Intégration numérique des équations différentielles avec condition initiale

On cherche $y : [a,b] \to \mathbb{R}^n$ fonction dérivable solution de :

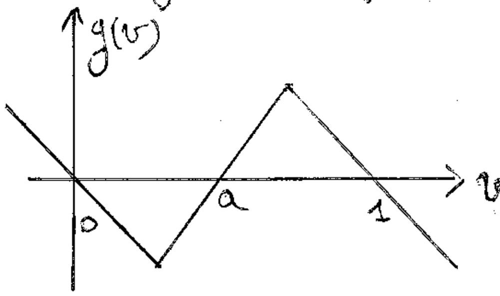
$$(P) \begin{cases} \dfrac{dy}{dt} = f(t,y) , & t \in [a,b] & (E) \\ y(t_0) = y_0 , & t_0 \in [a,b[, \; y_0 \in \mathbb{R}^n & (CI) \end{cases}$$

(P) est appelé un "problème de Cauchy".
Il est composé de l'équation différentielle (E) et de la condition initiale (CI).

<u>Exemple</u> : équation de Nagumo discrète

$$\begin{cases} y_1' = k\,(y_2 - 2y_1 + V(t)) + g(y_1) \\ y_i' = k\,(y_{i+1} - 2y_i + y_{i-1}) + g(y_i) , & 2 \le i \le n-1 \\ y_n' = k\,(y_{n-1} - y_n) + g(y_n) \end{cases}$$

où g est une fonction non linéaire du type :



et $V(t)$ un signal d'entrée $\left(V \in C^0(\mathbb{R}^+, \mathbb{R}) \right)$

C'est une équation de réaction-diffusion discrète qui a des applications multiples (propagation de l'influx nerveux, circuits, ...).

Ici $y(t) = \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix}$, $f(t,y) = \begin{pmatrix} f_1(t,y) \\ f_2(y) \\ \vdots \\ f_n(y) \end{pmatrix}$,

$f_1(t,y) = k(y_2 - 2y_1 + V(t)) + g(y_1)$, $f_i(y) = k\,(y_{i+1} - 2y_i + y_{i-1}) + g(y_i)$, $2 \le i \le n-1$

$f_n(y) = k\,(y_{n-1} - y_n) + g(y_n)$

On fait l'hypothèse suivante (H) sur f :

(H)
- $f : [a,b] \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ est continue
- f est lipschitzienne par rapport à y uniformément par rapport à $t \in [a,b]$, c'est à dire :

  $\exists L \geq 0 \;/\; \forall y_1, y_2 \in \mathbb{R}^n, \;\; \forall t \in [a,b],$

  $$\| f(t,y_1) - f(t,y_2) \| \leq L \| y_1 - y_2 \|$$

<u>Remarque</u> on considère une norme arbitraire sur $\mathbb{R}^n$.
Si (H) est vérifiée pour une norme, alors (H) est vraie pour toute norme par équivalence des normes sur $\mathbb{R}^n$.

On a alors le résultat suivant :

<u>Théorème de Cauchy-Lipschitz global :</u>

Sous l'hypothèse (H), le problème de Cauchy (P) admet une solution unique $y \in C^1([a,b], \mathbb{R}^n)$.
De plus, si f est C^k alors y est C^{k+1}.

Ce théorème se démontre en appliquant le théorème du point fixe contractant. En général, la solution y n'est pas connue explicitement. Il est donc important de pouvoir calculer y numériquement.

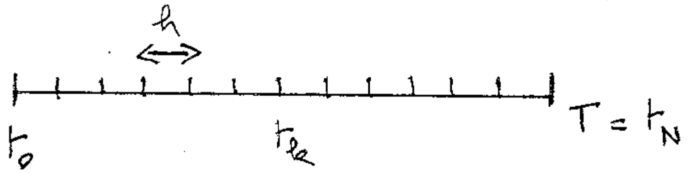
<u>Remarque</u> Les méthodes numériques que nous allons décrire restent valables lorsque f est C^k ($k \geq 1$) sans être forcément lipschitzienne en y comme dans l'hypothèse (H) (exemple : $f(y) = y^2$).
Cependant la solution y n'est pas forcément définie sur $[a,b]$ tout entier (seulement sur un intervalle ouvert autour de t_0).

