

Remarque (2) Si Σ_{xx} inversible, alors

$$\Sigma_{yx} \Sigma_{xx}^{-1} \in \operatorname{argmin}_D F(D)$$

Pour $F(D) = \sum_{i=1}^L \|y_i - D x_i\|^2$

preuve: On note $f_i(D) = \|y_i - D x_i\|^2$

On a $\nabla f_i(D) = 2(y_i - D x_i) x_i^T$ $(f_i(D+D')) = \|y_i - D x_i - D' x_i\|^2$
 $= f_i(D) + 2\langle y_i - D x_i, D' x_i \rangle + o(\|D'\|)$

Donc $\sum_i \nabla f_i(D) = 0 \Leftrightarrow \sum_{i=1}^L y_i x_i^T - D x_i x_i^T = 0$

$$\Leftrightarrow D \Sigma_{xx} = \Sigma_{yx}$$

$$\Leftrightarrow D = \Sigma_{yx} \Sigma_{xx}^{-1}$$

De plus, F est quadratique et positive donc tout point critique est un minimiseur. □

Remarque (3) La projection ^{de x} sur l'e.v. généré par les colonnes de Π vaut $P_\Pi(x) = \Pi (\Pi^T \Pi)^{-1} \Pi^T x$

preuve: On a $P_\Pi(x) = \Pi \operatorname{argmin}_{y \in \mathbb{R}^T} \|x - \Pi y\|^2$ $= \operatorname{argmin}_{y \in \mathbb{R}^T} \|x - \Pi y\|^2$

or on note $f(y) = \|x - \Pi y\|^2$, comme

f est quadratique et définie positive

$$\nabla f(y) = 2 \Pi^T (\Pi y - x)$$

$$\text{et } \nabla^2 f(y) = 2 \Pi^T \Pi$$

$$\text{car } f(y + y') = \|\Pi y - x + \Pi y'\|^2 = f(y) + 2 \langle \Pi^T (\Pi y - x), y' \rangle + \frac{1}{2} \langle 2 \Pi^T \Pi y', y' \rangle$$

On obtient que f a un Hessian défini positif. Son unique minimiseur correspond à son point critique vérifiant

$$\Pi^T \Pi y = \Pi^T x \quad \text{i.e.} \quad y = (\Pi^T \Pi)^{-1} \Pi^T x$$

□ Rabdi, Hammik Paragraphe lin.

Lemme 1: Si Σ_{xx} inversible et $\text{rg}(A) = n$,

(A, B) pt-critique de E

$$AB \Sigma_{xx} B^T = \Sigma_{yx} B^T \quad \text{et} \quad B = (A^T A)^{-1} A^T \Sigma_{yx} \Sigma_{xx}^{-1}$$

preuve:

On a

$$\begin{aligned} E(A, B) &= \sum_{i=1}^L \|y_i - ABx_i - AB'x_i\|^2 \\ &= E(A, B) - 2 \sum_{i=1}^L \langle y_i - ABx_i, AB'x_i \rangle + o(\|B'\|) \\ &= E(A, B) + \left\langle 2 \sum_{i=1}^L -A^T (y_i - ABx_i) x_i^T, B' \right\rangle + o(\|B'\|) \end{aligned}$$

Donc $\frac{\partial E}{\partial B}(A, B) = 2(A^T AB \Sigma_{xx} - A^T \Sigma_{yx})$

$$\begin{aligned} E(A + A', B) &= \sum_{i=1}^L \|y_i - ABx_i - (A + A')x_i\|^2 \\ &= E(A, B) - 2 \sum_{i=1}^L \langle y_i - ABx_i, A'x_i \rangle + o(\|A'\|) \\ &= E(A, B) + 2 \sum_{i=1}^L \langle (ABx_i - y_i) x_i^T B^T, A' \rangle + o(\|A'\|) \end{aligned}$$

