Faculté des Sciences
Aix*Marseille Université

CTES
Centre de Télé-Enseignement Sciences

LICENCE 3 MATHEMATIQUES – INFORMATIQUE. MATHEMATIQUES GENERALES. L3MiMG.

Expédition dans la semaine n°	Etape	Code UE	N° d'envoi de l'UE
49	**2L3MAT**	**SMI5U4T**	**3**

Nom de l'UE : Analyse numérique et optimisation

Le cours contient 3 chapitres (systèmes linéaires, systèmes non linéaires, optimisation). Pour chaque semaine, il est proposé d'étudier une partie du cours, de faire des exercices (corrigés) et, éventuellement, de réaliser un TP en scilab (scilab peut être pris sur le web, http://www.scilab.org/fr). Les TP sont fortement conseillés mais non obligatoires. Deux devoirs sont à rendre afin de bénéficier d'une note de contrôle continu.
note finale=max(note-examen, 1/3(2 note-examen + note-contrôle-continu)).

- Contenu de l'envoi : Polycopié, sections 1.6, 2.1 et 2.2.1.TP 5. TP-PageRank

- Guide du travail à effectuer

Semaine 1 :
Etudier le paragraphe 1.6.1 (méthodes de la puissance et puissance inverse)
Exercices proposés (avec corrigés) :
72 (méthode de la puissance) et 74 questions 1-3 (orthogonalisation de Gram-Schmidt)

Semaine 2 :
Etudier le paragraphe 1.6.2, méthode QR.
Exercices proposés (avec corrigés) : 76 (sur la méthode QR)
Faire le TP 5 (feuilles 1 et 2)

Semaine 3 :
Etudier le paragraphe 2.1, Rappel et notations de Calcul Différentiel
Exercices proposés (avec corrigés) : 77 et 78 (rappels de calcul différentiel).

Semaine 4 :
Etudier le paragraphe 2.2.1, point fixe de contraction
Exercice proposés (avec corrigé) : 79 et 80 (sur des méthodes de point fixe)
Faire le TP 5 (feuilles 3 et 4).
La partie de programmation du projet PageRank (très intéressant) n'est pas à rendre
La partie théorique fait partie du deuxième devoir (à rendre en mars)
Il s'agit donc des questions 1 (ne faire que 1 seul exemple), 3, 5, 7, 8, 9, 11
Si cela vous semble trop difficile, n'hésitez pas a me poser des questions

-Coordonnées de l'enseignant responsable de l'envoi

T. Gallouet, CMI, 39 rue Joliot Curie, 13453 marseille cedex 13
email : thierry.gallouet@univ-amu.fr

Vous pouvez aussi consulter la page web: http://www.cmi.univ-mrs.fr/~gallouet/tele.d/anum.d
et me poser des questions par email

1.6 Valeurs propres et vecteurs propres

Les techniques de recherche des éléments propres, c.à.d. des valeurs et vecteurs propres (voir Définition 1.2 page 7) d'une matrice sont essentielles dans de nombreux domaines d'application, par exemple en dynamique des structures : la recherche des modes propres d'une structure peut s'avérer importante pour le dimensionnement de structures sous contraintes dynamiques ; elle est essentielle dans la compréhension des phénomènes acoustiques.
On peut se demander pourquoi on parle dans ce chapitre, intitulé "systèmes linéaires" du problème de recherche des valeurs propres : il s'agit en effet d'un problème non linéaire, les valeurs propres étant les solutions du polynôme caractéristique, qui est un polynôme de degré n, où n est la dimension de la matrice. Il n'est malheureusement pas possible de calculer numériquement les valeurs propres comme les racines du polynôme caractéristique, car cet algorithme est instable : une petite perturbation sur les coefficients du polynôme peut entraîner une erreur très grande sur les racines (voir par exemple le chapitre 5 du polycopié d'E. Hairer, cité dans l'introduction de ce cours, en ligne sur le web). De nombreux algorithmes ont été développés pour le calcul des valeurs propres et vecteurs propres. Ces méthodes sont en fait assez semblables aux méthodes de résolution de systèmes linéaires. Dans le cadre de ce cours, nous nous restreignons à deux méthodes très connues : la méthode de la puissance (et son adaptation de la puissance inverse), et la méthode dite QR.

1.6.1 Méthode de la puissance et de la puissance inverse

Pour expliquer l'algorithme de la puissance, commençons par un exemple simple. Prenons par exemple la matrice

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

dont les valeurs propres sont 1 et 3, et les vecteurs propres associés $\boldsymbol{f}^{(1)} = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ et $\boldsymbol{f}^{(2)} = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Partons de $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ et faisons tourner scilab en itérant les instructions suivantes :

```
-->x =  A * x ; x = x/norm(x)
```

ce qui correspond à la construction de la suite

$$\boldsymbol{x}^{(0)} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|},\ \boldsymbol{x}^{(1)} = \frac{A\boldsymbol{x}^{(0)}}{\|A\boldsymbol{x}^{(0)}\|}, \cdots, \boldsymbol{x}^{(k+1)} = \frac{A\boldsymbol{x}^{(k)}}{\|A\boldsymbol{x}^{(k)}\|} \tag{1.141}$$

où $\|\boldsymbol{x}\|$ désigne la norme euclidienne.
On obtient les résultats suivants :

0.8944272	0.7808688	0.7327935	0.7157819	0.7100107	0.7080761	0.7074300
- 0.4472136	- 0.6246950	- 0.6804511	- 0.6983239	- 0.7061361	- 0.7067834	- 0.706999

On voit clairement sur cet exemple que la suite $\boldsymbol{x}^{(k)}$ converge vers $\boldsymbol{f}_2 = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ lorsque $k \to +\infty$. Si maintenant on fait tourner Scilab en lui demandant de calculer ensuite le produit scalaire de $A\boldsymbol{x}$ avec $\boldsymbol{x}$:

```
-->x= A*x;x=x/norm(x);mu=(A*x)'*x
```

ce qui correspond au calcul de la suite $\mu_k = A\boldsymbol{x}^{(k)} \cdot \boldsymbol{x}^{(k)}, k \geq 0$, on obtient la suite :

$$2.8, 2.9756098, 2.9972603, 2.9996952, 2.9999661, ...$$

qui a tout l'air de converger vers 3 ! En fait on a le théorème suivant, qui montre que dans un certain nombre de cas, on a effectivement convergence de l'algorithme vers la valeur propre dite dominante (celle qui correspond au rayon spectral).

Théorème 1.61 (Convergence de la méthode de la puissance)**.** *Soit A une matrice de $\mathcal{M}_n(\mathbb{C})$. On note $\lambda_1, \cdots, \lambda_n$ les valeurs propres de A, $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ une base orthonormée de trigonalisation de A telle que $A\boldsymbol{f}_n = \lambda_n \boldsymbol{f}_n$. On suppose que la valeur propre λ_n est dominante, c.à.d. que*

$$|\lambda_n| > |\lambda_{n-1}| \geq \cdots \geq |\lambda_1|,$$

et on suppose de plus que $\lambda_n \in \mathbb{R}$. Soit $\boldsymbol{x}^{(0)} \notin Vect(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_{n-1})$. Alors, la suite de vecteurs $\boldsymbol{x}_{2k}$ définie par (1.141) *converge vers un vecteur unitaire qui est vecteur propre de A pour la valeur propre dominante λ_n.*
De plus, si la norme choisie dans l'algorithme (1.141) *est la norme 2, alors la suite $(A\boldsymbol{x}_k \cdot \boldsymbol{x}_k)_{k\in\mathbb{N}}$ converge vers λ_n lorsque $k \to +\infty$.*

Démonstration. La démonstration de ce résultat fait l'objet de l'exercice 72 dans le cas plus simple où A est une matrice symétrique, et donc diagonalisable dans $\mathbb{R}$. □

La méthode de la puissance souffre de plusieurs inconvénients :

1. Elle ne permet de calculer que la plus grande valeur propre. Or très souvent, on veut pouvoir calculer la plus petite valeur propre.
2. De plus, elle ne peut converger que si cette valeur propre est simple.
3. Enfin, même dans le cas où elle est simple, si le rapport des deux plus grandes valeurs propres est proche de 1, la méthode va converger trop lentement.

De manière assez miraculeuse, il existe un remède à chacun de ces maux :

1. Pour calculer plusieurs valeurs propres simultanément, on procède par blocs : on part de p vecteurs orthogonaux $\boldsymbol{x}_1^{(0)}, \ldots \boldsymbol{x}_p^{(0)}$ (au lieu d'un seul). Une itération de la méthode consiste alors à multiplier les p vecteurs par A et à les orthogonaliser par Gram-Schmidt. En répétant cette itération, on approche, si tout se passe bien, p valeurs propres et vecteurs propres de A, et la vitesse de convergence de la méthode est maintenant $\frac{\lambda_{n-p}}{\lambda_n}$.
2. Si l'on veut calculer la plus petite valeur propre, on applique la méthode de la puissance à A^{-1}. On a alors convergence (toujours si tout se passe bien) de $A^{-1}\boldsymbol{x}_k \cdot \boldsymbol{x}_k$ vers $1/|\lambda_1|$. Bien sûr, la mise en oeuvre effective ne s'effectue pas avec l'inverse de A, mais en effectuant une décomposition LU de A qui permet ensuite la résolution du système linéaire $A\tilde{\boldsymbol{x}}_{k+1} = \boldsymbol{x}^{(k)}$ (et $\boldsymbol{x}_{k+1} = \tilde{\boldsymbol{x}}_{k+1}/\|\tilde{\boldsymbol{x}}_{k+1}\|$).
3. Enfin, pour accélérer la convergence de la méthode, on utilise une translation sur A, qui permet de se rapprocher de la valeur propre que l'on veut effectivement calculer. Voir à ce propos l'exercice 73.

1.6.2 Méthode QR

Toute matrice A peut se décomposer sous la forme $A = QR$, où Q est une matrice orthogonale et R une matrice triangulaire supérieure. Dans le cas où A est inversible, cette décomposition est unique. On a donc le théorème suivant :

Théorème 1.62 (Décomposition QR d'une matrice)**.** *Soit $A \in \mathcal{M}_n(\mathbb{R})$. Alors il existe Q matrice orthogonale et R matrice triangulaire supérieure à coefficients diagonaux positifs ou nuls tels que $A = QR$. Si la matrice A est inversible, alors cette décomposition est unique.*

La démonstration est effectuée dans le cas inversible dans la question 1 de l'exercice 76. La décomposition QR d'une matrice A inversible s'obtient de manière très simple par la méthode de Gram-Schmidt, qui permet de

construire une base orthonormée $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ (les colonnes de la matrice Q), à partir de n vecteurs vecteurs indépendants $a_1, \ldots, a_n$ (les colonnes de la matrice A). On se reportera à l'exercice 74 pour un éventuel raffraichissement de mémoire sur Gram-Schmidt. Dans le cas où A n'est pas inversible (et même non carrée), la décomposition existe mais n'est pas unique. La démonstration dans le cadre général se trouve dans le livre de Ph. Ciarlet conseillé en début de ce cours.
L'algorithme QR pour la recherche des valeurs propres d'une matrice est extêmement simple : Si A est une matrice inversible, on pose $A_0 = A$, on effectue la décomposition QR de A : $A = A_0 = Q_0 R_0$ et on calcule $A_1 = R_0 Q_0$. Comme le produit de matrices n'est pas commutatif, les matrices A_0 et A_1 ne sont pas égales, mais en revanche elles sont semblables ; en effet, grâce à l'associativité du produit matriciel, on a :

$$A_1 = R_0 Q_0 = (Q_0^{-1} Q_0) R_0 Q_0 = Q_0^{-1} (Q_0 R_0) Q_0 = Q_0^{-1} A Q_0.$$

Les matrices A_0 et A_1 ont donc même valeurs propres.
On recommence alors l'opération : à l'itération k, on effectue la décomposition QR de A_k : $A_k = Q_k R_k$ et on calcule $A_{k+1} = R_k Q_k$.
Par miracle, pour la plupart des matrices, les coefficients diagonaux de la matrice A_k tendent vers les valeurs propres de la matrice A, et, pour une matrice symétrique, les colonnes de la matrice Q_k vers les vecteurs propres associés. On sait démontrer cette convergence pour certaines matrices ; on pourra trouver par exemple dans les livres de Serre ou Hubbard-Hubert la démonstration sous une hypothèse assez technique et difficile à vérifier en pratique ; l'exercice 76 donne la démonstration (avec la même hypothèse technique) pour le cas plus simple d'une matrice symétrique définie positive.
Pour améliorer la convergence de l'algorithme QR, on utilise souvent la technique dite de "shift" (translation en français). A l'itération n, au lieu d'effectuer la décomposition QR de la matrice A_n, on travaille sur la matrice $A_n - bI$, où b est choisi proche de la plus grande valeur propre. En général on choisit le coefficient $b = a_{nn}^{(k)}$.
L'exercice 75 donne un exemple de l'application de la méthode QR avec shift.

1.6.3 Exercices (valeurs propres, vecteurs propres)

Exercice 72 (Méthode de la puissance)**.** *Suggestions en page 141, corrigé en page 142*

1. Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique (non nulle). Soit $\lambda_n \in \mathbb{R}$ valeur propre de A t.q. $|\lambda_n| = \rho(A)$ et soit $\boldsymbol{y}^{(0)} \in \mathbb{R}^n$. On suppose que $-\lambda_n$ n'est pas une valeur propre de A et que $\boldsymbol{y}^{(0)}$ n'est pas orthogonal à $\mathrm{Ker}(A - \lambda_n Id)$, ce qui revient à dire que lorsqu'on écrit le $\boldsymbol{y}^{(0)}$ dans une base formée de vecteurs propres de A, la composante sur sous-espace propre associé à λ_n est non nulle. (L'espace $\mathbb{R}^n$ est muni de la norme euclidienne.) On définit la suite $(\boldsymbol{y}^{(k)})_{k\in\mathbb{N}}$ par $\boldsymbol{y}^{(k+1)} = A\boldsymbol{y}^{(k)}$ pour $k \in \mathbb{N}$. Montrer que
   (a) $\dfrac{\boldsymbol{y}^{(k)}}{(\lambda_n)^k} \to \boldsymbol{y}$, quand $k \to \infty$, avec $\boldsymbol{y} \neq 0$ et $A\boldsymbol{y} = \lambda_n \boldsymbol{y}$.
   (b) $\dfrac{\|\boldsymbol{y}^{(k+1)}\|}{\|\boldsymbol{y}^{(k)}\|} \to \rho(A)$ quand $k \to \infty$.
   (c) $\frac{1}{\|\boldsymbol{y}^{2k}\|}\boldsymbol{y}^{2k} \to \boldsymbol{x}$ quand $k \to \infty$ avec $A\boldsymbol{x} = \lambda_n \boldsymbol{x}$ et $\|\boldsymbol{x}\| = 1$.

   Cette méthode de calcul de la plus grande valeur propre s'appelle "méthode de la puissance".
2. Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice inversible et $\boldsymbol{b} \in \mathbb{R}^n$. Pour calculer $\boldsymbol{x}$ t.q. $A\boldsymbol{x} = \boldsymbol{b}$, on considère un méthode itérative : on se donne un choix initial $\boldsymbol{x}^{(0)}$, et on construit la suite $\boldsymbol{x}^{(k)}$ telle que $\boldsymbol{x}^{(k+1)} = B\boldsymbol{x}^{(k)} + \boldsymbol{c}$ avec $\boldsymbol{c} = (Id - B)A^{-1}\boldsymbol{b}$, et on suppose B symétrique. On rappelle que si $\rho(B) < 1$, la suite $(\boldsymbol{y}^{(k)})_{k\in\mathbb{N}}$ tend vers $\boldsymbol{x}$. Montrer que, sauf cas particuliers à préciser,
   (a) $\frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|} \to \rho(B)$ quand $k \to \infty$ (ceci donne une estimation de la vitesse de convergence de la méthode itérative).
   (b) $\frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)}\|} \to \rho(B)$ quand $k \to \infty$ (ceci permet d'estimer $\rho(B)$ au cours des itérations).

Exercice 73 (Méthode de la puissance inverse avec shift)**.** *Suggestions en page 141. Corrigé en page 143.*

Soient $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique et $\lambda_1, \ldots, \lambda_p$ ($p \leq n$) les valeurs propres de A. Soit $i \in \{1, \ldots, p\}$, on cherche à calculer λ_i. Soit $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$. On suppose que $\boldsymbol{x}^{(0)}$ n'est pas orthogonal à $\mathrm{Ker}(A - \lambda_i Id)$. On suppose également connaître $\mu \in \mathbb{R}$ t.q. $0 < |\mu - \lambda_i| < |\mu - \lambda_j|$ pour tout $j \neq i$. On définit la suite $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ par $(A - \mu Id)\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)}$ pour $k \in \mathbb{N}$.

1. Vérifier que la construction de la suite revient à appliquer la méthode de la puissance à la matrice $A - \mu Id)^{-1}$.
2. Montrer que $\boldsymbol{x}^{(k)}(\lambda_i - \mu)^k \to \boldsymbol{x}$, quand $k \to \infty$, où $\boldsymbol{x}$ est un vecteur propre associé à la valeur propre λ_i, c.à.d. $\boldsymbol{x} \neq 0$ et $A\boldsymbol{x} = \lambda_i \boldsymbol{x}$.
3. Montrer que $\frac{\|\boldsymbol{x}^{(k+1)}\|}{\|\boldsymbol{x}^{(k)}\|} \to \frac{1}{|\mu - \lambda_i|}$ quand $k \to \infty$.

Exercice 74 (Orthogonalisation de Gram-Schmidt)**.** *Corrigé en page 143*

Soient $\boldsymbol{u}$ et $\boldsymbol{v}$ deux vecteurs de $\mathbb{R}^n$, $u \neq 0$. On rappelle que la projection orthogonale $\mathrm{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ du vecteur $\boldsymbol{v}$ sur la droite vectorielle engendrée par $\boldsymbol{u}$ peut s'écrire de la manière suivante :

$$\mathrm{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = \frac{\boldsymbol{v} \cdot \boldsymbol{u}}{\boldsymbol{u} \cdot \boldsymbol{u}} \boldsymbol{u},$$

où $\boldsymbol{u} \cdot \boldsymbol{v}$ désigne le produit scalaire des vecteurs $\boldsymbol{u}$ et $\boldsymbol{v}$. On note $\|\cdot\|$ la norme euclidienne sur $\mathbb{R}^n$.

1. Soient $(\boldsymbol{a_1}, \ldots, \boldsymbol{a_n})$ une base de $\mathbb{R}^n$. On rappelle qu'à partir de cette base, on peut obtenir une base orthogonale $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ et une base orthonormale $(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)$ par le procédé de Gram-Schmidt qui s'écrit :
Pour $k = 1, \ldots, n$,

$$\boldsymbol{v}_k = \boldsymbol{a}_k - \sum_{j=1}^{k-1} \frac{a_k \cdot v_j}{v_j \cdot v_j} v_j, \qquad \boldsymbol{q}_k = \frac{\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|}. \tag{1.142}$$

1. Montrer par récurrence que la famille $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ est une base orthogonale de $\mathbb{R}^n$.

2. Soient A la matrice carrée d'ordre n dont les colonnes sont les vecteurs $\boldsymbol{a}_j$ et Q la matrice carrée d'ordre N dont les colonnes sont les vecteurs $\boldsymbol{q}_j$ définis par le procédé de Gram-Schmidt (1.142), ce qu'on note :

$$A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \ldots & \boldsymbol{a}_n \end{bmatrix}, \qquad Q = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \ldots & \boldsymbol{q}_n \end{bmatrix}.$$

Montrer que

$$\boldsymbol{a}_k = \|\boldsymbol{v}_k\| \boldsymbol{q}_k + \sum_{j=1}^{k-1} \frac{\boldsymbol{a}_k \cdot \boldsymbol{v}_j}{\|\boldsymbol{v}_j\|} \boldsymbol{q}_j.$$

En déduire que $A = QR$, où R est une matrice triangulaire supérieure dont les coefficients diagonaux sont positifs.

3. Montrer que pour toute matrice $A \in \mathcal{M}_n(\mathbb{R})$ inversible, on peut construire une matrice orthogonale Q (c.à. d. telle que $QQ^t = \mathrm{Id}$) et une matrice triangulaire supérieure R à coefficients diagonaux positifs telles que $A = QR$.

4. Donner la décomposition QR de $A = \begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix}$.

5. On considère maintenant l'algorithme suivant (où l'on stocke la matrice Q orthogonale cherchée dans la matrice A de départ (qui est donc écrasée)

Algorithme 1.63 (Gram-Schmidt modifié)**.**

Pour $k = 1, \ldots, n$,

Calcul de la norme de $\boldsymbol{a}_k$

$r_{kk} := (\sum_{i=1}^n a_{ik}^2)^{\frac{1}{2}}$

Normalisation

Pour $\ell = 1, \ldots, n$

$a_{\ell k} := a_{\ell k} / r_{kk}$

Fin pour ℓ

Pour $j = k+1, \ldots, n$

Produit scalaire correspondant à $q_k \cdot a_j$

$r_{kj} := \sum_{i=1}^{n} a_{ik} a_{ij}$

On soustrait la projection de a_k *sur* q_j *sur tous les vecteurs de* A *après* k.