<u>Approximation numérique de y :</u> sur l'intervalle $[t_0, T] \subset [a,b]$.

On note $h = \frac{T-t_0}{N}$ le pas de discrétisation avec $N \in \mathbb{N}^*$,

et $t_k = t_0 + kh$ $(0 \leq k \leq N)$ les temps discrets où on approche $y(t)$.

On note y_k une approximation de $y(t_k)$.



Nous allons étudier différents schémas numériques où y_{k+1} est déterminé à partir de y_k. Ces schémas sont appelés <u>méthodes à un pas</u> (ou à pas séparés).

On construit souvent ces schémas à partir d'approximations des dérivées temporelles par différences finies, ou en approchant des intégrales par des formules de quadrature numérique.

<u>Exemples de schémas à un pas :</u>

<u>a) Euler explicite</u> :

On suppose $f \in C^1$, on a donc $y \in C^2$.

En utilisant la formule de Taylor, on a lorsque $h \to 0$:

$$y'(t_k) = \frac{y(t_k+h) - y(t_k)}{h} + O(h)$$

$$= \frac{y(t_{k+1}) - y(t_k)}{h} + O(h)$$

$$= f(t_k, y(t_k))$$ pour y solution de l'équation différentielle (E).

Le schéma d'Euler explicite s'obtient en négligeant le terme en $O(h)$:
$$\frac{y_{k+1} - y_k}{h} = f(t_k, y_k).$$

Le schéma d'Euler explicite s'écrit donc :

$$y_{k+1} = y_k + h\, f(t_k, y_k) \quad , \quad k = 0, 1, \dots, N-1$$

On peut montrer le résultat de convergence suivant :

<u>Théorème:</u> Si f vérifie (H) et $y \in C^2([t_0, T])$, alors pour le schéma d'Euler explicite :

$$\max_{0 \leq k \leq N} \| y_k - y(t_k) \| \leq C h$$

où $C = \dfrac{e^{L(T-t_0)} - 1}{2L} \times \sup_{[t_0,T]} \| y'' \|$

L'erreur étant $O(h)$, on dit que le schéma d'Euler explicite est <u>d'ordre 1</u>. Ce résultat de convergence est un cas particulier de l'étude générale de la convergence des méthodes à un pas que nous allons décrire plus loin.

b) point milieu / Runge :

On suppose $f \in C^2$, on a donc $y \in C^3$. Avec la formule de Taylor :

$$\frac{y(t_{k+1}) - y(t_k)}{h} = \frac{y\left((t_k + \frac{h}{2}) + \frac{h}{2}\right) - y\left((t_k + \frac{h}{2}) - \frac{h}{2}\right)}{2 \times \frac{h}{2}} = y'\left(t_k + \frac{h}{2}\right) + O(h^2)$$

$$= f\left(t_k + \frac{h}{2}, y(t_k + \frac{h}{2})\right) + O(h^2) \qquad \text{pour } y \text{ solution de (E)}$$

$$= f\left(t_k + \frac{h}{2}, y(t_k) + \frac{h}{2} y'(t_k) + O(h^2)\right) + O(h^2)$$

$$= f\left(t_k + \frac{h}{2}, y(t_k) + \frac{h}{2} f(t_k, y(t_k))\right) + O(h^2)$$

(puisque $y'(t_k) = f(t_k, y(t_k))$ et f est lipschitzienne en y).

Le schéma du point milieu s'obtient en négligeant le reste en $O(h^2)$ (erreur plus faible que pour Euler explicite).

On note $\phi(t_k, y_k, h) = f\left(t_k + \frac{h}{2}, y_k + \frac{h}{2} f(t_k, y_k)\right)$.

