

Diplôme Postgrade en Statistique  
La régression PLS

22 juin 2004

Réalisé par **Séverine Vancolen**  
Supervisé par le Professeur Yadolah Dodge

Groupe de Statistique  
Université de Neuchâtel  
Suisse

## Table des matières

<b>1</b>	<b>Introduction - Problématique</b>	<b>4</b>
1.1	Présentation . . . . .	4
1.2	Historique . . . . .	5
1.3	Comparaison . . . . .	5
1.3.1	La Régression Linéaire Multiple (MLR) . . . . .	6
1.3.2	La régression sur composantes principales . . . . .	6
1.3.3	La régression Ridge . . . . .	6
1.3.4	Relation avec les autres techniques . . . . .	6
<b>2</b>	<b>Interêt de la Régression PLS</b>	<b>7</b>
2.1	Défaillances du modèle linéaire . . . . .	7
2.2	Les approches pour contourner le problème de la multico- linéarité . . . . .	8
2.2.1	La Régression "Pas à Pas" . . . . .	8
2.2.2	La Régression Ridge . . . . .	8
2.2.3	Les méthodes factorielles . . . . .	8
<b>3</b>	<b>La méthode PLS linéaire</b>	<b>10</b>
3.1	Modèle de base . . . . .	10
3.2	La méthode . . . . .	11
3.2.1	Description de la $k^{ieme}$ étape . . . . .	11
3.2.2	Problème d'optimisation . . . . .	12
3.2.3	Calcul des gradients . . . . .	12
3.2.4	Calcul de $\lambda$ et $\mu$ . . . . .	13
3.2.5	Calcul de $w$ et $q$ . . . . .	13
3.2.6	Remarque . . . . .	13
3.2.7	Propriétés des composantes $t_1, \dots, t_A$ . . . . .	14
3.2.8	Le modèle PLS . . . . .	14
3.2.9	Cas particuliers . . . . .	15
3.3	Les algorithmes PLS . . . . .	16
3.3.1	L'algorithme PLS1 NIPALS . . . . .	16
3.3.2	L'algorithme de régression classique PLS2 (PLS mul- tivariable, $c > 1$ ) . . . . .	16
3.3.3	L'algorithme SIMPLS . . . . .	17
<b>4</b>	<b>Choix du nombre de composantes</b>	<b>18</b>
4.1	Critère 1 : Le "FIT" ou "Critère d'Ajustement" . . . . .	18
4.2	Critère 2 : Le "PRESS" ou "Critère de Validation Croisée" . . . . .	18
4.3	Critère 3 : "Critère de Validation Externe" et Prédiction . . . . .	19
<b>5</b>	<b>Les logiciels</b>	<b>19</b>

<b>6</b>	<b>Les extensions de la méthode PLS</b>	<b>19</b>
6.1	Les extensions linéaires . . . . .	19
6.1.1	"Discrimination Partial Least Squares" . . . . .	20
6.1.2	Une version robuste de l'algorithme SIMPLS . . . . .	20
6.1.3	"Partial Least Squares Proportional Hazard Regression" . . . . .	21
6.1.4	Extensions Multitableaux de la régression PLS . . . . .	21
6.1.5	"Partial Quantile Regression" . . . . .	22
6.2	Les méthodes PLS non linéaires . . . . .	22
6.2.1	PLS quadratique . . . . .	22
6.2.2	Non Linear PLS : NLPLS . . . . .	23
6.2.3	SPLine PLS : SPLPLS . . . . .	23
6.2.4	Continuum Non Linear PLS : CNLPLS . . . . .	23
6.2.5	Implicit Non Linear latent variable Regression : INLR . . . . .	23
6.2.6	La régression Partial Least Squares Splines : PLSS . . . . .	23
6.2.7	"Kernel Partial Least Squares Regression" . . . . .	24
<b>7</b>	<b>Conclusion</b>	<b>24</b>

# 1 Introduction - Problématique

De nombreux problèmes industriels peuvent être décrits sous la forme d'un modèle de régression, où l'on possède des variables  $X$  sur lesquelles on peut plus ou moins agir et des variables  $Y$  que l'on ne peut qu'observer. L'objectif est alors de décrire les relations entre  $Y$  et  $X$ , en l'absence de modèle théorique. Le problème est que le nombre de variables  $X$  est souvent très important par rapport au nombre d'observations. La régression PLS (Partial Least Squares Regression) est une méthode d'analyse des données spécifiquement construite pour l'étude de ce type de problème.

## 1.1 Présentation

La régression PLS est donc une extension du modèle de régression linéaire multiple. Dans sa forme la plus simple, un modèle linéaire spécifie la relation (linéaire) entre une (ou des) variables dépendantes (réponses)  $Y$  et un ensemble de variables prédictives  $X$  telles que :

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$$

où les  $b_i$  sont les coefficients de régression.

Alors par exemple, on peut estimer (prédire) le poids d'une personne en fonction de sa taille ou de son sexe. On peut utiliser une régression linéaire pour estimer les coefficients de régression respectifs à partir d'un échantillon de données mesurant la taille et le poids et en observant le sexe des individus.

Dans beaucoup de problèmes d'analyses de données, l'estimation de la relation linéaire entre deux variables est adéquate pour décrire les données observées et pour faire de bonnes prévisions pour de nouvelles observations.

Le modèle de régression multiple a été étendu de plusieurs façons afin de s'adapter aux problèmes d'analyse de données plus sophistiqués. Il sert donc de base pour de nombreuses méthodes multivariées comme l'analyse discriminante, la régression sur composantes principales (PCR) et la corrélation canonique.

La régression PLS est une technique récente qui généralise et combine les caractéristiques de l'analyse sur composantes principales et de la régression multiple.

Elle est particulièrement utile quand on a besoin de prédire un ensemble de variables dépendantes à partir d'un ensemble très grand de variables explicatives (prédicteurs) qui peuvent être très fortement corrélées entre elles.

Quand les prédicteurs sont peu nombreux, non significativement colinéaires et ont une relation connue avec les réponses, alors la régression linéaire multiple est la meilleure méthode pour utiliser les données. Cependant si l'une de ces trois conditions n'est pas vérifiée, la régression linéaire multiple peut être inefficace et inappropriée.

PLS est donc une méthode pour construire des modèles de prédiction quand les facteurs sont nombreux et très colinéaires.

Notons que cette méthode met l'accent sur la prédiction de la réponse et pas nécessairement sur la mise en évidence d'une relation entre les variables. Ce qui signifie que PLS n'est pas appropriée pour désigner les variables ayant un effet négligeable sur la réponse, mais quand le but est la prédiction et qu'il n'y a pas besoin de limiter le nombre de variables mesurées, PLS est un outil très utile.

## 1.2 Historique

La régression PLS tire son origine des sciences sociales (plus précisément des sciences économiques, Herman Wold 1966 [28]) mais devient très populaire en chimie grâce au fils d'Herman, Svante.

La régression PLS est née de l'association de l'algorithme NIPALS (Non linear Iterative Partial Least Squares) développé par H. Wold [28] pour l'analyse sur composantes principales et de l'approche PLS proposée par H. Wold [29] pour l'estimation des modèles d'équations structurelles sur les variables latentes. Il en résulte une représentation "classique" de la régression PLS sous la forme d'un algorithme (remarquons qu'il y a plusieurs versions possibles de l'algorithme NIPALS aboutissant aux mêmes résultats).

Une méthode d'estimation alternative pour les composantes de la régression est l'algorithme SIMPLS de de Jong [4].

**1966** : Herman Wold [28] publie un ouvrage introduisant le PLS, alors appelé NonLinear Iterative Partial Least Squares NIPALS.

**1983** : Svante Wold (le fils d'Herman) et Harold Mertens [34] adaptent NIPALS au problème de régression avec trop de prédicteurs et appellent PLS cette adaptation de l'algorithme.

**1990** : Stone et Brooks [18] introduisent le PLS dans le contexte du "Continuum Regression" (elle ajoute un paramètre continu  $\lambda$  autorisant la méthode de modélisation à varier continuellement entre MLR "Multiple Linear Regression" ( $\lambda = 0$ ), PLS ( $\lambda = 0.5$ ) et PCR "Principal Component Regression" ( $\lambda = 1$ )). Ce qui signifie une première projection sérieuse du PLS dans un contexte statistique.

Plus récemment la littérature concernant la méthode PLS s'est largement étoffée, surtout dans le journal "Chemometric". De plus un ouvrage entièrement en français dédié à cette méthode a été écrit par Tenenhaus en 1998 [21].

