Notes de cours - Préparation à l'agrégation

Équations différentielles : approximation numérique

Université de Rennes 1, ENS Rennes

Rozenn Texier-Picard *

Table des matières

Cadre de travail

On considère un problème de Cauchy de la forme

$$y'(t) = f(t, y(t)),\ y(0) = y_0. \tag{1}$$

où f sera supposée de classe C^1. Pour simplifier, on suppose $t_0 = 0$. Remarquons que y est solution de (1) sur l'intervalle $[0,T]$ si et seulement si elle vérifie la formulation intégrale

$$\forall t \in [0,T], y(t) = y_0 + \int_0^t f(s, y(s))ds. \tag{2}$$

Ainsi, pour approcher les solutions de (1) sur un intervalle $[0,T]$, il suffit d'utiliser une méthode d'intégration numérique pour le membre de droite de (2). Aussi, dans un premier temps, nous donnons quelques notions de base sur l'intégration numérique. Ensuite, nous présentons les méthodes classiques pour l'approximation des EDO, et nous introduisons les notions de convergence, consistance, stabilité.

*ENS Rennes, av. Robert Schuman, F-35170 Bruz, France ; rozenn.texier@ens-rennes.fr

1 Méthodes élémentaires pour l'intégration numérique

Dans cette section, on se place sur un intervalle réel $[a, b]$ (borné ou éventuellement non borné) et on considère une fonction u définie sur l'intervalle réel $[a, b]$ telle que l'intégrale

$$I(u) = \int_a^b u(t)dt$$

soit convergente. On souhaite approcher cette intégrale par une formule de quadrature prenant la forme

$$Q(u) = \sum_{i=0}^{n} w_i u(x_i),$$

où n désigne un entier naturel, les $(x_i)_{0\leq i\leq n}$ sont des points bien choisis de $[a, b]$ qu'on appellera nœuds de quadrature, et les w_i sont des réels non nuls bien choisis que l'on appellera poids de la quadrature. On note enfin l'erreur de quadrature

$$E(u) = I(u) - Q(u).$$

Définition 1.1. Ordre d'une quadrature

L'ordre d'une formule de quadrature est le plus grand entier naturel $N \geq 0$ tel que la quadrature soit exacte pour tout élément de $\mathbb{R}_N[X]$.

1.1 Sur $[0, 1]$ - quadrature élémentaire

On a souvent recours à des méthodes de quadrature élémentaire consistant à remplacer u par un polynôme d'interpolation de Lagrange aux nœuds considérés. Par définition, ces méthodes sont exactes pour des polynômes de degré inférieur ou égal à $n-1$ ou n est le nombre de nœuds, mais des considérations de symétrie peuvent parfois conduire à un ordre plus grand. On obtient en particulier les méthodes présentées dans la Table 1.

1.2 Sur $[a, b]$ - quadrature composée

En pratique, lorsque ces méthodes sont choisies, l'intervalle $[a, b]$ est au préalable subdivisé en m sous-intervalles $[a_k, a_{k+1}]$, $0 \leq k \leq m-1$. On note $\sigma = (a_0, a_1, \cdots, a_m)$ la subdivision. La méthode élémentaire va être appliquée sur chaque sous-intervalle de la subdivision.

Définition 1.2. Quadrature composée

On appelle quadrature composée une quadrature sur $[a, b]$ obtenue via la réalisation de la quadrature élémentaire Q sur des sous-intervalles de $[a, b]$ rapportés à l'intervalle de référence $[0, 1]$ par la formule

$$Q_\sigma(u) = \sum_{k=0}^{m-1} (a_{k+1} - a_k) Q(t \mapsto u(a_k + (a_{k+1} - a_k)t)).$$

Méthode	Nœuds	Poids	Ordre
Rectangles à gauche	0	1	0
Rectangles à droite	1	1	0
Point milieu	$\frac{1}{2}$	1	1
Trapèzes	0,1	$\frac{1}{2}, \frac{1}{2}$	1
Simpson	$0, \frac{1}{2}, 1$	$\frac{1}{6}, \frac{4}{6}, \frac{1}{6}$	3

TABLE 1 – Quelques quadratures élémentaires classiques sur l'intervalle $[0, 1]$.