Pour $i = k+1, \ldots, n$,

$a_{ij} := a_{ij} - a_{ik} r_{kj}$

Fin pour i

Fin pour j

Montrer que la matrice A résultant de cet algorithme est identique à la matrice Q donnée par la méthode de Gram-Schmidt, et que la matrice R est celle de Gram-Schmidt. (Cet algorithme est celui qui est effectivement implanté, car il est plus stable que le calcul par le procédé de Gram-Schmidt original.)

Exercice 75 (Méthode QR avec shift)**.** Soit $A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & 0 \end{bmatrix}$

1. Calculer les valeurs propres de la matrice A.
2. Effectuer la décomposition QR de la matrice A.
3. Calculer $A_1 = RQ$ et $\tilde{A}_1 = RQ - bId$ où b est le terme a_{22}^1 de la matrice A_1
4. Effectuer la décomposition QR de A_1 et $\tilde{A}_1$, et calculer les matrices $A_2 = R_1 Q_1$ et $\tilde{A}_2 = \tilde{R}_1 \tilde{Q}_1$.

Exercice 76 (Méthode QR pour la recherche de valeurs propres)**.** *Corrigé en page 144*

Soit A une matrice inversible. Pour trouver les valeurs propres de A, on propose la méthode suivante, dite "méthode QR" : On pose $A_1 = A$ et on construit une matrice orthogonale Q_1 et une matrice triangulaire supérieure R_1 telles que $A_1 = Q_1 R_1$ (par exemple par l'algorithme de Gram-Schmidt). On pose alors $A_2 = R_1 Q_1$, qui est aussi une matrice inversible. On construit ensuite une matrice orthogonale Q_2 et une matrice triangulaire supérieure R_2 telles que $A_2 = Q_2 R_2$ et on pose $A_3 = R_3 Q_3$. On continue et on construit une suite de matrices A_k telles que :

$$A_1 = A = Q_1 R_1, \; R_1 Q_1 = A_2 = Q_2 R_2, \; \ldots, \; R_k Q_k = A_k = Q_{k+1} R_{k+1}. \tag{1.143}$$

Dans de nombreux cas, cette construction permet d'obenir les valeurs propres de la matrice A sur la diagonale des matrices A_k. Nous allons démontrer que ceci est vrai pour le cas particulier des matrices symétriques définies positives dont les valeurs propres sont simples et vérifiant l'hypothèse (1.145) (on peut le montrer pour une classe plus large de matrices).

On suppose à partir de maintenant que A est une matrice symétrique définie positive qui admet n valeurs propres (strictement positives) vérifiant $\lambda_1 < \lambda_2 < \ldots < \lambda_n$. On a donc :

$$A = P \Lambda P^t, \text{ avec } \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \text{ et } P \text{ est une matrice orthogonale.} \tag{1.144}$$

(La notation $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ désigne la matrice diagonale dont les termes diagonaux sont λ_1,... ,λ_n).

On suppose de plus que

$$P^t \text{ admet une décomposition } LU \text{ et que les coefficients diagonaux de } U \text{ sont strictement positifs.} \tag{1.145}$$

On va montrer que A_k tend vers $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

2. Soient Q_i et R_i les matrices orthogonales et triangulaires supérieures définies par (1.143).
2.1 Montrer que $A^2 = \tilde{Q}_2 \tilde{R}_2$ avec $\tilde{Q}_k = Q_1 Q_2$ et $\tilde{R}_k = R_2 R_1$.

2.2 Montrer, par récurrence sur k, que

$$A^k = \tilde{Q}_k \tilde{R}_k, \tag{1.146}$$

avec

$$\tilde{Q}_k = Q_1 Q_2 \dots Q_{k-1} Q_k \text{ et } \tilde{R}_k = R_k R_{k-1} \dots R_2 R_1. \tag{1.147}$$

2.3 Justifier brièvement le fait que $\tilde{Q}_k$ est une matrice orthogonale et $\tilde{R}_k$ est une matrice triangulaire à coefficients diagonaux positifs.

3. Soit $M_k = \Lambda^k L \Lambda^{-k}$.

3.1 Montrer que $PM_k = \tilde{Q}_k T_k$ où $T_k = \tilde{R}_k U^{-1} \Lambda^{-k}$ est une matrice triangulaire supérieure dont les coefficients diagonaux sont positifs.

3.2 Calculer les coefficients de M_k en fonction de ceux de L et des valeurs propres de A.

3.3 En déduire que M_k tend vers la matrice identité et que $\tilde{Q}_k T_k$ tend vers P lorsque $k \to +\infty$.

4. Soient $(B_k)_{k\in\mathbb{N}}$ et $(C_k)_{k\in\mathbb{N}}$ deux suites de matrices telles que les matrices B_k sont orthogonales et les matrices C_k triangulaires supérieures et de coefficients diagonaux positifs. On va montrer que si $B_k C_k$ tend vers la matrice orthogonale B lorsque k tend vers l'infini alors B_k tend vers B et C_k tend vers l'identité lorsque k tend vers l'infini.
On suppose donc que $B_k C_k$ tend vers la matrice orthogonale B. On note $b_1, b_2, \dots, b_n$ les colonnes de la matrice B et $b_1^{(k)}, b_2^{(k)}, \dots, b_n^{(k)}$ les colonnes de la matrice B_k, ou encore :

$$B = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}, \qquad B_k = \begin{bmatrix} b_1^{(k)} & b_2^{(k)} & \dots & b_n^{(k)} \end{bmatrix}.$$

et on note $c_{i,j}^{(k)}$ les coefficients de C_k.

4.1 Montrer que la première colonne de $B_k C_k$ est égale à $c_{1,1}^{(k)} b_1^{(k)}$. En déduire que $c_{1,1}^{(k)} \to 1$ et que $b_1^{(k)} \to b_1$.

4.2 Montrer que la seconde colonne de $B_k C_k$ est égale à $c_{1,2}^{(k)} b_1^{(k)} + c_{2,2}^{(k)} b_2^{(k)}$. En déduire que $c_{1,2}^{(k)} \to 0$, puis que $c_{2,2}^{(k)} \to 1$ et que $b_2^{(k)} \to b_2$.

4.3 Montrer que lorsque $k \to +\infty$, on a $c_{i,j}^{(k)} \to 0$ si $i \neq j$, puis que $c_{i,i}^{(k)} \to 1$ et $b_i^{(k)} \to b_i$.

4.4 En déduire que B_k tend B et C_k tend vers l'identité lorsque k tend vers l'infini.

5. Déduire des questions 3 et 4 que $\tilde{Q}_k$ tend vers P et T_k tend vers Id lorsque $k \to +\infty$.

6. Montrer que $\tilde{R}_k (\tilde{R}_{k-1})^{-1} = T_k \Lambda T_{k-1}$. En déduire que R_k et A_k tendent vers Λ.

1.6.4 Suggestions

Exercice 72 page 138 (Méthode de la puissance pour calculer le rayon spectral de A.)

1. Décomposer $\boldsymbol{x}^{(0)}$ sur une base de vecteurs propres orthonormée de A, et utiliser le fait que $-\lambda_n$ n'est pas valeur propre.
2. a/ Raisonner avec $\boldsymbol{y}^{(k)} = \boldsymbol{x}^{(k)} - \boldsymbol{x}$ où $\boldsymbol{x}$ est la solution de $A\boldsymbol{x} = \boldsymbol{b}$ et appliquer la question 1.
b/ Raisonner avec $\boldsymbol{y}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$.

Exercice 73 page 138 (Méthode de la puissance inverse)

Appliquer l'exercice précédent à la matrice $B = (A - \mu Id)^{-1}$.

1.6.5 Corrigés

Exercice 72 page 138 (Méthode de la puissance pour calculer le rayon spectral de A)

1. Comme A est une matrice symétrique (non nulle), A est diagonalisable dans $\mathbb{R}$. Soit $(f_1, \ldots, f_n)$ une base orthonormée de $\mathbb{R}^n$ formée de vecteurs propres de A associée aux valeurs propres $\lambda_1, \ldots, \lambda_n$ (qui sont réelles). On décompose $y^{(0)}$ sur $(f_i)_{i=1,\ldots,n}$: $y^{(0)} = \sum_{i=1}^n \alpha_i f_i$. On a donc $Ay^{(0)} = \sum_{i=1}^n \lambda_i \alpha_i f_i$ et $A^k y^{(0)} = \sum_{i=1}^n \lambda_i^k \alpha_i f_i$.
   On en déduit :
   $$\frac{y^{(k)}}{\lambda_n^k} = \sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_n}\right)^k \alpha_i f_i.$$
   Comme $-\lambda_n$ n'est pas valeur propre,
   $$\lim_{k\to+\infty} (\frac{\lambda_i}{\lambda_n})^k = 0 \text{ si } \lambda_i \neq \lambda_n. \tag{1.148}$$
   Soient $\lambda_1, \ldots, \lambda_p$ les valeurs propres différentes de λ_n, et $\lambda_{p+1}, \ldots, \lambda_n = \lambda_n$. On a donc
   $\lim_{n\to+\infty} \frac{y^{(k)}}{\lambda_n^k} = \sum_{i=p+1}^n \alpha_i f_i = y$, avec $Ay = \lambda_n y$.
   De plus, $y \neq 0$: en effet, $y^{(0)} \notin (Ker(A-\lambda_n Id))^\perp = Vect\{f_1, \ldots, f_p\}$, et donc il existe $i \in \{p+1, \ldots, n\}$ tel que $\alpha_i \neq 0$.
   Pour montrer (b), remarquons que
   $$\frac{\|y^{(k+1)}\|}{\|y^{(k)}\|} = |\lambda_n| \frac{\|\frac{y^{(k+1)}}{\lambda_n^{k+1}}\|}{\|\frac{y^{(k)}}{\lambda_n^k}\|} \to |\lambda_n| \frac{\|y\|}{\|y\|} = |\lambda_n| \text{ lorsque } k \to +\infty.$$
   Enfin, $\frac{y^{(2k)}}{\|y^{(2k)}\|} = \frac{y^{(2k)}}{\lambda_n^{2k}} \frac{\lambda_n^{2k}}{\|y^{(2k)}\|}$ et $\lim_{k\to+\infty} \frac{\|y^{(2k)}\|}{\lambda_n^{2k}} = \|y\|$.
   On a donc $\lim_{k\to+\infty} \frac{y^{(2k)}}{\|y^{(2k)}\|} = x$, avec $x = \frac{y}{\|y\|}$.

2. a) La méthode I s'écrit à partir de $x^{(0)}$ connu : $x^{(k+1)} = Bx^{(k)} + c$ pour $k \geq 1$, avec $c = (I - B)A^{-1}b$.
   On a donc
   $$\begin{aligned} x^{(k+1)} - x &= Bx^{(k)} + (Id - B)x - x \\ &= B(x^{(k)} - x). \end{aligned} \tag{1.149}$$
   Si $y^{(k)} = x^{(k)} - x$, on a donc $y^{(k+1)} = By^{(k)}$, et d'après la question 1a) si $y^{(0)} \not\perp \mathrm{Ker}(B - \mu_n Id)$ où μ_n est la plus grande valeur propre de B, (avec $|\mu_n| = \rho(B) et - \mu_n$ non valeur propre), alors
   $$\frac{\|y^{(k+1)}\|}{\|y^{(k)}\|} \longrightarrow \rho(B) \text{ lorsque } k \to +\infty,$$
   c'est–à–dire
   $$\frac{\|x^{(k+1)} - x\|}{\|x^{(k)} - x\|} \longrightarrow \rho(B) \text{ lorsque } k \to +\infty.$$
   b) On applique maintenant 1a) à $y^{(k)} = x^{(k+1)} - x^{(k)}$ avec
   $$y^{(0)} = x^{(1)} - x^{(0)} \text{ où } x^{(1)} = Ax^{(0)}.$$
   On demande que $x^{(1)} - x^{(0)} \notin \mathrm{Ker}(B - \mu_n Id)^\perp$ comme en a), et on a bien $y^{(k+1)} = By^{(k)}$, donc $\frac{\|y^{(k+1)}\|}{\|y^{(k)}\|} \longrightarrow \rho(B)$ lorsque $k \to +\infty$.

Exercice 73 page 138 (Méthode de la puissance inverse avec shift)

Comme $0 < |\mu - \lambda_i| < |\mu - \lambda_j|$ pour tout $j \neq i$, la matrice $A - \mu Id$ est inversible. On peut donc appliquer l'exercice 72 à la matrice $B = (A - \mu Id)^{-1}$. Les valeurs propres de B sont les valeurs de $\dfrac{1}{\lambda_j - \mu}$, $j = 1, \ldots, n$, où les λ_j sont les valeurs propres de A.

Comme $|\mu - \lambda_i| < |\mu - \lambda_j|, \ \forall j \neq i$, on a $\rho(B) = \dfrac{1}{|\lambda_i - \mu|}$.

Or, $\dfrac{1}{\lambda_i - \mu}$ est valeur propre de B et $\dfrac{1}{\mu - \lambda_i}$ ne l'est pas. En effet, si $\dfrac{1}{\mu - \lambda_i}$ était valeur propre, il existerait j tel que $\dfrac{1}{\mu - \lambda_i} = \dfrac{1}{\lambda_j - \mu}$, ce qui est impossible car $|\mu - \lambda_i| < |\mu - \lambda_j|$ pour $j \neq i$. Donc $\rho(B) = \dfrac{1}{\lambda_i - \mu}$.

On a également $\mathrm{Ker}(B - \dfrac{1}{\lambda_i - \mu} Id) = \mathrm{Ker}(A - \lambda_i Id)$, donc

$$x^{(0)} \notin (\mathrm{Ker}(B - \frac{1}{\lambda_i - \mu} Id))^{\perp} = (\mathrm{Ker}(A - \lambda_i Id)^{\perp}.$$

On peut donc appliquer l'exercice 72 page 138 qui donne 1 et 2.

Exercice 74 page 139 (Orthogonalisation par Gram-Schmidt)

1. Par définition de la projection orthogonale, on a $v_1 \cdot v_2 = a_1 \cdot (a_2 - \mathrm{proj}_{a_1}(a_2)) = 0$.
Supposons la récurrence vraie au rang $k-1$ et montrons que v_k est orthogonal à tous les v_i pour $i = 1, \ldots, k-1$.
Par définition, $v_k = a_k - \sum_{j=1}^{k-1} \frac{a_k \cdot v_j}{v_j \cdot v_j} v_j$, et donc

$$v_k \cdot v_i = a_k \cdot v_i - \sum_{j=1}^{k-1} \frac{a_k \cdot v_j}{v_j \cdot v_j} v_j \cdot v_i = a_k \cdot v_i - a_k \cdot v_i$$

par hypothèse de récurrence. On en déduit que $v_k \cdot v_i = 0$ et donc que la famille $(v_1, \ldots v_n)$ est une base orthogonale.

2. De la relation (1.142), on déduit que :

$$a_k = v_k + \sum_{j=1}^{k-1} \frac{a_k \cdot v_j}{v_j \cdot v_j} v_j,$$

et comme $v_j = \|v_j\| q_j$, on a bien :

$$a_k = \|v_k\| q_k + \sum_{j=1}^{k-1} \frac{a_k \cdot v_j}{\|v_j\|} q_j.$$

La k-ième colonne de A est donc une combinaison linéaire de la k-ème colonne de Q affectée du poids $\|v_k\|$ et des $k-1$ premières affectées des poids $\frac{a_k \cdot v_j}{\|v_j\|}$. Ceci s'écrit sous forme matricielle $A = QR$ où R est une matrice carrée dont les coefficients sont $R_{k,k} = \|v_k\|$, $R_{j,k} = \frac{a_k \cdot v_j}{\|v_j\|}$ si $j < k$, et $R_{j,k} = 0$ si $j > k$. La matrice R est donc bien triangulaire supérieure et à coefficients diagonaux positifs.

3. Si A est inversible, par le procédé de Gram-Schmidt (1.142) on construit la matrice $Q = \begin{bmatrix} q_1 & q_2 & \ldots & q_n \end{bmatrix}$, et par la question 2, on sait construire une matrice R triangulaire supérieure à coefficients diagonaux positifs $A = QR$.

4. On a $a_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ et donc $q_1 = \frac{1}{2} \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}$

Puis $a_2 = \begin{bmatrix}4\\0\end{bmatrix}$ et donc $v_2 = a_2 - \frac{a_2\cdot v_1}{v_1\cdot v_1}v_1 = \begin{bmatrix}4\\0\end{bmatrix} - \frac{4}{2}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}2\\-2\end{bmatrix}$. Donc $\boldsymbol{q}_2 = \frac{1}{2}\begin{bmatrix}\sqrt{2}\\-\sqrt{2}\end{bmatrix}$, et $Q = \frac{1}{2}\begin{bmatrix}\sqrt{2} & \sqrt{2}\\ \sqrt{2} & -\sqrt{2}\end{bmatrix}$.
Enfin, $R = \begin{bmatrix}\|v_1\| & \frac{a_2\cdot v_1}{\|v_1\|}\\ 0 & \|v_1\|\end{bmatrix} = \begin{bmatrix}\sqrt{2} & 2\sqrt{2}\\ 0 & 2\sqrt{2}\end{bmatrix}$, et $Q = \frac{1}{2}\begin{bmatrix}\sqrt{2} & \sqrt{2}\\ \sqrt{2} & -\sqrt{2}\end{bmatrix}$.

Exercice 76 page 140 (Méthode QR pour la recherche de valeurs propres)

1.1 Par définition et associativité du produit des matrices,

$$A^2 = (Q_1R_1)(Q_1R_1) = Q_1(R_1Q_1)R_1 = Q_1(R_1Q_1)R_1 = Q_1(Q_2R_2)R_1 = (Q_1Q_2)(R_2R_1) = \tilde{Q}_2\tilde{R}_2$$

avec $\tilde{Q}_2 = Q_1Q_2$ et $\tilde{R}_2 = R_1R_2$.

1.2 La propriété est vraie pour $k = 2$. Supposons la vraie jusqu'au rang $k-1$ et montrons là au rang k. Par définition, $A^k = A^{k-1}A$ et donc par hypothèse de récurrence, $A^k = \tilde{Q}_{k-1}\tilde{R}_{k-1}A$. On en déduit que :

$$\begin{aligned}
A^k &= \tilde{Q}_{k-1}\tilde{R}_{k-1}Q_1R_1\\
&= Q_1\dots Q_{k-1}R_{k-1}\dots R_2(R_1Q_1)R_1\\
&= Q_1\dots Q_{k-1}R_{k-1}\dots R_2(Q_2R_2)R_1\\
&= Q_1\dots Q_{k-1}R_{k-1}\dots (R_2Q_2)R_2R_1\\
&= Q_1\dots Q_{k-1}R_{k-1}\dots R_3(Q_3R_3)R_2R_1\\
&\vdots \qquad\qquad \vdots\\
&= Q_1\dots Q_{k-1}R_{k-1}\dots R_j(Q_jR_j)R_{j-1}\dots R_2R_1\\
&= Q_1\dots Q_{k-1}R_{k-1}\dots R_{j+1}(R_jQ_j)R_{j-1}\dots R_2R_1\\
&= Q_1\dots Q_{k-1}R_{k-1}\dots R_{j+1}(Q_{j+1}R_j)R_{j-1}\dots R_2R_1\\
&= Q_1\dots Q_{k-1}R_{k-1}(Q_{k-1}R_{k-1})R_{k-2}\dots R_2R_1\\
&= Q_1\dots Q_{k-1}(R_{k-1}Q_{k-1})R_{k-1}R_{k-2}\dots R_2R_1\\
&= Q_1\dots Q_{k-1}(Q_kR_k)R_{k-1}R_{k-2}\dots R_2R_1\\
&= \tilde{Q}_k\tilde{R}_k
\end{aligned}$$

1.3 La matrice $\tilde{Q}_k$ est un produit de matrices orthogonales et elle est donc orthogonale. (On rappelle que si P et Q sont des matrices orthogonales, c.à.d. $P^{-1} = P^t$ et $Q^{-1} = Q^t$, alors $(PQ)^{-1} = Q^{-1}P^{-1} = Q^tP^t = (PQ)^t$ et donc PQ est orthogonale.)

De même, le produit de deux matrices triangulaires supérieures à coefficients diagonaux positifs est encore une matrice triangulaire supérieure à coefficients diagonaux positifs.

