

# TUTORIAL DA TEORIA DOS QUADRADOS MÍNIMOS PARCIAIS COM UMA ABORDAGEM GEOMÉTRICA

Fernanda Gomes da SILVEIRA<sup>1 2</sup>  
Lucas Monteiro CHAVES<sup>3</sup>  
Fabyano Fonseca e SILVA<sup>4</sup>

- RESUMO: A abordagem geométrica do método dos Quadrados Mínimos Parciais (PLS) é natural e intuitiva. Esta abordagem explicita as relações existentes entre os métodos de regressão PLS, Quadrados Mínimos Ordinários (OLS) e Componentes Principais (PCR). Em termos de projeções ortogonais, são explicadas as etapas na construção dos algoritmos PLS, populacional e amostral, bem como as construções relativas à regressão em componentes principais.
- PALAVRAS-CHAVE: Projeções; vetores aleatórios; regressão; componentes principais.

## 1 Introdução

O método mais utilizado para obter equações de predição é o método dos quadrados mínimos. Contudo, quando existem fortes relações lineares entre as covariáveis ou, quando o número de covariáveis é maior que o número de observações em uma regressão múltipla, tem-se um problema denominado multicolinearidade, ou um problema de quase-colinearidade. Nesta situação, as estimativas dos coeficientes de regressão por quadrados mínimos tendem a ser instáveis, geralmente com grandes erros-padrão, podendo resultar em inferências errôneas (GARTHWAITE, 1994).

---

<sup>1</sup>Universidade Federal de Lavras – UFLA, Programa de Pós-graduação em Estatística e Experimentação Agropecuária, CEP: 37200-000, Lavras, MG, Brasil. E-mail: *fernanda.gomes@ifmg.edu.br*

<sup>2</sup>Instituto Federal de Minas Gerais – IFMG, Campus Bambuí, Departamento de Ciências e Linguagens, CEP: 38900-000, Bambuí, MG, Brasil.

<sup>3</sup>Universidade Federal de Lavras – UFLA, Departamento de Ciências Exatas, CEP: 37200-000, Lavras, MG, Brasil. E-mail: *lucas@dex.ufla.br*

<sup>4</sup>Universidade Federal de Viçosa – UFV, Departamento de Zootecnia, CEP: 36570-900, Viçosa, MG, Brasil. E-mail: *fabyanofonseca@ufv.br*

Uma forma de contornar o problema da multicolinearidade é a aplicação de métodos de redução de dimensão nas covariáveis. Dentre estes métodos destacam-se, a regressão via componentes principais (*Principal Components Regression* - PCR) e a regressão via método dos Quadrados Mínimos Parciais (*Partial Least Squares* - PLS). Pela sua aplicabilidade, a regressão PLS vem sendo muito utilizada em áreas como quimiometria e genética (AZEVEDO et al., 2013; SINGH et al., 2013).

O método dos quadrados mínimos parciais é geralmente apresentado na forma de um algoritmo, o que torna fácil a sua implementação. No entanto, a teoria subjacente ao método PLS é bastante complexa, utilizando diversas construções de álgebra linear e vários conceitos estatísticos de regressão. Muitos estudos são encontrados na literatura com aplicações utilizando o método PLS, mas são raras as referências em que se aborda a teoria do PLS, em especial a abordagem geométrica. Trabalhos contemplando uma abordagem geométrica são úteis para o entendimento e desenvolvimento do método.

As referências mais acessíveis relacionadas ao método PLS e à descrição dos algoritmos são os artigos Helland (1990), Hoskuldsson (1988), Garthwaite (1994) e Phatak e Jong (1997). Enquanto os dois primeiros dão um enfoque algébrico, explicitando as propriedades do método, os dois últimos têm um enfoque geométrico, mas com poucas demonstrações. Assim, este trabalho pretende ser uma referência em português para aqueles que querem compreender os aspectos teóricos do PLS. Nenhum novo resultado é apresentado, porém, a abordagem totalmente geométrica pode ser considerada como uma contribuição, pelo menos no aspecto didático, à teoria. Além disso, este trabalho busca unir as abordagens dos quatro autores, fazendo ainda, um paralelo entre o PLS Populacional e o PLS Amostral. Como o método PLS pode ser, de certa forma, visto como uma generalização da análise de componentes principais, esta também é apresentada de forma geométrica.

