Le théorème de Perron-Frobenius, les chaines de Markov et un célèbre moteur de recherche

Bachir Bekka

Février 2007

Le théorème de Perron-Frobenius a d'importantes applications en probabilités (chaines de Markov) et en théorie des graphes. Une belle et récente application est l'algorithme PageRank avec lequel Google classe les pages web.

1 Le théorème de Perron-Frobenius

Une matrice $A = (a_{ij})_{1\leq i,j\leq n} \in M_n(\mathbf{R})$ est dite *positive* et on écrit $A \geq 0$ (respectivement *strictement positive* et on écrit $A > 0$) si $a_{ij} \geq 0$ (respectivement $a_{ij} > 0$) pour tous $1 \leq i, j \leq n$. On parle de même d'un vecteur positif (respectivement strictement positif) $x \in \mathbf{R}^n$. Pour $x, y \in \mathbf{R}^n$, on écrit $x \geq y$ (respectivement $x > y$) si $x - y \geq 0$ (respectivement $x - y > 0$)

La preuve du lemme suivant est un exercice facile.

Lemme 1 *Pour une matrice $A = (a_{ij})_{1\leq i,j\leq n} \in M_n(\mathbf{R})$, on a :*
(i) $A \geq 0$ si et seulement si $Ax \geq 0$ pour tout $x \in \mathbf{R}^n$ tel que $x \geq 0$;
(ii) $A > 0$ si et seulement si $Ax > 0$ pour tout $x \in \mathbf{R}^n \setminus \{0\}$ tel que $x \geq 0$

On note $\sigma(A)$ le *spectre* de A, c-à-d l'ensemble des valeurs propres (complexes) de A. Le *rayon spectral* $r(A)$ de A est $\max_{\lambda\in\sigma(A)} |\lambda|$.

Une valeur propre λ de A est *simple* si α est une racine simple du polynôme caractérstique de A et si $\dim_{\mathbf{C}} \ker(A - \lambda I_n) = 1$. Elle est dite *dominante* si, pour tout $\alpha \in \sigma(A) \setminus \{\lambda\}$, on a $|\lambda| > |\alpha|$.

Définition 2 Soit $A \in M_n(\mathbf{R}), A \geq 0$.
(i) A est dite *primitive* s'il existe $k \in \mathbf{N}$ tel que $A^k > 0$.

(ii) A est dite *irréductible* si, pour tout couple (i,j), il existe $k \in \mathbf{N}$ tel que $(A^k)_{ij} > 0$.

Il est clair qu'une matrice primitive est irréductible. L'exemple $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ montre que la réciproque est fausse.

Dans la suite, $\mathbf{C}^n$ sera muni de la norme $\|x\|_1 = \sum_{i=1}^n |x_i|$ et $M_n(\mathbf{C})$ de la norme d'opérateur associée.

Théorème 3 *Soit $A = (a_{ij})_{1 \le i,j \le n} \in M_n(\mathbf{R})$ une matrice positive et $r = r(A)$ son rayon spectral.*

(i) ***(Perron)*** *Supposons que A est primitive. Alors $r > 0$, r est une valeur propre dominante et simple de A et il existe un unique vecteur x^+ strictement positif tel que $Ax^+ = rx^+$ et $\|x^+\|_1 = 1$.*

(ii) ***(Frobenius)*** *Supposons que A est irréductible. Alors $r > 0$, r est une valeur propre simple de A et il existe un unique vecteur x^+ strictement positif tel que $Ax^+ = rx^+$ et $\|x^+\|_1 = 1$.*

La seule différence entre les cas (i) et (ii) est que, dans le deuxième cas, r n'est pas nécessairement une valeur propre dominante: il peut y avoir d'autres valeurs propres de même module (comme dans l'exemple de la matrice A ci-dessus). Le vecteur propre x^+ est appelé le *vecteur de Perron-Frobenius* de A.

Nous nous contenterons de donner une preuve de la partie (i) due à Perron qui est souvent suffisante pour les applications.

Preuve de (i) Comme A est primitive, il existe $k \in \mathbf{N}$ tel que $A^k > 0$.

