

Equations différentielles Ordinaires

Une grande part des problèmes scientifiques voient leur modélisation déboucher sur une (ou plusieurs) équation(s) différentielle(s) en fonction de la variable temps (évolution du système), et/ou de variables spatiales. Lorsqu'une seule variable est en jeu (temps t , une coordonnée de l'espace), on parle d'*Equation Différentielle Ordinaire (ODE)*. Dans le cas très fréquent en physique où plusieurs variables entrent en jeu (champs scalaires ou vectoriels fonction de l'espace et du temps¹), on parle d'*Equation aux Dérivées Partielles (PDE)*. C'est le cas par exemple des phénomènes ondulatoires. Ce cours sera restreint aux ODE, et nous aborderons les EDP en SPE.

Quand l'équation est simple (linéaire d'ordre n , à variables séparables...) ou peut s'y ramener (équations de *Bernoulli*, de *Ricatti*, de *Lagrange*...), la résolution analytique est possible. Mais nombre de modélisations conduisent à des équations non linéaires pour lesquelles aucune solution analytique n'est connue. Ce cours présente des méthodes pour en déterminer une solution approchée.

On commence par montrer que toutes les EDO ou systèmes d'EDO de tous ordres peuvent se mettre sous la forme d'une EDO **vectorielle du premier ordre**², ce qui permet de leur appliquer les mêmes méthodes. Ensuite on rappelle le cadre mathématique permettant d'obtenir l'**existence et l'unicité** d'une solution. Lorsque l'on tente d'approcher cette solution par une méthode numérique, différents problèmes se posent :

- L'approximation choisie doit être assez précise pour que la solution approchée **converge** vers la solution exacte : diminuer le pas d'intégration h fait tendre l'**erreur** vers 0, d'autant plus vite que l'**ordre** p est grand ($O(h^p)$);
- Pour cela, la méthode numérique doit nécessairement être **stable**, c'est-à-dire que la solution approchée ne doit pas diverger sous l'effet des erreurs d'arrondis inhérentes à l'utilisation des flottants à la place des réels.
- La méthode doit être efficace en terme de **temps de calcul** : ne pas requérir un pas exagérément petit, c'est-à-dire un nombre d'itérations exagérément grand, pour une précision donnée.

L'exemple élémentaire qui servira de support pour le cours est l'EDO suivante, notée (E_0) dont la solution exacte est connue :

$$(E_0) : \boxed{\dot{y}(t) + y(t) = 0 \quad \text{avec} \quad y(0) = 1} \quad \text{qui conduit à} \quad \boxed{y_e(t) = e^{-t}}.$$

1. Mécaniques des milieux continus dont les fluides, thermodynamique, électromagnétisme, mécanique quantique, théorie quantique des champs...

2. Par commodité, on notera les vecteurs en gras et horizontalement : $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$.

I. Résultats fondamentaux sur les EDO

I.1. Une EDO est toujours du premier ordre... !

DÉFINITION : Equation différentielle ordinaire du premier ordre^a

Soit un ouvert U de $\mathbb{R} \times \mathbb{R}^m$ et une application $\mathbf{F} : U \mapsto \mathbb{R}^m$ **continue**.
On considère l'équation différentielle

$$\mathbf{Y}'(t) = \mathbf{F}(t, \mathbf{Y}(t)) \quad (E)$$

On appelle solution de (E) sur l'intervalle $I \subset \mathbb{R}$ une application **dérivable** $\mathbf{Y} : I \rightarrow \mathbb{R}^m$ telle que $\forall t \in I$ on ait $(t, \mathbf{Y}(t)) \in U$, et vérifiant (E) pour tout $t \in I$.

a. Dans la suite, la variable t représente aussi bien le temps que n'importe quelle autre variable (coordonnée notamment). Sauf cas particulier, le symbole $'$ est utilisé pour désigner de façon générale la dérivation par rapport à cette variable t .

Exemples : (on omet d'écrire la variable t pour alléger...)

- Problèmes scalaires du premier ordre ($m = 1$) :

$$\frac{du}{dt} = -\frac{1}{\tau}(u(t) - u_\infty) \quad \rightarrow \quad Y(t) = u(t) \quad \text{et} \quad Y'(t) = -\frac{1}{\tau}(Y(t) - u_\infty)$$

$$\frac{dp}{dz}(z) = -\frac{Mg}{RT(z)}p(z) \quad \rightarrow \quad Y(z) = p(z) \quad \text{et} \quad Y'(z) = -\frac{Mg}{RT(z)}Y(z)$$

- Problème vectoriel 2D du premier ordre ($m = 2$) :

Mouvement avec frottement quadratique.

$$\begin{cases} \dot{v}_x = -\alpha \sqrt{v_x^2 + v_y^2} \cdot v_x \\ \dot{v}_y = -\alpha \sqrt{v_x^2 + v_y^2} \cdot v_y \end{cases} \quad \rightarrow \quad \begin{cases} \mathbf{Y} = (Y_1, Y_2) = (v_x, v_y) \\ \mathbf{Y}' = -\alpha \|\mathbf{Y}\| \cdot (Y_1, Y_2) \end{cases}$$

- Problème scalaire du second ordre ($m = 2$) :

Chute libre dans le champ de pesanteur avec vitesse initiale verticale :

$$\ddot{z} = -g \quad \rightarrow \quad \begin{cases} \mathbf{Y} = (Y_1, Y_2) = (z, \dot{z}) \\ \mathbf{Y}' = (Y_2, -g) \end{cases}$$

