

## Schémas de résolution à un pas

On s'intéresse ici à la résolution du problème de Cauchy suivant:

$$\begin{aligned} y'(t) &= f(t, y) \\ y(t_0) &= y_0 \end{aligned} \quad (\text{où } y_0 \in \mathbb{R}^n, \text{ donné comme condition initiale})$$

où  $y$  est une fonction suffisamment régulière, à valeurs dans  $\mathbb{R}^n$ , que nous cherchons à déterminer sur un certain intervalle,  $t$  un réel (symbolisant souvent le temps), et enfin  $f$  une fonction de  $\mathbb{R} \times \mathbb{R}^n$  dans  $\mathbb{R}^n$ .

Le domaine des équations différentielles est complexe, et nous ne savons, à l'heure actuelle, résoudre exactement que quelques unes d'entre elles. Néanmoins, il s'avère suffisant pour nombre d'applications d'approcher la solution théorique grâce à des méthodes itératives.

Le point de vue abordé est le suivant : on cherche à approcher la solution théorique du problème sur un intervalle donné du type  $[t_o; t_o + T]$ , et on se donne une subdivision  $t_o, t_1, t_2, \dots, t_N = t_o + T$  de cet intervalle. On pose, pour  $0 \leq n \leq N - 1$ ,  $h_n = t_{n+1} - t_n$ .

A partir d'une condition initiale  $y(t_0) = y_0$ , on se propose de construire une suite finie de valeurs  $y_n = y(t_n)$  qui vont respectivement approcher les valeurs prises par une solution théorique  $z$ , aux temps  $t_n$ .

Dans la suite, je prends la fonction  $y$  cherchée à valeurs dans  $\mathbb{R}$  et je prendrai parfois une subdivision régulière de pas  $h$  pour simplifier les démarches.

**Définition 1** (*méthodes à un pas*)

On appelle méthode à 1 pas une méthode permettant de calculer  $y_{n+1}$  à partir de la seule valeur précédente  $y_n$ .

Les méthodes à 1 pas peuvent s'écrire sous la forme suivante :

$y_{n+1} = y_n + h_n \varphi(t_n, y_n, h_n)$  avec toujours  $0 \leq n \leq N$  où  $\varphi: [t_0; t_0 + T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  est une fonction qu'on suppose continue, qui ne sera définie en réalité sur une partie plus petite du précédent domaine, liée aux besoins de l'équation différentielle.

**Remarque:** Nous connaissons déjà une méthode à un pas: il s'agit de la méthode d'Euler dite explicite, où  $\varphi(t, y, h) = f(t, y)$  est. La suite  $(y_n)$  ainsi définie par la relation  $y_{n+1} = y_n + h_n f(t_n, y_n)$ .

Elle repose sur la définition de la dérivée suivante :  $\lim(h_n \rightarrow 0) \frac{y_{n+1} - y_n}{h_n} = y'(t_n) = f(t_n, y_n)$

...valide ici si tant est bien sûr que la fonction  $f$  soit « suffisamment régulière » comme nous le disions dans l'introduction. On peut donc percevoir le fait que la fonction  $f$  doive remplir certaines conditions pour permettre l'utilisation de schémas à un pas comme celui présenté...

**Définition 2** (*erreur de consistance*)

L'erreur de consistance  $e_n$  relative à une solution exacte  $z$  est l'erreur

$$e_n = z(t_{n+1}) - y_{n+1} \text{ pour } 0 \leq n \leq N, \text{ comme d'habitude, en prenant } y_n = z(t_n)$$

Cette erreur ne met qu'en jeu une unique étape de l'algorithme : on suppose que le dernier terme connu de notre suite  $(y_n)$  est très exactement la valeur de  $z(t_n)$ , et on veut voir quelle erreur notre méthode commet en produisant ainsi le terme  $y_{n+1}$ , censé approcher  $z(t_{n+1})$ .

Par définition, l'erreur de consistance peut se ré-écrire  $e_n = z(t_{n+1}) - z(t_n) - h_n \varphi(t_n, z(t_n), h_n)$

Il est nécessaire de s'interroger sur cet écart car nous avons pour but d'approcher une solution exacte avec tout de même une certaine précision. De plus, nous construisons, itérativement, une nouvelle valeur approchée  $y_{n+1}$  à partir d'une autre valeur approchée  $y_n$ : autrement dit, plus nous avançons dans l'algorithme, plus les erreurs s'accumulent: on peut se demander si l'approximation ne va pas être de plus en plus mauvaise ou aléatoire (suivant que les  $e_n$  se compensent ou aillent dans le même sens...) au bout d'un certain nombre d'étapes...

