

Chapitre V

Valeurs et Vecteurs Propres

Les premiers vecteurs et valeurs propres viennent des équations différentielles (Lagrange 1759, *théorie du son*), la théorie des perturbations séculaires des orbites des 6 planètes (connues à l'époque, Lagrange 1781, *Oeuvres V*, p. 125-490), et les axes principales d'inertie d'un corps solide (Euler 1758, Lagrange 1788). Aujourd'hui, la théorie des valeurs et vecteurs propres, formalisée par Cayley 1858, est indispensable dans toutes les branches de la science, en particulier pour la solution des systèmes des équations différentielles linéaires, la diagonalisation des formes quadratiques et opérateurs autoadjoints, en théorie de stabilité, pour les questions de convergence de processus itératifs, et en physique et chimie (mécanique, circuits, cinétique chimique, équation de Schrödinger).

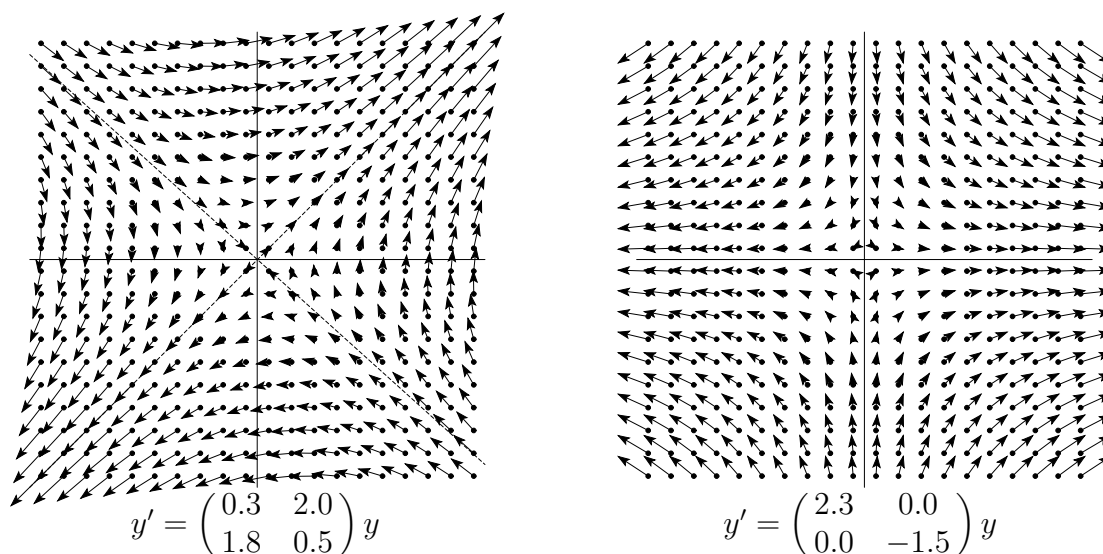


FIG. V.1: Une application linéaire comme champ de vecteurs (à gauche) ; transformée sur la base des vecteurs propres (à droite).

Observons en figure V.1 (à gauche) le champ de vecteurs d'une équation différentielle $y' = Ay$. Deux directions sautent aux yeux : ce sont les directions où le vecteur Av prend la même direction que le vecteur v , i.e., où

$$Av = \lambda v \quad \text{ou} \quad (A - \lambda I)v = 0. \quad (0.1)$$

Si cette équation est vérifiée, λ s'appelle *valeur propre* de la matrice A et v est le *vecteur propre* correspondant. L'équation (0.1) possède une solution v non nulle si et seulement si

$$\chi_A(\lambda) = \det(A - \lambda I) = 0. \quad (0.2)$$

Le polynôme $\chi_A(\lambda)$ est le *polynôme caractéristique* de la matrice A . Les valeurs propres de A sont alors les zéros du polynôme caractéristique.

Bibliographie sur ce chapitre

Tous les livres cités dans le chapitre IV, et en plus . . .

B.N. Parlett (1980): *The Symmetric Eigenvalue Problem*. Prentice-Hall, Englewood Cliffs, NJ.

B.T. Smith, J.M. Boyle, Y. Ikebe, V.C. Klema & C.B. Moler (1970): *Matrix Eigensystem Routines: EISPACK Guide*. 2nd ed., Springer-Verlag, New York.

J.H. Wilkinson (1965): *The Algebraic Eigenvalue Problem*. Clarendon Press. [MA 65/72]

J.H. Wilkinson & C. Reinsch (1971): *Handbook for Automatic Computation, Volume II, Linear Algebra*. Springer-Verlag, New York.

