3.6

Travail demandé :

Il vous est demandé d'étudier puis de présenter le texte joint à travers un exposé de synthèse d'une durée comprise entre 15 et 20 minutes.

Si l'étude de la totalité du dossier et la préparation d'un exposé cohérent dans la durée impartie ne vous paraît pas possible, vous pouvez décider de vous limiter à une partie du dossier.

Remarques générales :

1. Les textes proposés, quelle que soit leur origine, peuvent présenter des défauts (coquilles typographiques, négligences ou sous-entendus de l'auteur, voire erreurs...) qui, sauf exception, n'ont pas été corrigés.

2. Les textes proposés peuvent contenir des exercices qu'il n'est pas demandé de résoudre. Néanmoins, vous pouvez vous aider des énoncés de ces exercices pour enrichir votre exposé.

3. Vous pouvez annoter les documents qui vous sont fournis. Vos annotations ne seront pas regardées par l'examinateur.

6 MÉTHODES DE CALCUL DES VALEURS PROPRES ET DES VECTEURS PROPRES

Introduction

Pour calculer des approximations de l'ensemble des valeurs propres d'une matrice A, une idée couramment exploitée consiste à construire une suite de matrices $(P_k)_{k \geq 1}$ telle que les matrices $P_k^{-1}AP_k$ "convergent" (dans un sens à préciser ; en effet, il ne s'agit pas toujours d'une véritable convergence) vers une matrice de valeurs propres connues, c'est-à-dire diagonale ou triangulaire.

Cette idée est à la base de la *méthode de Jacobi* pour les matrices *symétriques* (*cf.* paragraphe 6.1), où les matrices P_k sont des produits de matrices orthogonales "élémentaires" très simples à construire. On peut alors montrer que (théorème 6.1-2)

$$\lim_{k \to \infty} P_k^{-1}AP_k = \text{diag } (\lambda_i),$$

où les nombres λ_i sont les valeurs propres de la matrice A (à une permutation près). Lorsque ces dernières sont toutes distinctes, on établit que les vecteurs colonnes des matrices P_k constituent de surcroît des approximations des vecteurs propres de la matrice A (théorème 6.1-3).

Pour des matrices *quelconques*, la remarquable *méthode* QR, dont le principe est décrit au paragraphe 6.3, relève de la même idée. Utilisant à chaque itération de la méthode la factorisation QR des matrices (*cf.* paragraphe 4.5), on construit une suite de matrices (P_k) telle que, moyennant certaines hypothèses assez restrictives (*cf.* théorème 6.3-1),

$$\lim_{k \to \infty} (P_k^{-1}AP_k)_{ij} = 0 \quad \text{pour} \quad j < i,$$

$$\lim_{k \to \infty} (P_k^{-1}AP_k)_{ii} = \lambda_i,$$

les scalaires λ_i étant les valeurs propres de la matrice A, sans qu'on puisse rien dire de la convergence des éléments $(P_k^{-1}AP_k)$, $i < j$; il s'agit donc seulement d'une "pseudo-convergence" de la suite $(P_k^{-1}AP_k)$. On trouvera également au paragraphe 6.3 quelques compléments sur la convergence de la méthode (sans démonstration) et sur sa mise en œuvre pratique (mise préliminaire sous forme de Hessenberg, méthode QR avec translations, etc.).

On notera l'absence de méthodes de calculs de l'ensemble des valeurs propres des *matrices quelconques* à partir du *polynôme caractéristique*. C'est d'ailleurs *exactement l'inverse* qui se produit : pour calculer l'ensemble des racines d'un polynôme de degré élevé, il est en effet courant d'appliquer la méthode QR à sa compagne (qui est précisément déjà une matrice de Hessenberg : *cf.* paragraphe 2.1) !

D'autres méthodes permettent de calculer seulement certaines valeurs propres sélectionnées. C'est le cas notamment de la *méthode de Givens-Householder* que nous étudions au paragraphe 6.2, et qui s'applique aux matrices *symétriques* : on commence par réduire une telle matrice à la forme *tridiagonale* à l'aide de *matrices de Householder* (introduites au paragraphe 4.5), puis un ingénieux procédé de *Givens* permet le calcul approché, avec une précision arbitraire, d'une valeur propre *de rang donné* d'une telle matrice.

On décrit enfin au paragraphe 6.4 un procédé très répandu de calcul d'un vecteur propre correspondant à une valeur propre particulière dont on connaît déjà une approximation, appelé *méthode de la puissance inverse*, et on donne (Théorème 6.4-1) des conditions suffisantes pour sa convergence.

6.1. La méthode de Jacobi

Cette méthode s'emploie lorsque l'on cherche *toutes les valeurs propres*, et (éventuellement) *tous les vecteurs propres*, d'une matrice *symétrique*. Elle s'applique bien aux matrices *pleines*.

Une matrice symétrique A est diagonalisable : il existe une matrice orthogonale O telle que (théorème 1.2-1)

$$O^T A O = \text{diag}\ (\lambda_1, \lambda_2, \ldots, \lambda_n),$$

où les nombres λ_i sont les valeurs propres, multiplicités comprises, de la matrice A. On rappelle que les vecteurs colonnes de la matrice O forment un ensemble orthonormal de vecteurs propres, le i-ème vecteur colonne étant un vecteur propre associé à la valeur propre λ_i.

Partant de la matrice $A = A_1$, la méthode de Jacobi consiste à construire une suite $(\Omega_k)_{k \geq 1}$ de matrices orthogonales "élémentaires" (dont la forme, très simple, est donnée au prochain théorème), en s'arrangeant pour que la suite de matrices (encore symétriques)

$$A_{k+1} = \Omega_k^T A_k \Omega_k = (\Omega_1 \Omega_2 \ldots \Omega_k)^T A (\Omega_1 \Omega_2 \ldots \Omega_k), \quad k \geq 1,$$

converge vers la matrice diag (λ_i), à une permutation des indices près (théorème 6.1-2). On peut alors s'attendre (au moins dans certains cas ; *cf* théorème 6.1-3) à la convergence de la suite des matrices orthogonales

$$O_k = \Omega_1 \Omega_2 \ldots \Omega_k$$

vers une matrice orthogonale dont les colonnes forment un ensemble orthonormal de vecteurs propres de la matrice A.

