Notes de cours
Algèbre Linéaire et Analyse Matricielle
3BIM INSA Lyon
2015-2016
v06122015

Ces notes de cours sont inspirées des livres de Algèbre des matrices de J Fresnel [1] et Numerical Linear Algebra de LN Trefethen et D Bau III [2].

6 Algorithmes pour les valeurs propres

Les valeurs propres d'une matrice sont les racines du polynôme caractéristique. Trouver les valeurs propres d'une matrice est équivalent à trouver les racines de son polynôme. Pour les matrices de taille $n \geq 5$, il n'existe en général pas d'expressions closes pour les racines du polynôme caractéristique basées sur des expressions primaires (additions, soustractions, multiplications, divisions et racines). Ce résultat implique que *les méthodes pour trouver les valeurs propres d'une matrice doivent être itératives.*

Une façon de calculer les valeurs propres serait de calculer les racines du polynôme caractéristique en utilisant une méthode numérique de calcul des racines, comme `roots` en Matlab ou `numpy.roots` en Python. Mais trouver les racines d'un polynôme est en général un **problème mal conditionné**. Le conditionnement d'un problème n'a pas été défini, mais un **problème mal conditionné** est un problème pour lequel un petit changement dans les données peut induire un changement non contrôlé dans les résultats.

6.1 Polynôme de Wilkinson

Pour illustrer le problème avec le calcul des racines d'un polynôme, on peut prendre l'exemple classique du **polynôme de Wilkinson**,

$$w(x) = \prod_{i=1}^{20}(x-i) = (x-1)(x-2)\cdots(x-20).$$

Ce polynôme de degré 20 est exprimé sous forme factorisé, et les racines sont $x_i = i$, $i = 1, ..., 20$. Si le problème de trouver les racines d'un polynôme est bien conditionné, on s'attend à ce que un petit changement dans les coefficient du polynôme mène à un petit changement dans les valeurs des racines. Malheureusement ce n'est pas le cas: le polynôme x^2 a une racine double en $x = 0$, mais $x^2 - \epsilon$ a une paire de racine $x = \pm\sqrt{(\epsilon)}$. Pour epsilon petit, $\epsilon \ll \sqrt{\epsilon}$ et donc les racines sont très "loin" de 0 même quand les deux polynômes (x^2 et $x^2 - \epsilon$) sont proches.

Ce type de problème peut se produire même quand les racines du polynôme sont bien espacés, comme dans le cas du polynôme de Wilkinson. Le coefficient de x^{19} du polynôme de Wilkinson est -210. Si on modifie le coefficient de $\epsilon = 2^{-23}$, les racines du polynôme peuvent dévier significativement. Par exemple, la onzième racine devient $10.09549 + 0.64215i$, au lieu de 11.

6.2 Quotient de Rayleigh

Pour une matrice réelle et symétrique et un vecteur réel x non-nul, le **quotient de Rayleigh** est le scalaire défini par

$$r(x) = \frac{{}^t x A x}{{}^t x x}.$$

Si x est un vecteur propre de A, alors $r(x)$ est une valeur propre de A. Le quotient de Rayleigh est le scalaire qui "ressemble" le plus à une valeur propre, dans le sens des moindres carrés. Quelle est la valeur α qui minimise $||Ax - \alpha x||_2$? Le problème de moindres carrés associé est $x\alpha = Ax$ (problème $m \times 1$ avec x la matrice, α l'inconnue et Ax le vecteur membre de droite). La solution du problème de moindres carrés est donnée par l'équation normale:

$$\begin{aligned} \alpha = x^+(Ax) &= ({}^t x x)^{-1} {}^t x (Ax), \\ &= \frac{{}^t x A x}{{}^t x x}, \end{aligned}$$

ce qui est exactement le quotient de Rayleigh. De plus, si l'on prend un vecteur x en tant que variable, le gradient du quotient de Rayleigh par est

$$\nabla r(x) = \frac{2}{{}^t x x}\big(Ax - r(x)x\big).$$

Le gradient de $r = 0$ si et seulement si x est un vecteur propre. De plus, si q est un vecteur propre de A, alors

$$r(x) - r(q) = O(||x - q||^2).$$

Autrement dit, si x est une approximation du vecteur propre q, alors $r(x)$ est une approximation quadratique de la valeur propre $r(q)$.

