4.2 Analyse d'un schéma numérique pour la résolution d'une équation de réaction-diffusion (206, 218)

Dans ce développement bonus, qui peut mettre le jury à vos pieds pour les leçons 206 et 218 (j'y crois mdr), je détaille l'analyse d'un schéma numérique semi-implicite pour résoudre une équation de réaction-diffusion sur lequel je suis tombé en oral blanc. On s'intéresse à l'EDP suivante, d'inconnue $u \in \mathscr{C}^2(\mathbb{R}^+ \times [-L, L], \mathbb{R})$:

$$\left\{\begin{array}{rcll} \partial_t u(t,x) & = & \nu \partial^2_{xx} u(t,x) + g(u(t,x)) & \forall (t,x) \in \mathbb{R}^+ \times [-L,L], \\ \partial_x u(t,-L) & = & 0 & \forall t \in \mathbb{R}^+, \\ \partial_x u(t,L) & = & 0 & \forall t \in \mathbb{R}^+, \\ u(0,x) & = & u_0(x) & \forall x \in [-L,L] \end{array}\right. \tag{R-D}$$

où $\nu > 0$ est un coefficient de diffusion, $g : \mathbb{R} \longrightarrow \mathbb{R}$ est une fonction de classe $\mathscr{C}^1$ telle que $g(0) = g(1) = 0$, et positive sur $[0,1]$, typiquement la fonciton $g : u \mapsto u(1-u)$ ou $g : u \mapsto u\left(1-u^2\right)$ et où u_0 est une condition initiale de classe $\mathscr{C}^2$ sur $[-L, L]$ telle que $u_0'(-L) = u_0'(L) = 0$ (il s'agit de conditions de compatibilité avec les conditions aux bords de Neumann) et $0 \leqslant u_0 \leqslant 1$ (on travaille avec des "concentrations" de population). Cette équation modélise le comportement d'une population sur un segment $[-L, L]$. Si le jury demande si l'équation ci-dessus admet une unique solution, on peut dire que l'équation sans le terme non-linéaire $g(u(t,x))$ est l'équation de la chaleur, qui a une unique solution, y compris avec les conditions aux bords de Neumann, et qu'on peut effectuer un argument de point fixe sur la formule de Duhamel associée à l'équation de la chaleur pour obtenir l'existence et l'unicité de la solution. En tous cas, le but du développement n'est pas de justifier l'existence et l'unicité de la solution, mais de faire l'analyse numérique d'un schéma aux différences finies semi-implicite. On se donne alors une subdivision régulière de l'intervalle $[-L, L]$ à N_x points :

$$\forall j \in [\![1, N_x]\!], \quad x_j := -L + (j-1)\Delta x$$

où $\Delta x := \frac{2L}{N_x - 1}$, et une subdivision d'un intervalle de temps $[0, T]$ à N_t points :

$$\forall n \in [\![1, N_t]\!], \quad t_n := n\Delta t$$

où $\Delta t := \frac{T}{N_t}$. On veut alors approcher les quantités $u(t_n, x_j)$ de la solution de l'équation R-D par des quantités u_j^n vérifiant le schéma suivant :

$$\left\{\begin{array}{rcll} \dfrac{u_j^{n+1} - u_j^n}{\Delta t} & = & \nu \dfrac{u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}}{\Delta x^2} + g\left(u_j^n\right) & \forall (j,n) \in [\![1, N_x]\!] \times [\![0, N_t - 1]\!], \\ u_0^{n+1} & = & u_1^{n+1} & \forall n \in [\![0, N_t - 1]\!], \\ u_{N_x+1}^{n+1} & = & u_{N_x}^{n+1} & \forall n \in [\![0, N_t - 1]\!], \\ u_j^0 & = & u_0(x_j) & \forall j \in [\![1, N_x]\!]. \end{array}\right. \tag{D-F}$$

On a alors le résultat suivant :

Théorème 4.1 (Convergence et propriétés qualitatives du schéma)**.** Le schéma D-F est convergent d'ordre 1 en Δt et d'ordre 2 en Δx. De plus, si $\Delta t \leqslant \frac{1}{\|g'\|_{\infty,[0,1]}}$, alors la solution (u_j^n) du schéma vérifie :

$$\forall n \in [\![0, N_t]\!],\ \forall j \in [\![1, N_x]\!], \quad 0 \leqslant u_j^n \leqslant 1.$$

Démonstration. **Étape 1 : Réécriture du schéma sous forme d'un système linéaire**

