Travaux Pratiques Matlab

Analyse en Composantes Principales

Exercice 1 : Un nuage de points presque coplanaires

On considère un nuage de 5 points de $\mathbb{R}^3$ dont les coordonnées sont les suivantes :

$$P_1(2,-1,1), \quad P_2(-1,1,-1), \quad P_3(3,1,0), \quad P_4(-2,-4,1), \quad P_5(-2,3,-1),$$

1) Afficher le nuage de points dans un repère à trois dimensions (commande `plot3`).
   On constate, en faisant tourner le repère, que les points sont presque coplanaires (cf. Fig. 1).

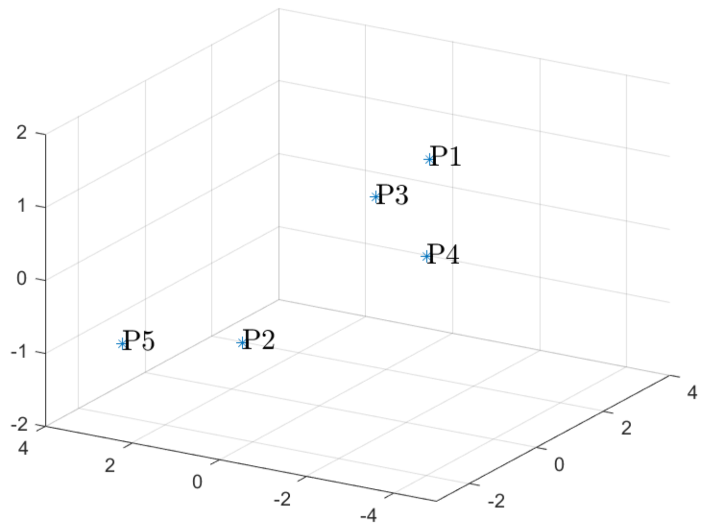

Figure 1 : Cinq points presque coplanaires

2) On souhaite alors déterminer l'équation du plan passant au plus proche des 5 points du nuage. Pour cela on propose d'aborder le problème à l'aide de l'ACP (en reprenant les notations du cours) :
   - Identifier les variables et les individus puis en déduire les valeurs de n et m
   - Construire la matrice des données (on la notera `Y`) et affecter les valeurs à `n` et `m` (commande `size`)
   - Construire la matrice des données centrée (on la notera `X`, utiliser la commande `mean`, pas de boucle for).
     Que constatez-vous ?
   - Construire la matrice de covariance (on la notera `M`)
   - Diagonaliser la matrice de covariance.
   - Placer les valeurs propres dans un tableau, que l'on notera `lambda`, dans l'ordre décroissant puis placer dans une matrice, que l'on notera `P`, les vecteur propres en respectant l'ordre des valeurs propres (commandes `flipud` et `fliplr`).
   - Afficher la courbe des valeurs propres (cf. Fig. 2)

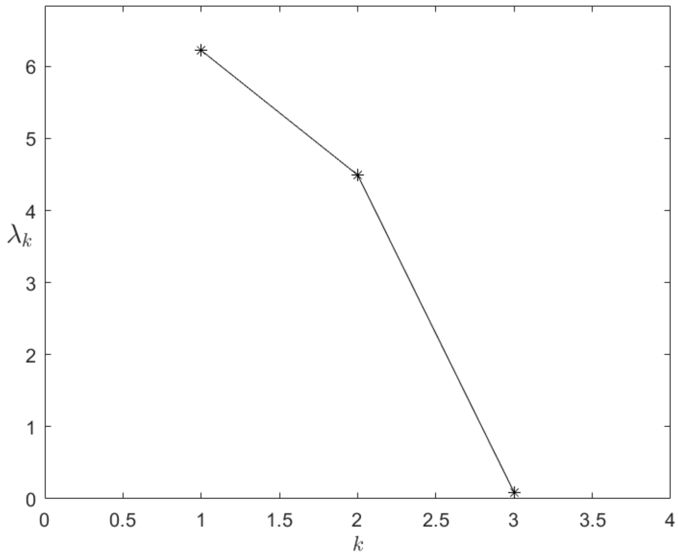

Figure 2 : Valeurs propres de la matrice de covariance

- Calculer le taux d'inertie expliquée pour chaque axe factoriel.
- Afficher la matrice `P`. Que représentent les vecteurs colonnes de cette matrice ?
- Calculer et afficher la matrice des composantes principales (que l'on notera `Xstar`).
- Afficher le nuage de points projeté sur les deux premier axes factoriels (cf. Fig. 3)