Donc $\frac{\partial E}{\partial A}(A, B) = 2(AB \Sigma_{xx} B^T - \Sigma_{yx} B^T)$

Finalement, pour tout A, B

$$\begin{cases} \frac{\partial E}{\partial A}(A, B) = 0 \\ \frac{\partial E}{\partial B}(A, B) = 0 \end{cases} \quad (\Rightarrow) \quad \begin{cases} AB \Sigma_{xx} B^T = \Sigma_{yx} B^T \\ A^T A B \Sigma_{xx} = A^T \Sigma_{yx} \end{cases}$$

Comme $A^T A$ et Σ_{xx} sont inversibles,

$$\Leftrightarrow \begin{cases} AB \Sigma_{xx} B^T = \Sigma_{yx} B^T \\ B = (A^T A)^{-1} A^T \Sigma_{yx} \Sigma_{xx}^{-1} \end{cases}$$

$$A = \begin{pmatrix} \vdots \\ \vdots \end{pmatrix}_{n \times 2}$$



Lemme 2: $\bar{\Sigma}_{xx}$ inversible et $\text{rg}(A) = n$, alors

i) (A, B) pt critique du 1^{er} ordre ss.

$$\text{ii)} \begin{cases} AB = P_A \bar{\Sigma}_{yx} \bar{\Sigma}_{xx}^{-1} \\ P_A \bar{\Sigma} = P_A \bar{\Sigma} P_A = \bar{\Sigma} P_A \end{cases}$$

preuve:

* Prop ii) $\Rightarrow$ i) : a) En utilisant le lemme 1 et l'item 3 de la remarque, on a pour tout point-critique

$$AB = A(A^T A)^{-1} A^T \bar{\Sigma}_{yx} \bar{\Sigma}_{xx}^{-1} = P_A \bar{\Sigma}_{yx} \bar{\Sigma}_{xx}^{-1}$$

b) En utilisant le lemme 1, on sait que

$$AB \bar{\Sigma}_{xx} B^T = \bar{\Sigma}_{yx} B^T$$

Donc

$$AB \bar{\Sigma}_{xx} B^T A^T = \bar{\Sigma}_{yx} B^T A^T$$

En utilisant a) et le fait que $\bar{\Sigma}_{xx}$, $\bar{\Sigma}_{xx}^{-1}$, P_A sont symétriques ainsi que $\bar{\Sigma}_{yx}^T = \bar{\Sigma}_{xy}$, on obtient

$$P_A \bar{\Sigma}_{yx} \bar{\Sigma}_{xx}^{-1} \bar{\Sigma}_{xx} \bar{\Sigma}_{xx}^{-1} \bar{\Sigma}_{xy} P_A = \bar{\Sigma}_{yx} \bar{\Sigma}_{xx}^{-1} \bar{\Sigma}_{xy} P_A$$

c'est à dire

$$P_A \bar{\Sigma} P_A = \bar{\Sigma} P_A$$

Comme $\bar{\Sigma}$ et P_A sont symétriques, on a aussi

$$\begin{aligned} P_A \bar{\Sigma} &= P_A^T \bar{\Sigma}^T = (\bar{\Sigma} P_A)^T \\ &= (P_A \bar{\Sigma} P_A)^T \\ &= P_A^T \bar{\Sigma}^T P_A^T \\ &= P_A \bar{\Sigma} P_A = \bar{\Sigma} P_A \end{aligned}$$