## 1.3 Comparaison

Comparé à d'autres méthodes de régression pour des données colinéaires, il a été établi que le plus grand avantage de PLSR est que "l'information dans

la variable  $Y$  est utilisée”, mais c’est une notion qu’il est difficile d’établir statistiquement. Un avantage plus évident est que la méthode rend possible la combinaison de la prédiction avec l’étude d’une structure jointe latente dans les variables  $X$  et  $Y$ . Ainsi, la méthode demande souvent moins de composantes que PCR pour donner une bonne prédiction. Avec la facilité d’implémentation, cela a rendu cette méthode très populaire en chimie, en particulier en spectroscopie, mais les applications dans d’autres domaines sont aussi très répandues.

### 1.3.1 La Régression Linéaire Multiple (MLR)

Si le nombre de facteurs  $A$  extrait est plus grand ou égal au rang de l’espace de l’échantillon  $X$ , alors PLS est équivalent MLR.

### 1.3.2 La régression sur composantes principales

PLS est une technique quantitative de décomposition spectrale étroitement liée à la régression sur composantes principale (PCR). Cependant, dans PLS, la décomposition est faite d’une manière légèrement différente.

Au lieu de décomposer d’abord la matrice spectrale  $X$  en un ensemble de vecteurs propres et de scores, et de les régresser contre les  $Y$  dans une étape séparée, PLS utilise l’information de  $Y$  en même temps que le processus de décomposition.

Ceci implique que les variables expliquant le mieux les  $Y$  seront plus fortement pondérées. De plus, les vecteurs propres et les scores calculés en utilisant PLS seront différents de ceux de PCR. L’idée principale de PLS est de donner le plus d’information possible sur  $Y$  dans les premiers vecteurs construits.

### 1.3.3 La régression Ridge

La régression Ridge est une technique originaire des statistiques permettant de manipuler la colinéarité en régression. Elle est probablement la plus grande rivale de PLS en terme de flexibilité et de robustesse du modèle prédictif. Mais cette méthode n’a pas de procédure de réduction de dimension (extraire linéairement peu de facteurs latents qui sont les plus utiles pour modéliser la réponse).

### 1.3.4 Relation avec les autres techniques

La régression PLS est évidemment liée à la corrélation canonique et à l’analyse des facteurs multiples. Ces relations sont explorées en détail par Tenenhaus (1998, [21]) et Pagès et Tenenhaus (2001, [13]). La principale originalité de la régression PLS est de préserver l’asymétrie de la relation

entre les prédicteurs et les variables dépendantes, contrairement aux autres techniques qui les traitent symétriquement.

## 2 Interêt de la Régression PLS

La régression PLS pour le modèle linéaire s'applique au champ d'application où l'on possède peu d'observations sur des variables très corrélées et en très grand nombre, c'est le cas pathologique de la régression linéaire, là où les méthodes habituelles ne fonctionnent pas ( $n$  étant très petit on ne peut pas tendre vers l'infini).

### 2.1 Défaillances du modèle linéaire

Supposons que l'on a :

- $p$  variables explicatives
- $c$  variables réponses
- un échantillon de  $n$  observations

Nous allons considérer dans un premier temps  $c = 1$

On appelle :

- $T_{n \times p}$  la matrice des observations sur les variables explicatives centrées
- $y_{n \times 1}$  les observations sur les variables réponses centrées.

On suppose ici que  $\text{rang}(T) = p$  (ce qui est faux en général)

Soit  $X_{n \times p}$  la matrice des variables centrées réduites, c'est à dire  $X^j$  la  $j^{\text{ieme}}$  colonne de  $X$  s'écrit :

$$X^j = \frac{T^j}{\sqrt{\sum_{i=1}^n (T_i^j)^2}}$$

Soit  $V$  la matrice des covariances empiriques ( $V = X^T X$ ).

Le modèle linéaire s'écrit :

$$Y = X\beta + \epsilon$$

où  $\epsilon \sim N(0, \sigma^2 I)$ .

Alors l'estimateur de  $\beta$  s'écrit :

$$\hat{\beta} = (X X^T)^{-1} X^T Y$$

et  $E(\hat{\beta}) = \beta$ .

On sait alors que l'erreur quadratique moyenne vaut :

$$MSE = E(\|\hat{\beta} - \beta\|_2^2) = \sigma^2 \text{trace}(V^{-1})$$

or on sait aussi que l'ensemble des valeurs propres de  $V^{-1}$  :  $\lambda(V^{-1})$ , peut s'écrire  $\lambda(V^{-1}) = \frac{1}{\lambda(V)}$ . Alors l'erreur quadratique moyenne se réécrit :

$$MSE = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j(V)}$$

Or avec les problèmes de multicolinéarité,  $\lambda_j(V) \sim 0$  et donc la MSE devient très grande et les coordonnées de  $\hat{\beta}$  ne sont plus interprétables.

On peut facilement montrer que  $trace(V^{-1}) = \sum_{j=1}^p \frac{1}{1-R_{-j}^2}$  où  $R_{-j}^2$  est le coefficient  $R^2$  quand on effectue la régression de la variable explicative  $X^j$  sur les autres variables.

Or  $R_{-j}^2$  sera voisin de 1 si  $X^j$  est très corrélé avec les autres variables, donc  $1 - R_{-j}^2$  sera proche de 0 et la MSE sera infinie.

On comprend maintenant mieux pourquoi le problème de multicolinéarité influe sur la variance de l'estimateur.

## 2.2 Les approches pour contourner le problème de la multicolinéarité

### 2.2.1 La Régression "Pas à Pas"

On va sélectionner seulement les variables les plus influentes dans le modèle.

### 2.2.2 La Régression Ridge

**Idée :** Au lieu d'inverser  $V$ , on inverse  $(X^T X + \lambda I_p)$  (On perturbe la matrice de covariance) où  $\lambda$  est un superparamètre de la méthode qu'il faudra déterminer.

- L'avantage est que même si  $rang(X) < p$ ,  $(X^T X + \lambda I_p)$  est toujours inversible.
- Le problème est tout d'abord dans le choix de  $\lambda$  (il faudra faire de la validation croisée). Et il est illusoire de croire qu'avec un seul paramètre  $\lambda$  on pourra régler le problème de multicolinéarité, il y a des cas où ça ne marche pas.

### 2.2.3 Les méthodes factorielles

On fait de la réduction de dimension. On construit un petit nombre de variables synthétiques (variables latentes ou composantes principales) qui sont des combinaisons linéaires des variables naturelles (elles résument ces variables naturelles).

**1<sup>ere</sup> méthode factorielle : La PCR** On fait une ACP (Analyse sur Composantes Principales) d'ordre  $k$  du tableau  $X$  qui va donner  $C_{n \times k}$  le tableau des composantes principales tel que les nouvelles variables dans  $C$  ne sont pas corrélées.

Ensuite on effectue la régression  $\hat{Y} = P_c Y$ , où  $P_c = C(C^T C)^{-1} C^T$ .

Propriétés des  $C^j$  :

- centrées
- non corrélées



–  $var(C^j) = \lambda_j(X^T X)$  où  $\lambda_j$  est la  $j^{eme}$  valeur propre de la matrice  $X^T X$ .

On a  $C^j = XV^j$  et  $X^T XV^j = \lambda_j V^j$ .

C'est à dire  $C^j$  est combinaison linéaire des colonnes de  $X$ ,  $C^j = X^1 V_1^j + \dots + X^p V_p^j$ .

Comme  $P_c = C(C^T C)^{-1} C^T$ , on retrouve bien un modèle linéaire en les variables  $X_1, \dots, X^p$ .

Les variables latentes sont choisies par le critère du minimum de la variance.

Inconvénients :

– Les 1<sup>eres</sup> composantes principales ne sont pas forcément celles qui expliqueront le mieux le  $Y$ .

– Le choix de  $k$  n'est pas évident (il faut un deuxième critère)

Il vaudrait donc mieux choisir les composantes principales les plus corrélés avec les  $Y$  et cela revient à du Pas à Pas.

2<sup>ieme</sup> **méthode factorielle : La régression PLS** ("Partial Least Squares" ou "Projection on to Latent Structure")

Avec les méthodes factorielles, on obtient  $k$  variables latentes, placées dans la matrice  $C_{n \times k}$  pour l'ACP, et on obtient  $\hat{Y} = P_c Y$ .

