Université de Rennes 2024-2025
Agrégation externe

Compléments sur l'approximation numérique de l'équation des ondes

Miguel Rodrigues

Ces notes de cours sont consacrées à l'approximation numérique des solutions de l'équation des ondes. Une partie du matériel discuté ici peut être trouvé dans les livres classiques d'analyse numérique, notamment dans [Fil13, Chapitre 7].

1 Premiers commentaires

On cherche à calculer des approximations de solutions de l'équation des ondes. On rapelle que l'on peut ramener les conditions de bord de Dirichlet et/ou de Neumann inhomogènes à leurs version homogènes par simple soustraction d'une fonction vérifiant les conditions de bord (ce qui modifie le terme source de l'équation), et que l'on peut ramener ces conditions homogènes à des conditions périodiques par prolongement pair/impair puis périodisation.
Une fois que l'on sait traiter le problème périodique sans terme source on peut passer au problème avec terme source par variation de la constante/formule de Duhamel (continue ou discrète). Notons cependant que lorsque l'on discrétise la formule de Duhamel on introduit une nouvelle erreur, due à la formule de quadrature choisie.
Par la suite nous nous contenterons de traiter le cas périodique sans terme source. Par homotéthie en espace on peut fixer la période à 1 (quitte à changer la vitesse) ce que nous ferons également désormais.
Comme on dispose de deux formules explicites on peut les utiliser pour construire une approximation numérique :

— en discrétisant la formule en espace (dite formule de d'Alembert) par une méthode de quadrature ;
— en utilisant la formule de Fourier, on obtient une méthode spectrale (d'ordre infini en espace et en temps quand le terme source est nul).

La première est rendue peu pratique par le fait que l'intégrale n'est pas facilement restreinte à une intégrale sur un intervalle de longueur 1.
Nous allons nous focaliser sur l'analyse de la méthode des différences finies. Pour préparer l'analyse rappelons que pour une vitesse $c > 0$ donnée le problème $\mathbf{Z}$-périodique sans terme source se récrit

$$\begin{aligned}
& & \partial_t^2 u - c^2 \partial_x^2 u = 0 & \qquad \text{sur }]0,T[\times]0,1[\,,\\
u(0,\cdot) &= u_0\,, & \partial_t u(0,\cdot) = u_1 & \qquad \text{sur }]0,1[\,,\\
u(\cdot,0) &= u(\cdot,1)\,, & \partial_x u(\cdot,0) = \partial_x u(\cdot,1) & \qquad \text{sur }]0,T[\,.
\end{aligned}$$

Au niveau discret il sera également utile de récrire l'équation des ondes comme un système. Il existe au moins trois choix naturels. Le premier, directement inspiré des équations différentiels ordinaires consiste à poser $w = \partial_t u$ et à récrire l'équation comme

$$\partial_t u = w\,, \qquad \partial_t w = c^2 \partial_x^2 u\,.$$

Le deuxième, inspiré de la factorisation $\partial_t^2 - c^2\partial_x^2 = (\partial_t + c\partial_x)(\partial_t - c\partial_x)$, consiste à poser $v = \partial_t u + c\,\partial_x u$ de sorte que l'équation se récrit comme

$$\partial_t v - c\,\partial_x v = 0\,, \qquad\qquad \partial_t u + c\,\partial_x u = v\,.$$

Le troisième, inspiré par la structure hamiltonienne, vise à symétriser le système obtenu en posant $(q,p) = (c\partial_x u, \partial_t u)$ de sorte que

$$\partial_t q - c\,\partial_x p = 0\,, \qquad\qquad \partial_t p - c\,\partial_x q = 0\,,$$

c'est-à-dire

$$\partial_t \begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} 0 & c\,\partial_x \\ c\,\partial_x & 0 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}.$$

Avec des bonnes conditions de bord (par exemple périodiques), l'opérateur

$$\begin{pmatrix} 0 & c\,\partial_x \\ c\,\partial_x & 0 \end{pmatrix}$$

est anti-symétrique et l'on retrouve que la norme L^2 de (q,p) est constante en temps. On reconstruit u à partir de (q,p) par $u(t,\cdot) = u_0 + \int_0^t p(s,\cdot)\,\mathrm{d}s$.

