



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 1

2009

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Partie I

1°) Matrice à diagonale strictement dominante.

Question adaptée au programme actuel : les matrices et les vecteurs sont réels. La preuve ne fait intervenir que la valeur absolue et l'inégalité triangulaire.

a) Puisque $X \neq 0$ et que $|x_1|$ est la plus grande valeur absolue des coordonnées, $|x_1| > 0$. La première ligne de $AX = 0$ donne

$$a_{1,1}x_1 = - \sum_{j=2}^n a_{1,j}x_j.$$

Par l'inégalité triangulaire,

$$|a_{1,1}| |x_1| \leq \sum_{j=2}^n |a_{1,j}| |x_j| \leq \left(\sum_{j=2}^n |a_{1,j}| \right) |x_1|.$$

On peut diviser par $|x_1| > 0$, ce qui donne

$$|a_{1,1}| \leq \sum_{j=2}^n |a_{1,j}|,$$

contraire à la dominance stricte. Il ne peut donc exister un tel vecteur.

Pour $n = 1$, la somme est vide et vaut 0; le raisonnement reste valable.

b) L'ensemble fini des valeurs $|x_i|$ possède un maximum. Choisissons i_0 tel que $|x_{i_0}| = \max_i |x_i| > 0$. La ligne d'indice i_0 de $AX = 0$ donne

$$|a_{i_0,i_0}| |x_{i_0}| \leq \sum_{j \neq i_0} |a_{i_0,j}| |x_j| \leq \left(\sum_{j \neq i_0} |a_{i_0,j}| \right) |x_{i_0}|.$$

La division par $|x_{i_0}|$ contredit la dominance stricte de cette ligne. Ainsi $AX = 0$ impose $X = 0$.

- c) La question b) montre que $\text{Ker } A = \{0\}$. L'endomorphisme associé est injectif sur l'espace de dimension finie $\mathbb{R}^n$, donc bijectif par le théorème du rang. La matrice A est inversible.

2°) Encadrement de Gershgorin pour les valeurs propres réelles.

Question adaptée au programme actuel : les disques sont remplacés par des intervalles réels. Le résultat porte sur toutes les valeurs propres réelles ; il ne présume pas qu'une matrice réelle quelconque soit diagonalisable sur $\mathbb{R}$.

a) Méthode 1 — Utilisation de la dominance stricte.

Soit $\lambda \in \mathbb{R}$ n'appartenant pas à D . Alors, pour tout i ,

$$|\lambda - a_{i,i}| > \sum_{j \neq i} |a_{i,j}|.$$

La matrice $A - \lambda I_n$ est à diagonale strictement dominante. Par la question 1°, elle est inversible, donc λ n'est pas valeur propre. Par contraposée, toute valeur propre réelle appartient à D .

Méthode 2 — Utilisation directe d'un vecteur propre.

Soit λ une valeur propre réelle et $X \neq 0$ tel que $AX = \lambda X$. Choisissons i_0 tel que $|x_{i_0}| = \max_j |x_j| > 0$. La ligne correspondante donne

$$(\lambda - a_{i_0,i_0})x_{i_0} = \sum_{j \neq i_0} a_{i_0,j}x_j.$$

Par l'inégalité triangulaire, $|\lambda - a_{i_0, i_0}| |x_{i_0}| \leq r_{i_0} |x_{i_0}|$. En divisant par $|x_{i_0}| > 0$, on obtient $\lambda \in D_{i_0} \subset D$.

La première méthode applique le résultat précédent ; la seconde fournit directement l'intervalle associé à une coordonnée maximale du vecteur propre.

b) Code adapté en Python. Question adaptée au programme actuel.

Les lignes et colonnes de NumPy sont indexées de 0 à 2. Les boucles `range(3)` parcourent ces trois indices. Les saisies sont supposées être des représentations valides de nombres réels.

Python

```
import numpy as np

def saisir_matrice():
    A = np.zeros((3, 3))
    for i in range(3):
        for j in range(3):
            A[i, j] = float(input("Coefficient de la matrice : "))
    return A

def gershgorin(A):
    centres = np.zeros(3)
    rayons = np.zeros(3)
    for i in range(3):
        centres[i] = A[i, i]
        for j in range(3):
            if j != i:
                rayons[i] = rayons[i] + abs(A[i, j])
    return centres, rayons
```

Les neuf saisies sont effectuées ligne par ligne, dans l'ordre de la première ligne à la troisième. La fonction de saisie renvoie une matrice NumPy de dimension 3×3 . La seconde fonction renvoie deux tableaux : les coefficients diagonaux et les sommes des

valeurs absolues des coefficients hors diagonale de chaque ligne.