2.1 Par définition, $PM_k = P\Lambda^kL\Lambda^{-k} = P\Lambda^kP^tP^{-t}L\Lambda^{-k} = A^kP^{-t}L\Lambda^{-k}$.
Mais $A^k = \tilde{Q}_k\tilde{R}_k$ et $P^t = LU$, et donc :
$PM_k = \tilde{Q}_k\tilde{R}_kU^{-1}\Lambda^{-k} = \tilde{Q}_kT_k$ où $T_k = \tilde{R}_kU^{-1}\Lambda^{-k}$. La matrice T_k est bien triangulaire supérieure à coefficients diagonaux positifs, car c'est un produit de matrices triangulaires supérieures à coefficients diagonaux positifs.
2.2

$$(M_k)_{i,j} = (\Lambda^kL\Lambda^{-k})_{i,j} = \begin{cases} L_{i,i} & \text{si } i = j,\\ \frac{\lambda_j^k}{\lambda_i^k}Li,j & \text{si } i > j,\\ 0 & \text{sinon.}\end{cases}$$

2.3 On déduit facilement de la question précédente que, lorsque $k \to +\infty$, $(M_k)_{i,j} \to 0$ si $i \neq j$ et $(M_k)_{i,i} \to 1$ et donc que M_k tend vers la matrice identité et que $\tilde{Q}_kT_k$ tend vers P lorsque $k \to +\infty$.

3.1 Par définition, $(B_kC_k)_{i,1} = \sum_{\ell=1,n}(B_k)_{i,\ell}(C_k)_{\ell,1} = (B_k)_{i,1}(C_k)_{1,1}$ car C_k est triangulaire supérieure. Donc la première colonne de B_kC_k est bien égale à $c_{1,1}^{(k)}\boldsymbol{b}_1^{(k)}$.

Comme B_kC_k tend vers B, la première colonne $\boldsymbol{b}_1^{(k)}$ de B_kC_k tend vers la première colonne de B, c'est -à-dire

$$c_{1,1}^{(k)}\boldsymbol{b}_1^{(k)} \to \boldsymbol{b}_1 \text{ lorsque } k \to \infty.$$

Comme les matrices B et B_k sont des matrices orthogonales, leurs vecteurs colonnes sont de norme 1, et donc

$$|c_{1,1}^{(k)}| = \|c_{1,1}^{(k)}\boldsymbol{b}_1^{(k)}\| \to \|\boldsymbol{b}_1\| = 1 \text{ lorsque } k \to \infty.$$

On en déduit que $|c_{1,1}^{(k)}| \to 1$ lorsque $k \to +\infty$, et donc $\lim_{k\to+\infty} c_{1,1}^{(k)} = \pm 1$. Or, par hypothèse, la matrice $C^{(k)}$ a tous ses coefficients diagonaux positifs, on a donc bien $c_{1,1}^{(k)} \to 1$ lorsque $k \to +\infty$. Par conséquent, on a $\boldsymbol{b}_1^{(k)} \to \boldsymbol{b}_1$ lorsque $k \to \infty$.

3.2 Comme C_k est triangulaire supérieure, on a :

$$(B_kC_k)_{i,2} = \sum_{\ell=1,n}(B_k)_{i,\ell}(C_k)_{\ell,2} = (B_k)_{i,1}(C_k)_{1,1} + (B_k)_{i,2}(C_k)_{2,1},$$

et donc la seconde colonne de B_kC_k est bien égale à $c_{1,2}^{(k)}\boldsymbol{b}_1^{(k)} + c_{2,2}^{(k)}\boldsymbol{b}_2^{(k)}$.
On a donc

$$c_{1,2}^{(k)}\boldsymbol{b}_1^{(k)} + c_{2,2}^{(k)}\boldsymbol{b}_2^{(k)} \to \boldsymbol{b}_2 \text{ lorsque } k \to +\infty. \tag{1.150}$$

La matrice B_k est orthogonale, et donc $\boldsymbol{b}_1^{(k)} \cdot \boldsymbol{b}_1^{(k)} = 1$ et $\boldsymbol{b}_1^{(k)} \cdot \boldsymbol{b}_2^{(k)} = 0$. De plus, par la question précédente, $\boldsymbol{b}_1^{(k)} \to \boldsymbol{b}_1$ lorsque $k \to +\infty$, On a donc, en prenant le produit scalaire du membre de gauche de (1.150) avec $\boldsymbol{b}_1^{(k)}$,

$$c_{1,2}^{(k)} = \left(c_{1,2}^{(k)}\boldsymbol{b}_1^{(k)} + c_{2,2}^{(k)}\boldsymbol{b}_2^{(k)}\right) \cdot \boldsymbol{b}_1^{(k)} \to \boldsymbol{b}_2 \cdot \boldsymbol{b}_1 = 0 \text{ lorsque } k \to +\infty.$$

Comme $c_{1,2}^{(k)} \to 0$ et $\boldsymbol{b}_1^{(k)} \to \boldsymbol{b}_1$ on obtient par (1.150) que

$$c_{2,2}^{(k)}\boldsymbol{b}_2^{(k)} \to \boldsymbol{b}_2 \text{ lorsque } k \to +\infty.$$

Le même raisonnement que celui de la question précédente nous donne alors que $c_{2,2}^{(k)} \to 1$ et $\boldsymbol{b}_2^{(k)} \to \boldsymbol{b}_2$ lorsque $k \to +\infty$.

3.3 On sait déjà par les deux questions précédentes que ces assertions sont vraies pour $i = 1$ et 2. Supposons qu'elles sont vérifiées jusqu'au rang $i-1$, et montrons que $c_{i,j}^{(k)} \to 0$ si $i \neq j$, puis que $c_{i,i}^{(k)} \to 1$ et $\boldsymbol{b}_i^{(k)} \to \boldsymbol{b}_i$.
Comme C_k est triangulaire supérieure, on a :

$$(B_kC_k)_{i,j} = \sum_{\ell=1,n}(B_k)_{i,\ell}(C_k)_{\ell,j} = \sum_{\ell=1}^{j-1}(B_k)_{i,\ell}(C_k)_{\ell,j} + (B_k)_{i,j}(C_k)_{j,j},$$

et donc la j-ème colonne de B_kC_k est égale à $\sum_{\ell=1}^{j-1} c_{\ell,j}^{(k)}\boldsymbol{b}_\ell^{(k)} + c_{j,j}^{(k)}\boldsymbol{b}_j^{(k)}$. On a donc

$$\sum_{\ell=1}^{j-1} c_{\ell,j}^{(k)}\boldsymbol{b}_\ell^{(k)} + c_{j,j}^{(k)}\boldsymbol{b}_j^{(k)} \to \boldsymbol{b}_j \text{ lorsque } k \to +\infty. \tag{1.151}$$

La matrice B_k est orthogonale, et donc $\boldsymbol{b}_i^{(k)} \cdot \boldsymbol{b}_j^{(k)} = \delta_{i,j}$. De plus, par hypothèse de récurrence, on sait que $\boldsymbol{b}_\ell^{(k)} \to \boldsymbol{b}_\ell$ pour tout $\ell \leq j-1$. En prenant le produit scalaire du membre de gauche de (1.151) avec $\boldsymbol{b}_m^{(k)}$, pour $m < j$, on obtient

$$c_{m,j}^{(k)} = \left(\sum_{\ell=1}^{j-1} c_{\ell,j}^{(k)}\boldsymbol{b}_\ell^{(k)} + c_{j,j}^{(k)}\boldsymbol{b}_j^{(k)}\right) \cdot \boldsymbol{b}_m^{(k)} \to \boldsymbol{b}_m \cdot \boldsymbol{b}_j = 0 \text{ lorsque } k \to +\infty.$$

On déduit alors de (1.151) que $c_{j,j}^{(k)} \boldsymbol{b}_j^{(k)} \to \boldsymbol{b}_j$ lorsque $k \to +\infty$, et le même raisonnement que celui de la question 4.1 nous donne alors que $c_{j,j}^{(k)} \to 1$ et $\boldsymbol{b}_j^{(k)} \to \boldsymbol{b}_j$ lorsque $k \to +\infty$.
ce qui conclut le raisonnement par récurrence.

3.4 En déduire que B_k tend B et C_k tend vers l'identité lorsque k tend vers l'infini.
On a montré aux trois questions précédentes que la j-ième colonne de B_k tend vers la j-ième colonne de B, et que $c_{i,j}^{(k)} \to \delta_{i,j}$ lorque k tend vers $+\infty$. On a donc bien le résultat demandé.

4. D'après la question 3, $\tilde{Q}_k T_k$ tend vers P, et d'après la question 4, comme $\tilde{Q}_k$ est orthogonale et T_k triangulaire supérieure à coefficients positifs, on a bien $\tilde{Q}_k$ qui tend vers P et T_k qui tend vers Id lorsque $k \to +\infty$.

5. On a $\tilde{R}_k = T_k \Lambda^k U$ et donc $\tilde{R}_k (\tilde{R}_{k-1})^{-1} = T_k \Lambda^k U U^{-1} \Lambda^{-k+1} T_{k-1} = T_k \Lambda T_{k-1}$.
Comme T_k tend vers Id, on a $R_k = \tilde{R}_k (\tilde{R}_{k-1})^{-1}$ qui tend vers Λ. De plus, $A_k = Q_k R_k$, où $Q_k = \tilde{Q}_k (\tilde{Q}_{k-1})^{-1}$ tend vers Id et R_k tend vers Λ. Donc A_k tend vers Λ.

Chapitre 2

Systèmes non linéaires

Dans le premier chapitre, on a étudié quelques méthodes de résolution de systèmes linéaires en dimension finie. L'objectif est maintenant de développer des méthodes de résolution de systèmes non linéaires, toujours en dimension finie. On se donne $g \in C(\mathbb{R}^n, \mathbb{R}^n)$ et on cherche x dans $\mathbb{R}^n$ solution de :

$$\begin{cases} x \in \mathbb{R}^n \\ g(x) = 0. \end{cases} \tag{2.1}$$

Au Chapitre I on a étudié des méthodes de résolution du système (2.1) dans le cas particulier $g(x) = Ax - b$, $A \in \mathcal{M}_n(\mathbb{R})$, $b \in \mathbb{R}^n$. On va maintenant étendre le champ d'étude au cas où g n'est pas forcément affine. On étudiera deux familles de méthodes pour la résolution approchée du système (2.1) :

— les méthodes de point fixe : point fixe de contraction et point fixe de monotonie

— les méthodes de type Newton [1].

2.1 Rappels et notations de calcul différentiel

Le premier chapitre faisait appel à vos connaissances en algèbre linéaire. Ce chapitre-ci, ainsi que le suivant (optimisation) s'appuieront sur vos connaissances en calcul différentiel, et nous allons donc réviser les quelques notions qui nous seront utiles.

Définition 2.1 (Application différentiable)**.** *Soient E et F des espaces vectoriels normés, f une application de E dans F et $x \in E$. On rappelle que f est différentiable en x s'il existe $T_x \in \mathcal{L}(E, F)$ (où $\mathcal{L}(E, F)$ est l'ensemble des applications linéaires continues de E dans F) telle que*

$$f(x + h) = f(x) + T_x(h) + \|h\|_E \varepsilon(h) \text{ avec } \varepsilon(h) \to 0 \text{ quand } h \to 0. \tag{2.2}$$

L'application T_x est alors unique [2] et on note $Df(x) = T_x \in \mathcal{L}(E, F)$ la différentielle de f au point x. Si f est différentiable en tout point de E, alors on appelle différentielle de f l'application $Df = E \to \mathcal{L}(E, F)$ qui à $x \in E$ associe l'application linéaire continue $Df(x)$ de E dans F.

Remarquons tout de suite que si f est une application linéaire continue de E dans F, alors f est différentiable, et $Df = f$. En effet, si f est linéaire, $f(x + h) - f(x) = f(h)$, et donc l'égalité (2.2) est vérifiée avec $T_x = f$ et $\varepsilon = 0$.
Voyons maintenant quelques cas particuliers d'espaces E et F :

1. Isaac Newton, 1643 - 1727, né d'une famille de fermiers, est un philosophe, mathématicien, physicien, alchimiste et astronome anglais. Figure emblématique des sciences, il est surtout reconnu pour sa théorie de la gravitation universelle et la création, en concurrence avec Leibniz, du calcul infinitésimal.

Cas où $E = \mathbb{R}$ et $F = \mathbb{R}$ Si f est une fonction de $\mathbb{R}$ dans $\mathbb{R}$, dire que f est différentiable en x revient à dire que f est dérivable en x. En effet, dire que f est dérivable en x revient à dire que

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \text{ existe, et } \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = f'(x),$$

ce qui s'écrit encore

$$\frac{f(x+h)-f(x)}{h} = f'(x) + \varepsilon(h), \text{ avec } \varepsilon(h) \to 0 \text{ lorsque } h \to 0,$$

c'est-à-dire

$$f(x+h) - f(x) = T_x(h) + h\varepsilon(h), \text{ avec } T_x(h) = f'(x)h,$$

ce qui revient à dire que f est différentiable en x, et que sa différentielle en x est l'application linéaire $T_x : \mathbb{R} \to \mathbb{R}$, qui à h associe $f'(x)h$. On a ainsi vérifié que pour une fonction de $\mathbb{R}$ dans $\mathbb{R}$, la notion de différentielle coïncide avec celle de dérivée.

Exemple 2.2. *Prenons $f : \mathbb{R} \to \mathbb{R}$ définie par $f(x) = \sin x$. Alors f est dérivable en tout point et sa dérivée vaut $f'(x) = \cos x$. La fonction f est donc aussi différentiable en tout point. La différentielle de f au point x est l'application linéaire $Df(x)$ qui à $h \in \mathbb{R}$ associe $Df(x)(h) = \cos x\, h$. La différentielle de f est l'application de $\mathbb{R}$ dans $\mathcal{L}(\mathbb{R}, \mathbb{R})$, qui à x associe $Df(x)$ (qui est donc elle même une application linéaire).*

Cas où $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$ Soit $f : \mathbb{R}^n \to \mathbb{R}^p$, $x \in \mathbb{R}^n$ et supposons que f est différentiable en x ; alors $Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$; par caractérisation d'une application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^n$, il existe une unique matrice $J_f(x) \in \mathcal{M}_{p,n}(\mathbb{R})$ telle que

$$\underbrace{Df(x)(y)}_{\in \mathbb{R}^p} = \underbrace{J_f(x)y}_{\in \mathbb{R}^p}, \ \forall y \in \mathbb{R}^n.$$

On confond alors souvent l'application linéaire $Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ avec la matrice $J_f(x) \in \mathcal{M}_{p,n}(\mathbb{R})$ qui la représente, qu'on appelle **matrice jacobienne** de f au point x et qu'on note J_f. On écrit donc :

$$J_f(x) = Df(x) = (a_{i,j})_{1\le i\le p, 1\le j\le n} \text{ où } a_{i,j} = \partial_j f_i(x),$$

∂_j désignant la dérivée partielle par rapport à la j-ème variable.
Notons que si $n = p = 1$, la fonction f est de $\mathbb{R}$ dans $\mathbb{R}$ et la matrice jacobienne en x n'est autre que la dérivée en x : $J_f(x) = f'(x)$. On confond dans cette écriture la matrice $J_f(x)$ qui est de taille 1×1 avec le scalaire $f'(x)$.

Exemple 2.3. *Prenons $n = 3$ et $p = 2$; soit $f : \mathbb{R}^3 \to \mathbb{R}^2$ définie par :*

$$f(x) = \begin{pmatrix} x_1^2 + x_2^3 + x_3^4 \\ \\ 2x_1 - x_2 \end{pmatrix}, \ \forall x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Soit $h \in \mathbb{R}^3$ de composantes (h_1, h_2, h_3). Pour calculer la différentielle de f (en x appliquée à h), on peut calculer $f(x+h) - f(x)$:

$$\begin{aligned} f(x+h) - f(x) &= \begin{bmatrix} (x_1+h_1)^2 - x_1^2 + (x_2+h_2)^3 - x_2^3 + (x_3+h_3)^4 - x_3^4 \\ \\ 2(x_1+h_1) - 2x_1 - 2(x_2+h_2) + 2x_2 \end{bmatrix} \\ &= \begin{bmatrix} 2x_1h_1 + h_1^2 +' 3x_2^2h_2 + 3x_2h_2^2 + h_2^3 + -4x_3^3h_3 + -4x_3^2h_3^2 + h_3^4 \\ \\ 2h_1 - 2 + h_2 \end{bmatrix} \end{aligned}$$

et on peut ainsi vérifier l'égalité (2.2) *avec :*

$$Df(x)h = \begin{bmatrix} 2x_1h_1 + 3x_2^2h_2 + 4x_3^3h_3 \\ 2h_1 - h_2 \end{bmatrix}$$

et donc, avec les notations précédentes,

$$J_f(x) = \begin{bmatrix} 2x_1 & 3x_2^2 & 4x_3^3 \\ 2 & -1 & 0 \end{bmatrix}$$

Bien sûr, dans la pratique, on n'a pas besoin de calculer la différentielle en effectuant la différence $f(x+h)-f(x)$. On peut directement calculer les dériées partielles pour calculer la matrice jacobienne J_f.

Cas où $E = \mathbb{R}^n$, $F = \mathbb{R}$ C'est en fait un sous-cas du paragraphe précédent, puisqu'on est ici dans le cas $p = 1$. Soit $x \in \mathbb{R}^n$ et f une fonction de E dans F différentiable en x ; on a donc $J_f(x) \in \mathcal{M}_{1,n}(\mathbb{R})$: J_f est une matrice ligne. On définit le **gradient** de f en x comme le vecteur de $\mathbb{R}^n$ dont les composantes sont les coefficients de la matrice colonne $(J_f(x))^t$, ce qu'on écrit, avec un abus de notation, $\nabla f(x) = (J_f(x))^t \in \mathbb{R}^n$. (L'abus de notation est dû au fait qu'à gauche, il s'agit d'un vecteur de $\mathbb{R}^n$, et à droite, une matrice $n \times 1$, qui sont des objets mathématiques différents, mais qu'on identifie pour alléger les notations). Pour $(x, y) \in (\mathbb{R}^n)^2$, on a donc

$$Df(x)(y) = J_f(x)y = \sum_{j=i}^{n} \partial_j f(x) y_j = \nabla f(x) \cdot y \text{ où } \nabla f(x) = \begin{bmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{bmatrix} \in \mathbb{R}^n.$$

Attention, lorsque l'on écrit $J_f(x)y$ il s'agit d'un *produit matrice vecteur*, alors que lorsqu'on écrit $\nabla f(x) \cdot y$, il s'agit du *produit scalaire entre les vecteurs* $\nabla f(x)$ et y, qu'on peut aussi écrire $\nabla(f(x))^t y$.

Cas où E est un espace de Hilbert et $F = \mathbb{R}$. On généralise ici le cas présenté au paragraphe précédent. Soit $f : E \to \mathbb{R}$ différentiable en $x \in E$. Alors $Df(x) \in \mathcal{L}(E, \mathbb{R}) = E'$, où E' désigne le dual topologique de E, c.à.d. l'ensemble des formes linéaires continues sur E. Par le théorème de représentation de Riesz, il existe un unique $u \in E$ tel que $Df(x)(y) = (u|y)_E$ pour tout $y \in E$, où $(.|.)_E$ désigne le produit scalaire sur E. On appelle encore gradient de f en x ce vecteur u. On a donc $u = \nabla f(x) \in E$ et pour $y \in E$, $Df(x)(y) = (\nabla f(x)|y)_E$.

Différentielle d'ordre 2, matrice hessienne.

Revenons maintenant au cas général de deux espaces vectoriels normés E et F, et supposons maintenant que $f \in C^2(E, F)$. Le fait que $f \in C^2(E, F)$ signifie que $Df \in C^1(E, \mathcal{L}(E, F))$. Par définition, on a $D^2 f(x) \in \mathcal{L}(E, \mathcal{L}(E, F))$ et donc pour $y \in E$, $D^2 f(x)(y) \in \mathcal{L}(E, F)$, et pour $z \in E$, $D^2 f(x)(y)(z) \in F$.
Considérons maintenant le cas particulier $E = \mathbb{R}^n$ et $F = \mathbb{R}$. On a :

$$f \in C^2(\mathbb{R}^n, \mathbb{R}) \Leftrightarrow \left[f \in C^1(\mathbb{R}^n, \mathbb{R}) \text{ et } \nabla f \in C^1(\mathbb{R}^n, \mathbb{R}^n) \right].$$

et

$$D^2 f(x) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$$

Mais à toute application linéaire $\varphi \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$, on peut associer de manière unique une forme bilinéaire ϕ sur $\mathbb{R}^n$ de la manière suivante :

$$\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \tag{2.3}$$

$$(u, v) \mapsto \phi(u, v) = \underbrace{(\varphi(u))}_{\in \mathcal{L}(\mathbb{R}^n, \mathbb{R})} \underbrace{(v)}_{\in \mathbb{R}^n}. \tag{2.4}$$

On dit qu'il existe une isométrie canonique (un isomorphisme qui conserve la norme) entre l'espace vectoriel normé $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$ et l'espace des formes bilinéaires sur $\mathbb{R}^n$.
On appelle matrice hessienne de f et on note $H_f(x)$ la matrice de la forme bilinéaire ainsi associée à l'application linéaire $D^2 f(x) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$.
On a donc $D^2 f(x)(y)(z) = y^t H_f(x) z$. La matrice hessienne $H_f(x)$ peut se calculer à l'aide des dérivées partielles : $H_f(x) = (b_{i,j})_{i,j=1\ldots N} \in \mathcal{M}_n(\mathbb{R})$ où $b_{i,j} = \partial^2_{i,j} f(x)$ et $\partial^2_{i,j}$ désigne la dérivée partielle par rapport

à la variable i de la dérivée partielle par rapport à la variable j. Notons que par définition (toujours avec l'abus de notation qui consiste à identifier les applications linéaires avec les matrices qui les représentent), $Dg(x)$ est la matrice jacobienne de $g = \nabla f$ en x.