Exemples - cas d'une subdivision régulière : $a_k = a + kh$, $h = \frac{b-a}{m}$. On notera alors Q_h la quadrature composée.

- Rectangles à gauche : $Q_h(u) = h \sum_{k=0}^{m-1} u(a + kh)$
- Rectangles à droite : $Q_h(u) = h \sum_{k=1}^{m} u(a + kh)$
- Point milieu : $Q_h(u) = h \sum_{k=0}^{m-1} u(a + kh/2)$
- Trapèzes : $Q_h(u) = \frac{h}{2} \sum_{k=0}^{m-1} (u(a + kh) + u(a + (k + 1)h))$
- etc.

Ici, la quadrature composée hérite de l'ordre de la quadrature élémentaire.

1.3 Analyse d'erreur - Méthode des rectangles à gauche

Si σ est une subdivision de $[a, b]$, on note $h = \max_k |a_{k+1} - a_k|$ le pas de la subdivision.

Proposition 1.3. Convergence - méthode des rectangles

Soit Q_σ la quadrature composée déduite de la méthode des rectangles à gauche sur une subdivision σ de $[a, b]$. On a alors convergence lorsque le pas h de la subdivision tend vers 0, ceci pour toute fonction u Riemann-intégrable sur $[a, b]$: $\lim Q_\sigma(u) = I(u)$. De plus, pour toute fonction $u \in C^1([a, b])$, il existe une constante $C > 0$ indépendante de h telle que

$$|I(u) - Q_\sigma(u)| \leq Ch.$$

Preuve : Considérons d'abord une fonction $v \in C^1([0,1])$. On sait que pour tout $x \in [0,1]$, il existe $\xi_x \in [0,x]$ tel que $v(x) - v(0) = v'(\xi_x)x$. Ainsi, en notant Q la quadrature des rectangles à gauche,

$$\left| \int_0^1 v(x)dx - Q(v) \right| = \left| \int_0^1 xv'(\xi_x)dx \right| \leq \frac{1}{2}\|v'\|_\infty.$$

Soit maintenant $u \in C^1([a,b])$ et notons $v_k(x) = u(a_k + (a_{k+1} - a_k)x)$, $x \in [0,1]$. On a alors

$$\left| \int_a^b u(x)dx - Q_\sigma(u) \right| \leq \sum_{k=0}^{m-1} (a_{k+1} - a_k) \left| \int_0^1 v_k(x) - Q(v_k)dx \right| \leq \frac{1}{2} \sum_{k=0}^{k-1} \|v_k'\|_\infty (a_{k+1} - a_k) \leq \frac{k}{2}\|v'\|_\infty h.$$

Remarque : la même preuve s'adapte pour une méthode interpolatoire de degré supérieur.

2 Méthodes classiques pour l'approximation des EDO

Elles découlent des méthodes classiques pour l'intégration numérique. On suppose ici, pour simplifier, que l'intervalle d'étude $[0,T]$ est subdivisé avec un pas constant $h = T/N$, et on note $t_n = nh$. L'objectif est de construire y_n une approximation de $y(t_n)$ au moyen de méthodes dites à un pas, c'est-à-dire telles que y_{n+1} se calcule uniquement à partir de y_n, t_n, h. Voir la Table 2.

Les méthodes d'Euler explicite, Heun, Runge-Kutta sont explicites : y_{n+1} s'obtient à partir de y_n par une série de calculs explicites que l'on peut exprimer sous la forme : $y_{n+1} = y_n + h\phi(t_n, y_n, h)$, où ϕ dépend de la méthode. A l'inverse, les méthodes d'Euler implicite et de Crank Nicolson sont implicites : elles nécessitent à chaque étape la résolution d'une équation. Si f est non linéaire, cette équation sera résolue en pratique de façon approchée, par exemple par une méthode de Newton.