Le principe de chaque transformation

$$A_k \to A_{k+1} = \Omega_k^T A_k \Omega_k, \quad k \geq 1,$$

est d'annuler deux éléments hors-diagonaux en position symétrique, soit $(A_k)_{pq}$ et $(A_k)_{qp}$, de la matrice A_k, suivant un procédé très simple, que nous allons décrire et étudier. Pour alléger l'écriture, nous posons "provisoirement"

$$A_k = A = (a_{ij}), \quad A_{k+1} = B = (b_{ij}), \quad \Omega_k = \Omega.$$

Le choix *effectif* du couple (p, q) sera décrit plus loin.

Théorème 6.1-1. *Soit p et q deux entiers vérifiant $1 \leqslant p < q \leqslant n$, et θ un nombre réel, auxquels on associe la matrice orthogonale*

$$\Omega = \begin{pmatrix} & & p & & q & & \\ & & \downarrow & & \downarrow & & \\ 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & \cos\theta & & \sin\theta & & \leftarrow p \\ & & & & 1 & & & \\ & & & & & 1 & & \\ & & & & & & 1 & \\ & & & (-\sin\theta) & & \cos\theta & & \leftarrow q \\ & & & & & & & 1 \\ & & & & & & 1 & \end{pmatrix}$$

(1) Si $A = (a_{ij})$ *est une matrice symétrique, la matrice*

$$B = \Omega^T A \Omega = (b_{ij}),$$

également symétrique, vérifie

$$\sum_{i,j=1}^{n} b_{ij}^2 = \sum_{i,j=1}^{n} a_{ij}^2.$$

(2) Si $a_{pq} \neq 0$, *il existe une et une seule valeur du nombre* θ *dans l'ensemble* $]-\frac{\pi}{4}, 0[\cup]0, \frac{\pi}{4}]$ *telle que*

$$b_{pq} = 0;$$

c'est la seule solution, dans le même ensemble, de l'équation

$$\operatorname{cotg} 2\theta = \frac{a_{qq} - a_{pp}}{2a_{pq}}.$$

Le nombre θ *étant ainsi choisi,*

$$\sum_{i=1}^{n} b_{ii}^2 = \sum_{i=1}^{n} a_{ii}^2 + 2a_{pq}^2.$$

DÉMONSTRATION. (1) Il est facile de vérifier que la matrice Ω est orthogonale. La norme $\|\cdot\|_E$ étant invariante par transformation orthogonale (théorème 1.4-4),

$$\sum_{i,j} a_{ij}^2 = \|A\|_E^2 = \|O^T A O\|_E^2 = \sum_{i,j} b_{ij}^2.$$

(2) On remarque ensuite que la transformation portant sur les éléments d'indices $(p, p), (p, q), (q, p), (q, q)$ s'écrit sous la forme

$$\begin{pmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

et le même raisonnement qu'en (1) montre que

$$a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2 = b_{pp}^2 + b_{qq}^2 + 2b_{pq}^2$$

pour toute valeur du nombre θ. Comme

$$b_{pq} = b_{qp} = a_{pq} \cos 2\theta + \frac{a_{pp} - a_{qq}}{2} \sin 2\theta,$$

le choix du nombre θ indiqué dans l'énoncé entraîne

$$b_{pq} = b_{qp} = 0, \quad \text{et donc} \quad a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2 = b_{pp}^2 + b_{qq}^2.$$

Comme par ailleurs $a_{ii} = b_{ii}$ pour $i \neq p$ et $i \neq q$, la démonstration est complète. ■

REMARQUES. (1) La matrice Ω représente une rotation d'angle θ dans le plan des p-ème et q-ème vecteurs de base, ce qui est une autre façon de voir qu'elle est orthogonale.

(2) On note également que seules les p-ème et q-ème lignes et colonnes de la matrice A sont modifiées dans la transformation $A \to B = \Omega^T A \Omega$. De façon plus précise, pour toute valeur de l'angle θ,

$$\begin{cases} b_{ij} = a_{ij} & \text{si} \quad i \neq p, q \quad \text{et} \quad j \neq p, q, \\ b_{pi} = a_{pi} \cos\theta - a_{qi} \sin\theta & \text{si} \quad i \neq p, q, \\ b_{qi} = a_{pi} \sin\theta + a_{qi} \cos\theta & \text{si} \quad i \neq p, q, \\ b_{pp} = a_{pp} \cos^2\theta + a_{qq} \sin^2\theta - a_{pq} \sin 2\theta, \\ b_{qq} = a_{pp} \sin^2\theta + a_{qq} \cos^2\theta + a_{pq} \sin 2\theta, \\ b_{pq} = b_{qp} = a_{pq} \cos 2\theta + \dfrac{a_{pp} - a_{qq}}{2} \sin 2\theta. \end{cases}$$

(3) Grâce aux relations existant entre les fonctions trigonométriques, les éléments de la matrice B sont, malgré les apparences, déterminés par des *relations algébriques* à partir des éléments de la matrice A ; on calcule successivement les quantités suivantes :

$$\varkappa = \frac{a_{qq} - a_{pp}}{2a_{pq}} \quad (= \operatorname{cotg} 2\theta),$$

$$\begin{cases} \text{ou bien} \quad t = \text{la racine de plus petit module} \quad \left(\text{car } t = \operatorname{tg}\theta \quad \text{et} \quad |\theta| \leqslant \frac{\pi}{4}\right) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{du trinôme} \quad t^2 + 2\varkappa t - 1 = 0, \quad \text{si} \quad \varkappa \neq 0, \\ \text{ou bien} \quad t = 1 \quad \text{si} \quad \varkappa = 0 \end{cases}$$

$$c = \frac{1}{\sqrt{1+t^2}} \quad (= \cos\theta),$$

$$s = \frac{t}{\sqrt{1+t^2}} \quad (= \sin\theta).$$

Les formules donnant les éléments de la matrice B peuvent alors être réécrites sous la forme suivante (le vérifier)

$$\begin{aligned} b_{pi} &= ca_{pi} - sa_{qi}, \quad i \neq p, q, \\ b_{qi} &= ca_{qi} + sa_{pi}, \quad i \neq p, q, \\ b_{pp} &= a_{pp} - ta_{pq}, \\ b_{qq} &= a_{qq} + ta_{pq}, \end{aligned}$$

■

Décrivons maintenant une étape de la méthode de Jacobi. La matrice $A_k = (a_{ij}^k)$ étant connue, on choisit un couple (p, q), $p \neq q$, pour lequel l'élément a_{pq}^k n'est pas nul ; puis on construit la matrice Ω_k comme la matrice Ω du théorème 6.1-1, l'angle θ_k étant choisi dans l'ensemble $]-\frac{\pi}{4}, 0[\cup]0, \frac{\pi}{4}]$ de telle façon que

$$\operatorname{cotg} 2\theta_k = \frac{a_{qq}^k - a_{pp}^k}{2a_{pq}^k},$$

et l'on pose

$$A_{k+1} = \Omega_k^T A_k \Omega_k = (a_{ij}^{k+1}).$$

On distingue trois stratégies pour le choix du couple (p, q).

