

UPMC L3 Mathématiques
LM336 Méthodes numériques pour les équations différentielles

Examen première session, 13 mai 2014

Durée : 2 heures. Malgré le fil conducteur cinématographique, les trois parties du problème sont en réalité pratiquement indépendantes les unes des autres. L'ensemble semblant un peu long, il est recommandé de tout lire avant de se lancer et de choisir par où commencer.

Pas de documents autorisés. Pas de calculatrice, pas de téléphone portable allumé : tous les gadgets électroniques susceptibles de stocker ou de transmettre de l'information doivent être éteints et rangés dans les sacs. En particulier, tout téléphone portable trouvé allumé en cours d'épreuve sera saisi (article 22 du décret n° 92-657 du 13 juillet 1992).

Question de cours

Définir la consistance d'un schéma explicite à un pas de la forme $y_{n+1} = y_n + hF(t_n, y_n, h)$ destiné à approcher numériquement la solution d'un problème de Cauchy $y'(t) = f(t, y(t))$, $y(0) = y_0$. Donner une condition suffisante de consistance portant sur la fonction F .

Problème, partie I

À la suite de la destruction de leur navette spatiale par des débris à haute vitesse, deux astronautes, le commandant G et le docteur S , se retrouvent à la dérive en orbite autour de la Terre. Pour simplifier, et même si cela doit nuire à la crédibilité scientifique de l'intrigue, on suppose que G et S sont ponctuels, et l'on note $y_G(t) \in \mathbb{R}^3$ et $y_S(t) \in \mathbb{R}^3$ la position de chacun d'entre eux à l'instant $t \in \mathbb{R}_+$. De même, on néglige les effets de la gravitation terrestre. On note m_G et m_S les masses respectives des deux astronautes et $y_{G,0}$, $v_{G,0}$, $y_{S,0}$ et $v_{S,0}$ leur position et vitesse respective à l'instant $t = 0$.

Le scénario précise que G et S sont reliés par un câble de longueur au repos L qui n'exerce aucune force quand il n'est pas tendu, et se comporte comme un ressort de raideur $k > 0$ quand il est tendu. En d'autres termes, la force exercée par le câble sur G est nulle quand $\|y_G(t) - y_S(t)\| \leq L$ et vaut $k(\|y_G(t) - y_S(t)\| - L) \frac{y_S(t) - y_G(t)}{\|y_G(t) - y_S(t)\|}$ dans le cas contraire. Le câble exerce une force exactement opposée sur S . Il est considéré être de masse négligeable. Enfin, le commandant G dispose de propulseurs qui exercent sur lui une force de la forme $g(t)a$ où $g: \mathbb{R}_+ \rightarrow \mathbb{R}$ est une fonction continue donnée et $a \in \mathbb{R}^3$ un vecteur donné. On néglige la diminution de masse de G due au fonctionnement des propulseurs.

On notera $x_+ = \max(x, 0)$ pour tout $x \in \mathbb{R}$ la fonction partie positive.

a. Montrer que le problème de Cauchy satisfait par le système formé de G , S et de leur câble est de la forme

$$\begin{cases} m_G \ddot{y}_G(t) = g(t)a + f(y_G(t) - y_S(t)) \\ m_S \ddot{y}_S(t) = -f(y_G(t) - y_S(t)) \\ y_G(0) = y_{G,0}, y_S(0) = y_{S,0}, \dot{y}_G(0) = v_{G,0}, \dot{y}_S(0) = v_{S,0}, \end{cases} \quad (1)$$

où $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ est donnée par $f(z) = u(z)v(z)$, où $u(z) = -k(\|z\| - L)_+$ et $v(z) = \frac{z}{\|z\|}$ si $\|z\| \geq L$, $v(z) = \frac{z}{L}$ si $\|z\| < L$.

b. Réécrire le problème de Cauchy (1) sous la forme standard $Y'(t) = F(t, Y(t))$, $Y(0) = Y_0$, pour un certain $F: \mathbb{R}_+ \times \mathbb{R}^{12} \rightarrow \mathbb{R}^{12}$ que l'on identifiera.

- c. Montrer que l'application $x \rightarrow x_+$ est globalement lipschitzienne de \mathbb{R} dans \mathbb{R} puis que l'application u est globalement lipschitzienne de \mathbb{R}^3 dans \mathbb{R} .
- d. Montrer que l'application v est globalement lipschitzienne de constante $\frac{2}{L}$ de \mathbb{R}^3 à valeurs dans \mathbb{R}^3 .
- e. Montrer que le produit d'une application globalement lipschitzienne de \mathbb{R}^3 dans \mathbb{R} avec une application globalement lipschitzienne de \mathbb{R}^3 dans \mathbb{R}^3 est une application localement lipschitzienne de \mathbb{R}^3 dans \mathbb{R}^3 . En déduire que la fonction F est localement lipschitzienne par rapport à Y , uniformément par rapport à t sur $\mathbb{R}_+ \times \mathbb{R}^{12}$ (on pourra procéder par quatre blocs de dimension 3 et utiliser la norme sur \mathbb{R}^{12} définie par $\|Y\|_{\mathbb{R}^{12}} = \sum_{i=1}^4 \|Y_i\|$, où $Y = (Y_1^t \ Y_2^t \ Y_3^t \ Y_4^t)^t$, $Y_i \in \mathbb{R}^3$, et la norme figurant dans la somme est la norme euclidienne sur \mathbb{R}^3).
- f. En déduire que le problème de Cauchy (1) admet une solution maximale et une seule.

Problème, partie II

On pose $y_I(t) = \frac{1}{m_G+m_S}(m_G y_G(t) + m_S y_S(t))$ et $d(t) = y_G(t) - y_S(t)$.

- a. Montrer que $\ddot{y}_I(t) = \frac{1}{m_G+m_S}g(t)a$ et que $\ddot{d}(t) = D(t, d(t))$ avec $D(t, d) = \frac{g(t)}{m_G}a + \frac{1}{M}f(d)$ pour tout $t \in \mathbb{R}_+$, $d \in \mathbb{R}^3$, où $M = \frac{m_G m_S}{m_G+m_S}$. Quel est l'intérêt de ce changement de variables ?
- b. Montrer que $y_I(t) = \frac{1}{m_G+m_S}(\int_0^t (t-s)g(s)ds)a + t\dot{y}_I(0) + y_I(0)$.
- c. Avec son flegme légendaire, le commandant G informe le docteur S qu'il va se retrouver à court de carburant à l'instant $t_0 > 0$, donc $g(t) = 0$ pour tout $t \geq t_0$. Sachant que G a fixé le vecteur a en direction de la station spatiale internationale à $t = 0$, nos héros ont-ils la moindre chance d'atteindre cette station ?

Problème, partie III

La production a refait ses comptes et s'est tout d'un coup aperçue qu'il revient beaucoup trop cher de tourner dans l'espace : elle va finalement tout faire en images de synthèse. Vous êtes embauchés pour calculer des approximations numériques réalistes de l'évolution du vecteur d . Pour vous faire la main, vous commencez par un cas simple : les deux astronautes se déplacent sur une même droite, d est donc à valeurs scalaires, $a = 1$, $m_G = m_S = 1$ et $L = 1$. Pour simplifier encore plus, vous enlevez une dérivée en temps et considérez donc le problème de Cauchy

$$\begin{cases} d'(t) = g(t) + f_1(d(t)), \\ d(0) = d_0, \end{cases} \quad (2)$$

avec $f_1(d) = -k \text{signe}(d)(|d| - 1)_+$, où $\text{signe}(d) = 1$ si $d \geq 0$, $\text{signe}(d) = -1$ si $d < 0$.