Pendule :

$$\ddot{\theta} = -\omega_0^2 \sin \theta \quad \rightarrow \quad \begin{cases} \mathbf{Y} &= (Y_1, Y_2) = (\theta, \dot{\theta}) \\ \mathbf{Y}' &= (Y_2, -\omega_0^2 \sin Y_1). \end{cases}$$

- Problème vectoriel 2D du second ordre ($m = 4$) :
Système masse-ressort en mouvement plan sans frottement.

$$\begin{cases} \ddot{x} = -k(\ell - \ell_0) \frac{x}{\ell} = -k\left(1 - \frac{\ell_0}{\ell}\right)x \\ \ddot{y} = -k(\ell - \ell_0) \frac{y}{\ell} = -k\left(1 - \frac{\ell_0}{\ell}\right)y \end{cases} \quad \text{avec} \quad \ell = \sqrt{x^2 + y^2}$$

On pose $\mathbf{Y} = (x, y, \dot{x}, \dot{y})$ et

$$\mathbf{Y}' = \left(Y_3, Y_4, -k \left(1 - \frac{\ell_0}{\sqrt{Y_1^2 + Y_2^2}} \right) Y_1, -k \left(1 - \frac{\ell_0}{\sqrt{Y_1^2 + Y_2^2}} \right) Y_2 \right).$$

I.2. Problème de Cauchy : existence des solutions

DÉFINITION : Un **problème de Cauchy** est une EDO du premier ordre munie d'une **condition initiale** $\mathbf{Y}(t_0) = \mathbf{Y}_0$ imposée.

THÉORÈME : Existence des solutions

Un problème de Cauchy admet au moins une solution.

REMARQUE : Pour une EDO scalaire d'ordre n en la fonction $y(t)$, le vecteur $\mathbf{Y} = (y(t), y'(t), \dots, y^{(n-1)}(t))$ contient les dérivées d'ordre 0 à $n-1$. Donc le problème de Cauchy requiert la connaissance au temps t_0 de $y(t_0)$ et de ses $n-1$ premières dérivées : $\mathbf{Y}(t_0) = (y(t_0), y'(t_0), \dots, y^{(n-1)}(t_0))$.

I.3. Unicité de la solution

Il ne suffit pas toujours de donner une condition initiale pour disposer d'une solution unique, ce qui peut être déroutant. Heureusement il existe une condition suffisante pour obtenir l'unicité, mais requiert d'être un peu plus contraignant sur la fonction \mathbf{F} .

DÉFINITION : \mathbf{F} est localement **lipschitzienne** en \mathbf{Y}

$\forall(t_0, \mathbf{Y}_0 = \mathbf{Y}(t_0)) \in U$, il existe $k = k(t_0) \geq 0$ et un compact $C_0 \subset U$ contenant (t_0, \mathbf{Y}_0) tels que :

$$\forall(t, \mathbf{Y}_1) \in C_0, \forall(t, \mathbf{Y}_2) \in C_0, \quad \|\mathbf{F}(t, \mathbf{Y}_1) - \mathbf{F}(t, \mathbf{Y}_2)\| \leq k \|\mathbf{Y}_1 - \mathbf{Y}_2\|$$

THÉORÈME : Cauchy-Lipschitz

Si \mathbf{F} est localement lipschitzienne en \mathbf{Y} , alors le problème de Cauchy de condition initiale (t_0, \mathbf{Y}_0) admet une unique solution.

REMARQUES :

- Une condition suffisante pour que \mathbf{F} soit localement lipschitzienne est qu'elle soit de classe \mathcal{C}^1 .
- Un corollaire de ce théorème est que deux solutions démarrant avec deux conditions initiales distinctes n'auront jamais aucun point commun. Autrement dit les *courbes intégrales*³ ne se croisent jamais.

II. Exemples de méthodes numériques à 1 pas

L'objectif est d'obtenir une solution approchée au problème de Cauchy pour une discrétisation temporelle

$$t_0 < t_1 < \dots < t_{N-1} < t_N = t_0 + T$$

de l'intervalle $[t_0, T]$ sur lequel on souhaite résoudre l'équation. La fonction $\mathbf{F}(t, \mathbf{Y})$ traduit l'évolution de \mathbf{Y} , c'est-à-dire sa pente. Bien que les propos qui suivent sont applicables de façon générale au cas d'un pas de temps $h_n = t_{n+1} - t_n$ variable⁴, on se restreint pour simplifier au cas d'un pas de temps constant :

$$h = t_{n+1} - t_n$$

Dans cette partie, on présente ici les méthodes les plus simples, qui appartiennent à la catégorie plus générale des méthodes à 1 pas (cf section III). Elles reposent sur une approximation de la dérivée par différence finie avant d'ordre 1. Ainsi, en notant $\mathbf{Y}_k = \mathbf{Y}(t_k)$, à l'instant t_n on a :

$$\frac{d\mathbf{Y}}{dt}(t_n) \simeq \frac{\mathbf{Y}_{n+1} - \mathbf{Y}_n}{h}.$$

Les méthodes à 1 pas diffèrent ensuite pas la façon d'évaluer approximativement le second membre du problème de Cauchy : $\mathbf{F}(t, \mathbf{Y}(t))$.

3. Les courbes intégrales sont simplement les ensembles de points $M(t)$ dont les coordonnées sont celles du vecteur $\mathbf{Y}(t)$ solution de l'équation (E). Dans le cas où $\mathbf{Y}(t)$ est scalaire il s'agit de courbes du plan.