Or si  $k = rang(X)$ , on aura  $Im(C) = Im(X)$ , c'est à dire les variables latentes non corrélées forment une base de  $Im(X)$  et je retrouve exactement la régression linéaire :

$$P_c = P_X$$

et

$$\hat{Y}_{PLS} = \hat{Y}_{PCR} = \hat{Y}_{OLS}.$$

L'idée de la régression PLS est donc de prendre les variables latentes les plus corrélées avec  $Y$  et de les projeter.

PLS munit les observations de poids  $p_1, \dots, p_n$  et l'on notera  $D = diag(p_1, \dots, p_n)$  avec  $p_i > 0$  et  $\sum_i p_i = 1$ .

Par défaut on prendra  $D = \frac{1}{n}$ , c'est à dire poids équirépartis, et on supposera les variables centrées par rapport à ces poids et réduites. C'est à dire  $1_n^T D X = O_{1 \times p}$ ,  $1_n^T D Y = O_{1 \times c}$  et les matrices de covariances empiriques s'écrivent  $X^T D X$ ,  $Y^T D Y$  et  $X^T D Y$ .

**Le modèle PLS linéaire s'écrit alors :**

$$\hat{Y}^j = \sum_{k=1}^p X^k \hat{\beta}_k^j(A)$$

$j = 1, \dots, c$ , et il dépend du nombre  $A$  de composantes principales PLS à estimer.

Il faut alors choisir le superparamètre  $A$  avec beaucoup de précautions.

**Remarque :**  $PLS(X, Y = X) = ACP(X)$ .

## 3 La méthode PLS linéaire

### 3.1 Modèle de base

#### Principe

Comme dans la régression linéaire multiple, le but principal de la régression PLS est de construire un modèle linéaire

$$Y = XB + E$$

où  $B_{p \times c}$   $c$  coefficients de régression,  $E_{n \times c}$  terme de bruit pour le modèle.

Usuellement les variables dans  $X$  et  $Y$  sont centrées en soustrayant leur moyenne, et réduites en divisant par leur écart type. La régression en composantes principales et la régression PLS produisent toutes les deux des facteurs de scores comme des combinaisons linéaires des variables prédictives originelles, de telle manière qu'il n'y ait pas de corrélation entre les facteurs scores utilisés par le modèle de régression prédictive.

Par exemple, supposons que nous ayons un ensemble de données pour des variables réponses  $Y$  et un grand nombre de variables prédictives  $X$ , dont certaines sont très fortement corrélées. Une régression utilisant l'extraction des facteurs pour ce type de données calcule la matrice de facteurs score  $T = XW$  pour une matrice de poids appropriée  $W$  et alors on considère le modèle de régression linéaire  $Y = TQ + E$  où  $Q$  est une matrice des coefficients de régression pour  $T$  et  $E$  un terme d'erreur. Une fois les  $Q$  calculés, le modèle de régression ci-dessus est équivalent à  $Y = XB + E$  où  $B = WQ$  qui peut être utilisé comme un modèle de régression prédictive.

**Remarque** La régression sur composantes principales et la régression PLS diffèrent dans la méthode utilisée pour extraire les facteurs de scores. En résumé, la régression sur composantes principales produit une matrice de poids  $W$  reflétant la structure de covariances entre les variables prédictives alors que la régression PLS produit une matrice de poids  $W$  reflétant les structures de covariance entre les prédicteurs et les réponses.

Pour établir le modèle, la régression PLS produit une matrice de poids  $W_{p \times c}$  pour  $X$  telle que  $T = XW$ , c'est à dire les colonnes de  $W$  sont des vecteurs de poids pour les colonnes de  $X$  produisant la matrice de facteurs de score  $T_{n \times c}$  correspondante. Ces poids sont calculés de telle façon qu'ils maximisent la covariance entre la réponse et les facteurs de score correspondants.

La procédure OLS (Ordinary Least Squares) pour la régression de  $Y$  sur  $T$  est alors utilisée pour produire  $Q$ , tel que  $Y = TQ + E$ .

Une fois  $Q$  calculé, nous avons  $Y = XB + E$  où  $B = WQ$  et le modèle de prédiction est complet.

Une matrice supplémentaire nécessaire pour une description complète de la procédure de régression PLS est la matrice  $P$  des facteurs qui donne le modèle  $X = TP + F$  où  $F$  est la partie non expliquée du score de  $X$ .

### 3.2 La méthode

La méthode PLS est une méthode de régression linéaire de  $c$  variables réponses sur  $p$  variables explicatives toutes mesurées sur les mêmes  $n$  individus. Les tableaux des observations, notés respectivement  $Y$  et  $X$ , de dimensions  $n \times c$  et  $n \times p$ , sont supposés centrés et éventuellement réduits par rapport aux poids  $(p_1, \dots, p_n)$ . On note  $D = \text{diag}(p_1, \dots, p_n)$  la matrice diagonale des poids.

L'intérêt de la méthode comparée à la régression sur composantes principales (RCP), réside dans le fait que les composantes PLS sur les  $X$ , notées  $t$ , sont calculées "dans le même temps" que des régressions partielles sont exécutées. Cette simultanéité leur confère un meilleur pouvoir prédictif que celles de la RCP. La question est donc d'examiner comment cette simultanéité est mise en oeuvre.

Notons  $E_0 = X$  et  $F_0 = Y$  les tableaux centrés et réduits au sens de  $D$  qui en général est égal à  $\frac{1}{n}I_n$ . La méthode procède par étapes succesives permettant le calcul des composantes principales. On notera  $A$  le nombre total d'étapes, c'est à dire de composantes indicées par  $k = 1, \dots, A$ .

#### 3.2.1 Description de la $k^{\text{ième}}$ étape

Notons  $t = E_{k-1}w$  et  $u = F_{k-1}q$ , les combinaisons linéaires colonnes des matrices centrées  $E_{k-1}$  et  $F_{k-1}$ , associées respectivement aux vecteurs des poids  $w$  et  $q$ . La covariance entre  $t$  et  $u$  s'écrit comme le  $D$ -produit scalaire

$$\text{cov}(t, u) = (t, u)_D = w^T E_{k-1}^T D F_{k-1} q.$$

Le carré de la  $D$ -norme associée fournit la variance  $\|t\|_D^2 = \text{var}(t)$ .

L'étape  $k$ , se décompose en deux parties. La première fournit les composantes  $t_k = E_{k-1}w_k$  et  $u_k = F_{k-1}q_k$  par le calcul des poids optimaux  $w_k$  et  $q_k$ . La deuxième actualise les matrices des prédicteurs et des réponses  $E_k$  et  $F_k$  comme résidus de la régression sur  $t_k$ .

Calcul des poids	$(w_k, q_k) = \text{argmax}_{\text{cov}(t, u) = w^T E_{k-1}^T D F_{k-1} q}$ sous les contraintes $\ w\ ^2 = \ q\ ^2 = 1,$
Actualisation	$E_k = E_{k-1} - P_{t_k} E_{k-1}$ $F_k = F_{k-1} - P_{t_k} F_{k-1},$

où  $P_{t_k} = \frac{t_k t_k^T D}{\text{var}(t_k)}$  est la matrice  $n \times n$  de projection  $D$ -orthogonale sur  $t_k$ .

Remarquons que le critère à optimiser, la covariance, est un compromis entre le critère de l'Analyse des Corrélations Canoniques, la corrélation, et celui de l'ACP sur chacun des tableaux, la racine carrée de la variance.

$$\begin{array}{llll}
\text{cov}(t_k, u_k) & = \text{cov}(X_{k-1}w_k, Y_{k-1}q_k) & & \\
\text{cov}^2(X_{k-1}w_k, Y_{k-1}q_k) & = \text{cor}^2(X_{k-1}w_k, Y_{k-1}q_k) & \text{var}(X_{k-1}w_k) & \text{var}(Y_{k-1}q_k) \\
& \text{Analyse Canonique} & \text{ACP de } X & \text{ACP de } Y
\end{array}$$

### 3.2.2 Problème d'optimisation

La fonction qu'il faut maximiser est :

$$\begin{array}{ll}
\mathfrak{R}^p \times \mathfrak{R}^c \mapsto & \mathfrak{R} \\
(w, q) \mapsto & \phi(w, q) = \text{cov}(t, u) \\
& = t^T D u \\
& = u^T D t
\end{array}$$

C'est à dire géométriquement :  $\phi(w, q) = \|t\|_D \|u\|_D \cos(t, u)$  et statistiquement :  $\phi(w, q) = \sigma(t)\sigma(u)r(t, u)$  où  $\sigma(t) = \|t\|_D = \sqrt{t^T D t}$  et  $\sigma(u) = \|u\|_D = \sqrt{u^T D u}$ .

