

M1 - Initiation à la Recherche

Applications des matrices aux moteurs de recherche.

Durée : 3h.

Les calculatrices ne sont pas autorisées, ni les communications entre vous. Si vous constatez ce qui vous semble être une erreur d'énoncé, il faut l'indiquer et poursuivre au mieux.

Partie I : Analyse d'activités de Terminale S

Cette partie est liée aux documents fournis pages 2 et 3. On ne demande pas de répondre aux questions de ces activités, mais aux questions suivantes :

I.1. Donner le nom de l'algorithme décrit dans l'activité, le nom de son inventeur, le nom de l'entreprise avec laquelle il travaillait, et l'année (à peu près) de son invention.

I.2. Énoncer et démontrer la formule des probabilités totales.

I.3. Avec nos connaissances du supérieur, on souhaite prendre du recul par rapport à l'activité 1, et donner un peu plus de formalisme à l'énoncé. On introduit l'espace probabilisé (Ω, p) , la suite de variables aléatoires $(Y_n)_{n \in \mathbb{N}}$ qui prend ses valeurs dans l'ensemble $\{1, 2, 3, 4, 5\}$, et dont la valeur prise correspond à la page sur laquelle se trouve l'utilisateur après n clics, comme décrit dans l'activité 1. On définit une suite de matrices colonnes $(Z_n)_{n \in \mathbb{N}}$ de $\mathcal{M}_{5,1}(\mathbb{R})$ de sorte que $\forall n \in \mathbb{N}, \forall i \in \llbracket 1, 5 \rrbracket, (Z_n)_{i,1} = p(Y_n = i)$. Montrer qu'il existe une matrice $B \in \mathcal{M}_{5,5}(\mathbb{R})$, que l'on déterminera, telle que $\forall n \in \mathbb{N}, Z_{n+1} = BZ_n$. On démontrera précisément l'égalité (comme d'habitude bien sûr !).

I.4. A la question 4.b de l'activité 2, un élève répond :

On calcule que $I_4 - 0,85A$ est différent de zéro : on peut alors diviser des deux côtés et on a
$$X = \frac{0,15C}{I_4 - 0,85A}.$$

Quel est le principal problème dans la réponse de l'élève ? Proposer un exemple pour l'aider à comprendre son erreur.

Activités d'exploration

1



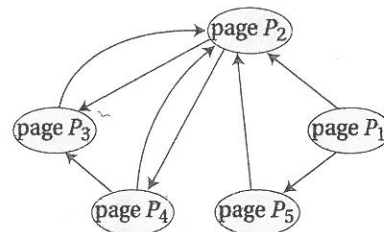
Comment un moteur de recherche classe les pages Web

Réinvestir : Une matrice pour décrire l'évolution d'un processus et les opérations sur les matrices.

Explorer : La convergence d'un processus vers un état stable.

On a représenté ci-contre cinq pages Internet comprenant des articles sur le sujet « les élèves de Terminale S font-ils des études scientifiques ? ».

Sur certaines des pages, un lien pointe vers une autre des pages ; ce lien est matérialisé par la flèche orientée.



- 1 Laquelle de ces pages semble être *a priori* la plus pertinente pour le sujet traité ?
- 2 On suppose qu'un utilisateur qui arrive sur l'une des ces pages suivra de façon équiprobable un des liens de cette page vers les autres pages. Établir la matrice A donnant les probabilités de passage d'une page à une autre par un utilisateur.
- 3 On suppose qu'un utilisateur est situé au départ sur la page P_1 .
 - a. À l'aide de la calculatrice ou d'un logiciel de calcul formel, compléter le tableau qui donne les probabilités d'arrivées successives sur chaque page après 1 clic, 2 clics, 5 clics et 10 clics.
 - b. Quelle est la page qui a la plus grande probabilité d'être fréquentée après 10 clics ? Comparer avec la réponse donnée à la question 1.
 - c. Observer l'évolution de ces probabilités lorsque le nombre n de clics devient de plus en plus grand. Vers quelles probabilités limites de fréquentation de chaque page semble tendre la répartition ?
- 4 Pour un utilisateur situé au départ sur une autre page que la page 1, observer l'évolution des mêmes probabilités. Que constate-t-on ?
- 5 On note X la matrice colonne formée des probabilités limites de fréquentation conjecturées à la question 3 c. Montrer que si la répartition des utilisateurs au départ se fait suivant ces probabilités, alors la répartition reste stable à chaque clic.
 Cette matrice colonne est appelée **état stable du système**.
 Écrire l'égalité matricielle vérifiée par X .
- 6 On définit la pertinence de la page P_i par sa probabilité p_i dans la matrice colonne stable $X = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix}$.
 - a. Quel classement, selon la pertinence, de ces 5 pages peut-on proposer pour le sujet traité ?
 - b. Dédurre de l'égalité matricielle vérifiée par X que $p_2 = \frac{1}{2}p_1 + p_3 + \frac{1}{2}p_4 + p_5$.
 - c. Quelle relation vérifie la pertinence d'une page ?
 - d. De cette manière, retrouver les relations entre les pertinences de chacune des pages pour le sujet traité à partir des données du graphe.