• *1ère etape:* On peut supposer que $A > 0$. En effet, supposons le théoréme prouvé pour $B = A^k$. Le spectre de B est $\sigma(B) = \{\lambda^k \,:\, \lambda \in \sigma(A)\}$ et un vecteur propre de A pour une valeur propre λ est un vecteur propre de B pour la valeur propre λ^k. Le rayon spectral de B est $s = r^k$. Donc $r > 0$ et r est une valeur propre de A : comme s est une valeur propre de B, il existe une valeur propre λ de A telle que $\lambda^k = s$; soit x un vecteur propre de A pour la valeur propre λ. Comme x est un vecteur propre de B pour la valeur propre s, on peut supposer que x est l'unique vecteur propre $x^+ > 0$ de B avec $\|x^+\|_1 = 1$ pour la valeur propre s et il s'ensuit que $\lambda > 0$ et donc $\lambda = r$. Le nombre r est une valeur propre simple de A : si $x, y \in \mathbf{C}^n$ sont tels que $Ax = rx$ et $Ay = ry$, alors $Bx = r^k x$ et $By = r^k y$ et il s'ensuit que x et y sont colinéaires. C'est une valeur propre dominante de A: si λ est une valeur propre de A de module r, alors λ^k est une valeur propre de B de module $r^k = s$, et donc $\lambda^k = s$. Comme plus haut, il s'ensuit que $\lambda > 0$ et donc $\lambda = r$.

On suppose, à partir de maintenant que $A > 0$. Par le Lemme 1, on a $Ax > 0$ pour tout $x \geq 0, x \neq 0$. On observera que ceci implique que, si $x \in \mathbf{R}^n$ avec $x \geq 0$ est un vecteur propre A, alors $x > 0$.

On considère le compact $X = \{x \in \mathbf{R}^n \, : \, x \geq 0, \|x\|_1 = 1\}$ de $\mathbf{R}^n$. Pour tout $x \in X$, l'ensemble $\{t \in \mathbf{R}^+ \, : \, tx \leq Ax\}$ est non vide, borné (par $\|A\|$) et fermé. On définit une fonction $\theta : X \to \mathbf{R}^+$ par

$$\theta(x) = \max\{t \in \mathbf{R}^+ \, : \, tx \leq Ax\}, \qquad x \in X.$$

On observera que $\theta(x) > 0$ car $Ax > 0$.

La fonction θ est bornée (par $\|A\|$); soit $r_0 = \sup_{x \in X} \theta(x) \in]0, +\infty[$.

• *2ème etape:* Tout $x \in X$ tel que $\theta(x) = r_0$ est un vecteur propre de A de valeur propre r_0. En effet, supposons, par l'absurde, que $Ax \neq r_0 x$. Par définition de $\theta(x)$, on a $Ax - r_0 x = Ax - \theta(x)x \geq 0$. D'où $A(Ax - r_0 x) > 0$ car $A > 0$. Alors $A(Ax - (r_0 + \varepsilon)x) \geq 0$ pour $\varepsilon > 0$ suffisamment petit. Soit $y = \dfrac{1}{\|Ax\|_1} Ax \in X$. Alors

$$Ay - (r_0 + \varepsilon)y = \frac{1}{\|Ax\|_1} A(Ax - (r_0 + \varepsilon)x) \geq 0$$

et donc $\theta(y) \geq r_0 + \varepsilon$. Ceci contredit la définition de r_0.

• *3ème etape:* Il existe un vecteur $x^+ \in X$ tel $\theta(x^+) = r_0$. En effet, soit $(x^{(k)})_k$ une suite dans X telle que $\lim_k \theta(x^{(k)}) = r_0$. Comme X est compact, $(x^{(k)})_k$ possède une sous-suite convergente, notée de nouveau $(x^{(k)})_k$. Soit $x^+ = \lim_k x^{(k)} \in X$. On a $\theta(x^{(k)})x^{(k)} \leq Ax^{(k)}$, par définition de θ. D'où, par passage à la limite $r_0 x^+ \leq Ax^+$ et donc $r_0 \leq \theta(x^+)$. Il s'ensuit, par définition de r_0, que $\theta(x^+) = r_0$.

• *4ème etape:* Pour toute valeur propre $\lambda \in \mathbf{C}$ de A, on a $|\lambda| \leq r_0$; en particulier r_0 est égal au rayon spectral $r = r(A)$. En effet, soit $v = (v_1, \ldots, v_n)^t \in \mathbf{C}^n$ un vecteur propre pour λ avec $\|v\|_1 = 1$. Alors, pour tout $i \in \{1, \ldots, n\}$, on $\lambda v_i = \sum_{j=1}^n a_{ij} v_j$ et donc

$$|\lambda||v_i| \leq \sum_{j=1}^n a_{ij}|v_j|.$$

Pour le vecteur $|v| = (|v_1|, \ldots, |v_n|)^t \in X$, on a donc $|\lambda||v| \leq A|v|$. Ceci implique que $|\lambda| \leq \theta(|v|)$ et donc $|\lambda| \leq r_0$, par définition de r_0.

• *5ème etape:* r est une valeur propre dominante et simple. En effet, soit $\lambda \in \mathbf{C}$ une valeur propre de A avec $|\lambda| = r$. Soit $v = (v_1, \ldots, v_n)^t \in \mathbf{C}^n$ un vecteur propre pour λ avec $\|v\|_1 = 1$. Il suffit de montrer que v est un multiple du vecteur x^+ introduit plus haut; par la 2ème étape, x^+ est un vecteur propre de valeur propre r.