V.1 La condition du calcul des valeurs propres

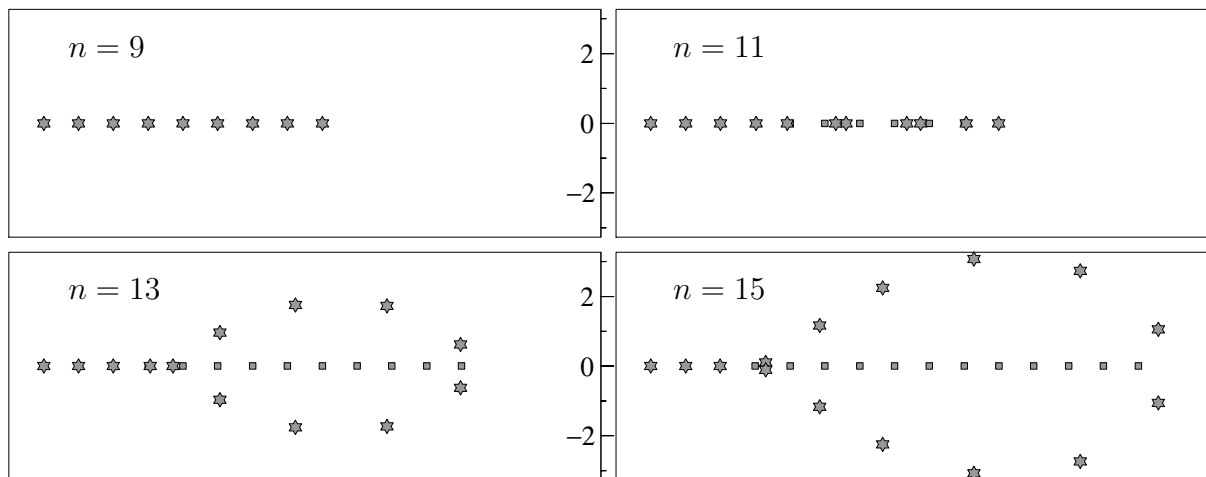


FIG. V.2: Zéros de (1.2) avec des coefficients perturbés

Un algorithme dangereux. La première méthode (déjà utilisée par Lagrange) pour calculer les valeurs propres d'une matrice A est la suivante: *calculer d'abord les coefficients du polynôme caractéristique $\chi_A(\lambda)$ et déterminer ensuite les zéros de ce polynôme*. Si la dimension de A est très petite (disons $n \leq 3$) ou si l'on fait le calcul en arithmétique exacte, cet algorithme peut être très utile. Par contre, si l'on fait le calcul en virgule flottante, cet algorithme peut donner des mauvaises surprises :

Considérons un **exemple de Wilkinson**, i.e., le problème de calculer les valeurs propres de la matrice diagonale

$$A = \text{diag}(1, 2, 3, \dots, n) \quad (1.1)$$

dont le polynôme caractéristique est

$$\begin{aligned} \chi_A(\lambda) &= (1 - \lambda)(2 - \lambda)(3 - \lambda) \cdot \dots \cdot (n - \lambda) \\ &= (-1)^n \lambda^n + a_{n-1} \lambda^{n-1} + a_{n-2} \lambda^{n-2} + \dots + a_1 \lambda + a_0. \end{aligned} \quad (1.2)$$

Les coefficients sont calculés en double précision, puis transformés en simple précision $\hat{a}_i = a_i(1 + \epsilon_i)$ avec $|\epsilon_i| \leq \text{eps}$ ($\text{eps} \approx 6 \cdot 10^{-8}$). Ensuite on calcule le mieux possible les zéros de ce polynôme. Les résultats, totalement faux si n dépasse 10, sont dessinés dans la Fig. V.1 pour $n = 9, 11, 13, 15$.

Conclusion. Eviter le calcul des coefficients du polynôme caractéristique, il s'agit d'un *algorithme numériquement instable* !

Valeurs propres bien conditionnées.

On pourrait encore croire, que les mauvais résultats ci-dessus étaient dûs à la mauvaise condition du problème. Cela n'est pas le cas, comme le montre le théorème suivant :

Théorème 1.1 (Gershgorin) (*Isv. Akad. Nauk 1941*). Soit A une matrice $n \times n$ (avec des éléments dans \mathbb{R} ou dans \mathbb{C}). Si λ est une valeur propre de A , alors il existe un indice i tel que

$$|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad (1.3)$$

c.-à-d. que toutes les valeurs propres de A se trouvent dans l'union des disques

$$D_i = \{ \lambda ; |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \}.$$

Démonstration. Soit $v \neq 0$ un vecteur propre et choisissons l'indice i tel que $|v_i| \geq |v_j|$ pour tout j . La ligne i de l'équation $Av = \lambda v$ donne

$$\sum_{j \neq i} a_{ij} v_j = (\lambda - a_{ii}) v_i.$$

En divisant par v_i et en utilisant l'inégalité de triangle, on obtient

$$|\lambda - a_{ii}| = \left| \sum_{j \neq i} a_{ij} \cdot \frac{v_j}{v_i} \right| \leq \sum_{j \neq i} |a_{ij}|. \quad \square$$

Addendum. On voit encore, par un argument de continuité (en faisant tendre les éléments en dehors de la diagonale vers zéro), que

*Si les disques de Gershgorin sont tous disjoints,
alors chacun contient exactement une valeur propre.*

Théorème 1.2 Soit A une matrice diagonalisable, i.e., il existe T avec $T^{-1}AT = \text{diag}(\lambda_1, \dots, \lambda_n)$, et soit $B = A + E$. Alors pour chaque valeur propre μ de $A + E$ existe un λ_i avec

$$|\mu - \lambda_i| \leq \kappa_\infty(T) \cdot \|E\|_\infty.$$

Démonstration. Nous transformons la matrice $B = A + E$ par la même matrice, qui transforme A sur forme diagonale :

$$T^{-1}BT = \text{diag}(\lambda_1, \dots, \lambda_n) + \underbrace{T^{-1}ET}_C.$$