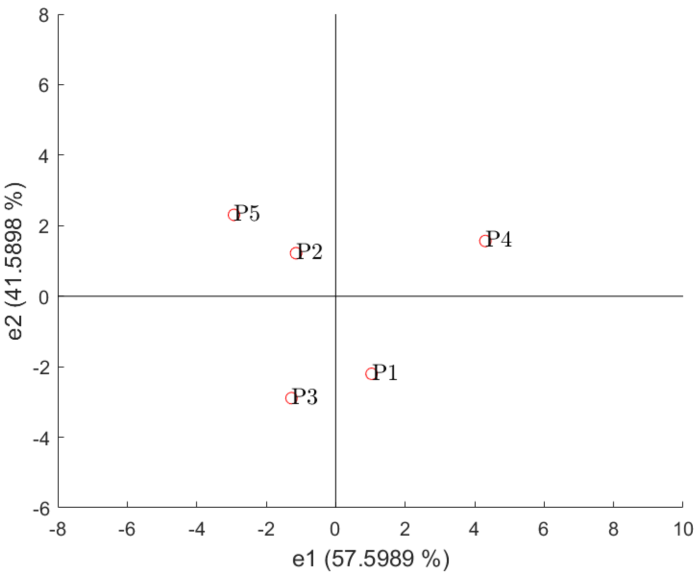

Figure 3 : Projection des 5 points sur le plan factoriel principal

- L'équation du plan défini par les deux premiers axes factoriels $\mathbf{e}_1$ et $\mathbf{e}_2$ a pour équation $ax + by + cz + d = 0$ où a, b et c sont les composantes d'un vecteur normal au plan factoriel $(\mathbf{e}_1, \mathbf{e}_2)$.
  - Quelle est la valeur de d ?
  - Calculer à l'aide de Matlab les valeurs de a, b et c
  - Afficher (cf. Fig. 4), dans le repère 3D de la question 1, le plan d'équation $ax + by + cz + d = 0$ (commandes `meshgrid` et `surf`)

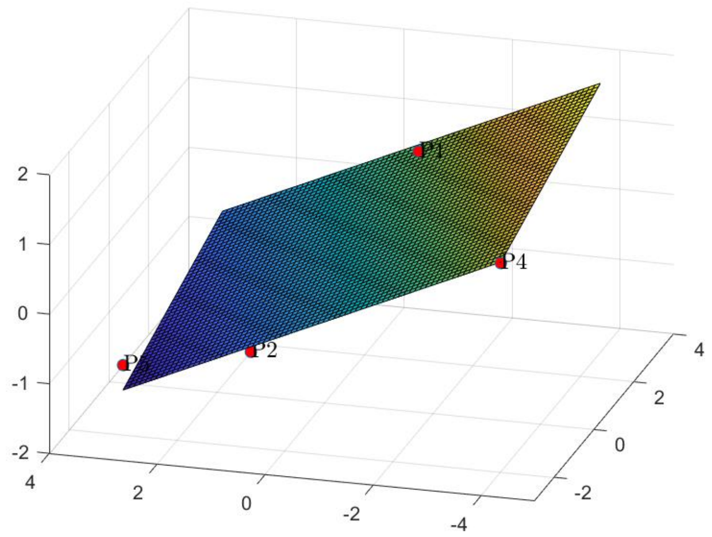

Figure 4 : Plan passant au plus près des 5 points

3) Exécuter à nouveau votre programme en remplacer le nuage de points précédents par les points suivants et vérifier si tout fonctionne correctement :

$P_1(2,1,3)$, $P_2(-2,5,-3)$, $P_3(3,-1,0)$, $P_4(5,-1,3)$, $P_5(-2,3,-1)$, $P_6(1,2,-4)$, $P_7(-4,-2,1)$

Exercice 2 : Habitudes alimentaires

Utiliser l'analyse en composante principale des données présenter dans le tableau ci-dessous afin de voir s'il existe un moyen de distinguer les pays selon leurs habitudes alimentaires.

