# L'ESGI : l'école d'informatique N°1 en alternance ...

# **Mathématiques Avancées pour le Big Data avec R**

# **Calcul matriciel + séries chronologiques + ACP**

**Réalisé par :** SALHAB Charbel

BENHADDAD Lamia

AMEGAH Komlan Godwin

**Exercice 01 : Cas d’une matrice non diagonalisable**

**A number and smiley face

AI-generated content may be incorrect.**

2. Les valeurs propres de A sont : λ₁ = 2 (multiplicité algébrique 2), λ₂ = 1 (multiplicité 1).

Les vecteurs propres associés à λ = 2 sont colinéaires (identiques), donc on n’a qu’un seul vecteur propre linéairement indépendant pour cette valeur propre.

Il n’est donc pas possible de diagonaliser A car elle ne possède pas une base complète de vecteurs propres.

**Conclusion :** Une matrice n’est diagonalisable que si le nombre de vecteurs propres (linéairement indépendants) = dimension de la matrice, A n’est pas diagonalisable.

3. Matrice de Jordan et matrice de passage :

a. Valeurs propres et vecteurs propres :

- La matrice A a pour valeurs propres λ1=2 tel que m(λ1=2) =2, et λ2=1 tel que m(λ2=1) =1

- Le nombre de vecteurs propres linéairement indépendants associés à λ=2 est inférieur à sa multiplicité algébrique, ce qui signifie que A n’est pas

diagonalisable.

b. Matrice de Jordan JJJ :

- La matrice de Jordan associe à A a une forme presque diagonale qui présente des blocs de Jordan correspondant aux valeurs propres.

- Pour λ=2 le bloc de Jordan est un bloc 2×2 avec un 1 sur la super diagonale.

* Une image contenant croquis, diagramme, conception  Le contenu généré par l’IA peut être incorrect.La matrice de Jordan est donc :

c. Matrice de passage P :

- La matrice P est constituée de colonnes formées par les vecteurs propres et vecteurs généralisés de A.

- Le vecteur généralisé associé à λ=2 est obtenu en résolvant

**(A−2I) v\_gen =vp** où vp est un vecteur propre.

- La matrice P construite permet d’effectuer le changement de base entre la base canonique et la base de Jordan.

d. Vérification :

- On vérifie que **A = P J P −1** en calculant le produit **PJP−1**

- Les résultats numériques montrent une très bonne approximation (écarts liés aux arrondis numériques), confirmant que J est la forme de Jordan de A et P la matrice de passage correspondante.

**Conclusion :**

La matrice A n’est pas diagonalisable car elle possède une valeur propre avec multiplicité algébrique strictement supérieure à la multiplicité géométrique. La

décomposition en forme de Jordan met en évidence la structure de A via des blocs de

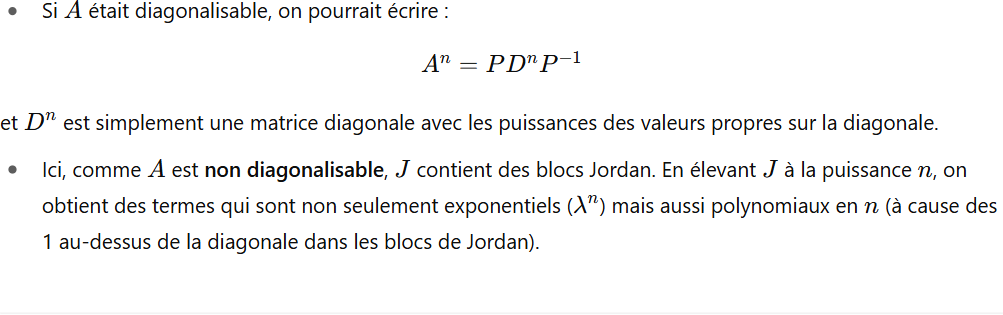
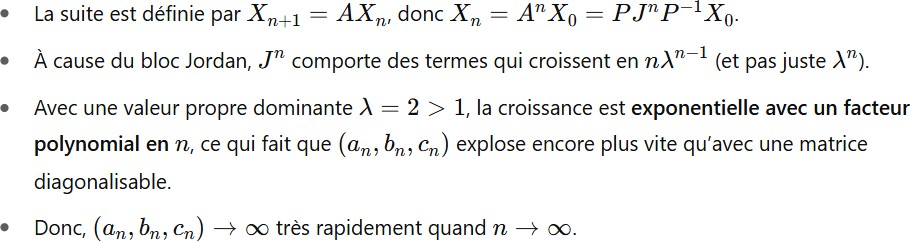
Jordan, et le changement de base associé permet d’étudier les propriétés spectrales de A de façon optimale.

