TD : Analyse en composantes principales

Echauffements.

- Retrouver la formule du centre de masse d'un nuage de points S en cherchant le point qui minimise la somme des carrés des distances aux points de S.
- Quelle est la propriété du centre de masse d'un polygone convexe ?
- Comme est défini le centre de masse d'un triangle ? en option : retrouver la formule à partir d'une intégrale double sur le triangle canonique (coordonnées (0; 0), (1; 0) et (0; 1)).
- Comme est défini le centre de masse d'un polygone convexe ?
- Comme est défini le centre de masse d'un polygone non-convexe ?

Interprétation géométrique. Lors d'une analyse en composante principale d'un nuage de points on cherche à minimiser la somme des carrés des distances entre les points et leur projection orthogonale sur un sous-espace linéaire. Expliquer *avec un argument géométrique* pourquoi ceci revient à maximiser l'inertie du nuage projeté (démarrer avec un exemple simple : nuage en 2D, et droite comme sous-espace linéaire, l'inertie équivaut à la variance).

Matrices diagonalisables. Une matrice *symétrique* réelle possède une base de vecteurs propres orthogonaux réels et ses valeurs propres sont elles aussi réelles. Une matrice symétrique définie positive est une matrice symétrique dont les valeurs propres sont strictement positives.

Supposons un nuage de points S en 3D, et la matrice de covariance C de S. Quelle est la dimension de C ? Supposons que l'on connaisse les 3 valeurs de propres de S.

- Dans quel cas les trois valeurs propres sont nulles ? qu'advient-il des valeurs propres et vecteurs propres de la matrice ? (attention : cas particulier).
- Deux valeurs propres sont nulles : qu'en déduit-on des composantes principales et de l'inertie du nuage projeté ? qu'advient-il si on cherche la droite ou le plan qui approchent au mieux le nuage de points ? comment construire le plan ou la droite en fonction des vecteurs propres ?
- Mêmes questions lorsqu'une valeur propre est nulle. Quelle est l'inertie le long de la normale au meilleur plan ?
- Mêmes questions lorsque les trois valeurs propres sont égales.
- Mêmes questions lorsque les deux plus grandes valeurs propres sont égales.
- Mêmes questions lorsque les deux plus petites valeurs propres sont égales.

- A partir des valeurs propres, comment définir un indice de *fiabilité* du meilleur plan, ou de la meilleure droite ?

Matrice de covariance de deux points. Supposons un nuage de points en 2D formé de deux points aux coordonnées (-1 ; 0) et (1 ; 0). En déduire la matrice de covariance et les composantes principales.

Covariance d'un rectangle. Supposons un domaine Ω défini par une forme de rectangle centré à l'origine, de largeur 4 et hauteur 2. Calculer en forme close (via une intégrale double) la matrice de covariance de Ω. En déduire les deux directions des composantes principales.

Covariance d'un carré. Supposons un domaine Ω défini par une forme de carré (côté de longueur 2) centré à l'origine. Quel est le centre de masse de Ω ? Calculer en forme close (via une intégrale double) la matrice de covariance de Ω. En déduire les composantes principales. Qu'en déduisez-vous ?

Covariance du triangle canonique. Supposons un domaine Ω défini par le triangle canonique de coordonnées (0; 0), (1; 0) et (0; 1).

- Calculer en forme close la matrice des moments par rapport à l'origine.
- Généraliser cette formule pour obtenir la matrice de covariance d'un polygone convexe.

Réduction de dimension.

- A partir de l'analyse en composantes principales**,** décrire un algorithme pour réduire la dimension d'un nuage de points.
- Quels calculs sont mis en jeu ?
- Comment choisir automatiquement la nouvelle dimension ?