

Compte-rendu en *Méthodes numériques de base* Résultats sur l'identification de conductivité

Aurélien PEPIN, Antonin KLOPP-TOSSER

2 mai 2017

1 Méthode des différences finies

QUESTION 1. L'écriture sous forme matricielle du θ -schéma (10) exprime l'itération $k + 1$ du vecteur U en fonction de l'itération k . On isole donc, dans le θ -schéma, les termes en $u_i^{(k+1)}$.

$$\begin{aligned} & u_i^{(k+1)} - \mu\theta \left(C_{i+1/2} u_{i+1}^{(k+1)} - (C_{i+1/2} + C_{i-1/2}) u_i^{(k+1)} + C_{i-1/2} u_{i-1}^{(k+1)} \right) \\ &= \mu(1 - \theta) \left(C_{i+1/2} u_{i+1}^k - (C_{i+1/2} + C_{i-1/2}) u_i^k + C_{i-1/2} u_{i-1}^k \right) + u_i^k \\ \\ &\iff (1 + \theta\mu(C_{i+1/2} + C_{i-1/2})) u_i^{(k+1)} - \theta\mu(C_{i+1/2} u_{i+1}^{(k+1)} - C_{i-1/2} u_{i-1}^{(k+1)}) \\ &= (1 + (\theta - 1)\mu(C_{i+1/2} + C_{i-1/2})) u_i^k - (\theta - 1)\mu C_{i+1/2} u_{i+1}^k - (\theta - 1)\mu C_{i-1/2} u_{i-1}^k \end{aligned}$$

Sachant que $U^{(k)}$ est la matrice-colonne des $u_i^{(k)} \forall i \in \llbracket 1; n \rrbracket$, on identifie les termes un à un dans le membre de gauche puis dans le membre de droite :

$$\left(I + \theta\mu \begin{bmatrix} C_{1+1/2} + C_{1-1/2} & -C_{1+1/2} & & & 0 \\ & -C_{2-1/2} & \ddots & \ddots & \\ & & \ddots & \ddots & -C_{(n-1)+1/2} \\ 0 & & & -C_{n-1/2} & C_{n+1/2} + C_{n-1/2} \end{bmatrix} \right) \begin{bmatrix} u_1^{(k+1)} \\ u_2^{(k+1)} \\ \vdots \\ u_n^{(k+1)} \end{bmatrix} = \begin{bmatrix} u_1^k \\ u_2^k \\ \vdots \\ u_n^k \end{bmatrix} + \mu \begin{bmatrix} B_1^{(k)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$A \qquad U^{(k+1)} \qquad B^{(k)}$

$$\left(I + (\theta - 1)\mu \begin{bmatrix} C_{1+1/2} + C_{1-1/2} & -C_{1+1/2} & & & 0 \\ & -C_{2-1/2} & \ddots & \ddots & \\ & & \ddots & \ddots & -C_{(n-1)+1/2} \\ 0 & & & -C_{n-1/2} & C_{n+1/2} + C_{n-1/2} \end{bmatrix} \right) \begin{bmatrix} u_1^{(k)} \\ u_2^{(k)} \\ \vdots \\ u_n^{(k)} \end{bmatrix} = \begin{bmatrix} u_1^{(k+1)} \\ u_2^{(k+1)} \\ \vdots \\ u_n^{(k+1)} \end{bmatrix} + \mu \begin{bmatrix} B_1^{(k)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$A \qquad U^{(k)} \qquad B^{(k)}$

$B_1^{(k)}$ exprime les conditions aux limites. Pour $i = 1$, le terme $(A)_{i,i-1} = (A)_{1,0}$ n'est pas défini et vaut 0 par convention. Pour conserver l'égalité, on ajoute dans B un coefficient simulé :

$$B_1^{(k)} = C_{1/2} (\theta u_0(t_k) + (1 - \theta) u_0(t_k + 1))$$

QUESTION 2. Une matrice M est dite définie positive si, pour toute matrice-colonne non nulle x à coefficients réels, ${}^t x M x > 0$. Soit $x = (x_1, x_2, \dots, x_n)$ non nul :

$${}^t x A x = \sum_{i=1}^n (C_{i+1/2} + C_{i-1/2}) x_i^2 - 2 \sum_{i=1}^{n-1} C_{i+1/2} x_i x_{i+1}$$


On pose $j = i - 1$ en développant le premier terme :

$$\begin{aligned} {}^t x A x &= \sum_{i=1}^n C_{i+1/2} x_i^2 + \sum_{j=0}^{n-1} C_{j+1/2} x_{j+1}^2 - 2 \sum_{i=1}^{n-1} C_{i+1/2} x_i x_{i+1} \\ &= \sum_{i=1}^{n-1} C_{i+1/2} (x_i - x_{i+1})^2 + C_{1/2} x_{1/2}^2 + C_{n+1/2} x_n^2 \end{aligned} \quad (1)$$

La fonction $C(x)$ est strictement positive par définition donc ${}^t x A x > 0 \forall x$ non nul.
La matrice A est bien définie positive.