Maintenant, par Gershgorin,

$$|\mu - \lambda_i| \leq \max\left(\sum |c_{ij}|\right) \leq \|T^{-1}ET\|_\infty \leq \|T^{-1}\|_\infty \|E\|_\infty \|T\|_\infty.$$

□

Exemples.

$$A = \begin{pmatrix} 1 & & & \\ & 2 & & \\ & & \ddots & \\ & & & 10 \end{pmatrix} \quad \text{bien cond. ;} \quad A = \begin{pmatrix} 1 & 10 & & \\ & 2 & 10 & \\ & & \ddots & 10 \\ & & & 10 \end{pmatrix} \quad \text{mal cond.}$$

Différentiabilité d'une valeur propre particulière.**Théorème 1.3** *Soit*

- λ une valeur propre simple de A
- u vecteur propre de gauche de A (i.e., $u^t A = \lambda u^t$)
- v vecteur propre de droite de A (i.e., $Av = \lambda v$)
- $\lambda(\epsilon)$ valeur propre corrs. de $A + \epsilon E$ (ϵ petit)

$$\Rightarrow \lambda(\epsilon) = \lambda + \epsilon \frac{u^t E v}{u^t v} + \mathcal{O}(\epsilon^2).$$

Démonstration. Pour ϵ petit nous posons

$$(A + \epsilon E)(v + \epsilon w) = (\lambda + \epsilon \mu)(v + \epsilon w).$$

Ainsi, μ sera la dérivée directionnelle de la valeur propre λ par rapport à la direction E et w sera la dérivée directionnelle du vecteur propre v .

On développe les produits et on isole les termes en ϵ :

$$Aw + Ev = \lambda w + \mu v.$$

Ceci est un système linéaire pour w , dont la matrice $A - \lambda I$ est singulière ! Donc, seulement pour une valeur particulière de μ il y aura une solution non nulle. Pour la trouver, multiplions cette équation à gauche par u^t :

$$u^t Aw + u^t Ev = \lambda u^t w + \mu u^t v \quad \Rightarrow \quad u^t Ev = \mu u^t v$$

car $u^t A = \lambda u^t$. □

On peut lire dans la formule de ce théorème : le plus le vecteur propre de droite est parallèle au vecteur propre de gauche, le mieux la valeur propre correspondante est *bien conditionnée* (p.ex., pour les matrices symétriques les deux vecteurs sont identiques) ; le plus ils se rapprochent de l'orthogonalité, le plus la valeur propre est *mal conditionnée*.

Exemple. Prenons la valeur propre $\lambda = 1$ de la matrice

$$A = \begin{pmatrix} 1 & \alpha \\ 0 & 2 \end{pmatrix} \quad \text{où} \quad v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad u = \frac{1}{\sqrt{1 + \alpha^2}} \begin{pmatrix} 1 \\ -\alpha \end{pmatrix}$$

Dans cette situation, la formule ci-dessus nous donne $\mu - \lambda = \epsilon \cdot (e_{11} - \alpha e_{21}) + \mathcal{O}(\epsilon^2)$ et le calcul de λ est mal conditionné si α est grand.

Valeurs propres mal conditionnées.

Exemple 1.4 Considérons la matrice (boîte de Jordan)

$$A = \left(\begin{array}{cccc} \lambda_1 & 1 & & \\ & \lambda_1 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_1 \end{array} \right) \Bigg\} n \quad (1.4)$$

Le polynôme caractéristique de $A + \epsilon E$ satisfait

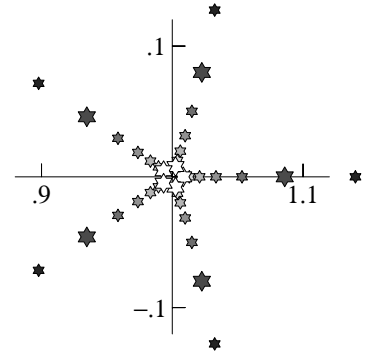
$$\det(A + \epsilon E - \lambda I) = (\lambda_1 - \lambda)^n - (-1)^n \cdot \epsilon \cdot e_{n1} + \mathcal{O}(\epsilon^2) + \mathcal{O}(\epsilon \cdot |\lambda_1 - \lambda|).$$

Si $e_{n1} \neq 0$, les termes $\mathcal{O}(\epsilon^2)$ et $\mathcal{O}(\epsilon \cdot |\lambda_1 - \lambda|)$ sont négligeables par rapport à $\epsilon \cdot e_{n1}$. Les valeurs propres de $A + \epsilon E$ sont alors approximativement données par les racines de

$$(\lambda_1 - \lambda)^n - (-1)^n \cdot \epsilon \cdot e_{n1} = 0, \quad \text{c.-à-d.} \quad \lambda \approx \lambda_1 + (\epsilon \cdot e_{n1})^{1/n} \quad (1.5)$$

(observer que $(\epsilon \cdot e_{n1})^{1/n}$ donne n valeurs complexes distinctes – multiples des racines de l'unité).