## 2 Componentes principais

A análise de componentes principais (*Principal Component Analysis* - PCA) é um dos métodos de redução de dimensionalidade, visando solucionar eventuais problemas de multicolinearidade da matriz  $\mathbf{X}$  e a necessidade de um número excessivo de amostras para a construção de um modelo viável, sem acarretar na perda significativa de informações presentes nos dados (OTTO, 1999).

A PCA é obtida com base nos autovalores e autovetores da matriz de covariâncias. As coordenadas dos autovetores determinam combinações lineares das variáveis originais e são conhecidos como *loadings*. Portanto, o número de componentes principais é igual ao número de autovetores, sendo este, igual ao número de covariáveis. Um autovalor associado a um autovetor é a variância da componente principal relativa a esse autovetor. Assim, as componentes principais relativas aos autovetores associados aos maiores autovalores, contêm a maior parte da explicação da variabilidade dos dados. Portanto, a análise pode ser feita com um número de componentes principais menor que o número de variáveis, levando a uma redução significativa da dimensionalidade na análise.

### 3 Regressão por componentes principais

A regressão via componentes principais foi introduzida por Kendall (1957) e Hotelling (1957). Este método faz uso dos procedimentos empregados na análise de componentes principais visando contornar alguns problemas apresentados pela regressão linear múltipla.

O método PCR consiste, basicamente, em eliminar componentes que não contribuam na explicação da variância presente nos dados, o que reduz a dimensionalidade dos dados originais. Após a escolha do número ótimo de componentes a serem utilizados, os coeficientes de regressão podem ser determinados por meio do método dos quadrados mínimos ordinários.

Como em qualquer regressão múltipla, a escolha do número de covariáveis a serem incluídas no modelo é de extrema importância, tendo em vista evitar a perda de informações relevantes. Segundo Roggo et al. (2007), cada componente descreve uma fração da variação total contida nos dados, tornando possível a determinação do número ótimo de componentes a serem incluídos na regressão.

Regressão em componentes principais é colocado neste trabalho em razão das semelhanças com o PLS. Na linha deste trabalho, a exposição da teoria do PCR será geométrica e, portanto, diferente da abordagem usual na literatura.

Os componentes são obtidos pelas direções no subespaço gerado pelas observações das covariáveis que fornecem a maior variabilidade dos dados. Na construção dos componentes, utiliza-se apenas as covariáveis, e estes componentes representam as novas variáveis a serem utilizadas na regressão. Os coeficientes da regressão do vetor de dados nas componentes principais são estimados pelo método dos quadrados mínimos.

Geometricamente, será demonstrado que os componentes são representados como imagens dos autovetores da matriz  $\mathbf{X}'\mathbf{X}$ . Obtendo o subespaço gerado por estes componentes, a regressão é feita pela projeção do vetor de dados neste subespaço. Também é demonstrado que o vetor de coeficientes estimado via Regressão em Componentes Principais,  $\hat{\beta}_{\text{PCR}}$ , é um estimador de encolhimento.

### 4 Regressão por quadrados mínimos parciais

O método dos quadrados mínimos parciais (PLS) foi introduzido por Wold em 1975, sendo considerado útil para a construção de equações de predições em situações nas quais se tem um grande número de variáveis explicativas e um número relativamente pequeno de dados amostrais (HOSKULDSSON, 1988).

Resumidamente, a ideia geral do PLS é formar componentes, isto é, combinações lineares das covariáveis, que capturem a maior quantidade de informação possível disposta nas covariáveis ( $X_1, \dots, X_p$ ) para predizer as variáveis respostas ( $Y_1, \dots, Y_k$ ). O método PLS apresenta similaridades com o tradicional método de Regressão via Componentes Principais, sendo a maior diferença entre eles dada pelo fato do PCR levar em consideração, na construção dos componentes, apenas as covariáveis, enquanto que o PLS também leva em consideração as variáveis

respostas. No caso de apenas uma variável resposta, o PLS é univariado, e para o caso de várias variáveis respostas, tem-se o PLS multivariado (GARTHWAITE, 1994).

A teoria dos Quadrados Mínimos Parciais é uma ferramenta para resolver problemas de regressão nos quais as variáveis preditoras (ou covariáveis) apresentam um alto nível de colinearidade. Neste caso, a matriz das covariáveis  $\mathbf{X}$  pode ser de posto incompleto. Outra situação, muito comum em aplicações na área de química, ocorre quando o número de covariáveis é muito maior que o número de observações. Em ambas as situações, as equações normais para obtenção do estimador de quadrados mínimos admitem infinitas soluções. Para a escolha de uma estimativa em particular, um método possível é o uso de inversas generalizadas. No entanto, a escolha desta inversa é, de certa forma, arbitrária, gerando situações indesejáveis para as equações de predição obtidas. No método de regressão utilizando PLS, é obtida uma solução única para os parâmetros de regressão.