Alors la fonction de Lagrange associée à ce problème est :

$$\begin{array}{ll}
\mathfrak{R}^p \times \mathfrak{R}^c \times \mathfrak{R} \times \mathfrak{R} \mapsto & \mathfrak{R} \\
(w, q, \lambda, \mu) \mapsto & L(w, q, \lambda, \mu) = \phi(w, q) + \frac{\lambda}{2}(1 - \|w\|_2^2) + \frac{\mu}{2}(1 - \|q\|_2^2)
\end{array}$$

Le problème de maximisation de  $\phi$  sous contrainte est équivalent au problème de maximisation de  $L$  sans contrainte. pour cela il faut calculer le gradient de  $L$  et résoudre le système suivant :

$$\begin{array}{ll}
\nabla_w L & = \nabla_w \phi(w, q) - \frac{\lambda}{2} \nabla_w \|w\|_2^2 = 0 \\
\nabla_q L & = \nabla_q \phi(w, q) - \frac{\mu}{2} \nabla_q \|q\|_2^2 = 0 \\
\frac{dL}{d\lambda} & = 1 - \|w\|_2^2 = 0 \\
\frac{dL}{d\mu} & = 1 - \|q\|_2^2 = 0
\end{array}$$

où  $\nabla_w L$  est le gradient de  $L$  par rapport  $w$

$\nabla_q L$  est le gradient de  $L$  par rapport  $q$

et les 2 dernières lignes du système sont les contraintes.

### 3.2.3 Calcul des gradients

$\phi(w, q) = w^T X_{k-1}^T D Y_{k-1} q$  d'où  $\nabla_q \phi(w, q) = (w^T X_{k-1}^T D Y_{k-1})^T = Y_{k-1}^T D X_{k-1} w$   
Et  $\phi(w, q) = q^T Y_{k-1}^T D X_{k-1} w$  d'où  $\nabla_w \phi(w, q) = (q^T Y_{k-1}^T D X_{k-1})^T = X_{k-1}^T D Y_{k-1} q$   
De plus  $\|w\|_2^2 = w^T w$  et  $\|q\|_2^2 = q^T q$  alors  $\nabla_w \|w\|_2^2 = 2w$  et  $\nabla_q \|q\|_2^2 = 2q$

Alors le système devient :

$$\nabla_w L = X_{k-1}^T D Y_{k-1} q - \lambda w = 0 \quad p \text{ équations} \quad (1)$$

$$\nabla_q L = Y_{k-1}^T D X_{k-1} w - \mu q = 0 \quad c \text{ équations} \quad (2)$$

$$\frac{dL}{d\lambda} = 1 - \|w\|_2^2 = 0 \quad 1 \text{ équation} \quad (3)$$

$$\frac{dL}{d\mu} = 1 - \|q\|_2^2 = 0 \quad 1 \text{ équation} \quad (4)$$

On a donc un système de  $p + c + 2$  équations à  $p + c + 2$  inconnus.

### 3.2.4 Calcul de $\lambda$ et $\mu$

$$w^T * (1) = w^T X_{k-1}^T D Y_{k-1} q - \lambda w^T w = 0$$

Or par (3)  $w^T w = 1$  d'où :

$$\boxed{w^T X_{k-1}^T D Y_{k-1} q = \lambda}$$

$$\text{De même } q^T * (2) = q^T Y_{k-1}^T D X_{k-1} w - \mu q^T q = 0$$

Or par (4)  $q^T q = 1$  d'où :

$$\boxed{q^T Y_{k-1}^T D X_{k-1} w = \mu}$$

$$\Rightarrow \boxed{\lambda = \mu}$$

Donc à l'optimum, on sait que  $\lambda = \mu = \phi(w, q)$ .

### 3.2.5 Calcul de $w$ et $q$

On appelle (1) et (2) les formules de transitions car on peut calculer  $w$  et  $q$  l'un en fonction de l'autre alors qu'ils n'appartiennent pas au même espace.

$$\mu * (1) + X_{k-1}^T D Y_{k-1} \mu q - \lambda \mu w = 0.$$

Or  $\lambda = \mu$  et d'après (2)  $\mu q = Y_{k-1}^T D X_{k-1} w$ .

$$\text{D'où } \boxed{X_{k-1}^T D Y_{k-1} Y_{k-1}^T D X_{k-1} w = \lambda^2 w}.$$

De même  $\lambda * (2)$  et (1) +  $\lambda = \mu$  alors

$$\boxed{Y_{k-1}^T D X_{k-1} X_{k-1}^T D Y_{k-1} q = \lambda^2 q}.$$

(les deux équations encadrées ci-dessus sont appelées **Equations aux dérivées partielles**)

Parmi toutes les solutions du système (cad parmi les vecteurs propre et les valeurs propres), je choisis la valeur propre la plus grande  $\lambda^2$  pour les 2 matrices précédentes  $[X_{k-1}^T D Y_{k-1} Y_{k-1}^T D X_{k-1} w$  ( $p \times p$ ) et  $Y_{k-1}^T D X_{k-1} X_{k-1}^T D Y_{k-1} q$  ( $c \times c$ )].

Pourquoi dois-je choisir la plus grande valeur propre  $\lambda$ ? car  $\lambda = \phi(w, q)$  et que je cherche à maximiser la fonction  $\phi$ .

### 3.2.6 Remarque

En général, il est plus facile de calculer les valeurs propres de la  $2^{ieme}$  matrice  $Y_{k-1}^T D X_{k-1} X_{k-1}^T D Y_{k-1} q$  ( $c \times c$ ) car  $c < p$  et souvent  $c = 1$ , ce sera donc un scalaire et en prenant  $q = 1$ ,  $\lambda^2$  sera ce scalaire (alors en connaissant  $q$ , j'obtiens  $w$ ).

**Problème :** Comment résoudre ce système et trouver la plus grande valeur propre?

Il faudra utiliser un algorithme.

En général,  $D = \frac{1}{n} I_n$ , ce qui simplifie les calculs.

### 3.2.7 Propriétés des composantes $t_1, \dots, t_A$

Les formules d'actualisation des variables conduisent à la relation

$$(t_k, t_l)_D = (t_k, u_l)_D = 0, \forall l > k.$$

La non corrélation ou  $D$ -orthogonalité, mutuelle entre les composantes  $t_1, \dots, t_A$  a de multiples conséquences. On montre ainsi par récurrence que  $t_k$  appartient à  $ImX$  espace vectoriel engendré par les prédicteurs. Plus précisément,  $t_k = X\alpha_k$  avec

$$\begin{aligned} \alpha_1 &= w_1 \\ \alpha_k &= [I_p - \sum_{j=1}^{k-1} \frac{\alpha_j \alpha_j^T}{\|t_j\|_D^2} X^T D X] w_k, \forall k > 1. \end{aligned}$$

La non corrélation implique en outre que  $\sum_{k=1}^{A-1} P_{t_k} = P_{T_A}$ , où  $P_{T_A} = T_A(T_A^T D T_A)^{-1} T_A^T D$  est le projecteur orthogonal sur la matrice  $T_A = [t_1 \parallel \dots \parallel t_A]$ .

Les deux derniers résultats permettent de considérer  $P_{T_A}$  comme le projecteur sur le sous espace de  $ImX$  engendré par les composantes  $t_1, \dots, t_A$ . Dans le cas particulier où  $A = rang(X)$ ,  $P_{T_A} = P_X$ .

Enfin, dernière conséquence de la non corrélation, la décomposition des variances totales des réponses et des prédicteurs fournit deux critères pour le choix du nombre  $A$  de composantes.

### 3.2.8 Le modèle PLS

Les formules d'actualisation entraînent l'écriture des modèles linéaires :

$$\begin{aligned} X &= E_0 = \sum_{k=1}^A \hat{X}_k + E_A = \hat{X}_A + E_A \\ Y &= F_0 = \sum_{k=1}^A \hat{Y}_k + F_A = \hat{Y}_A + F_A \end{aligned}$$

où  $\hat{X}_k = P_{t_k} E_{k-1}$  et  $\hat{Y}_k = P_{t_k} F_{k-1}$  sont les modèles partiels de rang 1.  $\hat{X}_A$  est l'approximation de  $X$  avec une erreur  $E_A$ , idem pour  $\hat{Y}_A$ .