Les deux fonctions ont été exécutées : neuf saisies contrôlées ont permis de vérifier la disposition des coefficients ; sur la matrice de la question c), les sorties sont les centres $(1, 2, 1)$ et les rayons $(1, 2, 1)$. Le code calcule des bornes ; cette vérification informatique ne remplace pas la preuve de a).

c) Étude de l'exemple.

- (i) La matrice A est réelle et symétrique. Le théorème spectral s'applique : elle est diagonalisable sur $\mathbb{R}$, et même dans une base orthonormée. Aucun calcul de valeur propre n'est nécessaire pour cette conclusion.
- (ii) Les intervalles sont $D_1 = [0, 2]$, $D_2 = [0, 4]$ et $D_3 = [0, 2]$. Leur réunion est $D = [0, 4]$. Toutes les valeurs propres de cette matrice symétrique sont réelles ; par a), elles appartiennent à $[0, 4]$.
- (iii) On recherche des vecteurs propres en utilisant la structure de la matrice, puis on vérifie directement leurs images.

Les sommes des lignes étant nulles, $v_0 = {}^t(1, 1, 1)$ vérifie $Av_0 = 0$.

Pour $v_1 = {}^t(1, 0, -1)$, le produit donne $Av_1 = {}^t(1, 0, -1) = v_1$.

Pour $v_3 = {}^t(1, -2, 1)$, le produit donne $Av_3 = {}^t(3, -6, 3) = 3v_3$.

Ces trois vecteurs propres, associés aux valeurs distinctes 0, 1 et 3, sont indépendants. Ils forment donc une base de $\mathbb{R}^3$. En posant

$$T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{pmatrix}, \quad D = \text{diag}(0, 1, 3),$$

on obtient $AT = TD$ et $A = TDT^{-1}$.

Leurs produits scalaires deux à deux sont nuls, et leurs normes valent respectivement $\sqrt{3}$, $\sqrt{2}$ et $\sqrt{6}$. On peut donc aussi poser

$$O = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}.$$

Ses colonnes sont orthonormées, donc $O^{-1} = {}^tO$ et $A = OD^tO$. Les valeurs propres sont exactement 0, 1 et 3, en accord avec l'encadrement précédent.

3°) La propriété ($\mathcal{P}$).

a) Puisque les coefficients hors diagonale sont négatifs ou nuls,

$$\sum_{j \neq i} |a_{i,j}| = - \sum_{j \neq i} a_{i,j}.$$

La positivité de la somme de la ligne donne

$$a_{i,i} > - \sum_{j \neq i} a_{i,j} = \sum_{j \neq i} |a_{i,j}|.$$

Comme $a_{i,i} > 0$, son module vaut lui-même. La matrice est donc à diagonale strictement dominante et inversible par 1° c).

b) Posons $x_{i_0} = \min_i x_i$. Si cette valeur était négative, on aurait, pour tout $j \neq i_0$,
 $x_j \geq x_{i_0}$ et $a_{i_0,j} \leq 0$, donc $a_{i_0,j}x_j \leq a_{i_0,j}x_{i_0}$.

La coordonnée d'indice i_0 de AX satisferait

$$(AX)_{i_0} = a_{i_0, i_0} x_{i_0} + \sum_{j \neq i_0} a_{i_0, j} x_j \leq x_{i_0} \sum_{j=1}^n a_{i_0, j} < 0,$$

car la somme de la ligne est strictement positive. Cela contredit l'hypothèse $AX \geq 0$ coordonnée par coordonnée. Ainsi $\min_i x_i \geq 0$, donc toutes les coordonnées de X sont positives ou nulles.

- c) La colonne $B_j = {}^t(b_{1,j}, \dots, b_{n,j})$ de A^{-1} est $A^{-1}e_j$, où e_j est le vecteur canonique d'indice j . Par conséquent $AB_j = AA^{-1}e_j = e_j$.
- d) Le vecteur e_j a toutes ses coordonnées positives ou nulles. La question b), appliquée à B_j , donne $b_{i,j} \geq 0$ pour tout i . Comme j est arbitraire, tous les coefficients de A^{-1} sont positifs ou nuls.
- e) Pour $\alpha > 0$, les coefficients diagonaux de A_α sont $1 + \alpha, 2 + \alpha, 1 + \alpha$, donc strictement positifs. Les coefficients hors diagonale sont inchangés et négatifs ou nuls. Les sommes des lignes valent toutes $\alpha > 0$. Ainsi A_α vérifie $(\mathcal{P})$ et est inversible.

Méthode 1 — Utilisation de la diagonalisation précédente.