2 Factorisation de Cholesky dans le cas tridiagonal

QUESTION 3. Phase de factorisation de Cholesky.

 Se référer à : Q3_Factorisation_Cholesky.sce

La décomposition de Cholesky permet ensuite de résoudre des systèmes linéaires en une phase de descente puis une phase de remontée.

QUESTION 4. Première phase de descente.

 Se référer à : Q4_Descente_Cholesky.sce

QUESTION 5. Seconde phase de remontée.

 Se référer à : Q5_Remontee_Cholesky.sce

On peut appliquer dans l'ordre ces trois fonctions pour résoudre l'exemple ci-dessous, noté $MX = B$. Le TD 5 a montré que la matrice M était aussi symétrique définie positive donc elle admet une unique factorisation de Cholesky.

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} \text{ où la solution est } X = \begin{pmatrix} 350 \\ 340 \\ 310 \\ 250 \\ 210 \end{pmatrix}$$

 Pour calculer et vérifier cet exemple, cf. Q5_Exemple_Cholesky.sce

3 Problème stationnaire

QUESTION 6. L'équation (9) de l'énoncé fournit une approximation du système (13) :

$$\frac{\partial}{\partial x} [C(x) \frac{\partial u}{\partial x}] (x_i, t_k) \approx \frac{C_{i+1/2} u_{i+1}^{(k)} - (C_{i+1/2} + C_{i-1/2}) u_i^{(k)} + C_{i-1/2} u_{i-1}^{(k)}}{\delta_x^2}$$

En écrivant ce système sous forme matricielle pour tout $\forall i \in \llbracket 1; n \rrbracket$, on obtient :

$$\frac{\partial}{\partial x} [C(x) \frac{\partial u}{\partial x}] = \frac{1}{\delta_x^2} (-Au + B) = 0 \iff Au = B$$

où, par identification :

- A est la matrice tridiagonale, symétrique et définie positive de la **question 1** ;
- u est la matrice-colonne des $u_i^{(k)}$ à l'itération k ;
- B est la matrice-colonne des conditions aux limites déterminée de la façon suivante :

Pour $i = 1$.

$$\begin{aligned} (Au)_1 &= -C_{3/2} u_2^{(k)} + (C_{3/2} + C_{1/2}) u_1^{(k)} \\ \iff (B)_1 &= C_{1/2} u_0 \end{aligned}$$

Pour $i > 1$.


$$\begin{aligned} (Au)_i &= -C_{i+1/2} u_{i+1}^{(k)} + (C_{i+1/2} + C_{i-1/2}) u_i^{(k)} - C_{i-1/2} u_{i-1}^{(k)} \\ &= \frac{\delta_x^2}{\delta_x^2} (-C_{i+1/2} u_{i+1}^{(k)} + (C_{i+1/2} + C_{i-1/2}) u_i^{(k)} - C_{i-1/2} u_{i-1}^{(k)}) \\ &\simeq \delta_x^2 \cdot 0 \\ \iff (B)_i &= 0 \quad \forall i \in \llbracket 2; n \rrbracket \end{aligned}$$

La matrice-colonne B est celle de la **question 1** avec la fonction u_0 constante.

Puisque A est définie positive, elle n'admet pas 0 comme valeur propre.

Elle est donc inversible et le système $Au = B$ possède par conséquent une unique solution.

QUESTION 7. Solution exacte du problème stationnaire et convergence.

 Se référer à : `Q7_Probleme_stationnaire.sce`

$$\begin{aligned} \frac{\partial}{\partial x} [C(x) \frac{\partial u}{\partial x}] = 0 &\iff -\frac{1}{l} \exp(-\frac{x}{l}) \frac{\partial u}{\partial x} + \exp(-\frac{x}{l}) \frac{\partial^2 u}{\partial x^2} = 0 \\ &\iff \frac{\partial^2 u}{\partial x^2} - \frac{1}{l} \frac{\partial u}{\partial x} = 0 \end{aligned}$$

donc $u(x)$ est de la forme :

