Application de méthodes d'optimisation à la résolution des EDP

Jacob Leygonie et Claire Lasserre

15 janvier 2017

1 Question I.2.1: Déterminer une relation sur ρ_k

Notons $r_k := \nabla J(uk)$, où la fonction J est définie sur $\mathbb{R}^n$ par
$J(u) = \frac{1}{2} < Au, u > -bu$.
Remarquons que $\forall u, \nabla J(u) = Au - b$, de sorte que la dérivée de la fonction $F(\rho) := J(u_k - \rho r_k)$ se dérive en

$$Au_k - b - \rho A r_k = r_k - \rho A r_k$$

En égalisant ce terme à 0 pour maximiser $F(\rho)$, et en faisant le produit scalaire du résultat avec r_k, il vient directement que:

$$\rho = \frac{< r_k, r_k >}{< Ar_k, r_k >}$$

qui est bien défini car A est définie positive.

2 Question II.2.1: Écrire le processus itératif

On réitère un même procédé pour obtenir β_{k+1} à partir de β_k:
-En l'occurence on minimise bien une fonctionnelle du type $||J(u)|| = \frac{1}{2}||f(u)||^2$ puisque $J(\beta) = \frac{1}{2}\sum_{i=1}^{m} ||y_i - g(x_i, \beta)||^2$.
-On fait l'approximation suggérée par l'énoncé, que $\nabla^2 J(\beta_k) = Df^T(\beta_k) Df(\beta_k)$ où
$Df(\beta_k)$ est la jacobienne de f en β_k et $f =$

$$\begin{pmatrix} f_1 \\ f_2 \\ . \\ . \\ f_m \end{pmatrix}$$

où $f_i(\beta) = y_i - g(x_i, \beta)$.
-On inverse cette matrice (pour le moment sans se soucier de la complexité).
-On applique la descente de Newton:

$$\beta_{k+1} = \beta_k - (\nabla^2 J(\beta_k))^{-1} \nabla J(\beta_k)$$

-On réitère le processus N fois en tout (il s'agit d'un paramètre de la fonction; on aurait également pu choisir de réitérer le processus tant que le gradient dépasse un certain seuil en valeur absolue).

3 Question III.1.1: Consistance des schémas implicites et explicites

On considère les schémas explicites et implicites respectivement, pour l'équation de convection-diffusion 1D:

$$\frac{u_i^{n+1} - u_i^n}{\triangle t} + a\frac{u_{i+1}^n - u_{i-1}^n}{2\triangle x} - b\frac{u_{i+1}^n + u_{i-1}^n - 2u_i^n}{\triangle x^2} = 0 \tag{1}$$

$$\frac{u_i^{n+1} - u_i^n}{\triangle t} + a\frac{u_{i+1}^{n+1} - u_{i-1}^{n+1}}{2\triangle x} - b\frac{u_{i+1}^{n+1} + u_{i-1}^{n+1} - 2u_i^{n+1}}{\triangle x^2} = 0 \tag{2}$$

Effectuons un développement limité des différents termes:

$$\frac{u_i^{n+1} - u_i^n}{\triangle t} = \frac{\partial u}{\partial t}(x_i, t_n) + O(\Delta t) \tag{3}$$

$$a\frac{u_{i+1}^n - u_{i-1}^n}{2\triangle x} = a\frac{\partial u}{\partial x}(x_i, t_n) + O(\Delta x^2) \tag{4}$$

$$-b\frac{u_{i+1}^n + u_{i-1}^n - 2u_i^n}{\triangle x^2} = -b\frac{\partial^2 u}{\partial^2 x}(x_i, t_n) + O(\Delta x^2) \tag{5}$$

En sommant les trois termes et en utilisant l'équation de convection-diffusion, on obtient que le schéma explicite est consistant d'ordre 1 en temps et 2 en espace.
Le cas du schéma implicite se traite exactement de la même façon et l'on obtient les mêmes résultats.

4 Question III.1.2: Stabilité des schémas en norme ℓ^∞

Commençons par le schéma explicite. Pour établir des CFL suffisante, on utilise la relation de récurrence en voyant celle-ci comme une combinaison convexe des composantes courantes. En effet, si l'on a (en posant $\alpha := a\frac{\triangle t}{\triangle x}$ et$\beta := b\frac{\triangle t}{2\triangle x^2}$):

$$0 \leqslant 1 - 2\beta \leqslant 1 \tag{6}$$

$$0 \leqslant \alpha + \beta \leqslant 1 \tag{7}$$

$$0 \leqslant \beta - \alpha \leqslant 1 \tag{8}$$

alors la relation de récurrence définit une combinaison convexe des éléments courants, si bien que le principe du maximum discret nous assure que le schéma est stable. Remarquons que l'on peut ramener ces trois équations aux deux suivantes:

$$0 \leqslant \beta - \alpha \tag{9}$$

$$0 \leqslant 1 - 2\beta \tag{10}$$

D'où la CFL canditate: $\left\{ \begin{array}{lcl} \frac{\triangle t}{\triangle x^2} & \leqslant & \frac{1}{2b} \\ \triangle x & \leqslant & \frac{2b}{a} \end{array} \right.$