Le schéma du point milieu s'écrit : $\dfrac{y_{k+1} - y_k}{h} = \phi(t_k, y_k, h)$

c'est à dire :

$$y_{k+1} = y_k + h\,\phi(t_k, y_k, h), \quad k = 0, 1, \dots, N-1$$

Nous verrons que ce schéma est d'ordre 2, ie $\max\limits_{0 \leq k \leq N} \| y_k - y(t_k) \| = O(h^2)$.

c) Euler implicite :

On suppose $f \in C^1$, on a donc $y \in C^2$, et par la formule de Taylor :

$$\begin{aligned} y'(t_{k+1}) &= \frac{y(t_{k+1}) - y(t_{k+1} - h)}{h} + O(h) \\ &= \frac{y(t_{k+1}) - y(t_k)}{h} + O(h) \\ &= f(t_{k+1}, y(t_{k+1})) \quad \text{pour } y \text{ solution de } (E). \end{aligned}$$

En négligeant le terme en $O(h)$, on obtient le schéma d'Euler implicite : $\dfrac{y_{k+1} - y_k}{h} = f(t_{k+1}, y_{k+1})$.

Le calcul de y_{k+1} à partir de y_k nécessite donc la résolution d'une équation algébrique ; on parle donc de schéma implicite. Chaque itération est donc plus complexe / coûteuse que pour un schéma explicite. Mais les schémas implicites sont souvent plus indiqués en présence d'échelles de temps différentes (problèmes "raides") et pour éviter des instabilités numériques conduisant à une divergence de la solution avec le temps.

Le schéma d'Euler implicite peut se reformuler comme suit :

$$\begin{cases} y_{k+1} = y_k + h k_1 \\ k_1 = f(t_{k+1}, y_{k+1}) \end{cases}$$

qui équivaut à :

$$\begin{cases} y_{k+1} = y_k + h\, k_1 & i) \\ k_1 = f(t_k + h, y_k + h\, k_1) & ii) \end{cases}$$

On cherche alors k_1 solution de l'équation ii), puis y_{k+1} est déterminé par i).

Puisque f vérifie (H), l'application $k_1 \mapsto f(t_k+h, y_k + h k_1)$ est contractante lorsque $h < \frac{1}{L}$:

$$\forall k_1, \tilde{k}_1 \in \mathbb{R}^n, \quad \| f(t_k+h, y_k + h k_1) - f(t_k+h, y_k + h \tilde{k}_1) \| \leq \underbrace{Lh}_{<1} \| k_1 - \tilde{k}_1 \|$$

Alors ii) admet une solution unique k_1 par le théorème du point fixe contractant (qui fournit aussi une méthode itérative de calcul de k_1). On note $k_1 = \phi(t_k, y_k, h)$ cette solution.

On a donc : $\quad y_{k+1} = y_k + h\,\phi(t_k, y_k, h), \quad 0 \leq k \leq N-1,$

où ϕ est une fonction implicite.

<u>Etude générale des méthodes à un pas :</u>

L'intervalle $[t_0, T]$ est fixé.
On considère un schéma (S) de la forme :

$$(S) \quad y_{k+1} = y_k + h\,\phi(t_k, y_k, h), \quad 0 \leq k \leq N-1$$

avec $\phi : [t_0, T] \times \mathbb{R}^n \times [0, h_m] \to \mathbb{R}^n$.

<u>Définition</u> Le schéma (S) est convergent si

$$\forall y_0 \in \mathbb{R}^n, \quad \max_{0 \leq k \leq N} \| y_k - y(t_k) \| \xrightarrow[h \to 0]{} 0,$$

où y désigne la solution de (P)

$E_k := y(t_k) - y_k$ est l'erreur entre les solutions exacte et numérique pour $y(t_0) = y_0$, ou <u>"erreur globale"</u>. Le schéma est <u>convergent d'ordre p</u> lorsque $\max_{0 \leq k \leq N} \| E_k \| = O(h^p)$

La convergence d'un schéma résulte de deux propriétés : consistance et stabilité, que nous allons définir.