4. C'est parfois indispensable, notamment pour aborder une discontinuité.

II.1. Méthode d'Euler (explicite)

a. Principe

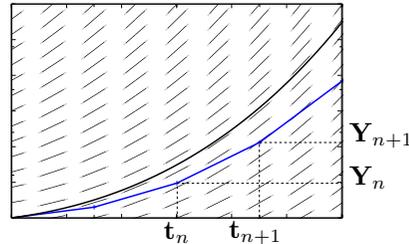
Le second membre est simplement évalué au temps t_n . Ainsi, la pente permettant de passer de \mathbf{Y}_n à \mathbf{Y}_{n+1} est $\mathbf{F}(t_n, \mathbf{Y}_n)$:

$$\frac{\mathbf{Y}_{n+1} - \mathbf{Y}_n}{h} = \mathbf{F}(t_n, \mathbf{Y}_n)$$

Ainsi, on obtient la récurrence

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + h\mathbf{F}(t_n, \mathbf{Y}_n)$$

comme l'illustre la figure ci-contre où le champ des pentes \mathbf{F} est tracé.



b. Application

L'application du schéma d'Euler (explicite) au problème (E_0) conduit à la récurrence

$$Y_{n+1} = (1 - h)Y_n$$

qui est une progression géométrique de raison $1 - h$. Le choix du pas d'intégration va donner lieu à divers comportements qui sont réunis dans les courbes de la Fig. 1 pour une intégration sur l'intervalle de temps $[0, 10]$. On constate que :

- si $h \geq 2$, la méthode est instable et/ou ne convergera pas vers la valeur attendue ;
- si $1 < h < 2$, la méthode converge lentement en oscillant autour de la valeur ;
- si $0 < h < 1$, la méthode converge et l'erreur est d'autant plus faible que le pas de temps est petit.

Pour évaluer quantitativement les performances de la méthode, on doit quantifier l'erreur par rapport à la solution exacte.

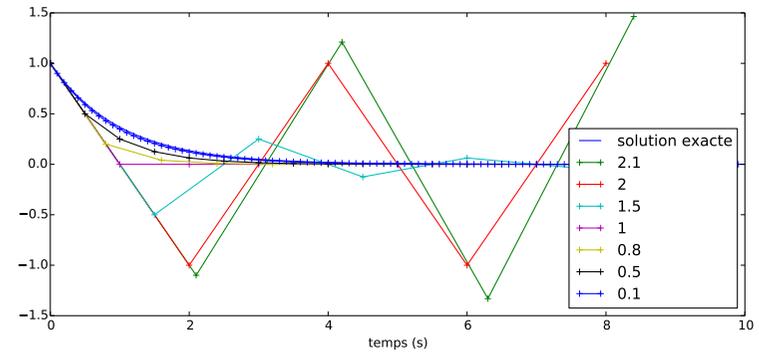


FIGURE 1 – Résolutions numériques de $\dot{y}(t) + y(t) = 0$ avec $y(0) = 1$ pour le schéma d'Euler explicite.

II.2. Erreur de consistance et erreur globale

DÉFINITION : Erreur de consistance (erreur locale)

Soit \mathbf{Z} une **solution exacte** du problème de Cauchy. L'erreur de consistance est celle produite au cours d'une **itération** de la méthode numérique à un pas, à **partir de la valeur exacte** $\mathbf{Y}_n = \mathbf{Z}(t_n)$:

$$\mathbf{e}_n = \mathbf{Z}(t_{n+1}) - \mathbf{Y}_{n+1}$$

DÉFINITION : Erreur globale ^a

$$\max_{0 \leq n \leq N} |\mathbf{Y}_n - \mathbf{Z}(t_n)|$$

DÉFINITION : Une méthode est **convergente** si pour toute solution exacte \mathbf{Z} , l'erreur globale tend vers 0 quand $h \rightarrow 0$ et $\mathbf{Y}_0 - \mathbf{Z}(t_0) \rightarrow 0$.

^a. On notera $||$ pour la valeur absolue d'un scalaire comme pour la norme d'un vecteur.

On illustre ces notions sur les figures ci-dessous.

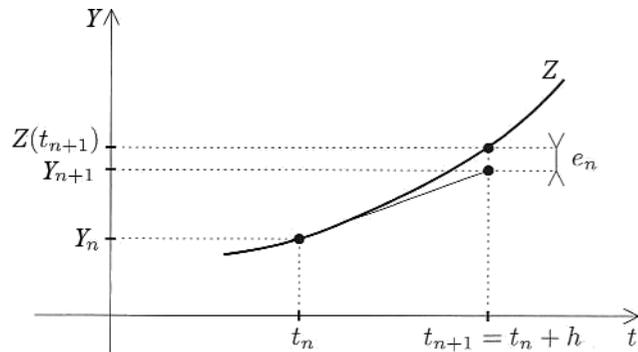


FIGURE 2 – Erreur de consistance dans le cas de la méthode d'Euler explicite.

Comme le montre le schéma ci-dessous, l'erreur de consistance n'a a priori que peu de rapport avec l'erreur globale. Toutefois, sous des hypothèses convenables de régularité, l'ordre de grandeur de la première permettra de connaître l'ordre de grandeur de la seconde (cf section III.2.a).