$$u(x) = \alpha + \beta \exp(\frac{x}{l})$$

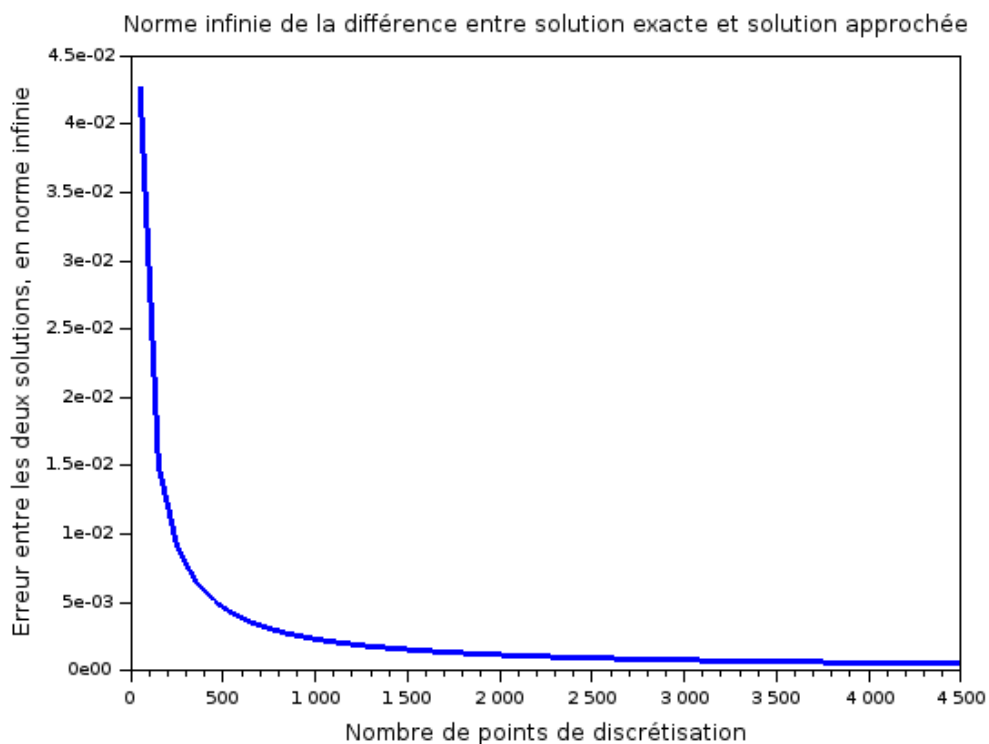
Détermination des constantes α et β . À partir des conditions initiales, on résout :

$$\begin{cases} u(-l) = \alpha + \beta e^{-1} = u_0 = 1 \\ u(l) = \alpha + \beta e = 0 \end{cases} \iff \begin{cases} \beta = \frac{1}{e^{-1}-e} \\ \alpha = -\frac{e}{e^{-1}-e} \end{cases}$$

Finalement, la solution exacte de l'équation différentielle (13) est :

$$u(x) = -\frac{e}{e^{-1}-e} + \frac{1}{e^{-1}-e} \exp\left(\frac{x}{l}\right) = \frac{e(e - e^{\frac{x}{l}})}{e^2 - 1}$$

Il est possible de mesurer la précision de la méthode des différences finies en comparant cette solution exacte avec la solution approchée pour différents pas $\delta_x \rightarrow 0$.



Plus le nombre de points de discrétisation x_i augmente, plus le pas δ_x tend vers 0. La solution numérique converge vers la solution exacte quand $\delta_x \rightarrow 0$.

4 Évolution d'une donnée initiale

QUESTION 8. On cherche à montrer que $\rho(M^{-1}N)$ est strictement inférieur à 1. On applique pour cela à M^{-1} et à N le résultat suivant, démontré plus bas.

Proposition. Soit P un polynôme d'endomorphisme (ou de matrice), soit S une matrice. Si v est vecteur propre de S pour la valeur propre λ , alors il est vecteur propre de $P(S)$ pour la valeur $P(\lambda)$.

Soient P et Q deux polynômes de matrice. Dire en effet que v est vecteur propre de S associé à la valeur propre λ revient à écrire :

$$\begin{aligned}
Sv = \lambda v &\iff S^p v = S^{p-1} \cdot Sv = S^{p-1} \cdot \lambda v = \lambda^p v \quad \forall p \in \mathbb{N} \\
&\iff P(S)v = \sum_{k=0}^p a_k S^k(v) = \sum_{k=0}^p a_k \lambda^k v = P(\lambda)v \\
&\iff P^{-1}(S)v = \frac{1}{P(\lambda)}v \\
&\iff P^{-1}(S) Q(S)v = \frac{Q(\lambda)}{P(\lambda)}v
\end{aligned}$$

Soit x_A un vecteur propre de A et λ_A sa valeur propre associée.

M^{-1} et N peuvent être exprimés comme des polynômes de matrice :

$$\begin{aligned}
M^{-1} = \theta\mu A + I &\iff M^{-1}(A) = \theta\mu A^1 + A^0 \\
N = (\theta - 1)\mu A + I &\iff N(A) = (\theta - 1)\mu A^1 + A^0
\end{aligned}$$

On en déduit donc la forme des valeurs propres de M et de N en fonction de celles de A :

$$\lambda_M = 1 + \theta\mu\lambda_A \quad \text{et} \quad \lambda_N = 1 + (\theta - 1)\mu\lambda_A$$

Les valeurs propres de $M^{-1}N$ sont donc de la forme :

$$\frac{\lambda_N}{\lambda_M} = \frac{1 + (\theta - 1)\mu\lambda_A}{1 + \theta\mu\lambda_A} = \frac{1 + \theta\mu\lambda_A - \mu\lambda_A}{1 + \theta\mu\lambda_A}$$

Or, $\mu > 0$ et A est définie positive donc $\lambda_A > 0$, d'où $\mu\lambda_A > 0$.