On considère les schémas d'Euler explicite $d_{e,n+1} = d_{e,n} + h(g(t_n) + f_1(d_{e,n}))$, $d_{e,0} = d_0$, et implicite $d_{i,n+1} = d_{i,n} + h(g(t_n) + f_1(d_{i,n+1}))$, $d_{i,0} = d_0$, $h = \frac{1}{N}$ est le pas de temps avec $N \in \mathbb{N}^*$, supposé assez petit pour que le schéma implicite soit bien défini.

- a. Indiquer rapidement pourquoi le problème (2) a une solution et une seule définie sur \mathbb{R}_+ .
- b. Calculer explicitement cette solution pour $d_0 = 0$ et $g(t) = 1$ pour tout t . On pourra distinguer ce qui se passe pour $t \leq 1$ et ce qui se passe pour $t \geq 1$.
- c. Sous les hypothèses de **b**, montrer que pour tout $n \leq N$, $d_{e,n} = d_{i,n} = d(t_n)$.
- d. Toujours sous les hypothèses de **b**, calculer $d_{e,N+1}$ et $d_{i,N+1}$. En supposant que le câble est très raide, c'est-à-dire que k est très grand, quelle méthode semble la meilleure du point de vue de ces deux valeurs ?
- e. Revenant au cas général, peut-on avoir d de classe au moins C^2 si d prend les valeurs 1 ou -1 ? Est-il alors a priori raisonnable d'appliquer un schéma d'ordre plus élevé ?
- f. On remplace f_1 par une approximation f_2 de classe C^2 , non spécifiée, globalement lipschitzienne de constante de Lipschitz K , et l'on suppose également g de classe C^2 . Montrer que d est alors de classe C^3 .
- g. On considère le schéma de Crank-Nicolson,

$$d_{CN,n+1} = d_{CN,n} + \frac{h}{2} (g(t_n) + f_2(d_{CN,n}) + g(t_{n+1}) + f_2(d_{CN,n+1})), \quad d_{CN,0} = d_0.$$

Montrer qu'il est bien défini pour h suffisamment petit.

- h. Son erreur de consistance est donnée par

$$\varepsilon_n = d(t_{n+1}) - d(t_n) - \frac{h}{2} (g(t_n) + f_2(d(t_n)) + g(t_{n+1}) + f_2(d(t_{n+1}))).$$

Montrer que ce schéma est d'ordre 2.

- i. Montrer que le schéma de Crank-Nicolson est stable pour h assez petit. En déduire qu'il est convergent et donner une estimation de l'erreur.

Correction problème, partie I

Deux remarques générales pour commencer : si on connaît son cours (mais je ne peux pas imaginer un seul instant que l'on ne connaisse pas son cours), on expédie la question de cours en deux-trois minutes et on engrange rapidement un nombre non négligeable de points. En plus, on a la satisfaction intellectuelle de connaître le cours, ce qui est important en soi. Deuxième remarque, il est totalement inutile de recopier l'énoncé, c'est une perte de temps.

Le corrigé qui suit est relativement complet, et mériterait une note de, disons en toute modestie, 120/60, mais comme c'est moi qui note, c'est peut-être un peu biaisé. Dans les contraintes du temps limité, on se satisfait de moins pour ce qui concerne la rédaction des détails... Le corrigé contient aussi quelques commentaires au fil de l'eau inspirés par les aspects les moins inspirants de certaines copies. Mais pas tous, comme définir une fonction sur $\mathbb{R}_+ \times \mathbb{R}^{12}$ par une formule donnant $F(t, Y(t))$ (donc une fonction seulement de t) alors qu'on ne sait même pas encore que $t \mapsto Y(t)$ existe. J'ai déjà longuement déploré tout cela dans les corrigés de partiels et d'examens disponibles sur Sakai, sans grand succès apparent.

a. La première équation est la loi de Newton appliquée à G : l'accélération de G , c'est-à-dire \ddot{y}_G , multipliée par la masse de G (constante) est égale à la somme des forces appliquées à G . La force venant du propulseur est donnée par $g(t)a$ et celle exercée sur G par le câble est donnée par $k(\|y_G(t) - y_S(t)\| - L)_+ v(y_S(t) - y_G(t))$. En effet, par définition de la partie positive, on a $x_+ = x$ si $x \geq 0$ et $x_+ = 0$ si $x \leq 0$. Par conséquent, $(\|y_G(t) - y_S(t)\| - L)_+ = 0$ quand $\|y_G(t) - y_S(t)\| - L \leq 0$, c'est-à-dire quand $\|y_G(t) - y_S(t)\| \leq L$, auquel cas la valeur de v n'a aucune importance (celle donnée par l'énoncé a le bon goût d'être définie partout vu qu'on ne divise jamais par 0, continue et même globalement lipschitzienne comme on le verra plus tard, mais on aurait pu en prendre une autre) puisque cette valeur est multipliée par 0 pour bien donner une force nulle. Et dans le cas contraire $(\|y_G(t) - y_S(t)\| - L)_+ = \|y_G(t) - y_S(t)\| - L$ quand $\|y_G(t) - y_S(t)\| \geq L$, c'est-à-dire exactement là où $v(z) = \frac{z}{\|z\|}$. Même chose pour S sans force du propulseur et avec un signe opposé pour la force exercée par le câble. Enfin, on réécrit simplement les conditions initiales. C'est un système du second ordre à valeurs dans $\mathbb{R}^3 \times \mathbb{R}^3 = \mathbb{R}^6$, couplé en les deux inconnues y_G et y_S .

b. Un système du second ordre à valeurs dans \mathbb{R}^6 se réécrit naturellement en système du premier ordre à valeurs dans $\mathbb{R}^6 \times \mathbb{R}^6 = \mathbb{R}^{12}$. On pose classiquement

$$Y(t) = \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \\ Y_4(t) \end{pmatrix} = \begin{pmatrix} y_G(t) \\ y_S(t) \\ \dot{y}_G(t) \\ \dot{y}_S(t) \end{pmatrix} \in \mathbb{R}^{12},$$

avec quatre blocs Y_i , $i = 1, \dots, 4$, de dimension 3. Il vient donc pour l'EDO

$$Y'(t) = \begin{pmatrix} \dot{y}_G(t) \\ \dot{y}_S(t) \\ \ddot{y}_G(t) \\ \ddot{y}_S(t) \end{pmatrix} = \begin{pmatrix} \dot{y}_G(t) \\ \dot{y}_S(t) \\ \frac{1}{m_G}(g(t)a + f(y_G(t) - y_S(t))) \\ -\frac{1}{m_S}f(y_G(t) - y_S(t)) \end{pmatrix} = \begin{pmatrix} Y_3(t) \\ Y_4(t) \\ \frac{1}{m_G}(g(t)a + f(Y_1(t) - Y_2(t))) \\ -\frac{1}{m_S}f(Y_1(t) - Y_2(t)) \end{pmatrix},$$

d'où la fonction second membre que l'on lit sur la formule précédente en supprimant la dépendance en t des Y_i

$$\forall (t, Y) \in \mathbb{R}_+ \times \mathbb{R}^{12}, \quad F(t, Y) = \begin{pmatrix} Y_3 \\ Y_4 \\ \frac{1}{m_G}(g(t)a + f(Y_1 - Y_2)) \\ -\frac{1}{m_S}f(Y_1 - Y_2) \end{pmatrix}.$$