Posons, pour $n \in [\![0, N_t]\!]$ notre vecteur des valeurs approchées au temps t_n $U^n := (u_j^n)_{j \in [\![1, N_x]\!]} \in \mathbb{R}^{N_x}$. On a alors que U^n vérifie la relation de récurrence suivante :

$$\forall n \in [\![0, N_t - 1]\!], \quad MU^{n+1} = U^n + \Delta t G^n,$$

où $M \in \mathscr{M}_{N_x}(\mathbb{R})$ désigne notre matrice d'itération :

$$M = I_n + \frac{\nu\Delta t}{\Delta x^2}\begin{pmatrix} 1 & -1 & & & \mathbf{O} \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ \mathbf{O} & & & -1 & 1 \end{pmatrix}$$

et G^n désigne le vecteur $\left(g\left(u_j^n\right)\right)_{j\in[\![1,N_x]\!]}$.

Étape 2 : Bonne définition du schéma

Le vecteur U^{n+1} est défini *implicitement*. Ainsi, il n'est *a priori* pas bien défini ! Pour montrer que U^{n+1} est bien défini, montrons que M est inversible. Pour cela, on montre que $M \in \mathscr{S}_{N_x}^{++}(\mathbb{R})$:

$$\forall V \in \mathbb{R}^{N_x}, \quad V^T M V = \sum_{i=1}^{N_x} v_i^2 + \frac{\nu\Delta t}{\Delta x^2}\left(v_1^2 + v_{N_x}^2 + 2\sum_{j=2}^{N_x-1} v_j^2 - 2\sum_{j=2}^{N_x} v_{j-1}v_j\right).$$

On va faire apparaître des carrés des différences $v_j - v_{j-1}$. Pour cela, on extrait une des deux sommes $\displaystyle\sum_{j=2}^{N_x-1} v_j^2$ et on effectue un changement d'indice :

$$\begin{aligned} \forall V \in \mathbb{R}^{N_x}, \quad V^T M V &= \sum_{i=1}^{N_x} v_i^2 + \frac{\nu\Delta t}{\Delta x^2}\left(v_1^2 + v_{N_x}^2 + \sum_{j=2}^{N_x-1} v_j^2 + \sum_{j=3}^{N_x} v_{j-1}^2 - \sum_{j=2}^{N_x} 2v_jv_{j-1}\right) \\ &= \sum_{i=1}^{N_x} v_i^2 + \frac{\nu\Delta t}{\Delta x^2}\left(\sum_{j=2}^{N_x} v_j^2 + \sum_{j=2}^{N_x} v_{j-1}^2 - \sum_{j=2}^{N_x} 2v_jv_{j-1}\right) \\ &= \sum_{i=1}^{N_x} v_i^2 + \frac{\nu\Delta t}{\Delta x^2}\sum_{j=2}^{N_x}(v_j - v_{j-1})^2. \end{aligned}$$

On a donc bien que M est symétrique définie positive !

Étape 3 : Consistance du schéma (à bien détailler pour la leçon 218, à passer vite pour la leçon 206)

Pour la consistance, on suppose que $u_0 \in \mathscr{C}^4([-L,L],\mathbb{R})$ et que $g \in \mathscr{C}^4([0,1],\mathbb{R})$ (pas sûr que ce soit nécessaire) afin de garantir le fait que la solution u soit de classe $\mathscr{C}^4$ sur $\mathbb{R}^+ \times [-L,L]$. On va appliquer la formule de Taylor-Lagrange pour obtenir nos estimations d'erreur de consistance :

1. Taylor-Lagrange à $u(t_{n+1},x_j)$:

$$\forall (n,j) \in [\![0,N_t-1]\!]\times[\![1,N_x]\!],\ \exists \tau_{n,j} \in [t_n,t_{n+1}], \quad u(t_{n+1},x_j) = u(t_n,x_j) + \Delta t\partial_t u(t_n,x_j) + \frac{\Delta t^2}{2}\partial_{tt}^2 u(\tau_{n,j},x_j).$$

2. Taylor-Lagrange à $u(t_{n+1},x_{j+1})$:

$$\begin{aligned} \forall (j,n) \in [\![1,N_x-1]\!]\times[\![0,N_t-1]\!],\ \exists \xi_{j,n}^+ \in [x_j,x_{j+1}], \quad u(t_{n+1},x_{j+1}) &= u(t_{n+1},x_j) + \Delta x\partial_x u(t_{n+1},x_j) + \\ &\frac{\Delta x^2}{2}\partial_{xx}^2 u(t_{n+1},x_j) + \\ &\frac{\Delta x^3}{6}\partial_{x^3}^3 u(t_{n+1},x_j) + \\ &\frac{\Delta x^4}{24}\partial_{x^4}^4 u\left(t_{n+1},\xi_{j,n}^+\right). \end{aligned}$$