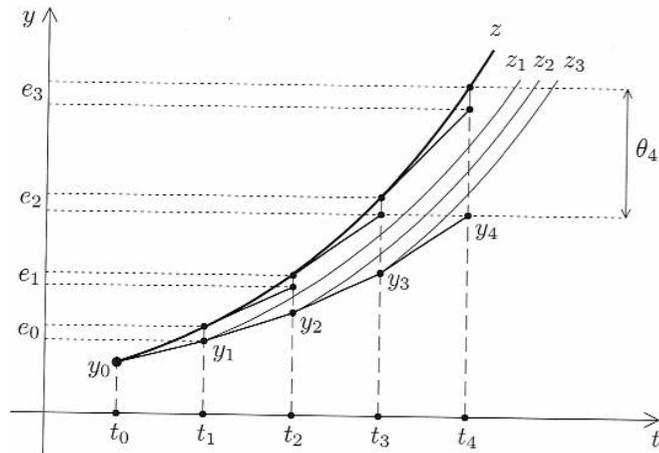


FIGURE 3 – Formation de l'erreur globale dans le cas de la méthode d'Euler explicite (les fonctions z , z_1 , z_2 et z_3 représentent les solutions exactes passant par les points (t_0, y_0) et (t_j, y_j) , $j = 1, 2, 3$).

Application à la méthode d'Euler (explicite)

Le tableau suivant montre l'évolution de l'erreur globale et du temps de calcul.

h	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
N pas de temps	10^2	10^3	10^4	10^5	10^6	10^7
Erreur globale	4×10^{-1}	5×10^{-2}	5×10^{-3}	5×10^{-4}	5×10^{-5}	5×10^{-6}
Temps (s)	5×10^{-3}	2×10^{-2}	8×10^{-2}	6×10^{-1}	5	50

- **L'erreur globale évolue linéairement en fonction du pas de temps h .** On parle alors de **méthode d'ordre 1**. Il faut descendre à un pas de temps très faible avant d'obtenir un niveau d'erreur satisfaisant.
- **Le temps de calcul évolue linéairement avec h : la complexité est en $O(N)$** car il y a une seule boucle.

II.3. Méthode du point milieu

La méthode du point milieu consiste à évaluer la pente entre les deux points par une valeur proche de la pente de la corde, à savoir **la pente évaluée au milieu de l'intervalle**, ce qui donne la récurrence suivante :

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + h \mathbf{F} \left(t_n + \frac{h}{2}, \mathbf{Y}_n + \frac{h}{2} \mathbf{F}(t_n, \mathbf{Y}_n) \right)$$

On voit ainsi que le principe repose sur la **construction d'un point intermédiaire** entre les points n et $n + 1$, qu'on notera symboliquement $n + \frac{1}{2}$:

- l'instant intermédiaire est choisi au milieu :

$$t_{n+\frac{1}{2}} = t_n + \frac{h}{2} = \frac{1}{2} (t_n + t_{n+1})$$

- la valeur de la fonction à cet instant est approchée comme dans la méthode d'Euler :

$$\mathbf{Y}_{n+\frac{1}{2}} = \mathbf{Y}_n + \frac{h}{2} \mathbf{F}(t_n, \mathbf{Y}_n)$$

L'avantage de cette évaluation plus précise de la dérivée \mathbf{Y}' est d'**obtenir une méthode d'ordre plus élevé, sans avoir à connaître les dérivées successives de \mathbf{F}** . On peut montrer que la méthode du point milieu est **d'ordre 2**, contrairement à celle d'Euler, qui est d'ordre 1.

II.4. Choix du pas

Le choix du pas d'intégration h répond à plusieurs contraintes contradictoires :

- pour une méthode d'ordre donné p , **plus le pas h est petit**, plus l'erreur de consistance (locale) sera faible.
- pour un domaine d'intégration fixé, **plus le pas h est petit**, plus le nombre d'étapes de calcul est élevé, donc **plus les erreurs d'arrondi cumulées sont importantes**.
- pour un domaine d'intégration fixé, **plus le pas h est petit**, plus le nombre d'étapes de calcul est élevé, donc **plus le temps de calcul est long**.

Notamment, il est remarquable qu'en général, **l'erreur globale commence par décroître lorsque h décroît, puis se met à augmenter en dessous d'un certain pas h_0 optimal. Il n'est donc pas toujours judicieux de réduire le pas.**

Critère de Shannon

En pratique, le choix du pas détermine une frontière pour la dynamique représentable, c'est-à-dire **les fréquences (temporelles ou spatiales) que l'on peut résoudre**. Cette limite est donnée, dans le cas d'un pas constant h , par le *critère de Shannon* :

$$f_{\max} = \frac{1}{2} f_e = \frac{1}{2h}$$

Typiquement, on fait en sorte d'avoir toujours au moins 10 points par période pour pouvoir évaluer correctement les variations à une certaine fréquence.

En pratique, le choix du pas de temps fait souvent appel à une démarche empirique et demande de l'expérience.

Contrôle du pas : méthodes à pas variable h_n

Il est souvent difficile de prévoir le pas h_{\max} à ne pas dépasser pour que l'erreur globale ne dépasse pas une tolérance fixée ε . L'utilisation d'algorithmes à pas variables⁵ présente de ce point-de-vue des avantages majeurs.

- En utilisant une méthode d'estimation⁶ à *la volée* de l'erreur de consistance e_n , on peut adapter le pas h_n en cours d'intégration.
- L'approche d'une discontinuité ou d'une singularité de l'équation différentielle ne peut se faire en général qu'avec une réduction importante du pas.

5. On généralise les méthodes présentées précédemment en utilisant un pas variable h_n .

6. On ne peut connaître exactement l'erreur de consistance puisqu'on ne connaît en général pas la solution exacte de façon analytique.