Éviter ici les notations du style \dot{y}_G qui évoquent une dérivée temporelle pour désigner des variables qui ne dépendent pas du temps. Pour la donnée initiale, on obtient

$$Y(0) = Y_0 = \begin{pmatrix} y_{G,0} \\ y_{S,0} \\ v_{G,0} \\ v_{S,0} \end{pmatrix}.$$

On pouvait mélanger les blocs de \mathbb{R}^3 dans l'ordre que l'on voulait, mais je n'ai vu qu'une petite partie des 4! possibilités (voire 12! possibilités si l'on écrit toutes les composantes scalaires).

c. On peut remarquer (mais ce n'est pas obligatoire, on pouvait aussi bien procéder par disjonction des cas, avec trois cas à considérer) que $x_+ = \frac{1}{2}(x + |x|)$. Par conséquent, pour tout couple de réels x et y

$$x_+ - y_+ = \frac{1}{2}(x - y + |x| - |y|),$$

d'où en prenant la valeur absolue et en utilisant deux fois l'inégalité triangulaire

$$|x_+ - y_+| \leq \frac{1}{2}(|x - y| + ||x| - |y||) \leq |x - y|.$$

C'est donc une fonction 1-lipschitzienne sur \mathbb{R} , ce qui est d'ailleurs évident en regardant son graphe. À ce propos, quasiment personne ne fait de dessin, alors que c'est une puissante aide au raisonnement de façon générale, et que cela peut parfois permettre d'éviter d'écrire des bêtises. C'est bien dommage.

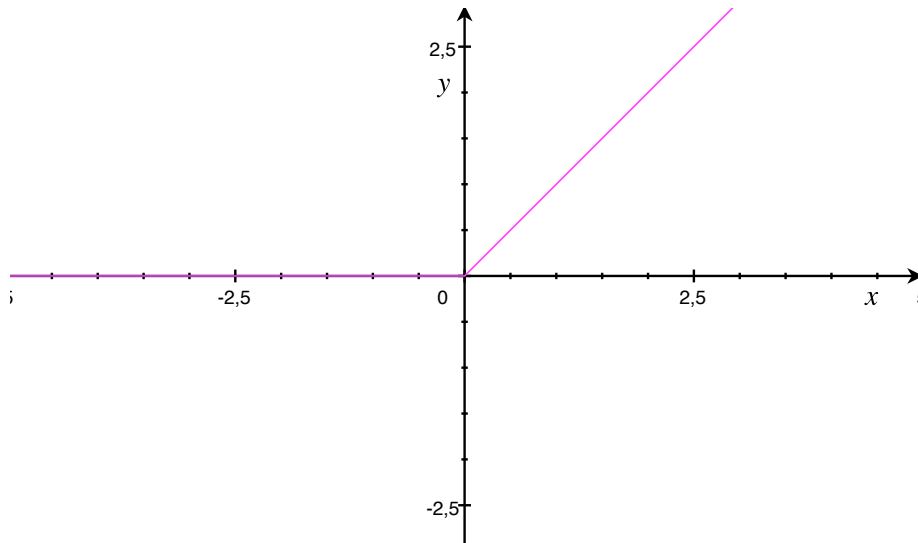


FIGURE 1 – Le graphe de la fonction partie positive.

Comme toujours, il vaut mieux éviter de dériver des fonctions non dérivables pour parvenir à ses fins. Et non, le max n'est pas une application linéaire ou bilinéaire.

On a par conséquent

$$|(\|z_1\| - L)_+ - (\|z_2\| - L)_+| \leq \left| \|z_1\| - L - (\|z_2\| - L) \right| = \left| \|z_1\| - \|z_2\| \right| \leq \|z_1 - z_2\|,$$

par le résultat précédent pour la première inégalité, et l'inégalité triangulaire à la dernière étape, d'où l'on déduit que u est k -lipschitzienne (on rappelle que $k > 0$, sinon u est $|k|$ -lipschitzienne. Le signe de k n'a pas grande importance en fait, hors de son interprétation physique).

d. Soit $A = \{z \in \mathbb{R}^3; \|z\| \geq L\}$. On distingue trois cas. Si z_1 et z_2 appartiennent à A , on peut écrire

$$\begin{aligned} \|v(z_1) - v(z_2)\| &= \left\| \frac{z_1}{\|z_1\|} - \frac{z_2}{\|z_2\|} \right\| = \frac{\| \|z_2\| z_1 - \|z_1\| z_2 \|}{\|z_1\| \|z_2\|} = \frac{\| (\|z_2\| - \|z_1\|) z_1 + \|z_1\| (z_1 - z_2) \|}{\|z_1\| \|z_2\|} \\ &\leq \frac{\| (\|z_2\| - \|z_1\|) z_1 \| + \| \|z_1\| (z_1 - z_2) \|}{\|z_1\| \|z_2\|} \\ &= \frac{\|z_2\| - \|z_1\| \|z_1\| + \|z_1\| \|z_1 - z_2\|}{\|z_1\| \|z_2\|} = \frac{\|z_2\| - \|z_1\| + \|z_1 - z_2\|}{\|z_2\|} \\ &\leq \frac{2}{L} \|z_1 - z_2\|, \end{aligned}$$

encore par l'inégalité triangulaire (deux fois), l'homogénéité positive de la norme et le fait que $\|z_2\| \geq L$.

Si maintenant $z_1 \in A$ et $z_2 \notin A$, alors

$$\begin{aligned} \|v(z_1) - v(z_2)\| &= \left\| \frac{z_1}{\|z_1\|} - \frac{z_2}{L} \right\| = \frac{\| [Lz_1 - \|z_1\| z_2] \|}{L \|z_1\|} = \frac{\| (L - \|z_1\|) z_1 + \|z_1\| (z_1 - z_2) \|}{L \|z_1\|} \\ &\leq \frac{|L - \|z_1\|| + \|z_1 - z_2\|}{L} \\ &\leq \frac{2}{L} \|z_1 - z_2\|, \end{aligned}$$

car $|L - \|z_1\|| = \|z_1\| - L \leq \|z_1\| - \|z_2\|$, vu que $\|z_2\| \leq L$, et bien sûr $\|z_1\| - \|z_2\| \leq \|z_1 - z_2\|$. Pas besoin de regarder le cas $z_1 \notin A$ et $z_2 \in A$ qui découle du précédent en échangeant les variables.

Enfin si $z_1 \notin A$ et $z_2 \notin A$, alors

$$\|v(z_1) - v(z_2)\| = \left\| \frac{z_1}{L} - \frac{z_2}{L} \right\| = \frac{1}{L} \|z_1 - z_2\|,$$

et l'on conclut par le fait que $\frac{2}{L} \geq \frac{1}{L}$.