Comme plus haut, pour $|v| = (|v_1|, \ldots, |v_n|)^t \in X$, on a $r = |\lambda| \leq \theta(|v|)$ et ainsi $\theta(|v|) = r$. Par la 2ème étape, $|v|$ est donc un vecteur propre de valeur propre r. c-à-d, pour tout $i \in \{1, \ldots, n\}$, on a $r|v_i| \leq \sum_{j=1}^n a_{ij}|v_j|$. D'autre part, $Av = \lambda v$ implique que $|\sum_{j=1}^n a_{ij}v_j| = r|v_i|$; comme $|\sum_{j=1}^n a_{ij}v_j| \leq \sum_{j=1}^n a_{ij}|v_j|$, il s'ensuit que

$$|\sum_{j=1}^n a_{ij}v_j| = \sum_{j=1}^n a_{ij}|v_j|, \qquad 1 \leq i \leq n. \tag{$*$}$$

Comme $a_{ij} > 0$, cette égalité, pour $i = 1$ par exemple, ne peut avoir lieu que si les vecteurs v et $|v|$ de $\mathbf{C}^n$ sont colinéaires (écrire $v_j = e^{i\theta_j}|v_j|$ avec $\theta_j \in [0, 2\pi[$, développer $|\sum_{j=1}^n a_{1j}v_j|^2$ et comparer avec $(\sum_{j=1}^n a_{1j}|v_j|)^2$).

Il nous reste à montrer que $|v|$ et $x^+ = (x_1, \ldots, x_n)$ sont colinéaires. Soit $t = \max_{1\leq j\leq n}\{-x_j/|v_j|\}$. Alors, pour le vecteur $y = x^+ + t|v|$, on a $y \geq 0$ et $y_i = 0$ pour un i tel que $t = -x_i/|v_i|$. Comme y est un vecteur propre de A, on a nécessairement $y = 0$ (car, comme y est positif, $y \neq 0$ impliquerait $y > 0$). Donc $|v|$ et x^+ sont colinéaires. ■

2 Matrices stochastiques, chaines de Markov et graphes

Une matrice positive $P = (p_{ij})_{1\leq i,j\leq n}$ est dite *stochastique* si, pour $1 \leq i \leq n$, on a $\sum_{1\leq j\leq n} p_{ij} = 1$.

Un *graphe* (stochastique) fini $\mathcal{G} = (E, P)$ est la donnée d'un ensemble fini $E = \{x_1, \ldots, x_n\}$ de *sommets* et d'une fonction $P : V \times V \to \mathbf{R}^+$ telle que la matrice $(P(x_i, x_j)_{ij})_{1\leq i,j\leq n}$ soit stochastique. Les *arêtes* de $\mathcal{G}$ sont les couples (x_i, x_j) tels que $P(x_i, x_j) > 0$.

Réciproquement, à une matrice stochastique $P = (p_{ij})_{1\leq i,j\leq n} \in M_n(\mathbf{R})$ est associé un graphe $\mathcal{G}(P)$ avec $E = \{1, 2, \ldots, n\}$ comme ensemble de sommets et l'ensemble des couples $(i,j) \in V \times V$ avec $p_{ij} \neq 0$ comme ensemble d'arêtes.

Soit $(X_n)_n$ une chaîne de Markov homogène avec ensemble d'états $E = \{x_1, \ldots, x_n\}$ et matrice de transition $P = (p_{ij})_{1 \leq i,j \leq n}$; on rappelle que ceci signifie que la probabilité de transition $\mathbf{P}(X_{n+1} = x_j | X_n = x_i)$ est égale à p_{ij}.

Soit μ une mesure de probabilité sur E On écrit μ_i pour $\mu(x_i)$. On dit que μ est *invariante* pour P si le vecteur ligne $\mu = (\mu_1, \ldots, \mu_n)$ satisfait à l'équation $\mu = \mu P$, c-à-d si le vecteur colonne μ est un vecteur propre de P^t pour la valeur propre 1. Les mesures invariantes représentent les états d'équilibre du système.

Comme P est stochastique, 1 est valeur propre de P (le vecteur $(1, \ldots, 1)$ est vecteur propre); donc 1 est valeur propre de P^t également. Comme $\|P^t\| = 1$, le rayon spectral de P^t est 1.