L'actualisation des variables et la non corrélation des composantes conduisent à écrire plus simplement les modèles partiels :  $\hat{X}_k = P_{t_k} X$  et  $\hat{Y}_k = P_{t_k} Y$ . La non corrélation des composantes permet d'une part la décomposition de la variance totale

$$var(Y) \sum_{j=1}^c var(Y^j) = \sum var(\hat{Y}_k) + var(F_A)$$

pour ce qui concerne les réponses. D'autre part, elle conduit à l'écriture définitive des modèles PLS en fonction des composantes

$$\hat{Y}_A = P_{T_A} Y$$

$$\hat{X}_A = P_{T_A} X.$$

Le projecteur s'écrit aussi

$$P_{T_A} = \sum_{k=1}^A \frac{X \alpha_k \alpha_k^T X^T D}{\|t_k\|_D^2}$$

ce qui implique que le modèle PLS est linéaire en les variables explicatives initiales

$$\hat{Y}_A = X \hat{\beta}_A$$

avec

$$\hat{\beta}_A = \sum_{k=1}^A \frac{\alpha_k \alpha_k^T}{\|t_k\|_D^2} X^T D Y.$$

### 3.2.9 Cas particuliers

- Si  $A = \text{rang}(X)$ ,  $PLS(X, Y) = OLS(X, Y)$ .  
Lorsque  $A = \text{rang}(X)$ ,  $t_1, \dots, t_A$  forment une base  $D$ -orthogonale de  $\text{Im}(X)$ . Alors,  $\hat{X}_A = X$  et  $E_A = 0$ .  $X$  est entièrement reconstitué. De plus,  $\hat{Y}_A = P_X Y$ , ce qui signifie que la régression PLS linéaire est équivalente à  $c$  régressions linéaires multiples aux moindres carrés usuels. En règle générale, la régression PLS multiréponses ne conduit pas aux mêmes modèles que ceux obtenus par  $c$  régressions PLS séparées.
- $PLS(X, X) = ACP(X)$ .  
Quand  $Y = X$ , les composantes  $t_k$  et  $u_k$  sont identiques. Le problème qui était de maximiser la covariance entre  $t$  et  $u$ , revient alors à maximiser la variance de  $t$  sous la contrainte  $\|w\|^2 = 1$ . Alors,  $\lambda_1 = \text{var}(t_1)$  est la plus grande valeur propre de  $V = X^T D X$ , matrice des covariances de  $X$ . On peut montrer que  $\lambda_2 = \text{var}(t_2), \dots, \lambda_k = \text{var}(t_k)$ , pour tout  $k = 1, \dots, A$ , et que, par récurrence, les  $w_k$ , qui sont de norme 1, sont les vecteurs propres associés aux valeurs propres  $\lambda_k$  de  $V$  classées en ordre décroissant. On retrouve donc l'Analyse en Composantes Principales de  $X$ .

## 3.3 Les algorithmes PLS

L'algorithme de régression PLS1 possède de nombreuses propriétés mathématiques. Ces propriétés justifient l'algorithme et permettent de le généraliser au cas de plusieurs  $Y$  (régression PLS2).

### 3.3.1 L'algorithme PLS1 NIPALS

Notons  $y$  le vecteur des valeurs de la variable dépendante centrées-réduites et  $X$  la matrice des valeurs des variables explicatives  $x_1, \dots, x_p$

centrées-réduites. La matrice  $X$  est de rang  $a$ . Voici un algorithme de régression PLS1 qui permet la prise en compte des données manquantes selon le principe NIPALS.

**Etape 1 :**  $X_0 = X; y_0 = y$

**Etape 2 :** Pour  $k=1,2,\dots,s$  :

**Etape 2.1 :**  $w_k = \frac{X_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}$

**Etape 2.2 :** Normer  $w_k$  à 1

**Etape 2.3 :**  $t_k = \frac{X_{k-1} w_k}{w_k^T w_k}$

**Etape 2.4 :**  $p_k = \frac{X_{k-1}^T t_k}{t_k^T t_k}$

**Etape 2.5 :**  $X_k = X_{k-1} - t_k p_k^T$

**Etape 2.6 :**  $q_k = \frac{y_{k-1}^T t_k}{t_k^T t_k}$

**Etape 2.7 :**  $u_k = \frac{y_{k-1}}{q_k}$

**Etape 2.8 :**  $y_k = y_{k-1} - q_k t_k$

**Commentaires** Lorsqu'il n'y a pas de données manquantes, on peut remplacer les étapes 2.1 et 2.2 par  $w_k = \frac{X_{k-1}^T y_{k-1}}{\|X_{k-1}^T y_{k-1}\|}$  et l'étape 2.3 par  $t_k = X_{k-1} w_k$ .

### 3.3.2 L'algorithme de régression classique PLS2 (PLS multivarié, $c > 1$ )

Supposons la matrice  $X$  de rang  $a$ .

L'algorithme PLS2 prend la forme suivante :

**Etape 1 :** Posons  $X_0 = X$  et  $Y_0 = Y$

**Etape 2 :** Pour  $k = 1, 2, \dots, a$

**Etape 2.1 :**  $u_k =$  première colonne de  $Y_{k-1}$

**Etape 2.2 :** Répéter jusqu'à convergence de  $w_k$

-  $w_k = \frac{X_{k-1}^T u_k}{u_k^T u_k}$

- Normer  $w_k$  à 1

-  $t_k = \frac{X_{k-1} w_k}{w_k^T w_k}$

-  $q_k = \frac{Y_{k-1}^T t_k}{t_k^T t_k}$

-  $u_k = \frac{Y_{k-1}^T q_k}{q_k^T q_k}$

**Etape 2.3 :**  $P_k = \frac{X_{k-1}^T t_k}{t_k^T t_k}$

**Etape 2.4 :**  $X_k = X_{k-1} - t_k P_k^T$

**Etape 2.5 :**  $Y_k = Y_{k-1} - t_k q_k^T$



### 3.3.3 L'algorithme SIMPLS

Voici une présentation de l'algorithme SIMPLS proposé par de Jong (1993, [4]) pour relier un groupe de réponses  $Y$  à un groupe de prédicteurs  $X$ . L'algorithme SIMPLS est équivalent à la régression PLS lorsque  $Y$  se limite à une seule variable et donne des résultats très proches dans le cas général.

$\forall k = 1, \dots, a$  posons  $A_0 = X^T Y, M_0 = X^T X, C_0 = I$  et  $a$  donné :

**Etape 1 :** Calculer  $q_k$  le vecteur dominant de  $A_k^T A_k$

**Etape 2 :** -  $w_k = A_k q_k$

-  $c_k = w_k^T M_k w_k$

-  $w_k = \frac{w_k}{\sqrt{c_k}}$

- et mettre  $w_k$  en colonne dans la matrice  $W$ .

**Etape 3 :**  $p_k = M_k w_k$  et mettre  $p_k$  en colonne dans la matrice  $P$ .

**Etape 4 :**  $q_k = A_k^T w_k$  et mettre  $q_k$  en colonne dans la matrice  $Q$ .

**Etape 5 :** -  $v_k = c_k P_k$

-  $v_k = \frac{v_k}{\|v_k\|}$

**Etape 6 :** -  $c_{k+1} = c_k - v_k v_k^T$

-  $M_{k+1} = M_k - p_k p_k^T$

**Etape 7 :**  $A_{k-1} = C_k A_k$

Comme dans NIPALS, le  $T$  de SIMPLS est calculé comme  $T = XW$  et  $B$  pour la régression de  $Y$  sur  $X$  comme  $B = WQ^T$ .

## 4 Choix du nombre de composantes

La borne supérieure de  $A$  sera évidemment le rang de la matrice de départ, c'est à dire  $A \leq \text{rang}(X)$ . On aura des critères permettant de déterminer  $A$  de façon raisonnable.

### Critères du choix de $A$

**CR1** Ajustement ("FIT") sur les données

**CR2** Prédiction interne ("Validation Croisée")

**CR3** Prédiction externe sur un jeu de données test ("Validation Externe")

On ne veut pas avoir un surajustement donc le FIT ne devra pas être trop grand ( $A$  trop grand) car cela dégradera la prédiction. Donc il faut une balance entre le critère "ajustement" et le critère "prédiction" (interne ou externe).