Puisque l'essentiel de notre analyse des schémas est faite en Fourier, notons aussi qu'avec $(q,p) = (c\partial_x u, \partial_t u)$, pour $k \in \mathbf{Z}$, on a [1]

$$\begin{pmatrix} c_k(q(t,\cdot)) \\ c_k(p(t,\cdot)) \end{pmatrix} = \begin{pmatrix} \cos(2\pi\,k\,t) & \mathrm{i}\sin(2\pi\,k\,t) \\ \mathrm{i}\sin(2\pi\,k\,t) & \cos(2\pi\,k\,t) \end{pmatrix} \begin{pmatrix} c_k(q(0,\cdot)) \\ c_k(p(0,\cdot)) \end{pmatrix}$$

où la matrice

$$\begin{pmatrix} \cos(2\pi\,k\,t) & \mathrm{i}\sin(2\pi\,k\,t) \\ \mathrm{i}\sin(2\pi\,k\,t) & \cos(2\pi\,k\,t) \end{pmatrix}$$

est unitaire (c'est-à-dire qu'elle définit une isométrie).

2 Méthode des différences finies

La méthode des différences finies est moins efficace que la méthode spectrale mais plus robuste. En particulier, bien que nous n'en fassions l'analyse qu'avec des pas équidistants et des conditions de bord périodiques, le schéma fonctionne bien plus largement.

Étant donnés des pas d'espace et de temps Δx et Δt tels que $N_x := 1/\Delta x \in \mathbf{N}^*$, on note N_t la partie entière de $T/\Delta t$ et on introduit les nœuds de subdivisions $x_j = j\Delta x$, $j \in [\![1, N_x]\!]$, et $t_n = n\Delta t$, $n \in [\![0, N_t]\!]$. On souhaite calculer u_j^n, des valeurs en (t_n, x_j) censées fournir une approximation des valeurs $u(t_n, x_j)$ de la solution. Insistons encore une fois sur le fait que bien que nous ne le marquions pas par des indices et des exposants, (u_j^n, t_n, x_j) ne dépendent pas que de (j,n) mais aussi de $(\Delta t, \Delta x)$ ou de manière équivalente de $(\Delta t, N_x)$.

Les deux schémas les plus simples sont

$$u_j^{n+1} - 2u_j^n + u_j^{n-1} = c^2\frac{(\Delta t)^2}{(\Delta x)^2}\left(u_{j+1}^{n-1} - 2u_j^{n-1} + u_{j-1}^{n-1}\right), \qquad j \in [\![1, N_x]\!], \qquad n \in [\![1, N_t - 1]\!]\,, \tag{2.1}$$

1. Concernant Fourier sous toutes ses formes on utilise les conventions des compléments assosicés.

et

$$u_j^{n+1} - 2u_j^n + u_j^{n-1} = c^2 \frac{(\Delta t)^2}{(\Delta x)^2} \left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right), \qquad j \in [\![1, N_x]\!], \qquad n \in [\![1, N_t - 1]\!], \tag{2.2}$$

où les valeurs en dehors de $j \in [\![1, N_x]\!]$ sont obtenues par périodicité, respectivement par $u_0^{n-1} := u_{N_x}^{n-1}$, $u_{N_x+1}^{n-1} := u_1^{n-1}$ et $u_0^n := u_{N_x}^n$, $u_{N_x+1}^n := u_1^n$. Pour préparer l'analyse de stabilité fixons $\lambda := \Delta t / \Delta x$.

Remarque 1 *En appliquant un schéma aux différences finis décentrés amont sur les équations du système (u, v) mentionné ci-dessus, on obtient un autre schéma naturel*

$$v_j^{n+1} = v_j^n + c\lambda\,(v_{j+1}^n - v_j^n), \qquad\qquad u_j^{n+1} = u_j^n - c\lambda\,(u_j^n - u_{j-1}^n) + \Delta t v_j^n\,.$$

Le schéma ainsi obtenu est stable sous condition $0 < c\lambda < 1$. Cela donne un schéma d'ordre 1 en espace et en temps. L'analyse complète est un peu plus simple que celle que nous allons menée pour (2.1)/(2.2).