Dans cette circonstance, il convient d'arrêter l'algorithme avant de traverser la discontinuité, faute de quoi les erreurs deviennent imprévisibles. Le calcul du pas h_n à appliquer sert alors de test d'arrêt.

III. Approche générale

III.1. Résolution numérique d'un problème de Cauchy

On s'appuie sur la forme intégrale de l'équation différentielle, qui sur un sous-intervalle $[t_{n-r}, t_{n+1}]$ conduit à :

$$\int_{t_{n-r}}^{t_{n+1}} \frac{d\mathbf{Y}}{dt} dt = \int_{t_{n-r}}^{t_{n+1}} \mathbf{F}(t, \mathbf{Y}) dt \iff \mathbf{Y}_{n+1} = \mathbf{Y}_{n-r} + \int_{t_{n-r}}^{t_{n+1}} \mathbf{F}(t, \mathbf{Y}) dt$$

Ainsi, les méthodes de résolution numérique du problème de Cauchy sont basées sur des techniques d'estimation de l'intégrale de la fonction \mathbf{F} (cf chapitre précédent).

On distingue :

Méthodes explicites : on calcule \mathbf{Y}_{n+1} à l'aide des valeurs aux instants précédents $\mathbf{Y}_n, \mathbf{Y}_{n-1}, \dots$, qui sont connues au moment du calcul de \mathbf{Y}_{n+1} .

- Méthodes à un pas ($r = 0$) :

$$\boxed{\mathbf{Y}_{n+1} = \mathbf{Y}_n + h \Phi(t_n, \mathbf{Y}_n, h)} \text{ avec } \mathbf{Y}_0 \text{ imposé.}$$

La fonction Φ permet de calculer la pente de la droite permettant de passer de \mathbf{Y}_n à \mathbf{Y}_{n+1} , dont l'évaluation sera plus ou moins pertinente.

- Méthodes à pas multiples :

$$\boxed{\mathbf{Y}_{n+1} = \mathbf{Y}_{n-r} + \Psi(t_n, \mathbf{Y}_n, \dots, t_{n-r}, \mathbf{Y}_{n-r}, h)}$$

Méthodes implicites : le calcul de \mathbf{Y}_{n+1} requiert la connaissance de \mathbf{Y}_{n+1}, \dots

$$\boxed{\mathbf{Y}_{n+1} = \mathbf{Y}_{n-r} + \Psi(t_{n+1}, \mathbf{Y}_{n+1}, t_n, \mathbf{Y}_n, \dots, t_{n-r}, \mathbf{Y}_{n-r}, h)}$$

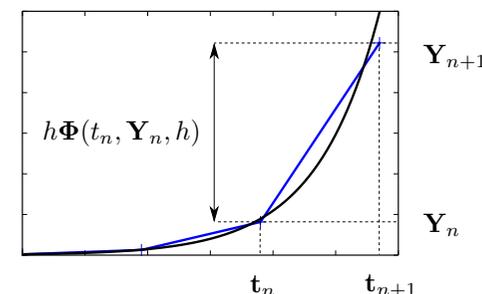
On doit donc résoudre une équation $f(\mathbf{Y}_{n+1}) = 0$ pour trouver \mathbf{Y}_{n+1} .

Les méthodes à pas multiples sont plus sophistiquées et donc plus performantes⁷.

7. En guise d'exemple, nous avons vu qu'il était plus précis d'exprimer une dérivée par une différence finie centrée d'ordre 2 : $D_n \approx \frac{\mathbf{Y}_{n+1} - \mathbf{Y}_{n-1}}{2h}$

Les méthodes implicites peuvent être utiles dans certains contextes (notamment pour résoudre des problèmes de stabilité...). Mais elles nécessitent des temps de calcul supérieurs donc sont à éviter si possible.

Dans le cadre de ce cours, on se restreint au cas des **méthodes explicites à 1 pas**.



Application à la méthode d'Euler (explicite)

La première façon d'approximer l'intégrale est de réaliser une méthode des rectangles à gauche :

$$\int_{t_n}^{t_{n+1}} \mathbf{F}(t, \mathbf{Y}) dt = h \mathbf{F}(t_n, \mathbf{Y}_n).$$

Ainsi, la pente permettant de passer de \mathbf{Y}_n à \mathbf{Y}_{n+1} via la récurrence

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + h \Phi(t_n, \mathbf{Y}_n)$$

est $\Phi(t_n, \mathbf{Y}_n) = \mathbf{F}(t_n, \mathbf{Y}_n)$, ce qui conduit à la méthode d'Euler (explicite).

Application à la méthode du point milieu

La méthode du point milieu consiste à évaluer l'intégrale presque comme dans la méthode des trapèzes :

$$\mathbf{Y}_{n+1} - \mathbf{Y}_n = \int_{t_n}^{t_{n+1}} \mathbf{F}(t, \mathbf{Y}) dt = h \mathbf{F}\left(t_n + \frac{h}{2}, \mathbf{Y}_n + \frac{h}{2} \mathbf{F}(t_n, \mathbf{Y}_n)\right), \text{ ou}$$

$$\Phi(t_n, \mathbf{Y}_n, h) = \mathbf{F}\left(t_n + \frac{h}{2}, \mathbf{Y}_n + \frac{h}{2} \mathbf{F}(t_n, \mathbf{Y}_n)\right)$$

III.2. Etude générale des méthodes à 1 pas

a. Méthodes consistantes, stables et convergentes

DÉFINITION : Une méthode est dite **consistante** si pour toute solution exacte \mathbf{Z} la somme des erreurs de consistance relatives à \mathbf{Z} tend vers 0 quand h (ou h_{\max}) tend vers 0 :

$$\sum_{0 \leq n \leq N} |\mathbf{e}_n| \xrightarrow{h \rightarrow 0} 0$$

Dans la pratique, le calcul récurrent des points \mathbf{Y}_n est entaché d'erreurs d'arrondi ε_n . Pour que les calculs soient significatifs, il est indispensable que la propagation de ces erreurs reste contrôlable.