4. Calcul des suites (an,bn,cn) :

En utilisant la décomposition de Jordan A=PJP−, on exprime ce système sous forme matricielle et on calcule explicitement **Xn=(an,bn,cn) T= A^n** . X0 donc

**Xn=P. Jn. P−1.X0**

**Conclusion :**

**Résultats numériques :**

**En résumé**

- Non diagonalisable = présence de blocs Jordan

- Blocs Jordan = puissance n dans l’expression de J^n

- Puissance n combinée à valeur propre >1= croissance très rapide, plus forte qu’une simple exponentielle

- Conclusion : la suite diverge (explose) quand

**Exercice 02 : Etude de cas (1)**

A table with numbers and letters

AI-generated content may be incorrect.

1. Un modèle additif est pertinent ici, car l’amplitude des variations saisonnières semble constante dans le temps.

A graph with lines and dots

AI-generated content may be incorrect.

1. A graph with lines and dots

   AI-generated content may be incorrect.La droite d’ajustement affine :
2. L’analyse des moyennes mobiles et des résidus montre une tendance globale stable avec une forte hausse des ventes au quatrième trimestre chaque année, révélant un effet saisonnier marqué lié aux fêtes, comme en témoignent les écarts positifs significatifs entre les ventes réelles et la tendance à cette période.

A screenshot of a computer

AI-generated content may be incorrect.

1. Les coefficients saisonniers centrés confirment une forte hausse systématique au 4e trimestre (+5.11) et, une fois retirés, la série désaisonnalisée révèle une progression plus régulière des ventes sous-jacente, mettant en évidence une tendance globale croissante.

A screen shot of a computer

AI-generated content may be incorrect.

Malgré une croissance modérée à long terme, la **saisonnalité reste le facteur dominant**, avec des hausses régulières probablement liées à des événements calendaires (comme Noël). La prévision à T=28 reflète bien cette dynamique.

A graph with lines and dots

AI-generated content may be incorrect.

**Exercice 03 : Etude de cas (2)**

A screenshot of a paper

AI-generated content may be incorrect.

1. Un modèle multiplicatif est pertinent ici, car l’amplitude des variations saisonnières augmente avec le niveau de la tendance.

A graph with lines and dots

AI-generated content may be incorrect.

1. La droite d’ajustement affine :

A graph with lines and dots

AI-generated content may be incorrect.

1. Les moyennes mobiles et la tendance révèlent une forte croissance sous-jacente, tandis que les résidus très dispersés (allant de -376.50 à +533.75) mettent en évidence une saisonnalité marquée et irrégulière, justifiant l’usage d’un modèle multiplicatif.

A screenshot of a computer

AI-generated content may be incorrect.

1. Les coefficients saisonniers centrés confirment une très forte saisonnalité positive au troisième trimestre (+442.58) et une négative au premier (-179.84), et leur retrait permet de révéler une série désaisonnalisée cohérente, montrant une croissance régulière des ventes hors effets saisonniers.

A screenshot of a computer

AI-generated content may be incorrect.

1. La prévision pour le troisième trimestre 2025, obtenue en combinant la valeur de tendance au temps t=31t=31 avec le fort coefficient saisonnier positif du T3, s’élève à 1273,1, ce qui s’inscrit dans la continuité d’une dynamique de croissance et d’un pic saisonnier prononcé observé chaque année.

A graph with blue lines and red dots

AI-generated content may be incorrect.

**Exercice 4 : Palindrome**

**Exercice 5 : Salles de cinéma**

Partie 1 : Première Approche

1. Matrice de corrélation :

A screenshot of a computer screen

AI-generated content may be incorrect.

1. La matrice de corrélation montre des liens très forts entre les variables liées à la fréquentation des cinémas : le nombre d’entrées, les recettes, le nombre de séances, de salles, d’établissements et de fauteuils sont tous fortement corrélés (coefficients > 0.9 pour plusieurs d’entre eux). Cela reflète une dynamique logique : plus un département est équipé en cinémas, plus il enregistre d’entrées et de recettes.

La population est également fortement corrélée à ces variables, traduisant un effet de taille : les départements plus peuplés disposent généralement de plus d’infrastructures cinématographiques.

Enfin, les variables multi (multiplexes) et artes (salles art et essai) présentent une corrélation modérée (0.55), ce qui indique qu’elles représentent deux types d’offres cinématographiques distincts.

1. En comparant les données brutes de Paris avec les autres départements, Paris se distingue par des valeurs exceptionnellement élevées sur presque toutes les variables. Avec 30,4 millions d’entrées, 192,2 millions d’euros de recettes, 698 000 séances, 368 salles, 72752 fauteuils et 38 salles d’art et essai, Paris affiche une concentration cinématographique inégalée à l’échelle nationale.