<u>Définition</u> : Le schéma (S) est consistant avec (E) si toute solution y de (E) sur $[t_0, T]$ vérifie :

$$\sup_{t \in [t_0, T-h]} \left\| \frac{y(t+h) - y(t)}{h} - \phi(t, y(t), h) \right\| \xrightarrow[h \to 0]{} 0$$

On appelle $\varepsilon(t) = \dfrac{y(t+h) - y(t)}{h} - \phi(t, y(t), h)$ <u>l'erreur de consistance</u> ou de troncature.

Si $\sup_{[t_0, T-h]} \| \varepsilon(t) \| = O(h^p)$, le schéma est consistant d'ordre p.

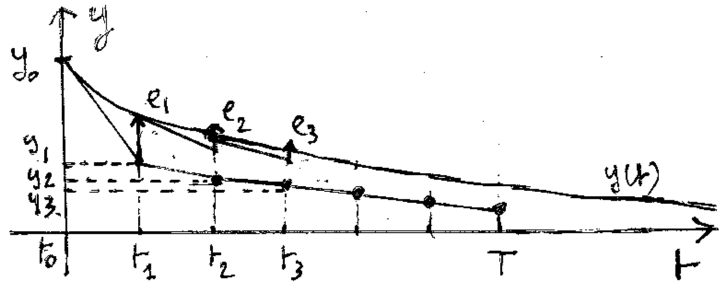
<u>Remarque</u> "consistant" est un anglicisme et signifie ici "cohérent". (S) est consistant si toute solution de (E) est presque solution de (S), à un petit reste près (l'erreur de consistance).

Il est intéressant de relier l'erreur de consistance à <u>l'erreur locale</u>

$$e_{k+1} = h\, \varepsilon(t_k) = y(t_k + h) - [\, y(t_k) + h\, \phi(t_k, y(t_k), h) \,].$$

e_{k+1} est la différence en $t = t_k + h$ entre les solutions exacte et numérique pour la condition initiale $y(t_k)$ en $t = t_k$.

Représentation de l'erreur locale :



En pratique, on évalue l'ordre de consistance d'un schéma à l'aide d'un développement de Taylor des solutions de l'équation différentielle (E). (comme dans les exemples a), b), c) précédents)

Formule de Taylor avec reste intégral :

Soient $y \in C^{n+1}([a,b], \mathbb{R}^m)$ et t, $t+h \in [a,b]$. Alors :

$$y(t+h) = \sum_{k=0}^{n} \frac{y^{(k)}(t)}{k!} h^k + \frac{h^{n+1}}{n!} \int_0^1 (1-s)^n y^{(n+1)}(t+sh)\, ds$$

* Pour le schéma d'Euler explicite avec $f \in C^1$:

On a $\phi = f$ donc : $\varepsilon(t) = \frac{y(t+h) - y(t)}{h} - f(t, y(t)) = \frac{y(t+h) - y(t)}{h} - y'(t)$

Puisque $y \in C^2$, on obtient par un développement de Taylor :

$y(t+h) = y(t) + h y'(t) + h^2 \int_0^1 (1-s)\, y''(t+sh)\, ds$ donc

$\varepsilon(t) = h \int_0^1 (1-s)\, y''(t+sh)\, ds$.

Donc $\sup_{t \in [t_0, T-h]} \|\varepsilon(t)\| \leq \frac{h}{2} \sup_{[t_0,T]} \|y''\|$.

Le schéma d'Euler explicite est donc consistant d'ordre 1.

* Par le même type de calcul, lorsque $f \in C^2$, le schéma du point milieu est consistant d'ordre 2

(reprendre le calcul de l'exemple b) en prenant en compte le reste intégral)

* Vérifier en exercice que le schéma d'Euler implicite est consistant d'ordre 1 lorsque $f \in C^1$

(reprendre le calcul de l'exemple c)).

On a de plus le résultat suivant :

Proposition 1 On suppose l'application ϕ continue. Le schéma (S) est consistant si et seulement si $\phi(t, y, 0) = f(t, y)$ $\forall t \in [t_0, T]$, $\forall y \in \mathbb{R}^n$.

preuve de la proposition 1 :

- supposons (S) consistant, et montrons que $\forall \tilde{t} \in [t_0, T]$, $\forall \tilde{y} \in \mathbb{R}^n$ on a $\phi(\tilde{t}, \tilde{y}, 0) = f(\tilde{t}, \tilde{y})$.

