# Équations différentielles - Cours no 6
# Approximation numérique

## 1 Introduction

De très nombreux problèmes scientifiques sont mis en équation à l'aide d'un système d'équations différentielles $\dot{x}(t) = f(t, x(t))$ (voir par exemple le mouvement à deux corps dans le cours d'introduction). Au moment de l'application numérique, il est nécessaire (connaissant une valeur initiale) de pouvoir calculer une ou plusieurs valeurs $x(T_1)$, $x(T_2)$... L'objet de ce cours est l'étude des méthodes qui permettent d'opérer ce calcul (à la main ou plutôt à l'aide d'un ordinateur de nos jours).
L'exemple le plus élémentaire est la méthode d'Euler : soit $f \in C^1(\mathbb{R} \times \mathbb{R}^d)$, globalement Lipschitz en sa deuxième variable. Pour calculer la solution au temps $T > 0$ du Problème de Cauchy

$$\left\{ \begin{array}{lcl} \dot{x}(t) & = & f(t, x(t)), \\ x(0) & = & x_0, \end{array} \right. \tag{1}$$

on subdivise l'intervalle $[0, T]$ en $0 = t_0 < t_1 < \cdots < t_N = T$ et, pour $t$ voisin de $t_n$, on utilise les approximations

$$\dot{x}(t) \simeq \frac{x(t_{n+1}) - x(t_n)}{t_{n+1} - t_n}, \quad f(t, x(t)) \simeq f(t_n, x(t_n)).$$

En reportant dans l'équation différentielle, on aboutit à la méthode d'Euler :

$$x_{n+1} = x_n + h_n f(t_n, x_n), \text{ où } x_n = x(t_n), h_n = t_{n+1} - t_n.$$

En commençant avec $x_0$ à $n = 0$, on calcule de proche en proche (récursivement) $x_1$, $x_2$, etc. jusqu'à $x_N$ qui est sensé fournir une valeur approchée de $x(T)$.

On verra que, lorsque $h := \sup_{0 \leq n \leq N-1} h_n \to 0$, $x_N \to x(T)$. Cela montre l'existence d'une méthode d'approximation numérique. On va aussi répondre aux questions suivantes : existe-t-il des méthodes arbitrairement précise ? Existe-t-il des méthodes respectant les propriétés de l'équation (symétrie, conservation...) ?

Le plan du cours est le suivant. Dans le chapitre 2, on continue l'exemple ci-dessus en décrivant d'autres méthodes d'approximation numérique. Dans le chapitre 3, on analyse une classe de schémas (schémas dits à un pas). Dans le chapitre 4, on étudie quelques méthodes symplectiques.

**Remarque importante :** il aurait été avisé d'assortir ce cours d'un cours (de base) de programmation (fonctionnement de l'ordinateur, représentation des nombres, operations élémentaires, conditionnement, temps, coût de calcul, etc.) Ce n'est pas le cas présentement. On s'est donc abstenu de commentaires sur l'effectivité réelle des méthodes rencontrées dans le cours (voir la remarque à la fin du chapitre 3 toutefois), c'est-à-dire dans quelle mesure elles satisfont des contraintes de calcul sur machine. Le dernier chapitre sur les méthodes symplectiques pose le problème de l'élaboration de méthode numérique sous une contrainte plus simple à appréhender : celle des propriétés géométriques de l'équation différentielle.

# 2 Exemples

Soit $x$ la solution du problème de Cauchy (1) et $0 = t_0 < t_1 < \cdots < t_N = T$ une subdivision de l'intervalle $[0, T]$. Par intégration, on a

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

En utilisant des méthodes de calcul approché de l'intégrale $\int_{t_n}^{t_{n+1}} g(t)dt$, $g(t) := f(t, x(t))$, on obtient les méthodes suivantes.

**Méthode d'Euler :** rectangle à gauche. On approche

$$\int_{t_n}^{t_{n+1}} g(t)dt \simeq (t_{n+1} - t_n)g(t_n).$$

On obtient la méthode d'Euler déjà décrite

$$x_{n+1} = x_n + h_n f(t_n, x_n), \text{ où } h_n = t_{n+1} - t_n.$$

**Méthode d'Euler implicite :** rectangle à droite. On approche

$$\int_{t_n}^{t_{n+1}} g(t)dt \simeq (t_{n+1} - t_n)g(t_{n+1}).$$

On obtient la méthode d'Euler implicite

$$x_{n+1} = x_n + h_n f(t_{n+1}, x_{n+1}), \text{ où } h_n = t_{n+1} - t_n.$$

On déduit $x_{n+1}$ de $x_n$en inversant la fonction $y \mapsto y - h_n f(t_{n+1}, y)$.