On peut montrer avec plus de travail¹ qu'en fait la meilleure constante de Lipschitz de v est bien $\frac{1}{L}$, mais sûrement pas de la façon dont un bon nombre de copies ont cru l'obtenir, avec des majorations plutôt sauvages...

e. Soient $u: \mathbb{R}^3 \rightarrow \mathbb{R}$ et $v: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ les deux applications en question, L_u et L_v leur constante de Lipschitz respective. Pour tout couple de vecteurs y_1 et y_2 de \mathbb{R}^3 , on écrit

$$u(y_1)v(y_1) - u(y_2)v(y_2) = u(y_1)(v(y_1) - v(y_2)) + (u(y_1) - u(y_2))v(y_2),$$

1. À moins qu'il n'y ait une preuve plus astucieuse à laquelle je n'ai pas pensé, ce qui est loin d'être exclu.

d'où en prenant les normes et en utilisant l'inégalité triangulaire

$$\begin{aligned}\|u(y_1)v(y_1) - u(y_2)v(y_2)\| &\leq |u(y_1)|\|v(y_1) - v(y_2)\| + |u(y_1) - u(y_2)|\|v(y_2)\| \\ &\leq (L_v|u(y_1)| + L_u\|v(y_2)\|)\|y_1 - y_2\|.\end{aligned}$$

Les fonctions u et v étant lipschitziennes, elles sont continues et donc bornées sur toute boule fermée \bar{B} de \mathbb{R}^3 , car une telle boule est compacte. Il s'ensuit que si y_1 et y_2 appartiennent à \bar{B} , alors

$$\|u(y_1)v(y_1) - u(y_2)v(y_2)\| \leq (L_v \max_{\bar{B}} |u| + L_u \max_{\bar{B}} \|v\|)\|y_1 - y_2\|,$$

et la fonction uv est localement lipschitzienne. Le caractère local mais a priori pas global vient du fait que la constante $L_v \max_{\bar{B}} |u| + L_u \max_{\bar{B}} \|v\|$ dépend de la boule \bar{B} , et n'a aucune raison d'être majorée sur \mathbb{R}^3 entier. En particulier si $|u|$ ou $\|v\|$ n'est pas majorée. Considérons par exemple l'exemple $u(y) = v(y) = y$ de \mathbb{R} dans \mathbb{R} .

Appliquons cela à F . Pour tous $t \in \mathbb{R}$, $Y, Z \in \mathbb{R}^{12}$, on a

$$F(t, Y) - F(t, Z) = \begin{pmatrix} Y_3 - Z_3 \\ Y_4 - Z_4 \\ \frac{1}{m_G}(f(Y_1 - Y_2) - f(Z_1 - Z_2)) \\ -\frac{1}{m_S}(f(Y_1 - Y_2) - f(Z_1 - Z_2)) \end{pmatrix}.$$

Le membre de droite ne dépend pas de t , donc tout sera uniforme par rapport à t . Utilisons la norme suggérée dans l'énoncé (elle simplifie les choses, mais on peut aussi prendre n'importe quelle autre norme si l'on préfère).

$$\|F(t, Y) - F(t, Z)\|_{\mathbb{R}^{12}} = \|Y_3 - Z_3\| + \|Y_4 - Z_4\| + \left(\frac{1}{m_G} + \frac{1}{m_S}\right)\|f(Y_1 - Y_2) - f(Z_1 - Z_2)\|.$$

On voit qu'il suffit d'estimer le terme $\|f(Y_1 - Y_2) - f(Z_1 - Z_2)\|$. Comme $f(z) = u(z)v(z)$, on applique le résultat précédent. Pour toute boule fermée \bar{B} de \mathbb{R}^{12} , si Y et Z appartiennent à \bar{B} , alors $Y_1 - Y_2$ et $Z_1 - Z_2$ appartiennent à un compact de \mathbb{R}^3 (comme image continue du compact \bar{B} par l'application continue $\mathbb{R}^{12} \rightarrow \mathbb{R}^3$, $Y \mapsto Y_1 - Y_2$) que l'on peut donc inclure dans une boule fermée de \mathbb{R}^3 et l'on peut alors écrire

$$\|f(Y_1 - Y_2) - f(Z_1 - Z_2)\| \leq K\|Y_1 - Y_2 - (Z_1 - Z_2)\| \leq K(\|Y_1 - Z_1\| + \|Y_2 - Z_2\|)$$

pour un certain K dépendant de la boule de \mathbb{R}^3 dans laquelle $Y_1 - Y_2$ et $Z_1 - Z_2$ se retrouvent. Par conséquent, pour tous Y et Z dans la boule \bar{B} de \mathbb{R}^{12} de départ,

$$\|F(t, Y) - F(t, Z)\|_{\mathbb{R}^{12}} \leq \max\left(1, K\left(\frac{1}{m_G} + \frac{1}{m_S}\right)\right)\|Y - Z\|_{\mathbb{R}^{12}}.$$

f. Le second membre est continu sur $\mathbb{R}_+ \times \mathbb{R}^{12}$ (ne pas l'oublier sous prétexte que ce n'est pas écrit à la question **e**) puisque g est continue, et localement lipschitzien par rapport à Y , uniformément par

rapport à t , donc le théorème de Cauchy-Lipschitz local nous assure de l'existence et de l'unicité d'une solution locale pour toute donnée initiale. On sait par ailleurs que cette solution locale peut être prolongée en une solution maximale, qui est aussi unique, mais on ne sait pas sur quel intervalle a priori, et en fait on ne sait rien de plus.

En travaillant un peu plus finement, on peut montrer que la fonction f est non seulement localement lipschitzienne, mais en fait globalement $3|k|$ -lipschitzienne sur \mathbb{R}^3 , ce qui implique que F est aussi globalement lipschitzienne par rapport à Y , uniformément par rapport à t , et le théorème de Cauchy-Lipschitz global nous donne l'existence et l'unicité sur tout intervalle compact $[0, T]$. Cette unicité permet d'étendre la solution à \mathbb{R}_+ tout entier comme dans le partiel. On peut également introduire l'énergie du système et s'en servir pour borner la solution en tout temps fini à l'aide du lemme de Grönwall², une autre façon d'obtenir l'existence globale sur \mathbb{R}_+ .

Correction problème, partie II

a. Calcul immédiat. L'intérêt est de découpler le problème couplé initial en deux problèmes indépendants l'un de l'autre. En effet, la connaissance de y_I et de d permet immédiatement de reconstruire y_G et y_S . Le premier problème revient juste à une double intégration, ce n'est plus une EDO à proprement parler, et le deuxième est un problème de Cauchy du second ordre à valeurs dans \mathbb{R}^3 , donc de dimension 6 moitié de celle du problème de départ. On simplifie donc pas mal. Incidemment, cette réduction est standard pour tous les problèmes à deux corps : on décompose le mouvement en deux mouvements indépendants l'un de l'autre : celui du centre de gravité et le déplacement relatif des deux corps.

b. On note que la fonction

$$\begin{aligned} z(t) &= \frac{1}{m_G + m_S} \left(\int_0^t (t-s)g(s) ds \right) a + t\dot{y}_I(0) + y_I(0) \\ &= \frac{1}{m_G + m_S} \left(t \int_0^t g(s) ds - \int_0^t sg(s) ds \right) a + t\dot{y}_I(0) + y_I(0) \end{aligned}$$

est de classe C^2 avec

$$\dot{z}(t) = \frac{1}{m_G + m_S} \left(\int_0^t g(s) ds + tg(t) - tg(t) \right) a + \dot{y}_I(0) = \frac{1}{m_G + m_S} \left(\int_0^t g(s) ds \right) a + \dot{y}_I(0)$$

puis

$$\ddot{z}(t) = \frac{1}{m_G + m_S} g(t)a$$

avec

$$z(0) = y_I(0), \quad \dot{z}(0) = \dot{y}_I(0).$$

2. Une version un peu plus sophistiquée que celles du cours.

C'est donc bien que $y_I(t) = z(t)$ pour tout t . On pouvait aussi partir dans l'autre sens de \ddot{y} et l'intégrer deux fois. Pour passer de l'intégrale double à l'intégrale simple ci-dessus, on fait alors soit une intégration par parties, soit on utilise le théorème de Fubini. On pouvait aussi, mais c'est plus exotique, reconnaître une formule de Taylor avec reste intégral.