On peut s'interroger également sur l'erreur de consistance maximale commise de l'approximation sur l'intervalle de départ (ou un de ses sous-intervalles)...

**Définition 3** (*consistance d'une méthode*)

On dit qu'une méthode est consistante si, lorsque les  $h_n$  tendent vers 0, la quantité  $\sum(|e_n|)$  tend aussi vers 0.

En gros, une méthode est consistante si en théorie, la suite des  $y_n$  tend bien vers ce qu'on veut, à savoir les  $z(t_n)$ . C'est à dire que nous avons bien visé de manière à atteindre la cible, en théorie : notre méthode est adaptée.

**Remarque :** Si les  $h_n$  tendent vers 0, le nombre d'itérations de l'algorithme tend vers l'infini.

Un autre paramètre à surveiller est ce qu'on appelle la stabilité de l'algorithme. Celui-ci pointe tout simplement du doigt les capacités calculatoires des machines utilisées: le fait qu'à chaque étape, des erreurs d'arrondis naitront (et qu'en plus elles se cumuleront lors de la progression de l'algorithme !).

Même si en théorie un schéma est consistant, si son fonctionnement est incapable d'entacher l'accumulation / la propagation des erreurs d'arrondis, on risque tout bonnement de s'écarter plus ou moins rapidement du résultat théorique qu'on vise (qui est peut être atteint en théorie si le schéma est consistant) et d'atterrir à côté de la cible.

#### Définition 4 (stabilité d'une méthode)

On dit qu'une méthode est stable s'il existe une constante  $S \geq 0$  (appelée constante de stabilité) telle que pour toutes suites  $(y_n)$  et  $(\tilde{y}_n)$  définies par

$$\begin{aligned} y_{n+1} &= y_n + h_n \varphi(t_n, y_n, h_n) \\ \text{et } \tilde{y}_{n+1} &= \tilde{y}_n + h_n \varphi(t_n, \tilde{y}_n, h_n) + \varepsilon_n \quad \text{on ait } \max(|\tilde{y}_n - y_n|) \leq S(|\tilde{y}_0 - y_0| + \sum |\varepsilon_n|) \text{ pour } 0 \leq n \leq N \end{aligned}$$

C'est à dire que, pour une même méthode donnée, si une petite perturbation (ou plusieurs) intervient dans le calcul d'une suite d'approximation par rapport à une autre, l'écart maximal observé, au final, reste contrôlé.

Ce n'est pas parce qu'à un moment donné « on s'écarter un peu » qu'on va atterrir totalement à côté de la cible au final!

Par exemple, si ces deux suites partent de la même condition initiale mais que la première permet un calcul avec plus de chiffres significatifs que la seconde (peut-être du aux capacités machines), l'écart final restera néanmoins contrôlé: c'est à dire que la seconde suite restera relativement de la première, malgré les plus grosses erreurs d'arrondies qui l'entachent: la propagation de ces erreurs d'arrondis est limitée.

Autre exemple, la condition initiale peut faire l'objet d'une valeur approchée, obtenue par une quelconque méthode numérique, légèrement différente pour nos deux suites.

Si la méthode à un pas employée est stable, alors cette « perturbation » initiale aura des répercussions limitées sur la différence qu'on pourra constater au final, une fois nos deux suites créées.

#### Définition 5 (convergence d'une méthode)

On dit que la méthode est convergente si pour toute solution exacte  $z$ , la suite des  $y_n$  comme définie plus haut vérifie  $\max|(z(t_n) - y_n)| \rightarrow 0$  lorsque  $y_0 \rightarrow z(t_0)$  et  $h_{\max} \rightarrow 0$

Cette quantité est l'erreur globale (de la suite  $y_n$  par rapport à la solution exacte). C'est l'erreur qui nous intéresse le plus puisque c'est l'erreur maximale commise entre le phénomène réel qu'on essaie de décrire et la modélisation mathématique (ainsi que son application) qu'on a mis en oeuvre pour l'approximer.