**Méthode de Runge :** point milieu. On approche

$$\int_{t_n}^{t_{n+1}} g(t)dt \simeq (t_{n+1} - t_n)g\left(\frac{t_{n+1} + t_n}{2}\right).$$

On obtient

$$x(t_{n+1}) \simeq x(t_n) + h_n f\left(t_n + \frac{h_n}{2}, x\left(\frac{t_{n+1}+t_n}{2}\right)\right), \text{ où } h_n = t_{n+1} - t_n.$$

En remplaçant (par la méthode d'Euler)

$$x\left(\frac{t_{n+1}+t_n}{2}\right) \simeq x(t_n) + \frac{h_n}{2} f(t_n, x(t_n)),$$

on obtient la méthode de Runge

$$x_{n+1} = x_n + h_n f\left(t_n + \frac{h_n}{2}, x_n + \frac{h_n}{2} f(t_n, x_n)\right).$$

**Méthode de Heun :** on utilise la formule de quadrature suivante :

$$\int_{t_n}^{t_{n+1}} g(t)dt \simeq (t_{n+1} - t_n)\left[\frac{1}{4}g(t_n) + \frac{3}{4}g\left(\frac{1}{3}t_n + \frac{2}{3}t_{n+1}\right)\right].$$

On obtient (en notant $h_n = t_{n+1} - t_n$)

$$x(t_{n+1}) \simeq x(t_n) + h_n\left[\frac{1}{4}f(t_n, x(t_n)) + \frac{3}{4}f\left(t_n + \frac{2}{3}h_n, x\left(t_n + \frac{2}{3}h_n\right)\right)\right].$$

En remplaçant (par la méthode de Runge)

$$x\left(t_n + \frac{2}{3}h_n\right) \simeq x(t_n) + \frac{2}{3}h_n f\left(t_n + \frac{h_n}{3}, x(t_n) + \frac{h_n}{3}f(t_n, x(t_n))\right),$$

on obtient la méthode de Heun

$$x_{n+1} = x_n + h_n\left[\frac{1}{4}f(t_n, x_n) + \frac{3}{4}f\left(t_n + \frac{2}{3}h_n, x_n + \frac{2}{3}h_n f\left(t_n + \frac{h_n}{3}, x_n + \frac{h_n}{3}f(t_n, x_n)\right)\right)\right].$$

On peut généraliser ce procédé, on obtient des méthodes de plus en plus emboîtées.

**Définition 1 (Kutta)** *Soit $s \in \mathbb{N}$. L'approximation numérique de la solution du problème de Cauchy (1) est une méthode de Runge-Kutta à $s$ étage s'il existe $(b_i)_{1,s}$, $(c_i)_{2,s}$ et $(a_{ij})_{1\leq j<i\leq s}$ tels que $x_{n+1}$ est calculé à partir de $x_n$ par les formules*

$$\begin{aligned}
k_1 &= f(t_n, x_n),\\
k_2 &= f(t_n + c_2 h_n, x_n + h_n a_{21} k_1)),\\
k_3 &= f(t_n + c_3 h_n, x_n + h_n(a_{31}k_1 + a_{32}k_2),\\
&\cdots\\
k_s &= f(t_n + c_s h_n, x_n + h_n(a_{s1}k_1 + \ldots + a_{s\,s-1}k_{s-1})),\\
x_{n+1} &= x_n + h_n(b_1 k_1 + \ldots + b_s k_s),
\end{aligned}$$

*où $h_n := t_{n+1} - t_n$.*

Une telle méthode est désignée par un tableau (où $c_1 = 0$) $\begin{array}{c|c} c_i & a_{ij} \\ \hline & b_j \end{array}$.

**Exemples :** Voici les tableaux de quelque méthodes.

Euler : $\begin{array}{c|c} 0 & \\ \hline & 1 \end{array}$

Runge : $\begin{array}{c|cc} 0 & & \\ 1/2 & 1/2 & \\ \hline & 1 & 0 \end{array}$

Heun : $\begin{array}{c|ccc} 0 & & & \\ 1/3 & 1/3 & & \\ 2/3 & 0 & 2/3 & \\ \hline & 1/4 & 0 & 3/4 \end{array}$.