Le quotient est donc strictement inférieur à 1.

De plus, $\theta \in [0; 1]$ donc le quotient est toujours strictement supérieur à 1.

Toutes les valeurs propres de $M^{-1}N$ sont comprises dans l'intervalle $] -1; 1[$ donc :

$$\rho(M^{-1}N) = \max_i |\lambda_i| < 1$$


QUESTION 9. D'après le chapitre 3 du cours, la méthode de Crank-Nicolson converge si et seulement si $\rho(M^{-1}N) < 1$. La **question 8** a montré que c'était le cas.

Soit X la limite des $U^{(k)}$. Pour $k \rightarrow +\infty$, on a :

$$\begin{aligned}
MX = NX + \mu B &\iff (M - N)X = \mu B \\
&\iff (I + \theta\mu A - I - (\theta - 1)\mu A)X = \mu B \\
&\iff \mu AX = \mu B \\
&\iff AX = B
\end{aligned}$$

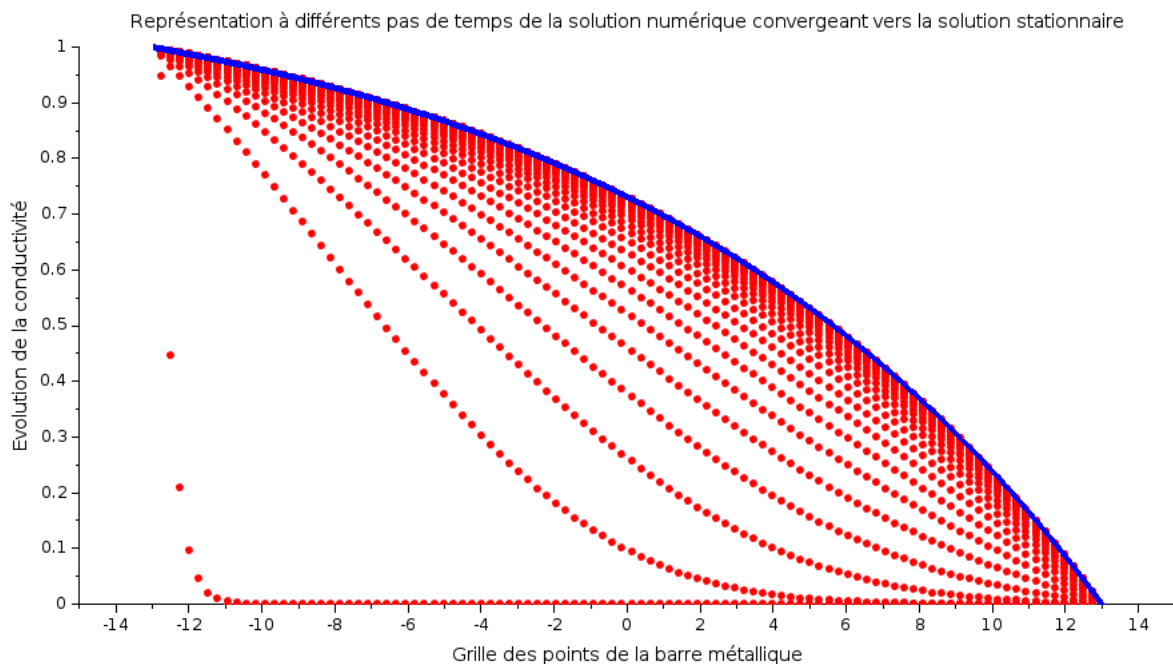
A et B sont les matrices de la **question 6** donc la solution discrétisée converge bien vers la solution stationnaire.

QUESTION 10. Dans le graphique ci-dessous, on trace en bleu la solution stationnaire trouvée dans la **question 6**. Puis on trace en rouge la solution approchée à différents pas t_k .

 Se référer à : Q10_CrankNicolson.sce

On choisit les paramètres suivants :


- $n = 100$, nombre de points de discrétisation de la grille ;
- $n_t = 6000$, nombre d'itérations assez grand pour approcher de près la solution exacte ;
- $T = 500$, ce qui donne un pas de temps $\delta_t \simeq 0.083$.



Quand $k \rightarrow +\infty$, $U^{(k)}$ tend effectivement vers la solution stationnaire. La convergence de la méthode de Crank-Nicolson est rapide : sur le graphique, la deuxième mesure, déjà éloignée de $U^{(0)}$, a été relevée à $k = 200$. Chaque courbe rouge correspond à 200 itérations supplémentaires.