Avec les vecteurs propres unitaires $w_0 = v_0/\sqrt{3}$, $w_1 = v_1/\sqrt{2}$ et $w_3 = v_3/\sqrt{6}$, on a

$$A_\alpha^{-1} = \frac{1}{\alpha} w_0 {}^t(w_0) + \frac{1}{\alpha + 1} w_1 {}^t(w_1) + \frac{1}{\alpha + 3} w_3 {}^t(w_3).$$

Les trois dénominateurs sont strictement positifs. En développant ces trois matrices, on

obtient

$$A_{\alpha}^{-1} = \frac{1}{3\alpha} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \frac{1}{2(\alpha+1)} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} + \frac{1}{6(\alpha+3)} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}.$$

Pour montrer la simplification, le coefficient d'indice (1, 1) vaut

$$\frac{1}{3\alpha} + \frac{1}{2(\alpha+1)} + \frac{1}{6(\alpha+3)} = \frac{\alpha^2 + 3\alpha + 1}{\alpha(\alpha+1)(\alpha+3)}.$$

Celui d'indice (1, 2) vaut $\frac{1}{3\alpha} - \frac{1}{3(\alpha+3)} = \frac{\alpha+1}{\alpha(\alpha+1)(\alpha+3)}$. Celui d'indice (1, 3) vaut

$$\frac{1}{3\alpha} - \frac{1}{2(\alpha+1)} + \frac{1}{6(\alpha+3)} = \frac{1}{\alpha(\alpha+1)(\alpha+3)}.$$

$$\frac{1}{3\alpha} + \frac{2}{3(\alpha+3)} = \frac{(\alpha+1)^2}{\alpha(\alpha+1)(\alpha+3)}.$$

Par symétrie, les autres coefficients s'en déduisent. Finalement

$$A_{\alpha}^{-1} = \frac{1}{\alpha(\alpha+1)(\alpha+3)} \begin{pmatrix} \alpha^2 + 3\alpha + 1 & \alpha + 1 & 1 \\ \alpha + 1 & (\alpha + 1)^2 & \alpha + 1 \\ 1 & \alpha + 1 & \alpha^2 + 3\alpha + 1 \end{pmatrix}.$$

Méthode 2 — Résolution d'un système par élimination.

Posons $t = \alpha + 1$ et $d = \alpha(\alpha + 1)(\alpha + 3) > 0$. Pour résoudre $A_{\alpha}X = Y$, avec $Y = {}^t(y_1, y_2, y_3)$, écrivons

$$tx_1 - x_2 = y_1, \quad -x_1 + (t+1)x_2 - x_3 = y_2, \quad -x_2 + tx_3 = y_3.$$

Les première et troisième équations donnent $x_1 = (y_1 + x_2)/t$ et $x_3 = (y_3 + x_2)/t$, car $t > 0$. La deuxième devient

$$(t(t+1) - 2)x_2 = ty_2 + y_1 + y_3.$$

Or $t(t+1) - 2 = \alpha(\alpha+3) > 0$. Par conséquent

$$x_2 = \frac{ty_1 + t^2y_2 + ty_3}{d}.$$

En remplaçant dans les expressions de x_1 et x_3 , on obtient

$$x_1 = \frac{(\alpha^2 + 3\alpha + 1)y_1 + ty_2 + y_3}{d}, \quad x_3 = \frac{y_1 + ty_2 + (\alpha^2 + 3\alpha + 1)y_3}{d}.$$

Ces coefficients donnent exactement la même matrice inverse. La première méthode utilise les espaces propres; la seconde fournit le résultat par des opérations de résolution de système. Tous les coefficients de la formule finale sont strictement positifs pour $\alpha > 0$, ce qui précise ici la conclusion générale de d).

Partie II — Convergence de suites de matrices

1°) Généralités.

- a)** Si $m(X_k - X) \rightarrow 0$, alors, pour tout i , $0 \leq |x_i^{(k)} - x_i| \leq m(X_k - X)$. Le théorème d'encadrement donne $x_i^{(k)} \rightarrow x_i$. C'est la convergence de X_k vers X au sens défini dans l'énoncé.

Réciproquement, supposons que chaque coordonnée converge. Pour tout $\varepsilon > 0$ et chaque i , il existe K_i tel que, pour $k \geq K_i$, $|x_i^{(k)} - x_i| < \varepsilon$. Comme il n'y a qu'un nombre fini d'indices, $K = \max_i K_i$ existe. Pour $k \geq K$, toutes les différences sont inférieures à ε , donc $m(X_k - X) < \varepsilon$. Ainsi $m(X_k - X) \rightarrow 0$.

- b)** Si $s(M_k - M) \rightarrow 0$, chaque terme vérifie $0 \leq |m_{i,j}^{(k)} - m_{i,j}| \leq s(M_k - M)$, donc converge vers 0. Ainsi $M_k \rightarrow M$.