(i) *Méthode de Jacobi classique* : on choisit l'un des couples pour lesquels

$$|a_{pq}^k| = \max_{i \neq j} |a_{ij}^k|.$$

Bien entendu, les couples (p, q) ainsi déterminés *varient avec l'entier k* : c'est uniquement pour alléger l'écriture que nous ne rappelons pas cette dépendance.

(ii) *Méthode de Jacobi cyclique* : la recherche du plus grand module des éléments hors-diagonaux prenant un temps assez long dans la méthode de Jacobi classique, on annule successivement *tous* les éléments hors-diagonaux par un *balayage cyclique*, toujours le même ; par exemple. on choisit les couples $(p. q)$ dans l'ordre suivant :

$$(1, 2), (1, 3), \ldots, (1, n); \quad (2, 3), \ldots, (2, n); \quad \ldots; (n-1, n).$$

Naturellement, si l'un des éléments "balayés" est déjà nul, on passe au suivant (du point de vue matriciel, ceci équivaut au choix $\theta_k = 0$, soit $\Omega_k = I$).

(iii) *Méthode de Jacobi avec seuil* : on procède comme dans le cas de la méthode de Jacobi cyclique mais on omet d'annuler les éléments hors-diagonaux dont le module est inférieur à un certain seuil, qui diminue avec chaque balayage ; il semble en effet inutile d'annuler des éléments hors-diagonaux "déjà" très petits en valeur absolue, alors que d'autres sont encore d'un ordre de grandeur beaucoup plus élevé.

Attention ! Quelle que soit la stratégie choisie, il est bien évident que les éléments annulés à une étape donnée peuvent être ensuite remplacés par des éléments non nuls. Sinon, on obtiendrait la réduction à une matrice diagonale en un nombre fini d'itérations, ce qui est impossible, comme nous l'avons déjà indiqué au paragraphe 2.1.

Nous allons étudier la convergence de la méthode de Jacobi en nous bornant au cas le plus simple (celui de la méthode classique) et sans chercher d'estimations de l'erreur ; les lecteurs intéressés par des compléments concernant la convergence des méthodes de Jacobi classique et cyclique se reporteront avec profit aux articles [1][2][3][4]. On peut d'ores et déjà remarquer que la convergence est rendue possible par le fait que *la somme des carrés de tous les éléments des matrices* A_k *reste constante, alors qu'à chaque étape la somme des carrés des éléments diagonaux s'augmente de la somme des carrés des deux éléments annulés* (c'est ce qu'exprime le théorème 6.1-1). On peut donc espérer que les matrices A_k vont converger vers une matrice diagonale, et que cette matrice diagonale

---

[1] HENRICI, P., On the speed of convergence of cyclic and quasicyclic Jacobi methods for computing eigenvalues of Hermitian matrices, *J. Soc. Indust. Appl. Math.*, **6** (1958), 144-162.

[2] HENRICI, P. ; ZIMMERMANN, K., An estimate for the norms of certain cyclic Jacobi operators, *Linear Algebra and Appl.*, **1** (1968), 489-501.

[3] van KEMPEN, H. P. M., On the convergence of the classical Jacobi method for real symmetric matrices with non-distinct eigenvalues, *Numer. Math.*, **9** (1966), 11-18.

[4] van KEMPEN, H. P. M., On the quadratic convergence of the special cyclic Jacobi method, *Numer. Math.*, **9** (1966), 19-22.

sera précisément la matrice diag (λ_i) à une permutation des indices près ; c'est effectivement ce que nous démontrons dans le théorème qui suit (l'approximation des vecteurs propres est examinée séparément).

Pour éviter des situations triviales, on suppose dans ce qui suit que $\max_{i \neq j} |a_{ij}^k| > 0$ pour tout entier $k \geqslant 1$. On rappelle que $\mathfrak{S}_n$ désigne l'ensemble des permutations de l'ensemble $\{1, 2, \ldots, n\}$.

Théorème 6.1-2 ***(convergence des valeurs propres pour la méthode de Jacobi classique). La suite*** $(A_k)_{k \geqslant 1}$ ***de matrices obtenues par la méthode de Jacobi classique est convergente, et***

$$\lim_{k \to \infty} A_k = \operatorname{diag}(\lambda_{\sigma(i)})$$

pour une permutation convenable $\sigma \in \mathfrak{S}_n$.

DÉMONSTRATION. (i) Commençons par démontrer un lemme, qui joue un rôle essentiel dans la présente démonstration aussi bien que dans celle du prochain théorème : *soit* X *un espace vectoriel normé de dimension finie, et* (x_k) *une suite bornée dans* X, *admettant un nombre fini de valeurs d'adhérence, et telle que* $\lim_{k \to \infty} \|x_{k+1} - x_k\| = 0$. *Alors la suite* (x_k) *est convergente* (vers l'une de ses valeurs d'adhérence, naturellement ...).

Soit en effet a_μ, $1 \leqslant \mu \leqslant M$, les valeurs d'adhérence de la suite. Pour tout $\varepsilon > 0$, il existe donc un entier $l(\varepsilon)$ tel que

$$k \geqslant l(\varepsilon) \Rightarrow x_k \in \bigcup_{\mu=1}^{M} B(a_\mu ; \varepsilon),$$

en notant $B(a ; \varrho) = \{x \in X ; \|x - a\| < \varrho\}$. Sinon, il existerait une suite extraite $(x_{k'})$ telle que

$$x_{k'} \notin \left\{ \bigcup_{\mu=1}^{M} B(a_\mu, \varepsilon) \right\} \quad \text{pour tout} \quad k' \geqslant 1.$$

Comme cette suite extraite est encore une suite bornée d'un espace de dimension finie, il existerait une nouvelle suite extraite $(x_{k''})$ qui convergerait vers un point $x'' \notin \left\{ \bigcup_{\mu=1}^{M} B(a_\mu ; \varepsilon) \right\}$. Mais alors x'' serait une valeur d'adhérence de la suite (x_k) distincte des points a_μ, ce qui contredit l'hypothèse.