Soit y la solution de (E) pour la condition initiale $y(\tilde{t}) = \tilde{y}$.

$$\begin{aligned}
\phi(\tilde{t}, \tilde{y}, 0) - f(\tilde{t}, \tilde{y}) &= \phi(\tilde{t}, y(\tilde{t}), 0) - y'(\tilde{t}) \\
&= \lim_{h \to 0} \left(\phi(\tilde{t}, y(\tilde{t}), h) - \frac{y(\tilde{t}+h) - y(\tilde{t})}{h} \right) \\
&= 0 \quad \text{car (S) est consistant.}
\end{aligned}$$

- supposons que $\phi_{|h=0} = f$ et montrons que (S) est consistant

Soit y une solution de (E). Alors :

$$\begin{aligned}
\varepsilon(t) = \frac{y(t+h) - y(t)}{h} - \phi(t, y(t), h) &= \frac{1}{h} \int_t^{t+h} y'(\tau)\, d\tau - \phi(t, y(t), h) \\
&= \frac{1}{h} \int_t^{t+h} f(\tau, y(\tau))\, d\tau - \phi(t, y(t), h) \\
&= \frac{1}{h} \int_t^{t+h} \left(f(\tau, y(\tau)) - f(t, y(t)) \right) d\tau \\
&\quad + \phi(t, y(t), 0) - \phi(t, y(t), h)
\end{aligned}$$

Donc par inégalité triangulaire :

$$\sup_{t \in [t_0, T-h]} \| \varepsilon(t) \| \leq \sup_{\substack{\tau, t \in [t_0, T] \\ |\tau - t| \leq h}} \| f(\tau, y(\tau)) - f(t, y(t)) \| + \sup_{t \in [t_0, T]} \| \phi(t, y(t), 0) - \phi(t, y(t), h) \|$$

$\longrightarrow 0$ quand $h \to 0$ par continuité uniforme des applications $\begin{array}{l}[t_0, T] \to \mathbb{R}^n \\ t \mapsto f(t, y(t))\end{array}$, $\begin{array}{l}[t_0, T] \times [0, h_m] \to \mathbb{R}^n \\ (t, h) \mapsto \phi(t, y(t), h)\end{array}$. $\square$

On définit maintenant une seconde propriété importante pour la convergence d'un schéma à un pas : la <u>stabilité par rapport aux erreurs</u>.

Pour cela, on considère une perturbation de (S) (on perturbe ϕ et la condition initiale) :

$$\begin{cases} z_{k+1} = z_k + h\left[\phi(t_k, z_k, h) + \varepsilon_k\right], & 0 \leq k \leq N-1 \\ z_0 \simeq y_0 \end{cases}$$

<u>Définition</u> : Le schéma (S) est stable par rapport aux erreurs sur l'intervalle $[t_0, T]$ s'il existe $M, h_{max}, \varepsilon_{max}, \eta > 0$ / $\forall h < h_{max}$, $\forall$ perturbation (ε_k) vérifiant $\max\limits_{0 \leq k \leq N-1} \|\varepsilon_k\| < \varepsilon_{max}$, $\forall y_0, z_0 \in \mathbb{R}^n$ avec $\|y_0 - z_0\| < \eta$, on a :

$$\max_{0 \leq k \leq N} \|y_k - z_k\| \leq M \left(\max_{0 \leq k \leq N-1} \|\varepsilon_k\| + \|y_0 - z_0\| \right)$$

Lorsqu'on perturbe le schéma numérique, cette condition donne une borne sur la perturbation de la solution numérique proportionnelle à la taille de la perturbation, avec une constante M indépendante de h.

On a le résultat de convergence suivant, souvent appelé "Théorème fondamental" :

<u>Théorème</u>

Si le schéma (S) est stable et consistant, alors il est convergent. De plus, si le schéma est consistant à l'ordre p alors il est convergent d'ordre p.