5 Étude du problème inverse

QUESTION 11. Mesure du flux à $t_{inter} = \frac{2}{3}T$ et $t_{fin} = T$.

 Se référer à : Q11_Flux.sce

La fonction `flux(x_d)` est gourmande en temps. Sa complexité temporelle est en $O(n_t \cdot n)$ mais elle cache trois étapes en $O(n)$: le calcul du second membre, la descente et la remontée.

On en optimise pour cela quelques parties comme le calcul du vecteur $B^{(k)}$ ramené au calcul de sa seule composante non nulle $B_1^{(k)}$. On économise ainsi de l'espace mémoire et n itérations pour l'affectation.

Annexes

6 Sensibilisation à l'arithmétique machine

Exercice 1

Avec $x = 1^{30}$ et $y = 1^{-8}$, les résultats attendus sont $z = w = 1$.

Or, ces calculs sous **Scilab** donnent bien $w = 1$ mais aussi $z = 0$. Étudions la commande :

```
--> x + y
```

Le résultat obtenu est erroné, il vaut x . En effet, le résultat de l'addition requiert **38 chiffres significatifs** ce qui est supérieur à la précision arbitraire permise par la commande `format(20)`. Seuls les premiers chiffres sont pris en compte, c'est une conséquence de la limite de la taille de la mantisse avec la norme IEEE 754 utilisée dans **Scilab**.

Même s'ils devraient mathématiquement être égaux à 1, l'ordre des opérations dans z et dans w altère le résultat. Le fait de mettre des parenthèses va indiquer à la machine quelle opération faire en premier.

- Dans le cas de z , $(y + x) = x$ à cause du manque de précision et donc $(y + x) - x = 0$;
- Dans le cas de w , $(x - x) = 0$ donc $\frac{y+(x-x)}{y} = 1$.

Exercice 2

La figure 1 ci-dessous présente en bleu la valeur attendue du calcul de $f(x)$ et en vert la valeur calculée par **Scilab** sur l'intervalle $[0, 4]$. Si $f(x) = x$ pour tout x de l'intervalle, on voit que ce n'est pas ce qu'on obtient par le calcul. En effet :

- sur l'intervalle $[0, 1]$, $f(x) = 0$;
- sur l'intervalle $[1, 4]$, $f(x) = 1$.

Méthode. Nous avons calculé y et $f(x)$ grâce à une boucle de 128 itérations. Sur l'intervalle $[0, 1]$, le calcul de y donne un résultat très proche de 1 mais strictement inférieur. Ceci est dû à

FIGURE 1 – Graphique de l'exercice 2

la précision machine et aux algorithmes de calcul de la racine carrée qui ne fournissent pas une valeur exacte mais une valeur approchée par des polynômes.

Le problème est le même sur l'intervalle $[1, 4]$ où les écarts de précision introduits par la racine carrée sont amplifiés par la mise au carré. Le calcul de y donne un résultat égal à 1 alors qu'il devrait être strictement supérieur.

Ainsi, sur l'intervalle $[0, 1]$, $y^{256} = 0$, alors que sur l'intervalle $[1, 4]$, $y = 1$ donc $y^{256} = 1$.

Exercice 3

1. Grâce à une intégration par parties, on établit les égalités suivantes :

$$\begin{aligned} I_n &= \int_0^1 x^n e^x dx \\ &= [x^n e^x]_0^1 - \int_0^1 n x^{n-1} e^x dx \\ &= e - n \int_0^1 x^{n-1} e^x dx \\ &= e - n I_{n-1} \end{aligned}$$

Ce qui conduit à une suite définie par :

$$\begin{cases} I_n &= e - n I_{n-1} \\ I_0 &= e - 1 \end{cases}$$

Évaluation. On implémente la récurrence dans **Scilab** grâce à une fonction récursive. Le résultat de l'appel de la fonction `integrale1(20)` vaut environ -129.264. La fonction $x^n e^x$ est positive entre 0 et 1 quel que soit n . L'intégrale d'une fonction positive étant aussi positive, on en déduit que ce résultat est faux.

La valeur de I_n devient en effet très petite au fur et à mesure que n croît. La précision de calcul limitée fausse les résultats, il faut donc choisir une autre méthode.

