+q}}{h}.$$

d'où

$$\int_{t_{n+q-1}}^{t_{n+q}} Q_{n,q}(t)\,dt = \sum_{j=0}^{q} \nabla^j f_{n+q} \int_{t_{n+q-1}}^{t_{n+q}} \binom{s+j-1}{j} dt$$

$$s = \frac{t - t_{n+q}}{h}.$$

$$(30)\quad y_{n+q} - y_{n+q-1} = h \sum_{j=0}^{q} \gamma_j^* \nabla^j f_{n+q}$$

$$(31)\quad \gamma_j^* = \int_0^1 \binom{s+j-2}{j} ds.$$

en utilisant la formule

$$\binom{n+k}{k} = \binom{n+k-1}{k} + \binom{n+k-1}{k-1}.$$

on montre que

$$(32)\quad \gamma_j^* = \gamma_j - \gamma_{j-1} \qquad j \geq 1$$

et l'on a

$$\gamma_0^* = 1.$$

on peut donc ainsi calculer les γ_j^*.

Les β_j^* s'en déduisent comme dans le cas d'Adams - Bashforth :

$$\sum_{j=0}^{q} \gamma_j^* \nabla^j f_{n+q} = \sum_{j=0}^{q} \gamma_j^* \sum_{i=0}^{j} (-1)^i \binom{j}{i} f_{n+q-i}$$

$$= \sum_{i=0}^{q} (-1)^i \left[\sum_{j=i}^{q} \gamma_j^* \binom{j}{i} \right] f_{n+q-i}$$

si bien que

$$\beta_{q-i,q}^* = (-1)^i \sum_{j=i}^{q} \binom{j}{i} \gamma_j^* \tag{33}$$

pour

$q = 1$, on obtient :

$$\gamma_0^* = 1 \qquad \gamma_1^* = -\frac{1}{2}.$$

$$\beta_{0,1}^* = \frac{1}{2} \qquad \beta_{1,1}^* = \frac{1}{2}$$

$$y_{n+1} - y_n = \frac{1}{2}(f_n + f_{n+1})$$

C'est la méthode de Runge Kutta .

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

ou méthode d'Euler amélioré.

on obtient alors le tableau

	$b^*_{0,q}$	$b^*_{1,q}$	$b^*_{2,q}$	$b^*_{3,q}$	$b^*_{4,q}$.
$q=1$	$\frac{1}{2}$	$\frac{1}{2}$			
$q=2$	$-\frac{1}{12}$	$\frac{2}{3}$	$\frac{5}{12}$		
$q=3$	$\frac{1}{24}$	$-\frac{5}{24}$	$\frac{19}{24}$	$\frac{9}{24}$	
$q=4$.	$-\frac{19}{720}$	$+\frac{53}{360}$	$-\frac{11}{30}$	$+\frac{323}{360}$.	$\frac{251}{720}$

Théorème 7 : les méthodes d'Adams-Moulton à q pas sont d'ordre $\geq q+1$.

La démonstration est identique à celle du théorème 6.

Les méthodes implicites sont donc, à nombre de pas égal, d'ordre plus élevé que les méthodes explicites.

Théorème 8 : les méthodes d'Adams sont zero-stable.

Démonstration : le polynôme $\rho(\xi)$ est ici

(34) $$\rho(\xi) = \xi^q - \xi^{q-1}$$

le résultat est donc un simple corollaire du théorème 4.

Méthodes prédicteurs - correcteurs

Nous avons vu que pour un nombre donné de pas, les méthodes d'Adams-Moulton (implicites) sont plus précises que les méthodes d'Adams-Bashforth (explicites). D'autre part, nous verrons qu'elles ont de meilleures propriétés de stabilité absolue. D'un point de vue théorique, elles sont donc plus intéressantes. Par contre, elles nécessitent d'utiliser un schéma itératif.

$$y_{n+q}^{[s]} = h\beta_q f(t_{n+q}, y_{n+q}^{[s-1]}) + \sum_{0}^{q-1} (-a_i y_{n+i} + h\beta_i f_{n+i})$$