Pages	Après 0 clic	Après 1 clic	Après 2 clics	Après 5 clics	Après 10 clics
P_1	1				
P_2	0				
P_3	0				
P_4	0				
P_5	0				

2 Amélioration du procédé de classement du moteur de recherche

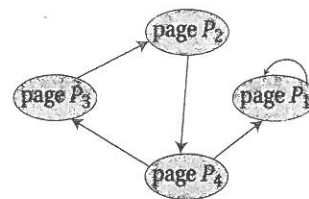
Réinvestir : Une matrice pour décrire l'évolution d'un processus et les opérations sur les matrices.

Explorer : Une condition particulière pour la convergence.

NOTE

Cette situation arrive très fréquemment en réalité puisque sur les milliards de pages Web existantes, les liens pointant d'une page vers une autre sont relativement peu fréquents.

Ce graphe donne les relations entre quatre pages Web.



1 Pour un utilisateur situé au départ page 1, que se passe-t-il ?

2 L'algorithme suivant pallie le problème en modélisant la situation de la façon suivante :

« Pour un utilisateur qui sur un thème donné arrive sur une page Web :

- dans 15 % des cas, il choisit au hasard de façon équirépartie une page quelconque y compris la page en cours ;
- dans 85 % des cas, il suit de façon équirépartie un des liens proposés par la page en cours vers une autre page. »

Pour n entier positif, on note X_n la matrice colonne donnant la répartition des probabilités de fréquentation des quatre pages après n clics.

On note A la matrice de transition donnant les probabilités de passage d'une page à une autre

par les liens représentés sur le graphe ci-dessus et C la matrice colonne égale à

$$\begin{pmatrix} 0,25 \\ 0,25 \\ 0,25 \\ 0,25 \end{pmatrix}.$$

Justifier que la suite de matrices colonnes (X_n) vérifie la relation : $X_{n+1} = 0,85AX_n + 0,15C$.

Un phénomène d'évolution

Dans cette activité, ce **phénomène d'évolution** se modélise par une relation de la forme $X_{n+1} = AX_n + B$ où A est une matrice carrée de taille p , B une matrice colonne de dimension p et (X_n) une suite de vecteurs colonnes de dimension p .

3 On considère toujours un utilisateur situé au départ sur la page 1. À l'aide de la calculatrice ou d'un logiciel de calcul formel, compléter le tableau ci-dessous qui donne les probabilités d'arrivées successives sur chaque page après 1 clic, 2 clics, 5 clics ou 10 clics.

Pages	Après 0 clic	Après 1 clic	Après 2 clics	Après 5 clics	Après 10 clics
P_1	1				
P_2	0				
P_3	0				
P_4	0				

4 On suppose que, quelle que soit la page de départ, la suite des matrices colonnes (X_n) converge vers une matrice colonne X (état stable du système) vérifiant $X = 0,85AX + 0,15C$.

- Montrer que X vérifie l'égalité matricielle : $(I_4 - 0,85A)X = 0,15C$.
- À quelle condition peut-on alors déterminer X de façon unique ?
- Trouver X à l'aide de la calculatrice, puis comparer avec les résultats du tableau.

L'état stable X associé à l'algorithme définit la pertinence des pages associées à une information comme dans l'**activité 1**, il s'ensuit un classement des pages. En pratique, la recherche de l'état stable ne se fait pas par la résolution de l'équation matricielle (les dimensions trop importantes rendent les calculs très longs). On approche l'état stable très rapidement par une suite de matrices colonnes construites à partir de l'algorithme comme dans cette activité.