Expérience numérique. Prenons la matrice (1.4) avec $\lambda_1 = 1$ et $n = 5$. Les éléments de la matrice E sont des nombres aléatoires dans l'intervalle $[-1, 1]$. Le dessin ci-contre montre les 5 valeurs propres de $A + \epsilon E$ pour $\epsilon = 10^{-4}, 10^{-5}, \dots, 10^{-10}$. L'erreur est $\approx 10^{-1}$ pour $\epsilon = 10^{-5}$ et $\approx 10^{-2}$ pour $\epsilon = 10^{-10}$, ce qui correspond à la formule (1.5) pour $n = 5$. Ce sacré pentagone n'arrête pas de nous poursuivre depuis le début du cours de Géométrie I!...



Conséquence. Si la dimension n d'une boîte de Jordan est plus grande que 1, le calcul de la valeur propre de cette matrice est *très mal conditionné*.

Après avoir montré l'insuffisance des méthodes fondées sur le polynôme caractéristique, il nous reste à voir ce qu'on peut leur substituer. La surprise est que les idées les plus simples sont les meilleures ; nous commençons par chercher à simplifier la matrice donnée :

V.2 Transformation sous forme de Hessenberg (ou tridiagonale)

Soit

$$Av = \lambda v \quad \text{et} \quad v = Tu \quad \Rightarrow \quad T^{-1}ATu = \lambda u,$$

donc, les valeurs propres de A et de $T^{-1}AT$ sont les mêmes et les vecteurs propres v_i se transforment par $v_i = Tu_i$. Le but de ce paragraphe est de trouver une matrice T telle que $T^{-1}AT$ devienne "plus simple".

Première idée naïve. Partons de la matrice

$$\begin{pmatrix} 3 & 2 & 1 \\ 2 & -1 & -2 \\ 1 & -2 & 4 \end{pmatrix}.$$

On essaie de la transformer sous forme triangulaire à l'aide de deux réflexions de Householder $W = I - 2uu^T$ ¹:

$$W_1 A \xrightarrow{\quad} \begin{pmatrix} -3.74 & -.53 & -.80 \\ .00 & -1.75 & -2.53 \\ .00 & -2.38 & 3.73 \end{pmatrix} \quad W_2 W_1 A \xrightarrow{\quad} \begin{pmatrix} -3.74 & -.53 & -.80 \\ .00 & 2.95 & -1.50 \\ .00 & .00 & 4.26 \end{pmatrix}$$

ou $Q^T A = R$. Ces beaux zéros gagnés ont malheureusement détruit les valeurs propres !! Pour préserver ces valeurs propres, nous devons multiplier ce résultat **à droite** par $Q = W_1 W_2$ (rappel: les W sont orthogonaux et symétriques). Une multiplication à droite effectue les *mêmes* opérations, mais sur les **lignes** de la matrice A :

$$W_2 W_1 A W_1 \xrightarrow{\quad} \begin{pmatrix} 3.50 & 1.61 & .27 \\ -1.18 & 2.60 & -1.67 \\ -1.14 & -.34 & 4.09 \end{pmatrix} \quad W_2 W_1 A W_1 W_2 \xrightarrow{\quad} \begin{pmatrix} 3.50 & -1.18 & -1.14 \\ -1.18 & -.20 & -3.09 \\ -1.14 & -3.09 & 2.70 \end{pmatrix} .$$

Cette dernière matrice possède de nouveau les mêmes valeurs propres que A , mais, hélas, les zéros ont disparus !

Première idée sérieuse. Nous répétons le procédé :

$$\mapsto \begin{pmatrix} 3.79 & .43 & 1.08 \\ .43 & 1.61 & -3.46 \\ 1.08 & -3.46 & .60 \end{pmatrix} \mapsto \begin{pmatrix} 3.97 & -.07 & .85 \\ -.07 & 3.24 & -2.72 \\ .85 & -2.72 & -1.21 \end{pmatrix} \mapsto \begin{pmatrix} 4.12 & -.33 & .58 \\ -.33 & 3.99 & -1.71 \\ .58 & -1.71 & -2.10 \end{pmatrix}$$

et après 10 et 15 itérations

$$\dots \mapsto \begin{pmatrix} 4.68 & -.30 & .02 \\ -.30 & 3.91 & -.07 \\ .02 & -.07 & -2.58 \end{pmatrix} \mapsto \dots \mapsto \begin{pmatrix} 4.77 & -.11 & .00 \\ -.11 & 3.82 & -.01 \\ .00 & -.01 & -2.58 \end{pmatrix}$$

Nous constatons que la matrice converge vers une matrice triangulaire (ici même diagonale), où il suffit de cueillir les valeurs propres 4.8, 3.8, -2.6 sur la diagonale, sans aucun polynôme caractéristique ! Le Théorème de Gershgorin nous fournit encore des estimations pour les erreurs. C'est le procédé (pour des matrices de rotation) de Jacobi (Crelle J. vol. 30, 1846) et (pour les matrices LR) de Rutishauser (1958). Les vieux livres d'analyse numérique font ici un théorème et une preuve de convergence. Mais nous allons faire mieux :