$y_{n+q}^{[0]}$ donné. pour $|h\beta_q L| < 1$

Elles ne seront intéressantes d'un point de vue numérique que si le nombre d'itération est petit. Pour cela il faut avoir eu $y_{n+q}^{[0]}$ une (déjà) bonne approximation de y_{n+q}, et ne faire si possible qu'une itération. La méthode la plus simple consiste à calculer $y_{n+q}^{[0]}$ par une méthode explicite à q^* pas (phase de prédiction) puis à corriger cette valeur :

notons y^*_{n+q} la solution de

$$(35)\quad y^*_{n+q} + \sum_{i=q-q^*}^{q-1} \alpha^*_i y_{n+i} = h \sum_{i=q-q^*}^{q-1} \beta^*_i f_{n+i}$$

on utilise alors la méthode implicite (ρ,σ) à q pas comme méthode de correction, ~~exclamation~~. On calcule donc y_{n+q} à l'aide de la formule

$$(36)\quad \sum_{i=0}^{q} \alpha_i y_{n+i} = h \beta_q f(t_{n+q}, y^*_{n+q}) + h \sum_{0}^{q-1} \beta_i f_{n+i}$$

La question est maintenant de choisir la méthode explicite (c'est-à-dire ses polynômes ρ^* et σ^* et le nombre de pas q^*) de façon que la méthode de prédiction correction ait les mêmes propriétés de stabilité et de convergence que la méthode implicite de départ

Théorème 9 : On suppose que la méthode implicite est 0-stable et d'ordre p. Alors si la méthode de prédiction (35) est d'ordre $p^* \geq p-1$, la méthode de ~~de~~ prédiction-correction est zéro stable et d'ordre p.

Application : méthode prédicteur : Adams-Bashforth à $q+1$ pas
méthode correcteur : Adams-Moulton à q pas

ex : M.P : $y^*_{n+3} - y_{n+2} = \frac{h}{12}(5f_n - 16 f_{n+1} + 23 f_{n+2})$

M.C $y_{n+3} - y^*_{n+3} = \frac{h}{12}(-f_{n+1} + 8 f_{n+2} + 5 f_{n+3})$.

3. Stabilité absolue.

Considérons l'équation différentielle scalaire

(37) $\quad y' = -\lambda y. \qquad \lambda \in$

Appliquée à (37) la méthode (ρ, σ) s'écrit

(38) $\quad \sum_{i=0}^{q} (\alpha_i + \lambda h \beta_i) y_{n+i} = 0.$

Posons $z = \lambda h$ et désignons par $y_n(z)$ la solution de l'équation aux différences

(39) $\quad \sum_{i=0}^{q} (\alpha_i + z\beta_i) y_{n+i}(z) = 0.$

associée aux données initiales

$$y_i(z) = y_i \qquad 0 \le i \le q-1$$

Définition 4 : On appelle domaine de stabilité absolue de la méthode (ρ, σ) l'ensemble des z dans $\mathbb{R}_*^+$ pour lesquels (39) admet une solution unique $(y_n(z))_{n \ge 0}$, telle que

$$\sup_{n \ge 0} |y_n(z)| < +\infty$$

quelles que soient les conditions initiales $y_0 \dots y_{q-1}$.

Introduisons alors le polynôme

$$\pi(\zeta, z) = \rho(\zeta) + z\sigma(\zeta).$$

et notons $\xi_i(z), 1 \leq i \leq q$ les racines de Π.
on a alors le

Théorème 10 : le point z appartient au domaine de stabilité absolue de la méthode si et seulement si :

(i) $|\xi_i(z)| \leq 1 \quad 1 \leq i \leq q$.

(ii) les racines de module 1 sont simples.

Nous ne démontrerons pas ce résultat, il est analogue au théorème 4. Remarquons que la 0-stabilité signifie que 0 appartient au domaine de stabilité absolue.

Il est plus important de savoir comment déterminer le domaine d'absolue stabilité de la méthode. Pour cela on utilise le théorème de Schurr. Soit ψ un polynôme à coefficients (éventuellement) complexes de degré q :

$$\psi(r) = c_q r^q + c_{q-1} r^{q-1} + \cdots + c_1 r + c_0$$

on suppose que c_0 et c_q sont tous deux non nuls.

Définition 5 : ψ est un polynôme de Schurr si toutes ses racines sont de module strictement inférieur à 1.

On définit alors le conjugué de φ comme :

$$\hat{\varphi}(r) = \overline{c_0} r^q + \overline{c_1} r^{q-1} + \dots + \overline{c_{q-1}} r + \overline{c_q}$$

et

$$\varphi_1(r) = \frac{1}{r}\left[\hat{\varphi}(0)\varphi(r) - \varphi(0)\hat{\varphi}(r)\right]$$

Théorème 11 (Schurr) : φ est un polynôme de Schurr si et seulement si $|\hat{\varphi}(0)| > |\varphi(0)|$ et φ_1 est un polynôme de Schurr.

En itérant le procédé, on est alors amené à chercher une relation entre les coefficients d'une équation du premier degré pour que le module de la racine soit inférieur à 1. si l'on ne regarde que les racines réelles, cela donne un critère simple.