A ideia básica do PLS é regressar os vetores respostas  $\mathbf{Y} = (Y_1, \dots, Y_k)$ , em cada um dos vetores das covariáveis e, através de escolha de pesos, obter combinações lineares destas regressões. Desta forma, a matriz de dados  $\mathbf{Y}$  é explicado em termos de um número menor de vetores. A escolha destas combinações lineares é justificada estatisticamente, pela maximização de covariâncias.

A estimativa do vetor de parâmetros através do método PLS,  $\hat{\beta}_{\text{PLS}}$ , admite uma expressão como uma transformação linear da estimativa  $\hat{\beta}_{\text{OLS}}$ , invariante em relação ao  $\hat{\beta}_{\text{OLS}}$  escolhido. O método possui uma natureza geométrica sendo obtido por meio de projeções oblíquas, diferentemente do método dos quadrados mínimos, que utiliza projeções ortogonais.

## 5 Regressão por componentes principais: uma abordagem geométrica

Esta seção tem como referência o artigo Phatak e Jong (1997). Para o problema de regressão em que se tem o modelo  $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ , a ideia básica é obter a direção, no subespaço  $\text{Im}\mathbf{X}$ , que fornece a maior variabilidade de  $\mathbf{Y}$ .

Neste caso, o problema de multiplicadores de Lagrange agora é mais complicado uma vez que se têm duas restrições ( $\|\beta\| = 1$  e subespaço  $\text{Im}\mathbf{X}$ ). No entanto, pode-se modificar o problema para maximizar  $\text{var}(\mathbf{X}\beta \cdot \mathbf{Y})$  restrito a  $\beta'\beta = 1$ . Isto é, maximizar  $(\mathbf{X}\beta)'\Sigma(\mathbf{X}\beta)$ , restrito a  $\|\beta\| = 1$ , ou seja,  $\beta'\mathbf{X}'\Sigma\mathbf{X}\beta$ , restrito a  $\|\beta\| = 1$ , como mostrado na Figura 1.

Este problema é semelhante ao anteriormente feito, com a direção de maior variação dada agora no espaço paramétrico. As direções são dadas pelos autovetores de  $\mathbf{X}'\Sigma\mathbf{X}$ . O caso mais simples ocorre quando  $\Sigma = \sigma^2\mathbf{I}$  (observações não correlacionadas e variância constante). Neste caso as direções são dadas pelos autovetores de  $\mathbf{X}'\mathbf{X}$ .

Sejam  $\gamma_1, \gamma_2, \dots, \gamma_p$  os autovetores e  $\lambda_1, \lambda_2, \dots, \lambda_p$  os respectivos autovalores, da matriz  $\mathbf{X}'\mathbf{X}$ . Assim,  $\mathbf{X}'\mathbf{X}\gamma_i = \lambda_i\gamma_i$  com  $\gamma_i'\gamma_i = 1$  e  $\gamma_i'\gamma_j = 0$  para  $i \neq j : 1, \dots, p$  (Figura 2).

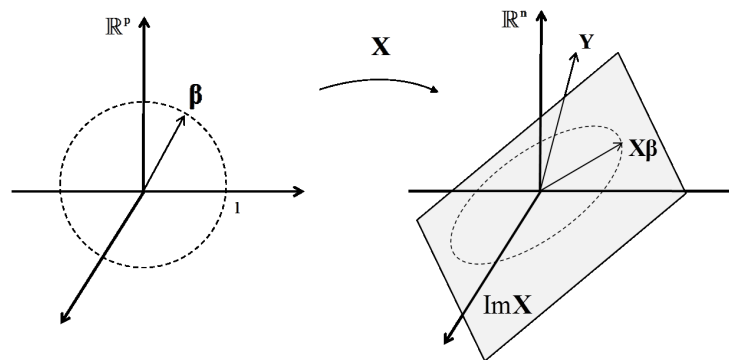


Figura 1 - Representação geométrica da maximização da  $\text{var}(\mathbf{X}\boldsymbol{\beta} \cdot \mathbf{Y})$  restrito a  $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ .

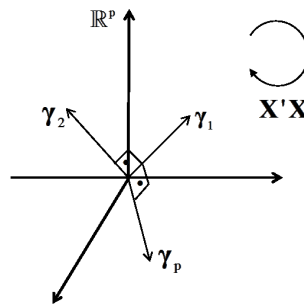


Figura 2 - Representação geométrica dos autovetores  $\gamma_1, \gamma_2, \dots, \gamma_p$  da matriz  $\mathbf{X}'\mathbf{X}$ .

Note que

$$\text{var}(\mathbf{X}\boldsymbol{\gamma}_i \cdot \mathbf{Y}) = (\mathbf{X}\boldsymbol{\gamma}_i)' (\sigma^2 \mathbf{I}) (\mathbf{X}\boldsymbol{\gamma}_i) = \sigma^2 \boldsymbol{\gamma}_i' \mathbf{X}'\mathbf{X}\boldsymbol{\gamma}_i = \sigma^2 \boldsymbol{\gamma}_i' (\lambda_i \boldsymbol{\gamma}_i) = \sigma^2 \lambda_i$$

e que os vetores

$\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \dots, \mathbf{X}\boldsymbol{\gamma}_p$  são ortogonais, pois:

$$\mathbf{X}\mathbf{X}' (\mathbf{X}\boldsymbol{\gamma}_i) = \mathbf{X} (\mathbf{X}'\mathbf{X}) \boldsymbol{\gamma}_i = \mathbf{X} \lambda_i \boldsymbol{\gamma}_i = \lambda_i (\mathbf{X}\boldsymbol{\gamma}_i).$$

Assim, os  $\mathbf{X}\boldsymbol{\gamma}_i$  são autovetores de  $\mathbf{X}\mathbf{X}'$ , portanto ortogonais entre si.

Ordenando os autovalores  $\lambda_1 > \lambda_2 > \dots > \lambda_p$ , define-se como componentes principais as direções dadas por  $\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \mathbf{X}\boldsymbol{\gamma}_3, \dots$ . Como estas direções apresentam uma ordem decrescente de variação do vetor aleatório  $\mathbf{Y}$ , elas “explicam” bem o

vetor  $\mathbf{Y}$ . A ideia, então, é formar os subespaços (Figura 3):

$$\begin{aligned} W_1 &= \text{span} \{ \mathbf{X}\gamma_1 \} \\ W_2 &= \text{span} \{ \mathbf{X}\gamma_1, \mathbf{X}\gamma_2 \} \\ W_3 &= \text{span} \{ \mathbf{X}\gamma_1, \mathbf{X}\gamma_2, \mathbf{X}\gamma_3 \} \\ &\dots \\ W_m &= \text{span} \{ \mathbf{X}\gamma_1, \mathbf{X}\gamma_2, \mathbf{X}\gamma_3, \dots, \mathbf{X}\gamma_m \} \\ \text{Im}\mathbf{X} &= \text{span} \{ \mathbf{X}\gamma_1, \mathbf{X}\gamma_2, \mathbf{X}\gamma_3, \dots, \mathbf{X}\gamma_p \}. \end{aligned}$$

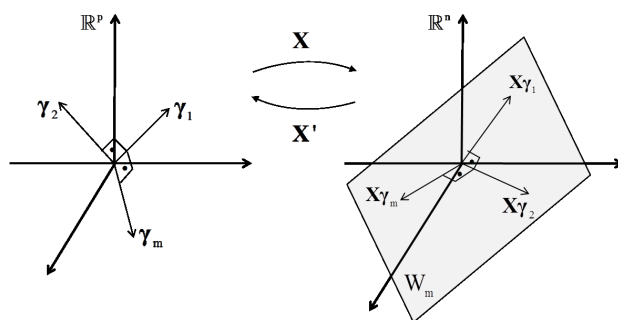


Figura 3 - Subespaço  $W_m$  gerado pelos  $m$ -primeiros autovetores da matriz  $\mathbf{X}'\mathbf{X}$ .

Como estas direções explicam a variabilidade de  $\mathbf{Y}$ , a ideia é projetar  $\mathbf{Y}$  ortogonalmente no subespaço  $W_m$ , isto é, regredir  $\mathbf{Y}$  aos  $m$ -primeiros componentes principais. A projeção define, então, o vetor das estimativas dos parâmetros de regressão denominado  $\hat{\beta}_{\text{PCR}}$ , cujas equações normais são dadas por  $P_{W_m}\mathbf{Y} = \mathbf{X}\hat{\beta}_{\text{PCR}}$ . Geometricamente, fica como representado na Figura 4.