c. Pour $t \geq t_0$, il vient de la formule précédente que

$$y_I(t) = \frac{1}{m_G + m_S} \left(\int_0^{t_0} (t-s)g(s) ds \right) a + t\dot{y}_I(0) + y_I(0).$$

Le point

$$\frac{1}{m_G + m_S} \left(\int_0^{t_0} (t-s)g(s) ds \right) a + y_I(0)$$

se trouve à tout instant sur la droite qui joint la position initiale du centre de gravité des astronautes à la station spatiale. La position de y_I s'obtient en lui ajoutant le vecteur $t\dot{y}_I(0)$. Si $\dot{y}_I(0) = \frac{1}{m_G + m_S}(m_G v_{G,0} + m_S v_{S,0})$ n'est pas colinéaire à a , c'est-à-dire si la vitesse initiale de leur centre de gravité ne pointe pas exactement vers la station, ils n'y arriveront donc jamais. Or cette vitesse est un petit peu aléatoire, vu les circonstances dramatiques. Notons que pour $t \geq t_0$, la fonction y_I devient affine par rapport à t , c'est-à-dire que le mouvement du centre de gravité des deux astronautes devient rectiligne uniforme, ce qui est normal puisqu'alors $\ddot{y}_I(t) = 0$.

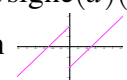
Notons que même sans panne de carburant, ils n'y arriveront pas plus pour la même raison. Dans la vraie vie naturellement, le vecteur a n'est pas fixé, mais varie en fonction des manipulations de G sur son propulseur. Il n'en reste pas moins qu'il est totalement invraisemblable que G , qui est lui aussi secoué par le câble, puisse être capable de planifier rien qu'au feeling une trajectoire les amenant à bon port.

Correction problème, partie III

a. La fonction second membre $(t, d) \mapsto g(t) + f_1(d)$ est k -lipschitzienne par rapport à d sur $\mathbb{R}_+ \times \mathbb{R}$. En effet, f_1 est continue affine par morceaux, valant $f_1(d) = k(d+1)$ pour $d \leq -1$, $f_1(d) = 0$ pour $-1 \leq d \leq 1$ et $f_1(d) = k(d-1)$ pour $d \geq 1$, voir Figure 2. Une disjonction des cas demanderait donc de distinguer 6 cas, mais en fait on peut faire mieux puisque l'on sait déjà que $d \mapsto (|d| - 1)_+$ est 1-lipschitzienne. Le seul cas restant alors à traiter est celui où $d_1 \leq 0 \leq d_2$. Comme on a alors $kd_1 \leq f_1(d_1) \leq 0 \leq f_1(d_2) \leq kd_2$, il vient

$$|f_1(d_1) - f_1(d_2)| = f_1(d_2) - f_1(d_1) \leq kd_2 - kd_1 = k|d_2 - d_1|.$$

Donc on a existence globale et unicité sur tout intervalle compact $[0, T]$, donc existence globale sur \mathbb{R}_+ , cf. corrigé du partiel pour ce prolongement de la solution à \mathbb{R}_+ .

De nombreuses copies ont oublié la partie positive, considérant que $f_1(d) = k \operatorname{signe}(d)(|d| - 1)$. Mais cette fonction est discontinue en 0, comme on le voit sur un petit dessin  ou bien

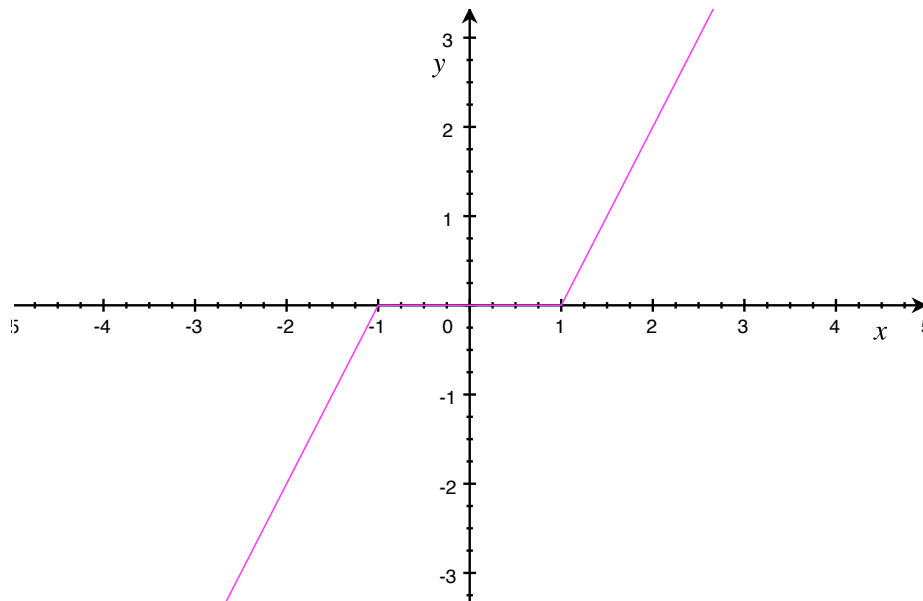


FIGURE 2 – Le graphe de la fonction f_1 avec $k = 2$.

en écrivant ses valeurs. Donc on ne peut certainement pas la déclarer lipschitzienne ex abrupto et appliquer un théorème de Cauchy-Lipschitz sous quelque version que ce soit. En fait, un problème de Cauchy du genre de celui du **b** peut très bien ne même pas avoir de solution avec ce second membre discontinu. En effet, prenant par exemple $k = 1$, $g(t) = 1$ et $d_0 = -1$, s'il en avait une, alors celle-ci serait initialement strictement négative, donc satisferait $d'(t) = 1 + d(t) + 1 = 2 + d(t)$ sur un certain intervalle $[0, \tau]$, c'est-à-dire $d(t) = e^t - 2$ sur $[0, \ln 2]$. En particulier, $d'(t) = e^t$ sur cet intervalle a pour limite 2 quand $t \rightarrow \ln 2^-$. Mais en ce point, on a $d(t) = 0$, par conséquent par l'EDO, $d'(\ln 2) = 1 + d(\ln 2) - 1 = 0$, ce qui est contradictoire avec ce qui précède. Il n'y a donc pas de solution de ce problème de Cauchy sur un intervalle plus grand que $[0, \ln 2]$. La difficulté n'est pas une explosion du type solution maximale non globale, mais que l'EDO à second membre discontinu impose une non dérivabilité, ce qui la contredit elle-même.