Notons qu'à une méthode d'intégration donnée, il peut correspondre plusieurs méthodes pour les EDO, selon qu'on explicite ou implicite les valeurs de f aux différents nœuds (Heun vs Crank-Nicolson), et selon la façon dont on approche ces valeurs dans le cas d'une méthode explicite. En particulier, la performance d'une méthode de calcul d'intégrale ne se retrouve dans la méthode EDO que sous certaines conditions (ex : Simpson et RK4).

Remarque : la méthode d'Euler explicite peut aussi se comprendre très simplement sur un dessin.

Rectangles à gauche $\int_{t_n}^{t_{n+1}} f(t,y(t))\mathrm{d}t$ approché par $hf(t_n,y(t_n))$	Euler explicite (ou progressive) (1768) $y_{n+1} = y_n + hf(t_n,y_n)$ $\phi(t,y,h) = f(t,y)$
Rectangles à droite $\int_{t_n}^{t_{n+1}} f(t,y(t))\mathrm{d}t$ approché par $hf(t_{n+1},y(t_{n+1}))$	Euler implicite (ou rétrograde) $y_{n+1} = y_n + hf(t_{n+1},y_{n+1})$
Rectangles à point milieu $\int_{t_n}^{t_{n+1}} f(t,y(t))\mathrm{d}t$ approché par $hf\left(t_n+\frac{h}{2}, y(t_n+\frac{h}{2})\right)$	Runge (1895) $y_{n+1} = y_n + hf\left(t_n+\frac{h}{2}, y_n+\frac{h}{2}f(t_n,y_n)\right)$ $\phi(t,y,h) = f(t+\frac{h}{2}, y+\frac{h}{2}f(t,y))$
Trapèzes $\int_{t_n}^{t_{n+1}} f(t,y(t))\mathrm{d}t$ approché par $\frac{h}{2}\left(f(t_n,y(t_n)) + f(t_{n+1},y(t_{n+1}))\right)$	Heun (1900) $y_{n+1} = y_n + \frac{h}{2}\left[f(t_n,y_n) + f(t_{n+1}, y_n + hf(t_n,y_n))\right]$ Crank Nicolson (1947) $y_{n+1} = y_n + \frac{h}{2}\left[f(t_n,y_n) + f(t_{n+1},y_{n+1})\right]$
Simpson $\int_{t_n}^{t_{n+1}} f(t,y(t))\mathrm{d}t$ approché par $\frac{h}{6}\left(u(t_n) + 4u(t_n+\frac{h}{2}) + u(t_{n+1})\right)$ où $u(t) = f(t,y(t))$	Runge-Kutta 4 (1901) $k_1^n = f(t_n,y_n),$ $k_2^n = f\left(t_n+\frac{h}{2}, y_n+\frac{h}{2}k_1^n\right)$ $k_3^n = f\left(t_n+\frac{h}{2}, y_n+\frac{h}{2}k_2^n\right)$ $k_4^n = f\left(t_{n+1}, y_n+hk_3^n\right)$ $y_{n+1} = y_n + \frac{h}{6}(k_1^n + 2k_2^n + 2k_3^n + k_4^n)$

TABLE 2 – Quelques méthodes classiques pour l'approximation des EDO.

3 Convergence de la méthode d'Euler explicite *

Considérons la solution y d'un problème de Cauchy (2), et son approximation par la méthode d'Euler explicite pour une subdivision à pas constant $h = \frac{T}{N}$. On a alors, pour tout $0 \leq n \leq N-1$:

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f(y(s))ds \tag{3}$$

$$y_{n+1} = y_n + hf(y_n). \tag{4}$$

Notons $e_n = y(t_n) - y_n$, $0 \leq n \leq N$.