Deuxième idée : Forme de Hessenberg (Hessenberg, pendant la deuxième guerre mondiale, non publié, Givens 1956 pour les rotations). Si on demande moins, on obtient plus ! On laisse tranquille les premiers coefficients, et on fait seulement les réflexions sur les coordonnées $2, \dots, n$. Ainsi, les zéros en place $3, \dots, n$ ne vont pas être troublés par les multiplications à droite:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \xrightarrow{W_1 \cdot A} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ \boxed{\alpha} & \boxed{*} & \boxed{*} \\ \boxed{0} & \boxed{*} & \boxed{*} \end{pmatrix} \xrightarrow{W_1 \cdot A \cdot W_1} \begin{pmatrix} a_{11} & \boxed{*} & \boxed{*} \\ \alpha & \boxed{*} & \boxed{*} \\ 0 & \boxed{*} & \boxed{*} \end{pmatrix} .$$

pour notre exemple

$$W_1 \cdot A \xrightarrow{\quad} \begin{pmatrix} 3.00 & 2.00 & 1.00 \\ -2.24 & 1.79 & .00 \\ .00 & -1.34 & 4.47 \end{pmatrix} \quad W_1 \cdot A \cdot W_1 \xrightarrow{\quad} \begin{pmatrix} 3.00 & -2.24 & .00 \\ -2.24 & -1.60 & -.80 \\ .00 & -.80 & 4.60 \end{pmatrix} .$$

¹que nous écrivons maintenant avec W , car "Householder" et "Hessenberg" commencent par la même lettre.

En dimension supérieure, on répète pour les colonnes suivantes :

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ \alpha & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix}}_{W_1 A} \longrightarrow \underbrace{\begin{pmatrix} a_{11} & * & * & * \\ \alpha & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix}}_{W_1 A W_1} \longrightarrow \underbrace{\begin{pmatrix} a_{11} & * & * & * \\ \alpha & * & * & * \\ 0 & \beta & * & * \\ 0 & 0 & * & * \end{pmatrix}}_{W_2 W_1 A W_1} \longrightarrow \underbrace{\begin{pmatrix} a_{11} & * & * & * \\ \alpha & * & * & * \\ 0 & \beta & * & * \\ 0 & 0 & * & * \end{pmatrix}}_{W_2 W_1 A W_1 W_2}$$

etc. Nous avons un triple avantage :

- l'algorithme est le même pour lignes et colonnes, facile à programmer ;
- il n'y a pas de problème de recherche de pivot ;
- si A est symétrique, alors $W_2 W_1 A W_1 W_2$ est aussi symétrique, et donc **tridiagonale**

et nous résumons les théorèmes :

Théorème 2.1 *Chaque matrix A se laisse transformer sous forme de Hessenberg*

$$T^{-1}AT = H = \begin{pmatrix} * & * & \cdots & \cdots & * \\ * & * & \cdots & \cdots & * \\ & * & \cdots & \cdots & * \\ & & \cdots & \cdots & * \\ & & & * & * \end{pmatrix}, \tag{2.1}$$

c.-à-d., $h_{ij} = 0$ pour $i > j + 1$, ayant les mêmes valeurs propres que A .

Théorème 2.2 *Si A est une matrice symétrique, la matrice $H = Q^t A Q$, obtenue par l'algorithme de Householder, est aussi symétrique. Ceci implique que H est automatiquement tridiagonale et symétrique.*

V.3 Méthode de bisection pour des matrices tridiagonales

Considérons une matrice symétrique tridiagonale

$$A = \begin{pmatrix} d_1 & e_2 & & & \\ e_2 & d_2 & e_3 & \cdots & \\ & e_3 & \ddots & \ddots & e_n \\ & & \ddots & \ddots & e_n & d_n \end{pmatrix}. \tag{3.1}$$

On observe tout d'abord que si un élément e_i est nul, la matrice A est déjà décomposée en deux sous-matrices du même type, qui ensemble fournissent les valeurs propres de A .

On peut donc supposer, sans restreindre la généralité, que

$$e_i \neq 0 \quad \text{pour} \quad i = 2, \dots, n. \tag{3.2}$$

Pour cette matrice, il est possible de calculer la valeur $\chi_A(\lambda)$ du polynôme caractéristique sans connaître ses coefficients. En effet, si l'on pose

$$A_1 = (d_1), \quad A_2 = \begin{pmatrix} d_1 & e_2 \\ e_2 & d_2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} d_1 & e_2 & \\ e_2 & d_2 & e_3 \\ & e_3 & d_3 \end{pmatrix}, \quad \dots$$

et si l'on définit $p_i(\lambda) := \det(A_i - \lambda I)$, on obtient

$$\begin{aligned} p_0(\lambda) &= 1 \\ p_1(\lambda) &= d_1 - \lambda \\ p_i(\lambda) &= (d_i - \lambda)p_{i-1}(\lambda) - e_i^2 p_{i-2}(\lambda), \quad i = 2, \dots, n. \end{aligned} \quad (3.3)$$

La formule de récurrence dans (3.3) est obtenue en développant le déterminant de la matrice $A_i - \lambda I$ par rapport à la dernière ligne (ou colonne).

Gravissimum conséquence. On lit dans la formule de récurrence pour chaque $i \in \{1, \dots, n-1\}$:

$$\text{Si } p_i(\hat{\lambda}) = 0, \text{ alors } p_{i-1}(\hat{\lambda}) \text{ et } p_{i+1}(\hat{\lambda}) \text{ ont signes opposés.} \quad (3.4)$$

Cette propriété fondamentale d'une "suite de Sturm" a permis au genevois Charles Sturm de résoudre en 1829 entièrement le problème de la localisation des racines réelles d'un polynôme et ainsi d'établir sa gloire de mathématicien à Paris.