# 3 Approximation numérique par des méthodes à un pas

Soit $\Phi \in C(\mathbb{R} \times \mathbb{R}^d \times [0,1])$, soit $0 = t_0 < t_1 < \cdots < t_N = T$ une subdivision de l'intervalle $[0,T]$ et $h$ son pas :

$$h := \max_{0 \le n < N} h_n, \quad h_n = t_{n+1} - t_n.$$

Soit $x_h(T)$ le $N$-ième élément de la suite définie par

$$x_{n+1} = x_n + h_n \Phi(t_n, x_n, h_n), \quad 0 \le n \le N-1. \tag{2}$$

On note $x$ la solution du Problème de Cauchy (1). On va montrer le résultat suivant : il existe $\Phi$ telle que $|x_h(T) - x(T)| \to 0$ lorsque $h \to 0$, et même : $\forall p \in \mathbb{N}$, il existe $\Phi$ telle que $|x_h(T) - x(T)| = o(h^p)$ (précision arbitraire).

## 3.1 Stabilité et consistance

**Définition 2 (Consistance)** *Soit $x$ une solution de $\dot{x}(t) = f(t, x(t))$. L'erreur de consistance locale (au temps $t_n$) relative à $x$ est*

$$\varepsilon_n(x) = x(t_{n+1}) - (x(t_n) + h_n \Phi(t_n, x(t_n), h_n)).$$

*L'erreur de consistance relative à $x$ est*

$$\varepsilon_h(x) = \sum_{n=0}^{N-1} |\varepsilon_n(x)|.$$

*On dit que la méthode est* consistante *si, pour tout $x$ solution de $\dot{x}(t) = f(t, x(t))$, $\varepsilon_h(x) \to 0$ lorsque $h \to 0$.*

**Définition 3 (Stabilité)** *On dit que la méthode est* stable *s'il existe $M > 0$ indépendant de $h$ tel que pour tout $(x_n)$ et $(y_n)$ satisfaisant*

$$\left\{ \begin{array}{rcl} x_{n+1} & = & x_n + h_n \Phi(t_n, x_n, h_n), \quad 0 \leq n \leq N-1, \\ y_{n+1} & = & y_n + h_n \Phi(t_n, y_n, h_n) + \varepsilon_n, \quad 0 \leq n \leq N-1, \end{array} \right.$$

*on a*

$$\max_{0 \leq n \leq N} |x_n - y_n| \leq M \left( |x_0 - y_0| + \sum_{0 \leq n < N} |\varepsilon_n| \right).$$

**Théorème 1 (Convergence)** *Si la méthode est stable et consistante, alors elle est convergente :*

$$\lim_{h \to 0} |x_h(T) - x(T)| = 0.$$

**Preuve du Théorème 1 :** Posons $y_n := x(t_n)$, de sorte que $|x_h(T) - x(T)| = |x_N - y_N|$ et

$$y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n) + \varepsilon_n(x), \quad 0 \leq n \leq N-1,$$

où $\varepsilon_n(x)$ est l'erreur de consistance locale (relative à $x$). La stabilité de la méthode donne alors

$$\max_{0 \leq n \leq N} |x_n - y_n| \leq M \left( |x_0 - y_0| + \sum_{0 \leq n < N} |\varepsilon_n(x)| \right),$$

c'est-à-dire (comme $x_0 = y_0$)

$$\max_{0 \leq n \leq N} |x_n - y_n| \leq M \varepsilon_h(x).$$

On a donc $\max_{0 \leq n \leq N} |x_n - y_n| \to 0$ lorsque $h \to 0$. ∎

**Proposition 1 (Consistance)** *La méthode est consistante si, et seulement si,*

$$\Phi(t, x, 0) = f(t, x)$$

*pour tout $t \in [0, T]$, $x \in \mathbb{R}^d$.*

**Application :** une méthode de Runge-Kutta est consistante dès que $\sum_i b_i = 1$. C'est le cas des méthodes décrites dans le paragraphe "Exemples" (Euler explicite, Runge, Heun, RK4).