2.1 Stabilité

Nous allons montrer que le premier schéma est instable alors que le second schéma est stable (en un certain sens) sous une certaine condition CFL.
Avant de nous lancer notons qu'il y a une difficulté liée au fait que l'évolution, comme système d'ordre 1 en temps, est en fait donnée par un système de deux équations et qu'il faut choisir les normes sur les coordonnées de manière compatible. Ainsi on ne peut pas propager $\|u(t,\cdot)\|_{L^2}$ tout seul, mais pas plus $\|(u, \partial_t u)(t,\cdot)\|_{L^2}$ alors que l'on peut propager $\|(u, \partial_x u, \partial_t u)(t,\cdot)\|_{L^2}$ ou $\|(\partial_x u, \partial_t u)(t,\cdot)\|_{L^2}$. Cette difficulté se retrouve aussi au niveau discret parce que l'on souhaite de l'uniformité dans la discrétisation. Cependant au niveau continu lorsque l'on rate le choix de norme L^2 cela ne doit coûter qu'un nombre fini de dérivées spatiales en plus sur la donnée initiale. Par analogie, au niveau discret si l'on rate le choix de norme cela ne doit coûter qu'une perte d'une puissance de Δx.
Laissons cela de côté pour l'instant et menons une analyse directe de stabilité des relations de récurrences d'ordre 2 données par (2.1) et (2.2). Posons $U^n := (u_1^n, \cdots, u_{N_x}^n)$, puis $C^n := \mathcal{F}_{N_x}(U^n)$. En Fourier, le schéma (2.1) devient, pour tout $k \in [\![-\lfloor \frac{N_x}{2} \rfloor, N_x - \lfloor \frac{N_x}{2} \rfloor - 1]\!]$ et tout $n \in [\![0, N_t - 2]\!]$

$$\begin{pmatrix} C_k^{n+1} \\ C_k^{n+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 + 2c^2\lambda^2\,(\cos(2\pi k \Delta x) - 1) & 2 \end{pmatrix} \begin{pmatrix} C_k^n \\ C_k^{n+1} \end{pmatrix}.$$

La stabilité L^2 de (2.1) est équivalente à l'existence d'une constante C_T telle que

$$\left\| \begin{pmatrix} 0 & 1 \\ -1 + 2c^2\lambda^2\,(\cos(2\pi k \Delta x) - 1) & 2 \end{pmatrix}^n \right\| \leq C_T\,, \quad k \in [\![-\lfloor \tfrac{N_x}{2} \rfloor, N_x - \lfloor \tfrac{N_x}{2} \rfloor - 1]\!]\,, \quad 0 \leq (n-2)\Delta t \leq T\,.$$

Les valeurs propres de la k-matrice d'itération (dont on considère la puissance n) sont

$$1 \pm \mathrm{i}\,2\lambda\,c\,\sqrt{\tfrac{1}{2}(1 - \cos(2\pi k \Delta x))}\,.$$

En choisissant k proche de $\frac{N_x}{2}$, on obtient des valeurs proches de $1 \pm \mathrm{i}\,2\lambda\,c$ qui sont de module $\sqrt{1 + 4\lambda^2\,c^2}$ strictement supérieur à 1. En choisissant n proche de $T/\Delta t$, on déduit une borne inférieure proche de $\sqrt{1 + 4\lambda^2\,c^2}^{T/\Delta t} = \sqrt{1 + 4\lambda^2\,c^2}^{T/(\lambda\,\Delta x)}$. Cela montre l'instabilité L^2.
Notons que la croissance démontrée est exponentielle en Δx et ne peut donc pas être réparée par un meilleur choix de normes L^2.

Tournons-nous vers le schéma (2.2). La k-matrice d'itération est maintenant

$$\begin{pmatrix} 0 & 1 \\ -1 & 2 + 2c^2\lambda^2\left(\cos(2\pi k\Delta x) - 1\right) \end{pmatrix}$$

dont les valeurs propres sont

$$m_\lambda(k) \pm \sqrt{(m_\lambda(k))^2 - 1}$$

où

$$m_\lambda(k) := 1 - c^2\lambda^2\left(1 - \cos(2\pi k\Delta x)\right),$$

pour une certaine détermination de la racine carrée. Avec k proche de $\frac{N_x}{2}$, $m_\lambda(k)$ est proche de $1-2c^2\lambda^2$, qui donne des valeurs propres proches de

$$1 - 2c^2\lambda^2 \pm \sqrt{(1 - 2c^2\lambda^2)^2 - 1}\,.$$

Quand $\lambda c > 1$, $1 - 2c^2\lambda^2 < -1$ et au moins l'une des valeurs propres est réelle et strictement inférieure à $1 - 2c^2\lambda^2$. Cela donne une instabilité exponentielle quand $\lambda c > 1$.
Si $\lambda c \leq 1$, alors pour tout k, $-1 \leq m_\lambda(k) \leq 1$ et les valeurs propres