Partie II : matrices stochastiques.

Dans ce qui suit, n est un entier supérieur ou égal à 1. On note I_n la matrice identité de rang n . On dit qu'une matrice $M \in \mathcal{M}_{n,n}(\mathbb{R})$ est stochastique si $\begin{cases} \forall (i, j) \in \llbracket 1, n \rrbracket^2, & M_{i,j} \geq 0 \\ \forall i \in \llbracket 1, n \rrbracket, & \sum_{j=1}^n M_{i,j} = 1 \end{cases}$. On note $\mathcal{S}_n$ l'ensemble des matrices stochastiques ainsi défini.

Soit p, q, r des éléments de $\mathbb{N}^*$. On rappelle qu'une suite $(A_k)_{k \in \mathbb{N}}$ d'éléments de $\mathcal{M}_{p,q}(\mathbb{R})$ est dite convergente vers $A \in \mathcal{M}_{p,q}(\mathbb{R})$ si chacune des suites des composantes de la suite (A_k) converge vers la composante correspondante dans A . Autrement dit, on dit que (A_k) converge vers A si pour tout $i \in \llbracket 1, p \rrbracket$ et $j \in \llbracket 1, q \rrbracket$ on a $\left((A_k)_{i,j} \right)_{k \in \mathbb{N}}$ qui converge vers $(A)_{i,j}$.

On rappelle que si deux suites $(A_k)_{k \in \mathbb{N}}$ et $(B_k)_{k \in \mathbb{N}}$ d'éléments de $\mathcal{M}_{p,q}(\mathbb{R})$ et $\mathcal{M}_{q,r}(\mathbb{R})$ convergent respectivement vers un élément $A \in \mathcal{M}_{p,q}(\mathbb{R})$ et un élément $B \in \mathcal{M}_{q,r}(\mathbb{R})$, alors la suite des produits matriciels $(A_k B_k)$ converge vers le produit des limites AB .

Enfin, pour tout $n \in \mathbb{N}^*$ et $X \in \mathcal{M}_{n,1}(\mathbb{R})$, on note $\|X\|_\infty = \max\{|X_{i,1}| \mid i \in \llbracket 1, n \rrbracket\}$. Dans tout le problème, on assimilera $\mathcal{M}_{n,1}(\mathbb{R})$ à $\mathbb{R}^n$, et une matrice $A \in \mathcal{M}_{n,n}(\mathbb{R})$ à l'endomorphisme de $\mathbb{R}^n$ canoniquement associé $X \mapsto AX$.

II.1. Pour tout $n \in \mathbb{N}^*$, on considère le vecteur colonne $V_n \in \mathcal{M}_{n,1}(\mathbb{R})$ dont toutes les composantes valent 1. Montrer que $M \in \mathcal{M}_{n,n}(\mathbb{R})$ est stochastique si et seulement si $\begin{cases} \forall (i, j) \in \llbracket 1, n \rrbracket^2, & M_{i,j} \geq 0 \\ MV_n = V_n \end{cases}$.

En déduire une valeur propre commune à toutes les matrices stochastiques.

II.2. En vous ramenant à la définition, montrer que le produit de deux matrices stochastiques est stochastique de deux façons :

II.2.1 en utilisant le résultat d'une question précédente, et

II.2.2 en se ramenant à la définition d'une matrice stochastique, et du produit de deux matrices.

II.3. Pour répondre à la question suivante :

Question : On considère $n \in \mathbb{N}^*$ et $A \in \mathcal{S}_n$. Montrer que pour tout $k \in \mathbb{N}$, la matrice A^k est stochastique.

un étudiant propose :

Réponse : On fait une récurrence : la matrice A^0 qui est l'identité est stochastique donc la proposition est vraie au rang 0. Supposons que la proposition soit vraie au rang n pour tout $n \in \mathbb{N}^*$, donc A^n est stochastique pour tout $n \in \mathbb{N}^*$. On a donc $A \times A^n$ qui est stochastique, car c'est le produit de deux matrices stochastiques. Donc A^{n+1} est stochastique, et la proposition au rang $n+1$ est démontrée.

Analyser rapidement sa réponse et relever les principaux problèmes.