À ce stade, nous avons trouvé une CFL suffisante. Pour montrer qu'elle est nécessaire, nous avons envisagé deux approches:
-En supposant par l'absurde que l'une des CFL n'est pas vérifiée, on essaye de choisir un vecteur initial qui fasse décoller l'évolution de la solution du schéma. Classiquement, on choisit des vecteurs alternés tels que

$$\begin{pmatrix} (-1) \\ 1 \\ . \\ . \\ (-1)^j \\ . \\ . \\ (-1)^m \end{pmatrix}$$

On peut également penser à un vecteur à unique composante non nulle, ou bien un vecteur avec les mêmes composantes, ou encore des solutions plus compliquées. Nous avons essayé maintes possibilités, notamment en s'inspirant des cas de convection ou de diffusion pure pour choisir des vecteurs usuels et traiter leurs itérés en considérant la matrice de passage de notre schéma comme la somme des matrices de passage du schéma pour la convection pure et du schéma pour la diffusion pure. Néanmoins, l'inter-corrélation des deux termes et les conditions aux bords de Dirichlet rendent impossibles le calcul.
-En essayant de réduire la matrice de passage. En effet, si l'on parvient à diagonaliser la matrice, on aura une CFL nécessaire et suffisante portant sur la norme de la plus grande valeur propre. Passer par une formule de récurrence sur la forme du polynôme caractéristique n'aboutit pas. Il existe en revanche des techniques de diagonalisation des matrices tri diagonales, notamment si les produits des termes de diagonale supérieure avec ceux de la diagonale inférieure sont positifs. Mais cette hypothèse revient en fait à supposer une CFL, et cette méthode est donc obsolète. Une autre considération est d'itérer la matrice sur les vecteurs propres de la matrice de passage de l'équation de convection ou celle de l'équation de diffusion. L'inter corrélation empêche encore une fois d'obtenir de bons résultats.

En définitive, prouver le caractère nécessaire de la CFL nécessite un approfondissement qui sort certainement du champ de nos connaissances. Il est néanmoins très probable que si cela est possible, cela passe par l'investigation plus précise des deux voies décrites ci-dessus.

5 Question III.1.2: Stabilité des schémas en norme ℓ^2

Commençons par le schéma explicite. Développons en série de Fourrier les composantes de $u(x)$:

$$u_j^n(x) = \sum_{-\infty}^{+\infty} e^{2ik\pi x} \widehat{u}_j^n(k)$$

En posant $v_n(x) := u_n(x + \triangle x)$ on a que les coefficient de Fourrier de v^n sont les

$$\widehat{v}_j^n(k) = \widehat{u}_j^n(k) e^{2ik\pi \triangle x}$$

En passant tous les membres au temps t_n de l'équation explicite dans le membre de droite, et en passant aux coefficients de Fourrier, il vient que:

$$\begin{aligned}
\forall j, \widehat{u}_j^{n+1}(k) &= (1 - 2b\frac{\triangle t}{\triangle x^2})\widehat{u}_j^n(k) + (a\frac{\triangle t}{2\triangle x} + b\frac{\triangle t}{\triangle x^2})e^{-2ik\pi\triangle x}\widehat{u}_j^n(k) + (-a\frac{\triangle t}{2\triangle x} + b\frac{\triangle t}{\triangle x^2})e^{2ik\pi\triangle x}\widehat{u}_j^n(k) \\
&= [(1 - 4b\frac{\triangle t}{\triangle x^2} sin^2(k\pi\triangle x)) - ia\frac{\triangle t}{2\triangle x} sin(2k\pi\triangle x)]\widehat{u}_j^n(k) \\
&= [(1 - 4b\frac{\triangle t}{\triangle x^2} sin^2(k\pi\triangle x)) - ia\frac{\triangle t}{2\triangle x} 2sin(k\pi\triangle x)cos(k\pi\triangle x)]\widehat{u}_j^n(k) \\
&=: A(k)\widehat{u}_j^n(k)
\end{aligned}$$

Par le théorème de Plancherel,

$$\int_{-\infty}^{\infty} |u^n(x)|^2 dx = \sum_{k=-\infty}^{+\infty} |\widehat{u}^n(k)|^2$$