Le choix particulier

$$\varepsilon_0 = \frac{1}{3} \min_{\mu \neq \mu'} \|a_\mu - a_{\mu'}\| > 0$$

montre l'existence d'un entier l_0 tel que

$$k \geqslant l_0 \Rightarrow \begin{cases} x_k \in \bigcup_{\mu=1}^{M} B(a_\mu ; \varepsilon_0), \\ \|x_{k+1} - x_k\| \leqslant \varepsilon_0. \end{cases}$$

De la sorte,

$$x_k \in B(a_\mu ; \varepsilon_0) \Rightarrow x_{k+1} \in B(a_\mu ; \varepsilon_0) \quad \text{pour tout} \quad k \geqslant l_0,$$

et le lemme est démontré.

(ii) Revenant à la démonstration proprement dite de la convergence de la méthode de Jacobi classique, *posons, pour tout entier* $k \geqslant 1$,

$$A_k = (a_{ij}^k) = D_k + B_k, \quad \text{avec} \quad D_k \overset{\text{déf}}{=} \operatorname{diag}(a_{ii}^k),$$

et montrons dans un premier temps que $\lim_{k\to\infty} \mathrm{B}_k = 0$. Les nombres

$$\varepsilon_k = \sum_{i\neq j} |a_{ij}^k|^2 = ||\mathrm{B}_k||_E^2, \quad k \geqslant 1,$$

vérifient d'une part (*cf.* théorème 6.1-1)

$$\varepsilon_{k+1} = \varepsilon_k - 2|a_{pq}^k|^2,$$

et, d'autre part (par définition de la stratégie adoptée dans la méthode de Jacobi classique),

$$\varepsilon_k \leqslant n(n-1)|a_{pq}^k|^2,$$

puisqu'il y a $n(n-1)$ éléments hors-diagonaux. Combinant ces relations, on obtient

$$\varepsilon_{k+1} \leqslant \left(1 - \frac{2}{n(n-1)}\right)\varepsilon_k,$$

ce qui montre que $\lim_{k\to\infty} \varepsilon_k = 0$.

(iii) Montrons que *la suite* (D_k) *n'a qu'un nombre fini de valeurs d'adhérence, qui sont nécessairement de la forme diag* $(\lambda_{\sigma(i)})$, $\sigma \in \mathfrak{S}_n$. On rappelle que $\lambda_1, \lambda_2, \ldots, \lambda_n$ désignent les n valeurs propres de la matrice A, supposées rangées une fois pour toutes dans un ordre déterminé, mais arbitraire.

Si $(\mathrm{D}_{k'})$ est une suite extraite convergeant vers une matrice D, alors on a aussi

$$\lim_{k'\to\infty} \mathrm{A}_{k'} = \mathrm{D} \text{ car } \mathrm{A}_{k'} = \mathrm{D}_{k'} + \mathrm{B}_{k'}, \text{ et } \lim_{k'\to\infty} \mathrm{B}_{k'} = 0,$$

et donc (considérer les coefficients du polynôme caractéristique)

$$\text{dét}\,(\mathrm{D} - \lambda\mathrm{I}) = \lim_{k'\to\infty} \text{dét}\,(\mathrm{A}_{k'} - \lambda\mathrm{I}) \quad \text{pour tout} \quad \lambda \in \mathbf{C}.$$

Comme

$$\text{dét}\,(\mathrm{A}_{k'} - \lambda\mathrm{I}) = \text{dét}\,(\mathrm{A} - \lambda\mathrm{I}) \quad \text{pour tout} \quad \lambda \in \mathbf{C},$$

puisque les matrices $\mathrm{A}_{k'}$ et A sont semblables, on déduit que les matrices A et $\mathrm{D} = \lim_{k'\to\infty} \mathrm{D}_{k'}$ ont le même polynôme caractéristique, et donc mêmes valeurs propres, multiplicités comprises.

Comme D est une matrice diagonale (c'est la limite d'une suite de matrices diagonales), il existe une permutation $\sigma \in \mathfrak{S}_n$ telle que

$$\mathrm{D} = \text{diag}\,(\lambda_{\sigma(i)}).$$

(iv) *Montrons que* $\lim_{k\to\infty} (\mathrm{D}_{k+1} - \mathrm{D}_k) = 0$. On vérifie que

$$a_{ii}^{k+1} - a_{ii}^k = \begin{cases} 0 & \text{si } i \neq p, q, \\ -\operatorname{tg}\theta_k a_{pq}^k & \text{si } i = p, \\ \operatorname{tg}\theta_k a_{pq}^k & \text{si } i = q. \end{cases}$$

Comme

$$|\theta_k| \leqslant \frac{\pi}{4} \quad \text{et} \quad |a_{pq}^k| \leqslant ||\mathrm{B}_k||_E \quad \text{avec} \lim_{k\to\infty} \mathrm{B}_k = 0,$$

la conclusion s'en déduit.

(v) *Conclusion.* La suite (D_k) est bornée, puisque

$$\|D_k\|_E \leqslant \|A_k\|_E = \|A\|_E$$

(*cf.* théorème 6.1-1). Alors d'après le lemme (partie (i) de la démonstration), la suite (D_k) converge, et sa limite, qui est l'une de ses valeurs d'adhérence, est nécessairement de la forme diag $(\lambda_{\sigma(i)})$, $\sigma \in \mathfrak{S}_n$ (*cf.* (iii)). Comme $A_k = D_k + B_k$, on déduit que la suite (A_k) est également convergente, avec $\lim_{k\to\infty} A_k = \lim_{k\to\infty} D_k$. ■

Passons maintenant à la convergence des vecteurs propres. On rappelle les relations

$$A_{k+1} = \Omega_k^T A_k \Omega_k = O_k^T A O_k, \quad \text{où} \quad O_k = \Omega_1 \Omega_2 \ldots \Omega_k.$$

Théorème 6.1-3 ***(convergence des vecteurs propres pour la méthode de Jacobi classique). On suppose que toutes les valeurs propres de la matrice*** A ***sont distinctes.***

Alors la suite $(O_k)_{k\geqslant 1}$ ***de matrices construites dans la méthode de Jacobi classique converge vers une matrice orthogonale dont les vecteurs colonnes constituent un eusemble orthonormal de vecteurs propres de la matrice*** A.