### 4.1 Critère 1 : Le "FIT" ou "Critère d'Ajustement"

Le nombre total  $A$  d'axes ou de composantes (la dimension du modèle) est le **super-paramètre** de la méthode. Lors de l'exécution de l'algorithme, il est utile de voir évoluer, à chaque nouvel axe construit, la reconstruction de la variance des prédictors et des réponses. La stratégie consiste à s'arrêter lorsque le pourcentage de la variance de  $X$  est suffisamment grand, pour un gain faible dans le pourcentage de variance de  $Y$ . Ce critère est donc basé sur l'ajustement des données.

En terme d'ajustement il est aussi intéressant de regarder l'évolution des coefficients du modèle de régression linéaire

$$\hat{Y}_A = X\hat{\beta}_A$$

en fonction de la valeur de  $A$ .

Pour une réponse  $j$  choisie, l'étendue des valeurs du vecteur  $\hat{\beta}_A^j$  reste stable jusqu'à une certaine valeur de  $A$ . Au delà de cette valeur  $A$ , pour certains prédictors fortement corrélés, les valeurs des coefficients  $\hat{\beta}_A^j$  semblent exploser. On choisit alors la plus petite valeur  $A$  à partir de laquelle ce phénomène apparaît.

### 4.2 Critère 2 : Le "PRESS" ou "Critère de Validation Croisée"

Après avoir enlevé un individu  $i$  (une ligne) aux matrices  $X$  et  $Y$  (ou 10% des individus si leur nombre est élevé), on calcule  $\hat{\beta}_A^{(i)}$ , qui représente la matrice des coefficients du modèle construit avec les individus restants. Puis, on calcule l'erreur de prédiction faite sur l'individu  $i$ ,  $E_A^{(i)} = Y_i - X_i\hat{\beta}_A^{(i)}$ . L'erreur de prédiction est définie par le PRESS, PREdiction Sum of Squares :

$$PRESS(A) = \frac{1}{n} \sum_{i=1}^n \|E_A^{(i)}\|^2.$$

On choisit donc  $A$  tel que  $PRESS(A)$  soit le plus petit possible.

### 4.3 Critère 3 : "Critère de Validation Externe" et Prédiction

Pour utiliser ce critère de choix de  $A$ , on doit posséder outre l'échantillon d'apprentissage  $(X, Y)$  sur les  $n$  individus, un échantillon test  $(X_{test}, Y_{test})$  sur  $N$  observations mesurées sur les mêmes variables. Différents modèles PLS basés sur  $(X, Y)$  sont calculés pour différentes dimensions. La valeur optimale de  $A$  correspond à la plus petite des erreurs de prédiction calculée sur  $(X_{test}, Y_{test})$ .

## 5 Les logiciels

La régression PLS nécessite des calculs sophistiqués et ses applications dépendent de la validité des logiciels. Pour les chimistes, deux programmes principaux sont utilisés :

- le premier SIMCA-P a été développé par Wold
- le deuxième UNSCRAMBLER a été développé par Martens.

Pour les images cérébrales, SPM, qui est l'un des programmes les plus utilisés dans ce domaine, a récemment intégré un module de régression PLS.

En dehors de ces domaines, SAS PROC PLS est probablement le programme le plus facilement disponible.

Plus récemment, les fonctions **pls** et **plscv** ont été développées sous Splus, et un programme Matlab est aussi disponible.

## 6 Les extensions de la méthode PLS

### 6.1 Les extensions linéaires

Depuis la découverte de la méthode PLS, de nombreux chercheurs ont tenté de l'améliorer ou de l'étendre à d'autres champs d'applications comme par exemple Sjöström et al. (1986, [17]) proposent la PLS-DA (PLS Discriminant Analysis) une méthode généralisant PLS au contexte de la discrimination.

Mais plus récemment de nombreux chercheurs se sont eux aussi penchés sur ce problème d'extensions de la méthode, comme par exemple :

- Vanden Branden et Hubert (2002, [23]) ont proposé une version plus robuste de l'algorithme SIMPLS.
- Nguyen et Rocke (2002, [12]) utilisent la régression PLS dans le domaine de la survie en proposant une méthode "Partial Least Squares Proportional Hazard Regression".
- Vivien et Sabatier (2000 [25], 2001 [26] et à paraître [27]) ont proposé des extensions multitableaux de la régression PLS en introduisant les

méthodes OMCIA-PLS (Orthogonal Multiple Co-Inertia Analysis Partial Least Squares) et GOMCIA-PLS (Generalized OMCIA-PLS).

- Dodge et Whittaker travaillent actuellement sur la "Partial Quantile Regression".

### 6.1.1 "Discrimination Partial Least Squares"

Dans les sciences médicales, biologiques, écologiques et dans de nombreux autres domaines, nous avons souvent besoin d'étudier des groupes que nous voulons séparer dans le but de fixer un rôle de décision. Dans la littérature de nombreuses méthodes ont été proposées pour le problème de la discrimination (Tomassone et al., 1988 [22] ; Saporta, 1990 [16]).

Dans certaines situations, une mauvaise utilisation de l'analyse discriminante PLS (PLS-DA) (Sjörström et al., 1986 [17]) peut donner une solution biaisée qui ne répondra pas au problème de discrimination.

Sabatier et al. (2002, [15]) proposent deux extensions simple :

- La première est proche de la généralisation de PLS proposée par Cazes (1997, [3]) mais pour le contexte de la discrimination.
- La deuxième proposition considère l'Analyse de Redondance PLS proposée par Tenenhaus (1995, [20]) en utilisant une métrique particulière.

### 6.1.2 Une version robuste de l'algorithme SIMPLS

Vanden Branden et Hubert (2002, [23]) ont proposé une version plus robuste de l'algorithme SIMPLS en "robustifiant" les deux étapes de l'algorithme.

Tout d'abord ils proposent d'obtenir les poids  $w_k$  et  $q_k$  comme la première paire des valeurs singulières de la décomposition en valeurs singulières d'une matrice de variance-covariance robuste. pour cela ils appliquent l'algorithme ROBPCA (Hubert et Rousseeuw, 2002 [11]) aux données  $(X, Y)$  et obtiennent une matrice de covariance  $\hat{\Sigma}$  robuste et une moyenne  $\hat{\mu}$  robuste :

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{XX} & \hat{\Sigma}_{XY} \\ \hat{\Sigma}_{YX} & \hat{\Sigma}_{YY} \end{pmatrix}$$

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_X \\ \hat{\mu}_Y \end{pmatrix}.$$

La matrice de covariance classique dans l'algorithme SIMPLS est alors remplacée par  $\hat{\Sigma}_{XY}$ . Les scores  $t_k$  et  $u_k$  sont alors obtenus comme une combinaison linéaire des matrices centrées robustes  $X$  et  $Y$  :

$$t_k = (X - 1_n \hat{\mu}_X^T) w_k$$

$$u_k = (Y - 1_n \hat{\mu}_Y^T) q_k.$$

Une deuxième "robustification" est appliquée à la régression des moindres carrés multivariée. Ils remplacent cette régression par une méthode de régression multivariée robuste appelée "MCD-Regression Method" (Rousseeuw et al., 2000 [14]).

### 6.1.3 "Partial Least Squares Proportional Hazard Regression"

Nguyen et Rocke (2002, [12]) utilisent la régression PLS dans le domaine de la survie en proposant une méthode "Partial Least Squares Proportional Hazard Regression". Ils démontrent comment les prédictions des probabilités de survie des patients peuvent être basées sur le modèle de régression à hasard proportionnel après avoir extrait les composantes des gènes par la méthode PLS.

### 6.1.4 Extensions Multitableaux de la régression PLS

Vivien et Sabatier (2000 [25], 2001 [26] et à paraître [27]) ont proposé des extensions multitableaux de la régression PLS en introduisant les méthodes OMCIA-PLS (Orthogonal Multiple Co-Inertia Analysis Partial Least Squares) et GOMCIA-PLS (Generalized OMCIA-PLS).

La situation est celle où l'on dispose de  $K$  groupes de variables notées  $Y_{k=1\dots K}$  que l'on désire expliquer "globalement" à l'aide d'un seul groupe de prédicteurs, noté  $X$ . Ces  $K + 1$  tableaux étant mesurés sur les mêmes  $n$  individus.

