Licence d'Informatique 3ème année
Analyse de Données Programmeur (C5-160512)
ACP – Solution Mathématique

Carl FRÉLICOT – Dpt Info / Lab MIA

Le problème de l'Analyse en Composantes Principales est un problème d'optimisation sous constrainte(s). Il s'agit de trouver pour la 1ère composante (puis la 2ème, etc), le vecteur directeur u_1 unitaire (puis u_2, etc) qui la définit, le critère à optimiser étant : la variance projetée qu'il faut maximiser.

On utilise pour cela la *méthode de Lagrange* qui permet [1], par substitution, d'optimiser une fonction à valeurs dans $\mathbb{R}$ de plusieurs variables[2] sous contrainte(s) d'égalité c.
Par exemple, le problème : $\max_u\{g(u)|c(u)=0\}$ est substitué par $\max_{u,\lambda}\{L(u,\lambda)=g(u)-\lambda\, c(u)\}$ où $\lambda\in\mathbb{R}$ est appelé *multiplicateur de Lagrange*. La solution est alors tout à fait standard[3] :

$$\begin{cases}\frac{d}{du}L(u,\lambda)=0\\ \frac{d}{d\lambda}L(u,\lambda)=0=c(u)\end{cases}$$

où $\frac{d}{dx}$ représente la dérivée par rapport à x. Si x est un vecteur, sa dérivée (vectorielle) est le vecteur des dérivées par rapport à chacune de ses composantes ; on l'appelle alors *gradient* et on le note parfois ∇. Dans notre cas, $\frac{d}{du}L=[\frac{d}{du_i}L]_{i=1,p}$. On montre[4] que :

- si a est un vecteur constant de $\mathbb{R}^p$ et $u\in\mathbb{R}^p$, alors $\frac{d}{d\,u}\,({}^t a\, u)=a$
- si u et $v\in\mathbb{R}^p$, et M une matrice carrée d'ordre p, alors $\frac{d}{d\,u}\,({}^t u\, M\, v)=M\, v$ et $\frac{d}{d\,u}\,({}^t v\, M\, u)={}^t M\, v$
- en conséquence, pour une forme quadratique $\frac{d}{d\,u}\,({}^t u\, M\, u)=M\, u+{}^t M\, u$,
- et si de plus M est symétrique, il vient $\frac{d}{d\,u}\,({}^t u\, M\, u)=2\, M\, u$.

Notez que vous connaissez ces formules en dimension $p=1$, c'est-à-dire lorsque u est un scalaire, et $M=m$ également. Par ex. la dernière devient $(m\, u^2)'=2\, m\, u$.

Pour l'ACP, on doit tout d'abord trouver u_1 qui maximise $g(u_1)={}^t u_1\, V\, u_1$, à valeurs dans $\mathbb{R}$, où V est la matrice de covariance[5], sous la contrainte $||u_1||=1$ que l'on écrit $c(u_1)=||u||-1=||u||^2-1={}^t u_1\, u_1-1=0$. On substitue donc par la recherche de u_1 qui maximise $L(u_1,\lambda_1)={}^t u_1\, V\, u_1-\lambda_1({}^t u_1\, u_1-1)$. La solution est alors :

$$\begin{cases}\frac{d}{du_1}L(u_1,\lambda_1)=0\\ \frac{d}{d\lambda_1}L(u_1,\lambda_1)=0\end{cases}\Leftrightarrow\begin{cases}2\, V\, u_1-2\,\lambda_1\, u_1=0\\ {}^t u_1\, u_1-1=0\end{cases}\Leftrightarrow\begin{cases}V\, u_1=\lambda_1\, u_1\\ {}^t u_1\, u_1=1\end{cases}$$

Dans la 2ème équation on retrouve la contrainte (u unitaire), et la 1ère indique que u_1 est vecteur propre[6] de V associé à la (plus grande, voir mon CM) valeur propre λ_1.

On doit ensuite trouver u_2 qui maximise $g(u_2)={}^t u_2\, V\, u_2$ sous les contraintes $||u_2||=1$ et $u_2\perp u_1\Leftrightarrow u_2\, V\, u_1=0$. Il suffit d'introduire deux multiplicateurs de Lagrange (un par contrainte) pour substituer le problème par la recherche de u_2 qui maximise $L(u_2,\lambda_1,\lambda_2)={}^t u_2\, V\, u_2-\lambda_2({}^t u_2\, u_2-1)-\lambda_1({}^t u_2\, V\, u_1)$. La solution est alors :

$$\begin{cases}\frac{d}{du_2}L(u_2,\lambda_1,\lambda_2)=0\\ \frac{d}{d\lambda_1}L(u_1,\lambda_1,\lambda_2)=0\\ \frac{d}{d\lambda_2}L(u_2,\lambda_1,\lambda_2)=0\end{cases}\Leftrightarrow\begin{cases}2\, V\, u_2-2\lambda_2\, u_2-\lambda_1\, V\, u_1=0\\ {}^t u_2\, u_2-1=0\\ {}^t u_2\, V\, u_1=0={}^t u_1\, V\, u_2\end{cases}$$

Comme u_1 est vecteur propre de V, la 3ème équation peut s'écrire : ${}^t u_2\,(\lambda\, u_1)=0\Leftrightarrow\lambda\,{}^t u_2\, u_1=0\Rightarrow{}^t u_2\, u_1=0$.

Pré-multiplions la 1ère par ${}^t u_1$: ${}^t u_1(2\, V\, u_2-2\lambda_2\, u_2-\lambda_1\, V\, u_1)=0\Leftrightarrow 2\,{}^t u_1\, V\, u_2-2\lambda_2\,{}^t u_1\, u_2-\lambda_1\,{}^t u_1\, V\, u_1=0$.
Le deux premiers termes étant nuls, il faut que $\lambda_1=0$.
Ce qui donne dans la 1ère équation : $2\, V\, u_2-2\lambda_2\, u_2=0\Leftrightarrow V\, u_2=\lambda_2\, u_2$, c'est-à-dire que u_2 est aussi vecteur propre de V, associé à la valeur propre λ_2. Comme u_2 doit maximiser ${}^t u_2 V u_2=\lambda_2\,{}^t u_2 u_2=\lambda_2$, il s'agit de prendre le vecteur propre associé à la 2ème plus grande valeur propre de V.

Et ainsi de suite jusqu'à la p-ième composante dont le vecteur directeur u_p est vecteur propre de V associé à la p-ième plus grande valeur propre, c'est-à-dire la plus petite. Les composantes principales d'un nuage X sont ainsi définies par les vecteurs propres de sa matrice de covariance, dans l'ordre décroissant des valeurs propres.

---

[1] sous conditions non développées ici, en particulier : dérivabilité par rapport à toutes les variables
[2] un vecteur $u\in\mathbb{R}^p$ est un objet constitué de p variables
[3] faites appel à vos souvenirs sur la recherche d'extrema d'une fonction $f(x)$
[4] dans un cours sur les dérivées vectorielles des formes quadratiques
[5] symétrique par définition
[6] voir un cours sur les éléments propres d'une matrice