* Prop ii) $\Rightarrow$ i) :

page linéaire
Baldi-Kornik 3

On suppose donc $AB = P_A \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1}$ et $P_A \bar{\Sigma} = P_A \bar{\Sigma} P_A = \bar{\Sigma} P_A$

a) En multipliant la première égalité par $(A^T A)^{-1} A^T$ (à gauche), et en utilisant $P_A = A (A^T A)^{-1} A^T$, on obtient

$$\cancel{(A^T A)^{-1} A^T A} B = \cancel{(A^T A)^{-1} A^T} A \cancel{(A^T A)^{-1} A^T} \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1}$$

qui donne $B = (A^T A)^{-1} A^T \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1}$

b) On veut que $P_A \bar{\Sigma} P_A = \bar{\Sigma} P_A$ c'est à dire

$$P_A \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1} \bar{\Sigma}_{XX} \bar{\Sigma}_{XX}^{-1} \bar{\Sigma}_{XY} P_A = \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1} \bar{\Sigma}_{XY} P_A$$

qui comme $AB = P_A \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1}$ (et donc $B^T A^T = \bar{\Sigma}_{XX}^{-1} \bar{\Sigma}_{YX} P_A$) et on obtient permet d'obtenir.

$$AB \bar{\Sigma}_{XX} B^T A^T = \bar{\Sigma}_{YX} B^T A^T$$

En multipliant à ^{droite} gauche par $A (A^T A)^{-1}$, on obtient

$$AB \bar{\Sigma}_{XX} B^T = \bar{\Sigma}_{YX} B^T$$



Prop 1: ~~Q.E.D.~~

Σ_{xx} inversible, $\text{rang}(A) = n$, les valeurs propres de Σ sont distinctes, ~~non nulles~~.

(A^T, B) p.s. critique du 1^{er} ordre

$\exists C \in \mathbb{R}^{n \times n}$ inversible et $Y \subset \{1, \dots, n\}$ de taille n ,

sq

$$A = U_Y C \quad \text{et} \quad B = C^{-1} U_Y^T \Sigma_{YX} \Sigma_{XX}^{-1}$$

preuve:

$\Leftarrow$

a) $\text{rang}(A) = \text{rang}(\Sigma_{YX} \Sigma_{XX}^{-1})$
 $U_Y^T U_Y = I_{n_Y} : \begin{pmatrix} U_Y^T \end{pmatrix} = I_{n_Y} \text{ donc } U_Y^T U_Y = P_{U_Y}$

et $AB = U_Y C C^{-1} U_Y^T \Sigma_{YX} \Sigma_{XX}^{-1}$
 $= P_{U_Y} \Sigma_{YX} \Sigma_{XX}^{-1}$
 comme $A = U_Y C$

On l'espace ^{vectoriel} g n r  par les colonnes de U_Y est le m me que celui g n r  par les colonnes de A . Donc

$$P_A = P_{U_Y}$$

Donc $AB = P_A \Sigma_{YX} \Sigma_{XX}^{-1}$

b) $\text{rang}(P_A \Sigma) = \text{rang}(P_A \Sigma P_A) = \text{rang}(P_A)$

$U_Y U_Y^T = P_{U_Y}$ ses ~~valeurs propres~~

elle a donc n_Y valeurs propres valant 1 que l'on peut associer aux colonnes de U d'indice dans Y

$n - n_Y$ valeurs propres valant 0 " " " " hors de Y .

Dis autrement,

$P_A = P_{U_Y} = U_Y U_Y^T = U \text{diag}(\mathbb{1}_Y) U^T$ o  $\text{diag}(\mathbb{1}_Y)$ est une matrice diagonale dont les  l ments diagonaux d'indice dans Y valent 1 et ceux hors de Y valent 0.

Baldi-Hornik ^{page 14}

On a donc, en notant Λ_y la matrice extraite de Λ dont seuls les indices dans Y sont conservés.

$$P_A \Sigma = P_{U_y} \Sigma = U \operatorname{diag}(\pi_y) U^T U \Lambda U^T = U \Lambda_y U^T$$

$$P_A \Sigma P_A = U \operatorname{diag}(\pi_y) U^T U \Lambda U^T U \operatorname{diag}(\pi_y) U^T \\ = U \Lambda_y U^T$$

$$\text{et } \Sigma P_A = U \Lambda U^T U \operatorname{diag}(\pi_y) U^T = U \Lambda_y U^T$$

$\Rightarrow$ Comme $P_A \Sigma = \Sigma P_A$ et P_A et Σ sont diagonalisables, elles sont orthogonalisables.