Définition 3.1. Schéma convergent

On dit que le schéma (4) converge si, dès lors que $e_0 = 0$, on a :

$$\lim_{h \to 0} \left(\max_{0 \leq n \leq N(h)} |e_n| \right) = 0.$$

En soustrayant (4) de (3), on obtient

$$\begin{aligned} e_{n+1} &= e_n + \int_{t_n}^{t_{n+1}} f(y(s))ds - hf(y_n) \\ &= e_n + \underbrace{\int_{t_n}^{t_{n+1}} f(y(s))ds - hf(y(t_n))}_{\epsilon_n} + \underbrace{hf(y(t_n)) - hf(y_n)}_{\eta_n} \end{aligned}$$

On appelle ϵ_n l'erreur de consistance locale (de l'anglais, *consistent*=cohérent, concordant) relative à la solution y : c'est l'erreur qu'on commettrait sur le pas de temps $[t_n, t_n + h]$ si y_n était égal à $y(t_n)$. Il est bon de repérer ces différentes sources d'erreur sur un dessin.

Pour que le schéma soit convergent, il faut à la fois
- contrôler l'erreur de consistance globale $\sum_n |\epsilon_n|$ (méthode *consistante*),
- s'assurer que l'erreur finale est contrôlée par les différentes sources d'erreur qui peuvent s'accumuler à chaque pas (méthode *stable*).

Définition 3.2. Méthode consistante

On dit que la méthode est consistante si, pour toute solution y (i.e. pour toute donnée initiale y_0),

$$\lim_{h\to 0}\left(\sum_{k=0}^{N(h)-1}|\epsilon_k|\right)=0.$$

Définition 3.3. Méthode stable

On dit que la méthode d'Euler explicite est stable s'il existe une constante $M>0$ indépendante de h telle que, étant données deux suites (y_n) et (z_n) vérifiant le schéma d'Euler explicite respectivement de manière exacte et de manière approchée :

$$\forall n\in\{0,\cdots,N-1\},\quad y_{n+1} = y_n + hf(t_n,y_n), \tag{5}$$
$$\forall n\in\{0,\cdots,N-1\},\quad z_{n+1} = z_n + hf(t_n,z_n)+\eta_n, \tag{6}$$

on ait :

$$\max_{n\in\mathbb{N}}|y_n-z_n|\le M\left(|y_0-z_0|+\sum_{k=0}^{N(h)-1}|\eta_k|\right).$$

Plus généralement, un schéma à un pas défini par la récurrence

$$y_{n+1}=y_n+h\phi(t_n,y_n,h),$$

est stable s'il existe une constante $M>0$ indépendante de h telle que, étant données deux suites (y_n) et (z_n) vérifiant :

$$\forall n\in\{0,\cdots,N-1\},\quad y_{n+1} = y_n + h\phi(t_n,y_n,h), \tag{7}$$
$$\forall n\in\{0,\cdots,N-1\},\quad z_{n+1} = z_n + h\phi(t_n,z_n,h)+\eta_n, \tag{8}$$

on ait :

$$\max_{n\in\mathbb{N}}|y_n-z_n|\le M\left(|y_0-z_0|+\sum_{k=0}^{N(h)-1}|\eta_k|\right).$$

Vérifions que la méthode d'Euler explicite vérifie ces deux propriétés.

Montrons que la méthode est consistante.

Supposons que la fonction f est C^1, alors la solution exacte y est de classe C^2. On pose $M_2 = \sup_{0\leq s\leq T} |y''(s)|$. Alors on a

$$\sum_{k=0}^{N-1} |\epsilon_k| \leq \frac{M_2}{2} Th. \tag{9}$$

Ainsi, la méthode d'Euler explicite est consistante.

Remarque : on peut se passer de l'hypothèse $f \in C^1$ en utilisant seulement son uniforme continuité sur un compact $[0,T] \times y([0,T])$. Les détails sont laissés en exercice.

Montrons que la méthode est stable (cas globalement lipschitzien).

Soient (y_n) et (z_n) deux suites comme dans la définition de la stabilité. On suppose ici que la fonction f est globalement lipschitzienne par rapport à la 2ème variable sur $[0,T] \times \Omega$, notons L sa constante de Lipschitz. On montre la majoration suivante :

$$\max_{0\leq n\leq N} |y_n - z_n| \leq e^{TL} \left(|y_0 - z_0| + \sum_{k=0}^{N-1} |\eta_k| \right). \tag{10}$$

On a bien montré la stabilité de la méthode, dès que f est globalement lipschitzienne par rapport à la variable y.