DÉFINITION : Une méthode est dite **stable** s'il existe une constante $S \geq 0$, appelée *constante de stabilité*, telle que pour toutes les suites (\mathbf{Y}_n) et les suites *perturbées* $(\tilde{\mathbf{Y}}_n)$ définies par

$$\begin{aligned} \mathbf{Y}_{n+1} &= \mathbf{Y}_n + h \Phi(t_n, \mathbf{Y}_n, h) & 0 \leq n < N \\ \tilde{\mathbf{Y}}_{n+1} &= \tilde{\mathbf{Y}}_n + h \Phi(t_n, \tilde{\mathbf{Y}}_n, h) + \varepsilon_n & 0 \leq n < N \end{aligned}$$

on ait

$$\max_{0 \leq n \leq N} |\tilde{\mathbf{Y}}_n - \mathbf{Y}_n| \leq S (|\tilde{\mathbf{Y}}_0 - \mathbf{Y}_0| + \sum_{0 \leq n \leq N} |\varepsilon_n|).$$

THÉORÈME : Si une méthode est stable et consistante, alors elle est convergente.

Démonstration - On applique la définition de la stabilité à la suite $(\tilde{\mathbf{Y}}_n) = (\mathbf{Z}(t_n))$: $\max_{0 \leq n \leq N} |\mathbf{Z}(t_n) - \mathbf{Y}_n| \leq S (|\mathbf{Z}(t_0) - \mathbf{Y}_0| + \sum_{0 \leq n \leq N} |\varepsilon_n|)$.

La propriété de consistance permet de conclure.

b. Propriétés importantes

THÉORÈME : Condition nécessaire et suffisante de consistance

La méthode à un pas définie par $\mathbf{Y}_{n+1} = \mathbf{Y}_n + h \Phi(t_n, \mathbf{Y}_n, h)$ est consistante si et seulement si

$$\forall (t, \mathbf{Y}) \in [t_0, t_0 + T] \times \mathbb{R}^m, \quad \Phi(t, \mathbf{Y}, 0) = \mathbf{F}(t, \mathbf{Y}).$$

Démonstration - Elle repose sur la continuité de $\Phi \dots$

THÉORÈME : Condition suffisante de stabilité

Si la fonction Φ est Λ -lipschitzienne en \mathbf{Y} sur $[t_0, t_0 + T]$, alors

- la méthode est stable ;
- on peut prendre pour constante de stabilité $S = e^{\Lambda T}$.

c. Notion d'ordre

DÉFINITION : Une méthode à un pas est dite **d'ordre** $\geq p$ si pour toute solution exacte \mathbf{Z} de l'équation différentielle

$$\mathbf{Y}'(t) = \mathbf{F}(t, \mathbf{Y}(t)) \quad \text{où } \mathbf{F} \text{ est de classe } \mathcal{C}^p,$$

il existe une constante $C \geq 0$ telle que l'erreur de consistance relative à \mathbf{Z} vérifie

$$|\mathbf{e}_n| \leq C h_n^{p+1}, \quad \forall n, \quad 0 \leq n < N.$$

La méthode est dite d'ordre p (exactement) si elle est d'ordre $\geq p$ mais pas d'ordre $\geq p + 1$.

THÉORÈME : Une méthode est consistante si et seulement si elle est d'ordre $p \geq 1$.

PROPRIÉTÉ : Majoration de l'erreur globale

Si la méthode est stable de constante de stabilité S , alors

$$\max_{0 \leq n \leq N} |\mathbf{Z}(t_n) - \mathbf{Y}_n| \leq S (|\mathbf{Z}(t_0) - \mathbf{Y}_0| + CT h^p)$$

où C est une constante liée à la norme des dérivées d'ordre p de \mathbf{F} et Φ .

L'erreur initiale $|\mathbf{Z}(t_0) - \mathbf{Y}_0|$ est généralement négligeable. **L'erreur globale donnée par une méthode stable d'ordre p est donc de l'ordre de grandeur de h^p** , avec une constante de proportionnalité SCT . Il est donc intéressant d'augmenter l'ordre de la méthode, à condition que les fonctions \mathbf{F} et Φ soient suffisamment régulières.

d. Application à la méthode d'Euler (explicite)

Consistance :

→ L'erreur de consistance est $\mathbf{e}_n = \mathbf{Z}(t_{n+1}) - \mathbf{Z}(t_n) - h\mathbf{F}(t_n, \mathbf{Z}(t_n))$.

→ Or le développement de Taylor-Lagrange donne : $\mathbf{Z}(t_n + h) = \mathbf{Z}(t_{n+1}) = \mathbf{Z}(t_n) + h\dot{\mathbf{Z}}(t_n) + \frac{h^2}{2}\ddot{\mathbf{Z}}(\xi)$ avec $\xi \in [t_n, t_n + h]$.

→ D'autre part $\dot{\mathbf{Z}} = \Phi(t_n, \mathbf{Z}(t_n), h) = \mathbf{F}(t_n, \mathbf{Z}(t_n))$.