Exemple 1) : A.B à 2 pas

$$y_{n+2} - y_{n+1} = \frac{\Delta t}{2}(3f_{n+1} - f_n)$$

$$\rho(r) = r^2 - r \quad , \quad \sigma(r) = \frac{3}{2}r - \frac{1}{2}.$$

$$\Pi(r, z) = r^2 + \left(\frac{3}{2}z - 1\right)r - \frac{z}{2}.$$

si $z \in \mathbb{R}$, $\quad \hat{\Pi}(r, z) = -\frac{z}{2}r^2 + \left(\frac{3}{2}z - 1\right)r + 1.$

La condition $|\hat{\Pi}(0,z)| > |\Pi(0,z)|$ s'écrit

$$1 > \left|\frac{z}{2}\right|$$

elle est remplie si et seulement si $\delta \in]-2,2[$

$$\Pi_1(r,z) = \left(1+\frac{z}{2}\right)\left[\left(1-\frac{z}{2}\right)r + \frac{3}{2}z - 1\right]$$

la seule racine de Π_1 est

$$r = \frac{1-\frac{3}{2}z}{1-\frac{z}{2}}.$$

et l'on a

$$|r|<1 \iff z \in]0,1[.$$

Π est donc un polynôme de ~~Guarego~~ Schurr si et seulement si $z \in]0,1[$. Pour $z=0$, $\Pi(r,0)=\rho(r)$, la racine $r=1$ est simple, et pour $z=1$, $\Pi(r,1)=(r+1)(r-\frac{1}{2})$, la racine $r=-1$ est simple. L'intervalle d'absolue stabilité est donc $[0,1]$

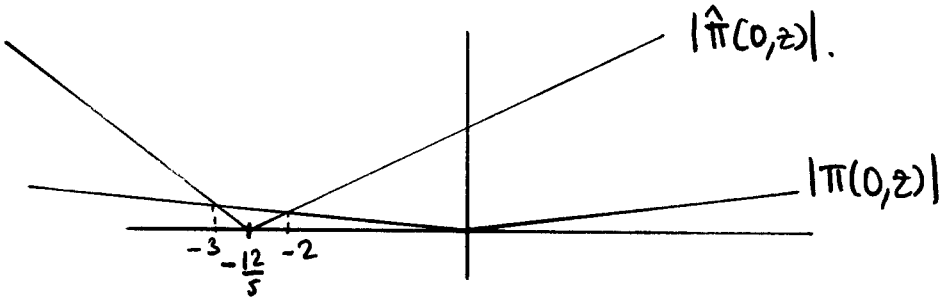
Exemple 2 : Méthode d'Adams-Moulton à 2 pas

$$y_{n+2} - y_{n+1} = \frac{1}{12}(5f_{n+2} + 8f_{n+1} - f_n)$$

$$\rho = r^2 - r, \quad z = \frac{1}{12}(5r^2 + 8r - 1).$$

$$\Pi(r,z) = \left(1+\frac{5}{12}z\right)r^2 + \left(-1+\frac{2}{3}z\right)r - \frac{z}{12}$$

$$|\Pi(0,z)| = \frac{|z|}{12} \qquad |\hat{\Pi}(0,z)| = \left|1+\frac{5z}{12}\right|.$$



d'où

$$|\hat{\Pi}(0,z)| > |\Pi(0,z)| \iff z \leq -3 \text{ ou } z \geq -2.$$

$$\Pi_1(r,z) = \left(1+\frac{z}{2}\right)\left[\left(1+\frac{z}{3}\right) r + \left(-1+\frac{2}{3}z\right)\right]$$

d'où

Π_1 est un polynôme de Schurr si et seulement si

$$0 < z < 6.$$

donc Π est un polynôme de Schurr si et seulement si

$$0 < z < 6$$

et de même le ~~polyt~~ domaine d'absolue stabilité est $[0,6]$. Il s'agit d'un phénomène général : le domaine d'absolue stabilité pour les méthodes d'Adams Moulton sont plus grands que pour la méthode d'Adams - Bashforth. Elles permettent donc de prendre une pas de temps plus grand.

q	1	2	3	4.
AB	$[0,2]$	$[0,1]$	$[0,\frac{6}{11}]$	$[0,\frac{3}{10}]$
AM	$[0,+\infty[$	$[0,6]$	$[0,3]$	$[0,\frac{90}{49}]$.

Remarque : contrairement aux méthodes de Runge-Kutta, le domaine d'absolue stabilité décroit lorsque l'ordre croit.