Montrons qu'une méthode consistante et stable est convergente.

Soit y la solution exacte et $(y_n)_{0\leq n\leq N}$ la suite des valeurs approchées construites par la méthode d'Euler explicite, ou toute autre méthode explicite à un pas, consistante et stable. On remarque que si on pose $z_n = y(t_n)$, la suite (z_n) vérifie de façon approchée le schéma considéré, avec une erreur au pas n qui est précisément l'erreur de consistance locale ϵ_n.

La méthode étant stable, on peut écrire :

$$\max_{0\leq n\leq N} |e_n| \leq M \left(|y_0 - z_0| + \sum_{k=0}^{N-1} |\epsilon_k| \right).$$

Supposons la donnée initiale connue exactement, i.e. $y_0 = z_0$. La consistance de la méthode montre alors que le membre de droite tend vers 0 quand le pas h tend vers 0, ainsi la méthode est convergente.

Dans le cas particulier de la méthode d'Euler explicite, les estimations (9) et (10) conduisent au résultat

$$\max_{0\leq n\leq N} |e_n| \leq e^{TL} \frac{M_2 T}{2} h.$$

Ainsi, la méthode est convergente, on dit qu'elle est d'ordre (global) 1.

Remarques :

- En affinant un peu ce raisonnement, on peut montrer que la condition de Lipschitz locale pour f suffit à assurer la convergence de la méthode.
- On montre de même que la méthode de Runge est d'ordre 2, i.e. on a une estimation du type

$$\max_{0\leq n\leq N} |e_n| \leq Ch^2.$$

La méthode de Runge Kutta 4 est d'ordre 4. Ces ordres peuvent être vérifiés numériquement, voir figure 5.

Définition 3.4. Ordre d'une méthode

Une méthode à 1 pas est d'ordre (global) au moins $p \geq 1$ si il existe K ne dépendant que de y tel que

$$\sum_{n=0}^{N(h)-1} |\epsilon_n| \leq Kh^p.$$

Théorème 3.5. Convergence d'ordre p

Si une méthode est stable et d'ordre p et si $f \in C^p$ alors

$$\forall n \in \mathbb{N}, |y(t_n) - y_n| \leq M\left(|y_0 - y_0^h| + Kh^p\right)$$

où M est comme dans la définition 3.3 et K est comme dans la définition 3.4.

4 En pratique : limitations des méthodes explicites *

4.1 Problèmes raides

Considérons le problème ci-dessous :

$$y'(t) = \lambda y(t), t \in [0,1], \quad y(0) = 1, \tag{11}$$

avec $\lambda = -10^4$. (Cela peut modéliser une cinétique chimique très rapide.) La solution exacte est donnée par

$$y(t) = e^{\lambda t}, t \in [0,1].$$

Elle décroît très vite vers 0. En particulier, $y(1) = e^{\lambda}$ est plus petit que la précision machine. Si le pas $h = 1/N$ n'est pas suffisamment petit, la méthode d'Euler explicite ne va pas bien se comporter : ici, la solution approchée par Euler explicite est donnée par

$$y_n^e = (1 + h\lambda)^n, 0 \leq n \leq N,$$

qui n'a pas le comportement attendu si $|1 + h\lambda| \geq 1$. En particulier,

$$\text{si } N = \frac{1}{h} = 100, \; y_N^e = (-99)^{100} >> e^{\lambda} = y(1).$$

On dit ici que la méthode explicite est instable.

Pour ce type de problèmes, il est conseillé d'utiliser une méthode implicite.

Plus généralement, pour un système différentiel linéaire

$$y' = Ay,$$

avec A diagonalisable de valeurs propres $\lambda_1, ..., \lambda_N \in \mathbb{C}$, la suite $(y_n^e)_{n\in\mathbb{N}}$ des solutions approchées par méthode d'Euler explicite est "stable" si et seulement si $\forall 1 \leq j \leq N, |1 + \lambda_j h| \leq 1$. (Pour un système non linéaire $y' = F(y)$, c'est le spectre de $DF(y_0)$ qui va intervenir.)

