

## Visualisation de données par ACP régularisée

L'ACP est souvent utilisée pour explorer et visualiser des données. Un modèle classique en ACP est le modèle à effets fixes de Caussinus (1986) qui suppose que les données sont générées comme une structure fixe ayant une représentation en rang inférieur entachée d'erreur.

Sous ce modèle, l'ACP ne fournit pas la meilleure estimation du signal en termes d'erreur quadratique moyenne. En suivant le même principe qu'en régression ridge, nous proposons une version régularisée de l'ACP qui permet de mieux retrouver la structure sous-jacente et ainsi d'obtenir des représentations graphiques plus proches de celles qui auraient été obtenues à partir du signal seulement. L'ACP régularisée revient à seuiller chaque valeur singulière par un terme qui correspond au ratio de la variance du signal sur la variance totale de la dimension associée. Nous montrerons aussi que ce terme peut s'obtenir par une présentation Bayésienne du modèle à effets fixes.

Les performances de l'ACP régularisée seront illustrées sur des simulations et sur un vrai jeu de données provenant du domaine de la génétique. Nous comparerons l'ACP régularisée à l'ACP classique ainsi qu'à une méthode dite de seuillage doux des valeurs singulières. Enfin nous montrerons aussi qu'il peut être intéressant de réaliser l'ACP régularisée avant d'utiliser une méthode de classification.