Peu de méthodes relatives à ce problème ont été développées. Wold et al. (1987, [31]) ont proposé une méthode appelée "Multiway PLS", MPLS, qui revient à faire la régression PLS usuelle d'un cube de variables ( $n \times p \times K$ ) réponses en le dépliant en une matrice de taille  $n \times pK$ , sur un cube de variables ( $n \times q \times K'$ ) déplié en matrice de taille  $n \times qK'$ .

Plus récemment, Bro(1996, [2]) a développé la PLS multi-linéaire, N-PLS, qui est basée sur la maximisation d'un critère de covariance généralisant celui de PLS usuelle au cas de cubes de données.

On peut aussi citer la méthode Wold et al. (1996, [33]), la PLS hiérarchie, H-PLS, qui est destinée à expliquer  $K$  sous groupes d'un ensemble de variables réponses par  $K'$  sous groupes d'un ensemble de prédicteurs. Elle calcule des composantes de type PLS pour chaque sous groupe de variables, et fait une régression PLS des composantes réponses sur les composantes explicatives, afin d'obtenir des composantes dites "globales".

La méthode proposée par Vivien et Sabatier (2001, [26]), l'Analyse de Co-Inertie Multiple Orthogonale-Partial Least Squares, ACIMO-PLS, maximise un critère qui peut être vu comme une généralisation à plusieurs tableaux de la régression PLS usuelle de deux tableaux telle que la définit Tenenhaus (1995, [20]). Il se trouve que la solution est un problème aux valeurs propres ce qui présente l'avantage d'être facile à résoudre compte tenu de la stabilité

des algorithmes existants. De plus, les solutions de l'ACIMO-PLS peuvent s'obtenir par le rajout de quelques éléments et de graphiques au programme PLS à deux tableaux. Elle est donc facile à obtenir.

GOMCIA-PLS (Vivien et Sabatier à paraître, [27]) est une nouvelle méthode de composante multiblock et de régression qui généralise OMCIA-PLS et sélectionne les variables en éliminant quelques groupes de variables "bruits" avant d'analyser les données.

### 6.1.5 "Partial Quantile Regression"

La régression décrit comment la valeur moyenne de la variable réponse varie avec les variables explicatives. La régression quantile décrit comment les quantiles de la réponse varient avec les variables explicatives.

Ainsi Dodge et Whittaker travaillent actuellement sur la "Partial Quantile Regression", dont la méthodologie est parallèle à la procédure PLS.

L'algorithme PQR :

- retient les bonnes propriétés de PLSR comme la qualité de la prédiction et la réduction de dimension.
- fournit une nouvelle information qui est le comportement des quantiles de la variable réponse.
- retient la structure algorithmique de PLSR mais remplace les opérateurs espérance et covariance par l'espérance quantile et la covariance quantile.

## 6.2 Les méthodes PLS non linéaires

La régression PLS linéaire étant efficace, dans les années 1980, les chimométriciens ont commencé à développer des méthodes de modélisation non-linéaires basées sur des extensions de la régression PLS.

La régression PLS usuelle contient plusieurs expressions de la linéarité :

- les modèles  $\hat{Y} = X\hat{B}$  sont linéaires en les variables et les coefficients.
- La composante PLS  $t_k = X_{k-1}w_k$  est linéaire en les variables (relation externe).
- On peut montrer que la composante PLS  $u_k$  est linéaire en  $t_k$  :  $u_k = b_k t_k + \text{erreur}$  (relation interne).

Voici une présentation de plusieurs méthodes non-linéaires qui ont été reprises en détails dans le travail de thèse de Myrtille Vivien en 2002 [24].

### 6.2.1 PLS quadratique

La première méthode de régression PLS non-linéaire a été proposée par Wold, Kettaneh-Wold & Skagerberg [32] : Quadratic PLS (QPLS). En régression PLS, il y a une relation linéaire entre les composantes  $t_k$  et  $u_k$ ,

c'est la relation interne. L'idée des auteurs est de délinéariser cette relation par un modèle polynômial d'ordre 2 à l'aide de nouvelles étapes dans l'algorithme de la régression PLS linéaire.

### **6.2.2 Non Linear PLS : NLPLS**

En 1990, Frank propose la méthode de régression non-linéaire dérivée de PLS, appelée Non-Linear PLS.

Il propose de délinéariser la relation interne de PLS à l'aide de lisseurs, à la manière SMART (Smooth Multiple Additive Regression Technique) (Friedman, 1984 [9]), pour approximer la relation interne. Avec SMART, chaque variable réponse est modélisée par une fonction non linéaire d'une combinaison linéaire des prédicteurs.

### **6.2.3 SPLine PLS : SPLPLS**

Wold (1992, [30]), qui s'intéresse toujours à la délinéarisation de la relation interne dans PLS, propose cette fois d'utiliser les fonctions splines linéaires, quadratiques, ou cubiques, plutôt qu'une fonction quadratique comme il l'avait précédemment fait. Cette méthode est désignée par SPL-PLS.

### **6.2.4 Continuum Non Linear PLS : CNLPLS**

Taavitsainen & Korhonen (1992, [19]) et Haario & Taavitsainen (1994, [10]) proposent d'utiliser des critères permettant d'obtenir un continuum de modèles de régression, pour exprimer une liaison non-linéaire entre les composantes PLS  $u_k$  et  $t_k$ .

### **6.2.5 Implicit Non Linear latent variable Regression : INLR**

Berglund & Wold (1997, [1]) proposent une méthode simple et rapide pour modéliser non linéairement un tableau  $Y$  par un tableau prédicteur  $X$ . Il s'agit d'étendre le tableau de prédicteurs avec les termes au carré, et éventuellement les termes d'ordre supérieurs, mais en omettant les termes croisés (car ceux-ci ont tendance à dominer les autres prédicteurs, ajoutent du bruit et engendrent de mauvais modèles). Cette méthode est appelée Implicit non Linear latent variable Regression.

### **6.2.6 La régression Partial Least Squares Splines : PLSS**

La première extension de PLS au modèle non linéaire additif, appelée Additive Spline PL ou ASPLS (Durand & Sabatier, 1997 [8]) est une méthode itérative qui utilise les splines ou polynômes par morceaux, comme dans l'extension non linéaire de l'Analyse en Composantes Principales sur Variables Instrumentales (Durand 1993, [5]).

La deuxième extension est appelée PLS Spline ou PLSS (Durand 1997 [6], 1999 [7]) et produit comme la précédente, un modèle spline additif. Elle présente l'avantage de mettre en oeuvre une méthode non itérative au sens où les composantes principales sont obtenues comme dans PLS par une technique de valeurs propres-vecteurs propres. En fait, PLSS n'est pas autre chose que la régression PLS linéaire entre la (ou les) réponse(s)  $Y$  et la matrice  $\mathcal{B}$  du codage de l'échantillon  $X$  des variables explicatives ou prédicteurs par les fonctions de base de l'espace vectoriel des fonctions splines. Le prix à payer pour le passage au non linéaire étant l'expansion de la dimension colonne de la nouvelle matrice de "design"  $\mathcal{B}$

$$PLSS(X, Y) = PLS(\mathcal{B}, Y).$$

### 6.2.7 "Kernel Partial Least Squares Regression"

L'idée de cette méthode est de transformer les données originales non linéairement dans un espace caractéristique  $\mathcal{F}$  dans lequel un modèle PLS linéaire est alors créé. Les propriétés du modèle PLS correspondant sont alors obtenues par une estimation appropriée des coefficients de régression dans  $\mathcal{F}$  et par la sélection d'une fonction kernel appropriée.

## 7 Conclusion

Depuis sa création, la méthode Partial Least Squares a été étudiée par de nombreux chercheurs et a été adaptée à de nombreuses situations. Cependant, il reste encore beaucoup de choses à développer.

On sait que la méthode PLS est la meilleure dans le cas où il y a plus de prédicteurs que d'observations et qu'il y a une forte colinéarité entre ces prédicteurs, mais comme dans le cas de la régression linéaire classique, il serait intéressant de pouvoir étudier plus précisément la qualité d'un modèle PLS lorsqu'il a été construit.

En développant des tests de validité et en examinant les résidus, on pourrait utiliser plus d'informations que celle donnée par le PRESS et le FIT (voir partie 4) et donc mieux juger de la qualité du modèle. Même si le contexte (peu d'observations, beaucoup de variables, très corrélées) permet de construire un modèle PLS, rien ne dit qu'il sera de bonne qualité.