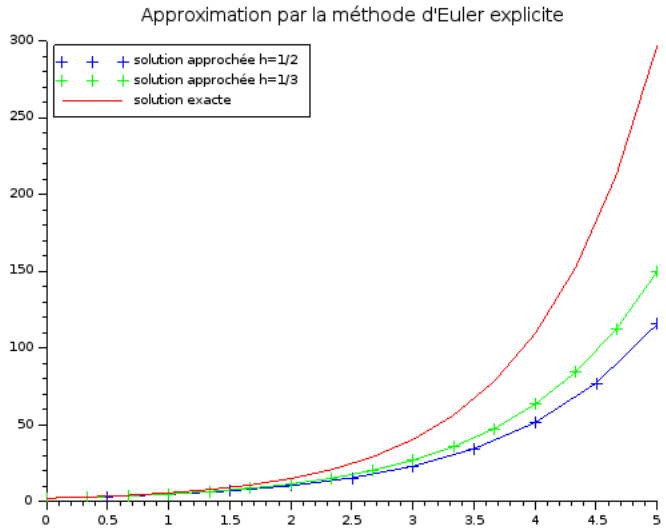

FIGURE 1 – Résolution approchée par la méthode d'Euler explicite pour l'équation $y' = y$, $y(0) = 2$.

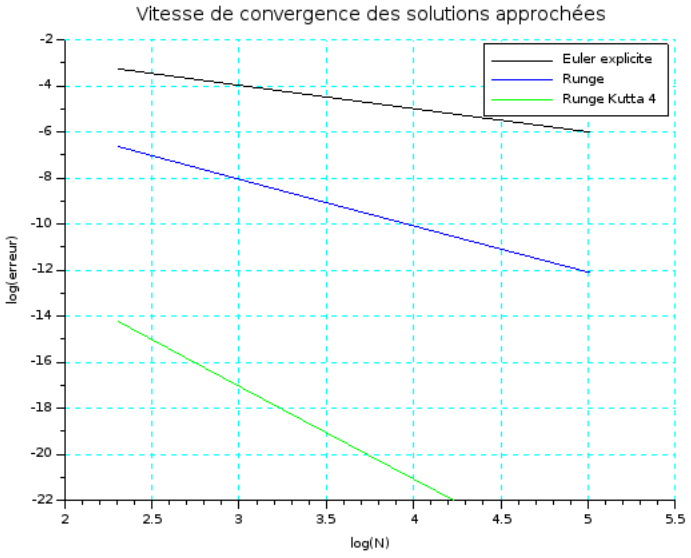

FIGURE 2 – Visualisation de l'ordre de convergence de trois méthodes numériques

4.2 Systèmes hamiltoniens

On considère le problème suivant, dit de l'oscillateur harmonique,

$$\left\{ \begin{array}{rcl} x' & = & -y \\ y' & = & x \end{array} \right. \quad \text{avec } x(0) = 1, y(0) = 0. \tag{12}$$

Remarquons que y vérifie l'équation du pendule linéarisé $y" = -y$. La solution est donnée par

$$x(t) = \cos t, \ y(t) = \sin t,$$

et la trajectoire dans le plan de phase est le cercle unité. La quantité $H(t) = x^2(t) + y^2(t)$ est conservée, on la nomme *hamiltonien* du système (12). (Elle représente l'énergie mécanique du pendule.)

Comparons les comportements de différentes méthodes pour ce problème. Un calcul simple nous montre que l'hamiltonien n'est pas conservé par les méthodes d'Euler. Plus précisément, on a les comportements suivants.