2. D'après le développement en série de e^x , on sait que :

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \implies I_n = \int_0^1 x^n e^x dx = \int_0^1 \sum_{n=0}^{+\infty} \frac{x^n}{n!} dx$$

La suite $u_n = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$ est positive et croissante. D'après le **théorème de convergence monotone**, il est possible d'intervertir la somme et l'intégrale :

$$\begin{aligned}
\int_0^1 \sum_{n=0}^{+\infty} \frac{x^n}{n!} dx &= \sum_{n=0}^{+\infty} \int_0^1 \frac{x^n}{n!} dx \\
&= \sum_{n=0}^{+\infty} \frac{1}{n!} \left[\frac{x^{21+n}}{21+n} \right]_0^1 \\
&= \sum_{n=0}^{+\infty} \frac{1}{n!(21+n)}
\end{aligned}$$

Évaluation. Le terme général de la série ci-dessus tend rapidement vers 0. On obtiendra donc une bonne approximation de l'intégrale sans calculer énormément de termes. Soit N le nombre de termes calculés. Pour $N = 10$, l'appel à `integrale2(N)` vaut environ 0.1238.

3. Le résultat de la fonction `integrale2` est plus crédible que celui de la fonction `integrale1`. Dans le calcul par récurrence, les erreurs de précision se propagent au fur et à mesure des appels et faussent beaucoup le résultat. Dans le calcul itératif, les erreurs de précision sont minimales et n'ont plus d'influence quand le terme tend vers zéro.

La fonction `integrate` de **Scilab** permet de vérifier que le calcul itératif est correct.

Exercice 4

En subdivisant l'intervalle d'intégration en N points, on peut obtenir une approximation de l'intégrale I_n par la méthode des rectangles à gauche ou à droite. On note I^G l'approximation par la gauche et I^D l'approximation par la droite.

- Pour $N = 100$ points, $I^G = 0.1106$ et $I^D = 0.1379$
- Pour $N = 1000$ points, $I^G = 0.1224$ et $I^D = 0.1251$
- Pour $N = 10000$ points, $I^G = 0.1236$ et $I^D = 0.1239$
- Pour $N = 100000$ points, $I^G = 0.1237$ et $I^D = 0.1238$

À gauche comme à droite, les valeurs de la méthode des rectangles tendent bien vers 0.1238.

FIGURE 2 – Graphique de l'exercice 4

7 Étude du phénomène de Gibbs

Exercice 5

Pour calculer la série de Fourier de la fonction $f(x)$, on détermine ses coefficients de Fourier.

Note. Comme la fonction f est impaire, les a_n **sont nuls**. En effet, la fonction $\cos(x)$ est paire et le produit d'une fonction paire et d'une fonction impaire reste impair. Puisque l'intervalle sur lequel on intègre est centré en zéro, la partie gauche (< 0) et la partie droite (> 0) de l'intégrale s'annulent.

On procède alors au calcul des $b_n(f)$.

$$\begin{aligned}
b_n(f) &= \frac{2}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t) \sin(2\pi \frac{n}{T} t) dt \\
&= 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t) \sin(2\pi n t) dt \text{ car } T = 1 \\
&= -2 \int_{-\frac{1}{2}}^0 \sin(2\pi n t) dt + 2 \int_0^{\frac{1}{2}} \sin(2\pi n t) dt \\
&= -2 \left[\frac{-\cos(2\pi n t)}{2\pi n} \right]_{-\frac{1}{2}}^0 + 2 \left[\frac{-\cos(2\pi n t)}{2\pi n} \right]_0^{\frac{1}{2}} \\
&= \frac{1}{\pi n} (1 - \cos(\pi n)) - \frac{1}{\pi n} (\cos(\pi n) - 1) \\
&= \frac{1}{\pi n} (1 - (-1)^n) - \frac{1}{\pi n} ((-1)^n - 1) \\
&= \frac{2}{\pi n} (1 - (-1)^n)
\end{aligned}$$

Grâce à ce résultat, la série de Fourier de la fonction f est :

$$\begin{aligned}
f(t) &= \sum_{n=0}^{+\infty} \frac{2}{\pi n} (1 - (-1)^n) \sin(2\pi n t) \\
&= \sum_{n=0}^{+\infty} \frac{4}{\pi(2n+1)} \sin(2(2n+1)\pi t) \text{ car } \forall n \in 2\mathbb{Z}, f(t) = 0
\end{aligned}$$

La fonction $f(x)$ est une fonction affine et dérivable par morceaux. Elle possède trois points de discontinuité : $\{-\frac{1}{2}, 0, \frac{1}{2}\}$. Sa série de Fourier est une fonction continue qui ne peut pas effectuer de « sauts », elle devient alors invalide en ces points. Ainsi, $f(0) = -1$, $f(-\frac{1}{2}) = -1$ et $f(\frac{1}{2}) = 1$ mais la série de Fourier passe par 0 pour ces trois points.

Exercice 6

La figure 3 illustre la représentation en série de Fourier *tronquée* de la fonction $f(x)$ définie dans l'exercice 5. La courbe bleue montre le résultat de la somme des 10 premiers **Ntermes** tandis que la courbe verte est générée à partir de 70 **Ntermes**. Plus le nombre de **Ntermes** croît, plus la série de Fourier approche la fonction $f(x)$.