Réciproquement, si chaque coefficient converge, chaque suite $|m_{i,j}^{(k)} - m_{i,j}|$ tend vers 0. La somme porte sur les n^2 indices d'un ensemble fini ; par les opérations sur les limites de sommes finies, leur somme $s(M_k - M)$ tend vers 0.

c) Pour chaque i , la coordonnée d'indice i de MX est

$$\sum_j m_{i,j} x_j.$$

L'inégalité triangulaire donne

$$|(MX)_i| \leq \sum_{j=1}^n |m_{i,j}| |x_j| \leq m(X) \sum_{j=1}^n |m_{i,j}| \leq m(X) s(M).$$

En prenant le maximum sur i , on obtient $m(MX) \leq s(M)m(X)$.

d) Fixons X . Par c), $0 \leq m(M_k X - MX) = m((M_k - M)X) \leq s(M_k - M)m(X)$. La question b) donne $s(M_k - M) \rightarrow 0$; $m(X)$ est une constante finie. Par encadrement, $m(M_k X - MX) \rightarrow 0$, puis la question a) donne $M_k X \rightarrow MX$.

e) Soit e_j le vecteur canonique d'indice j . Le vecteur $M_k e_j$ est la colonne d'indice j de M_k . L'hypothèse appliquée à $X = e_j$ entraîne donc la convergence de tous les coefficients de cette colonne vers ceux de $M e_j$.

En faisant varier j de 1 à n , on obtient la convergence de chaque coefficient de M_k vers celui de M . Ainsi $M_k \rightarrow M$.

f) Écrivons $M = (m_{i,\ell})$ et $N = (n_{\ell,j})$. Tous les indices des sommes sont finis. L'inégalité triangulaire donne

$$s(MN) \leq \sum_{i=1}^n \sum_{j=1}^n \sum_{\ell=1}^n |m_{i,\ell}| |n_{\ell,j}| = \sum_{\ell=1}^n \left(\sum_i |m_{i,\ell}| \right) \left(\sum_j |n_{\ell,j}| \right).$$

Pour chaque ℓ ,

$$\sum_j |n_{\ell,j}| \leq s(N).$$

Donc le dernier membre est au plus

$$s(N) \sum_{\ell,i} |m_{i,\ell}| = s(N)s(M).$$

On a bien $s(MN) \leq s(M)s(N)$.

g) Pour tout i , $|y_i + z_i| \leq |y_i| + |z_i| \leq m(Y) + m(Z)$. En prenant le maximum, on obtient $m(Y + Z) \leq m(Y) + m(Z)$.

Appliquons cette inégalité à $Y = (Y - Z) + Z$: $m(Y) - m(Z) \leq m(Y - Z)$. En échangeant Y et Z , on obtient $m(Z) - m(Y) \leq m(Z - Y) = m(Y - Z)$. Ces deux inégalités donnent $|m(Y) - m(Z)| \leq m(Y - Z)$.

2°) Convergence de la suite des inverses.

a) Pour X fixé, on utilise une identité de différence des inverses, vérifiée par multiplication :

$$M_k^{-1}X - M^{-1}X = M^{-1}(M - M_k)M_k^{-1}X.$$

En effet, le membre de droite vaut $M_k^{-1}X - M^{-1}M_kM_k^{-1}X = M_k^{-1}X - M^{-1}X$.

En appliquant deux fois II-1° c), on obtient

$$m(M_k^{-1}X - M^{-1}X) \leq s(M^{-1})s(M - M_k)m(M_k^{-1}X).$$

Posons $\eta_k = s(M^{-1})s(M - M_k) \geq 0$. L'inégalité triangulaire II-1° g) donne

$$m(M_k^{-1}X) \leq m(M^{-1}X) + m(M_k^{-1}X - M^{-1}X) \leq m(M^{-1}X) + \eta_k m(M_k^{-1}X).$$

En regroupant les termes, on obtient $m(M_k^{-1}X)(1 - \eta_k) \leq m(M^{-1}X)$. À ce stade, on n'a pas divisé par $1 - \eta_k$, dont le signe n'est pas encore contrôlé.

b) Comme $M_k \rightarrow M$, II-1° b) donne $s(M - M_k) \rightarrow 0$, donc $\eta_k \rightarrow 0$. Il existe k_0 tel que, pour $k > k_0$, $0 \leq \eta_k \leq \frac{1}{2}$.

On a alors $1 - \eta_k \geq \frac{1}{2} > 0$, et l'inégalité précédente entraîne $m(M_k^{-1}X) \leq 2m(M^{-1}X)$. Le même k_0 convient à tous les vecteurs X , car η_k ne dépend pas de X .

c) Pour $k > k_0$, les questions a) et b) donnent

$$0 \leq m(M_k^{-1}X - M^{-1}X) \leq 2\eta_k m(M^{-1}X).$$

Le membre de droite tend vers 0. Le théorème d'encadrement, puis II-1° a), montrent que $M_k^{-1}X \rightarrow M^{-1}X$ pour tout X .

d) Le résultat de c) est valable pour chaque $X \in \mathbb{R}^n$. La réciproque II-1° e), appliquée aux matrices M_k^{-1} et M^{-1} , donne donc $M_k^{-1} \rightarrow M^{-1}$ coefficient par coefficient.