$$m_\lambda(k) \pm \mathrm{i}\sqrt{1 - (m_\lambda(k))^2}$$

sont toutes de module 1. Cela ne suffit pas à exclure l'instabilité. Nous sommes dans un cas limite de l'analyse des puissances de matrice. Même si toutes les k-matrices d'itération étaient diagonalisables pour utiliser la diagonalisation il faudrait contrôler l'uniformité de la diagonalisation.
Commençons par étudier le cas $k = 0$. Alors $m_\lambda(0) = 1$ et la valeur propre est double, égale à 1, mais la matrice d'itération

$$\begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$$

n'est pas diagonale donc on a un bloc de Jordan de taille 2. La matrice d'itération pour $k = 0$ ne dépend pas de Δx et donc ses puissances n-ièmes croissent comme $\mathcal{O}(n)$. Cela montre une instabilité en $\mathcal{O}(1/\Delta x)$. C'est précisément le type de croissance que nous avions annoncé pour le cas où l'on a mal choisi les normes L^2. Pour se convaincre que cette instabilité n'est pas pertinente pour les approximations numériques de l'équation des ondes remarquons que par souci de consistance on n'utilise (2.2) qu'avec des données telles que $U^1 = U^0 + \mathcal{O}(\Delta t)$ donc $C_0^1 = C_0^0 + \mathcal{O}(\Delta t)$. Puisque $(1,1)$ est un vecteur propre de la matrice ci-dessus on déduit que pour de telles données la taille finale de C_0^n sera seulement en $\mathcal{O}(n\Delta t) = \mathcal{O}(1)$.
Retournons au problème de stabilité qui implique de contrôler toutes les fréquences (et pas seulement la fréquence $k = 0$). L'analyse de la fréquence $k = 0$ et l'analogie avec l'analyse continue suggèrent d'au moins commencer par remplacer l'étude de (U^n, U^{n+1}) par celle de $(U^n, (U^{n+1} - U^n)/\Delta t)$. Dans cette direction notons que

$$\begin{pmatrix} C_k^{n+1} \\ \dfrac{C_k^{n+2} - C_k^{n+1}}{\Delta t} \end{pmatrix} = \begin{pmatrix} 1 & \Delta t \\ 2c^2\lambda\,\frac{\cos(2\pi k\Delta x)-1}{\Delta x} & 1 + 2c^2\lambda^2\left(\cos(2\pi k\Delta x) - 1\right) \end{pmatrix} \begin{pmatrix} C_k^{n} \\ \dfrac{C_k^{n+1} - C_k^{n}}{\Delta t} \end{pmatrix}.$$

Cela résout effectivement le problème en $k = 0$ puisque la matrice d'itération correspondante est

$$\begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}$$

dont les puissances n-ièmes avec $0 \leq (n-2)\Delta t \leq T$ sont bornées. La présence d'un bloc de Jordan en $k=0$ montre que l'on peut pas espérer une diagonalisation uniforme pour $k \neq 0$ petit. Par ailleurs, au vu de l'analyse continue on s'attend effectivement à devoir de plus introduire une variable correspondant à la discrétisation de $c\partial_x u$.

Plutôt que de mener au bout une analyse spectrale des puissances, essayons plutôt quand $0 < c\lambda \leq 1$ de montrer la stabilité par une estimation d'énergie menée en espace, donnant une contrepartie discrète à la conservation de l'énergie continue. Pour cela commençons par remarquer que (2.2) se récrit

$$\frac{1}{\Delta t}\left(\frac{U^{n+2}-U^{n+1}}{\Delta t} - \frac{U^{n+1}-U^{n}}{\Delta t}\right) \;=\; -c^2 D_{\Delta x}^{*} D_{\Delta x} U^{n+1}$$

pour $0 \leq (n+2)\Delta t \leq T$, où $D_{\Delta x}$ est défini par

$$(D_{\Delta x}V)_j = \frac{V_{j+1}-V_j}{\Delta x}\,, \qquad 1 \leq j \leq N_x\,,$$

(avec prolongement périodique), et son adjoint $D_{\Delta x}^{*}$ est donné par

$$(D_{\Delta x}^{*}V)_j = -\frac{V_j - V_{j-1}}{\Delta x}\,, \qquad 1 \leq j \leq N_x\,,$$