Ainsi le schéma est bornée en norme ℓ^2 si et seulement si les coefficients d'amplification $A(k)$ sont inférieurs à 1 en norme. Or,

$$|A(k)|^2 \leqslant 1 \Leftrightarrow$$

$$(1 - 4b\frac{\triangle t}{\triangle x^2} sin^2(k\pi\triangle x))^2 + (a\frac{\triangle t}{\triangle x})^2 sin^2(k\pi\triangle x)cos^2(k\pi\triangle x) \leqslant 1 \quad \Leftrightarrow$$

Posons $\lambda := 4b\frac{\triangle t}{\triangle x^2}$, $\mu := a\frac{\triangle t}{\triangle x}$ et $X := sin^2(k\pi\triangle x)$.
En développant tous les termes et du fait que $sin^2(k\pi\triangle x)cos^2(k\pi\triangle x) = sin^2(k\pi\triangle x) - sin^4(k\pi\triangle x)$, l'inéquation précédente se réécrit:

$$\begin{aligned} X^2(\lambda^2-\mu^2) + (\mu^2-2\lambda)X + 1 \leqslant 1 \quad &\Leftrightarrow \\ X[X(\lambda^2-\mu^2) + (\mu^2-2\lambda)] \leqslant 1 \quad &\Leftrightarrow \\ X \leqslant \frac{2\lambda-\mu^2}{\lambda^2-\mu^2} \quad &\Leftrightarrow \\ 1 \leqslant \frac{2\lambda-\mu^2}{\lambda^2-\mu^2} \quad &\Leftrightarrow \\ \lambda \leqslant 2 \end{aligned}$$

car $0 \leqslant X \leqslant 1$. D'où la CFL pour le schéma explicite appliquée à l'équation de convection-diffusion :

$$\frac{\triangle x^2}{\triangle t} \geqslant 2b$$

Dans le cas du schéma implicite, on obtient exactement de la même manière un relation sur les coefficients de Fourrier:

$$\begin{aligned} \forall j, \widehat{u}_j^n(k) &= [(1 + 4b\frac{\triangle t}{\triangle x^2}sin^2(k\pi\triangle x)) + ia\frac{\triangle t}{2\triangle x}sin(2k\pi\triangle x)]\widehat{u}_j^{n+1}(k) \\ &=: B(k)\widehat{u}_j^{n+1}(k) \end{aligned}$$

Comme les deux termes sinusoïdaux ne sauraient s'annuler simultanément, on a bien $|B(k)| > 1$ donc $|B(k)|^{-1} < 1$. Le schéma implicite est donc inconditionnellement stable en norme ℓ^2.

6 Question III.2.1:Matrice de passage pour le schéma implicite

La relation de récurrence du schéma implicite permet d'écrire:

$$u_i^{n+1} + a\triangle t\frac{u_{i+1}^{n+1} - u_{i-1}^{n+1}}{2\triangle x} - b\triangle t\frac{u_{i+1}^{n+1} + u_{i-1}^{n+1} - 2u_i^{n+1}}{\triangle x^2} = u_i^n$$

Avec les condition aux limites de Dirichlet, cela s'écrit directement $U^n = MU^{n+1}$ avec M la matrice

$$\begin{pmatrix} 1 + 2b\frac{\triangle t}{\triangle x^2} & a\frac{\triangle t}{2\triangle x} - b\frac{\triangle t}{\triangle x^2} & 0 & \cdots & 0 \\ -a\frac{\triangle t}{2\triangle x} - b\frac{\triangle t}{\triangle x^2} & \ddots & \ddots & 0 & \\ 0 & \ddots & & & \\ \vdots & & & & \\ 0 & & & & \end{pmatrix}$$

7 Question III.2.2: Montrer que la matrice M est inversible

Montrer que M est inversible revient à montrer que M est injective. Posons $\alpha := a\frac{\triangle t}{2\triangle x}$ et $\beta := b\frac{\triangle t}{\triangle x^2}$.
En outre, on supposera que $\alpha \neq \beta$ sans quoi M est triangulaire inférieure à diagonale strictement positive donc inversible.
Donnons nous une combinaison linéaire des colonnes C_i de M nulle:

$$\exists\{\lambda_1, \cdots, \lambda_m\}, \lambda_1 C_1 + \cdots + \lambda_m C_m = 0$$