Supposons que P est irréductible; remarquons que cette condition équivaut à la connectivité du graphe $\mathcal{G}(P)$ associé à P : étant donés 2 états x_i, x_j, il existe un chemin dans le graphe $\mathcal{G}(P)$ allant de x_i vers x_j. Par le théorème de Perron-Frobenius, il existe alors une unique mesure de probabilité invariante μ sur E telle que $\mu_i > 0$ pour tout i.

Si P est primitive, alors, pour tout mesure de probablité μ_0 sur E (représentant l'état initial), on a $\lim_k \|\mu_0 P^k - \mu\|_1 = 0$ avec vitesse exponentielle; plus précisément, on a $\|\mu_0 P^k - \mu\|_1 = O(\rho^k)$ avec $\rho = |\lambda_1|$, où $\lambda_1 \in \sigma(P)$ est telle que $|\lambda_1| = \max\{|\lambda| : \lambda \in \sigma(P), \lambda \neq 1\}$; voir le document "Sur les matrices stochastiques" par Françoise Guimier (Documents pédagoqiques pour l'agrégation, Compléments d'analyse).

3 L'algorithme PageRank de Google

Le graphe du web est le graphe est le suivant:

- L'ensemble des sommets $E = \{x_1, \ldots, x_N)$ est l'ensemble des N pages web (actuellement, $N \sim 2.10^{10}$ d'après Google)
- Une arête part de x_i vers x_j si la page web x_i possède un lien pointant vers x_j.

L'algorithme PageRank de Google, développé en 1998 par S. Brin et L. Page (le "Page" de PageRank a un double sens!), classe ces pages en leur attribuant des poids pour en mesurer l'importance. L'importance d'une page

x_i est établie en fonction du nombre et de l'importance (sic!) des pages x_j qui pointent vers x_i. Plus précisément, imaginons-nous en train de naviguer ("surfer") au hasard sur le web: ayant choisi une page initiale au hasard, nous navigons vers un des liens disponibles, en le choisissant aléatoirement et de manière équiprobable. Arrivés sur la nouvelle page, nous procédons de la même manière, et ainsi de suite. Ceci définit une chaîne de Markov et l'espoir est qu'à la longue nous convergerons vers un état d'équilibre. La mesure de probabilité invariante correspondante (qui attribut des poids à toutes les pages web) est le classement voulu: le poids d'une page est proportionnel au nombre de fois où cette page a été visitée et ceci détermine son importance.

La *matrice des liens* $L = (l_{ij})_{1\leq,i,j\leq N}$ est définie par

$$l_{ij} = \begin{cases} 1/d_i & \text{si } x_i \text{ pointe vers } x_j \\ 0 & \text{sinon,} \end{cases}$$

où d_i est le nombre de liens sur la page x_i. Malheureusement, L n'est pas stochastique: il y a des pages qui ne pointent vers aucune autre page! Pour pallier ce défaut, on modifie L en définissant $l_{ij} = 1/N$ pour une telle page x_i.

Même ainsi modifiée, la matrice L n'est pas irréductible et, encore moins, primitive. Brin et Page ont introduit la *matrice de Google* pour un paramètre $\alpha \in]0, 1[$:

$$G = \alpha L + (1 - \alpha)E,$$

où $E = (e_{ij})_{1\leq,i,j\leq N}$ est la matrice avec $e_{ij} = 1/N$. La matrice G est stochastique et trivialement primitive (car $G > 0$). Il y a donc une unique mesure de probabilité invariante $\mu = (\mu(x_1), \ldots, \mu(x_N))$ pour G, appelé le *vecteur de Google*, qui est tout simplement le vecteur de Perron-Frobenius de G^t. C'est ce vecteur qui détermine le classement de votre page web.

Bien sûr, l'aspect crucial est le calcul effectif du vecteur μ. Rappelons que G est une matrice avec plusieurs milliards de coefficients. Le calcul de μ est fait au moyen d'une itération $\mu_{n+1} = \mu_n G$ avec une distribution initiale μ_0 quelconque. Comme G est primitive, cette itération converge avec une vitesse de l'ordre de α^n (voir la section précédente).

D'une part, le paramètre artificiel α doit être choisi proche de 1 pour mieux refléter la structure du web; d'autre part, si α est trop proche de 1, la convergence est fortement ralentie. Il semble qu'un bon compromis soit la valeur expérimentale $\alpha = 0,85$ utilisée actuellement par Google.

Bibliographie: Pour la preuve de (ii) du Théorème 3, on pourra consulter les ouvrages “The theory of matrices. Volume 2” de F. R. Gantmacher, Chelsea Publ.Co 1959, ou “Les matrices: théorie et pratique” de D. Serre, Dunod 2001.

Pour plus de détails sur les algorithmes des moteurs de recherche, voir le livre “Google’s PageRank and beyond: the science of search engine rankings” de A.N. Langville et C. D. Meyer , Princeton University Press 2006.