Quel que soit le nombre de termes calculé, il se produit toutefois aux points de discontinuité du signal de fortes oscillations issues du **phénomène de Gibbs**.

FIGURE 3 – Graphique de l'exercice 6

Le phénomène de Gibbs donne l'intuition que la convergence de la série de Fourier quand $n \rightarrow \infty$ ne peut pas être uniforme sur $[-\frac{1}{2}; \frac{1}{2}]$. Plus formellement, le *théorème de Dirichlet* permet de conclure que la série de Fourier de la fonction f (dérivable par morceaux) :

- converge simplement vers la fonction f ;
- converge uniformément vers f sur tout $[a, b]$ qui ne contient pas de point de discontinuité ;
- converge vers $\frac{f(t^+) + f(t^-)}{2}$ pour tout point t de discontinuité.

8 Théorème de Gershgorin

Exercice 7

1. On appelle disque de Gershgorin, le disque D_k défini par :

$$D_k = \{z \in \mathbb{C} : |z - a_{kk}| \leq \Lambda_k = \sum_{j=1, j \neq k}^N |a_{kj}|\}$$

Soit la matrice A , carrée de taille N , telle que $A = (a_{ij})_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}}$.

Soit λ un vecteur propre de A et $v = (v_1, v_2, \dots, v_N)$ le vecteur propre associé.

D'après l'égalité $(A - \lambda I)v = 0$, on a :

$$\forall i \in \llbracket 1, N \rrbracket, (a_{ii} - \lambda)v_i + \sum_{\substack{j=1 \\ i \neq j}}^N a_{ij}v_j = 0 \text{ d'où on déduit que } |a_{ii} - \lambda v_i| = \sum_{\substack{j=1 \\ i \neq j}}^N |a_{ij}v_j|$$

On choisit i tel que $|v_i| = \sup_{k \in \llbracket 1, N \rrbracket} |v_k|$. On peut former le quotient $\frac{v_j}{v_i}$ car v est un vecteur propre, donc il est non nul. On a alors $v_i = \sup_{k \in \llbracket 1, N \rrbracket} |v_k| \neq 0$.

$$\begin{aligned} |a_{ii} - \lambda| &= \left| \sum_{\substack{j=1 \\ i \neq j}}^N a_{ij} \frac{v_j}{v_i} \right| \\ &\leq \sum_{\substack{j=1 \\ i \neq j}}^N \left| a_{ij} \frac{v_j}{v_i} \right| \\ &\leq \sum_{\substack{j=1 \\ i \neq j}}^N |a_{ij}| \text{ étant donné que } \frac{v_j}{v_i} \leq 1 \end{aligned}$$

À partir de cette inégalité, on vérifie que $\exists i \in \llbracket 1, N \rrbracket, \lambda \in D_i$ soit $\lambda \in \bigcup_{k=1}^N D_k$.

2. Dans **Scilab**, on représente un disque de Gershgorin D_k dans le plan complexe où :
 - Le centre (x_O, y_O) du disque vaut $(\Re(a_{kk}), \Im(a_{kk}))$;
 - Le rayon r du disque vaut Λ_k .
 La fonction **spec** permet d'obtenir les valeurs propres de la matrice A . La question suivante montre un exemple de génération des disques grâce à **Scilab**.

3. Ce graphique montre que les valeurs propres sont bien dans l'union des disques de A .

FIGURE 4 – Application du théorème de Gershgorin sur la matrice A

4. Soit A une matrice **strictement dominante** telle que $A = (a_{ij})_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}}$.

Comme A est strictement dominante, on a : $\forall i \sum_{k \neq i} |a_{ik}| < |a_{ii}|$.

Soit λ une valeur propre de A . D'après la question 1, $\lambda \in \bigcup_{k=1}^N D_k$. Alors :

$$\begin{aligned} \exists k : |\lambda - a_{kk}| &\leq \Lambda_k = \sum_{j=1, j \neq k}^N |a_{kj}| \\ |\lambda - a_{kk}| &\leq \Lambda_k = \sum_{j=1, j \neq k}^N |a_{kj}| < |a_{kk}| \\ |\lambda - a_{kk}| &< |a_{kk}| \end{aligned}$$

On en déduit donc que $\lambda \neq 0$. Puisque toutes les valeurs propres de la matrice A sont différentes de 0, la matrice A est **inversible**.

9 RESTE

Soit $x = (x_1, x_2, \dots, x_n)$

$$\begin{aligned} x^T A x &= \sum_{i=1}^n (C_{i+1/2} + C_{i-1/2}) x_i^2 - 2 \sum_{i=1}^{n-1} C_{i+1/2} x_i x_{i+1} \\ &= \sum_{i=1}^n C_{i+1/2} x_i^2 + \sum_{j=0}^{n-1} C_{j+1/2} x_{j+1}^2 - 2 \sum_{i=1}^{n-1} C_{i+1/2} x_i x_{i+1} \quad \text{On pose } j = i-1 \\ &= \sum_{i=1}^{n-1} C_{i+1/2} (x_i - x_{i+1})^2 + C_{1/2} x_1^2 + C_{n+1/2} x_n^2 > 0 \end{aligned}$$

Comme $x^T A x > 0 \forall x \in \mathbb{R}^n$ alors la matrice A est symétrique définie positive.