3°) Par I-3° a), les matrices M_k sont toutes inversibles. Par I-3° d), les coefficients de M_k^{-1} sont tous positifs ou nuls. Les hypothèses de II-2° sont satisfaites : $M_k \rightarrow M$ et M est inversible. Ainsi $M_k^{-1} \rightarrow M^{-1}$.

Chaque coefficient de M^{-1} est donc la limite d'une suite de réels positifs ou nuls. Le passage à la limite dans les inégalités donne que ce coefficient est positif ou nul. Cela prouve la conclusion, même si M ne vérifie pas elle-même les inégalités strictes de ($\mathcal{P}$).

4°) Étude de la matrice tridiagonale.

a) Le produit scalaire canonique donne

$$\langle AX, X \rangle = 2 \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^{n-1} x_i x_{i+1}.$$

On développe la somme des différences au carré : pour $n \geq 2$,

$$\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 = x_1^2 + x_n^2 + 2 \sum_{i=2}^{n-1} x_i^2 - 2 \sum_{i=1}^{n-1} x_i x_{i+1}.$$

En ajoutant $x_1^2 + x_n^2$, on retrouve le produit scalaire. Ainsi

$$\langle AX, X \rangle = x_1^2 + x_n^2 + \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2.$$

Pour $n = 1$, la somme est vide et la formule donne $2x_1^2$, ce qui est bien $\langle AX, X \rangle$.

Si $AX = 0$, ce produit scalaire est nul. Tous les termes de la somme de carrés sont positifs ou nuls ; ils sont donc tous nuls. On a $x_1 = 0$, et toutes les différences successives sont nulles, d'où $x_i = 0$ pour chaque i . Ainsi $\text{Ker } A = \{0\}$ et A est inversible.

b) Pour $\alpha > 0$, les coefficients diagonaux de A_α valent $2 + \alpha > 0$, et ses coefficients hors diagonale valent -1 ou 0 .

Pour $n \geq 2$, les sommes des lignes extrêmes valent $1 + \alpha > 0$. Les sommes des lignes intérieures, lorsqu'elles existent, valent $\alpha > 0$. Pour $n = 1$, la seule somme de ligne vaut $2 + \alpha > 0$. La propriété $(\mathcal{P})$ est donc vérifiée dans tous les cas.

c) Pour $k \in \mathbb{N}$, posons $M_k = A + \frac{1}{k+1} I_n$. Comme $\frac{1}{k+1} > 0$, toutes les matrices M_k vérifient $(\mathcal{P})$ par b). Chaque coefficient converge vers le coefficient correspondant de A puisque $\frac{1}{k+1} \rightarrow 0$. Ainsi $M_k \rightarrow A$.

d) La matrice A est inversible par a), et la suite construite en c) converge vers A en vérifiant $(\mathcal{P})$. II-3° s'applique : tous les coefficients de A^{-1} sont positifs ou nuls.

Partie III — Résolution du système ($\mathcal{S}$)

1°) Existence et unicité.

a) Méthode 1 — Construction par deux primitives.

Posons $F_0(x) = \int_0^x f(t) dt$ et $G(x) = \int_0^x F_0(t) dt$. Le théorème fondamental du calcul intégral donne $F'_0 = f$ et $G'' = f$. Comme f est de classe $\mathcal{C}^2$, F_0 est de classe $\mathcal{C}^3$ et G de classe $\mathcal{C}^4$.

Définissons $u(x) = xG(1) - G(x)$. Cette fonction est de classe $\mathcal{C}^4$, satisfait $u'' = -f$ et vérifie $u(0) = 0$ puisque $G(0) = 0$, ainsi que $u(1) = G(1) - G(1) = 0$. Elle est donc solution.

Pour l'unicité, si u et v sont deux solutions, $w = u - v$ vérifie $w'' = 0$. Sa dérivée est constante par les accroissements finis, donc $w(x) = ax + b$. Les conditions $w(0) = w(1) = 0$ imposent $a = b = 0$. Ainsi $u = v$.

Une intégration par parties sur $[0, x]$ donne

$G(x) = xF_0(x) - \int_0^x tf(t) dt = \int_0^x (x-t)f(t) dt$. La formule de la solution peut donc aussi s'écrire

$$u(x) = x \int_0^1 (1-t)f(t) dt - \int_0^x (x-t)f(t) dt.$$

Méthode 2 — Une formule intégrale à deux termes.