**Preuve de la Proposition 1 :** on a

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

donc

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

Soit $\eta > 0$. Soit $\alpha$ un $\eta$-module d'uniforme continuité de $\Phi$ sur $[0,T] \times K \times [0,1]$, où $K = x([0,T])$. Soit $\beta$ un $\alpha$-module d'uniforme continuité de $x$ sur $[0,T]$ :

$$t, s \in [0,T], |t-s| < \beta \Rightarrow |x(t) - x(s)| < \alpha.$$

Quitte à diminuer $\beta$, on peut supposer $\beta < \alpha$. Pour $h < \beta$, on a alors, pour tout $0 \leq n < N$, pour tout $t \in [t_n, t_{n+1}]$,

$$|\Phi(t_n, x(t_n), h_n) - \Phi(t, x(t), 0)| \leq \varepsilon.$$

En particulier, en notant

$$\widetilde{\varepsilon}_n(x) := \int_{t_n}^{t_{n+1}} (f(t, x(t)) - \Phi(t, x(t), 0))dt, \quad \widetilde{\varepsilon}_h(x) = \sum_{n=0}^{N-1} |\widetilde{\varepsilon}_n(x)|,$$

on a $|\widetilde{\varepsilon}_n(x) - \varepsilon_n(x)| \leq (t_{n+1} - t_n)\eta$ et $|\widetilde{\varepsilon}_h(x) - \varepsilon_h(x)| \leq T\eta$. Par conséquent $\varepsilon_h(x)$ tends vers 0 si, et seulement si, $\widetilde{\varepsilon}_h(x) \to 0$ lorsque $h \to 0$.
Si $\Phi(t, x, 0) = f(t, x)$ pour tout $t, x$, alors $\widetilde{\varepsilon}_h(x) = 0$, c'est donc une condition suffisante à la consistance. Montrons que c'est une condition nécessaire : si $\widetilde{\varepsilon}_h(x) \to 0$ lorsque $h \to 0$, alors, en particulier,

$$\sum_{n=0}^{N-1} \widetilde{\varepsilon}_n(x) \to 0,$$

c'est-à-dire

$$\int_0^T (f(t, x(t)) - \Phi(t, x(t), 0))dt = 0.$$

Comme la consistance sur $[0,T]$ implique la consistance sur tout sous-intervalle $I \subset [0,T]$, on a en fait :

$$\int_I (f(t, x(t)) - \Phi(t, x(t), 0))dt = 0$$

pour tout sous-intervalle $I \subset [0,T]$. Cela montre que $f(t, x(t)) = \Phi(t, x(t), 0)$, $0 \leq t \leq T$. Si $t \in [0,T]$, $x_* \in \mathbb{R}^d$, alors il existe $x_0 \in \mathbb{R}^d$ tel que la solution $x$ du Problème de Cauchy (1) satisfait $x(t) = x_*$ (pourquoi ?). On en déduit $f(t, x_*) = \Phi(t, x_*, 0)$.

■

**Proposition 2 (Stabilité)** *Si $\Phi$ est L-globalement lipschitzienne en $x$ uniformément par rapport à $(t,h) \in [0,T] \times [0,1]$, alors la méthode est stable (avec $M = e^{LT}$).*

**Preuve de la proposition 2 :** L'hypothèse est : pour tout $(t,h) \in [0,T] \times [0,1]$, pour tout $x, y \in \mathbb{R}^d$,

$$|\Phi(t, x, h) - \Phi(t, y, h)| \leq L|x - y|.$$

Si

$$\left\{ \begin{array}{rcl} x_{n+1} & = & x_n + h_n \Phi(t_n, x_n, h_n), \quad 0 \leq n \leq N-1, \\ y_{n+1} & = & y_n + h_n \Phi(t_n, y_n, h_n) + \varepsilon_n, \quad 0 \leq n \leq N-1, \end{array} \right.$$

on a alors

$$|x_{n+1} - y_{n+1}| \leq (1 + Lh_n)|x_n - y_n| + |\varepsilon_n|.$$

En utilisant l'inégalité $1 + u \leq e^u$ avec $u = Lh_n$ et un raisonnement par récurrence, on déduit

$$|x_n - y_n| \leq e^{L(h_n + \cdots + h_0)}|x_0 - y_0| + \sum_{k=0}^{n-1} e^{L(h_k + \cdots + h_0)}|\varepsilon_k|,$$

c'est-à-dire

$$|x_n - y_n| \leq e^{Lt_{n+1}}|x_0 - y_0| + \sum_{k=0}^{n-1} e^{Lt_k}|\varepsilon_k|.$$

On en déduit le résultat. ∎

## 3.2 Ordre de la méthode

On a montré dans la preuve du Théorème de convergence (Théorème 1) l'estimation

$$\max_{0 \leq n \leq N} |x_n - x(t_n)| \leq M\varepsilon_h(x),$$

qui montre que, plus petit est $\varepsilon_h(x)$ avec $h$, meilleure est l'approximation numérique.