DÉMONSTRATION. Le lemme établi au point (i) de la démonstration précédente va à nouveau jouer un rôle crucial.

(i) Montrons que *la suite* (O_k) *n'a qu'un nombre fini de valeurs d'adhérence, qui sont nécessairement de la forme*

$$\left(\boxed{\pm p_{\sigma(1)}} \; \boxed{\pm p_{\sigma(2)}} \; \cdots \; \boxed{\pm p_{\sigma(n)}} \right), \quad \sigma \in \mathfrak{S}_n,$$

où $p_1, p_2, \ldots, p_n$ désignent les vecteurs colonnes de la matrice orthogonale O intervenant dans la relation

$$O^T A O = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Soit $(O_{k'})$ une suite extraite de la suite (O_k) convergeant vers une matrice (orthogonale) O'. D'après le théorème précédent, il existe une permutation $\sigma \in \mathfrak{S}_n$ telle que

$$\text{diag}\,(\lambda_{\sigma(i)}) = \lim_{k'\to\infty} A_{k'} = \lim_{k'\to\infty} (O_{k'}^T A O_{k'}) = (O')^T A O',$$

et l'assertion est démontrée. On notera que l'hypothèse selon laquelle les valeurs propres de la matrice A sont toutes *distinctes* est utilisée de façon essentielle pour conclure à l'existence d'un nombre fini de valeurs d'adhérence.

(ii) *Montrons que* $\lim_{k\to\infty} (O_{k+1} - O_k) = O$. Par construction, l'angle θ_k vérifie

$$\text{tg}\, 2\theta_k = \frac{2a_{pq}^k}{a_{qq}^k - a_{pp}^k}, \quad |\theta_k| \leqslant \frac{\pi}{4}.$$

Utilisant le théorème précédent et (de nouveau) le fait que toutes les valeurs propres de la matrice A sont *distinctes*, on conclut à l'existence d'un entier l tel que

$$k \geqslant l \Rightarrow |a_{qq}^k - a_{pp}^k| \geqslant \frac{1}{2} \min_{i\neq j} |\lambda_i - \lambda_j| > 0$$

(comme les couples (p, q) varient avec l'entier k, on ne peut pas affirmer que les suites (a_{pp}^k) et (a_{qq}^k) convergent). Comme par ailleurs $\lim_{k\to\infty} a_{pq}^k = 0$, on a donc établi que

$$\lim_{k\to 0} \theta_k = 0, \quad \text{et donc} \quad \lim_{k\to\infty} \Omega_k = \mathrm{I}$$

(se reporter à l'expression de la matrice Ω_k en fonction de l'angle θ_k). Comme enfin

$$\mathrm{O}_{k+1} - \mathrm{O}_k = \mathrm{O}_k(\Omega_{k+1} - \mathrm{I}),$$

l'assertion en découle, puisque *la suite* (O_k) *est bornée* (on rappelle que la norme $\|\cdot\|_2$ d'une matrice orthogonale vaut 1). On dispose donc de tous les éléments nécessaires à l'application du lemme, ce qui termine la démonstration. ■

Les démonstrations des théorèmes 6.1-2 et 6.1-3 données ci-dessus sont dues à Michel Crouzeix.

6.2. La méthode de Givens-Householder

C'est une méthode particulièrement bien adaptée à la recherche de *valeurs propres sélectionnées* d'une matrice *symétrique*, par exemple toutes les valeurs propres situées dans un intervalle déterminé à l'avance, ou bien les valeurs propres d'un rang donné (en les supposant ordonnées par ordre croissant), etc. . De plus, cette méthode permet d'approcher chaque valeur propre avec une précision variable, selon le rang de la valeur propre considérée, au gré de l'utilisateur. Par contre, cette méthode ne fournit pas les vecteurs propres correspondants, qu'il faudra chercher par des moyens séparés (*cf.* paragraphe 6.4).

La *méthode de Givens-Householder* comprend deux étapes :

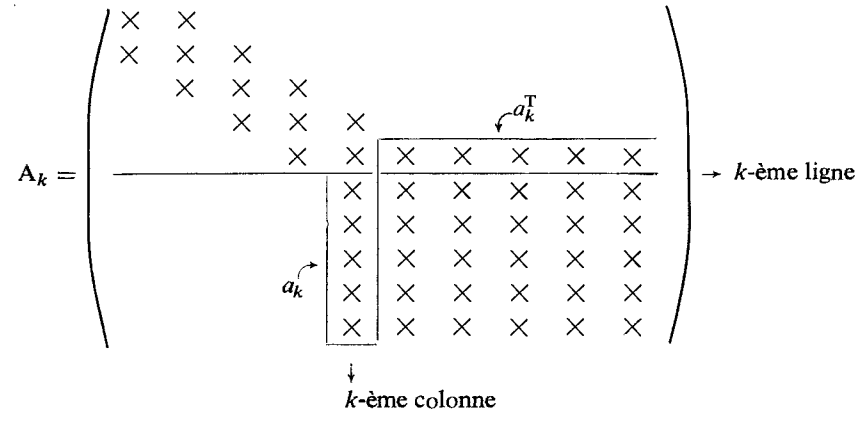
(i) Étant donné une matrice *symétrique* A, on détermine une matrice *orthogonale* P d'un type particulier telle que la matrice symétrique $\mathrm{P^TAP}$ soit *tridiagonale*. Cette étape, qui ne nécessite qu'un nombre *fini* d'opérations élémentaires, constitue la *méthode de Householder de réduction d'une matrice symétrique à la forme tridiagonale.*

(ii) On est ainsi ramené au *calcul des valeurs propres d'une matrice symétrique tridiagonale*, qui s'effectue par la *méthode de Givens* appelée encore *méthode de bissection.*