→ On en déduit que $\mathbf{e}_n = \frac{h^2}{2}\ddot{\mathbf{Z}}(\xi)$ avec $\xi \in [t_n, t_n + h]$, ou

$$\mathbf{e}_n = \frac{h^2}{2}\ddot{\mathbf{Z}}(t_n) + o(h_n^2).$$

Ainsi la méthode est d'ordre 1. Elle est donc consistante.

Stabilité : Si Φ est lipschitzienne, alors le schéma est stable.

Convergence : La méthode étant consistante, et stable si Φ est lipschitzienne en \mathbf{Y} , alors elle est **convergente si Φ est lipschitzienne en \mathbf{Y}** .

e. Application à la méthode du point milieu

Consistance : on peut montrer que la méthode du point milieu est d'ordre 2, contrairement à celle d'Euler, qui est d'ordre 1. Elle est donc **consistante**.

Stabilité et Convergence : La méthode étant consistante, **si Φ est lipschitzienne en \mathbf{Y}** , alors elle est stable et elle converge.

III.3. Méthode de Runge-Kutta

a. Principe

Les méthodes de Runge-Kutta peuvent être vues comme une généralisation de la méthode du point milieu, consistant en une évaluation plus précise de la pente

$$\mathbf{Y}'(t_n) \approx \frac{\mathbf{Y}_{n+1} - \mathbf{Y}_n}{h} = \Phi(t_n, \mathbf{Y}_n, h)$$

en utilisant **une série de q points intermédiaires** $(t_{n,i}, \mathbf{Y}_{n,i})$. Ainsi, on définit les instants

$$t_{n,i} = t_n + c_i h_n, \quad 1 \leq i \leq q, \quad c_i \in [0, 1].$$

En chacun de ces points on définit la pente

$$\mathbf{p}_{n,i} = \mathbf{F}(t_{n,i}, \mathbf{Y}_{n,i}),$$

où les $\mathbf{Y}_{n,i}$ sont construits par approximations successives sous la forme

$$\mathbf{Y}_{n,i} = \mathbf{Y}_n + h \sum_{1 \leq j < i} a_{i,j} \mathbf{p}_{n,j}.$$

Finalement on construit \mathbf{Y}_{n+1} selon ⁸

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + h \sum_{1 \leq j < q} b_j \mathbf{p}_{n,j}.$$

Exemple : Pour $q = 1$, le seul choix possible est $c_1 = 0$, $a_{1,1} = 0$, $b_1 = 1$, ce qui donne

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + h \mathbf{p}_{n,1} \quad \text{avec} \quad \mathbf{p}_{n,1} = \mathbf{F}(t_n, \mathbf{Y}_n).$$

On reconnaît la méthode d'Euler explicite.

L'avantage principal des méthodes de Runge Kutta est d'obtenir des ordres plus élevés, sans avoir à connaître les dérivées successives de \mathbf{F} .

b. Runge-Kutta d'ordre 2 ($q = 2$)

Sous sa forme condensée, la méthode RK pour $q = 2$ peut s'écrire ainsi :

$$\Phi(t_n, \mathbf{Y}_n, h) = \left(1 - \frac{1}{2\alpha}\right) \mathbf{F}(t_n, \mathbf{Y}_n) + \frac{1}{2\alpha} \mathbf{F}(t_n + \alpha h, \mathbf{Y}_n + \alpha h \mathbf{F}(t_n, \mathbf{Y}_n)).$$

PROPRIÉTÉ : La méthode RK à 1 point intermédiaire ($q = 2$) est d'ordre 2.

On peut citer les cas particuliers importants suivants.

- Pour $\alpha = \frac{1}{2}$, on retrouve la méthode du point milieu

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + h \mathbf{F}(t_n + \frac{h}{2}, \mathbf{Y}_n + \frac{h}{2} \mathbf{F}(t_n, \mathbf{Y}_n)),$$

qui est basée sur la méthode d'intégration du point milieu : $\int_0^1 g(t) dt \approx g(\frac{1}{2})$.

- Pour $\alpha = 1$, on obtient la *méthode de Heun*

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + h \left(\frac{1}{2} \mathbf{F}(t_n, \mathbf{Y}_n) + \frac{1}{2} \mathbf{F}(t_{n+1}, \mathbf{Y}_n + h \mathbf{F}(t_n, \mathbf{Y}_n)) \right),$$

basée sur la méthode d'intégration des trapèzes : $\int_0^1 g(t) dt \approx \frac{1}{2} (g(0) + g(1))$.

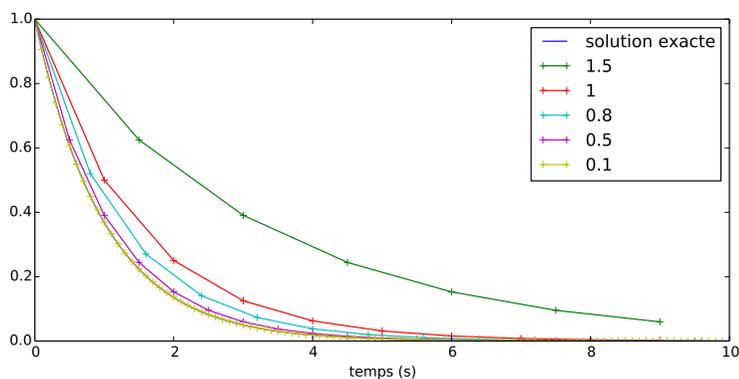


FIGURE 4 – Résolution numérique de $y'(t) + y(t) = 0$ avec $y(0) = 1$ par le schéma de Heun.