**Définition 4 (Ordre de la méthode)** *Soit $p \in \mathbb{N}$. On dit que la méthode a l'ordre $p$ si, pour tout $x$ solution de $\dot{x}(t) = f(t, x(t))$, on a $\varepsilon_h(x) = O(h^p)$.*

Ainsi, la méthode est consistante si, et seulement si, elle a l'ordre 0. On généralise maintenant la Proposition 1 qui caractérise les méthodes consistantes.

**Notation :** pour $f \in C^\infty(\mathbb{R} \times \mathbb{R}^d)$, on définit par récurrence

$$\begin{aligned}
d^0 f(t,x) =& f(t,x), \\
d^1 f(t,x) =& \frac{\partial f}{\partial t}(t,x) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(t,x) f_i(t,x), \\
& \cdots \\
d^{k+1} f(t,x) =& \frac{\partial d^k f}{\partial t}(t,x) + \sum_{i=1}^{d} \frac{\partial d^k f}{\partial x_i}(t,x) f_i(t,x),
\end{aligned}$$

de sorte que, si $\dot{x} = f(t,x)$, alors

$$x^{(k+1)}(t) = d^k f(t, x(t)).$$

**Proposition 3 (Ordre de la méthode)** *Soit $p \in \mathbb{N}$. Supposons $f \in C^\infty(\mathbb{R} \times \mathbb{R}^d)$ et $\Phi$ de classe $C^\infty$ par rapport à $h$. Alors la méthode a l'ordre $p$ si, et seulement si,*

$$\frac{\partial^k \Phi}{\partial h^k}(t, x, 0) = \frac{1}{k+1} d^k f(t, x)$$

*pour tout $t \in [0, T]$, $x \in \mathbb{R}^d$ et $k \in \{0, \ldots, p-1\}$.*

**Application :** Euler explicite a l'ordre 1, Runge l'ordre 2, Heun l'ordre 3, RK4 l'ordre 4 (voir TD).

**Preuve de la Proposition 3 :** c'est une généralisation de la preuve de la Proposition 1 ($p = 1$). Supposons $p \geq 2$. On écrit les développement de Taylor avec reste intégral

$$\begin{aligned} x(t_{n+1}) - x(t_n) &= \sum_{k=1}^{p} \frac{h_n^k}{k!} x^{(k)}(t_n) + \int_{t_n}^{t_{n+1}} \frac{(t-t_n)^p}{p!} x^{(p+1)}(t)dt \\ &= \sum_{k=1}^{p} \frac{h_n^k}{k!} d^{k-1} f(t_n, x(t_n)) + \int_{t_n}^{t_{n+1}} \frac{(t-t_n)^p}{p!} d^p f(t, x(t))dt \\ &= \sum_{k=0}^{p-1} \frac{h_n^{k+1}}{k!} \frac{1}{k+1} d^k f(t_n, x(t_n)) + \int_{t_n}^{t_{n+1}} \frac{(t-t_n)^p}{p!} d^p f(t, x(t))dt \end{aligned}$$

et

$$h_n \Phi(t_n, x(t_n), h_n) = \sum_{k=0}^{p-1} \frac{h_n^{k+1}}{k!} \frac{\partial^k \Phi}{\partial h^k}(t_n, x(t_n), 0) + h_n \int_0^{h_n} \frac{\theta^{p-1}}{(p-1)!} \frac{\partial^p \Phi}{\partial h^p}(t_n, x(t_n), \theta)d\theta$$

et on identifie les puissances de $h_n$. Les restes sont $O(h_n^{p+1})$; en les sommant de $n = 0$ à $N-1$, on obtient un $O(h_n^p) = O(h^p)$. ∎

Application : en posant

$$\Phi(x, t, h) := \sum_{k=0}^{p-1} \frac{h^k}{(k+1)!} d^k f(t, x)$$

on obtient la méthode de Taylor, qui a l'ordre $p$.