3. Taylor-Lagrange à $u(t_{n+1}, x_{j-1})$:

$$\begin{aligned}\forall (j,n) \in [\![2, N_x]\!] \times [\![0, N_t - 1]\!],\ \exists \xi^-_{j,n} \in [x_j, x_{j+1}], \quad u(t_{n+1}, x_{j-1}) &= u(t_{n+1}, x_j) - \Delta x \partial_x u(t_{n+1}, x_j) + \\ &\frac{\Delta x^2}{2} \partial^2_{xx} u(t_{n+1}, x_j) - \\ &\frac{\Delta x^3}{6} \partial^3_{x^3} u(t_{n+1}, x_j) + \\ &\frac{\Delta x^4}{24} \partial^4_{x^4} u\left(t_{n+1}, \xi^-_{j,n}\right).\end{aligned}$$

En combinant ces résultats, on obtient que pour tout $n \in [\![0, N_t - 1]\!]$ et pour tout $j \in [\![2, N_x - 1]\!]$:

$$\begin{aligned}&\frac{u(t_{n+1}, x_j) - u(t_n, x_j)}{\Delta t} - \nu \frac{u(t_{n+1}, x_{j+1}) - 2u(t_{n+1}, x_j) + u(t_{n+1}, x_{j-1})}{\Delta x^2} - g(u(t_n, x_j)) \\ = \ &\partial_t u(t_n, x_j) + \frac{\Delta t}{2} \partial^2_{tt} u(\tau_{n,j}, x_j) - \nu \partial^2_{xx} u(t_{n+1}, x_j) - \nu \frac{\Delta x^2}{24} \left(\partial^4_{x^4} u(t_{n+1}, \xi^+_{j,n}) + \partial^4_{x^4} u(t_{n+1}, \xi^-_{j,n})\right) - g(u(t_n, x_j)) \\ = \ &\frac{\Delta t}{2} \partial^2_{tt} u(\tau_{n,j}, x_j) - \nu \left(\partial^2_{xx} u(t_{n+1}, x_j) - \partial^2_{xx} u(t_n, x_j)\right) - \nu \frac{\Delta x^2}{24} \left(\partial^4_{x^4} u(t_{n+1}, \xi^+_{j,n}) + \partial^4_{x^4} u(t_{n+1}, \xi^-_{j,n})\right).\end{aligned}$$

En appliquant une dernière fois la formule de Taylor-Lagrange à $\partial^2_{xx}(t_{n+1}, x_j)$, on obtient, en notant ε^n_j l'erreur de consistance du schéma :

$$\forall (n,j) \in [\![0, N_t - 1]\!] \times [\![2, N_x - 1]\!], \quad |\varepsilon^n_j| \leqslant \Delta t \left(\frac{M_{tt}}{2} + \nu M_{txx}\right) + \frac{\Delta x^2}{12} \nu M_{x^4},$$

où on a noté :

$$M_{tt} := \|\partial^2_{tt} u\|_{\infty, [0,T] \times [-L,L]}, \quad M_{txx} := \|\partial^3_{txx} u\|_{\infty, [0,T] \times [-L,L]}, \quad M_{x^4} := \|\partial^4_{x^4}\|_{\infty, [0,T] \times [-L,L]}.$$

On n'a pas encore traité les bords, c'est-à-dire ε^n_1 et $\varepsilon^n_{N_x}$! On va voir que l'erreur, qu'on pourrait penser plus grossière, ne l'est au final pas :

$$\begin{aligned}\varepsilon^n_1 &= \frac{u(t_{n+1}, -L) - u(t_n, -L)}{\Delta t} - \nu \frac{u(t_{n+1}, x_2) - u(t_{n+1}, -L)}{\Delta x^2} - g(u(t_n, -L)) \\ &= \partial_t u(t_n, -L) + \frac{\Delta t}{2} \partial^2_{tt} u(\tau_{n,1}, -L) - \nu \frac{\partial_x u(t_n, -L)}{\Delta x} - \nu \partial^2_{xx} u(t_{n+1}, -L) - \nu \frac{\Delta x}{6} \partial^3_{x^3} u(t_n, -L) - \\ &\quad \nu \frac{\Delta x^2}{24} \partial^4_{x^4} u(t_{n+1}, \xi^+_{1,n}) - g(u(t_n, x_j)).\end{aligned}$$

Or, $\partial_x u(t_n, -L) = 0$: ce sont les conditions aux bords ! De plus, on remarque que, d'après l'équation :

$$\nu \partial^3_{x^3}(t_n, -L) = \partial^2_{xt} u(t_n, -L) - \partial_x u(t_n, -L) g'(u(t_n, -L)) = \partial_t \left(\partial_x u\right)(t_n, -L) - 0.$$