Commençons par la description de *la méthode de Householder.* Étant donné une matrice symétrique $\mathrm{A} = \mathrm{A}_1$, on détermine de proche en proche $(n-2)$ matrices orthogonales $\mathrm{H}_1, \mathrm{H}_2, \ldots, \mathrm{H}_{n-2}$ telles que les matrices

$$\mathrm{A}_k = \mathrm{H}_{k-1}^{\mathrm{T}}\mathrm{A}_{k-1}\mathrm{H}_{k-1} = (\mathrm{H}_1\mathrm{H}_2 \ldots \mathrm{H}_{k-1})^{\mathrm{T}}\mathrm{A}(\mathrm{H}_1\mathrm{H}_2 \ldots \mathrm{H}_{k-1}), \quad 1 \leqslant k \leqslant n-2,$$

soient de la forme



De la sorte, la matrice (encore symétrique)

$$A_{n-1} = (H_1 H_2 \dots H_{n-2})^T A (H_1 H_2 \dots H_{n-2})$$

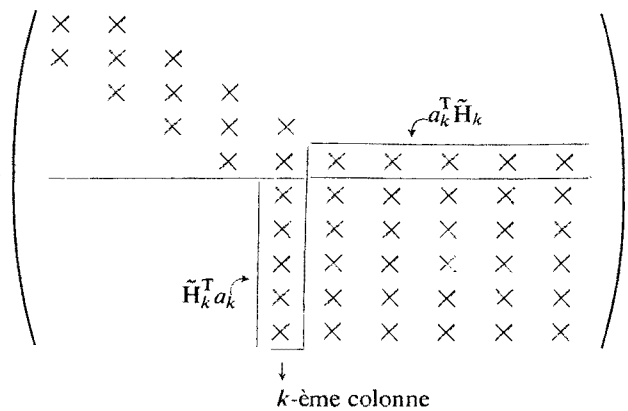
est tridiagonale d'une part, et semblable à la matrice donnée A, d'autre part, la matrice orthogonale P cherchée étant le produit $P = H_1 H_2 \dots H_{n-2}$.

Dans la présente méthode, chaque transformation $A_k \to A_{k+1} = H_k^T A_k H_k$ est effectuée à l'aide d'une matrice de la forme

$$H_k = \left(\begin{array}{c|c} I_k & 0 \\ \hline 0 & \tilde{H}_k \end{array} \right), \quad I_k = \text{matrice unité d'ordre } k.$$

Dans ces conditions, les propriétés de la multiplication par blocs des matrices montrent que la matrice $H_k^T A_k H_k$ est de la forme (le vérifier)

$H_k^T A_k H_k =$



où a_k désigne le vecteur colonne de $\mathbf{R}^{n-k}$ de composantes les éléments a_{ik}^k, $k+1 \leqslant i \leqslant n$, de la matrice $A_k = (a_{ij}^k)$, les éléments a_{ij}^k, $1 \leqslant i, j \leqslant k$, de la matrice A_k restant inchangés. *Il suffit donc de s'arranger pour que le vecteur* $\tilde{H}_k^T a_k \in \mathbf{R}^{n-k}$ *ait toutes ses composantes nulles sauf la première*, la matrice $\tilde{H}_k$ étant orthogonale et "aussi simple que possible à construire".

Cet objectif va être atteint très simplement à l'aide des *matrices de Householder* introduites au paragraphe 4.5 et qui, rappelons-le, sont des matrices à la fois symétriques et orthogonales de la forme particulière :

$$H(v) = I - 2\frac{vv^T}{v^T v}, \quad v : \text{vecteur non nul.}$$

Si $\sum_{i=k+2}^{n} |a_{ik}^k| > 0$, il résulte en effet du théorème 4.5-1 qu'il existe un vecteur $\tilde{v}_k \in \mathbf{R}^{n-k}$ tel que le vecteur $H(\tilde{v}_k) a_k \in \mathbf{R}^{n-k}$ ait toutes ses composantes nulles sauf la première. On posera donc $\tilde{H}_k = H(\tilde{v}_k)$. Naturellement, si $a_{ik}^k = 0$ pour $i = k+2, \dots, n$, on choisit $\tilde{H}_k = I$ (on rappelle qu'on a convenu de considérer la matrice unité également comme une matrice de Householder).

On dispose maintenant de tous les éléments nécessaires à la description du *passage de la matrice* A_k *à la matrice* $A_{k+1} = H_k^T A_k H_k$: si $a_{ik}^k = 0$ pour $i = k+2, \dots, n$, la matrice A_k est déjà de la forme A_{k+1}, et il suffit de poser $H_k = I$. Dans le cas contraire, on construit la matrice H_k à l'aide de la matrice $\tilde{H}_k = H(\tilde{v}_k)$, comme indiqué précédemment.

Il est d'ailleurs loisible de considérer la matrice H_k comme étant elle-même une matrice de Householder :

$$H_k = H(v_k) = I - 2\frac{v_k v_k^T}{v_k^T v_k},$$

le vecteur $v_k \in \mathbf{R}^n$ ayant ses k premières composantes nulles, les $(n-k)$ dernières étant celles du vecteur $\tilde{v}_k \in \mathbf{R}^{n-k}$.

De ces considérations et du théorème 4.5-1, il résulte qu'il existe deux choix possibles pour le vecteur v_k, à savoir

$$v_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a^k_{k+1,k} \pm \left(\sum_{i=k+1}^{n} |a^k_{ik}|^2 \right)^{1/2} \\ a^k_{k+2,k} \\ \vdots \\ a^k_{nk} \end{pmatrix},$$

le choix *effectif* du signe dans l'expression de la $(k+1)$-ème composante étant *celui de l'élément* $a^k_{k+1,k}$ (pour éviter des dénominateurs trop "petits").