Preuve du théorème fondamental :

On note y la solution du problème de Cauchy (P), y_k son approximation numérique, et on considère l'erreur de consistance $\varepsilon_k = \varepsilon(t_k) = \dfrac{y(t_{k+1}) - y(t_k)}{h} - \phi(t_k, y(t_k), h)$.

On a donc :
$$\begin{cases} y(t_{k+1}) = y(t_k) + h\left[\phi(t_k, y(t_k), h) + \varepsilon_k\right] \\ y_{k+1} = y_k + h\,\phi(t_k, y_k, h) \qquad (S) \\ y(t_0) = y_0 \end{cases}$$

La stabilité du schéma (S) par rapport aux erreurs donne :

$$\max_{0 \leq k \leq N} \| y_k - y(t_k) \| \leq M \max_{0 \leq k \leq N-1} \| \varepsilon_k \| \longrightarrow 0 \quad \text{quand } h \to 0$$

puisque le schéma (S) est consistant.

Par ailleurs, $\max_k \| \varepsilon_k \| = O(h^p)$ donne $\max_k \| y_k - y(t_k) \| = O(h^p)$.

$\square$

Pour montrer la stabilité de (S) par rapport aux erreurs, on a le critère suivant :

Proposition 2

On suppose ϕ lipschitzienne par rapport à y, uniformément en $t \in [t_0, T]$ et $h \in [0, h_{max}[$:

$\exists K \geq 0 \,/\, \forall y, z \in \mathbb{R}^n, \forall t \in [t_0, T], \forall h \in [0, h_{max}[$:

$\| \phi(t, y, h) - \phi(t, z, h) \| \leq K \| y - z \|$.

Alors le schéma (S) est stable par rapport aux erreurs.

Application : Si f vérifie (H), vérifier que les schémas d'Euler explicite, implicite et du point milieu sont stables par rapport aux erreurs.

Le résultat suivant est un corollaire du théorème fondamental et des propositions 1 et 2 :

Théorème :

Si ϕ est continue, lipschitzienne par rapport à y uniformément en t, h et vérifie $\phi_{|h=0} = f$, alors le schéma (S) est convergent.

En particulier, les schémas d'Euler explicite et implicite sont convergents d'ordre 1 si $f \in C^1$, et le schéma du point milieu est convergent d'ordre 2 si $f \in C^2$.

On donne maintenant la preuve du critère de stabilité :

preuve de la proposition 2 :

On considère le schéma perturbé et le schéma (S) :

$$z_{k+1} = z_k + h\left[\phi(t_k, z_k, h) + \varepsilon_k\right], \quad z_0 \text{ donné}$$

$$y_{k+1} = y_k + h\,\phi(t_k, y_k, h), \quad y_0 \text{ donné}$$

En utilisant l'inégalité triangulaire et le caractère lipschitzien de ϕ par rapport à y :

$$\|y_{k+1} - z_{k+1}\| \leq \|y_k - z_k\|(1 + hM) + h\|\varepsilon_k\|$$

$$\leq \|y_k - z_k\| e^{hM} + h\|\varepsilon_k\| \qquad \text{(convexité de la fonction exponentielle)}$$

En multipliant par $e^{-hM(k+1)}$:

$$e^{-hM(k+1)}\|y_{k+1} - z_{k+1}\| - e^{-hMk}\|y_k - z_k\| \leq e^{-hM(k+1)} h\|\varepsilon_k\|$$

En sommant sur $k = 0, \dots, p-1$:

$$e^{-hMp}\|y_p - z_p\| - \|y_0 - z_0\| \leq \sum_{k=0}^{p-1} e^{-hM(k+1)} h\|\varepsilon_k\|$$

Puisque $h = \frac{T-t_0}{N}$, on obtient pour tout $p \in \{0, 1, \dots, N\}$:

$$\|y_p - z_p\| \leq e^{(T-t_0)M}\left(\|y_0 - z_0\| + \max_{0 \leq k \leq N-1}\|\varepsilon_k\|(T - t_0)\right) \quad \square$$