	RFA	**France**	**Italie**	**Pays Bas**	**Belg. Lux.**	**Russie**	**Irlande**	**Danemark**	**Grèce**
Céréales	68	76	127	60	72	68	84	68	82
Riz	2	4	3	3	2	3	2	2	4
Pommes de terre	81	74	41	81	101	102	111	68	70
Sucre blanc	36	34	31	37	35	40	42	44	24
Légumes	64	115	172	82	60	76	83	57	237
Fruits	89	53	69	84	64	33	30	38	57
Vin	25	93	87	13	66	6	3	14	45
Huiles végétales	6	12	20	4	4	8	8	7	22
Margarine	7	3	1	12	10	6	4	12	1
Viande bov.	24	33	25	21	28	23	23	13	20
Viande porc.	58	38	24	41	41	26	32	51	18
Volailles	84	90	80	136	87	14	14	9	13
Lait et dériv.	85	91	84	135	89	134	187	159	64
Beurre	6	7	2	3	8	5	11	6	1
Fromages	12	18	13	13	10	6	3	10	8
Œufs	17	15	11	11	14	14	14	14	12

Quelques indications :

- Récupérer le fichier `Ex2_data.m` sur Teams
- S'inspirer fortement du travail effectué pour l'exercice 1, notamment pour les notations (qui sont celles du cours), les affichages, etc ...
- Mettre en évidence le taux d'inertie expliquée pour chaque axe factoriel.
- Ajouter des graphiques montrant la projection du nuage sur le plan factoriel principal $(\mathbf{e}_1, \mathbf{e}_2)$ mais également sur d'autres plan plus secondaires, et interpréter les résultats.
- Afficher le cercle des corrélations (cf. cours) non seulement sur le plan factoriel principal $(\mathbf{e}_1, \mathbf{e}_2)$ mais également sur d'autres plans plus secondaires, puis interpréter les résultats.

Exercice 3 : ACP et compression de données

On considère les signaux s_1 et s_2 définis sur $[0, 2\pi]$ par :

$$\begin{cases} s_1(t) = \sin(t) + \delta_1(t) \\ s_2(t) = \varepsilon \sin(t + \varphi) + \delta_2(t) \end{cases} \quad \text{avec } \varepsilon = \pm 1 \quad \text{et} \quad \varphi \in [0, \pi]$$

avec $\varepsilon = \pm 1$, $\varphi \in [0, \pi]$ et $\delta_1(t)$ et $\delta_2(t)$ des perturbations aléatoires d'amplitude $\delta_0 = 0.2$.

1) On suppose dans une premier temps que $\varepsilon = 1$ et $\varphi = 0$.

   a. Afficher, dans une même figure mais sur 2 graphiques différents (commandes `plot` et `subplot`), les deux signaux ainsi que le nuage de points (s_1, s_2) [voir Fig. 1] et donner une interprétation du nuage de points.

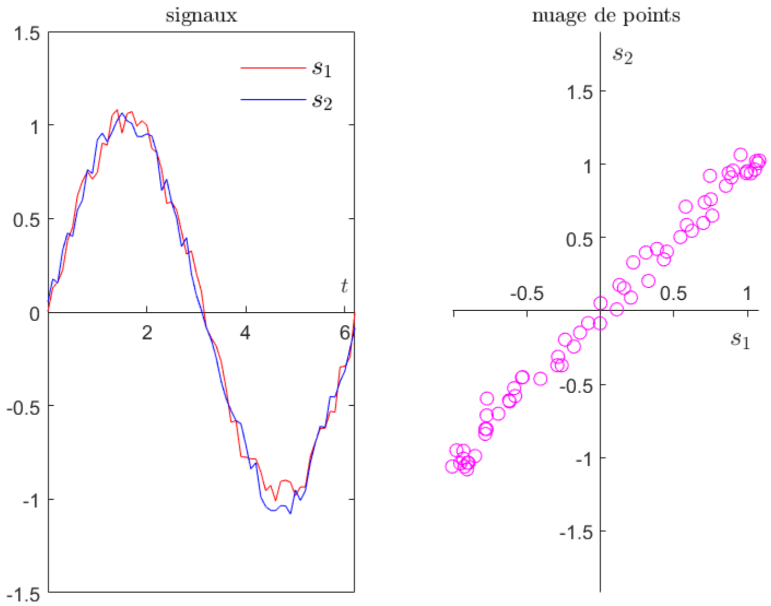

Figure 1 : Signaux s_1 et s_2 et nuage de points (s_1, s_2)