b. Par le théorème de Cauchy-Lipschitz global, on sait qu'il existe une solution et une seule, donc si on en trouve une par tout moyen légal ou illégal, mais de préférence légal, c'est que c'est la bonne. En particulier, si $d(0) = 0$ alors par continuité, la valeur absolue de d va rester inférieure à 1 sur un intervalle de temps $[0, \tau_1]$ a priori inconnu, sur lequel on a donc $f_1(d) = 0$ et d va y vérifier $d'(t) = 1$, soit $d(t) = t$. D'après ce qui vient d'être dit, cette solution est la bonne au moins jusqu'à $t = \tau_1 = 1$. Au delà de $t = 1$, elle n'est plus bonne puisqu'elle ne vérifie plus l'équation différentielle car $k \neq 0$ et $(t - 1)_+ = t - 1 > 0$. Comme $d'(1) = 1$, on va avoir $d(t) > 1$ sur un intervalle $]1, \tau_2[$ a priori inconnu. Sur cette intervalle, on a donc $d'(t) = 1 - k(d(t) - 1) = -kd(t) + 1 + k$. C'est une EDO linéaire à coefficient constant avec une solution particulière évidente $d_p(t) = 1 + \frac{1}{k}$, d'où

la solution générale $d_g(t) = Ce^{-kt} + 1 + \frac{1}{k}$. La condition initiale $d(1) = 1$ implique que $C = -\frac{e^k}{k}$, d'où $d(t) = -\frac{e^k}{k}e^{-kt} + 1 + \frac{1}{k}$ pour la solution du problème de Cauchy. Or il est évident que cette formule définit une fonction croissante sur $[1, +\infty[$, quel que soit le signe de k d'ailleurs, qui reste donc supérieure à 1 pour tout $t \geq 1$. C'est donc la bonne sur $[1, +\infty[$. Voir Figure 3.

L'oubli très courant de la partie positive dans la fonction second membre était ici aussi rédhibitoire, vu ce qui a été dit plus haut.

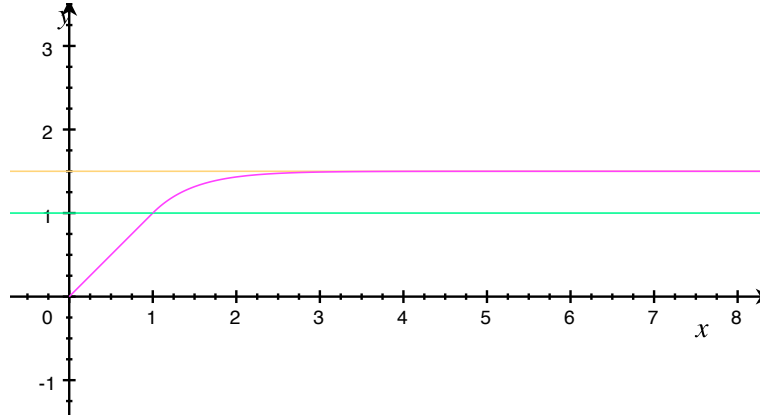


FIGURE 3 – Le graphe de la solution d avec $k = 2$. La valeur $d = 1$ est en vert, l'asymptote horizontale $d = 1 + \frac{1}{k}$ en orange.

c. Pour la méthode d'Euler explicite, tant que $|d_{e,n}| \leq 1$, on a $f_1(d_{e,n}) = 0$. Il vient donc $d_{e,n} = nh$ tant que $(n-1)h \leq 1$, c'est-à-dire $n \leq N+1$.

Montrons cela un peu plus proprement par récurrence. L'assertion à montrer est que pour tout $n \leq N+1$, $d_{e,n} = nh$. Elle est vérifiée par définition pour $n = 0$. Supposons donc la vraie pour un n tel que $n+1 \leq N+1$. Comme $n \leq N$, on a donc $0 \leq d_{e,n} = nh \leq 1$, ce qui implique bien que $f_1(d_{e,n}) = 0$. Par conséquent, $d_{e,n+1} = d_{e,n} + h(1 + f_1(d_{e,n})) = nh + h = (n+1)h$.

Pour la méthode d'Euler implicite, on a également $f_1(d_{i,n+1}) = 0$ tant que $|d_{i,n+1}| \leq 1$. Donc de même $d_{i,n} = hn$ tant que $n \leq N$ (il faut aussi écrire le raisonnement par récurrence, en utilisant le fait que h est supposé assez petit pour que l'équation implicite définissant $d_{i,n+1}$ ait une solution et une seule, donc si on en trouve une, c'est que c'est la bonne).

On a enfin $d(t_n) = t_n = nh$ tant que $t_n \leq 1$, c'est-à-dire $n \leq N$.

Même si l'on n'avait pas calculé la solution exacte, ce n'était quand même pas très difficile de regarder ce que pouvait bien donner $d_{e,0}$, puis $d_{e,1}$, puis $d_{e,2}$ et à partir de là de se dire, tiens, tiens, mais est-ce que pour n ça ne vaudrait pas...

d. On a déjà vu que $d_{e,N+1} = (N+1)h = 1 + h$. Pour la méthode d'Euler implicite, comme $d_{i,N+1} > 1$, on doit résoudre l'équation $d_{i,N+1} = d_{i,N} + h(1 + f_1(d_{i,N+1}))$, soit en d'autres termes $d_{i,N+1} = 1 + h(1 - k(d_{i,N+1} - 1))$, équation du premier degré dont l'unique solution est $d_{i,N+1} = 1 + \frac{h}{1+hk}$.

On a $d(t_{N+1}) = -\frac{e^k}{k} e^{-k(1+h)} + 1 + \frac{1}{k} = 1 + \frac{1-e^{-kh}}{k}$. Raisonnant grossièrement, on a donc $d_{e,N+1} = 1 + h$, $d_{i,N+1} = 1 + h - h^2 k + \dots$ et $d(t_{N+1}) = 1 + h - \frac{h^2 k}{2} + \dots$. Si $k > 0$ est grand, il n'est donc pas déraisonnable de penser que la méthode implicite est mieux adaptée, voir Figure 4.

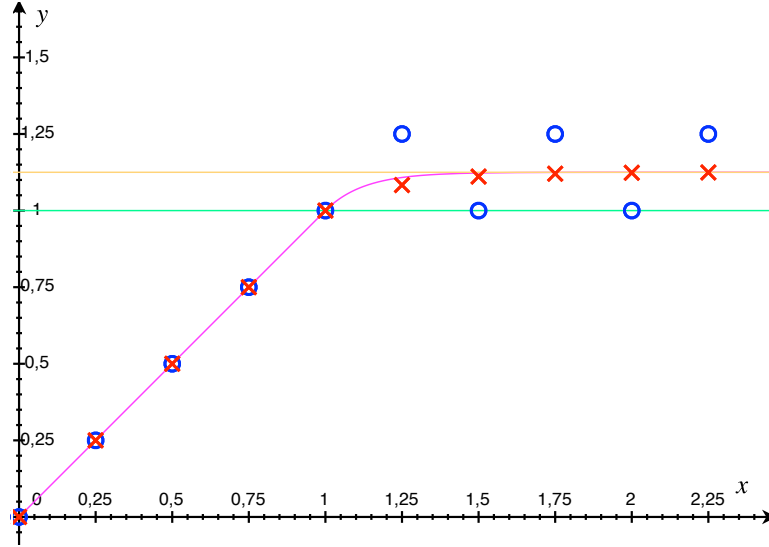


FIGURE 4 – Comparaison Euler explicite (ronds bleus), Euler implicite (croix rouges) avec $k = 8$, $h = \frac{1}{4}$.