Définissons $I(x) = \int_0^x tf(t) dt$, $J(x) = \int_x^1 (1-t)f(t) dt$, puis

$$u(x) = (1-x)I(x) + xJ(x).$$

Les intégrandes sont de classe $\mathcal{C}^2$. Le théorème fondamental donne $I'(x) = xf(x)$ et $J'(x) = -(1-x)f(x)$. En dérivant les produits,

$$u'(x) = -I(x) + (1-x)xf(x) + J(x) - x(1-x)f(x) = -I(x) + J(x),$$

puis $u''(x) = -xf(x) - (1-x)f(x) = -f(x)$. L'identité $u'' = -f$, avec $f \in \mathcal{C}^2$, assure $u \in \mathcal{C}^4$. Aux bornes, $I(0) = 0$ et $J(1) = 0$, donc $u(0) = u(1) = 0$.

Pour l'unicité, la différence de deux solutions a une dérivée seconde nulle, donc est affine ; ses deux valeurs nulles aux bornes la rendent identiquement nulle. Cette formule donne donc la même unique solution.

La première construction est fondée sur les primitives ; la seconde présente une expression dont les signes et les conditions aux bornes sont directement lisibles. Aucun échange entre dérivation et intégration n'est supposé : seules des primitives à borne variable sont dérivées par le théorème fondamental.

b) Méthode 1 — Concavité.

Si $f \geq 0$, on a $u'' = -f \leq 0$ sur $[0, 1]$. La fonction u est donc concave. Pour tout $x \in [0, 1]$, son graphe est au-dessus de la corde joignant les valeurs aux bornes :

$$u(x) = u((1-x)0 + x1) \geq (1-x)u(0) + xu(1) = 0.$$

Ainsi u est positive ou nulle.

Méthode 2 — Signe de la formule intégrale.

La seconde formule de a) donne

$$u(x) = (1-x) \int_0^x tf(t) dt + x \int_x^1 (1-t)f(t) dt.$$

Pour $x \in [0, 1]$, les facteurs x et $1-x$ sont positifs ou nuls. Sur les intervalles d'intégration, t , $1-t$ et $f(t)$ sont également positifs ou nuls. Par positivité de l'intégrale, les deux termes sont positifs ou nuls, donc $u(x) \geq 0$.

La première méthode est géométrique ; la seconde établit la positivité terme par terme.

c) Pour $f = 1$, l'équation donne $\tilde{u}'' = -1$. Deux intégrations fournissent

$$\tilde{u}(x) = -\frac{x^2}{2} + ax + b. \text{ La condition en } 0 \text{ donne } b = 0, \text{ puis celle en } 1 \text{ donne } a = \frac{1}{2}.$$

Donc

$$\tilde{u}(x) = \frac{x(1-x)}{2}.$$

Méthode 1 — Mise sous forme canonique pour le maximum.

On écrit $\tilde{u}(x) = \frac{1}{8} - \frac{1}{2}(x - \frac{1}{2})^2$. Le carré est positif ou nul et s'annule en $x = \frac{1}{2}$. Le maximum sur $[0, 1]$ est donc $\frac{1}{8}$, atteint en $\frac{1}{2}$.

Méthode 2 — Étude de la dérivée pour le maximum.

On a $\tilde{u}'(x) = \frac{1}{2} - x$, positive sur $[0, \frac{1}{2}[$ et négative sur $] \frac{1}{2}, 1]$. La fonction est donc croissante puis décroissante ; son maximum est $\tilde{u}(\frac{1}{2}) = \frac{1}{8}$. La forme canonique donne directement une borne globale ; la dérivée donne en plus les variations.

2°) Approximation de la dérivée seconde.

On utilise $x_i = ih = i/(n+1)$, avec $x_0 = 0$, $x_{n+1} = 1$ et un espacement constant $h > 0$.

a) Posons $K = \sup_{t \in [0,1]} |u^{(4)}(t)|$. Cette quantité est finie et atteinte, car $u^{(4)}$ est continue sur le segment $[0, 1]$.

L'inégalité de Taylor-Lagrange à l'ordre 3, appliquée sur les segments joignant x à $x+h$ et à $x-h$, donne

$$u(x+h) = u(x) + hu'(x) + \frac{h^2}{2}u''(x) + \frac{h^3}{6}u^{(3)}(x) + R_+,$$

$$u(x-h) = u(x) - hu'(x) + \frac{h^2}{2}u''(x) - \frac{h^3}{6}u^{(3)}(x) + R_-,$$

avec $|R_+| \leq Kh^4/24$ et $|R_-| \leq Kh^4/24$. Les segments considérés sont bien inclus dans $[0, 1]$ par l'hypothèse sur $x - h$ et $x + h$.