Remarque 2.4 (Sur les différentielles, gradient et Hessienne)**.** *Pour définir la différentielle d'une fonction f d'un expace vectoriel de dimension finie E dans $\mathbb{R}$, on a besoin d'une norme sur E.*
Si f est différentiable en $x \in E$, pour définir le gradient de f en x, on a besoin d'un produit scalaire sur E pour pouvoir utiliser le théorème de representation de Riesz mentionné plus haut. Le gradient est défini de manière unique par le produit scalaire, mais ses composantes dépendent de la base choisie.
Enfin, si f est deux fois différentiable en $x \in E$, on a besoin d'une base de E pour définir la matrice hessienne en x, et cette matrice hessienne dépend de la base choisie.

2.2 Les méthodes de point fixe

2.2.1 Point fixe de contraction

Soit $g \in C(\mathbb{R}^n, \mathbb{R}^n)$, on définit la fonction $f \in C(\mathbb{R}^n, \mathbb{R}^n)$ par $f(x) = x - g(x)$. On peut alors remarquer que $g(x) = 0$ si et seulement si $f(x) = x$. Résoudre le système non linéaire (2.1) revient donc à trouver un point fixe de f. Encore faut-il qu'un tel point fixe existe... On rappelle le théorème de point fixe bien connu :

Théorème 2.5 (Point fixe)**.** *Soit E un espace métrique complet, d la distance sur E, et $f : E \to E$ une fonction strictement contractante, c'est–à–dire telle qu'il existe $\kappa \in]0,1[$ tel que $d(f(x), f(y)) \leq \kappa d(x,y)$ pour tout $x, y \in E$. Alors il existe un unique point fixe $\bar{x} \in E$ qui vérifie $f(\bar{x}) = \bar{x}$. De plus si $x^{(0)} \in E$, et $x^{(k+1)} = f(x^{(k)})$, $\forall k \geq 0$, alors $x^{(k)} \to \bar{x}$ quand $n + \infty$.*

DÉMONSTRATION – *Etape 1 : Existence de $\bar{x}$ et convergence de la suite*
Soit $x^{(0)} \in E$ et $(x^{(k)})_{k\in\mathbb{N}}$ la suite définie par $x^{(k+1)} = f(x^{(k)})$ pour $k \geq 0$. On va montrer que :

1. la suite $(x^{(k)})_{k\in\mathbb{N}}$ est de Cauchy (donc convergente car E est complet),
2. $\lim_{n\to+\infty} x^{(k)} = \bar{x}$ est point fixe de f.

Par hypothèse, on sait que pour tout $k \geq 1$,

$$d(x^{(k+1)}, x^{(k)}) = d(f(x^{(k)}), f(x^{(k-1)})) \leq \kappa d(x^{(k)}, x^{(k-1)}).$$

Par récurrence sur k, on obtient que

$$d(x^{(k+1)}, x^{(k)}) \leq \kappa^k d(x^{(1)}, x^{(0)}),\ \forall k \geq 0.$$

Soit $k \geq 0$ et $p \geq 1$, on a donc :

$$\begin{aligned} d(x^{(k+p)}, x^{(k)}) &\leq d(x^{(k+p)}, x^{(k+p-1)}) + \cdots + d(x^{(k+1)}, x^{(k)}) \\ &\leq \sum_{q=1}^{p} d(x^{(k+q)}, x^{(k+q-1)}) \\ &\leq \sum_{q=1}^{p} \kappa^{k+q-1} d(x^{(1)}, x^{(0)}) \\ &\leq d(x^{(1)}, x^{(0)})\kappa^k (1 + \kappa + \ldots + \kappa^{p-1}) \\ &\leq d(x^{(1)}, x^{(0)}) \frac{\kappa^k}{1-\kappa} \longrightarrow 0 \text{ quand } k \to +\infty \text{ car } \kappa < 1. \end{aligned}$$

La suite $(x^{(k)})_{k\in\mathbb{N}}$ est donc de Cauchy, *i.e.* :

$$\forall \varepsilon > 0, \quad \exists k_\varepsilon \in \mathbb{N}\,; \quad \forall k \geq k_\varepsilon, \quad \forall \ p \geq 1 \quad d(x^{(k+p)}, x^{(k)}) \leq \varepsilon.$$

Comme E est complet, on a donc $x^{(k)} \longrightarrow \bar{x}$ dans E quand $k \to +\infty$. Comme la fonction f est strictement contractante, elle est continue, donc on a aussi $f(x^{(k)}) \longrightarrow f(\bar{x})$ dans E quand $k \to +\infty$. En passant à la limite dans l'égalité $x^{(k+1)} = f(x^{(k)})$, on en déduit que $\bar{x} = f(\bar{x})$.

Etape 2 : Unicité

Soit $\bar{x}$ et $\bar{y}$ des points fixes de f, qui satisfont donc $\bar{x} = f(\bar{x})$ et $\bar{y} = f(\bar{y})$. Alors $d(f(\bar{x}), f(\bar{y})) = d(\bar{x}, \bar{y}) \leq \kappa d(\bar{x}, \bar{y})$; comme $\kappa < 1$, ceci est impossible sauf si $\bar{x} = \bar{y}$.

■

La méthode du point fixe s'appelle aussi méthode des itérations successives. Dans le cadre de ce cours, nous prendrons $E = \mathbb{R}^n$, et la distance associée à la norme euclidienne, que nous noterons $|\cdot|$.

$$\forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n \times \mathbb{R}^n \text{ avec } \boldsymbol{x} = (x_1, \ldots, x_n), \boldsymbol{y} = (y_1, \ldots, y_n), d(\boldsymbol{x}, \boldsymbol{y}) = |\boldsymbol{x} - \boldsymbol{y}| = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}.$$

A titre d'illustration, essayons de la mettre en oeuvre pour trouver les points fixes de la fonction $x \mapsto x^2$.

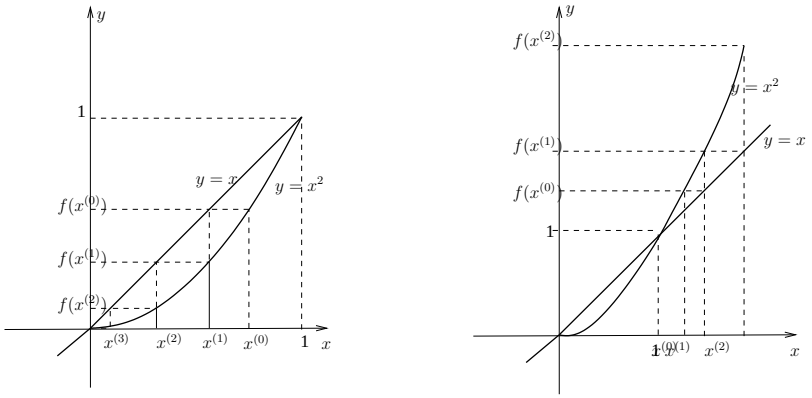


FIGURE 2.1: Comportement des itérés successifs du point fixe pour $x \mapsto x^2$– A gauche : $x^{(0)} < 1$, à droite : $x^{(0)} > 1$.

Pour la fonction $x \mapsto x^2$, on voit sur la figure 2.1, côté gauche, que si l'on part de $x = x^{(0)} < 1$, la méthode converge rapidement vers 0 ; or la fonction $x \mapsto x^2$ n'est strictement contractante que sur l'intervalle $]-\frac{1}{2}, \frac{1}{2}[$. Donc si $x = x^{(0)} \in]-\frac{1}{2}, \frac{1}{2}[$, on est dans les conditions d'application du théorème du point fixe. Mais en fait, la suite $(x^{(k)})_{k \in \mathbb{N}}$ définie par le point fixe converge pour tout $x^{(0)} \in]-1, 1[$; ceci est très facile à voir car $x^{(k)} = (x^{(k)})^2$ et on a donc convergence vers 0 si $|x| < 1$.

Par contre si l'on part de $x^{(0)} > 1$ (à droite sur la figure 2.1), on diverge rapidement : mais rien de surprenant à cela, puisque la fonction $x \mapsto x^2$ n'est pas contractante sur $[1, +\infty[$

Dans le cas de la fonction $x \mapsto \sqrt{x}$, on voit sur la figure 2.2 que les itérés convergent vers 1 que l'on parte à droite ou à gauche de $x = 1$; on peut même démontrer (exercice) que si $x^{(0)} > 0$, la suite $(x)_{k \in \mathbb{N}}$ converge vers 1 lorsque $k \to +\infty$. Pourtant la fonction $x \mapsto \sqrt{x}$ n'est contractante que pour $x > \frac{1}{4}$; mais on n'atteint jamais le point fixe 0, ce qui est moral, puisque la fonction n'est pas contractante en 0. On se rend compte encore sur cet exemple que le théorème du point fixe donne une condition suffisante de convergence, mais que cette condition n'est pas nécessaire.

Remarquons que l'hypothèse que f envoie E dans E est cruciale. Par exemple la fonction $f : x \mapsto \frac{1}{x}$ est lipschitzienne de rapport $k < 1$ sur $[1 + \varepsilon, +\infty[$ pour tout $\varepsilon > 0$ mais elle n'envoie pas $[1 + \varepsilon, +\infty[$ dans $[1 + \varepsilon, +\infty[$. La méthode du point fixe à partir du choix initial $x \neq 1$ donne la suite $x, \frac{1}{x}, x, \frac{1}{x}, \ldots, x, \frac{1}{x}$ qui ne converge pas.

Remarque 2.6 (Vitesse de convergence)**.** *Sous les hypothèses du théorème 2.5, $d(x^{(k+1)}, \bar{x}) = d(f(x^{(k)}), f(\bar{x})) \leq k d(x^{(k)}, \bar{x})$; donc si $x^{(k)} \neq \bar{x}$ alors $\frac{d(x^{(k+1)}, \bar{x})}{d(x^{(k)}, \bar{x})} \leq k$ (< 1), voir à ce sujet la définition 2.14. La convergence est donc au moins linéaire (même si de fait, cette méthode converge en général assez lentement).*

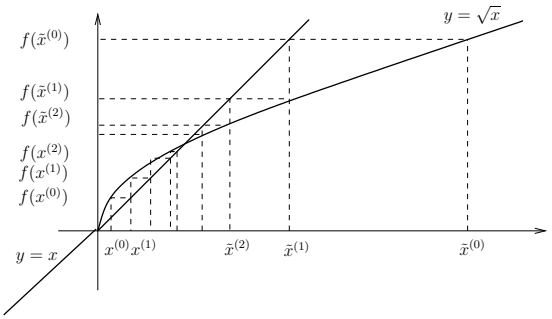


FIGURE 2.2: Comportement des itérés successifs du point fixe pour $x \mapsto \sqrt{x}$

Remarque 2.7 (Généralisation)**.** *Le théorème 2.5 se généralise en remplaçant l'hypothèse " f strictement contractante" par " il existe $k > 0$ tel que $f^{(k)} = \underbrace{f \circ f \circ \ldots \circ f}_{k \text{ fois}}$ est strictement contractante " (reprendre la démonstration du théorème pour le vérifier).*

La question qui vient alors naturellement est : que faire pour résoudre $g(x) = 0$ si la méthode du point fixe appliquée à la fonction $x \mapsto x - g(x)$ ne converge pas ? Dans ce cas, f n'est pas strictement contractante ; une idée possible est de pondérer la fonction g par un paramètre $\omega \neq 0$ et d'appliquer les itérations de point fixe à la fonction $f_\omega(x) = x - \omega g(x)$; on remarque là encore que x est encore solution du système (2.1) si et seulement si x est point fixe de $f_\omega(x)$. On aimerait dans ce cas trouver ω pour que f_ω soit strictement contractante, c.à.d. pour que

$$|f_\omega(x) - f_\omega(y)| = |x - y - \omega(g(x) - g(y))| \leq \kappa |x - y| \text{ pour } (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \text{ avec } \kappa < 1.$$

Or

$$\begin{aligned} |x - y - \omega(g(x) - g(y))|^2 &= \big(x - y - \omega(g(x) - g(y))\big) \cdot \big(x - y - \omega(g(x) - g(y))\big) \\ &= |x - y|^2 - 2(x - y) \cdot (\omega(g(x) - g(y))) + \omega^2 |g(x) - g(y)|^2. \end{aligned}$$

Supposons que g soit lipschitzienne, et soit $M > 0$ sa constante de Lipschitz :

$$|g(x) - g(y)| \leq M|x - y|, \ \forall x, y \in \mathbb{R}^n. \tag{2.5}$$

On a donc

$$|x - y - \omega(g(x) - g(y))|^2 \leq (1 + \omega^2 M^2)|x - y|^2 - 2(x - y) \cdot (\omega(g(x) - g(y)))$$

Or on veut $|x - y - \omega(g(x) - g(y))|^2 \leq \kappa |x - y|^2$, avec $\kappa < 1$. On a donc intérêt à ce que le terme $-2(x - y) \cdot (\omega(g(x) - g(y)))$ soit de la forme $-a|x - y|^2$ avec a strictement positif. Pour obtenir ceci, on va supposer de plus que :

$$\exists \alpha > 0 \text{ tel que } (g(x) - g(y)) \cdot (x - y) \geq \alpha |x - y|^2, \forall x, y \in \mathbb{R}^n, \tag{2.6}$$

On obtient alors :

$$|x - y - \omega(g(x) - g(y))|^2 \leq (1 + \omega^2 M^2 - 2\omega\alpha)|x - y|^2.$$

Et donc si $\omega \in]0, \frac{2\alpha}{M^2}[$, le polynôme $\omega^2 M^2 - 2\omega\alpha$ est strictement négatif : soit $-\mu$ (noter que $\mu \in]0, 1[$) et on obtient que

$$|x - y - \omega(g(x) - g(y))|^2 \leq (1 - \mu)|x - y|^2.$$

On peut donc énoncer le théorème suivant :

Théorème 2.8 (Point fixe de contraction avec relaxation)**.** *On désigne par $|\cdot|$ la norme euclidienne sur $\mathbb{R}^n$. Soit $g \in C(\mathbb{R}^n, \mathbb{R}^n)$ lipschitzienne de constante de Lipschitz $M > 0$, et telle que* (2.6) *est vérifiée : alors la fonction $f_\omega : x \mapsto x - \omega g(x)$ est strictement contractante si $0 < \omega < \dfrac{2\alpha}{M^2}$. Il existe donc un et un seul $\bar{x} \in \mathbb{R}^n$ tel que $g(\bar{x}) = 0$ et $x^{(k)} \to \bar{x}$ quand $n \to +\infty$ avec $x^{(k+1)} = f_\omega(x^{(k)}) = x^{(k)} - \omega g(x^{(k)})$.*

Remarque 2.9. *Le théorème 2.8 permet de montrer que sous les hypothèses* (2.6) *et* (2.5)*, et pour $\omega \in]0, \frac{2\alpha}{M^2}[$, on peut obtenir la solution de* (2.1) *en construisant la suite :*

$$\begin{cases} x^{(k+1)} = x^{(k)} - \omega g(x^{(k)}) \ n \geq 0, \\ x^{(0)} \in \mathbb{R}^n. \end{cases} \tag{2.7}$$

Or on peut aussi écrire cette suite de la manière suivante (avec $f(x) = x - g(x)$) :

$$\begin{cases} \tilde{x}^{(k+1)} &= f(x^{(k)}), \ \forall n \geq 0 \\ x^{(k+1)} &= \omega \tilde{x}^{(k+1)} + (1-\omega)x^{(k)}, \quad x^{(0)} \in \mathbb{R}^n. \end{cases} \tag{2.8}$$

En effet si $x^{(k+1)}$ est donné par la suite (2.8)*, alors*

$$x^{(k+1)} = \omega \tilde{x}^{(k+1)} + (1-\omega)x^{(k)} = \omega f(x^{(k)}) + (1-\omega)x^{(k)} = -\omega g(x^{(k)}) + x^{(k)}.$$

Le procédé de construction de la suite (2.8) *est l'algorithme de relaxation sur f.*

La proposition suivante donne une condition suffisante pour qu'une fonction vérifie les hypothèses (2.6) et (2.5).

Proposition 2.10. *Soit $h \in C^2(\mathbb{R}^n, \mathbb{R})$, et $(\lambda_i)_{i=1,n}$ les valeurs propres de la matrice hessienne de h. On suppose qu'il existe des réels strictement positifs α et M tels que*

$$\alpha \leq \lambda_i(x) \leq M, \ \forall i \in \{1 \dots n\}, \ \forall x \in \mathbb{R}^n.$$

(Notons que cette hypothèse est plausible puisque les valeurs propres de la matrice hessienne sont réelles). Alors la fonction $g = \nabla h$ (gradient de h) vérifie les hypothèses (2.6) *et* (2.5) *du théorème 2.8.*

DÉMONSTRATION – Montrons d'abord que l'hypothèse (2.6) est vérifiée. Soit $(x, y) \in (\mathbb{R}^n)^2$, on veut montrer que $(g(x) - g(y)) \cdot (x - y) \geq \alpha |x - y|^2$. On introduit pour cela la fonction $\varphi \in C^1(\mathbb{R}, \mathbb{R}^n)$ définie par :

$$\varphi(t) = g(x + t(y - x)).$$

On a donc

$$\varphi(1) - \varphi(0) = g(y) - g(x) = \int_0^1 \varphi'(t) dt.$$

Or $\varphi'(t) = Dg(x + t(y - x))(y - x)$. Donc

$$g(y) - g(x) = \int_0^1 Dg(x + t(y - x))(y - x) dt.$$

On en déduit que :

$$(g(y) - g(x)) \cdot (y - x) = \int_0^1 \left(Dg(x + t(y - x))(y - x) \cdot (y - x) \right) dt.$$

Comme $\lambda_i(x) \in [\alpha, M] \ \forall i \in \{1, \dots, n\}$, on a

$$\alpha |w|^2 \leq Dg(z)w \cdot w \leq M |w|^2 \text{ pour tout } w, z \in \mathbb{R}^n$$

On a donc :

$$(g(y) - g(x)) \cdot (y - x) \geq \int_0^1 \alpha |y - x|^2 dt = \alpha |y - x|^2$$

ce qui montre que l'hypothèse (2.6) est bien vérifiée.

Montrons maintenant que l'hypothèse (2.5) est vérifiée. On veut montrer que $|g(y) - g(x)| \leq M|y - x|$. Comme

$$g(y) - g(x) = \int_0^1 Dg(x + t(y - x))(y - x)dt,$$

on a

$$\begin{aligned} |g(y) - g(x)| &\leq \int_0^1 |Dg(x + t(y - x))(y - x)|dt \\ &\leq \int_0^1 |Dg(x + t(y - x))||y - x|dt, \end{aligned}$$

où $|.|$ est la norme sur $\mathcal{M}_n(\mathbb{R})$ induite par la norme euclidienne sur $\mathbb{R}^n$.
Or, comme $\lambda_i(x) \in [\alpha, M]$ pour tout $i = 1, \ldots, n$, la matrice $Dg(x + t(y - x))$ est symétrique définie positive et donc, d'après la proposition 1.30 page 63, son rayon spectral est égal à sa norme, pour la norme induite par la norme euclidienne. On a donc :

$$|Dg(x + t(y - x)| = \rho(Dg(x + t(y - x)) \leq M.$$

On a donc ainsi montré que : $|g(y) - g(x)| \leq M|y - x|$, ce qui termine la démonstration. ∎

2.2.2 Point fixe de monotonie

Dans de nombreux cas issus de la discrétisation d'équations aux dérivées partielles, le problème de résolution d'un problème non linéaire apparaît sous la forme $Ax = R(x)$ où A est une matrice carrée d'ordre n inversible, et $R \in C(\mathbb{R}^n, \mathbb{R}^n)$. On peut le réécrire sous la forme $x = A^{-1}R(x)$ et appliquer l'algorithme de point fixe sur la fonction $f : x \mapsto A^{-1}Rx$, ce qui donne comme itération : $x^{(k+1)} = A^{-1}R(x^{(k)})$. Si on pratique un point fixe avec relaxation, dont le paramètre de relaxation $\omega > 0$, alors l'itération s'écrit :

$$\tilde{x}^{(k+1)} = A^{-1}R(x^{(k)}), \qquad x^{(k+1)} = \omega\tilde{x}^{(k+1)} + (1 - \omega)x^{(k)}.$$

Si la matrice A possède une propriété dite "de monotonie", on peut montrer la convergence de l'algorithme du point fixe ; c'est l'objet du théorème suivant.