Il existe V unitaire telle que

$$P_A = V \Lambda_A V^T \quad \text{et} \quad \Sigma = V \Lambda_\Sigma V^T$$

pour Λ_A et Λ_Σ diagonales.

Comme les valeurs propres de Σ sont distinctes, quitte

à changer l'ordre des colonnes et le signe des colonnes de V ,

on peut supposer que $V = U$

Donc

$$P_A = U \Lambda_A U^T$$

Par ailleurs, comme P_A est une projection sur un e.v. de dimension n_1 , il existe $Y \subset \{1, \dots, n_2\}$ de taille n_1 , tel que

$$P_A = U \operatorname{diag}(\pi_y) U^T (= P_{U_y})$$

Donc l'espace vectoriel généré par les colonnes de A coïncide avec l'espace généré par les colonnes de U_y .

Donc il existe $C \in \mathbb{R}^{n_1 \times n_1}$ inversible telle que

$$A = U_y C$$

Comme

$$B = (A^T A)^{-1} A^T \Sigma_{yx} \Sigma_{xx}^{-1}$$

$$= (C^T U_y^T U_y C)^{-1} C^T U_y^T \Sigma_{yx} \Sigma_{xx}^{-1}$$

$$= \overbrace{C^{-1} (C^T)^{-1} C^T U_y^T \Sigma_{yx} \Sigma_{xx}^{-1}}^{\text{Id}} = \text{Baldi-Kornik}$$

page 6

prop 2: Sous les hypothèses de la proposition précédente.

- 1) Pour (A, B) un point critique du 1^{er} ordre de E
si pour Y définissant A, B on a

$$(1.1) \quad AB = P_{U_Y} \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1}$$

~~2)~~ (1.2) $E(A, B) = \text{Trace}(\bar{\Sigma}_{YX}) - \sum_{i \in Y} \lambda_i$

- 2) Les seuls pts critiques d'ordre 2 sont les minimiseurs globaux. On a alors $Y = \{1, \dots, m\}$.

preuve:

- (1.1) En utilisant la prop. 1, on a

$$A = U_Y C \quad \text{et} \quad B = C^T U_Y^T \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1}$$

On a donc

$$AB = U_Y U_Y^T \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1}$$

$$\text{or } U_Y^T U_Y = \begin{pmatrix} & \\ & \end{pmatrix} = \bar{I}_m$$

$$\text{donc } U_Y U_Y^T = P_{U_Y}$$

□ 1.1

- 1.2) Rappelons tout d'abord que pour

$$u, v \in \mathbb{R}^n \quad u^T v = \text{Trace } v u^T = v^T u$$

$$\frac{u^T}{u^T} v = \text{Trace } \frac{u^T}{v}$$

On a

$$\begin{aligned} E(A, B) &= \sum_{i=1}^L \|y_i - AB x_i\|^2 \\ &= \sum_{i=1}^L (y_i - AB x_i)^T (y_i - AB x_i) \\ &= \sum_{i=1}^L y_i^T y_i - 2 x_i^T B^T A^T y_i + x_i^T B^T A^T A B x_i \\ &= \sum_{i=1}^L y_i^T y_i - 2 \text{Tr}(B^T A^T y_i x_i^T) + \text{Tr}(B^T A^T A B x_i x_i^T) \\ &\stackrel{a}{=} \text{Tr} \bar{\Sigma}_{YX} - 2 \text{Tr}(B^T A^T \bar{\Sigma}_{YX}) + \text{Tr}(B^T A^T A B \bar{\Sigma}_{XX}) \end{aligned}$$