   b. Pour confirmer l'interprétation précédente, faire une analyse en composantes principales du nuage de points (s_1, s_2) et afficher la projection du nuage sur le plan factoriel principal $(\mathbf{e}_1, \mathbf{e}_2)$ en précisant le taux d'inertie expliquée des axes [voir Fig. 2]. Afficher également le cercle des corrélations [voir Fig. 3] et interpréter le résultat.

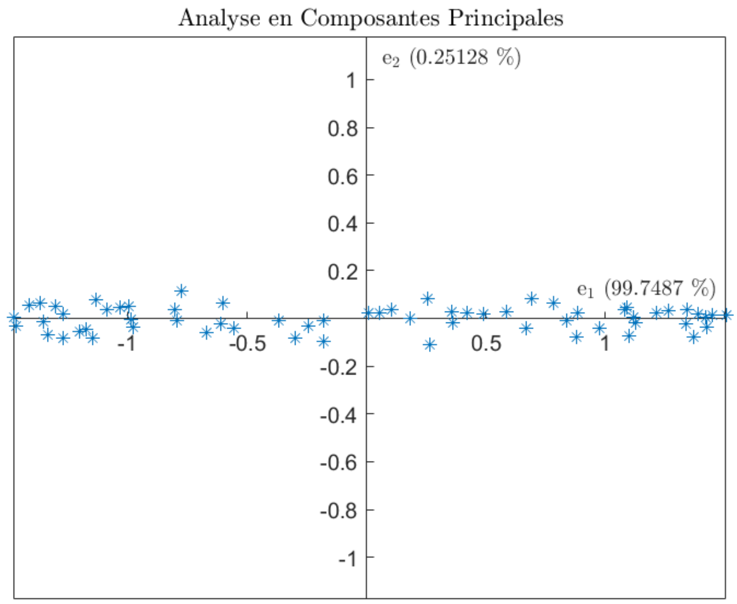

Figure 2 : Projection du nuage de points sur l'axe factoriel $(\mathbf{e}_1, \mathbf{e}_2)$

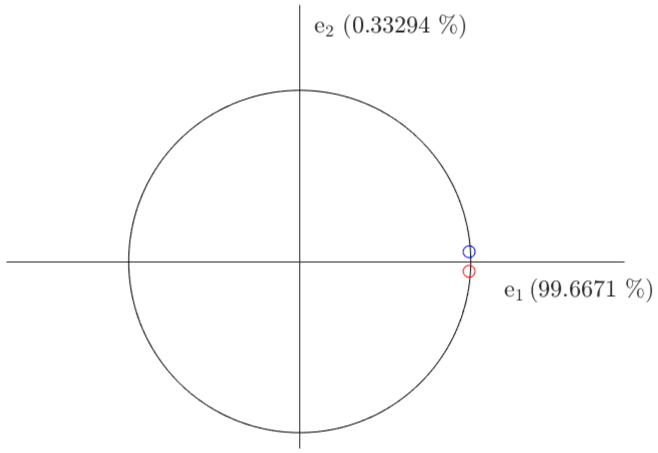

Figure 3 : Cercle des corrélations

c. Reconstruire le signal en ne considérant que la 1ère composante principale et afficher les signaux reconstruits $s_{1,\text{rec}}$ et $s_{2,\text{rec}}$ en superposition des signaux d'origine [voir Fig. 4].