En bref, si le pas d'itération tend vers 0 et qu'on à la bonne valeur de départ, en théorie on tend vers la solution exacte.

#### Théorème 1 (consistance + stabilité = convergence)

Si une méthode à 1 pas est consistante et stable, alors elle converge vers la solution théorique, au sens de la précédente définition.

Reformulé ainsi le théorème semble logique: si l'on « vise bien » (consistance) et qu'on ne dévie pas de sa trajectoire à cause de petites perturbations (stabilité) on arrive bien à atteindre la cible au final.

## II Quelques résultats

Les définitions précédentes ne sont pas nécessairement faciles à vérifier. Peut-être avons nous des conditions un peu plus pratiques pour témoigner de la stabilité ou de la consistance d'une méthode à 1 pas...

**Théorème 2** (*condition nécessaire et suffisante de consistance d'un schéma*)

La méthode à 1 pas faisant intervenir la fonction  $\varphi$  est consistante si et seulement si elle vérifie la condition suivante :  $\forall (t, y) \in [t_0, t_0 + T] \times \mathbb{R}, \varphi(t, y, 0) = f(t, y)$

**Remarque :** Une conclusion triviale de ce théorème est le fait que la méthode d'Euler explicite est consistante.

**Théorème 3** (*condition suffisante de stabilité*)

Pour que la méthode soit stable, il suffit que la fonction  $\varphi$  vérifie une condition de Lipschitz en  $y$ , c'est à dire qu'il existe une constante  $K$  positive telle que :

$$\forall t \in [t_0, t_0 + T], \forall (y_1, y_2) \in \mathbb{R}^2, \forall h \in \mathbb{R} \text{ on ait } |(\varphi(t, y_1, h) - \varphi(t, y_2, h))| \leq K \times |y_1 - y_2|.$$

Si tel est le cas, on peut prendre pour constante de stabilité la valeur  $S = e^{(KT)}$

(J'avoue ne pas comprendre cette décision, et je suppose que cette majoration a été l'objet d'une étude plus fine, qui a permis de déterminer une constante optimale)

**Remarque:** Dans le cas de la méthode d'Euler, puisque  $\varphi(t, y, h) = f(t, y)$ , on s'aperçoit que si  $f$  est Lipschitz en  $y$  alors  $\varphi$  l'est aussi. Comme la méthode d'Euler explicite est consistante, on vient donc de montrer une condition suffisante de convergence pour celle-ci : il suffit que  $f$  soit Lipschitzienne en  $y$ .

On rappelle que pour le peu que  $f$  soit dérivable une fois sur l'intervalle qui nous intéresse, cette condition est immédiatement démontrée par l'utilisation du théorème des accroissements finis.

## II Méthodes de Taylor et de Runge-Kutta

**Définition 6** (*méthode d'ordre  $p$* )

On dit qu'une méthode est consistante d'ordre  $p \leq 1$  s'il existe une constante  $C$  positive telle que la majoration suivante soit vérifiée:  $|(e_n)| \leq Ch_n^{p+1}$ ,  $\forall n$ ,  $0 \leq n \leq N$ .

**Remarque:** Bien entendu, si une méthode est consistante d'ordre  $p$ , elle est consistante (puisque les  $h_n$  tendent vers 0, la série de terme général  $|(e_n)|$  va converger normalement, car  $\sum Ch_n^{p+1}$  va converger).

## 1) Méthode de Taylor d'ordre $p$

Au début de ce document, je disais que nous connaissions déjà un schéma à un pas, qui était celui de la méthode d'Euler, qui reposait sur la définition de la dérivée...il ne s'agit là en fait que d'un point de départ à un ensemble plus vaste de méthodes. En effet, si tant est que la fonction  $f$  soit de classe  $C^p$ , la solution exacte  $z$  est de classe  $C^{p+1}$  et l'on peut effectuer un développement de Taylor à l'ordre  $p$ .

Il vient  $z(t_n + h_n) = z(t_n) + \sum_{k=1}^p \frac{1}{k!} h_n^k f^{[k-1]}(t_n, z(t_n)) + o(h_n^p)$ , la somme étant sur  $k$  allant de 1 à  $p$ .