Théorème 2.11 (Point fixe de monotonie)**.**
Soient $A \in \mathcal{M}_n(\mathbb{R})$ et $R \in C(\mathbb{R}^n, \mathbb{R}^n)$. On suppose que :

1. *La matrice A est une matrice d'inverse positive, ou IP-matrice (voir exercice 10), c'est-à-dire que A est inversible et tous les coefficients de A^{-1} sont positifs ou nuls, ce qui est équivalent à dire que :*

$$Ax \geq 0 \Rightarrow x \geq 0,$$

au sens composante par composante, c'est-à-dire

$$\big((Ax)_i \geq 0,\ \forall i = 1, \ldots, n\big) \Rightarrow \big(x_i \geq 0,\ \forall i = 1, \ldots, n\big).$$

2. *R est monotone, c'est-à-dire que si $x \geq y$ (composante par composante) alors $R(x) \geq R(y)$ (composante par composante).*
3. *0 est une sous-solution du problème, c'est-à-dire que $0 \leq R(0)$ et il existe $\tilde{x} \in \mathbb{R}^n$; $\tilde{x} \geq 0$ tel que $\tilde{x}$ est une sur-solution du problème, c'est-à-dire que $A\tilde{x} \geq R(\tilde{x})$.*

On pose $x^{(0)} = 0$ et $Ax^{(k+1)} = R(x^{(k)})$. On a alors :

1. $0 \leq x^{(k)} \leq \tilde{x},\ \forall k \in \mathbb{N}$,
2. $x^{(k+1)} \geq x^{(k)},\ \forall k \in \mathbb{N}$,
3. $x^{(k)} \to \bar{x}$ *quand* $k \to +\infty$ *et* $A\bar{x} = R(\bar{x})$.

DÉMONSTRATION – Comme A est inversible la suite $(x^{(k)})_{n\in\mathbb{N}}$ vérifiant

$$\begin{cases} x^{(0)} = 0, \\ Ax^{(k+1)} = R(x^{(k)}), \ \ k \geq 0 \end{cases}$$

est bien définie. On va montrer par récurrence sur k que $0 \leq x^{(k)} \leq \tilde{x}$ pour tout $k \geq 0$ et que $x^{(k)} \leq x^{(k+1)}$ pour tout $k \geq 0$.

1. Pour $k = 0$, on a $x^{(0)} = 0$ et donc $0 \leq x^{(0)} \leq \tilde{x}$ et $Ax^{(1)} = R(0) \geq 0$. On en déduit que $x^{(1)} \geq 0$ grâce aux hypothèses 1 et 3 et donc $x^{(1)} \geq x^{(0)} = 0$.
2. On suppose maintenant (hypothèse de récurrence) que $0 \leq x^{(p)} \leq \tilde{x}$ et $x^{(p)} \leq x^{(p+1)}$ pour tout $p \in \{0, \ldots, n-1\}$. On veut montrer que $0 \leq x^{(k)} \leq \tilde{x}$ et que $x^{(k)} \leq x^{(k+1)}$. Par hypothèse de récurrence pour $p = k-1$, on sait que $x^{(k)} \geq x^{(k-1)}$ et que $x^{(k-1)} \geq 0$. On a donc $x^{(k)} \geq 0$. Par hypothèse de récurrence, on a également que $x^{(k-1)} \leq \tilde{x}$ et grâce à l'hypothèse 2, on a donc $R(x^{(k-1)}) \leq R(\tilde{x})$. Par définition de la suite $(x^{(k)})_{k\in\mathbb{N}}$, on a $Ax^{(k)} = R(x^{(k-1)})$ et grâce à l'hypothèse 3, on sait que $A\tilde{x} \geq R(\tilde{x})$. On a donc : $A(\tilde{x} - x^{(k)}) \geq R(\tilde{x}) - R(x^{(k-1)}) \geq 0$. On en déduit alors (grâce à l'hypothèse 1) que $x^{(k)} \leq \tilde{x}$.
   De plus, comme $Ax^{(k)} = R(x^{(k-1)})$ et $Ax^{(k+1)} = R(x^{(k)})$, on a $A(x^{(k+1)} - x^{(k)}) = R(x^{(k)}) - R(x^{(k-1)}) \geq 0$ par l'hypothèse 2, et donc grâce à l'hypothèse 1, $x^{(k+1)} \geq x^{(k)}$.

On a donc ainsi montré (par récurrence) que

$$\begin{aligned} 0 \leq x^{(k)} \leq \tilde{x}, \quad \forall k \geq 0 \\ x^{(k)} \leq x^{(k+1)}, \quad \forall k \geq 0. \end{aligned}$$

Ces inégalités s'entendent composante par composante, c.à.d. que si $x^{(k)} = (x_1^{(k)} \ldots x_n^{(k)})^t \in \mathbb{R}^n$ et $\tilde{x} = (\tilde{x}_1 \ldots \tilde{x}_n)^t \in \mathbb{R}^n$, alors $0 \leq x_i^{(k)} \leq \tilde{x}_i$ et $x_i^{(k)} \leq x_i^{(k+1)}$, $\forall i \in \{1, \ldots, n\}$, et $\forall k \geq 0$.

Soit $i \in \{1, \ldots, n\}$; la suite $(x_i^{(k)})_{n\in\mathbb{N}} \subset \mathbb{R}$ est croissante et majorée par $\tilde{x}_i$ donc il existe $\bar{x}_i \in \mathbb{R}$ tel que $\bar{x}_i = \lim_{k\to+\infty} x_i^{(k)}$. Si on pose $\bar{x} = (\bar{x}_1 \ldots \bar{x}_n)^t \in \mathbb{R}^n$, on a donc $x^{(k)} \longrightarrow \bar{x}$ quand $k \to +\infty$.

Enfin, comme $Ax^{(k+1)} = R(x^{(k)})$ et comme R est continue, on obtient par passage à la limite lorsque $k \to +\infty$ que $A\bar{x} = R(\bar{x})$ et que $0 \leq \bar{x} \leq \tilde{x}$. ∎

L'hypothèse 1 du théorème 2.11 est vérifiée par exemple par les matrices A qu'on a obtenues par discrétisation par différences finies des opérateurs $-u''$ sur l'intervalle $]0,1[$ (voir page 11 et l'exercice 52) et Δu sur $]0,1[\times]0,1[$ (voir page 14).

Théorème 2.12 (Généralisation du précédent)**.**
Soit $A \in \mathcal{M}_n(\mathbb{R})$, $R \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, $R = (R_1, \ldots, R_n)^t$ tels que

1. *Pour tout $\beta \geq 0$ et pour tout $x \in \mathbb{R}^n$, $Ax + \beta x \geq 0 \Rightarrow x \geq 0$*
2. *$\dfrac{\partial R_i}{\partial x_j} \geq 0$, $\forall i, j$ t.q. $i \neq j$ (R_i est monotone croissante par rapport à la variable x_j si $j \neq i$) et $\exists \gamma > 0$, $-\gamma \leq \dfrac{\partial R_i}{\partial x_i} \leq 0$, $\forall x \in \mathbb{R}^n$, $\quad \forall i \in \{1, \ldots, n\}$ (R_i est monotone décroissante par rapport à la variable x_i).*
3. *$0 \leq R(0)$ (0 est sous-solution) et il existe $\tilde{x} \geq 0$ tel que $A(\tilde{x}) \geq R(\tilde{x})$ ($\tilde{x}$ est sur-solution).*

Soient $x^{(0)} = 0$, $\beta \geq \gamma$, et $(x^{(k)})_{n\in\mathbb{N}}$ la suite définie par $Ax^{(k+1)} + \beta x^{(k+1)} = R(x^{(k)}) + \beta x^{(k)}$. Cette suite converge vers $\bar{x} \in \mathbb{R}^n$ et $A\bar{x} = R(\bar{x})$. De plus, $0 \leq x^{(k)} \leq \tilde{x}$ $\forall n \in \mathbb{N}$ et $x^{(k)} \leq x^{(k+1)}$, $\forall n \in \mathbb{N}$.

DÉMONSTRATION – On se ramène au théorème précédent avec $A + \beta Id$ au lieu de A et $R + \beta$ au lieu de R. ∎

Remarque 2.13 (Point fixe de Brouwer)**.** *On s'est intéressé ici uniquement à des théorèmes de point fixe "constructifs", i.e. qui donnent un algorithme pour le déterminer. Il existe aussi un théorème de point fixe dans $\mathbb{R}^n$ avec des*

hypothèses beaucoup plus générales (mais le théorème est non constructif), c'est le théorème de Brouwer[3] *: si f est une fonction continue de la boule unité de $\mathbb{R}^n$ dans la boule unité, alors elle admet un point fixe dans la boule unité.*

2.2.3 Vitesse de convergence

Définition 2.14 (Vitesse de convergence)**.** *Soit $(x^{(k)})_{n\in\mathbb{N}} \in \mathbb{R}^n$ et $\bar{x} \in \mathbb{R}^n$. On suppose que $x^{(k)} \to \bar{x}$ lorsque $k \to +\infty$, que la suite est non stationnaire, c.à.d. que $x^{(k)} \neq \bar{x}$ pour tout $k \in \mathbb{N}$, et que*

$$\lim_{k\to+\infty} \frac{\|x^{(k+1)} - \bar{x}\|}{\|x^{(k)} - \bar{x}\|} = \beta \in [0,1]. \tag{2.9}$$

On s'intéresse à la "vitesse de convergence" de la suite $(x^{(k)})_{k\in\mathbb{N}}$. On dit que :

1. *La convergence est* **sous–linéaire** *si $\beta = 1$.*
2. *La convergence est* **au moins linéaire** *si $\beta \in [0,1[$.*
3. *La convergence* **est linéaire** *si $\beta \in]0,1[$.*
4. *La convergence est* **super linéaire** *si $\beta = 0$. Dans ce cas, on dit également que :*
   (a) *La convergence est* **au moins quadratique** *s'il existe $\gamma \in \mathbb{R}_+$ et il existe $n_0 \in \mathbb{N}$ tels que si $k \geq n_0$ alors $\|x^{(k+1)} - \bar{x}\| \leq \gamma\|x^{(k)} - \bar{x}\|^2$.*
   (b) *La convergence est* **quadratique** *si*

$$\lim_{k\to+\infty} \frac{\|x^{(k+1)} - \bar{x}\|}{\|x^{(k)} - \bar{x}\|^2} = \gamma > 0.$$

   Plus généralement, on dit que :
   (a) *La convergence est* **au moins d'ordre** *p s'il existe $\gamma \in \mathbb{R}_+$ et il existe $k_0 \in \mathbb{N}$ tels que si $k \geq k_0$ alors $\|x^{(k+1)} - \bar{x}\| \leq \gamma\|x^{(k)} - \bar{x}\|^p$.*
   (b) *La convergence est* **d'ordre** *p si*

$$\lim_{k\to+\infty} \frac{\|x^{(k+1)} - \bar{x}\|}{\|x^{(k)} - \bar{x}\|^p} = \gamma > 0.$$

Remarque 2.15 (Sur la vitesse de convergence des suites)**.**

— Remarquons d'abord que si une suite $(x^{(k)})_{k\in\mathbb{N}}$ de $\mathbb{R}^n$ converge vers $\bar{x}$ lorsque k tend vers l'infini, et qu'il existe β vérifiant (2.9), alors on a forcément $\beta \leq 1$. En effet, si la suite vérifie (2.9) avec $\beta > 1$, alors il existe $k_0 \in \mathbb{N}$ tel que si $k \geq k_0$, $|x_n - \bar{x}| \geq |x_{k_0} - \bar{x}|$ pour tout $k \geq k_0$, ce qui contredit la convergence.

— *Quelques exemples de suites qui convergent sous-linéairement* : $x_k = \frac{1}{\sqrt{k}}$, $x_k = \frac{1}{k}$, mais aussi, de manière moins intuitive : $x_k = \frac{1}{k^2}$. Toutes ces suites vérifient l'égalité (2.9) avec $\beta = 1$.

— Attention donc, contrairement à ce que pourrait suggérer son nom, la convergence linéaire (au sens donné ci-dessus), est déjà une convergence très rapide. Les suites géométriques définies par $x_k = \beta^k$ avec $\beta \in]0,1[$ sont des suites qui convergent linéairement (vers 0), car elles verifient évidemment bien (2.9) avec $\beta \in]0,1[$.

— la convergence quadratique est encore plus rapide ! Par exemple la suite définie par $x_{k+1} = x_k^2$ converge de manière quadratique pour un choix initial $x_0 \in]-1,1[$. Mais si par malheur le choix initial est en dehors

3. Luitzen Egbertus Jan Brouwer (1881-1966), mathématicien néerlandais.

de cet intervalle, la suite diverge alors très vite... de manière exponentielle, en fait (pour $x_0 > 1$, on a $x_k = e^{2k \ln x_0}$).

C'est le cas de la méthode de Newton, que nous allons introduire maintenant. Lorsqu'elle converge, elle converge très vite (nous démontrerons que la vitesse de convergence est quadratique). Mais lorsqu'elle diverge, elle diverge aussi très vite...

Pour construire des méthodes itératives qui convergent "super vite", nous allons donc essayer d'obtenir des vitesses de convergence super linéaires. C'est dans cet esprit que nous étudions dans la proposition suivante des conditions suffisantes de convergence de vitesse quadratique pour une méthode de type point fixe, dans le cas d'une fonction f de $\mathbb{R}$ dans $\mathbb{R}$.

Proposition 2.16 (Vitesse de convergence d'une méthode de point fixe)**.** *Soit $f \in C^1(\mathbb{R}, \mathbb{R})$; on suppose qu'il existe $\bar{x} \in \mathbb{R}$ tel que $f(\bar{x}) = \bar{x}$. On construit la suite*

$$x^{(0)} \in \mathbb{R}$$
$$x^{(k+1)} = f(x^{(k)}).$$

1. *Si on suppose que $f'(\bar{x}) \neq 0$ et $|f'(\bar{x})| < 1$, alors il existe $\alpha > 0$ tel que si $x^{(0)} \in I_\alpha = [\bar{x} - \alpha, \bar{x} + \alpha]$ on a $x^{(k)} \to \bar{x}$ lorsque $k \to +\infty$. De plus si $x^{(k)} \neq \bar{x}$ pour tout $k \in \mathbb{N}$, alors*

$$\frac{|x^{(k+1)} - \bar{x}|}{|x^{(k)} - \bar{x}|} \to |f'(\bar{x})| = \beta \text{ avec } \beta \in]0, 1[.$$

   La convergence est donc linéaire.

2. *Si on suppose maintenant que $f'(\bar{x}) = 0$ et $f \in C^2(\mathbb{R}, \mathbb{R})$, alors il existe $\alpha > 0$ tel que si $x^{(0)} \in I_\alpha = [\bar{x} - \alpha, \bar{x} + \alpha]$, alors $x^{(k)} \to \bar{x}$ quand $k \to +\infty$, et si $x^{(k)} \neq \bar{x}$, $\forall n \in \mathbb{N}$ alors*

$$\frac{|x^{(k+1)} - \bar{x}|}{|x^{(k)} - \bar{x}|^2} \to \beta = \frac{1}{2}|f''(\bar{x})|.$$

   Dans ce cas, la convergence est donc au moins quadratique.

DÉMONSTRATION –

1. Supposons que $|f'(\bar{x})| < 1$, et montrons qu'il existe $\alpha > 0$ tel que si $x^{(0)} \in I_\alpha$ alors $x^{(k)} \to \bar{x}$. Comme $f \in C^1(\mathbb{R}, \mathbb{R})$ il existe $\alpha > 0$ tel que $\gamma = \max_{x \in I_\alpha} |f'(x)| < 1$ (par continuité de f').

   On va maintenant montrer que $f : I_\alpha \to I_\alpha$ est strictement contractante, on pourra alors appliquer le théorème du point fixe à $f_{|I_\alpha}$, (I_α étant fermé), pour obtenir que $x^{(k)} \to \bar{x}$ où $\bar{x}$ est l'unique point fixe de $f_{|I_\alpha}$.

   Soit $x \in I_\alpha$; montrons d'abord que $f(x) \in I_\alpha$: comme $f \in C^1(\mathbb{R}, \mathbb{R})$, il existe $\xi \in]x, \bar{x}[$ tel que $|f(x) - \bar{x}| = |f(x) - f(\bar{x})| = |f'(\xi)||x - \bar{x}| \leq \gamma |x - \bar{x}| < \alpha$, ce qui prouve que $f(x) \in I_\alpha$. On vérifie alors que $f_{|I_\alpha}$ est strictement contractante en remarquant que pour tous $x, y \in I_\alpha$, $x < y$, il existe $\xi \in]x, y[(\subset I_\alpha)$ tel que $|f(x) - f(y)| = |f'(\xi)||x - y| \leq \gamma |x - y|$ avec $\gamma < 1$. On a ainsi montré que $x^{(k)} \to \bar{x}$ si $x^{(0)} \in I_\alpha$.

   Cherchons maintenant la vitesse de convergence de la suite. Supposons que $f'(\bar{x}) \neq 0$ et $x^{(k)} \neq \bar{x}$ pour tout $n \in \mathbb{N}$. Comme $x^{(k+1)} = f(x^{(k)})$ et $\bar{x} = f(\bar{x})$, on a $|x^{(k+1)} - \bar{x}| = |f(x^{(k)}) - f(\bar{x})|$. Comme $f \in C^1(\mathbb{R}, \mathbb{R})$, il existe $\xi_k \in]x^{(k)}, \bar{x}[$ ou $]\bar{x}, x^{(k)}[$, tel que $f(x^{(k)}) - f(\bar{x}) = f'(\xi_k)(x^{(k)} - \bar{x})$. On a donc

$$\frac{|x^{(k+1)} - \bar{x}|}{|x^{(k)} - \bar{x}|} = |f'(\xi_k)| \longrightarrow |f'(\bar{x})| \text{ car } x^{(k)} \to \bar{x} \text{ et } f' \text{ est continue.}$$

   On a donc une convergence linéaire.

2. Supposons maintenant que $f \in C^2(\mathbb{R}, \mathbb{R})$ et $f'(\bar{x}) = 0$. On sait déjà par ce qui précède qu'il existe $\alpha > 0$ tel que si $x^{(0)} \in I_\alpha$ alors $x^{(k)} \to \bar{x}$ lorsque $k \to +\infty$. On veut estimer la vitesse de convergence ; on suppose pour cela que $x^{(k)} \neq \bar{x}$ pour tout $k \in \mathbb{N}$. Comme $f \in C^2(\mathbb{R}, \mathbb{R})$, il existe $\xi_k \in]x^{(k)}, \bar{x}[$ tel que

$$f(x^{(k)}) - f(\bar{x}) = f'(\bar{x})(x^{(k)} - \bar{x}) + \frac{1}{2} f''(\xi_k)(x^{(k)} - \bar{x})^2.$$

On a donc : $x^{(k+1)} - \bar{x} = \frac{1}{2} f''(\xi_k)(x^{(k)} - \bar{x})^2$ ce qui entraîne, par continuité de f'', que

$$\frac{|x^{(k+1)} - \bar{x}|}{|x^{(k)} - \bar{x}|^2} = \frac{1}{2}|f''(\xi_k)| \longrightarrow \frac{1}{2}|f''(\bar{x})| \text{ quand } k \to +\infty.$$

La convergence est donc au moins quadratique.

■

2.2.4 Méthode de Newton dans $\mathbb{R}$

On va étudier dans le paragraphe suivant la méthode de Newton pour la résolution d'un système non linéaire. (En fait, il semble que l'idée de cette méthode revienne plutôt à Simpson [4] Donnons l'idée de la méthode de Newton dans le cas $n = 1$ à partir des résultats de la proposition précédente. Soit $g \in C^3(\mathbb{R}, \mathbb{R})$ et $\bar{x} \in \mathbb{R}$ tel que $g(\bar{x}) = 0$. On cherche une méthode de construction d'une suite $(x^{(k)})_{k\in\mathbb{N}} \subset \mathbb{R}^n$ qui converge vers $\bar{x}$ de manière quadratique. On pose

$$f(x) = x - h(x)g(x) \text{ avec } h \in C^2(\mathbb{R}, \mathbb{R}) \text{ tel que } h(x) \neq 0 \ \forall x \in \mathbb{R},$$

et on a donc

$$f(x) = x \Leftrightarrow g(x) = 0.$$

Si par miracle $f'(\bar{x}) = 0$, la méthode de point fixe sur f va donner (pour $x^{(0)} \in I_\alpha$ donné par la proposition 2.16) $(x^{(k)})_{n\in\mathbb{N}}$ tel que $x^{(k)} \to \bar{x}$ de manière au moins quadratique. Or on a $f'(x) = 1 - h'(x)g(x) - g'(x)h(x)$ et donc $f'(\bar{x}) = 1 - g'(\bar{x})h(\bar{x})$. Il suffit donc de prendre h tel que $h(\bar{x}) = \dfrac{1}{g'(\bar{x})}$. Ceci est possible si $g'(\bar{x}) \neq 0$.