(avec prolongement périodique). L'identité remarquable $\|A\|^2 - \|B\|^2 = \langle A+B, A-B\rangle$ est la contrepartie discrète de la formule de dérivation du carré d'une norme. On déduit ainsi que (2.2) implique

$$\left\|\frac{U^{n+2}-U^{n+1}}{\Delta t}\right\|^2 - \left\|\frac{U^{n+1}-U^{n}}{\Delta t}\right\|^2 \;=\; -c^2\left\langle D_{\Delta x}(U^{n+2}-U^n), D_{\Delta x}U^{n+1}\right\rangle$$

c'est-à-dire

$$\left\|\frac{U^{n+2}-U^{n+1}}{\Delta t}\right\|^2 + c^2\left\langle D_{\Delta x}U^{n+2}, D_{\Delta x}U^{n+1}\right\rangle \;=\; \left\|\frac{U^{n+1}-U^{n}}{\Delta t}\right\|^2 + c^2\left\langle D_{\Delta x}U^{n+1}, D_{\Delta x}U^{n}\right\rangle.$$

Ainsi en introduisant

$$\mathcal{E}_{\lambda,\Delta t}(A,B) := \left\|\frac{B-A}{\Delta t}\right\|^2 + c^2\left\langle D_{\Delta x}B, D_{\Delta x}A\right\rangle$$

l'on constate que $\mathcal{E}_{\lambda,\Delta t}(U^n, U^{n+1})$ est conservé par (2.2). Pour conclure l'analyse de stabilité nous observons que quand $0 < c\lambda < 1$, $\mathcal{E}_{\lambda,\Delta t}(A,B)$ équivaut à

$$\left\|\frac{B-A}{\Delta t}\right\|^2 + c^2\,\|D_{\Delta x}A\|^2\,.$$

Cela résulte du fait que

$$\begin{aligned}
\left|\mathcal{E}_{\lambda,\Delta t}(A,B) - \left(\left\|\frac{B-A}{\Delta t}\right\|^2 + c^2\,\|D_{\Delta x}A\|^2\right)\right| &= c^2\Delta t\left|\left\langle D_{\Delta x}\frac{B-A}{\Delta t}, D_{\Delta x}A\right\rangle\right| \\
&\leq 2\,c^2\lambda\left\|\frac{B-A}{\Delta t}\right\|\,\|D_{\Delta x}A\| \\
&\leq c\lambda\left(\left\|\frac{B-A}{\Delta t}\right\|^2 + c^2\,\|D_{\Delta x}A\|^2\right).
\end{aligned}$$

Ainsi on déduit pour (2.2) quand $0 < c\lambda < 1$ de la stabilité L^2 mesurée sur $\left(D_{\Delta x}U^n, \frac{U^{n+1}-U^n}{\Delta t}\right)$ (ou sur $\left(U^n, D_{\Delta x}U^n, \frac{U^{n+1}-U^n}{\Delta t}\right)$.

2.2 Consistance

Pour compléter l'analyse de convergence dans le cas stable, il faut mener une analyse de stabilité.
Ici on a deux types d'erreur de consistance

1. l'erreur intérieure due à (2.2) ;
2. l'erreur d'initialisation due au choix de U^1.

Le premier type est toujours présent et son analyse (dans les variables $\left(D_{\Delta x}U^n, \frac{U^{n+1}-U^n}{\Delta t}\right)$) donne une erreur d'ordre 2 en espace et en temps. Le second type est dû au fait que (2.2) est un schéma à deux pas en temps.
Nous n'avons pas jusqu'ici discuté du choix de U^1. Puisque le schéma est d'ordre 2 en temps dans les variables $\left(D_{\Delta x}U^n, \frac{U^{n+1}-U^n}{\Delta t}\right)$, pour ne pas réduire l'ordre il faut également garantir que $\frac{U^1-U^0}{\Delta t}$ est une approximation de $\left(\frac{u(\Delta t,x_j)-u_0(x_j)}{\Delta t}\right)_j$ à l'ordre 2. On peut l'obtenir en discrétisant en espace

$$\begin{aligned} u(\Delta t, x) &\overset{\Delta t \to 0}{=} u_0(x) + \Delta t\, u_1(x) + \frac{(\Delta t)^2}{2}\,\partial_t^2 u(0,x) + \mathcal{O}((\Delta t)^3) \\ &\overset{\Delta t \to 0}{=} u_0(x) + \Delta t\, u_1(x) + c^2\frac{(\Delta t)^2}{2}\,\partial_x^2 u_0(x) + \mathcal{O}((\Delta t)^3)\,. \end{aligned}$$

Avec un terme source dans l'équation, celui-ci apparaîtrait aussi ici.

Références

[Fil13] F. Filbet. *Analyse numérique - Algorithme et étude mathématique.* Dunod, 2013.