**Remarque :**

- La méthode de Taylor n'est en général pas utilisée en pratique car le calcul des valeurs $d^k f(t, x)$ est trop coûteux.
- Il manque des chapitres de bases dans ce cours d'introduction, en particulier sur les méthodes à pas variable et les méthodes multi-pas. On renvoit au polycopié de Ernst Hairer : http://www.unige.ch/~hairer/polycop.html.

# 4 Méthodes symplectiques

Soit $N \in \mathbb{N}^*$, soit $U, T \in C^1(\mathbb{R}^N; \mathbb{R})$. On s'intéresse à l'approximation numérique du système hamiltonien suivant (l'inconnue est $x := \begin{pmatrix} p \\ q \end{pmatrix} \in \mathbb{R}^{2N}$) :

$$\left\{ \begin{array}{rcrl} \dot{p} &=& - & \nabla U(q) \\ \dot{q} &=& & \nabla T(p) \end{array} \right. . \tag{3}$$

**Exemple 1 :** le pendule : $N = 1$, $T(p) = p^2/2$ (énergie cinétique), $U(q) = -\cos(q)$ (énergie potentielle).

**Exemple 2 :** cas linéaire : on note $\langle \cdot, \cdot \rangle$ le produit scalaire canonique sur $\mathbb{R}^N$. Soit $B, C \in \mathcal{M}_N(\mathbb{R})$, $T(p) = \frac{1}{2}\langle Cp, p\rangle$, $U(q) = \frac{1}{2}\langle Bq, q\rangle$. On a alors

$$\left\{ \begin{array}{rcl} \dot{p} &=& -\ Bq \\ \dot{q} &=& \quad Cp \end{array} \right. , \tag{4}$$

soit le système linéaire

$$\dot{x} = \mathcal{A}x, \quad x = \begin{pmatrix} p \\ q \end{pmatrix}, \quad \mathcal{A} = \begin{pmatrix} 0 & -B \\ C & 0 \end{pmatrix}.$$

Dorénavant, on se place dans ce dernier cas (même si tout ce qui suit s'applique au cas général de (3)). Le flot pour (4) est donné par $e^{t\mathcal{A}}$.

**Théorème 2** *Le flot de (4) conserve le hamiltonien $H : x = (p, q) \mapsto T(p) + U(q)$ et les "volumes" : pour tout $x \in \mathbb{R}^{2N}$,*

$$H(e^{t\mathcal{A}}x) = H(x)$$

*et, pour tout borélien $D$ de $\mathbb{R}^{2N}$,*

$$|e^{t\mathcal{A}}(D)| = |D|,$$

*où $|D|$ est la mesure de Lebesgue de $D$.*

**Preuve du Théorème 2 :** Le premier point a déjà été vu en cours (exercice). Pour le deuxième point, on a

$$|e^{t\mathcal{A}}(D)| = \det(e^{t\mathcal{A}})|D| = e^{t\mathrm{Tr}(\mathcal{A})}|D| = |D|$$

car $\mathrm{Tr}(\mathcal{A}) = 0$ puisque $\mathcal{A}$ a ses blocs diagonaux nuls. ∎

Le but est de construire une méthode numérique permettant d'approcher les solutions de (4) et respectant les deux propriétés de conservation de l'hamiltonien et des volumes. Si elle est préserve les volumes, on dit qu'elle est symplectique.

## 4.1 La méthode d'Euler

Soit $0 = t_0 < t_1 < \cdots < t_L = T$ une subdivision régulière de l'intervalle $[0, T]$ : $h_n = h = T/L$ pour tout $n$. En appliquant la méthode d'Euler, on définit : $x_{n+1} = x_n + h\mathcal{A}x_n = (I_{2N} + h\mathcal{A})x_n$.

**Conservation du volume :** pour $D$ borélien de $\mathbb{R}^{2N}$, on a

$$|(I_{2N} + h\mathcal{A})D| = \det(I_{2N} + h\mathcal{A})|D|.$$

**Lemme 1** *Si* $\det(I_{2N} + h\mathcal{A}) = 1$ *quel que soit* $h$ *dans un voisinage de* $0$*, alors* $\mathcal{A} = 0$*.*

Conclusion : pas de conservation du volume.