Exactement comme nous l'avons fait pour les racines des $P_k(t)$ (comparer Fig. I.12 à Fig. V.3), et basé sur les conditions limite

$$\text{sign } p_i(-\infty) = 1, \quad \text{sign } p_i(\infty) = (-1)^i, \quad (3.5)$$

nous trouvons :

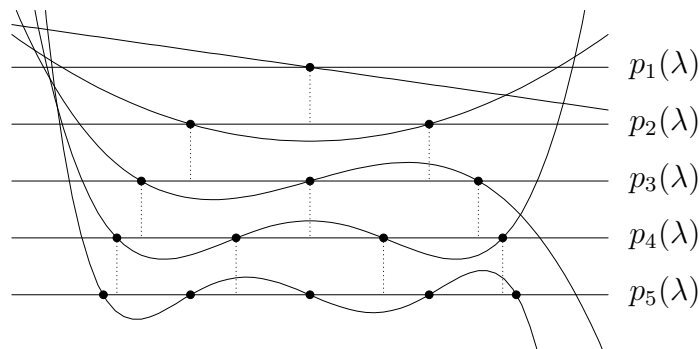


FIG. V.3: Suite de Sturm

Théorème 3.1 Toutes les racines de $p_i(\lambda)$ sont réelles, simples et séparées par celles de $p_{i-1}(\lambda)$. Si l'on définit

$$\omega(\lambda) = \text{nombre de changements de signes de } \{p_0(\lambda), p_1(\lambda), \dots, p_n(\lambda)\}, \quad (3.6)$$

et si $p_n(a) \neq 0$ et $p_n(b) \neq 0$, alors le polynôme $p_n(\lambda)$ possède exactement

$$\omega(b) - \omega(a) \quad (3.7)$$

zéros dans l'intervalle (a, b) .

Méthode de bisection. On obtient toutes les valeurs propres de A de la manière suivante:

- on cherche un intervalle $[a, b]$ qui contienne toutes les valeurs propres de A (p.ex., en appliquant le théorème de Gershgorin). On a donc que $\omega(a) = 0$ et $\omega(b) = n$.
- on pose $c = (a + b)/2$ et on calcule $\omega(c)$. Les différences $\omega(c) - \omega(a)$ et $\omega(b) - \omega(c)$ indiquent combien de valeurs propres de A sont dans $[a, c]$ et combien sont dans $[c, b]$.

- on continue à diviser les intervalles qui contiennent au moins une valeur propre de A .

On peut facilement modifier cet algorithme pour calculer la valeur propre la plus petite ou la 3ème plus grande valeur propre, etc.

Pour éviter un “overflow” dans le calcul de $p_n(\lambda)$ (si n et λ sont grands), il vaut mieux travailler avec

$$f_i(\lambda) := p_i(\lambda)/p_{i-1}(\lambda) \quad i = 1, \dots, n \quad (3.8)$$

et utiliser le fait que

$$\omega(\lambda) = \text{nombre d'éléments négatifs parmi } \{f_1(\lambda), \dots, f_n(\lambda)\} \quad (3.9)$$

(attention: si $p_{i-1}(\lambda)$ est zéro, on pose $f_i(\lambda) = -\infty$; cette valeur compte pour un élément négatif).

Pour une programmation de l'algorithme, on utilise la récurrence

$$\begin{aligned} f_1(\lambda) &= d_1 - \lambda \\ f_i(\lambda) &= d_i - \lambda - \begin{cases} e_i^2/f_{i-1}(\lambda) & \text{si } f_{i-1}(\lambda) \neq 0 \\ |e_i|/eps & \text{si } f_{i-1}(\lambda) = 0. \end{cases} \end{aligned} \quad (3.10)$$

La formule pour le cas $f_{i-1}(\lambda) \neq 0$ est une conséquence de (3.3). Si $f_{i-1}(\lambda) = 0$ (c.-à-d., $p_{i-1}(\lambda) = 0$), on remplace cette valeur par $|e_i| \cdot eps$. Ceci correspond à ajouter la perturbation $|e_i| \cdot eps$ à d_{i-1} .

V.4 Algorithme QR pour valeurs propres

La méthode QR, due à J.C.F. Francis et à V.N. Kublanovskaya, est la méthode la plus couramment utilisée pour le calcul de l'ensemble des valeurs propres ... (P.G. Ciarlet 1982)

... the QR iteration, and it forms the backbone of the most effective algorithm for computing the Schur decomposition. (G.H. Golub & C.F. van Loan 1989)

Cet algorithme a été développé indépendamment par J.G.F. Francis (1961) et par V.N. Kublanovskaya (1961). Un algorithme similaire, qui utilise la décomposition LR à la place de la décomposition QR, a été introduit par H. Rutishauser (1958).

Problème. Calculer les valeurs (et vecteurs) propres d'une matrice quelconque.

Premier pas. Transformons A sous forme de Hessenberg (voir section V.2).