On crée donc une suite  $y_{n+1} = y_n + \sum_{k=1}^p \frac{1}{k!} h_n^k f^{[k-1]}(t_n, y_n)$  qui définit la méthode de Taylor d'ordre  $p$ , et on identifie

$$\varphi(t, y, h) = \sum_{k=1}^p \frac{1}{k!} h^{k-1} f^{[k-1]}(t, y).$$

On peut vérifier que l'erreur de consistance est de l'ordre de  $h_n^{p+1}$ : ces méthodes sont donc consistantes d'ordre  $p$ . Comme  $f$  est de classe  $C^p$ , le TAF permet d'appliquer le théorème 3 : ces méthodes sont donc convergentes.

**Remarque :** L'expression des dérivées  $k$ -ièmes de  $\varphi$  en fonction de  $f$  s'avère rapidement douloureuse, bien que possible. Il suffit d'effectuer des différenciations successives de  $f$  en fonction de  $t$ .

Pour avoir essayé, on a vraiment pas envie d'aller voir que loin que la dérivée troisième, on imagine bien que les calculs machines vont rapidement devenir complexes, sans parler d'éventuels problèmes concernant l'évaluation de ces dérivées au point donné...

De plus, l'application correcte de cette méthode nécessite tout de même que  $f$  soit de classe  $C^p$ , ce qui n'est pas gagné.

Pour conclure, une remarque intéressante pour la suite: si la fonction  $f$  est de classe  $C^p$  alors on peut montrer avec un développement de Taylor à l'ordre  $p+1$  sur  $z$  que l'erreur de consistance  $e_n$  vérifie

$$e_n = \sum_{l=1}^p \frac{1}{l!} h_n^{l+1} \left( \frac{1}{l+1} f^{[l]}(t_n, y_n) - \frac{\partial^l \varphi}{\partial h^l}(t_n, y_n, 0) \right) + o(h_n^{p+1}) \text{ pour } l \text{ allant de } 0 \text{ à } n \text{ bien sûr.}$$

On en déduit donc le résultat suivant, découlant de la définition de l'ordre d'une méthode donnée plus haut :

**Proposition:** Si  $f$  est de classe  $C^p$ , la méthode à un pas associée est d'ordre  $p$  (au moins) si et seulement si  $\varphi$  est telle que  $\frac{1}{l+1} f^{[l]}(t, y) = \frac{\partial^l \varphi}{\partial h^l}(t, y, 0)$  pour  $0 \leq l \leq p-1$ .

( ce qui laisse bien  $e_n = o(h_n^{p+1})$  )

## 2) Méthodes de Runge-Kutta

Je ne vois pas l'intérêt de plagier le livre de Mr Demailly de la page 237 à la page 241 (à la louche). Je n'ai rien à ajouter.

Dans les grandes lignes, l'idée de ces méthodes est de calculer par récurrence les couples  $(t_n, y_n)$  en utilisant des évaluations de  $f$  et de ses pentes à des points intermédiaires (c'est à dire entre  $t_n$  et  $t_{n+1}$ , au sens large) pour approcher une intégrale, obtenue en intégrant formellement l'équation différentielle de départ.

On en connaît déjà une: celle d'Euler (explicite) qui utilise  $y_n$  et la pente en  $t_n$  correspondante (obtenue grâce à l'équation différentielle). C'est en réalité la méthode de Runge-Kutta d'ordre 1.

A partir de relations sur des coefficients (définis par les points intermédiaires et une méthode d'intégration théorique, d'ordre 0 -convergeant vers l'intégrale donc-) on produit tout un tas de méthodes à un pas qui coïncident avec nombre de méthodes d'intégration connues : ce sont les méthodes de Runge-Kutta.

On y retrouve la méthode d'Euler (explicite) [RK1], la méthode du point milieu et la méthode dite de Heûn [RK2], la méthode de Simpson [RK4] , et, - j'en fais le pari sans avoir vu ça marqué quelque part – de manière plus générale la méthode de Newton-Côtes d'ordre N [RKN?].

(Toutes ces méthodes d'approximation d'une intégrale faisaient l'objet d'une leçon de CAPES l'an dernier)

Nos conditions sur les coefficients cherchés étant larges, nous disposons en réalité de beaucoup de méthodes selon la valeur choisie pour fixer le premier coefficient, etc.