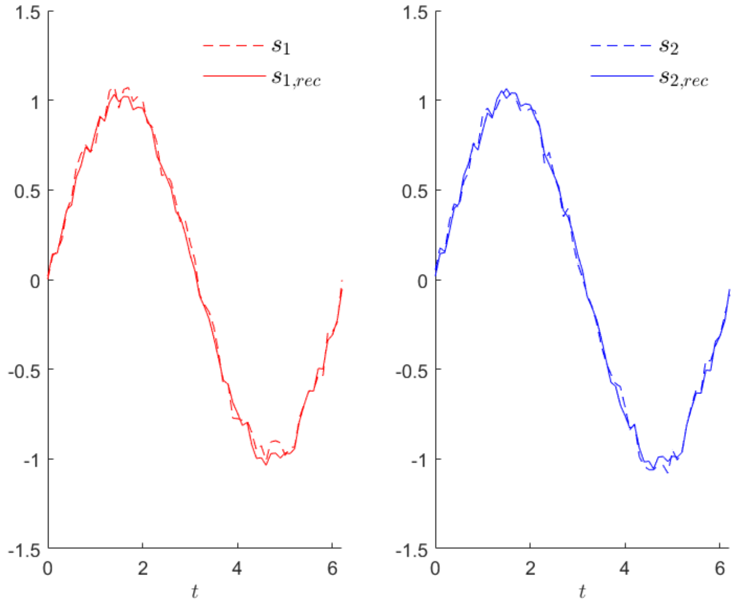

Figure 4 : Reconstruction des signaux à l'aide de la 1ère composante principale

d. Reprendre les questions *a*) *b*) et *c*) en prenant les valeurs $\varphi = \pi/2$ puis $\varphi = \pi$, puis prendre $\varepsilon = -1$ et exécuter à nouveau le programme avec les valeurs $\varphi = 0$, $\varphi = \pi/2$ et $\varphi = \pi$.

2) On souhaite refaire l'analyse précédente même cette fois-ci sur 20 signaux dont les données sont dans le fichier `signaux.mat` disponible sur Teams. Copier ce fichier dans le même répertoire que le programme et afficher les 20 signaux à l'aides des quelques lignes de codes ci-dessous :

```
% chargement des signaux
load signaux;
Y=D;
[n,m]=size(Y); % n=768 abscisses, m=20 signaux

% affichage des 20 signaux
figure(1);
for i=1:m
    subplot(m,1,i);
    title('Signaux originaux','interpreter','latex');
    plot(Y(:,i));
    axis off;
end
```

Ajouter une figure contenant la courbe des valeurs propres de la matrice de covariance ainsi que la somme cumulée des taux d'inertie [voir Fig. 5].

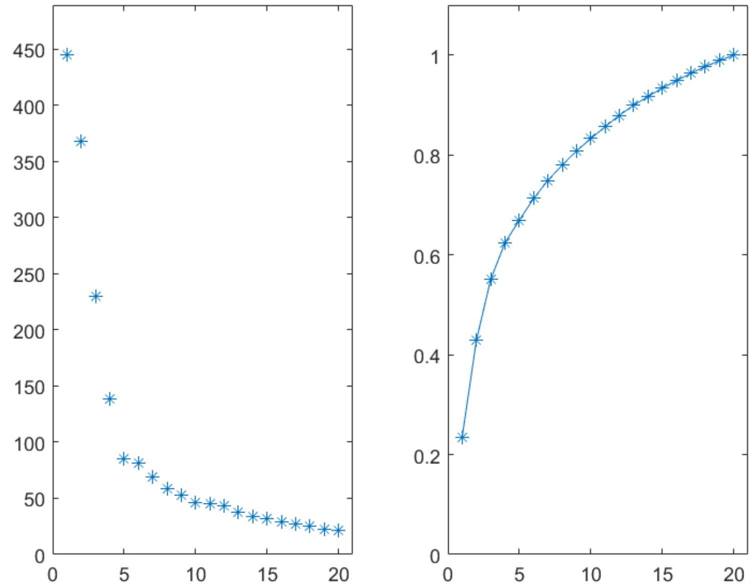

Figure 5 : Valeurs propres de la matrice de covariance (à gauche) et somme cumulée des taux d'inertie

Exercice 4 : Transformée de Karhunen-Loève (TKL)

La transformée de Karhunen-Loève est le nom que l'on donne à l'analyse en composantes principales lorsque cette dernière est appliquée au traitement des images.