D'après (9) : $\frac{\partial}{\partial x} [C(x) \frac{\partial u}{\partial x}] (x_i, t_k) = \frac{C_{i+1/2} u_{i+1}^{(k)} - (C_{i+1/2} + C_{i-1/2}) u_i^{(k)} + C_{i-1/2} u_{i-1}^{(k)}}{\delta_x^2}$ Pour résoudre (13), on résout 9 pour chaque coefficient u_i^k

$$\begin{pmatrix} u_0 C_{1/2} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Question 6.

D'après l'équation (9),

$$\frac{\partial}{\partial x} [C(x) \frac{\partial u}{\partial x}] (x_i, t_k) \approx \frac{C_{i+1/2} u_{i+1}^{(k)} - (C_{i+1/2} + C_{i-1/2}) u_i^{(k)} + C_{i-1/2} u_{i-1}^{(k)}}{\delta_x^2}$$

Si on écrit ce système, $\forall i$ sous forme matricielle, on obtient :

$$\frac{\partial}{\partial x} [C(x) \frac{\partial u}{\partial x}] = \frac{1}{\delta^2} (-A u + \begin{pmatrix} u_0 C_{1/2} \\ 0 \\ 0 \\ 0 \end{pmatrix}) \quad \text{car } u(-l) = u_0 \text{ avec } u = (u_1^{(k)}, \dots, u_n^{(k)}) \text{ et } A \text{ la matrice}$$

trouvée à la question 1.

Donc $A u = B$, avec $B = (b_0, \dots, b_n)$ où, $b_0 = C_{1/2} u_0$ et $b_i = 0 \forall i \in [2, n]$

Ce système admet une solution unique car A est définie et donc inversible.

Question 7.

On résout l'équation :

$$\begin{cases} \frac{\partial}{\partial x}[C(x)\frac{\partial u}{\partial x}] = 0 \\ u(-l) = u_0 \\ u(l) = 0 \end{cases}$$

$$\frac{\partial}{\partial x}[C(x)\frac{\partial u}{\partial x}] = 0 \Rightarrow -\frac{1}{l}e^{-\frac{x}{l}}\frac{\partial u}{\partial x} + e^{-\frac{x}{l}}\frac{\partial^2 u}{\partial x^2} = 0 \Rightarrow \frac{\partial^2 u}{\partial x^2} - \frac{1}{l}\frac{\partial u}{\partial x} = 0$$

$$u(x) = \alpha + \beta e^{\frac{x}{l}}$$

Calcul des constantes :

$$\begin{cases} u(-l) = \alpha + \beta e^{-1} = u_0 = 1 \\ u(l) = \alpha + \beta e = 0 \end{cases}$$

$$\begin{cases} \beta = \frac{1}{e^{-1}-e} \\ \alpha = -\frac{e}{e^{-1}-e} \end{cases}$$

$$\text{Donc : } u(x) = -\frac{e}{e^{-1}-e} + \frac{1}{e^{-1}-e}e^{\frac{x}{l}}$$

Question 8.

$$M = I + \theta\mu A \text{ et } N = I + (\theta - 1)\mu A$$

M et N sont donc des polynomes de matrices.

Donc, soit x_A un vecteur propre de A et λ_A une valeur propre de A, alors x_A est aussi un vecteur propre de M avec comme valeur propre $M\lambda_A$

Les valeurs propres de M sont $\lambda_M = 1 + \theta\mu\lambda_A$.

Les valeurs propres de N sont $\lambda_N = 1 + (\theta - 1)\mu\lambda_A$

Les valeurs propres de $M^{-1}N$ sont $\frac{\lambda_N}{\lambda_M} = \frac{1+(\theta-1)\mu\lambda_A}{1+\theta\mu\lambda_A} \forall \lambda_A$. $1 + (\theta - 1)\mu\lambda_A < 1 + \theta\mu\lambda_A$

Donc $\rho(M^{-1}N) = \max(\frac{\lambda_N}{\lambda_M}) < 1$

Question 9.

$MU^{(k+1)} = NU^{(k)} + \mu B$ Cette méthode de décomposition converge uniquement si $\rho(M^{-1}N) < 1$, ce qui est le cas ici (question 8).

Cette décomposition converge vers $Ax = b$, avec $A = M - N$.

$$M - N = I + \theta\mu A - I - (\theta - 1)\mu A = \mu A$$

$\mu Ax = \mu B$ $Ax = B$ On retrouve ici les matrices A et B de la question (6).