On les représente souvent sous la forme d'un tableau, qui fait apparaître visuellement une matrice triangulaire inférieure.

La méthode la plus utilisée est la méthode RK4, pour son bon compromis entre précision et quantité calculatoire. J'ai pu la tester sur quelques exemples (en la faisant religieusement sous OpenOfficeCalc) sur un PDF que je vous ai déjà donné en pièce jointe dans le précédent mail.

Demailly démontre clairement (grâce à un petit lemme pratique qu'il démontre aussi) que les méthodes de Runge-Kutta sont stables si  $f$  est (encore une fois) lipschitzienne en  $y$ .

Il montre également que par définition des coefficients et de la méthode d'intégration, les conditions du théorème 2 sont immédiatement vérifiées, et que donc ces méthodes sont toujours consistantes.

Au final, si  $f$  est lipschitzienne, toutes les méthodes de Runge-Kutta convergent. Il ne reste qu'à choisir.

Concernant l'ordre de ces méthodes, on peut utiliser la proposition de la partie précédente pour le déterminer. C'est là qu'un choix de coefficient risque de s'avérer plus judicieux qu'un autre.

On montre que la méthode d'Euler explicite (RK1) est d'ordre 1, que la méthode du point milieu (issue possible de RK2) est d'ordre 2, et que la méthode RK4 liée à l'intégration de Simpson est d'ordre 4...

### 3) Méthode d'Euler implicite

Nous avons démontré que la méthode d'Euler (dite explicite), définissant la suite des  $y_n$  par  $y_{n+1} = y_n + h_n f(t_n, y_n)$  convergeait sous condition (par exemple celle vérifiée par le théorème 3).

Cette méthode, comme les autres que nous avons proposées jusqu'à maintenant, s'avère réellement efficace lorsque les pas de la subdivision sont très petits (donnant lieu du coup à un plus grand nombre d'étapes).

On en voit notamment les limites lorsque, pour  $c < 0$  on essaie de l'appliquer pour approcher  $e^c$  (au travers d'un problème de Cauchy, comme montré au début de ce document): si on fait l'essai, on voit que si  $h$  n'est « pas assez petit » alors la suite  $(y_n)$  engendrée ne coïncide même pas avec le comportement monotone de l'exponentielle.

On voit donc un aperçu des limites de la modélisation qu'on peut avoir avec certaines méthodes, qui nécessitent un pas « suffisamment petit » pour bien faire le boulot.

Cette fois-ci, considérons notre problème initial, avec  $y$  à valeurs dans  $\mathbb{R}^n$ . Nos méthodes construisent donc une suite de vecteurs de  $\mathbb{R}^n$  approchant la solution exacte théorique du problème en une subdivision de  $[t_o, t_o + T]$

Il est très fréquent, dans le domaine des réactions chimiques, de faire face à des équations différentielles du type  $y' = Ay + f(t)$  où  $A$  est une matrice carrée de taille  $n$  constante, et  $f$  une fonction de  $\mathbb{R}$  dans  $\mathbb{R}^n$ .

Le cours de licence assure qu'une solution exacte de ce problème s'écrit sous la forme  $z(t) = \sum c_k e^{\lambda_k t} z_k + \phi(t)$  ( $k$  allant de 1 à  $n$ ) où les  $\lambda_k$  sont les valeurs propres de  $A$  et  $z_k$  les vecteurs propres associés. Les  $c_k$  sont quand à eux des constantes à déterminer grâce à la condition initiale, et  $\phi(t)$  une solution particulière (qu'on arrive bien sûr pas à trouver de manière générale).

L'exponentielle a une croissance plutôt radicale. Selon les valeurs propres  $\lambda_k$ , certaines coordonnées des vecteurs formant risquent de converger à des vitesses très différentes, notamment si on voit arriver des valeurs propres très éloignées, ou encore des valeurs propres négatives (le comportement de l'exponentielle change alors totalement).

On a vu par l'exemple introductif de cette partie que le choix de  $h$  doit être « suffisamment petit » pour qu'on puisse être satisfait du résultat.

Mais les coordonnées de notre suite vectorielle convergeant à des vitesses très différentes, si on prend  $h$  « suffisamment » petit, on risque de mettre très longtemps pour voir converger les fameuses coordonnées « lentes » : le nombre d'étapes risque d'être extrêmement élevé, et les erreurs d'arrondi vont s'accumuler.