Un autre problème sur lequel il faudrait se pencher est celui des "outliers". Comment sont-ils traités par la méthode, et quand peut-on dire qu'une observation est aberrante ? Dans le cas de la régression linéaires, un certain nombre de procédures ont été développées, mais sont-elles utilisables dans le cas de la régression PLS ou faut-il en développer d'autres ?

De nombreuses autres questions concernant la régression PLS restent en suspend et la méthode a encore beaucoup de progrès à faire.



## Références

- [1] Berglund, A. & Wold, S. (1997), *INLR, implicit non-linear latent variable regression*, Journal of Chemometrics, 11, 141-156.
- [2] Bro, R. (1996), *Multiway Calibration - Multilinear PLS*, Journal of Chemometrics, 10, 47-61.
- [3] Cazes, P. (1997), *Adaptation de la régression PLS au cas de la régression après analyse des correspondances multiples*, Revue de Statistique Appliquées, XLV, 89-99.
- [4] de Jong, S. (1993), *SIMPLS : An alternative approach to partial least squares regression*, Chemometrics and Intelligent Laboratory Systems, vol. 18, pp. 251-263.
- [5] Durand, J. F. (1993), *Generalized principal component analysis with respect to instrumental variables via univariate spline transformations*, Computational Statistics & Data Analysis, 16, 423-440.
- [6] Durand, J.F (1997), *Additive modeling of multivariate data by spline functions*, Rapport de recherche n97-04, Unité de Biométrie.
- [7] Durand, J. F. (1999), *PLS and multivariate additive spline modeling*, Les méthodes PLS, Symposium International PLS'99, Tenenhaus, M. & Morineau, A. (Eds), CISIA-CERESTA, 1-20.
- [8] Durand, J. F. & Sabatier, R. (1997), *Additive Splines for PLS regression*, Journal of the American Statistical Association, Vol. 92, 440, 1546-1554.
- [9] Friedman, J.H. (1984), *SMART-Technical report n1*, Department of Statistics, Stanford University.
- [10] Haario, H. & Taavitsainen, V.M. (1994), *Nonlinear data Analysis II. Examples on new link functions and optimization aspects*, Chemometrics and Intelligent Laboratory Systems, 23, 51-64.
- [11] Hubert, M. & Rousseeuw, P. J. (2002), *ROBPCA : a new approach to robust principal component analysis*, submitted.
- [12] Nguyen, D. V. & Rocke, D. M. (2002), *Partial Least Squares proportional hazard regression for application to DNA microarray survival data*, Bioinformatics, Vol. 18, p. 1625-1632.
- [13] Pagès, J. & Tenenhaus, M. (2001), *Multiple factor analysis combined with PLS path modeling. Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgments*, Chemometrics and Intelligent Laboratory Systems, 58, 261-273.
- [14] Rousseeuw, P. J., Van Aelst, S., Van Driessen, K. & Agollo, A. (2000), *Robust Multivariate Regression*, submitted.

- [15] Sabatier, R. Vivien, M. & Amenta, P. (2002), *Two approaches for discriminant partial least squares*, 26th Annual Conference of the Gesellschaft für Klassifikation, July 22-24. Mannheim, Germany.
- [16] Saporta, G. (1990), *Probabilité, Analyse de données et Statistique*, Editions Technip, Paris, 493p.
- [17] Sjöström, M., Wold, S. & Söderström, B. (1986), *PLS Discrimination Plots*. In E. S. GELSEMA and L. N. KANALS (Eds) : Pattern Recognition in Practice II. Elsevier, Amsterdam.
- [18] Stone, M. & Brooks, R. (1990), *Continuum regression : Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares, and principal components regression*, Journal of the Royal Statistical Society, Series B, 52(2), 237-269.
- [19] Taavitsainen, V. M. & Korhonen, P. (1992), *Nonlinear data analysis with latent variable*, Chemometrics and Intelligent Laboratory Systems, 14, 184-194.
- [20] Tenenhaus, M. (1995), *A partial least squares approach to multiple regression, redundancy analysis, and canonical analysis*, Les cahiers de la recherche de HEC, CR 550/1995.
- [21] Tenenhaus, M. (1998), *La régression PLS, théorie et pratique* Paris : Technip.
- [22] Tomassone, R., Danzart, M., Daudin, J. J. & Masson, J. P. (1988), *Discrimination et classement*, Masson, Paris, 172 P.
- [23] Vanden Branden, K. & Hubert, M. (2002), *A robustified version of the SIMPLS algorithm*, 26th Annual Conference of the Gesellschaft für Klassifikation, July 22-24. Mannheim, Germany.
- [24] Vivien, M. (2002), *Approches PLS linéaires et non linéaires pour la modélisation multi-tableaux. Théorie et Applications*. Thèse de Doctorat, Université Montpellier I, 2002.
- [25] Vivien, M. & Sabatier, R. (2000), *Nouvelle extension de la régression PLS*. Actes des 6<sup>èmes</sup> journées européennes agro-industrie et méthodes statistiques (pp. 5.0-5.9), Pau, 19-21 janvier.
- [26] Vivien, M. & Sabatier, R. (2001), *Une extension Multi-tableaux de la régression PLS*, Rev. Stat. Appliquée, XLIV(1), 31-54.
- [27] Vivien, M. & Sabatier, R. (A paraître), *Generalized Orthogonal Multiple Co-Inertia Analysis (PLS) : new multiblock component and regression method*, Journal of Chemometrics.
- [28] Wold, H. (1966), *Estimation of principal component and related models by iterative least squares*, *Multivariate Analysis*, ed. P.R. Krishnaiah, Newyork : Academic Press, 391-420.

- [29] Wold, H. (1975), *Soft Modelling by latent variables; the nonlinear iterative partial least squares approach*. In : Perspectives in Probability and Statistics, Gani, J. (Ed), (Papers in honour of M. S. Bartlett). London : Academic Press.
- [30] Wold, S. (1992), *Nonlinear Partial Least Squares Modelling II - Spline Inner Relation*, Chemometrics and Intelligent Laboratory Systems, 14, 71-84.
- [31] Wold, S., Geladi, P., Esbensen, K. & öhman, J. (1987), *Multivariate Principal Components - and PLS - Analysis*, Journal of Chemometrics, 1, 41-56.
- [32] Wold, S., Kettaneh-Wold, N. & Skagerberg, B. (1989), *Nonlinear PLS modeling*, Chemometrics and Intelligent Laboratory Systems, 7, 53-65.
- [33] Wold, S., Kettaneh, N. & Tjessem, K. (1996), *Hierarchical multiblock PLS and PC models for easier interpretation and as an alternative to-variable selection*, Journal of Chemometrics, 10, 463-486.
- [34] Wold, S., Martens, H. & Wold, H. (1983), *The multivariate calibration problem in chemistry solved by the PLS method*, in Proc. Conf. Matrix Pencils, Ruhe A. & Kågström B. (Eds), March 1982, Lecture Notes in Mathematics, Springer Verlag, Heidelberg, PP. 286-293.

#### **Lectures complémentaires**

- [35] Abdi, H. (in press 2003), *Partial Least Squares (PLS) Regression*. In M. Lewis-Beck, A. Bryman, T. Futing (Eds) : Encyclopedia for Research methods for the Social Sciences. Thousand Oaks (CA) : Sage.
- [36] Durand, J. F. , Roman, S. & Vivien, M. (1998), *Guide de la régression PLS linéaire sous S-PLUS*. Rapport de Recherche 98-06, Groupe de Biostatistique et d'Analyses des Systèmes, ENSA.M-INRA-UMII.
- [37] Durand, J. F. (2000), *La régression partial least squares spline-PLSS-Guide d'utilisation sous S-PLUS*, Rapport de Recherche 00-06, Groupe de Biostatistique et d'Analyses des Systèmes, ENSA.M-INRA-UMII.
- [38] Garthwaite, P.H. (1994), *An interpretation of partial least squares*, Journal of the American Statistical Association, 89-122.
- [39] Helland, I.S. (2001), *Some theoretical aspects of partial least squares regression*, Chemometrics and Intelligent Laboratory Systems, 58, 97-107.
- [40] Rosipal, R. & Trejo, L. S. (2001), *Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space*, Journal of Machine Learning Research 2, 97-123.
- [41] Tobias, R. D. (1997), *An introduction to partial least squares regression*, TS-505, SAS.Institute Ins, Cary.N.C, April 1997.

- [42] Wold, S., Sjöström, M. & Eriksson, L. (2001), PLS-regression : a basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems*, 58, 109-130.