II.4. Montrer qu'une limite de suites de matrices stochastiques est stochastique.

II.5. Soit $n \in \mathbb{N}^*$ et $M \in \mathcal{S}_n$.

II.5.1. Montrer que pour tout $X \in \mathbb{C}^n$ on a $\|MX\|_\infty \leq \|X\|_\infty$.

II.5.2. En déduire que toute valeur propre $\lambda \in \mathbb{C}$ de M vérifie $|\lambda| \leq 1$.

II.6. On considère la matrice stochastique

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}.$$

II.6.1 Calculer les valeurs propres de T et montrer que T est diagonalisable.

II.6.2 Donner une matrice diagonale D et une matrice carrée inversible P , d'ordre 3, telles qu'on ait $T = PDP^{-1}$.

II.6.3 Montrer que la suite $(T^k)_{k \geq 0}$ converge, et calculer sa limite.

II.7. Pour tout $n \in \mathbb{N}^*$ on considère $U_n \in \mathcal{S}_n$ la matrice stochastique dont tous les coefficients sont égaux. A partir du rang de U_n , montrer que U_n est diagonalisable, et en déduire le polynôme caractéristique de U_n .

II.8. Soit $n \in \mathbb{N}^*$ et $A \in \mathcal{S}_n$. On suppose que A est diagonalisable, et que la seule valeur propre de A de module 1 est 1. Montrer que la suite $(A^k)_{k \in \mathbb{N}}$ converge.

II.9. Soit $n \in \mathbb{N}^*$ et $A \in \mathcal{S}_n$. On suppose que la suite $(A^k)_{k \in \mathbb{N}}$ converge vers B .

II.9.1 Montrer que $B^2 = B$.

II.9.2 En déduire que $\mathbb{R}^3 = \text{Ker}(B - I_3) \oplus \text{Ker} B$.

Pour une partie de la démonstration, on pourra s'appuyer sur un raisonnement par analyse et synthèse.

II.9.3 Proposer un exemple d'énoncé d'exercice, sur le thème de son choix, avec une résolution qui s'appuie sur un raisonnement par analyse et synthèse.

L'exercice et sa résolution seront au niveau d'une classe de lycée.

Ê

Partie III : valeurs propres de module 1 d'une matrice stochastique.

Dans cette partie on considère $n \in \mathbb{N}^*$ et $A \in \mathcal{S}_n$ telle que $\forall (i, j) \in \llbracket 1, n \rrbracket^2, A_{i,j} > 0$. On considère également $\lambda \in \mathbb{C}$ une valeur propre de A telle que $|\lambda| = 1$. Soit $X \in \mathbb{C}^n$ un vecteur propre associé à cette valeur propre λ , et on note

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Enfin, soit $k \in \llbracket 1, n \rrbracket$ tel que $\|X\|_\infty = |x_k|$.

III.1 Montrer que pour tout $(z, z') \in \mathbb{C}^2$ on a $|z - z'| \geq ||z| - |z'|||$.

III.2 Soit $(z_i)_{i \in \llbracket 1, n \rrbracket} \in \mathbb{C}^n$. Montrer que $|\sum_{i=1}^n z_i| = \sum_{i=1}^n |z_i|$ si et seulement si il existe

$\theta \in [0, 2\pi[$ et $(\rho_i)_{i \in \llbracket 1, n \rrbracket} \in (\mathbb{R}_+)^n$ tels que $\forall j \in \llbracket 1, n \rrbracket, z_j = \rho_j e^{i\theta}$.

III.3 Montrer les égalités ou inégalités suivantes :

- A. $|x_k| = |\sum_{i=1}^n A_{k,i} x_i|$
- B. $|\sum_{i=1}^n A_{k,i} x_i| \leq \sum_{i=1}^n A_{k,i} |x_i|$
- C. $\sum_{i=1}^n A_{k,i} |x_i| \leq |x_k|$

III.4 Que peut-on en déduire ?

III.5 Montrer que pour tout $(i, j) \in \llbracket 1, n \rrbracket^2$ on a $x_i = x_j$.

III.6 Quel résultat avons-nous montré dans cette partie (relativement à λ et à l'espace propre associé ? (*Plus dur.*) En déduire que sous les conditions de cette partie, la suite $(A^k)_{k \in \mathbb{N}}$ converge vers une matrice dont on précisera le rang