Application

La solution approchée de l'EDO (E_0) pour $t \in [0, 10]$ est présentée sur la Fig. 4, pour différents pas de temps.

Le tableau suivant donne l'évolution de l'erreur globale ainsi que le temps de calcul pour différentes valeurs du pas de temps.

h	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
N pas de temps	10^2	10^3	10^4	10^5	10^6	10^7
Erreur	2×10^{-2}	2×10^{-4}	2×10^{-6}	2×10^{-8}	2×10^{-10}	2×10^{-12}
Temps (s)	$1,5 \times 10^{-3}$	$1,4 \times 10^{-2}$	$1,3 \times 10^{-1}$	1,2	12	120

- On vérifie bien que **la méthode est d'ordre 2** puisque l'erreur évolue quadratiquement en fonction du pas de temps : en divisant le pas de temps par 10, l'erreur diminue d'un facteur 100.
- Le temps de calcul est toujours linéaire car la complexité n'a pas augmenté (complexité en $O(N)$). Il est environ double de la méthode d'Euler explicite puisqu'il y a deux évaluations de $\mathbf{F}(t, \mathbf{Y})$.

8. Les coefficients $a_{i,j}$ et b_j correspondent toujours à une méthode classique d'approximation d'une intégrale.

c. Runge-Kutta d'ordre 4 (RK4)

La méthode «classique» RK à 3 points intermédiaires ($q = 4$) est résumée par l'algorithme suivant. On peut montrer qu'elle est **d'ordre 4**.

$$\left\{ \begin{array}{l} \mathbf{p}_{n,1} = \mathbf{F}(t_n, \mathbf{Y}_n) \quad (\text{donc } t_{n,1} = t_n) \\ t_{n,2} = t_n + \frac{1}{2} h \\ \mathbf{Y}_{n,2} = \mathbf{Y}_n + \frac{1}{2} h \mathbf{p}_{n,1} \\ \mathbf{p}_{n,2} = \mathbf{F}(t_{n,2}, \mathbf{Y}_{n,2}) \\ \mathbf{Y}_{n,3} = \mathbf{Y}_n + \frac{1}{2} h \mathbf{p}_{n,2} \\ \mathbf{p}_{n,3} = \mathbf{F}(t_{n,2}, \mathbf{Y}_{n,2}) \quad (t_{n,3} = t_{n,2}) \\ t_{n+1} = t_n + h \quad (t_{n,4} = t_{n+1}) \\ \mathbf{Y}_{n,4} = \mathbf{Y}_n + \frac{1}{2} h \mathbf{p}_{n,3} \\ \mathbf{p}_{n,4} = \mathbf{F}(t_{n+1}, \mathbf{Y}_{n,4}) \\ \mathbf{Y}_{n+1} = \mathbf{Y}_n + h \left(\frac{1}{6} \mathbf{p}_{n,1} + \frac{2}{6} \mathbf{p}_{n,2} + \frac{2}{6} \mathbf{p}_{n,3} + \frac{1}{6} \mathbf{p}_{n,4} \right) \end{array} \right.$$

Application

La solution approchée de (E_0) pour différents pas de temps est donnée sur la Fig. 5.

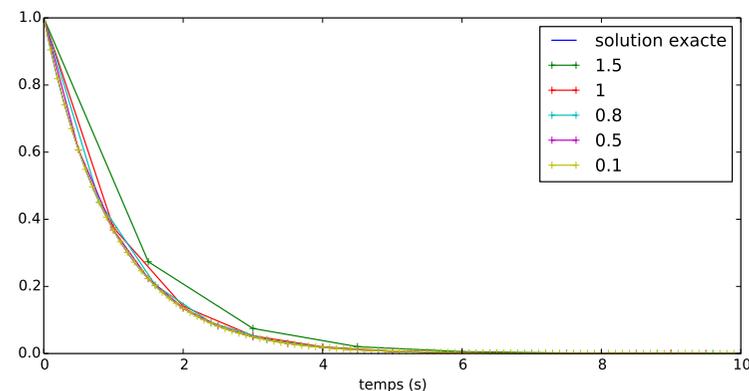


FIGURE 5 – Résolution numérique de $y'(t) + y(t) = 0$ avec $y(0) = 1$ par le schéma classique Runge-Kutta d'ordre 4.

Le tableau suivant donne l'erreur globale ainsi que le temps de calcul pour diffé-

rentes valeurs du pas de temps.

h	10^{-1}	10^{-2}	10^{-3}
N pas de temps	10^2	10^4	10^5
Erreur	$9,0 \times 10^{-6}$	$8,4 \times 10^{-10}$	$9,2 \times 10^{-14}$
Temps (s)	$1,7 \times 10^{-3}$	$1,6 \times 10^{-2}$	$2,5 \times 10^{-1}$

- On constate que l'erreur évolue à l'ordre 4 en fonction du pas de temps : en divisant le pas de temps par 10, l'erreur diminue d'un facteur 10 000.
- Le temps de calcul est toujours linéaire car la complexité n'a pas augmentée (toujours en $O(N)$).

La méthode RK4 classique est la «méthode reine» des méthodes à un pas : elle est très stable ^a et converge rapidement grâce à un ordre 4.

^a. La constante de stabilité est, toutes choses égales par ailleurs, petite.

Des méthodes d'ordre supérieur existent mais leur complexité les rend souvent moins praticables.