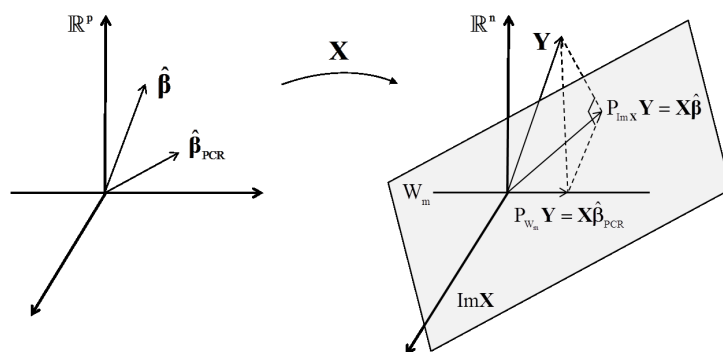


Figura 4 - Projeção ortogonal de  $\mathbf{Y}$  no subespaço  $W_m$  definindo o parâmetro de regressão  $\hat{\beta}_{\text{PCR}}$ .

O triângulo retângulo formado pelos vetores  $P_{\text{Im}\mathbf{X}}\mathbf{Y}$ ,  $P_{W_m}\mathbf{Y}$  e pelo vetor dado pela diferença entre eles, garante que  $\|P_{W_m}\mathbf{Y}\| \leq \|P_{\text{Im}\mathbf{X}}\mathbf{Y}\|$  e, apesar deste fato não implicar em  $\|\hat{\beta}_{\text{PCR}}\| < \|\hat{\beta}\|$ , o procedimento da regressão em componentes principais é claramente conservador, no sentido que fornece estimativas menores para o vetor de parâmetros  $\beta$ . De fato, será demonstrado que  $\|\hat{\beta}_{\text{PCR}}\| < \|\hat{\beta}\|$ . Resta então, obter expressões matriciais para as equações normais, isto é, expressar o projetor  $P_{W_m}$  matricialmente.

A construção será heurística baseada no argumento: seja  $f : A \rightarrow B$  uma função injetiva entre os conjuntos  $A$  e  $B$ . O que se quer é uma aplicação  $g : B \rightarrow B$  que seja uma projeção em  $\text{Im}f$ , isto é,  $g(f(x)) = f(x)$ ,  $\forall x \in A$  e  $g(g(y)) = g(y)$ ,  $\forall y \in B$ , como mostra a Figura 5.

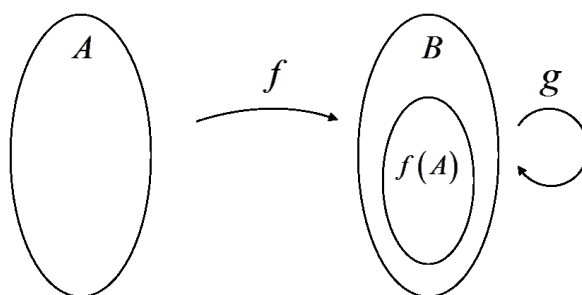


Figura 5 - Representação da função injetiva  $f$  entre os conjuntos  $A$  e  $B$  e da função  $g$  como uma projeção na  $\text{Im}f$ .

A ideia básica é que a composição de uma função injetiva com uma sobrejetiva define uma bijeção. Seja  $h : B \rightarrow A$  uma função sobrejetiva tal que  $h$  restrita à imagem de  $f$  seja injetiva,  $h \circ f : A \rightarrow A$  é uma bijeção (Figura 6).

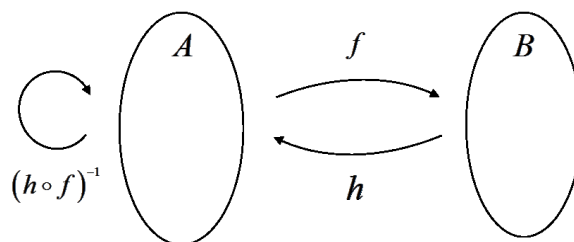


Figura 6 - Representação da composição de uma função  $h$  sobrejetiva com uma função  $f$  injetiva, definindo uma bijeção  $h \circ f$ .