Une fois le vecteur v_k déterminé, les éléments a^{k+1}_{ij}, $k+1 \leqslant i, j \leqslant n$, de la matrice $A_{k+1} = (a^{k+1}_{ij})$ sont obtenus de la façon suivante: on calcule successivement les vecteurs

$$w_k = (v_k^T v_k)^{-1/2} v_k,$$
$$q_k = 2(I - w_k w_k^T) A_k w_k,$$

de composantes (w^k_i) et (q^k_i), respectivement. Alors la matrice A_{k+1} prend la forme

$$A_{k+1} = A_k - w_k q_k^T - q_k w_k^T,$$

de sorte que

$$a^{k+1}_{ij} = a^k_{ij} - w^k_i q^k_j - q^k_i w^k_j, \quad k+1 \leqslant i, j \leqslant n.$$

REMARQUE. Le conditionnement $\mathrm{cond}_2(A)$ n'est pas modifié par cette méthode :

$$\mathrm{cond}_2(A) = \mathrm{cond}_2(A_k) \text{ pour } k = 1, \ldots, n-1,$$

puisqu'il est en effet invariant par transformation orthogonale (théorème 2.2-3) ; c'est là un intérêt de la méthode, du point de vue de la "stabilité numérique". ■

Notons au passage le résultat suivant, qui résume l'application de la méthode de Householder :

Théorème 6.2-1. ***Étant donné une matrice symétrique*** A, ***il existe une matrice*** P, ***produit de*** $(n-2)$ ***matrices de Householder, telle que la matrice*** P^TAP ***soit tridiagonale.*** ■

Passons ensuite à la description de la *méthode de Givens* de recherche des valeurs propres d'une matrice symétrique tridiagonale :

$$B = \begin{pmatrix} b_1 & c_1 & & & \\ c_1 & b_2 & c_2 & & \\ \ddots & \ddots & \ddots & \ddots & \\ & c_{n-2} & b_{n-1} & c_{n-1} & \\ & & c_{n-1} & b_n & \end{pmatrix}.$$

On observe tout d'abord que si l'un des éléments c_i est nul, la matrice B est déjà décomposée en deux sous-matrices diagonales du même type. On peut donc supposer, sans restreindre la généralité, que

$$c_i \neq 0, \quad 1 \leqslant i \leqslant n-1.$$

Nous allons commencer par établir que les racines des polynômes caractéristiques des sous-matrices

$$\mathbf{B}_i = \begin{pmatrix} b_1 & c_1 & & & \\ c_1 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & c_{i-2} & b_{i-1} & c_{i-1} \\ & & & c_{i-1} & b_i \end{pmatrix}, \quad 1 \leqslant i \leqslant n,$$

ont des propriétés d'"emboîtement" tout à fait remarquables, illustrées à la figure 6.2-1 ci-après.

Théorème 6.2-2. ***Les polynômes $p_i(\lambda)$, $\lambda \in \mathrm{R}$, définis pour $i = 0, 1, \ldots, n$, par les formules de récurrence :***

$$p_0(\lambda) = 1,$$
$$p_1(\lambda) = b_1 - \lambda,$$
$$p_i(\lambda) = (b_i - \lambda)p_{i-1}(\lambda) - c_{i-1}^2 p_{i-2}(\lambda), \quad 2 \leqslant i \leqslant n,$$

ont les propriétés suivantes :

(1) Le polynôme p_i est le polynôme caractéristique de la matrice $\mathbf{B}_i$, $1 \leqslant i \leqslant n$;

(2) $\lim_{\lambda \to -\infty} p_i(\lambda) = +\infty, \quad 1 \leqslant i \leqslant n$;

(3) $p_i(\lambda_0) = 0 \Rightarrow p_{i-1}(\lambda_0)p_{i+1}(\lambda_0) < 0, \quad 1 \leqslant i \leqslant n-1$;

(4) Le polynôme p_i possède i racines réelles distinctes, qui séparent les $(i+1)$ racines du polynôme p_{i+1}, $1 \leqslant i \leqslant n-1$, comme l'indique la figure 6.2-1.

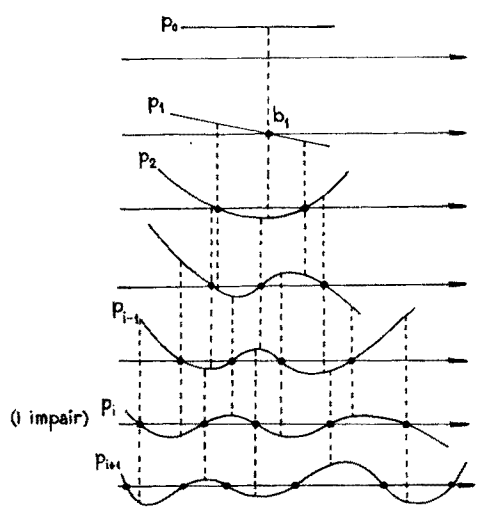


FIG. 6.2-1.

DÉMONSTRATION. (1) Cette propriété se vérifie en développant le déterminant de la matrice $(B_i - \lambda I)$ par rapport à la dernière ligne (ou colonne).

(2) On voit aisément par récurrence que le monôme de plus haut degré du polynôme p_i est $(-1)^i \lambda^i$.

(3) Supposons que $p_i(\lambda_0) = 0$ pour un entier vérifiant $1 \leqslant i \leqslant n-1$. De la relation de récurrence, on déduit

$$p_{i+1}(\lambda_0) = -c_i^2 p_{i-1}(\lambda_0),$$

ce qui entraîne (on a supposé $c_i \neq 0$) :

$$\begin{cases} \text{ou bien} & p_{i-1}(\lambda_0) p_{i+1}(\lambda_0) < 0, \\ \text{ou bien} & p_{i-1}(\lambda_0) = p_i(\lambda_0) = p_{i+1}(\lambda_0) = 0. \end{cases}$$

Dans la seconde éventualité, la relation de récurrence (jointe à nouveau à l'hypothèse $c_i \neq 0$, $1 \leqslant i \leqslant n-1$) montre que

$$p_i(\lambda_0) = p_{i-1}(\lambda_0) = \ldots = p_1(\lambda_0) = p_0(\lambda_0) = 0,$$

ce qui est exclu puisque $p_0(\lambda) = 1$.

(4) Cette propriété est une conséquence des propriétés (2) et (3). ■

Une suite de polynômes vérifiant les conditions (2), (3), (4) du théorème précédent est appelée *suite de Sturm*. La méthode de Givens repose sur une propriété tout à fait remarquable d'une telle suite, qui est de permettre un calcul immédiat du nombre des racines $< \mu$, $\mu \in \mathbf{R}$ donné, de chacun des polynômes de la suite, comme le montre le résultat ci-dessous.

Théorème 6.2-3. ***Soit i un entier vérifiant $1 \leqslant i \leqslant n$. Étant donné un nombre $\mu \in \mathbf{R}$, on pose***

$$\operatorname{sgn} p_i(\mu) = \begin{cases} \textit{signe de} & p_i(\mu) & \textit{si} & p_i(\mu) \neq 0, \\ \textit{signe de} & p_{i-1}(\mu) & \textit{si} & p_i(\mu) = 0. \end{cases}$$

Alors le nombre $N(i\,;\mu)$ de changements de signes entre éléments consécutifs de l'ensemble ordonné

$$E(i\,;\mu) = \{+, \operatorname{sgn} p_1(\mu), \operatorname{sgn} p_2(\mu), \ldots, \operatorname{sgn} p_i(\mu)\}$$

est égal au nombre de racines du polynôme p_i qui sont $< \mu$.