e. La fonction f_1 n'est pas dérivable en $d = \pm 1$, voir Figure 2. Si g est régulière et si d prend l'une de ces deux valeurs à un certain instant t , il s'ensuit que d' n'est pas dérivable à cet instant. On a donc au maximum d de classe C^1 a priori. Du coup, il se peut qu'il ne serve à rien d'utiliser un schéma d'ordre élevé, car l'estimation d'erreur correspondante requiert que d soit dérivable suffisamment de fois. Même pour le schéma d'Euler explicite, on ne peut pas appliquer aveuglément l'estimation en $O(h)$ du cours, car d n'est pas C^2 . Par contre, il reste vrai que ce schéma, ou tout schéma consistant d'ailleurs, est convergent. Simplement, on n'a pas l'assurance que la vitesse de convergence corresponde à l'ordre du schéma.

f. On arrondit donc les angles en introduisant f_2 . On note toujours d la solution du problème de Cauchy

$$\begin{cases} d'(t) = g(t) + f_2(d(t)), \\ d(0) = d_0, \end{cases} \quad (3)$$

même si ce n'est pas le même d qu'avant. Comme f_2 est globalement lipschitzienne, on a une solution globale. Comme f_2 et g sont toutes deux de classe C^2 , et que d est de classe C^1 comme toute solution d'EDO, on voit que d' est de classe C^1 par dérivation des fonctions composées pour

le deuxième terme. C'est-à-dire que d est de classe C^2 . Mais si d est de classe C^2 , encore par dérivation des fonctions composées, on a que d' est de classe C^2 , c'est-à-dire d de classe C^3 . Ça s'arrête là puisqu'on n'a pas d'hypothèse de régularité supplémentaire sur g et f_2 .

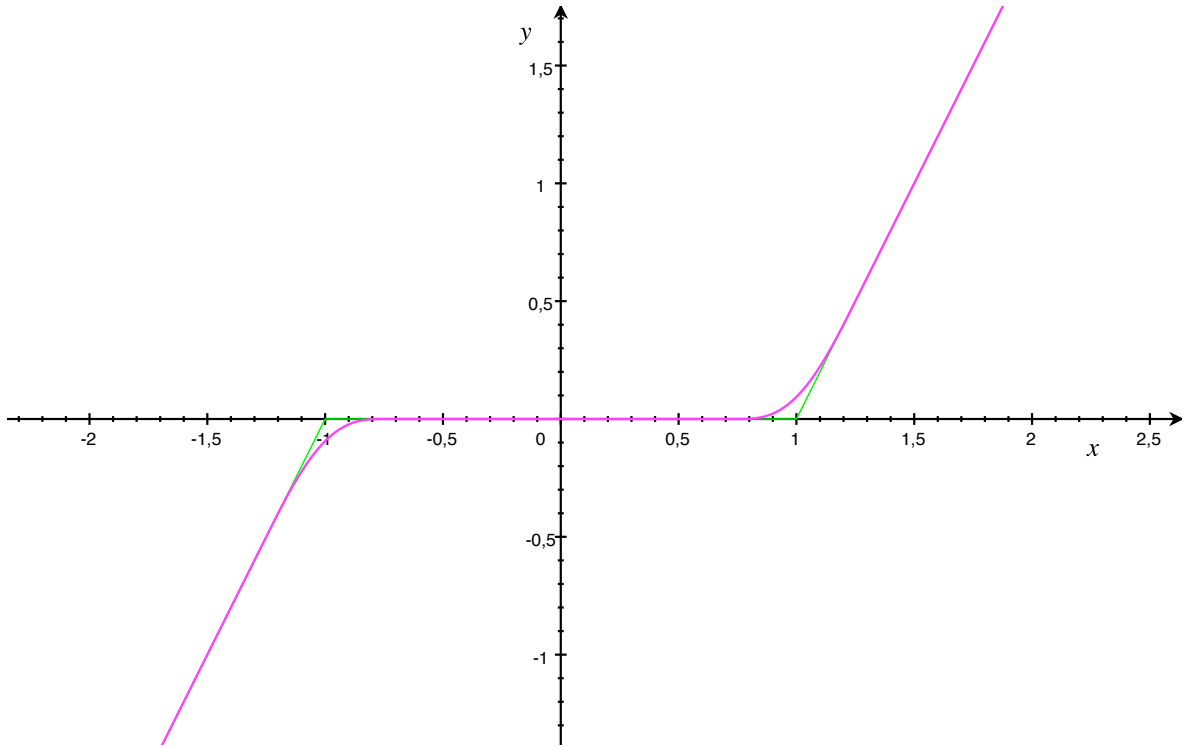


FIGURE 5 – Un joli arrondi possible de classe C^2 pour f_2 .

g. Chaque itération du schéma de Crank-Nicolson se présente comme un problème de point fixe $z = \varphi(z)$ avec

$$\varphi(z) = d_{CN,n} + \frac{h}{2}(g(t_n) + f_2(d_{CN,n}) + g(t_{n+1}) + f_2(z)),$$

application de l'espace métrique complet \mathbb{R} muni de sa distance usuelle dans lui-même. Comme dans le cours, il vient

$$\varphi(z_1) - \varphi(z_2) = \frac{h}{2}(f_2(z_1) - f_2(z_2)),$$

d'où

$$|\varphi(z_1) - \varphi(z_2)| = \frac{h}{2}|f_2(z_1) - f_2(z_2)| \leq \frac{hK}{2}|z_1 - z_2|.$$

Il s'ensuit que si $\frac{hK}{2} < 1$, c'est-à-dire si $h < \frac{2}{K}$, l'application φ est strictement contractante et admet donc un unique point fixe.

h. On se place sur un intervalle compact $[0, T]$. Avant de faire des développements de Taylor dans l'erreur de consistance, il vaut mieux la réécrire sous la forme

$$\varepsilon_n = d(t_{n+1}) - d(t_n) - \frac{h}{2}(d'(t_n) + d'(t_{n+1})).$$

Comme d est de classe C^3 , on a d'une part

$$d(t_{n+1}) - d(t_n) = hd'(t_n) + \frac{h^2}{2}d''(t_n) + \frac{h^3}{6}d'''(s_n)$$

pour un certain $s_n \in]t_n, t_{n+1}[$. D'un autre côté, on a aussi

$$d'(t_{n+1}) = d'(t_n) + hd''(t_n) + \frac{h^2}{2}d'''(u_n)$$

pour un certain $u_n \in]t_n, t_{n+1}[$. On s'aperçoit donc que

$$\varepsilon_n = h^3 \left(\frac{1}{6}d'''(s_n) - \frac{1}{4}d'''(u_n) \right).$$

Par conséquent, pour tout $n \leq T/h$,

$$|\varepsilon_n| \leq \left(\frac{5}{12} \max_{s \in [0, T]} |d'''(s)| \right) h^3,$$

et le schéma est d'ordre 2.