Define-se então,

$$g = f \circ (h \circ f)^{-1} \circ h. \quad (1)$$

$$\begin{aligned} g \circ g &= (f \circ (h \circ f)^{-1} \circ h) \circ (f \circ (h \circ f)^{-1} \circ h) \\ &= f \circ (h \circ f)^{-1} \circ (h \circ f) \circ (h \circ f)^{-1} \circ h \\ &= f \circ (h \circ f)^{-1} \circ h = g \quad \text{e} \end{aligned}$$

$$\begin{aligned} g(f(x)) &= f \circ (h \circ f)^{-1} \circ h(f(x)) \\ &= f \circ (h \circ f)^{-1} (h \circ f)(x) = f(x). \end{aligned}$$

Portanto,  $g$  é uma projeção em  $\text{Im}f$ , como procurado. Basta agora fazer a construção de  $g$  utilizando as funções lineares convenientes para se obter a projeção linear em  $W_m$ .

Sejam  $\xi_1 = X\gamma_1, \xi_2 = X\gamma_2, \dots, \xi_m = X\gamma_m$ , como na Figura 7.

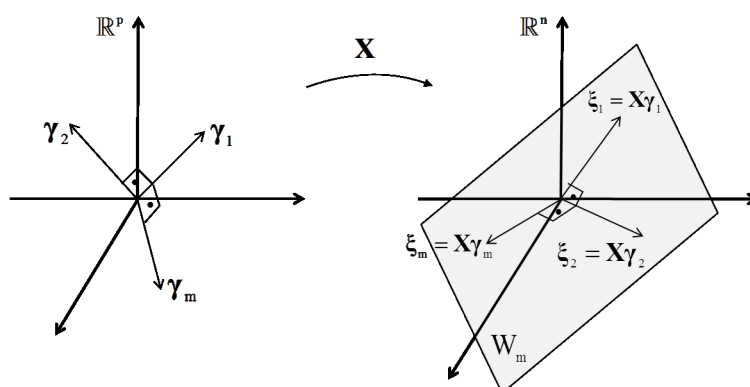


Figura 7 - Representação dos vetores  $\xi_1, \xi_2, \dots, \xi_m$  como imagem dos autovetores  $\gamma_1, \gamma_2, \dots, \gamma_m$  pela matriz  $X$ .

Seja  $\Xi_m$  a matriz  $n \times m$  cujas colunas são os vetores  $\xi_i$ ,

$$\Xi_m = [\xi_1, \xi_2, \dots, \xi_m].$$

A matriz  $\Xi_m$  define uma transformação linear  $\Xi_m : \mathbb{R}^m \rightarrow \mathbb{R}^n$  tal que  $\Xi_m e_i = \xi_i$ , em que  $e_i$  são os vetores canônicos. Desta forma,  $\Xi_m$  é injetiva de posto  $m$  e a imagem de  $\Xi_m$  é o subespaço  $W_m$  (Figura 8).



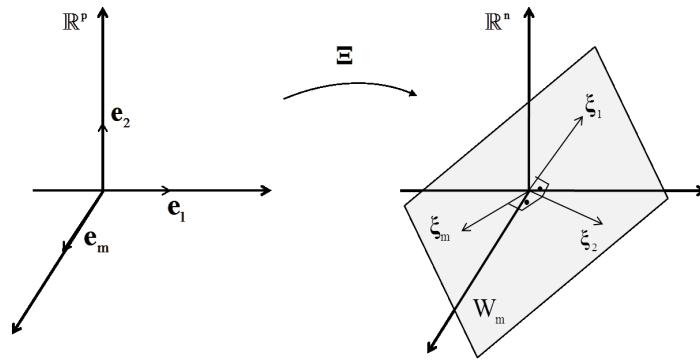


Figura 8 - Representação da transformação linear  $\Xi_m : \mathbb{R}^m \rightarrow \mathbb{R}^n$  em que  $\Xi_m e_i = \xi_i$ .