Or, par continuité des dérivées partielles en $-L$, on a que $\partial^2_{xt} u(t_n, -L) = 0$ comme dérivée en temps de $\partial_x u(\cdot, -L)$ qui est la fonction nulle ! Ainsi :

$$\varepsilon^n_1 = \frac{\Delta t}{2} \partial^2_{tt} u(\tau_{n,1}, -L) - \nu \left(\partial^2_{xx} u(t_{n+1}, -L) - \partial^2_{xx} u(t_n, -L)\right) - \nu \frac{\Delta x^2}{24} \partial^4_{x^4} u(t_{n+1}, \xi^+_{1,n})$$

On a donc également notre estimation d'erreur :

$$|\varepsilon^n_1| \leqslant \Delta t \left(\frac{M_t}{2} + \nu M_{txx}\right) + \Delta x^2 \frac{\nu}{24} M_{x^4}.$$

De la même manière, on montre que :

$$|\varepsilon^n_{N_x}| \leqslant \Delta t \left(\frac{M_t}{2} + \nu M_{txx}\right) + \Delta x^2 \frac{\nu}{24} M_{x^4}.$$

Ainsi, en notant E^n le vecteur $(\varepsilon_j^n)_{j\in[\![1,N_x]\!]}$, on obtient :

$$\forall n \in [\![0, N_t - 1]\!], \quad \|E^n\|_\infty \leqslant C_t \Delta t + C_x \Delta x^2$$

avec $C_t := \frac{M_t}{2} + \nu M_{txx}$ et $C_x := \frac{\nu}{12} M_{x^4}$.

Étape 4 : Stabilité et conclusion

Notons, pour $n \in [\![0, N_t]\!]$, U_{ex}^n le vecteur $(u(t_n, x_j))_{j\in[\![1,N_x]\!]}$. On a alors que U_{ex}^n vérifie le schéma perturbé suivant :

$$\forall n \in [\![0, N_t - 1]\!], \quad M U_{\text{ex}}^{n+1} = U_{\text{ex}}^n + \Delta t \left(G^n + E^n\right).$$

Ainsi, on a que la différence $U_{\text{ex}}^n - U^n$ vérifie :

$$\forall n \in [\![0, N_t - 1]\!], \quad U_{\text{ex}}^{n+1} - U^{n+1} = M^{-1} \left(U_{\text{ex}}^n - U^n + \Delta t E^n\right).$$

Une récurrence facile montre alors :

$$\forall n \in [\![0, N_t]\!], \quad U_{\text{ex}}^n - U^n = M^{-n} \underbrace{\left(U_{\text{ex}}^0 - U^0\right)}_{=0} + \sum_{k=0}^{n-1} M^{k-n} \Delta t E^k.$$

Afin d'estimer cette erreur en norme infini, il faut montrer que la norme subordonnée $|\!|\!|M|\!|\!|_\infty$ ne dépende ni de Δt, ni de Δx. Pour cela on montre le fait suivant :

Lemme 4.2 (À bien détailler pour la leçon 206, pas le temps pour la leçon 218 je pense)**.** La matrice M du schéma numérique est monotone.

Démonstration. Soit $B \in \mathbb{R}^{N_x}$ un vecteur positif et soit $V \in \mathbb{R}^{N_x}$. On note V^- le vecteur $(\underbrace{\min(0, v_j)}_{=:v_j^-})_{j\in[\![1,N_x]\!]}$. On observe le fait suivant :

$$(V^-)^T M V = \sum_{j=1}^{N_x} v_j^- v_j + \sum_{j=2}^{N_x} (v_j^- - v_{j-1}^-)(v_j - v_{j-1}) = \sum_{j=1}^{N_x} (v_j^-)^2 + \sum_{j=2}^{N_x} (v_j^- - v_{j-1}^-)(v_j - v_{j-1}).$$

Maintenant, si v_j et v_{j-1} sont positifs, alors $v_j^- = v_{j-1}^- = 0$, ce qui donne 0 dans la somme. Si v_j est positif et v_{j-1} est négatif, alors $v_j^- = 0$ et $-v_{j-1}^-$ est positif. Ainsi, on a :

$$(v_j^- - v_{j-1}^-)(v_j - v_{j-1}) = \underbrace{-v_{j-1}}_{\geqslant 0} \underbrace{(v_j - v_{j-1})}_{\geqslant -v_{j-1}} \geqslant v_{j-1}^2 \geqslant 0.$$