DÉMONSTRATION. On vérifie pour commencer que l'expression $\operatorname{sgn} p_i(\lambda)$ est définie sans ambiguïté dans tous les cas puisque $p_i(\lambda) = 0 \Rightarrow p_{i-1}(\lambda) \neq 0$ d'après le théorème précédent. Puisque

$$\mu \leqslant b_1 \Rightarrow \mathrm{E}_1(\mu) = \{+, +\} \Rightarrow \mathrm{N}(1\,;\mu) = 0,$$
$$b_1 < \mu \Rightarrow \mathrm{E}_1(\mu) = \{+, -\} \Rightarrow \mathrm{N}(1\,;\mu) = 1,$$

l'assertion est vraie pour $i = 1$.

Supposons la propriété vraie pour les entiers $1, 2, \ldots, i$ et montrons qu'elle est encore vraie pour l'entier $(i+1)$. Appelant $\lambda_1^i, \ldots, \lambda_i^i$ et $\lambda_1^{i+1}, \ldots, \lambda_{i+1}^{i+1}$ les racines, rangées par ordre croissant, des polynômes p_i et p_{i+1} respectivement, nous avons

$$\lambda_1^i < \ldots < \lambda_{N(i\,;\mu)}^i < \mu \leqslant \lambda_{N(i\,;\mu)+1}^i < \ldots < \lambda_i^i,$$

par définition de l'entier $N(i\,;\mu)$, et par ailleurs

$$\lambda_{N(i\,;\mu)}^i < \lambda_{N(i\,;\mu)+1}^{i+1} < \lambda_{N(i\,;\mu)+1}^i,$$

d'après le théorème précédent. Il suffit alors d'examiner les différents cas possibles (*cf.* figure 6.2-1) :

$\lambda^i_{N(i\,;\,\mu)} < \mu \leqslant \lambda^{i+1}_{N(i\,;\,\mu)+1} \Rightarrow \operatorname{sgn} p_{i+1}(\mu) = \operatorname{sgn} p_i(\mu) \Rightarrow N(i+1\,;\,\mu) = N(i\,;\,\mu)$;

$\lambda^{i+1}_{N(i\,;\,\mu)+1} < \mu < \lambda^i_{N(i\,;\,\mu)+1} \Rightarrow \operatorname{sgn} p_{i+1}(\mu) = -\operatorname{sgn} p_i(\mu) \Rightarrow N(i+1\,;\,\mu) = N(i,\,\mu)+1$;

$\lambda^i_{N(i\,;\,\mu)+1} = \mu \Rightarrow \operatorname{sgn} p_i(\mu) = \operatorname{sgn} p_{i-1}(\mu) = -\operatorname{sgn} p_{i+1}(\mu) \Rightarrow N(i+1\,;\,\mu) = N(i,\,\mu)+1$

(dans la dernière implication, on utilise de façon essentielle la propriété (3) du théorème précédent). ■

REMARQUE. Appelons $M(i\,;\,\mu)$ le nombre de paires consécutives de même signe ($\{+,+\}$ ou $\{-,-\}$) trouvées dans l'ensemble ordonné $E(i\,;\,\mu)$; par exemple, si

$$E(10\,;\,\mu) = \{+,\,+,\,-,\,+,\,-,\,-,\,+,\,-,\,-,\,-,\,+\},$$

$M(10\,;\,\mu) = 4$, tandis que $N(10\,;\,\mu) = 6$. On vérifie d'ailleurs facilement que

$$M(i\,;\,\mu)+N(i\,;\,\mu) = i, \quad \text{pour} \quad i = 1, \ldots, n,$$

de sorte que le nombre $M(i\,;\,\mu)$ est égal au nombre de racines du polynôme p_i qui sont $\geqslant \mu$. ■

Le résultat du théorème 6.2-3 permet d'approcher *d'aussi près qu'on veut* les valeurs propres de la matrice $\mathrm{B} = \mathrm{B}_n$, et même de *calculer directement une valeur propre de rang donné*. Supposons en effet qu'on souhaite approcher la i-ème valeur propre $\lambda_i = \lambda^n_i$ de la matrice B, l'entier i étant fixé (on suppose comme précédemment les valeurs propres $\lambda_1, \ldots, \lambda_n$ rangées par ordre croissant ; on rappelle qu'elles sont toutes distinctes d'après le théorème 6.2-2).

On commence par déterminer un intervalle $[a_0, b_0]$ dans lequel on est sûr de trouver cette valeur propre ; *par exemple*, on choisit $a_0 = -b_0 = \|\mathrm{B}\|_1$ ou $\|\mathrm{B}\|_\infty$ (théorème 1.4-3), ou bien on utilise le théorème de Gerschgorin (exercice 1.1-5). On note ensuite c_0 le milieu du segment $[a_0, b_0]$ puis on calcule l'entier $N(n\,;\,c_0)$ (avec les notations du théorème 6.2-3 ; on rappelle que p_n n'est autre que le polynôme caractéristique de la matrice B) ; alors

$$\begin{aligned} &- \text{ ou bien} \quad N(n\,;\,c_0) \geqslant i \quad \text{et} \quad \lambda_i \in [a_0, c_0[, \\ &- \text{ ou bien} \quad N(n\,;\,c_0) < i \quad \text{et} \quad \lambda_i \in [c_0, b_0], \end{aligned}$$

de sorte que l'on connaît un intervalle $[a_1, b_1]$, de longueur moitié de celle du précédent, dans lequel se trouve la valeur propre λ_i.

On détermine ainsi de proche en proche une suite d'intervalles emboîtés $[a_k, b_k]$, $k \geqslant 0$, tels que

$$\lambda_i \in [a_k, b_k] \quad \text{et} \quad b_k - a_k = 2^{-k}(b_0 - a_0), \quad k \geqslant 0,$$

de sorte que l'on peut effectivement encadrer la i-ème valeur propre λ_i avec une précision (théoriquement) arbitraire.