6.3 Méthode des puissances

Une méthode bien connue et utilisée pour calculer une valeur propre et est la méthode des puissances. La suite

$$\frac{x}{||x||}, \frac{Ax}{||Ax||}, \frac{A^2 x}{||A^2 x||}, \dots$$

converge, dans certaines conditions, vers le vecteur propre principal associée à la valeur propre maximale en valeur absolue.

Méthode des puissances.
Data: une matrice A carrée avec une valeur propre dominante réelle
Result: Un vecteur propre v et une valeur propre λ
$v^{(0)}$ un vecteur de norme 1;
for $k = 1$ **to**... **do**
    $w = Av^{(k-1)}$;
    $v^{(k)} = w/||w||$;
    $\lambda^{(k)} = {}^t v^{(k)} A v^{(k)}$;
end

Cet algorithme ne convergera si la valeur propre dominante n'est pas unique (par exemple si on a une paire de valeurs propres complexes), ou si le vecteur initial $v^{(0)}$ est orthogonal avec le vecteur propre recherché. C'est la première méthode itérative qu'on voit, mais cette méthode n'est pas très efficace. Voyons à quelle vitesse on converge. On suppose que A est diagonalisable et a des valeurs propres $|\lambda_1| > |\lambda_2| \geq ... \geq |\lambda_m|$. Alors A se décompose en valeurs propres $A = XDX^{-1}$, avec x_1 le vecteur propre associé à la plus grande valeur propre λ_1.

$$\begin{aligned}
v^{(k)} &= c_k A^k v^{(0)}, \\
&= c_k \big(XDX^{-1}\big)^k v^{(0)}, \\
&= c_k X D^k X^{-1} v^{(0)}, \\
&= c_k \lambda^k X \operatorname{diag}[1, \lambda_2/\lambda_1, ..., \lambda_m/\lambda_1] X^{-1} v^{(0)},
\end{aligned}$$

Quand $k \to \infty$, le terme $X \operatorname{diag}[1, \lambda_2/\lambda_1, ..., \lambda_m/\lambda_1] X^{-1} v^{(0)}$ converge vers

$$\begin{aligned}
X \operatorname{diag}[1, 0, ..., 0] X^{-1} v^{(0)} &= X \begin{pmatrix} {}^t x_1 \\ \mathbf{0} \end{pmatrix} v^{(0)}, \\
&= X \begin{pmatrix} {}^t x_1 v^{(0)} \\ \mathbf{0} \end{pmatrix}, \\
&= {}^t x_1 v^{(0)} x_1.
\end{aligned}$$

On obient donc que la convergence du vecteur propre se fait en $O(|\lambda_2/\lambda_1|^k)$. La convergence du quotient de Rayleigh se fait en $O(|\lambda_2/\lambda_1|^{2k})$.

6.4 Itération inverse

Pour n'importe quelle valeur μ non valeur propre de A, la matrice $(A - \mu I)^{-1}$ possède les mêmes vecteurs propres que A, et les valeurs propres sont $(\lambda - \mu)^{-1}$

pour λ valeur propre de A. Si μ est proche d'une valeur propre λ_j de A, alors $(\lambda_j - \mu)^{-1}$ sera une valeur propre beaucoup plus grande que $(\lambda_i - \mu)^{-1}$ pour $i \neq j$. Si on applique la méthode des puissance à $(A - \mu I)^{(-1)}$, on convergera rapidement.

Itération inverse
Data: une matrice A carrée telle que $(A - \mu I)^{-1}$ a une valeur propre dominante réelle
Result: Un vecteur propre v et une valeur propre λ
$v^{(0)}$ un vecteur de norme 1;
for $k = 1$ **to**... **do**
  Résoudre $(A - \mu I)w = v^{(k-1)}$;
  $v^{(k)} = w/||w||$;
  $\lambda^{(k)} = {}^t v^{(k)} A v^{(k)}$;
end

Avec l'itération inverse, on peut contrôler la convergence vers n'importe quelle valeur propre. De plus la convergence peut sera en $O(|\lambda_j - \mu|/|\lambda_s - \mu|^k)$, où λ_k est la deuxième valeur propre la plus proche de μ.