Considere agora  $\Xi'_m : \mathbb{R}^n \rightarrow \mathbb{R}^m$  a transformação linear definida pela matriz transposta de  $\Xi_m$ ,

$$\Xi'_m = \begin{bmatrix} \xi'_1 \\ \xi'_2 \\ \vdots \\ \xi'_m \end{bmatrix}.$$

Como o posto linha é igual ao posto coluna, tem-se que  $\Xi'_m$  é uma aplicação sobrejetiva.  $\Xi'_m$  restrito à  $\text{Im} \mathbf{X}$  deve ser injetiva. De fato,

$$\begin{bmatrix} \xi'_1 \\ \xi'_2 \\ \vdots \\ \xi'_i \\ \vdots \\ \xi'_m \end{bmatrix} [\xi_i] = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \|\xi_i\|^2 \\ \vdots \\ 0 \end{bmatrix}.$$

Podemos, então, definir o projetor

$$P_{W_m} = \Xi_m (\Xi'_m \Xi_m)^{-1} \Xi'_m,$$

de acordo com a equação (1). De onde seguem as equações normais do estimador de componentes principais,  $\hat{\beta}_{\text{PCR}}$  :

$$\mathbf{X} \hat{\beta}_{\text{PCR}} = P_{W_m} \mathbf{Y} = \Xi_m (\Xi'_m \Xi_m)^{-1} \Xi'_m \mathbf{Y}.$$

Pode-se obter uma relação entre os estimadores  $\hat{\beta}_{\text{PCR}}$  e  $\hat{\beta}_{\text{OLS}}$ . Seja  $\Gamma_m$  a matriz  $p \times m$  definida por  $\Gamma_m = [\gamma_1, \gamma_2, \dots, \gamma_m]$ . Tem-se que:

$$\mathbf{X}\Gamma_m = \mathbf{X}[\gamma_1, \gamma_2, \dots, \gamma_m] = [\mathbf{X}\gamma_1, \mathbf{X}\gamma_2, \dots, \mathbf{X}\gamma_m] = \Xi_m.$$

Substituindo  $\Xi_m = \mathbf{X}\Gamma_m$  na definição de  $\hat{\beta}_{\text{PCR}}$ ,

$$\begin{aligned} \mathbf{X}\hat{\beta}_{\text{PCR}} &= \mathbf{X}\Gamma_m (\Xi_m' \Xi_m)^{-1} \Gamma_m' \mathbf{X}' \mathbf{Y} \\ &= \mathbf{X}\Gamma_m (\Xi_m' \Xi_m)^{-1} \Gamma_m' (\mathbf{X}' \mathbf{X}) (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} \\ &= \mathbf{X}\Gamma_m (\Xi_m' \Xi_m)^{-1} (\mathbf{X}\Gamma_m)' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} \\ &= \mathbf{X}\Gamma_m (\Xi_m' \Xi_m)^{-1} (\mathbf{X}\Gamma_m)' \mathbf{X} \hat{\beta}_{\text{OLS}}. \end{aligned}$$

Como os  $\xi_i$  são ortogonais,  $\Xi_m' \Xi_m = \Lambda_m$  é uma matriz diagonal

$$\Lambda_m = \begin{bmatrix} \|\xi_1\|^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \|\xi_m\|^2 \end{bmatrix}.$$

Assim,

$$\begin{aligned} &\mathbf{X}\Gamma_m (\Xi_m' \Xi_m)^{-1} (\mathbf{X}\Gamma_m)' \\ &= \mathbf{X}\Gamma_m \Lambda_m^{-1} (\mathbf{X}\Gamma_m)' \\ &= [\xi_1, \xi_2, \dots, \xi_m] \begin{bmatrix} \|\xi_1\|^{-2} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \|\xi_m\|^{-2} \end{bmatrix} \begin{bmatrix} \xi_1' \\ \xi_2' \\ \vdots \\ \xi_m' \end{bmatrix} \\ &= [\xi_1, \xi_2, \dots, \xi_m] \begin{bmatrix} \xi_1' / \|\xi_1\|^2 \\ \xi_2' / \|\xi_2\|^2 \\ \vdots \\ \xi_m' / \|\xi_m\|^2 \end{bmatrix}. \end{aligned}$$

$$\text{Portanto, } \mathbf{X}\hat{\beta}_{\text{PCR}} = [\xi_1, \xi_2, \dots, \xi_m] \begin{bmatrix} \xi_1' / \|\xi_1\|^2 \\ \xi_2' / \|\xi_2\|^2 \\ \vdots \\ \xi_m' / \|\xi_m\|^2 \end{bmatrix} \mathbf{X}\hat{\beta}_{\text{OLS}}.$$