En résumé, si $g \in C^3(\mathbb{R}, \mathbb{R})$ est telle que $g'(\overline{x}) \neq 0$ et $g(\bar{x}) = 0$, on peut construire, pour x assez proche de $\bar{x}$, la fonction $f \in C^2(\mathbb{R}, \mathbb{R})$ définie par

$$f(x) = x - \frac{g(x)}{g'(x)}.$$

Grâce à la proposition 2.16, il existe $\alpha > 0$ tel que si $x^{(0)} \in I_\alpha$ alors la suite définie par

$$x^{(k+1)} = f(x^{(k)}) = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})}$$

converge vers $\bar{x}$ de manière au moins quadratique.

Remarquons que dans le cas $n = 1$, la suite de Newton peut s'obtenir naturellement en remplaçant l'équation $g(\overline{x}) = 0$ par $g(x^{(k+1)}) = 0$, et $g(x^{(k+1)})$ par le développement limité en x^k :

$$g(x^{(k+1)}) = g(x^{(k)}) + g'(x^{(k)})(x^{(k+1)} - x^{(k)}) + |x^{(k+1)} - x^{(k)}|\epsilon(x^{(k+1)} - x^{(k)}).$$

C'est le plus sûr moyen mnémotechnique pour retrouver l'itération de Newton :

$$g(x^{(k)}) + g'(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0 \text{ ou encore } g'(x^{(k)})(x^{(k+1)} - x^{(k)}) = -g(x^{(k)}). \tag{2.10}$$

Comparons sur un exemple les méthodes de point fixe et de Newton. On cherche le zéro de la fonction $g : x \mapsto x^2 - 3$ sur $\mathbb{R}_+$. Notons en passant que la construction de la suite $x^{(k)}$ par point fixe ou Newton permet l'approximation effective de $\sqrt{3}$. Si on applique le point fixe standard, la suite $x^{(k)}$ s'écrit

$$x^{(0)} \text{ donné ,}$$
$$x^{(k+1)} = x^{(k)} - (x^{(k)})^2 + 3.$$

4. Voir Nick Kollerstrom (1992). *Thomas Simpson and "Newton's method of approximation" : an enduring myth*, The British Journal for the History of Science, 25, pp 347-354 doi :10.1017/S0007087400029150 – Thomas Simpson est un mathématicien anglais du 18-ème siècle à qui on attribue généralement la méthode du même nom pour le calcul approché des intégrales, probablement à tort car celle-ci apparaît déjà dans les travaux de Kepler deux siècles plus tôt !

Si on applique le point fixe avec paramètre de relaxation ω, la suite $x^{(k)}$ s'écrit

$$x^{(0)} \text{ donné},$$
$$x^{(k+1)} = -x^{(k)} + \omega(-x^{(k)})^2 + 3)$$

Si maintenant on applique la méthode de Newton, la suite $x^{(k)}$ s'écrit

$$x^{(0)} \text{ donné},$$
$$x^{(k+1)} = -\frac{(x^{(k)})^2 - 3}{2x^{(k)}}.$$

Comparons les suites produites par scilab à partir de $x^{(0)} = 1$ par le point fixe standard, le point fixe avec relaxation ($\omega = .1$) et la méthode de Newton.

— **point fixe standard :** 1. 3. - 3 - 9 - 87 - 7653 - 58576059 - 3.431D+15 - 1.177D+31

— **point fixe avec relaxation :**
1. 1.2 1.356 1.4721264 1.5554108 1.6134805 1.6531486 1.6798586 1.6976661 1.7094591 1.717234 1.7223448 1.7256976 1.7278944 1.7293325 1.7302734 1.7308888 1.7312912 1.7315543 1.7317263 1.7318387 1.7319122 1.7319602 1.7319916 1.7320121 1.73204 1.7320437 1.7320462 1.7320478 1.7320488 1.7320495 1.73205 1.7320503 1.7320504 1.7320506 1.7320507 1.7320507 1.7320507 1.7320508

— **Newton :**
1. 2. 1.75 1.7321429 1.7320508 1.7320508

Remarque 2.17 (Attention à l'utilisation du théorème des accroissements finis…)**.** *On a fait grand usage du théorème des accroissements finis dans ce qui précède. Rappelons que sous la forme qu'on a utilisée, ce théorème n'est valide que pour les fonctions de* $\mathbb{R}$ *dans* $\mathbb{R}$*. On pourra s'en convaincre en considérant la fonction de* $\mathbb{R}$ *dans* $\mathbb{R}^2$ *définie par :*

$$\varphi(x) = \begin{bmatrix} \sin x \\ \cos x \end{bmatrix}.$$

On peut vérifier facilement qu'il n'existe pas de $\xi \in \mathbb{R}$ *tel que* $\varphi(2\pi) - \varphi(0) = 2\pi\varphi'(\xi)$.

2.2.5 Exercices (méthodes de point fixe)

Exercice 77 (Calcul différentiel)**.** *Suggestions en page 161, corrigé détaillé en page 161*

Soit $f \in C^2(\mathbb{R}^n, \mathbb{R})$.
1. Montrer que pour tout $x \in \mathbb{R}^n$, il existe un unique vecteur $a(x) \in \mathbb{R}^n$ tel que $Df(x)(h) = a(x) \cdot h$ pour tout $h \in \mathbb{R}^n$.
Montrer que $(a(x))_i = \partial_i f(x)$.

2. On pose $\nabla f(x) = (\partial_1 f(x), \ldots, \partial_1 f(x))^t$. Soit φ l'application définie de $\mathbb{R}^n$ dans $\mathbb{R}^n$ par $\varphi(x) = \nabla f(x)$. Montrer que $\varphi \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ et que $D\varphi(x)(y) = A(x)y$, où $(A(x))_{i,j} = \partial^2_{i,j} f(x)$.

Exercice 78 (Calcul différentiel, suite)**.** *Corrigé en page 162*

1. Soit $f \in C^2(\mathbb{R}^2, \mathbb{R})$ la fonction définie par $f(x_1, x_2) = ax_1 + bx_2 + cx_1x_2$, où a, b, et c sont trois réels fixés. Donner la définition et l'expression de $Df(x)$, $\nabla f(x)$,Df, $D^2 f(x)$, $H_f(x)$.
2. Même question pour la fonction $f \in C^2(\mathbb{R}^3, \mathbb{R})$ définie par $f(x_1, x_2, x_3) = x_1^2 + x_1^2x_2 + x_2 \sin(x_3)$.

Exercice 79 (Point fixe dans $\mathbb{R}$)**.** *Corrigé en page 163*

1. Etudier la convergence de la suite $(x^{(k)})_{k\in\mathbb{N}}$, définie par $x^{(0)} \in [0, 1]$ et $x^{(k+1)} = \cos\left(\frac{1}{1 + x^{(k)}}\right)$.

2. Soit $I = [0, 1]$, et $f : x \mapsto x^4$. Montrer que la suite des itérés de point fixe converge pour tout $x \in [0, 1]$ et donner la limite de la suite en fonction du choix initial $x^{(0)}$.

Exercice 80 (Point fixe et Newton). *Corrigé détaillé en page 163.*

1. On veut résoudre l'équation $2xe^x = 1$.
   (a) Vérifier que cette équation peut s'écrire sous forme de point fixe : $x = \frac{1}{2}e^{-x}$.
   (b) Ecrire l'algorithme de point fixe, et calculer les itérés x_0, x_1, x_2 et x_3 en partant depuis $x_0 = 1$.
   (c) Justifier la convergence de l'algorithme donné en (b).
2. On veut résoudre l'équation $x^2 - 2 = 0$, $x > 0$.
   (a) Vérifier que cette équation peut s'écrire sous forme de point fixe : $x = \frac{2}{x}$.
   (b) Ecrire l'algorithme de point fixe, et tracer sur un graphique les itérés x_0, x_1, x_2 et x_3 en partant de $x_0 = 1$ et $x_0 = 2$.
   (c) Essayer ensuite le point fixe sur $x = \frac{x^2+2}{2x}$. Pas très facile à deviner, n'est ce pas ?
   (d) Pour suivre les traces de Newton (ou plutôt Simpson, semble-t-il) : à x_n connu, écrire le développement limité de $g(x) = x^2 - 2$ entre $x^{(n)}$ et $x^{(n+1)}$, remplacer l'équation $g(\overline{x}) = 0$ par $g(x^{(n+1)}) = 0$, et $g(x^{(n+1)})$ par le développement limité en x^{n+1} , et en déduire l'approximation $x^{(n+1)} = x^{(n)} - \frac{g(x^{(n)})}{g'(x^{(n)})}$. Retrouver ainsi l'itération de la question précédente (pour $g(x) = x^2 - 2$).

Exercice 81 (Méthode de monotonie). *Suggestions en page 161, corrigé détaillé en page 164.*

On suppose que $f \in C^1(\mathbb{R}, \mathbb{R})$, $f(0) = 0$ et que f est croissante. On s'intéresse, pour $\lambda > 0$, au système non linéaire suivant de n équations à n inconnues (notées $u_1, \ldots, u_n$) :

$$\begin{array}{l} (Au)_i = \alpha_i f(u_i) + \lambda b_i \ \forall i \in \{1, \ldots, n\}, \\ u = (u_1, \ldots, u_n)^t \in \mathbb{R}^n, \end{array} \tag{2.11}$$

où $\alpha_i > 0$ pour tout $i \in \{1, \ldots, n\}$, $b_i \geq 0$ pour tout $i \in \{1, \ldots, n\}$ et $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice vérifiant

$$u \in \mathbb{R}^n, \ Au \geq 0 \Rightarrow u \geq 0. \tag{2.12}$$

On suppose qu'il existe $\mu > 0$ t.q. (2.11) ait une solution, notée $u^{(\mu)}$, pour $\lambda = \mu$. On suppose aussi que $u^{(\mu)} \geq 0$. Soit $0 < \lambda < \mu$. On définit la suite $(v^{(k)})_{n\in\mathbb{N}} \subset \mathbb{R}^n$ par $v^{(0)} = 0$ et, pour $n \geq 0$,

$$(Av^{(k+1)})_i = \alpha_i f(v_i^{(k)}) + \lambda b_i \ \forall i \in \{1, \ldots, n\}. \tag{2.13}$$

Montrer que la suite $(v^{(k)})_{n\in\mathbb{N}}$ est bien définie, convergente (dans $\mathbb{R}^n$) et que sa limite, notée $u^{(\lambda)}$, est solution de (2.11) (et vérifie $0 \leq u^{(\lambda)} \leq u^{(\mu)}$).

Exercice 82 (Point fixe amélioré). *Suggestions en page 161, Corrigé en page 165*

Soit $g \in C^3(\mathbb{R}, \mathbb{R})$ et $\overline{x} \in \mathbb{R}$ tels que $g(\overline{x}) = 0$ et $g'(\overline{x}) \neq 0$.
On se donne $\varphi \in C^1(\mathbb{R}, \mathbb{R})$ telle que $\varphi(\overline{x}) = \overline{x}$.
On considère l'algorithme suivant :

$$\left\{ \begin{array}{l} x_0 \in \mathbb{R}, \\ \\ x_{n+1} = h(x_n), n \geq 0. \end{array} \right. \tag{2.14}$$

avec $h(x) = x - \dfrac{g(x)}{g'(\varphi(x))}$

1) Montrer qu'il existe $\alpha > 0$ tel que si $x_0 \in [\overline{x} - \alpha, \overline{x} + \alpha] = I_\alpha$, alors la suite donnée par l'algorithme (2.14) est bien définie ; montrer que $x_n \to \overline{x}$ lorsque $n \to +\infty$.

On prend maintenant $x_0 \in I_\alpha$ où α est donné par la question 1.

2) Montrer que la convergence de la suite $(x_n)_{n\in\mathbb{N}}$ définie par l'algorithme (2.14) est au moins quadratique.
3) On suppose que φ' est lipschitzienne et que $\varphi'(\overline{x}) = \dfrac{1}{2}$. Montrer que la convergence de la suite $(x_k)_{k\in\mathbb{N}}$ définie par (2.14) est au moins cubique, c'est-à-dire qu'il existe $c \in \mathbb{R}_+$ tel que

$$|x_{k+1} - \overline{x}| \leq c|x_k - \overline{x}|^3, \qquad \forall k \geq 1.$$

4) Soit $\beta \in \mathbb{R}_+^*$ tel que $g'(x) \neq 0 \quad \forall x \in I_\beta =]\overline{x} - \beta, \overline{x} + \beta[$; montrer que si on prend φ telle que :

$$\varphi(x) = x - \frac{g(x)}{2g'(x)} \qquad \text{si } x \in I_\beta,$$

alors la suite définie par l'algorithme (2.14) converge de manière cubique.

Suggestions

Exercice 77 page 159 (Calcul différentiel) 1. Utiliser le fait que $Df(x)$ est une application linéaire et le théorème de Riesz. Appliquer ensuite la différentielle à un vecteur h bien choisi.
2. Mêmes idées...

Exercice 81 page 160 (Méthode de monotonie) Pour montrer que la suite $(v^{(k)})_{n\in\mathbb{N}}$ est bien définie, remarquer que la matrice A est inversible. Pour montrer qu'elle est convergente, montrer que les hypothèses du théorème du point fixe de monotonie vu en cours sont vérifiées.

Exercice 82 page 160 (Point fixe amélioré)
1) Montrer qu'on peut choisir α de manière à ce que $|h'(x)| < 1$ si $x \in I_\alpha$, et en déduire que $g'(\varphi(x_n) \neq 0$ si x_0 est bien choisi.
2) Remarquer que

$$|x_{k+1} - \overline{x}| = (x_k - \overline{x})(1 - \frac{g(x_k) - g(\overline{x})}{(x_k - \overline{x})g'(\varphi(x_k))}. \tag{2.15}$$

En déduire que

$$|x_{n+1} - \overline{x}| \leq \frac{1}{\varepsilon}|x_n - \overline{x}|^2 \sup_{x\in I_\alpha} |\varphi'(x)| \sup_{x\in I_\alpha} |g''(x)|.$$

3) Reprendre le même raisonnement avec des développements d'ordre supérieur.
4) Montrer que φ vérifie les hypothèses de la question 3).

Corrigés

Exercice 77 page 159 1. Par définition, $T = Df(x)$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^n$, qui s'écrit donc sous la forme : $T(h) = \sum_{i=1}^n a_i h_i = a \cdot h$. Or l'application T dépend de x, donc le vecteur a aussi.
Montrons maintenant que $(a(x))_i = \partial_i f(x)$, pour $1 \leq i \leq n$ Soit $h^{(i)} \in \mathbb{R}^n$ défini par $h_j^{(i)} = h\delta_{i,j}$, où $h > 0$ et $\delta_{i,j}$ désigne le symbole de Kronecker, i.e. $\delta_{i,j} = 1$ si $i = j$ et $\delta_{i,j} = 0$ sinon. En appliquant la définition de la différentielle avec $h^{(i)}$, on obtient :

$$f(x + h^{(i)}) - f(x) = Df(x)(h^{(i)}) + \|h^{(i)}\|\varepsilon(h^{(i)}),$$

c'est–à–dire :

$$f(x_1, \ldots, x_{i-1}, x_i + h, x_{i-1}, \ldots, x_n) - f(x_1, \ldots, x_n) = (a(x))_i h + h\varepsilon(h^{(i)}).$$

En divisant par h et en faisant tendre h vers 0, on obtient alors que $(a(x))_i = \partial_i f(x)$.

2. Comme $f \in C^2(\mathbb{R}^n, \mathbb{R})$, on a $\partial_i f \in C^1(\mathbb{R}^n, \mathbb{R})$, et donc $\varphi \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Comme $D\varphi(x)$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^n$, il existe une matrice $A(x)$ carrée d'ordre n telle que $D\varphi(x)(y) = A(x)y$

pour tout $y \in \mathbb{R}^n$. Il reste à montrer que $(A(x))_{i,j} = \partial^2_{i,j} f(x)$. Soit $h^{(i)} \in \mathbb{R}^n$ défini à la question préceédente, pour $i,\ j = 1, \ldots, n$, on a

$$(D\varphi(x)(h^{(j)}))_i = (A(x)h^{(j)})_i = \sum_{k=1}^{n} a_{i,k}(x)h^{(j)}_k = ha_{i,j}(x).$$

Or par définition de la différentielle,

$$\varphi_i(x + h^{(j)}) - \varphi_i(x) = (D\varphi(x)(h^{(j)}))_i + \|h^{(j)}\|\varepsilon_i(h^{(j)}),$$

ce qui entraîne, en divisant par h et en faisant tendre h vers 0 : $\partial_j\varphi_i(x) = a_{i,j}(x)$. Or $\varphi_i(x) = \partial_i f(x)$, et donc $(A(x))_{i,j} = a_{i,j}(x) = \partial^2_{i,j} f(x)$.

Exercice 78 page 159 (Calcul différentiel, suite)
1. $Df(x)$ est la différentielle de f en x, c'est-à-dire l'application linéaire telle que $f(x+h) - f(x) - Df(x)h = h\varepsilon(h)$ pour tout $h \in \mathbb{R}^2$, où $\varepsilon(h)$ tend vers 0 lorsque $|h|$ tend vers 0. Calculons les dérivées partielles de f.

$$\begin{aligned}\partial_1 f(x_1, x_2) &= a + cx_2,\\ \partial_2 f(x_1, x_2) &= b + cx_1.\end{aligned}$$

$Df(x)$ est donc l'application linéaire qui a $(h_1, h_2) \in \mathbb{R}^2$ associe $\partial_1 f(x_1, x_2)h_1 + \partial_2 f(x_1, x_2)h_2 = (a+cx_2)h_1 + (b + cx_1)h_2$.
Par définition du gradient, on a : $Df(x)h = \nabla f(x) \cdot h$ et

$$\nabla f(x) = \begin{bmatrix}\partial_1 f(x_1, x_2)\\ \partial_2 f(x_1, x_2)\end{bmatrix} = \begin{bmatrix}a + cx_2\\ b + cx_1\end{bmatrix}$$

Df est la différentielle de f, c'est-à-dire la fonction de $\mathbb{R}^2$ dans $\mathcal{L}(\mathbb{R}^2, \mathbb{R})$ qui a $x = (x_1, x_2)$ associe $Df(x)$ définie plus haut.
La différentielle d'ordre 2 de f en x est une application linéaire de $\mathbb{R}^2$ dans $\mathcal{L}(\mathbb{R}^2, \mathbb{R})$, telle que $Df(x + h) - Df(x) - D^2 f(x)(h) = |h|\varepsilon(h)$ pour tout $h \in \mathbb{R}^2$, où $\varepsilon(h)$ tend vers 0 lorsque $|h|$ tend vers 0 (noter que $\varepsilon(h) \in \mathcal{L}(\mathbb{R}^2, \mathbb{R})$). Elle vérifie, pour $h,\ y \in \mathbb{R}^2$, $D^2 f(x)(h)(y) = H_f(x)h \cdot y$, où $H_f(x)$ est la matrice hessienne de f en x, donnée par les dérivées partielles secondes : $H_f(x)_{i,j} = \partial^2_{i,j} f(x)$, pour $i, j = 1, \ldots, 3$.
Calculons maintenant les dérivées partielles secondes :

$$\begin{aligned}\partial^2_{1,1} f(x) = 0,\ \partial^2_{1,2} f(x) = c,\\ \partial^2_{2,1} f(x) = c,\ \partial^2_{2,2} f(x) = 0.\end{aligned}$$

2. Calculons les dérivées partielles de f.

$$\begin{aligned}\partial_1 f(x_1, x_2, x_3) &= 2x_1(1 + x_2),\\ \partial_2 f(x_1, x_2, x_3) &= x_1^2 + \sin(x_3),\\ \partial_3 f(x_1, x_2, x_3) &= x_2 \cos(x_3).\end{aligned}$$

On a donc $\nabla f(x) = (2x_1(1 + x_2), x_1^2 + \sin(x_3), -x_2 \cos x_3))^t$. L'application $Df(x)$ est une application linéaire de $\mathbb{R}^3$ dans $\mathbb{R}$, définie par

$$Df(x)(y) = (2x_1(1 + x_2))y_1 + (x_1^2 + \sin(x_3))y_2 - x_2 \cos(x_3)y_3. \tag{2.16}$$

L'application Df appartient à $C^1(\mathbb{R}^3, \mathcal{L}(\mathbb{R}^3, \mathbb{R}))$, et à $x \in \mathbb{R}^3$, elle associe $Df(x) \in \mathcal{L}(\mathbb{R}^3, \mathbb{R})$.
Calculons maintenant les dérivées partielles secondes :

$$\begin{array}{lll} \partial^2_{1,1} f(x) = 2(1+x_2), & \partial^2_{1,2} f(x) = 2x_1, & \partial^2_{1,3} f(x) = 0, \\ \partial^2_{2,1} f(x) = 2x_1, & \partial^2_{2,2} f(x) = 0, & \partial^2_{2,3} f(x) = \cos(x_3)), \\ \partial^2_{3,1} f(x) = 0, & \partial^2_{3,2} f(x) = \cos(x_3), & \partial^2_{3,3} f(x) = -x_2 \sin(x_3). \end{array}$$

La matrice $H_f(x)$ est définie par $H_f(x)_{i,j} = \partial^2_{i,j} f(x)$, pour $i, j = 1, \ldots, 3$. L'application $D^2 f(x)$ est une application linéaire de $\mathbb{R}^3$ dans $\mathcal{L}(\mathbb{R}^3, \mathbb{R})$, définie par $D^2 f(x)(y) = \psi_{x,y}$ et $(D^2 f(x)(y))(z) = \psi_{x,y}(z) = H_f(x)y \cdot z$. Enfin, l'application D^2 est une fonction continue de $\mathbb{R}^3$ dans $\mathcal{L}(\mathbb{R}^3, \mathcal{L}(\mathbb{R}^3, \mathbb{R}))$, définie par $D^2 f(x)(y) = \psi_{x,y}$ pour tout $x,\ y \in \mathbb{R}^3$.