Reduction

$$\begin{aligned}
 \text{On } B^T A^T A B \Sigma_{xx} &= B^T A^T P_{U_y} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xx} \\
 &= B^T A^T P_{U_y} \Sigma_{yx} \\
 &= B^T A^T \Sigma_{xx}^{-1} \Sigma_{xy} U_y U_y^T U_y U_y^T \Sigma_{yx} \\
 &= \Sigma_{xx}^{-1} \Sigma_{xy} U_y U_y^T \Sigma_{yx} \quad \text{Id} \\
 &= B^T A^T \Sigma_{yx}
 \end{aligned}$$

donc $E(AB) = \text{Tr} \Sigma_{yy} - \text{Tr}(B^T A^T \Sigma_{yx})$

et

$$\begin{aligned}
 \text{Tr}(B^T A^T \Sigma_{yx}) &= \text{Trace}(\Sigma_{xx}^{-1} \Sigma_{xy} P_{U_y} \Sigma_{yx}) \\
 &= \text{Trace}(\Sigma P_{U_y}) \\
 &= \text{Trace}(U \Lambda U^T U \text{diag}(\pi_y) U^T) \\
 &= \text{Trace}(U \Lambda_y U^T) \\
 &= \sum_{i \in y} \lambda_i
 \end{aligned}$$

D 1.2

- 2) Les seules minimas globales sont les points critiques pour lesquels $y = \{1, \dots, n\}$.
 du 1^{er} ordre cela permet de conclure si $n_1 \neq n_2$. Si $n_2 > n_1$.
 Or les points critiques correspondant à $y \neq \{1, \dots, n\}$
 ne sont pas des points critiques du second ordre.
 L'énoncé est une conséquence directe des précédents :

- 3) Il faut uniquement montrer $\text{min local} \Rightarrow \text{min global}$.
 On va montrer, en utilisant 2), qu'un point critique qui n'est pas un min global, n'est pas un min local.

Soit (A, B) un point critique du premier ordre associé à y tq

- $\exists i \in \{1, \dots, n_1\}$ avec $i \notin y$
- $\exists j \in \{n_1+1, \dots, n_2\}$ avec $j \in y$

et tq $\exists C \in \mathbb{R}^{n_1 \times n_2}$ tq $A = U_y C$ et $B = C^T U_y \Sigma_{yx} \Sigma_{xx}^{-1}$

Radici-Hornik
9

On note $u_t = (u_j + t u_i) / \sqrt{1+t^2}$

où u_j et u_i sont les colonnes j et i de U .

On considère U_t valant U_j , sauf la colonne j qui est remplacée par u_t .

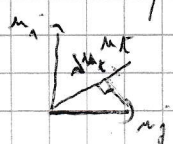
On pose $A_t = U_t C$ et $B_t = C^{-1} U_t^T \Sigma_{YX} \Sigma_{XX}^{-1}$

On a $U_t^T U_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = Id_n$ $P_{U_t} E(A_t, B_t) \leq F(A, B)$
 $t \neq 0$

Donc $P_{U_t} = U_t U_t^T$

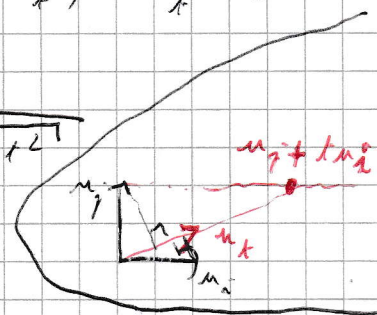
On a par ailleurs $P_{U_t} u_k = \begin{cases} 0 & \text{si } k \notin \{U\setminus\{i\}\} \\ u_k & \text{si } k \in \{U\setminus\{i\}\} \end{cases}$

De plus $P_{U_t} u_j = \lambda u_t$ pour λ tel que $(u_j - \lambda u_t)^T u_t = 0$



c-a-d. $\lambda = \frac{u_j^T u_t}{u_t^T u_t} = \frac{u_j^T u_j}{\sqrt{1+t^2}} = \frac{1}{\sqrt{1+t^2}}$