En additionnant, les termes d'ordres 1 et 3 s'annulent. Il reste

$$u(x+h) - 2u(x) + u(x-h) - h^2 u''(x) = R_+ + R_-.$$

Par l'inégalité triangulaire, le module est au plus $2Kh^4/24 = Kh^4/12$. C'est la majoration demandée.

b) Pour $1 \leq i \leq n$, les points $x_i - h = x_{i-1}$ et $x_i + h = x_{i+1}$ appartiennent à $[0, 1]$. On applique a) en x_i , puis on divise l'inégalité par $h^2 > 0$.

On obtient

$$\left| u''(x_i) - \frac{u(x_{i-1}) - 2u(x_i) + u(x_{i+1}))}{h^2} \right| \leq \frac{h^2}{12} K,$$

ce qui est l'inégalité annoncée.

3°) Cas de la fonction constante égale à 1.

a) La fonction $\tilde{u}$ est un polynôme de degré 2, donc $\tilde{u}^{(4)} = 0$. La borne de III-2° b) est nulle. Ainsi, pour tout i ,

$$\left| -1 - \frac{\tilde{u}(x_{i-1}) - 2\tilde{u}(x_i) + \tilde{u}(x_{i+1}))}{h^2} \right| = 0.$$

Par conséquent $\tilde{u}(x_{i-1}) - 2\tilde{u}(x_i) + \tilde{u}(x_{i+1})) = -h^2$. L'approximation est ici exacte.

b) Comme $\tilde{u}(0) = \tilde{u}(1) = 0$, les termes aux extrémités correspondent aux valeurs nulles absentes du vecteur $\tilde{U}$. Pour tout i ,

$$(A\tilde{U})_i = 2\tilde{u}(x_i) - \tilde{u}(x_{i-1}) - \tilde{u}(x_{i+1})) = h^2.$$

Cette écriture vaut aussi pour $n = 1$, où les deux voisins sont les bornes 0 et 1. On a donc $A\tilde{U} = h^2 \tilde{F}$, soit $\frac{1}{h^2} A\tilde{U} = \tilde{F}$.

c) L'inversibilité de A donne $\tilde{U} = h^2 A^{-1} \tilde{F}$. La coordonnée d'indice i s'écrit

$$\tilde{u}(x_i) = h^2 \sum_{j=1}^n b_{i,j}.$$

Comme $0 \leq \tilde{u}(x_i) \leq \frac{1}{8}$, on obtient

$$0 \leq \sum_j b_{i,j} \leq \frac{1}{8h^2} = \frac{(n+1)^2}{8}.$$

On peut même préciser la somme exacte à partir de III-1° c) :

$$\sum_{j=1}^n b_{i,j} = \frac{x_i(1-x_i)}{2h^2} = \frac{i(n+1-i)}{2}.$$

4°) On contrôle d'abord le résidu V , puis on utilise les coefficients positifs de A^{-1} pour majorer l'erreur.

Notons $W = {}^t(u(x_1), \dots, u(x_n))$. Par définition, $\Delta U = U - W$, $AU = h^2 F$ et $V = A\Delta U = h^2 F - AW$.

Les valeurs $u(0) = u(1) = 0$ permettent d'écrire, pour tout i ,

$$v_i = h^2 f(x_i) - 2u(x_i) + u(x_{i-1}) + u(x_{i+1}).$$

Comme $f = -u''$, cela donne

$$v_i = u(x_{i-1}) - 2u(x_i) + u(x_{i+1}) - h^2 u''(x_i).$$

L'inégalité de III-2° a) entraîne $|v_i| \leq \frac{h^4}{12} \sup_{[0,1]} |u^{(4)}|$. Or $u'' = -f$ et $f \in \mathcal{C}^2$ donnent

$u^{(4)} = -f''$. En posant $K_f = \sup_{t \in [0,1]} |f''(t)|$, qui est fini par continuité sur un segment, on obtient

$$|v_i| \leq \frac{h^4}{12} K_f.$$

Ensuite, $\Delta U = A^{-1}V$, donc

$$\Delta u_i = \sum_j b_{i,j} v_j.$$

Les coefficients $b_{i,j}$ sont positifs ou nuls par II-4° d). Par l'inégalité triangulaire et III-3° c),

$$|\Delta u_i| \leq \sum_j b_{i,j} |v_j| \leq \frac{h^4}{12} K_f \sum_j b_{i,j} \leq \frac{h^4}{12} K_f \frac{1}{8h^2} = \frac{h^2}{96} K_f.$$

Finalement, pour tous $1 \leq i \leq n$,

$$|u_i - u(x_i)| \leq \frac{\sup_{t \in [0,1]} |f''(t)|}{96(n+1)^2}.$$

Cette estimation est uniforme sur les points de la subdivision et ne suppose pas f positive.