Corrigé de l'exercice 80 page 160 (Point fixe dans $\mathbb{R}$)

1. On vérifie que l'application $f : \ x \mapsto \cos\left(\dfrac{1}{1+x}\right)$ est une application de $[0,1]$ dans lui-même qui est contractante. En effet, $0 < \dfrac{1}{1+x} \leq 1 \leq \frac{\pi}{2}$ pour tout $x \in [0,1]$, donc $f(x) \in [0,1]$ pour tout $x \in [0,1]$. De plus, $f'(x) = \dfrac{1}{(1+x)^2} \sin\left(\dfrac{1}{1+x}\right)$. On voit que $f'(x) \geq 0$ pour tout $x \in [0,1]$ et $f'(x) \leq \sin(1) < 1$. On peut donc appliquer le théorème de point fixe de Banach pour déduire que f admet un unique point fixe dans l'intervalle $[0,1]$ qui est limite de toutes les suites définies par $x^{(0)} \in [0,1]$, $x^{(k+1)} = f(x^{(k)})$.
2. La suite des itérés de point fixe est définie par $x_0 \in [0,1]$ et $x_{n+1} = (x_n)^4$.
   (a) Si $x_0 = 0$, la suite est stationnaire et égale à 0.
   (b) Si $x_0 = 1$, la suite est stationnaire et égale à 1.
   (c) Si $x_0 \in]0,1[$, on montre par une récurrence facile que
      i. $x_{n+1} < x_n$,
      ii. $x_{n+1} \in]0,1[$.

   On en déduit que la suite converge vers une limite ℓ, et en passant à la limite sur $x_{n+1} = (x_n)^4$, on obtient $\ell = 0$ ou 1. Comme $\ell \leq x_0 < 1$, on en dduit que $\ell = 0$.

Corrigé de l'exercice 80 page 160 (Point fixe et Newton)

1. Résolution de l'équation $2xe^x = 1$.
   (a) Comme e^x ne s'annule pas, l'équation $2xe^x = 1$ est équivalente à l'équation $x = \frac{1}{2}e^{-x}$, qui est sous forme point fixe $x = f(x)$ avec $f(x) = \frac{1}{2}e^{-x}$.
   (b) L'algorithme de point fixe s'écrit

$$x^{(0)} \text{ donné} \tag{2.17a}$$
$$x^{(k+1)} = f(x^{(k)}). \tag{2.17b}$$

   Scilab donne :

```
x  =   1.
x  =   0.1839397
x  =   0.4159930
x  =   0.3298425
```

   Notons que la suite n'est pas monotone.
   (c) On a $f'(x) = -\frac{1}{2}e^{-x}$ et donc $|f'(x)| \leq \frac{1}{2}$ pour $x \in [0,1]$. De plus $f(x) \in [0,1]$ si $x \in [0,1]$. L'application $x \mapsto f(x) = \frac{1}{2}e^{-x}$ est donc strictement contractante de $[0,1]$ dans $[0,1]$, et elle admet donc un point fixe, qui est limite de la suite construite par l'algorithme précédent.
2. Résolution de l'équation $x^2 - 2 = 0$.

(a) On se place sur l'intervalle $]0,4[$. L'équation $x^2-2=0$ est manifestement équivalente à l'équation $x=\frac{2}{x}$, qui est sous forme point fixe $x=f(x)$ avec $f(x)=\frac{2}{x}$.

(b) L'algorithme de point fixe s'écrit toujours (2.17), mais si on part de $x_0=1$ ou $x_0=2$, on obtient une suite cyclique (1, 2, 1, 2, 1, 2...) ou (2, 1, 2, 1, 2, 1, 2...) qui ne converge pas.

(c) Scilab donne

```
% x  =    1.
%  x  =    1.5
%  x  =    1.4166667
%  x  =    1.4142157
```

(d) Le développement limité de $g(x)=x^2-2$ entre $x^{(n)}$ et $x^{(n+1)}$ s'écrit :

$$g(x^{(n+1)})=g(x^{(n)})+(x^{(n+1)}-x^{(n)})g'(x^{(n)})+(x^{(n+1)}-x^{(n)})\varepsilon(x^{(n+1)}-x^{(n)}),$$

avec $\varepsilon(x)\to 0$ lorsque $x\to 0$. En écrivant qu'on cherche $x^{(n+1)}$ tel que $g(x^{(n+1)})=0$ et en négligeant le terme de reste du développement limité, on obtient :

$$0=g(x^{(n)})+(x^{(n+1)}-x^{(n)})g'(x^{(n)}),$$

Pour $g(x)=x^2-2$, on a $g'(x)=2x$ et donc l'équation précédente donne bien l'itération de la question précédente.

Corrigé de l'exercice 81 page 160 (Méthode de monotonie) Montrons que la suite $v^{(k)}$ est bien définie. Supposons $v^{(k)}$ connu ; alors $v^{(k+1)}$ est bien défini si le système

$$Av^{(k+1)}=d^{(k)},$$

où $d^{(x)}$ est défini par : $d_i^{(k)}=\alpha_i f(v_i^{(k)})+\lambda b_i$ pour $i=1,\dots,n$, admet une solution. Or, grâce au fait que $Av\geq 0\Rightarrow v\geq 0$, la matrice A est inversible, ce qui prouve l'existence et l'unicité de $v^{(k+1)}$.
Montrons maintenant que les hypothèses du théorème de convergence du point fixe de monotonie sont bien satisfaites.
On pose $R_i^{(\lambda)}(u)=\alpha_i f(u_i)+\lambda b_i$. Le système à résoudre s'écrit donc :

$$Au=R^{(\lambda)}(u)$$

Or 0 est sous–solution car $0\leq\alpha_i f(0)+\lambda b_i$ (grâce au fait que $f(0)=0$, $\lambda>0$ et $b_i\geq 0$).
Cherchons maintenant une sur–solution, c'est–à–dire $\tilde{u}\in\mathbb{R}^n$ tel que

$$\tilde{u}\geq R^{(\lambda)}(\tilde{u}).$$

Par hypothèse, il existe $\mu>0$ et $u^{(\mu)}\geq 0$ tel que

$$(Au^{(\mu)})_i=\alpha f(u_i^{(\mu)})+\mu b_i.$$

Comme $\lambda<\mu$ et $b_i\geq 0$, on a

$$(Au^{(\mu)})_i\geq\alpha_i f(u_i^{(\mu)})+\lambda b_i=R_i^{(\lambda)}(u^{(\mu)}).$$

Donc $u^{(\mu)}$ est sur–solution. Les hypothèses du théorème dont bien vérifiées, et donc $v^{(k)}\to\bar{u}$ lorsque $n\to+\infty$, où $\bar{u}$ est tel que $A\bar{u}=R(\bar{u})$.

Corrigé de l'exercice 82 page 160 (Point fixe amélioré)

1) La suite donnée par l'algorithme (2.14) est bien définie si pour tout $n \in \mathbb{N}$, $g' \circ \varphi(x_n) \neq 0$. Remarquons d'abord que $g' \circ \varphi(\overline{x}) \neq 0$. Or la fonction $g' \circ \varphi$ est continue ; pour $\varepsilon > 0$ fixé, il existe donc $\beta \in \mathbb{R}_+$ tel que $|g' \circ \varphi(x)| \geq \varepsilon$ pour tout $x \in [\overline{x} - \beta, \overline{x} + \beta] = I_\beta$. Remarquons ensuite que $h'(\overline{x}) = 1 - \frac{(g'(\bar{x}))^2}{(g'(\bar{x}))^2} = 0$. Or h' est aussi continue. On en déduit l'existence de $\gamma \in \mathbb{R}_+$ tel que $|h'(x)| < 1$ pour tout $x \in [\overline{x} - \gamma, \overline{x} + \gamma] = I_\gamma$.
Soit maintenant $\alpha = \min(\beta, \gamma)$; si $x_0 \in I_\alpha$, alors $g' \circ \varphi(x_0) \neq 0$. Comme h est strictement contractante sur I_α (et que $h(\overline{x}) = \overline{x}$), on en déduit que $x_1 \in I_\alpha$, et, par récurrence sur n, $x_n \in I_\alpha$ pour tout $n \in \mathbb{N}$ (et la suite est bien définie). De plus, comme h est strictement contractante sur I_α, le théorème du point fixe (théorème 2.5 page 150 donne la convergence de la suite $(x_n)_{n\in\mathbb{N}}$ vers $\bar{x}$.

2) Remarquons d'abord que si $\varphi \in C^2(\mathbb{R}, R)$, on peut directement appliquer la proposition 2.16 (item 2), car dans ce cas $h \in C^2(\mathbb{R}, \mathbb{R})$, puisqu'on a déjà vu que $h'(\overline{x}) = 0$. Effectuons maintenant le calcul dans le cas où l'on n'a que $\varphi \in C^1(\mathbb{R}, R)$. Calculons $|x_{k+1} - \overline{x}|$. Par définition de x_{k+1}, on a :

$$x_{k+1} - \overline{x} = x_k - \overline{x} - \frac{g(x_k)}{g'(\varphi(x_k))},$$

ce qui entraîne que

$$x_{n+1} - \overline{x} = (x_n - \overline{x}) \left(1 - \frac{g(x_n) - g(\overline{x})}{(x_n - \overline{x}) g'(\varphi(x_n))}\right). \tag{2.18}$$

Or il existe $\theta_n \in I(\overline{x}, x_n)$, où $I(\overline{x}, x_n)$ désigne l'intervalle d'extrémités $\overline{x}$ et x_n, tel que

$$\frac{g(x_n) - g(\overline{x})}{x_n - \overline{x}} = g'(\theta_n).$$

Mais comme $g \in C^3(\mathbb{R}, \mathbb{R})$ il existe $\zeta_n \in I(\theta_n, \varphi(x_n))$ tel que :

$$g'(\theta_n) = g'(\varphi(x_n)) + (\theta_n - \varphi(x_n)) g''(\zeta_n).$$

On en déduit que

$$x_{n+1} - \overline{x} = (x_n - \overline{x})(\theta_n - \varphi(x_n)) \frac{g''(\zeta_n)}{g'(\varphi(x_n))}. \tag{2.19}$$

Par inégalité triangulaire, on a :

$$|\theta_n - \varphi(x_n)| \leq |\theta_n - \overline{x}| + |\overline{x} - \varphi(x_n)| = |\theta_n - \overline{x}| + |\varphi(\overline{x}) - \varphi(x_n)|.$$

Comme $\theta_n \in I(\overline{x}, x_n)$, on a donc $|\theta_n - \overline{x}| \leq |x_n - \overline{x}|$; de plus : $|\varphi(\overline{x}) - \varphi(x_n)| \leq \sup_{x\in I_\alpha} |\varphi'(x)||x_n - \overline{x}|$. On en déduit que

$$|\theta_n - \varphi(x_n)| \leq |x_n - \overline{x}| \left(1 + \sup_{x\in I_\alpha} |\varphi'(x))|\right).$$

En reportant dans (2.19), on en déduit que :

$$|x_{n+1} - \overline{x}| \leq \frac{1}{\varepsilon} |x_n - \overline{x}|^2 \left(1 + \sup_{x\in I_\alpha} |\varphi'(x))|\right) \sup_{x\in I_\alpha} |g''(x)|,$$

où ε est donné à la question 1 par choix de α.
On a ainsi montré que la convergence de la suite $(x_n)_{n\in\mathbb{N}}$ définie par l'algorithme (2.14) est au moins quadratique.

3) Reprenons le calcul de la question précédente en montant en ordre sur les développements. Calculons $|x_{n+1} - \overline{x}|$. Ecrivons maintenant qu'il existe $\mu_n \in I(\overline{x}, x_n)$ tel que

$$g(x_n) = g(\overline{x}) + (x_n - \overline{x}) g'(\overline{x}) + \frac{1}{2}(x_n - \overline{x})^2 g''(\mu_n).$$

De (2.18), on en déduit que

$$x_{n+1} - \overline{x} = (x_n - \overline{x}) \left(1 - (x_n - \overline{x}) \frac{g'(\overline{x}) + \frac{1}{2}(x_n - \overline{x}) g''(\mu_n)}{(x_n - \overline{x}) g'(\varphi(x_n))} \right).$$

Or il existe $\nu_n \in I(\overline{x}, \varphi(x_n))$ tel que

$$g'(\varphi(x_n)) = g'(\overline{x}) + (\varphi(x_n) - \varphi(\overline{x})) g''(\nu_n).$$

On a donc :

$$x_{n+1} - \overline{x} = \frac{x_n - \overline{x}}{g'(\varphi(x_n))} \left((\varphi(x_n) - \varphi(\overline{x})) g''(\nu_n) - \frac{1}{2}(x_n - \overline{x}) g''(\mu_n) \right).$$

Ecrivons maintenant que $\varphi(x_n) = \varphi(\overline{x}) + \varphi'(\xi_n)(x_n - \overline{x})$, où $\xi_n \in I(\overline{x}, x_n)$. Comme φ' est lipschitzienne, on a $\varphi'(\xi_n) = \varphi'(\overline{x}) + \epsilon_n = \frac{1}{2} + \epsilon_n$, avec $|\epsilon_n| \leq M|x_n - \overline{x}|$, où M est la constante de Lipschitz de φ'. On a donc :

$$x_{n+1} - \overline{x} = \frac{x_n - \overline{x}}{g'(\varphi(x_n))} \left((x_n - \overline{x})(\frac{1}{2} + \epsilon_n) g''(\nu_n) - \frac{1}{2}(x_n - \overline{x}) g''(\mu_n) \right),$$

et donc (avec ε donné à la question 1 par choix de α) :

$$|x_{n+1} - \overline{x}| \leq \frac{1}{\varepsilon} |x_n - \overline{x}|^2 \left((\frac{1}{2}(g''(\nu_n) - g''(\mu_n)) + \epsilon_n g''(\nu_n) \right).$$

Mais de même, comme $g \in C^3(\mathbb{R}, \mathbb{R})$, et que μ_n et $\nu_n \in I(\overline{x}, x_n)$, on a

$$|g''(\mu_n) - g''(\nu_n)| \leq \sup_{x \in I_\alpha} |g'''(x)| |x_n - \overline{x}|.$$

On en déduit finalement que :

$$|x_{n+1} - \overline{x}| \leq C|x_n - \overline{x}|^3, \text{ avec } C = \frac{1}{2\varepsilon} \sup_{x \in I_\alpha} |g'''(x)| + \frac{M}{\varepsilon} \sup_{x \in I_\alpha} |g''(x)|.$$

4) Pour montrer que la suite définie par l'algorithme (2.14) converge de manière cubique, il suffit de montrer que φ vérifie les hypothèses de la question 3). On a évidemment $\varphi(\bar{x}) = \bar{x}$. Comme $g \in C^3(\mathbb{R}, \mathbb{R})$ et que $g'(x) \neq 0, \quad \forall x \in I_\beta$, on en déduit que $\varphi \in C^2(\mathbb{R}, \mathbb{R})$. De plus

$$\varphi'(\overline{x}) = 1 - \frac{1}{2} \frac{g'(\overline{x})^2 - g''(\overline{x}) g(\overline{x})}{g'(\overline{x})^2} = \frac{1}{2}.$$

La fonction φ vérifie donc bien les hypothèses de la question 3.

Université d'Aix-Marseille

3ème année, lic. Math. Géné.
Analyse numérique, semestre 5

TP scilab n°5

Méthode de la puissance

On considère nos deux problèmes favoris :

$$a = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \qquad b = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}$$

$$A_n = (n+1)^2 \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & & \vdots \\ 0 & 0 & -1 & 2 & \ddots & 0 \\ \vdots & & & \ddots & \ddots & -1 \\ 0 & \cdots & & 0 & -1 & 2 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R}) \quad B_n = \begin{pmatrix} \sin\left(2\pi\frac{1}{n+1}\right) \\ \vdots \\ \sin\left(2\pi\frac{i}{n+1}\right) \\ \vdots \\ \sin\left(2\pi\frac{n}{n+1}\right) \end{pmatrix}$$

On se propose de montrer pourquoi la méthode de Jacobi ne fonctionne pas dans un des problèmes indiqués plus haut. La méthode de la puissance permet d'approcher la plus grande valeur propre (en valeur absolue) d'une matrice A. L'idée consiste à construire une suite (X^k) de vecteurs telle que :

$$(1) \qquad \begin{cases} X^0 \in \mathbb{R}^n \setminus \{0\} \\ X^{k+1} = \frac{1}{\|AX^k\|_2} AX^k \end{cases}$$

Alors (AX^k, X^k) tend vers Λ (quand $k \to \infty$), où Λ est la plus grande valeur propre de A (en valeur absolue).

1. Ecrire une fonction scilab `meth_puiss` qui, a partir d'une matrice `A` et du nombre d'itérations `nb_iter`, renvoie une approximation de `Lambda`, la plus grande valeur propre (en valeur absolue) de `A`. On propose la syntaxe suivante pour `meth_puiss` :

```
function [Lambda]=meth_puiss(A,nb_iter),
...
endfunction;
```

2. A l'aide de `meth_puiss`, calculez le rayon spectral de $M^{-1}N$ pour les méthodes de Jacobi et de Gauss Seidel appliquées aux deux matrices a et A_n.
   Rappels : Pour la méthode de Jacobi : $M = D$ et $N = E + F$.
   Pour la méthode de Jacobi : $M = D - E$ et $N = F$.

3. Que se passe-til si l'on remplace la norme 2 par la norme 1 dans l'algorithme (1)

Exercice 2 : Méthode QR

La méthode QR est une méthode qui permet de retrouver, de manière approchée, les valeurs propres d'une matrice. Elle utilise la décomposition QR d'une matrice ($A = QR$ où Q est une matrice orthogonale et R une matrice triangulaire supérieure).

L'algorithme de la méthode QR est le suivant (en pratique, tol est une petite valeur) :

$B \leftarrow A$
Tant qu'il existe des coefficients sous la diagonale de B plus grand que tol (en valeur absolue),
    $[Q, R] \leftarrow qr(B)$
    $B \leftarrow RQ$
Fin de la boucle
Retourner B

On va tester la méthode QR sur différentes matrices diagonalisables, et étudier (numériquement) la convergence de la méthode :

1. Construire une matrice symétrique 4×4 en prenant une matrice M tirée au hasard (fonction random) et en écrivant $P = MM^t$. Vérifier que P est inversible.
2. On pose alors :
$$A_i = PD_iP^{-1}$$

$$\text{et } D_1 = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad D_2 = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix} \qquad D_3 = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

   (a) Ecrire une fonction Scilab `methQR` qui implémente l'algorithme de la méthode QR en prenant en entrée une matrice `A`. On pourra faire appel à la fonction Scilab `qr` qui réalise la décomposition QR d'une matrice.
   Testez ensuite la méthode sur les matrices A_1 et A_2 et A_3.

   (b) Modifiez la fonction de manière à ne pas faire plus de 1000 itérations.

   (c) Testez la fonction ainsi modifiée sur la matrice A_4 avec :
$$D_4 = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$
   Que remarquez-vous ? Calculez les valeurs propres de la sous-matrice composée des deux dernières colonnes et des deux dernières lignes du résultat de `methQR(A4)`. Que constatez-vous maintenant ?

Université d'Aix-Marseille

3ème année, lic. Math. Géné.
Analyse numérique, semestre 5

TP scilab n°5

Exercice 1 (Point fixe et Newton)

On cherche à calculer $\bar{x} = \sqrt{3}$,

1. par le point fixe sur la fonction $f(x) = x - x^2 + 3$,
2. par le point fixe avec relaxation sur la même fonction,
3. par la méthode de Newton.