Exemple : $N = 1$, $B = C = 1$ : $\det(I_2 + h\mathcal{A}) = 1 + h^2$.

**Conservation de l'hamiltonien :** Calcul :

$$H(x_{n+1}) = H(x_n) + h(-\langle CBq_n, p_n\rangle + \langle BCp_n, q_n\rangle) + h^2(\langle CBq_n, Bq_n\rangle + \langle BCp_n, Cp_n\rangle).$$

Supposons que le terme en facteur de $H$ est nul, i.e. $(CB)^* = BC$. Pour simplifier, on se place dans le cas

$$B = C, \quad B \text{ symétrique.} \tag{5}$$

On a alors

$$H(x_{n+1}) = H(x_n) + h^2(\langle B^3q_n, q_n\rangle + \langle B^3p_n, p_n\rangle).$$

**Lemme 2** *On a* $\langle B^3q, q\rangle + \langle B^3p, p\rangle = 0$ *quel que soit* $(p, q) \in \mathbb{R}^{2N}$ *si, et seulement si,* $B = 0$ *(i.e.* $\mathcal{A} = 0$*).*

Conclusion : pas de conservation de $H$.

Exemple : $N = 1$, $B = 1$ : $H(x_{n+1}) = (1 + h^2)H(x_n)$.

## 4.2 La méthode du point milieu

On fait l'hypothèse (5). On note $(\cdot|\cdot)$ le produit scalaire canonique sur $\mathbb{R}^{2N}$ : pour $x = (p, q)$ et $y = (w, z) \in \mathbb{R}^{2N}$, $(x|y) := \langle p, w\rangle + \langle q, z\rangle$. On introduit la matrice (symétrique)

$$\mathcal{B} := \begin{pmatrix} B & 0 \\ 0 & B \end{pmatrix}$$

de sorte que $H(x) = (\mathcal{B}x|x)$. Soit la méthode (dite du point milieu)

$$x_{n+1} = x_n + h\mathcal{A}\left(\frac{x_n + x_{n+1}}{2}\right),$$

c'est-à-dire

$$x_{n+1} = \left(I_{2N} - \frac{h}{2}\mathcal{A}\right)^{-1}\left(I_{2N} + \frac{h}{2}\mathcal{A}\right)x_n.$$

**Conservation du volume :** On rappelle que si $M, Q$ sont des matrices de $\mathcal{M}_N(\mathbb{R})$ alors la matrice bloc

$$\begin{pmatrix} I_N & M \\ Q & I_N \end{pmatrix}$$

a pour déterminant $\det(I_N - MQ)$. Une preuve est donnée par le calcul suivant :

$$\begin{pmatrix} I_N & M \\ Q & I_N \end{pmatrix}\begin{pmatrix} I_N & 0 \\ -Q & I_N \end{pmatrix} = \begin{pmatrix} I_N - MQ & M \\ 0 & I_N \end{pmatrix}.$$

On a donc

$$\det\left(I_{2N}-\frac{h}{2}\mathcal{A}\right)=\det\left(I_{2N}+\frac{h}{2}\mathcal{A}\right)=\det(I_N+4^{-1}h^2B^2).$$

On en déduit

$$\det\left(I_{2N}-\frac{h}{2}\mathcal{A}\right)^{-1}\left(I_{2N}+\frac{h}{2}\mathcal{A}\right)=1.$$

Conclusion : la méthode est symplectique.

**Conservation de l'hamiltonien :** On a

$$\begin{aligned}2(H(x_{n+1})-H(x_n)) &= (\mathcal{B}x_{n+1}|x_{n+1})-(\mathcal{B}x_n|x_n)\\ &= (\mathcal{B}(x_{n+1}-x_n)|x_{n+1}+x_n)\\ &= \frac{h}{2}(\mathcal{B}\mathcal{A}(x_{n+1}+x_n)|x_{n+1}+x_n)\\ &= 0 \text{ car } \mathcal{B}\mathcal{A} \text{ est antisymétrique.}\end{aligned}$$

On calcule en effet

$$\mathcal{B}\mathcal{A}=\begin{pmatrix}0 & -B^2\\ B^2 & 0\end{pmatrix}.$$

Conclusion : la méthode du point milieu préserve l'hamiltonien.