5°) Exemple $f(x) = \exp(\sin x)$.

a) La fonction est de classe $\mathcal{C}^2$ sur $[0, 1]$. La règle de dérivation d'une composée, puis celle d'un produit donnent

$$f'(x) = \cos x e^{\sin x}, \quad f''(x) = (\cos^2 x - \sin x) e^{\sin x}.$$

Sur $[0, 1]$, on a $|\sin x| \leq 1$, $\cos^2 x \leq 1$ et $e^{\sin x} \leq e$. Ainsi

$$|f''(x)| \leq (\cos^2 x + |\sin x|) e^{\sin x} \leq 2e.$$

La question 4° donne donc $|u_i - u(x_i)| \leq \frac{e}{48(n+1)^2}$. Une condition suffisante pour que cette borne soit au plus 10^{-4} est

$$(n+1)^2 \geq \frac{10^4 e}{48}.$$

Le seuil $\sqrt{10^4 e / 48}$ vaut environ 23,8. Le choix $n = 23$, qui donne $n + 1 = 24$, est donc suffisant. Sa validité peut être établie sans se fier à l'arrondi : $e < \frac{11}{4}$ entraîne

$$\frac{e}{48 \cdot 24^2} < \frac{11}{110592} < \frac{1}{10000},$$

car $110000 < 110592$.

Pour justifier la borne rationnelle utilisée, la série exponentielle donne

$$e = \sum_{k=0}^4 \frac{1}{k!} + \sum_{k=5}^{+\infty} \frac{1}{k!}.$$

La première somme vaut $\frac{65}{24}$. Pour $k = 5 + \ell$, on a $k! \geq 5! 6^\ell$, donc le reste est au plus

$$\frac{1}{120} \sum_{\ell=0}^{+\infty} (1/6)^\ell = \frac{1}{100}.$$

Ainsi $e \leq \frac{65}{24} + \frac{1}{100} < \frac{11}{4}.$

Le choix $n = 23$ garantit bien la précision demandée à tous les points, par une borne exacte.

b) Posons $g(t) = (t^2 + t - 1)e^t$ pour $t \geq 0$. Sa dérivée est

$$g'(t) = ((2t + 1) + (t^2 + t - 1))e^t = t(t + 3)e^t.$$

Elle est nulle en 0 et strictement positive pour $t > 0$. La fonction g est donc strictement croissante sur $[0, +\infty[$, avec $g(0) = -1$ et $g(t) \rightarrow +\infty$ lorsque $t \rightarrow +\infty$.

Puisque $e^t > 0$, son signe est celui de $t^2 + t - 1$. La racine positive vaut $\tau = (\sqrt{5} - 1)/2$. Ainsi g est négative sur $[0, \tau[$, nulle en τ et positive sur $]\tau, +\infty[$. En particulier, sur $[0, 1]$, $-1 = g(0) \leq g(t) \leq g(1) = e$.

Pour $x \in [0, 1]$, $0 \leq \sin x \leq 1$. L'identité $\cos^2 x = 1 - \sin^2 x$ donne

$$f''(x) = (1 - \sin^2 x - \sin x)e^{\sin x} = -g(\sin x).$$

Comme $g(\sin x) \in [-1, e]$ et $e > 1$, on a $|f''(x)| \leq e$. Cette borne améliore celle de a) d'un facteur 2.

L'estimation de 4° devient $|u_i - u(x_i)| \leq \frac{e}{96(n+1)^2}$. Il suffit donc de choisir

$(n+1)^2 \geq \frac{10^4 e}{96}$. Le seuil correspondant vaut environ 16,83, ce qui conduit à $n = 16$.

La même borne rationnelle donne une vérification exacte : pour $n = 16$,

$$\frac{e}{96 \cdot 17^2} < \frac{11}{110976} < \frac{1}{10000},$$

car $110000 < 110976$. Le choix $n = 16$, strictement meilleur que $n = 23$, garantit donc encore l'erreur demandée pour tous les indices i .

Ces deux valeurs sont des choix suffisants issus des majorations respectives ; elles ne prétendent pas être les plus petites valeurs possibles pour l'erreur réelle.

Relevé des adaptations

- Préambule et I-1° : restriction aux espaces vectoriels et matrices réels ; les arguments de dominance diagonale restent inchangés.
- I-2° a) et préambule associé : remplacement des disques par les intervalles $[a_{i,i} - r_i, a_{i,i} + r_i]$ et restriction de la conclusion aux valeurs propres réelles. L'application à la matrice symétrique de I-2° c) est préservée.
- I-2° b) : code adapté en Python ; les fonctions renvoient les centres et rayons des intervalles réels. Les commandes de saisie et de conversion sont expliquées dans le sujet.