On est face à ce qu'on appelle un problème raide.

Pour faire face à ce genre de problèmes, on fait généralement appel à ce qu'on nomme les méthodes implicites. Une méthode implicite à un pas est une méthode qui va construire elle aussi une suite  $(y_n)$ , mais cette fois-ci  $y_{n+1}$  ne sera pas simplement obtenu grâce à une évaluation de  $f$  en un point donnée et de  $y_n$ :

$y_{n+1}$  va être défini en fonction de  $y_n$ ,  $f$ , et... lui-même !

C'est à dire que pour obtenir  $y_{n+1}$  on doit résoudre une équation :  $y_{n+1}$  est dit défini de manière implicite.

Prenons l'exemple de la méthode d'Euler dite « implicite » (ou encore « rétrograde » dans certaines ouvrages).

Pour établir la méthode d'Euler explicite, on se base sur le bien fondé de la définition de la dérivée:

$$\lim(h_n \rightarrow 0) \frac{y_{n+1} - y_n}{h_n} = y'(t_n) = f(t_n, y_n)$$

Dans cette définition,  $y_{n+1}$  et  $y_n$  jouent des rôles symétriques : on pourrait tout aussi bien poser

$$\lim(h_n \rightarrow 0) \frac{y_{n+1} - y_n}{h_n} = y'(t_{n+1}) = f(t_{n+1}, y_{n+1}) \text{ et on défini ainsi le schéma à un pas suivant:}$$

$$y_{n+1} = y_n + h_n f(t_{n+1}, y_{n+1})$$

La première question qui doit venir à l'esprit concerne certainement la bonne définition de cette suite, et ensuite celle des propriétés de la méthode.

Michèle Schatzman, dans «Analyse numérique, une approche mathématique» démontre que, d'une part, le schéma d'Euler implicite peut s'écrire sous la forme  $y_{n+1} = y_n + h_n \varphi(t_n, y_n, h_n)$  (ce qui montre la bonne définition de la suite  $(y_n)$ ) mais elle montre aussi au passage le fait que la méthode d'Euler implicite est consistante, et dite « inconditionnellement stable ». Donc, elle converge.

A un moment donné, elle déclare que puisque  $f$  vérifie les conditions de Cauchy-Lipschitz (qu'est-ce que ça veut dire? Que  $f$  est Lipschitzienne en  $y$  ? ) et s'en sert pour toute la suite de sa preuve, ce que je ne saisis pas.

Elle établit à un moment une fonction  $G(s, u, h)$  et montre que celle-ci est Lipschitzienne en  $u$ . A partir de cette fonction, elle construit une fonction  $F$  qui va permettre d'exprimer notre suite de la manière « classique » comme celle donnée dans la définition d'une méthode à un pas.

Elle montre facilement que  $F(t, u, 0) = f(t, u)$ , ce qui nous donne la consistance de la méthode.

Elle montre enfin, que, d'autre part,  $F$  est lipschitzienne en  $u$ , ce qui nous donne la stabilité, et donc la convergence. La méthode est d'ordre 1.

Néanmoins, j'avoue ne pas comprendre quelles hypothèses nous avons ici sur  $f$  et en quoi on a une « stabilité inconditionnelle » : si  $f$  est lipschitzienne, j'ai du mal à voir ce qu'on gagne sur Euler explicite... Est-ce qu'on entend par là que l'algorithme reste stable avec des valeurs de  $h$  plus grandes?

Par rapport aux méthodes explicites, les méthodes implicites souffrent du désavantage qu'est la résolution (exacte ou approchée!) de l'équation permettant d'obtenir  $y_{n+1}$  en fonction de  $y_n$  (qui lui est connu).

L'équation peut être non linéaire, et donc pourrait ne pas permettre une résolution exacte : on pourrait donc se retrouver à devoir utiliser -par exemple- le théorème du point fixe pour avoir  $y_{n+1}$ , à chaque étape...

Si en plus on travaille dans  $\mathbb{R}^n$ , on a possiblement une matrice à inverser, ce qui est très coûteux.

En revanche il semblerait que ces méthodes soient bien adaptées pour observer l'évolution d'événements « lents », et c'est pour cela qu'on fait généralement appel à elles pour les problèmes raides.