Euler explicite	$x_{n+1}^2 + y_{n+1}^2 = (1+h^2)(x_n^2 + y_n^2)$.
Euler implicite	$x_{n+1}^2 + y_{n+1}^2 = (1+h^2)^{-1}(x_n^2 + y_n^2)$.
Runge Kutta 4	$x_{n+1}^2 + y_{n+1}^2 = \left(1 - \frac{h^6}{72} + \frac{h^8}{576}\right)(x_n^2 + y_n^2)$.
Crank-Nicolson	$x_{n+1}^2 + y_{n+1}^2 = x_n^2 + y_n^2$.

On remarque que seule la méthode de Crank-Nicolson préserve l'hamiltonien, ce qui est crucial lorsque le système doit être étudié sur des temps longs.

Une alternative entièrement explicite est d'utiliser un schéma d'Euler symplectique défini pour un système de la forme

$$y' = x, \quad x' = f(y)$$

par

$$\left\{ \begin{array}{rcl} y_{n+1} & = & y_n + hx_n \\ x_{n+1} & = & x_n + hf(y_n + hx_n) \end{array} \right.$$

Ce schéma ne préserve pas l'hamiltonien exact $H(x_n, y_n)$ mais il préserve un hamiltonien approché, dans le cas de l'oscillateur harmonique :

$$H_h(x, y) = x^2 + y^2 + hxy.$$

On a donc la garantie que pour des temps longs, la solution garde un comportement "raisonnable" par rapport au modèle physique étudié.

Références

[1] Crouzeix, Mignot, Analyse numérique des équations différentielles, Masson, Paris 1984.

[2] Hairer, Norsett, Wanner, Solving ordinary differential equations. I. Springer-Verlag, Berlin, 1993.

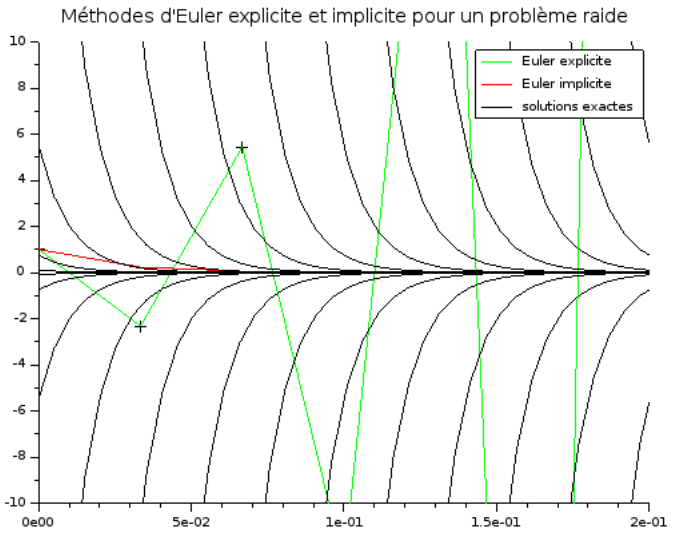

FIGURE 3 – Résolution approchée d'un problème raide $y' = -100y$, $y(0) = 1$.

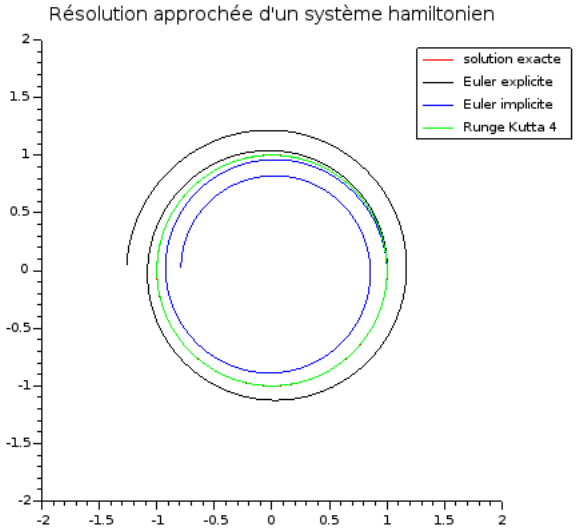

FIGURE 4 – Résolution exacte et approchée du système hamiltonien $x' = -y,\ y' = x$, $x(0) = 1$, $y(0) = 0$.