Sa particularité réside aussi dans sa densité urbaine extrême : il n’y a qu’une seule commune, mais elle concentre un nombre très élevé de salles et de structures. Cela fait de Paris un outlier statistique qui peut biaiser les analyses comme l’ACP ou les corrélations globales si on ne le traite pas à part.

Partie 2 : ACP Version 1

1. Les deux tracés du dessus avec le logiciel R :

A graph with lines and arrows

AI-generated content may be incorrect.

A diagram of a number of numbers

AI-generated content may be incorrect.

1. La première composante principale (78,1 % de l’inertie) exprime essentiellement le niveau global d’équipement cinématographique des départements : plus elle est élevée, plus le département est richement doté en salles, fauteuils, séances, etc.

La seconde composante (12,4 %) est portée principalement par la variable comm et distingue les départements selon la dispersion de l’offre sur le territoire (nombre de communes avec au moins un cinéma), indépendamment de l’intensité d’usage.

1. L’analyse du plan principal montre que Paris est un cas exceptionnel : il se situe très à l’écart des autres départements sur l’axe 1 en raison de son offre cinématographique extrêmement dense (plus grand nombre de séances, salles, fauteuils, etc.).

Il se distingue également sur l’axe 2, car il ne compte qu’une seule commune, concentrant toute l’activité cinématographique. Cette configuration unique fait de Paris un outlier qui influence fortement la structure de l’ACP.

1. Diviser les variables par la population permet de neutraliser l’effet de taille des départements et d’obtenir des indicateurs comparables, exprimés en valeur relative (par habitant). Cette approche met davantage en lumière les différences d’intensité d’offre ou de fréquentation, plutôt que de simples volumes, et permet une lecture plus équilibrée des résultats, notamment dans l’analyse en composantes principales.

Partie 3 : ACP Version 2

1. La nouvelle répartition de l’inertie montre que la première composante principale explique 3.71 / 7 ≈ 53 % de l’inertie totale, suivie par la deuxième (2.04 ≈ 29 %). Les deux premières composantes totalisent donc environ 82 % de l'information, ce qui est très satisfaisant.

Cette situation est nettement meilleure que dans l’ACP précédente, où le premier axe était très dominant (>78 %), ce qui masquait d'autres dimensions. Ici, l'ajout de pondérations par population et la normalisation rendent l’analyse plus équilibrée et interprétable.

1. Le premier axe est fortement influencé par les variables etab, salle, faut, sean et comm, avec des corrélations proches de -0.9 ou -1. Ces variables mesurent la densité d'équipement cinématographique par habitant.

Le deuxième axe est surtout influencé par multi (0.66) et artes (0.50), ce qui reflète un gradient entre cinéma commercial (multiplexes) et cinéma d’art et essai.

Ces corrélations indiquent qu’il n’y a pas d’effet direct de la taille (la population est déjà prise en compte via la pondération), mais plutôt des effets liés au type et à l’intensité de l’offre ciné par habitant.

A diagram of a pie chart

AI-generated content may be incorrect.

1. Critère utilisé : ce sont les départements avec des coordonnées extrêmes sur les axes ou une forte qualité de représentation (cos² élevé).

Sur l’axe 1, les départements comme D75 (Paris), D05, D32, D46, D23 ou D28 se distinguent par des valeurs extrêmes → ils contribuent fortement à la construction de cet axe.

Sur l’axe 2, les départements comme D74, D92, D80, ou D95 ont des coordonnées élevées et une bonne représentation → ils déterminent également ce second axe.

Il y a donc des départements sur-représentés dans la construction des axes : en particulier Paris (D75) et d’autres petits départements fortement équipés par habitant.

A diagram with many dots and lines

AI-generated content may be incorrect.

1. À partir des corrélations et des départements influents :

Axe 1 reflète l’intensité d’équipement cinématographique par habitant (plus on va à gauche, plus le département est bien équipé pour sa population).

Axe 2 oppose les départements orientés art et essai (artes) à ceux orientés multiplexes (multi). Il représente donc le type d’offre dominante par habitant.

1. Critère : qualité de représentation (colonne Axi1 :4) < 80 %.

Les départements comme :

* D28 (Eure-et-Loir) : 79.4 %
* D32 (Gers) : 79.9 %
* D80 (Somme) : 79.9 %

Ont une qualité de représentation inférieure à 80 % → ils sont moins bien projetés dans l’espace factoriel formé par les 4 premiers axes. Cela signifie que leur profil est plus complexe ou plus dispersé dans l’espace multidimensionnel.

A graph of a number of individuals

AI-generated content may be incorrect.