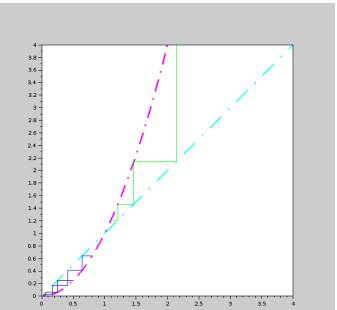
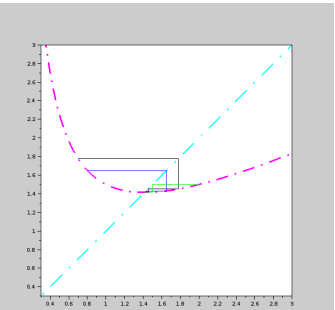
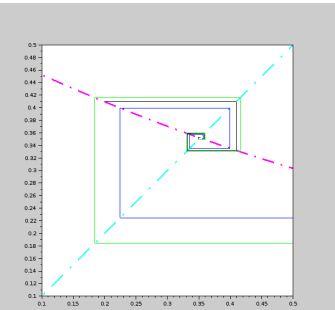
Programmer les trois méthodes et comparer les vitesses de convergence en partant de $x_0 = 1$.
Tester ensuite plusieurs initialisations possibles, et pour la méthode de relaxation, plusieurs paramètres.
Evaluer la vitesse de convergence des algorithmes en calculant, pour p bien choisi

$$\frac{\|x^{(k+1)} - \bar{x}\|}{\|x^{(k)} - \bar{x}\|^p}$$

Exercice 2 (Suites récurrentes - Visualisation)

On considère la suite réccurente $u_{n+1} = f(u_n)$ Reproduire les graphes ci-dessous qui correspondent aux fonctions

$$f_1(x) = x^2, \quad f_2(x) = \frac{2 + x^2}{2x}, \quad f_3(x) = \frac{e^{-x}}{2}.$$

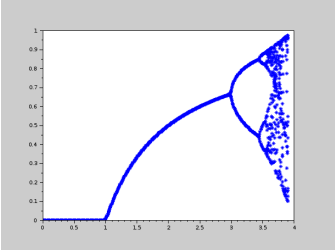
Exercice 3 (Suite de Feigenbaum)

On considère la suite :

$$u_{n+1}^{(\lambda)} = f_\lambda(u_n^{(\lambda)}) \text{ avec } \begin{cases} f_\lambda(x) = \lambda x(1-x), \\ \lambda \in [0, 4[, \\ u_0^{(\lambda)} \in [0, 1]. \end{cases} \tag{1}$$

1. Créer la fonction : $(x, \lambda) \mapsto f_\lambda(x)$.

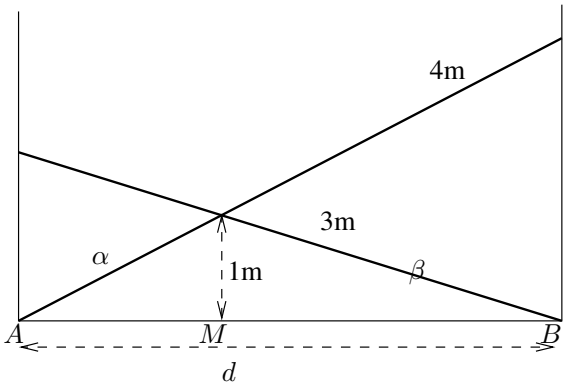
2. Créer une fonction qui renvoie pour λ, $u_0^{(\lambda)}$ et n, p fixés, $u_{n+i}^{(\lambda)}$ pour $i = 1, \cdots, p$.
3. Pour $\lambda \in [0, 4[$, $u_0^{(\lambda)} = 0.5$, visualiser les points $u_n^{(\lambda)}$ pour $n = 101, \ldots, 110$. On doit obtenir ce que l'on appelle la cascade de Feigenbaum :



4. Interpréter ce graphique.

Exercice 4 (Newton et les échelles. . .)

Soient deux échelles de longueurs respectives 3 et 4 m, posées contre deux murs verticaux selon la figure ci-contre. On sait que les échelles se croisent à 1m du sol, et on cherche à connaître la distance d entre les deux murs.



1. Montrer que le problème revient à déterminer x et y tels que

$$16x^2 = (x^2 + 1)(x + y)^2 \tag{2}$$

$$9y^2 = (y^2 + 1)(x + y)^2. \tag{3}$$

2. Ecrire une fonction sous scilab qui à $(x^{(0)}, y^{(0)})$ associe $(x^{(k)}, y^{(k)})$ le $k^{ième}$ itéré de l'algorithme de Newton pour la résolution du système (2)-(3).
3. Visualiser les premiers itérés $x^{(k)}$ et $y^{(k)}$ construits par la méthode de Newton en partant de $x^{(0)} = 1$ et $y^{(0)} = 1$.

Université d'Aix-Marseille **Analyse numérique, semestre 5**
Licence de maths, 3ème année

Projet PageRank

L'objectif de ce projet est d'étudier sur l'exemple du classement des pages web, la méthode de la puissance. Ce sujet est fortement inspiré d'un sujet d'agreg option B, 2008.
Le projet contient une partie à faire avec scilab, partie facultative, et une partie à rendre pour le deuxième devoir, qui comptera pour la note de contrôle continu.

La recherche d'informations pertinentes sur le Web est un des problèmes les plus cruciaux pour l'utilisation de de ce dernier. Des enjeux économiques colossaux sont en jeu, et diverses multinationales se livrent à de grandes manoeuvres. Le leader actuel de ce marché, Google, utilise pour déterminer la pertinence des références fournies, un certain nombre d'algorithmes dont certains sont des secrets industriels jalousement gardés, mais d'autres sont publics. On va s'intéresser ici à l'algorithme PageRank [1], lequel fait intervenir des valeurs propres et vecteurs propres d'une énorme matrice.

1 Le principe de l'algorithme PageRank

On peut considérer pour simplifier que le Web est une collection de $N \in \mathbb{N}$ pages, avec N très grand (de l'ordre de 10^{10} en octobre 2005). La plupart de ces pages incluent des liens hypertextes vers d'autres pages. On dit qu'elles pointent vers ces autres pages. L'idée de base utilisée par les moteurs de recherche pour classer les pages par ordre de pertinence décroissante consiste à considérer que plus une page est la cible de liens venant d'autres pages, c'est-à-dire plus il y a de pages qui pointent vers elle, plus elle a de chances d'être fiable et intéressante pour l'utilisateur final, et réciproquement. Il s'agit donc de quantifier cette idée, c'est-à-dire d'attribuer un score de pertinence à chaque page.

Pour ce faire, on représente le Web comme un graphe orienté : un *graphe* est un ensemble de points, dont certaines paires sont directement reliées par un "lien". Ces liens peuvent être orientés, c'est-à-dire qu'un lien entre deux points u et v relie soit u vers v, soit v vers u : dans ce cas, le graphe est dit orienté. Sinon, les liens sont symétriques, et le graphe est non-orienté. Les points sont généralement appelés sommets, et les liens "arêtes". On peut représenter le graphe par une matrice, qu'on appelle matrice d'adjacence : le coefficient de la i-ème ligne et j-ème colonne est 1 s'il existe une arête allant du sommet i au sommet j, et 0 sinon. Remarquer que si le graphe est non orienté, alors une arête est définie comme liant les sommets i et j sans ordre, et la matrice d'adjacence est symétrique.

1. Le nom "PageRank" vient du nom de Larry Page, l'un des inventeurs de Google.

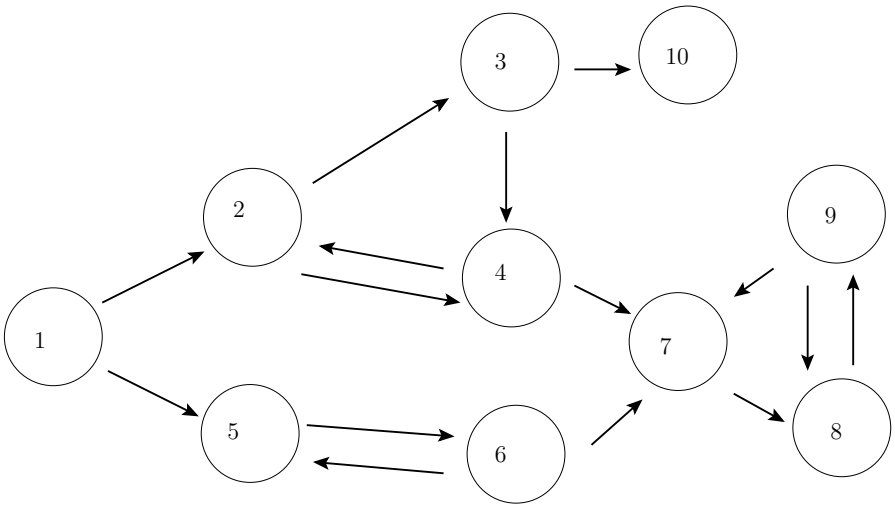


FIGURE 1 – Exemple 1

Pour représenter le Web comme un graphe, on se donne donc un ordre arbitraire sur l'ensemble des pages que l'on numérote ainsi de $i = 1$ à $i = N$: ce sont les sommets du graphe. La structure de connectivité du Web peut alors être représentée par la matrice d'adjacence C de taille $N \times N$ telle que $C_{ij} = 1$ si la page j pointe sur la page i, $C_{ij} = 0$ sinon. Remarquons que c'est un graphe non symétrique. Les liens d'une page sur elle-même ne sont pas significatifs, on pose donc $C_{ii} = 0$. Remarquons que la ligne i de la matrice C contient tous les liens significatifs qui pointent sur la page i, alors que la colonne j contient tous les liens significatifs qui partent de la page j.
Pour illustrer notre étude, on propose les trois exemples de graphe décrits sur les figures 1, 2 et 3. .

1. *Ecrire les matrices C_1, C_2 et C_3 associées respectivement aux exemples 1, 2 et 3.*

On souhaite attribuer à chaque page i un score $r_i \in \mathbb{R}_+^*$ de façon à pouvoir classer l'ensemble des pages par score décroissant et présenter à l'utilisateur une liste ainsi classée des pages correspondant à sa requête. L'algorithme PageRank part du principe qu'un lien de la page j pointant sur la page i contribue positivement au score de cette dernière, avec une pondération par le score r_j de la page dont est issu le lien (une page ayant un score élevé a ainsi plus de poids qu'une n'ayant qu'un score médiocre) et par le nombre total de liens présents sur ladite page $N_j = \sum_{k=1}^{N} C_{kj}$. On introduit donc la matrice Q définie par $Q_{ij} = \frac{C_{ij}}{N_j}$ si $N_j \neq 0$, $Q_{ij} = 0$ sinon.

2. *Déterminer les matrices Q_1, Q_2 et Q_3 associées aux trois exemples. Vérifier que la somme des coefficients des colonnes non nulles de Q vaut toujours* 1.
3. *Démontrer que toutes les colonnes non nulles de la matrice Q générale sont telles que la somme de leurs coefficients est égale à 1.*

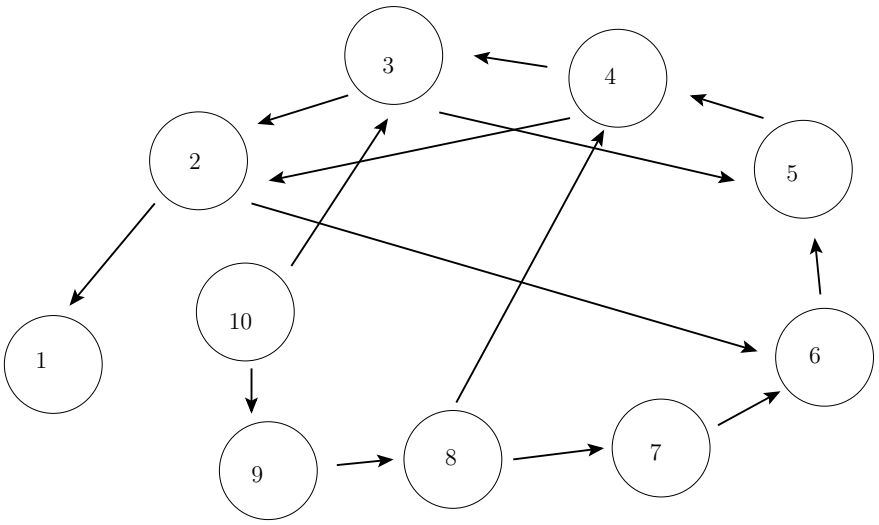


FIGURE 2 – Exemple 2

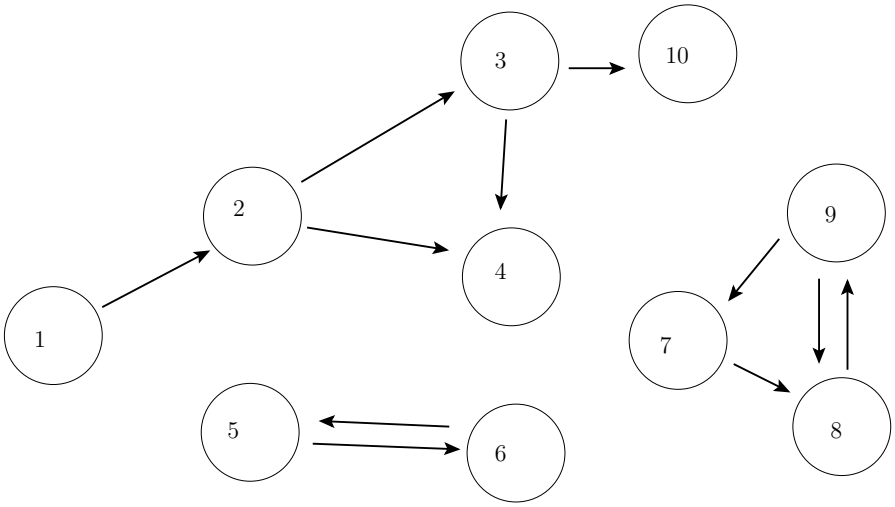


FIGURE 3 – Exemple 3

L'application des principes ci-dessus conduit donc à une équation pour le vecteur $r \in \mathbb{R}^N$ des scores des pages de la forme

$$r_i = \sum_{j=1}^{N} Q_{ij} r_j \text{ c'est à dire } r = Qr, \tag{1}$$

où Q est une matrice dont la somme des coefficients de chaque colonne non nulle est égale à 1. Le problème du classement des pages du Web se trouve ainsi ramené à la recherche d'un vecteur propre d'une énorme matrice, associé à la valeur propre 1 ! Mais il peut arriver que la matrice Q n'admette pas la valeur propre 1 ce qui invalide quelque peu la philosophie originale de l'algorithme.

4. *Vérifier en utilisant scilab que les matrices Q_1 et Q_3 admettent 1 comme valeur propre mais que ce n'est pas le cas de Q_2.*
5. *Soit Q la transposée d'une matrice stochastique*[2]
   (a) *Montrer que 1 est valeur propre de Q^t et donc de Q.*
   (b) *Montrer que $\rho(Q^t) = \rho(Q) = 1$.*

Pour faire en sorte que 1 soit valeur propre, on va donc modifier la matrice Q de manière à ce qu'elle soit la transposée d'une matrice stochastique, et donc en particulier qu'elle n'ait plus de colonne nulle. On considère pour cela les vecteurs $e = (1\,1 \cdots 1)^t \in \mathbb{R}^N$ et $d \in \mathbb{R}^N$ de composantes $d_j, j = 1, \ldots, N$, avec $d_j = 1$ si $N_j = 0$, $d_j = 0$ sinon. On définit alors la matrice

$$P = Q + \frac{1}{N} e d^t. \tag{2}$$

6. *Visualiser sur scilab les matrices P_1, P_2 et P_3 associées aux trois exemples et calculer leurs valeurs propres à l'aide de la commande spec.*
7. *Montrer que le passage de Q à P revient à remplacer les colonnes de zéros de Q par des colonnes dont toutes les composantes sont égales à $\frac{1}{N}$.*
8. *Montrer que P est bien la transposée d'une matrice stochastique, et en déduire que P admet bien la valeur propre 1 et que $\rho(P^t)$ et $\rho(P)$ sont égaux à 1.*

On remarque sur l'exemple 3 que la valeur propre 1 peut être multiple. Or, notre problème consiste à trouver un vecteur propre associé à cette valeur propre et il est alors préférable que 1 soit valeur propre simple. On va donc encore modifier la matrice de manière à ce que tous ses coefficients soient strictement positifs. En effet, on a le théorème suivant :

2. On appelle *matrice stochastique* une matrice dont tous les coefficients appartiennent à $[0, 1]$ et dont la somme des coefficients de chaque ligne vaut 1.

Théorème 1 (Perron-Frobenius, cas des matrices stochastiques) *Soit A une matrice transposée d'une matrice stochastique et qui est de plus strictement positive (c.à.d. dont tous les coefficients sont strictement positifs). Alors $\rho(A) = 1$, et 1 est une valeur propre simple ; de plus, il existe un vecteur propre r strictement positif associé à la valeur propre 1 .*

Ce théorème est important pour nous car il va nous permettre non seulement de construire une matrice avec 1 comme valeur propre simple, mais de plus d'obtenir le vecteur r donnant un classement des pages du web. La question 9 a pour objet de démontrer le théorème de Perron-Frobenius. En fait, sous les hypothèses de ce théorème, on sait déjà par les questions précédentes que $\rho(A) = 1$ et que 1 est valeur propre. Il reste à montrer que 1 est valeur propre simple et qu'il existe un vecteur propre r strictement positif associé à la valeur propre 1.

9. *Soit $A = B^t$ où B est une matrice stochastique strictement positive.*
   (a) *Montrer que le sous-espace propre associé à la valeur propre 1 de B est $\mathbb{R}e$.*
   (b) *En déduire que la valeur propre 1 de A est simple.*
   (c) *Soit f un vecteur propre de A pour la valeur propre 1 .*
      i. *Montrer que $|f_i| < \sum_j a_{ij}|f_j|$ sauf si les f_j sont tous de même signe.*
      ii. *En raisonnant sur $\sum_i |f_i|$, en déduire que les f_j sont tous de même signe.*

Dans le but d'obtenir une matrice strictement positive, on effectue alors une dernière modification sur la matrice en choisissant un nombre $0 < \alpha < 1$ et en posant

$$A_\alpha = \alpha P + (1-\alpha)\frac{1}{N}ee^t. \tag{3}$$

10. *Ecrire sous scilab les matrices $A_{\alpha,1}$, $A_{\alpha,2}$ et $A_{\alpha,3}$ associées aux trois exemples pour $\alpha = 0.1$ et $\alpha = 0.5$. Calculer les modules valeurs propres de ces matrices et classez les par ordre décroissant (on pourra utiliser la commande gsort de scilab).*
11. *Montrer que A_α est toujours la transposée d'une matrice stochastique, et qu'elle est de plus strictement positive. En déduire que $\rho(A_\alpha) = 1$, que 1 est une valeur propre simple de A_α et qu' il existe un vecteur propre r_α strictement positif associé à la valeur propre 1.*

Finalement, PageRank calcule un tel vecteur propre $r_\alpha \in \mathbb{R}^N$, normalisé d'une façon ou d'une autre, qui est tel que

$$r_\alpha = A_\alpha r_\alpha, \tag{4}$$

dont les N composantes fournissent le classement recherché des pages du Web. On remarquera que pour N grand, la matrice A_α n'est qu'une petite perturbation de la matrice Q. On sait combien cette stratégie s'est révélée efficace, puisque Google a totalement laminé les moteurs de recherche de première génération, comme Altavista, lesquels ont essentiellement disparu du paysage.

2 Calcul effectif du score des pages web

On décrit dans cette section des méthodes pour approcher ce vecteur r.

1. *Programmer la méthode de la puissance : pour* $r_0 \neq 0$

$$q_k = A_\alpha r_{k-1},\ r_k = \frac{q_k}{\|q_k\|_1}. \tag{5}$$

On normalisera les vecteurs en utilisant la norme 1.

Approcher le vecteur r *solution de* $r = A_\alpha r$ *pour nos trois exemples pour* $\alpha = 0.1$. *On pourra discuter le choix du test d'arrêt utilisé pour stopper l'algorithme.*

Vérifier numériquement que $\|r_k - r\|_2 \leq Cte|\alpha\mu_2|^k$ *où* μ_2 *est la seconde plus grande valeur propre (en module) de* P.

On remarque que les matrices A_α sont des matrices pleines alors que la matrice initiale Q était creuse. Il est en pratique hors de question d'assembler cette matrice A_α.

2. *Montrer que si* $z \in \mathbb{R}^N$, $z \geq 0$ *avec* $\|z\|_1 = 1$, *alors* $y = A_\alpha z = \alpha Q z + \frac{1-\|\alpha Q z\|_1}{N} e$ *et* $y \geq 0$.

On a ainsi ramené le calcul du produit matrice pleine-vecteur original à un produit matrice creuse-vecteur et à une évaluation de norme, effectivement calculables à l'échelle du Web (à condition de disposer de ressources informatiques conséquentes, quand même).

3. *Programmer alors l'algorithme*

```
1  Choisir r0
2  Tant que s >tol, faire
3    r1=alpha Q r0
4    beta=1-norm(r1,1)
5    r2=r1+beta/N*e
6    s=norm(r2-r0,1)
7    r0=r2
8   retourner r0
```

Construire un exemple de réseau de pages webs de "grande taille" $N = 1000$ *(la taille réelle du WWW est de plusieurs milliards !...) et comparer la vitesse de cet algorithme avec l'algorithme de la puissance.*