Maintenant, si $v_j \leqslant 0$ et $v_{j-1} \geqslant 0$, alors :

$$(v_j^- - v_{j-1}^-)(v_j - v_{j-1}) = \underbrace{v_j}_{\leqslant 0} \underbrace{(v_j - v_{j-1})}_{\leqslant v_j} \geqslant v_j^2 \geqslant 0.$$

Et si v_j et v_{j-1} sont tous deux négatifs, alors le terme dans la somme devient $(v_j - v_{j-1})^2 \geqslant 0$. Ainsi, on a dans tous les cas :

$$(V^-)^T M V \geqslant \sum_{j=1}^{N_x} (v_j^-)^2 \geqslant 0.$$

Ainsi, si V est tel que $MV = B$, alors :

$$0 \leqslant (V^-)^T MV = (V^-)^T B \leqslant 0 \quad \text{car } V^- \leqslant 0 \text{ et } B \geqslant 0.$$

Ainsi $(V^-)^T MV = 0$. Or, on l'avait minoré par $\sum_{j=1}^{N_x} (v_j^-)^2$! Ainsi $v_j^- = 0$ pour tout $j \in [\![1, N_x]\!]$ et donc V est positif ! On a bien que M est monotone. □

Maintenant, il ne reste plus qu'à dire que puisque M est monotone, M^{-1} est positive, et donc :

$$\|\!|M^{-1}|\!\|_\infty = \max_{1 \leqslant i \leqslant N_x} \sum_{j=1}^{N_x} \left| \left[M^{-1}\right]_{i,j} \right| = \max_{1 \leqslant i \leqslant N_x} \sum_{j=1}^{N_x} \left[M^{-1}\right]_{i,j} = \|M^{-1}\mathbf{1}\|_\infty$$

où $\mathbf{1}$ désigne le vecteur dont toutes les composantes sont égales à 1. Mais miracle ! On observe très facilement que $M\mathbf{1} = \mathbf{1}$! Et donc $M^{-1}\mathbf{1} = \mathbf{1}$! Ainsi :

$$\|\!|M^{-1}|\!\|_\infty = \|\mathbf{1}\|_\infty = 1.$$

Cela permet de conclure que :

$$\forall n \in [\![0, N_t]\!], \quad \|U_{\text{ex}}^n - U^n\|_\infty \leqslant \Delta t \sum_{k=0}^{n-1} \left\| E^k \right\|_\infty \leqslant n\Delta t \left(C_t \Delta t + C_x \Delta x^2\right) \leqslant T(C_t \Delta t + C_x \Delta x^2).$$

Cela montre que le schéma est convergent en norme infini, d'ordre 1 en Δt et d'ordre 2 en Δx !

Étape 5 : Propriétés qualitatives du schéma (si le temps le permet)

On montre que pour tout n, les composantes du vecteur U^n sont dans $[0, 1]$ si $\Delta t \leqslant \frac{1}{\|g'\|_{\infty,[0,1]}}$. On raisonne par récurrence, l'initialisation étant claire puisque u_0 est à valeurs dans $[0, 1]$. Pour l'hérédité, on a déjà :

$$MU^{n+1} = \underbrace{U^n}_{\substack{\geqslant 0 \text{ par hypothèse} \\ \text{de récurrence}}} + \underbrace{\Delta t G^n}_{\geqslant 0 \text{ par hypothèse}} \geqslant 0.$$

Ainsi, par monotonie de M, $U^{n+1} \geqslant 0$. Maintenant :

$$M(\mathbf{1} - U^{n+1}) = \mathbf{1} - U^n - \Delta t G^n$$

Or, par inégalité des accroissements finis, étant donné que $g(1) = 0$:

$$\forall j \in [\![1, N_x]\!], \quad g\left(u_j^n\right) \leqslant \|g'\|_{\infty,[0,1]} \underbrace{\left(1 - u_j^n\right)}_{\geqslant 0}.$$

Ainsi :

$$\forall j \in [\![1, N_x]\!], \quad 1 - u_j^n - \Delta t g\left(u_j^n\right) \geqslant 1 - u_j^n - \underbrace{\Delta t \|g'\|_{\infty,[0,1]}}_{\leqslant 1} \left(1 - u_j^n\right) \geqslant 1 - u_j^n - (1 - u_j^n) = 0$$

Et donc le vecteur $M(\mathbf{1} - U^{n+1})$ est positif ! Par monotonie de M, on a donc que $\mathbf{1} - U^{n+1}$ est positif, et donc on a bien montré :

$$0 \leqslant U^{n+1} \leqslant 1,$$

ce qui termine l'étude de ce schéma numérique ! □