Si λ_i est non nul, cela impose que λ_{i-1} ou λ_{i-2} sont non nuls,
en vertu du fait que seuls les colonnes C_{i-1} et C_{i-2} de M ont leurs $(i-1)^{ieme}$ composantes non nulles, capables d'annuler ces mêmes composantes dans C_i.
En remontant ce constat, on obtient λ_1 non nul (avec les mêmes considérations vers les coefficients croissant on montre que λ_m non nul).
En effet, ce qui a été dit ci-dessus permet par récurrence immédiate de dire que λ_2 ou λ_1 non nul. Mais si $\lambda_2 \neq 0$, afin d'annuler la 1^{ere} composante de la combinaison linéaire, il faut avoir $\lambda_1 \neq 0$.
Conséquemment, si l'on suppose dorénavant que l'un des λ_i est non nul, on posera sans restriction $\lambda_1 = 1$. Alors la première composante de la combinaison linéaire vaut $(1 + 2\beta) + \lambda_2(\alpha - \beta)$.
Ainsi nécessairement $\lambda_2 = -\frac{1+2\beta}{\alpha-\beta}$. Puis, pour $k \geqslant 2$, afin d'annuler la $(k-1)^{ieme}$ composante de C_k, on a la relation:

$$\lambda_k(\alpha - \beta) + \lambda_{k-1}(1 + 2\beta) + \lambda_{k-2}(-\alpha - \beta) = 0$$

On calcule le discriminant du trinôme associée à cette récurrence d'ordre 2, qui vaut $\triangle = 1 + 4\beta + 4\alpha^2 > 0$.
La solution générale s'écrit donc $\lambda_k = \sigma_1(\frac{-(1+2\beta)+\sqrt{\triangle}}{2(\alpha-\beta)})^k + \sigma_2(\frac{-(1+2\beta)-\sqrt{\triangle}}{2(\alpha-\beta)})^k$.
Les conditions initiales sur λ_1 et λ_2 permettent d'obtenir les constantes de la forme générale. Mais observons d'ors et déjà que si l'on se donne la dernière colonne de M, la seule colonne capable d'annuler la n^{eme} composante de celle-ci est la colonne C_{n-1}, avec la relation sous-jacente:

$$\lambda_n(1 + 2\beta) + \lambda_{n-1}(-\alpha - \beta) = 0$$

Passons sous silence le calcul fastidieux des constantes σ_1 et σ_2. Il n'est néanmoins pas difficile de voir que, les ayant calculées, on obtient alors une valeur bien déterminée des coefficients λ_n et λ_{n-1}, incompatible avec la relation ci-dessus.
Cela impose que les coefficients de la combinaison linéaire sont tous nuls et ainsi que la matrice M est inversible.

8 Question IV.1.1: Réécriture du schéma implicite linéaire sur l'équation de la chaleur non linéaire

On se donne le schéma implicite linéarisé suivant:

$$\frac{u_i^{n+1}-u_i^n}{\triangle t} - b\frac{u_{i+1}^{n+1}+u_{i-1}^{n+1}-2u_i^{n+1}}{\triangle x^2} + u_i^{n+1}(u_i^n)^3 = 0$$

avec les conditions aux bords de Dirichlet $u^n_{-i_{max}-1} = u^n_{i_{max}+1} = 0$. Passons les valeurs ultérieures sur le terme de droite ainsi que les valeurs courantes au cube:

$$u_i^{n+1}(\triangle t(u_i^n)^3 + 2\frac{b\triangle t}{\triangle x^2}+1) + u_{i-1}^{n+1}(\frac{-b\triangle t}{\triangle x^2}) + u_{i+1}^{n+1}(\frac{-b\triangle t}{\triangle x^2}) = u_i n$$

Le système se réécrit donc matriciellement, avec les conditions aux bords,

$$U^n = (A + \triangle t diag((u^n_{-i_{max}})^3, \cdots, (u^n_{i_{max}})^3))U^{n+1}$$

avec la matrice A triadiagonale usuelle de l'équation de la chaleur:

$$\begin{pmatrix} 1+2b\frac{\triangle t}{\triangle x^2} & -b\frac{\triangle t}{\triangle x^2} & 0 & \cdots & 0 \\ -b\frac{\triangle t}{\triangle x^2} & \ddots & \ddots & 0 & \\ 0 & \ddots & & & \\ \vdots & & & & \\ 0 & & & & \end{pmatrix}$$

9 Question IV.2.1: Réécriture du schéma explicite centré sur l'équation de la chaleur en régime stationnaire

On se donne, avec les mêmes conditions de Dirichlet au bords que dans la partie précédente, le schéma suivant:

$$\forall i \in [-i_{max}, i_{max}], -b\frac{u_{i+1}+u_{i-1}-2u_i}{\triangle x^2} + (u_i)^4 = Q_i$$

Cela revient à résoudre $G(u_{-i_{max}}, \cdots, u_{-i_{max}}) = 0$ où $G : \mathbb{R}^{2i_{max}+1} \to \mathbb{R}^{2i_{max}+1}$ a pour i^{ieme} composante $-b\frac{u_{i+1}+u_{i-1}-2u_i}{\triangle x^2} + (u_i)^4 - Q_i$. Cette fonction est évidemment C^2 car ces composantes sont toutes C^∞.