6.5 Itération du quotient de Rayleigh

Comme μ est une estimation initiale de la valeur propre, il est tentant d'utiliser le quotient de Rayleigh comme valeur de μ. Si on met à jour μ à après chaque itération, on devrait converger encore plus rapidement:

Itération du quotient de Rayleigh
Data: une matrice A carrée telle que $(A - \mu I)^{-1}$ a une valeur propre dominante réelle
Result: Un vecteur propre v et une valeur propre λ
$v^{(0)}$ un vecteur de norme 1;
$\lambda^{(0)} = {}^t v^{(0)} A v^{(0)}$ le quotient de Rayleigh;
for $k = 1$ **to**... **do**
  Résoudre $(A - \lambda^{(k-1)})w = v^{(k-1)}$;
  $v^{(k)} = w/||w||$;
  $\lambda^{(k)} = {}^t v^{(k)} A v^{(k)}$;
end

Cette méthode permet de passer d'une convergence quadratique pour les valeurs propres en convergence cubique: $|\lambda^{(k+1)} - \lambda_j| = O(|\lambda^{(k)} - \lambda_j|^3)$.

6.6 Factorisation de Schur

L'idée est de factoriser A de façon à révéler ses valeurs propres. Par exemple, si

$$A = XBX^{-1}$$

avec X inversible, alors les matrices A et B sont semblables et donc partagent les mêmes valeurs propres. Si B est diagonale ou même triangulaire, les valeurs propres se trouveront sur la diagonale. La diagonalisation n'existe pas toujours, mais *on peut toujours trouver une factorisation de la forme*

$$A = QTQ^*$$

avec Q unitaire et T triangulaire. Cette factorisation est appelée factorisation de Schur.

Théorème 6.1. *Toutes les matrices carrées admettent une factorisation de Schur.*

On montre l'existence d'une factorisation de Schur par induction sur la taille de la matrice. Pour une matrice 1×1, la factorisation est triviale. On suppose qu'une factorisation existe pour les matrice de taille $n-1$ et on prend une matrice A carrée de taille n, x un vecteur propre (qui existe toujours) et λ la valeur propre associée. On construit une matrice unitaire U

$$U = \left[\begin{array}{c|c} x & \cdots \end{array} \right]$$

Alors

$$U^*AU = \begin{bmatrix} \lambda & B \\ 0 & C \end{bmatrix}$$

avec B un vecteur ligne de taille $n-1$ et C une matrice carrée de taille $n-1$. Par hypothèse d'induction, C admet un factorisation $C = VTV^*$. Alors la matrice

$$Q = U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}.$$

est unitaire, et

$$
\begin{aligned}
Q^*AQ &= \left(U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}\right)^* A \left(U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}\right), \\
&= \begin{bmatrix} 1 & 0 \\ 0 & V^* \end{bmatrix} U^*AU \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}, \\
&= \begin{bmatrix} 1 & 0 \\ 0 & V^* \end{bmatrix} \begin{bmatrix} \lambda & B \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}, \\
&= \begin{bmatrix} \lambda & VB \\ 0 & V^*CV \end{bmatrix}, \\
&= \begin{bmatrix} \lambda & VB \\ 0 & T \end{bmatrix}.
\end{aligned}
$$

Cette dernière matrice est une matrice triangulaire et la factorisation de Schur est complète.

6.7 Algorithme QR pour les valeurs propres

L'idée est de calculer une factorisation de Schur à l'aide d'une méthode itérative. Soit A une matrice carrée et une décomposition QR de A. On définit $A_0 = A = Q_0R_0$. A l'étape k, on effectue une décomposition QR de matrice $A_k = Q_kR_k$ et on définit $A_{k+1} = R_kQ_k$ (notez l'ordre inversé de la multiplication RQ). Donc

$$A_{k+1} = R_kQ_k = (Q_k^*A_k)Q_k = Q_k^*A_kQ_k.$$

Toutes les matrices A_k sont semblable, et sont issues de transformations unitaires, qui sont très stables numériquement. Sous certaines conditions, les matrices A_k convergent vers une matrice triangulaire, celle de la décomposition de Schur (notez la forme $A_k = Q_kA_{k+1}Q_k^*$). Les valeurs propres sont alors révélées sur la diagonale de A_k. En pratique, on peut s'arrêter d'itérer quand A_k est presque triangulaire.

6.8

References

[1] Jean Fresnel. *Algebre des matrices*. Hermann, 2013.

[2] Lloyd N Trefethen and David Bau III. *Numerical linear algebra*, volume 50. Siam, 1997.