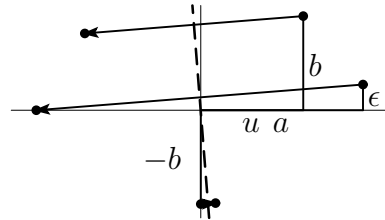
Deuxième pas. Revenons alors à notre “première idée naïve” : transformons notre matrice de Hessenberg par une suite de réflexions W_1, W_2, \dots , **qui sont toutes de dimension 2**, à la forme triangulaire (voir première ligne de (4.1)), suivie de réflexions sur les lignes W_1, W_2, \dots , qui, comme on sait, détruisent à nouveau les zéros gagnés, mais qui préservent la forme de Hessenberg (voir deuxième ligne de (4.1)) :

$$\begin{aligned} & \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ & a_{32} & a_{33} & a_{34} \\ & & a_{43} & a_{44} \end{pmatrix}}_A \quad \rightarrow \quad \underbrace{\begin{pmatrix} \alpha & * & * & * \\ 0 & * & * & * \\ & a_{32} & a_{33} & a_{34} \\ & & a_{43} & a_{44} \end{pmatrix}}_{W_1 A} \quad \rightarrow \quad \underbrace{\begin{pmatrix} \alpha & * & * & * \\ & \beta & * & * \\ & 0 & * & * \\ & & a_{43} & a_{44} \end{pmatrix}}_{W_2 W_1 A} \quad \rightarrow \quad \underbrace{\begin{pmatrix} \alpha & * & * & * \\ & \beta & * & * \\ & & \gamma & * \\ & & 0 & * \end{pmatrix}}_{W_3 W_2 W_1 A} \\ & \rightarrow \quad \underbrace{\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ & & \gamma & * \\ & & & * \end{pmatrix}}_{W_3 W_2 W_1 A W_1} \quad \rightarrow \quad \underbrace{\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ & & * & * \end{pmatrix}}_{W_3 W_2 W_1 A W_1 W_2} \quad \rightarrow \quad \underbrace{\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ & & * & * \end{pmatrix}}_{W_3 W_2 W_1 A W_1 W_2 W_3} \end{aligned} \quad (4.1)$$

Cette dernière matrice possède les mêmes valeurs propres que la première. Le miracle est alors, que si l'on répète cet algorithme, *ces matrices convergent (presque toujours si les valeurs propres sont réelles et pas de mêmes valeurs absolues) vers une matrice triangulaire, dont les valeurs propres se trouvent dans la diagonale en ordre décroissant $|\lambda_1| > |\lambda_2| > |\lambda_3| > \dots$. La vitesse de convergence dépend des rapports λ_i/λ_{i-1} .*

Explication. Supposons d'être déjà proche de la limite

$$A = \begin{pmatrix} a & u & * & * \\ \epsilon & b & v & * \\ & \zeta & c & w \\ & & \eta & d \end{pmatrix}$$



avec ϵ, ζ, η infiniment petits. Alors, par exemple, la première réflexion W_1 enverra (voir la petite figure)

$$\begin{pmatrix} a \\ \epsilon \end{pmatrix} \mapsto \begin{pmatrix} -a \\ 0 \end{pmatrix}, \quad \begin{pmatrix} u \\ b \end{pmatrix} \mapsto \begin{pmatrix} -u \\ b \end{pmatrix}, \quad \begin{pmatrix} 0 \\ -b \end{pmatrix} \mapsto \begin{pmatrix} \epsilon \frac{b}{a} \\ -b \end{pmatrix}.$$

Ainsi, la chaîne du bonheur dans (4.1) devient ici

$$\begin{aligned} &\mapsto \begin{pmatrix} -a & -u & * & * \\ 0 & b & v & * \\ & \zeta & c & w \\ & & \eta & d \end{pmatrix} \mapsto \begin{pmatrix} -a & -u & * & * \\ & -b & -v & * \\ & 0 & c & w \\ & & \eta & d \end{pmatrix} \mapsto \begin{pmatrix} -a & -u & * & * \\ & -b & -v & * \\ & & -c & -w \\ & & 0 & d \end{pmatrix} \\ &\mapsto \begin{pmatrix} a & -u & * & * \\ \epsilon \frac{b}{a} & -b & -v & * \\ & & -c & -w \\ & & & d \end{pmatrix} \mapsto \begin{pmatrix} a & u & * & * \\ \epsilon \frac{b}{a} & b & -v & * \\ & \zeta \frac{c}{b} & -c & -w \\ & & & d \end{pmatrix} \mapsto \begin{pmatrix} a & u & * & * \\ \epsilon \frac{b}{a} & b & v & * \\ & \zeta \frac{c}{b} & c & -w \\ & & -\eta \frac{d}{c} & d \end{pmatrix}. \end{aligned} \quad (4.2)$$

Exemple numérique. Appliquons la méthode QR à la matrice

$$A = \begin{pmatrix} 10 & 2 & 3 & 5 \\ 3 & 6 & 8 & 4 \\ & 5 & 4 & 3 \\ & & 4 & 3 \end{pmatrix}. \quad (4.3)$$

On constate que

$$\frac{a_{i+1,i}^{(k+1)}}{a_{i+1,i}^{(k)}} \approx \frac{\lambda_{i+1}}{\lambda_i} \quad (4.4)$$

($\lambda_1 \approx 14.3$, $\lambda_2 \approx 7.86$, $\lambda_3 \approx 2.70$, $\lambda_4 \approx -1.86$). Comme $|\lambda_{i+1}/\lambda_i| < 1$, les éléments $a_{i+1,i}^{(k)}$ convergent, pour $k \rightarrow \infty$, linéairement vers 0 (voir la Fig. V.3).