1) Une image couleur est un ensemble de 3 images obtenues à l'aide 3 filtres (canaux) de 3 couleurs différentes : rouge, vert et bleu. On souhaite construire, à partir d'une image couleur, une image unique à niveaux de gris dans laquelle figure un maximum d'information.
   - Récupérer l'image `ballon.tif` sur Teams (ou toute autre image couleur)
   - Construire la matrice des données (que l'on notera `Y`) comportant 3 colonnes (une par canal) et n lignes où n est le nombre de pixels d'une image ($n = H \times W$)
   - Faire les calculs de l'ACP (matrice des données centrée `X`, matrice de covariance `M`, calcul et affichage des valeurs propres dans l'ordre décroissant, matrice de projection `P`, matrice des composantes principales `Xstar`)
   - Calculer l'inertie expliquée par le premier axe factoriel
   - Construire la matrice (que l'on notera `Irec`) de dimension $H \times W$ à partir de la première composante principale (commande reshape)
   - Afficher l'image correspondant à la matrice `Irec` et commenter.

2) (d'après Stéphane Grousson) On peut généraliser la méthode précédente à des images multi-spectrales, comme par exemple ici, 6 images satellitaires d'une même vue mais acquises à travers des canaux fréquentiels différents (voir Fig. 6). Construire une seule image en conservant le maximum d'informations (récupérer sur Teams les fichiers des 6 images `i1.jpg`, …, `i6.jpg`).

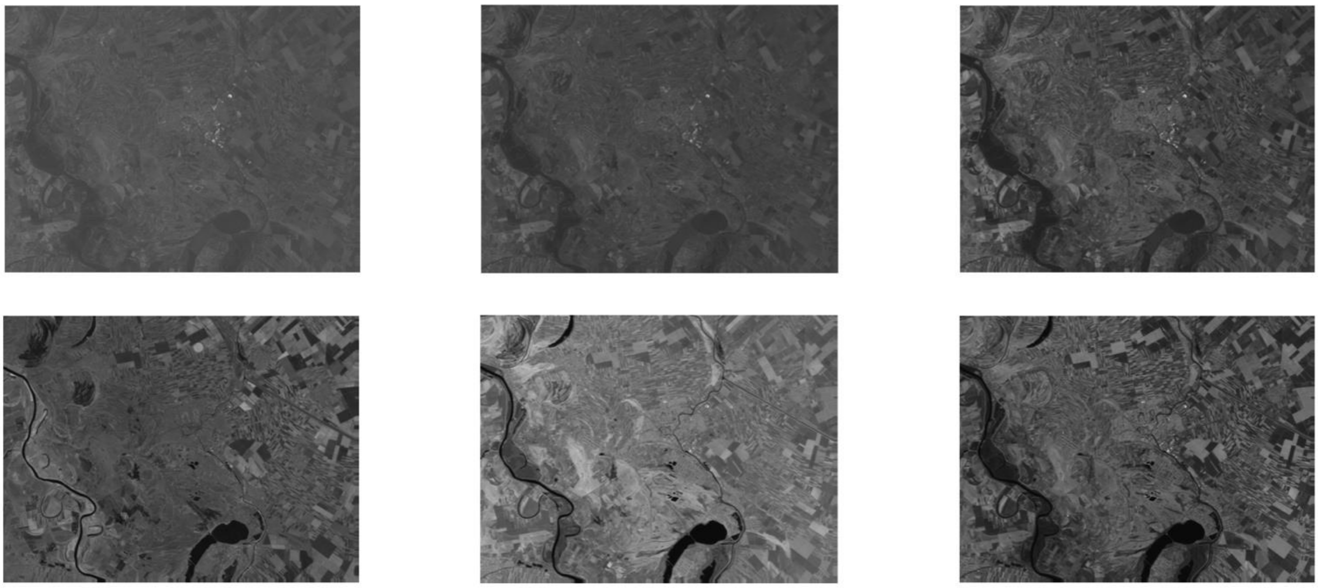

Figure 6 : Images satellitaires d'une même vue