Donc $P_{U_t} u_j = \frac{1}{1+t^2} u_j + \frac{t}{1+t^2} u_i$



De même $P_{U_t} u_i = \lambda' u_t$ pour λ' tel que $(u_i - \lambda' u_t)^T u_t = 0$

c-a-d. $\lambda' = \frac{u_i^T u_t}{u_t^T u_t} = \frac{t}{\sqrt{1+t^2}}$

Donc $P_{U_t} u_i = \frac{t}{1+t^2} u_j + \frac{t^2}{1+t^2} u_i$

Finalement

$P_{U_t} U = U \Sigma'$

c'est à dire $P_{U_t} = U \Sigma' U^T$

pour

$\Sigma' = \begin{pmatrix} 1 & & & \\ & \frac{t}{1+t^2} & & \\ & & \frac{1}{1+t^2} & \\ & & & \frac{t^2}{1+t^2} & \\ & & & & 1 \end{pmatrix}$

$\Sigma'_{k,l} = \begin{cases} 1 & \text{si } k=l \in \{U\setminus\{i\}\} \\ \frac{t}{1+t^2} & \text{si } l=j, k=i \\ \frac{1}{1+t^2} & \text{si } l=i, k=j \\ \frac{t^2}{1+t^2} & \text{si } l=i, k=i \\ 0 & \text{sinon} \end{cases}$

si $k=l \in \{U\setminus\{i\}\}$

si $l=j, k=i$ $\left. \begin{matrix} \text{car} \\ P_{U_t} u_j \end{matrix} \right\}$

si $l=i, k=j$ $\left. \begin{matrix} \text{car} \\ P_{U_t} u_i \end{matrix} \right\}$

si $l=i, k=i$ $\left. \begin{matrix} \text{car} \\ P_{U_t} u_i \end{matrix} \right\}$

sinon (Baldi-Hornik 9)

On a donc comme précédemment (voir preuve du #1, page 7 et 8)

$$A_t B_t = P_{U_t} \bar{\Sigma}_{YX} \bar{\Sigma}_{XX}^{-1}$$

$$E(A_t, B_t) = \text{tr} \bar{\Sigma}_{YY} - 2 \text{tr} (B_t^T A_t^T \bar{\Sigma}_{YX}) + \text{tr} (B_t^T A_t^T A_t B_t \bar{\Sigma}_{XX})$$

$$\text{et } B_t^T A_t^T A_t B_t \bar{\Sigma}_{XX} = B_t^T A_t^T \bar{\Sigma}_{YX}$$

moins $\text{tr} (B_t^T A_t^T \bar{\Sigma}_{YX}) = \text{tr} (\bar{\Sigma}_{XX}^{-1} \bar{\Sigma}_{XY} P_{U_t} \bar{\Sigma}_{YX})$

$$= \text{tr} (\bar{\Sigma} P_{U_t})$$

$$= \text{tr} (U \Lambda U^T U \bar{\Sigma}' U^T)$$

$$= \text{tr} (\Lambda \bar{\Sigma}')$$

$$= \sum_{k \in \bar{Y} \setminus \{j\}} \lambda_k + \lambda_j \frac{1}{1+t^2} + \lambda_i \frac{t^2}{1+t^2}$$

$$= \text{''} + (\lambda_j - \lambda_j t^2 + o(t^2)) + (\lambda_i t^2 + o(t^2))$$

$$= \sum_{k \in \bar{Y}} \lambda_k + t^2 (\lambda_i - \lambda_j) + o(t^2)$$

Finalement

$$E(A_t, B_t) = \text{tr} \bar{\Sigma}_{YY} - \sum_{k \in \bar{Y}} \lambda_k - t^2 (\lambda_i - \lambda_j) + o(t^2)$$

$$= E(A, B) - t^2 (\lambda_i - \lambda_j) + o(t^2)$$

comme $\lambda_i > \lambda_j$,

(A, B) n'est pas un min local.