Shift et déflation.

Nous allons tourner notre algorithme en une méthode puissante à l'aide de deux Lemmes simples :

Lemme 1. (shift)

$$\lambda \text{ val. prop. de } A \quad \Rightarrow \quad \lambda - p \text{ val. prop. de } A - pI; \text{ même vect. prop.}$$

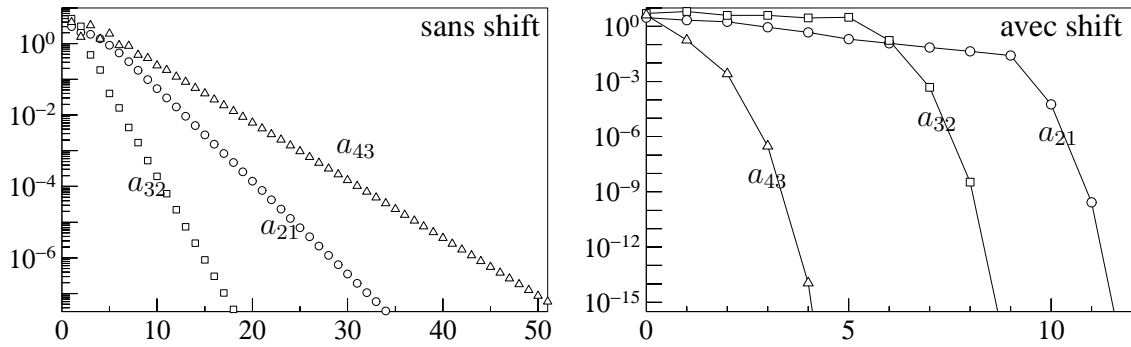


FIG. V.4: Convergence de la méthode QR (à gauche : sans shift, à droite : avec shift)

Lemme 2. (déflation)

$$A = \begin{pmatrix} B & D \\ C & \end{pmatrix} \quad (B \text{ et } C \text{ matrices carré}) \Rightarrow \begin{cases} \lambda \text{ val. prop. de } B \Rightarrow \lambda \text{ val. prop. de } A \\ \mu \text{ val. prop. de } C \Rightarrow \mu \text{ val. prop. de } A \end{cases}$$

Preuve.

- a) $Bv = \lambda v \Rightarrow \begin{pmatrix} B & D \\ C & \end{pmatrix} \begin{pmatrix} v \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} v \\ 0 \end{pmatrix} ;$
- b) $Cw = \mu w \Rightarrow \begin{pmatrix} B & D \\ C & \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = \mu \begin{pmatrix} v \\ w \end{pmatrix} \quad \text{si } Bv + Dw = \mu v ;$

si μ n'est pas valeur propre de B , il existe un vecteur v pour cette dernière condition ; si oui, on retourne au point (a). □

Algorithme. Nous voyons en (4.2), que $a_{n,n-1}$ deviendrait nul (plus précisément de l'ordre $\mathcal{O}(\eta^2)$), si par hasard le d (i.e., le a_{nn}) est zéro. L'idée est alors de faire un shift $A \mapsto A - pI$ avec $p = a_{nn}$ avant d'effectuer les transformations QR de (4.1). L'élément $a_{n,n-1}$ tendra alors très vite vers zéro et, par le Lemme de la déflation, la première valeur propre est trouvée (la valeur a_{nn} à laquelle il faut additionner tous les shifts). Puis on pose $n \mapsto n - 1$ et on continue l'algorithme avec la matrice réduite.

Exemple numérique. Nous avons appliqué l'algorithme QR à la matrice (4.3). La convergence de $a_{i+1,i}$ vers zéro est illustrée à droite de la Fig. V.4. Une comparaison avec la gauche nous montre que la convergence est beaucoup plus rapide (convergence quadratique). Après 5 itérations, on a $|a_{43}| \leq 10^{-15}$. Encore 4 itérations pour la matrice de dimension 3 donnent $|a_{32}| \leq 10^{-15}$. Il ne reste plus que 3 itérations à faire pour la matrice de dimension 2 pour avoir $|a_{21}| \leq 10^{-15}$. En tout, 12 itérations ont donné toutes les valeurs propres avec une précision de 15 chiffres.

Valeurs propres complexes et le "double shift" de Francis. L'algorithme ci-dessus a encore un petit défaut : il ne fonctionne pas pour les valeurs propres complexes (pour A réelle). On devrait faire des shifts avec un p complexe et l'algorithme précédent nécessite un calcul avec des matrices complexes. Mais pour les programmeurs de l'âge de la pierre informatique ² il était important d'éviter un tel calcul.

Idée de Francis (1961): soient $\lambda = \alpha \pm i\beta$ les valeurs propres de la sous-matrice des a_{ij} pour $n - 1 \leq i, j \leq n$. Alors on effectue un *double shift*, d'abord avec $p = \alpha + i\beta$, puis avec $p = \alpha - i\beta$. Après ces deux itérations QR, la matrice sera de nouveau réelle. Il existe une astuce pour calculer cette dernière matrice sans passer par le complexe. Pour plus de détails voir les livres mentionnés (ou le polycopié des années passées).

²... et pour les programmeurs modernes traitant des problèmes de taille gigantesque ...