Ainsi, la méthode du point milieu a les bonnes propriétés. Dans le cas non-linéaire (pendule par exemple), il peut être difficile de faire une inversion, comme c'est requis ici. La méthode qui suit est explicite (calcul direct), est symplectique, et "préserve presque" l'hamiltonien.

## 4.3 La méthode d'Euler symplectique

On calcule $x_{n+1}$ connaissant $x_n$ par

$$\left\{\begin{array}{rcl}p_{n+1} &=& p_n-hBq_n\\ q_{n+1} &=& q_n+hCp_{n+1}\end{array}\right. .$$

On peut le réécrire

$$\begin{pmatrix}p_{n+1}\\ q_{n+1}\end{pmatrix}=\begin{pmatrix}I_N & 0\\ hC & I_N\end{pmatrix}\begin{pmatrix}I_N & -hB\\ 0 & I_N\end{pmatrix}\begin{pmatrix}p_n\\ q_n\end{pmatrix}. \tag{6}$$

Les deux matrices en jeu ont pour déterminant 1 donc la méthode est symplectique.

**Lemme 3** *Supposons $B=C$, $B$ symétrique. L'hamiltonien modifié*

$$\widetilde{H}(x):=\frac{1}{2}(\langle Bp,p\rangle+\langle Bp,p\rangle-h\langle B^2p,q\rangle)$$

*est préservé.*

Exemple : $N = 1$, $B = 1$ ; la trajectoire ne reste pas sur le cercle $p^2 + q^2 = \text{Cst}$ mais sur l'ellipse (proche) $p^2 + q^2 - pq = \text{Cst}$.

**Preuve du Lemme 3 :** de (6), on tire

$$\begin{pmatrix} I_N & 0 \\ -hB & I_N \end{pmatrix} \begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = \begin{pmatrix} I_N & -hB \\ 0 & I_N \end{pmatrix} \begin{pmatrix} p_n \\ q_n \end{pmatrix}.$$

En multipliant par la matrice

$$\begin{pmatrix} B & 0 \\ 0 & B \end{pmatrix}$$

on en déduit

$$\begin{pmatrix} B & 0 \\ -hB^2 & B \end{pmatrix} \begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = \begin{pmatrix} B & -hB^2 \\ 0 & B \end{pmatrix} \begin{pmatrix} p_n \\ q_n \end{pmatrix}.$$

En multipliant d'abord par le vecteur ligne $(p_{n+1}, q_{n+1})$, on en déduit

$$(p_{n+1}, q_{n+1}) \begin{pmatrix} B & 0 \\ -hB^2 & B \end{pmatrix} \begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = (p_{n+1}, q_{n+1}) \begin{pmatrix} B & -hB^2 \\ 0 & B \end{pmatrix} \begin{pmatrix} p_n \\ q_n \end{pmatrix}.$$

En multipliant ensuite par le vecteur ligne $(p_n, q_n)$, on a

$$(p_n, q_n) \begin{pmatrix} B & 0 \\ -hB^2 & B \end{pmatrix} \begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = (p_n, q_n) \begin{pmatrix} B & -hB^2 \\ 0 & B \end{pmatrix} \begin{pmatrix} p_n \\ q_n \end{pmatrix}.$$

On conclut en observant les termes croisés dans les égalité précédentes sont égaux :

$$\begin{aligned} (p_{n+1}, q_{n+1}) \begin{pmatrix} B & -hB^2 \\ 0 & B \end{pmatrix} \begin{pmatrix} p_n \\ q_n \end{pmatrix} &= \langle Bp_{n+1}, p_n \rangle + \langle Bq_{n+1}, p_n \rangle - h\langle B^2 p_{n+1}, q_n \rangle \\ &= (p_n, q_n) \begin{pmatrix} B & 0 \\ -hB^2 & B \end{pmatrix} \begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix}, \end{aligned}$$

et donc

$$(p_{n+1}, q_{n+1}) \begin{pmatrix} B & 0 \\ -hB^2 & B \end{pmatrix} \begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = (p_n, q_n) \begin{pmatrix} B & -hB^2 \\ 0 & B \end{pmatrix} \begin{pmatrix} p_n \\ q_n \end{pmatrix},$$

i.e. $\widetilde{H}(x_{n+1}) = \widetilde{H}(x_n)$. ∎