On pouvait aussi garder l'expression avec g et f_2 en développant (correctement) tous les termes correspondants, mais c'était plus de travail. Les développements avec des restes en $O(h^p)$ sont acceptables également parce que l'on sait (ou en tout cas, il faut en avoir conscience) que les constantes cachées dans ces O sont uniformes par rapport à n et h , sur un intervalle compact $[0, T]$ donné, comme on le souhaite pour pouvoir conclure dans ces questions d'ordre.

i. Pour raccourcir l'écriture, notons simplement $d_n = d_{CN,n}$, de sorte que

$$d_{n+1} = d_n + \frac{h}{2}(g(t_n) + f_2(d_n) + g(t_{n+1}) + f_2(d_{n+1})), \quad d_0 = d_0.$$

On introduit la récurrence perturbée

$$\bar{d}_{n+1} = \bar{d}_n + \frac{h}{2}(g(t_n) + f_2(\bar{d}_n) + g(t_{n+1}) + f_2(\bar{d}_{n+1})) + \eta_{n+1}, \quad \bar{d}_0 = d_0 + \eta_0.$$

Prenant la différence des deux, il vient

$$\bar{d}_{n+1} - d_{n+1} = \bar{d}_n - d_n + \frac{h}{2}(f_2(\bar{d}_n) - f_2(d_n) + f_2(\bar{d}_{n+1}) - f_2(d_{n+1})) + \eta_{n+1}, \quad \bar{d}_0 - d_0 = \eta_0,$$

d'où

$$|\bar{d}_{n+1} - d_{n+1}| \leq |\bar{d}_n - d_n| + \frac{Kh}{2} (|\bar{d}_n - d_n| + |\bar{d}_{n+1} - d_{n+1}|) + |\eta_{n+1}|, \quad |\bar{d}_0 - d_0| = |\eta_0|,$$

d'où, dès que $h < \frac{2}{K}$,

$$|\bar{d}_{n+1} - d_{n+1}| \leq \frac{1 + \frac{Kh}{2}}{1 - \frac{Kh}{2}} |\bar{d}_n - d_n| + \frac{1}{1 - \frac{Kh}{2}} |\eta_{n+1}|, \quad |\bar{d}_0 - d_0| = |\eta_0|,$$

et l'on conclut avec le lemme de Grönwall discret dès que $h \leq h_0 < \frac{2}{K}$ en écrivant

$$\frac{1 + \frac{Kh}{2}}{1 - \frac{Kh}{2}} = 1 + \frac{Kh}{1 - \frac{Kh}{2}} \leq 1 + \frac{Kh}{1 - \frac{Kh_0}{2}} \quad \text{et} \quad \frac{1}{1 - \frac{Kh}{2}} \leq \frac{1}{1 - \frac{Kh_0}{2}}.$$

On en déduit de façon standard l'estimation d'erreur en $O(h^2)$ en prenant $\bar{d}_n = d(t_n)$.

Par curiosité, regardons ce qui se passe si l'on applique le schéma de Crank-Nicolson à l'EDO initiale avec f_1 non régulière. A priori, ce n'est pas une bonne idée. Néanmoins, le même calcul qu'en **d** donne $d_{N+1} = 1 + \frac{h}{1 + \frac{h}{2}} = 1 + h - \frac{h^2}{2} + \dots$, une valeur qui semble nettement meilleure que celle donnée par le schéma d'Euler implicite, voir Figure 6. Donc finalement, les choses ne sont peut-être pas aussi simples qu'elles en ont l'air... qu'en pensez-vous ?

Ci-dessous dans la Figure 7, un exemple de trajectoire du problème de départ (1), vue sous plusieurs angles, G est en bleu, S en rouge, en vert la direction de la station spatiale internationale (zoomer dans le pdf pour y voir plus clair). Calcul fait avec ode de scilab sans se casser la tête outre mesure.

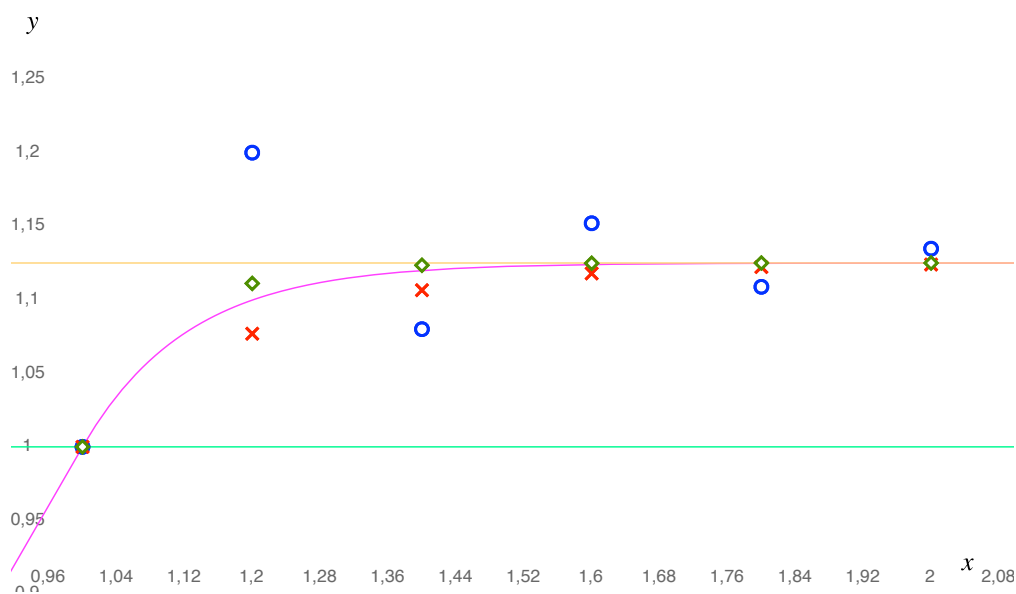


FIGURE 6 – Presque le même calcul que dans la Figure 4 mais avec $h = \frac{1}{5}$, Euler explicite = cercles bleus, Euler implicite = croix rouges, Crank-Nicolson = losanges verts. Pas si mal, Crank-Nicolson finalement.

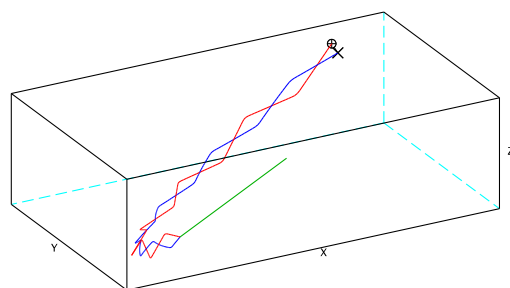
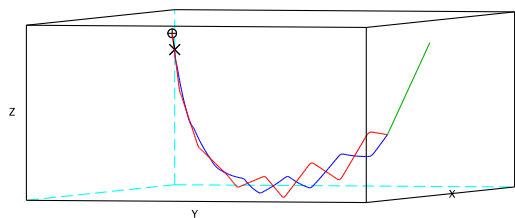
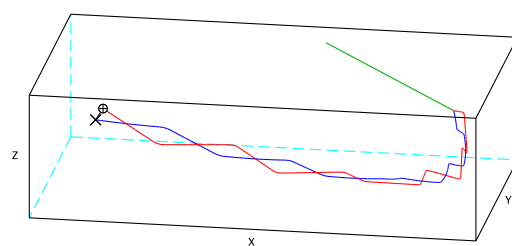
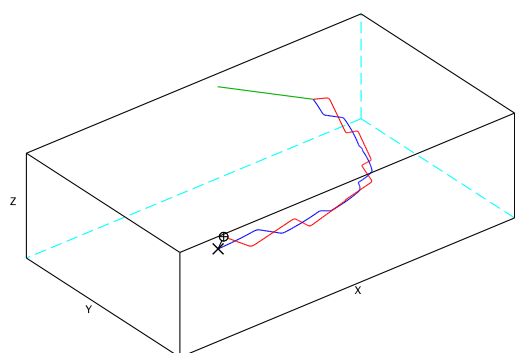


FIGURE 7 – Dans l'espace, on ne vous entendra pas crier.