Utilizando a decomposição de  $\mathbf{X}$  da forma:

$$\mathbf{X} = \xi_1 \gamma'_1 + \xi_2 \gamma'_2 + \dots + \xi_p \gamma'_p,$$

$$\begin{aligned} & [\xi_1, \xi_2, \dots, \xi_m] \begin{bmatrix} \xi'_1 / \|\xi_1\|^2 \\ \xi'_2 / \|\xi_2\|^2 \\ \vdots \\ \xi'_m / \|\xi_m\|^2 \end{bmatrix} (\xi_1 \gamma'_1 + \xi_2 \gamma'_2 + \dots + \xi_p \gamma'_p) \hat{\beta}_{OLS} \\ &= [\xi_1, \xi_2, \dots, \xi_m] \left\{ \sum_{i=1}^p \begin{bmatrix} \xi'_1 / \|\xi_1\|^2 \\ \xi'_2 / \|\xi_2\|^2 \\ \vdots \\ \xi'_m / \|\xi_m\|^2 \end{bmatrix} \xi_i \gamma'_i \right\} \hat{\beta}_{OLS} \\ &= [\xi_1, \xi_2, \dots, \xi_m] \left\{ \sum_{i=1}^p \frac{\xi'_i}{\|\xi_i\|^2} \xi_i \gamma'_i \right\} \hat{\beta}_{OLS} \\ &= [\xi_1, \xi_2, \dots, \xi_m] \Gamma'_m \hat{\beta}_{OLS} \\ &= \mathbf{X} \Gamma_m \Gamma'_m \hat{\beta}_{OLS}. \end{aligned}$$

Logo,  $\mathbf{X} \hat{\beta}_{PCR} = \mathbf{X} (\Gamma_m \Gamma'_m) \hat{\beta}_{OLS} \Rightarrow \hat{\beta}_{PCR} = (\Gamma_m \Gamma'_m) \hat{\beta}_{OLS}$ . Como os vetores  $\gamma_i$  são ortonormais,  $\Gamma'_m \Gamma_m = \mathbf{I}_m$  e  $\hat{\beta}_{PCR} = \Gamma_m (\Gamma'_m \Gamma_m)^{-1} \Gamma'_m \hat{\beta}_{OLS}$ , portanto,  $\hat{\beta}_{PCR}$  é obtido como a projeção ortogonal do vetor  $\hat{\beta}_{OLS}$  no subespaço gerado por  $\{\gamma_1, \gamma_2, \dots, \gamma_m\}$ , como observado na Figura 9.

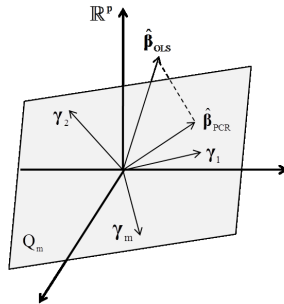


Figura 9 - Representação do parâmetro  $\hat{\beta}_{PCR}$  como a projeção ortogonal do vetor  $\hat{\beta}_{OLS}$  no subespaço gerado por  $\{\gamma_1, \gamma_2, \dots, \gamma_m\}$ .

Em particular, fica demonstrado que  $\|\hat{\beta}_{\text{PCR}}\| < \|\hat{\beta}_{\text{OLS}}\|$ , logo,  $\hat{\beta}_{\text{PCR}}$  é um estimador de encolhimento.

Outra maneira mais geométrica de se ver esta relação é: considerando  $\{\gamma_1, \gamma_2, \dots, \gamma_p\}$  como base no domínio e  $\{\xi_1, \xi_2, \dots, \xi_p\}$  como base no contradomínio. A transformação linear  $\mathbf{X}$  se expressa como a matriz identidade

$$\mathbf{X} \left( \sum_{i=1}^p a_i \gamma_i \right) = \sum_{i=1}^p a_i \xi_i.$$

Projetar  $\mathbf{Y}$  ortogonalmente na  $\text{Im}\mathbf{X}$  e depois projetar ortogonalmente em  $W_m$  é o mesmo que projetar ortogonalmente  $\mathbf{Y}$  em  $W_m$ . A demonstração deste fato segue da aplicação do Teorema de Pitágoras nos vários triângulos retângulos da Figura 10.

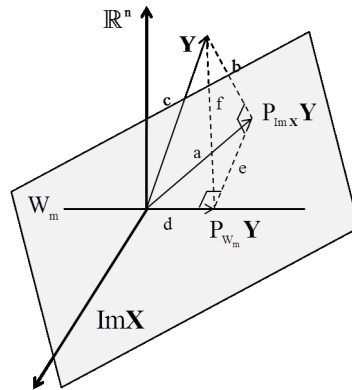


Figura 10 - Projeção do vetor de dados  $\mathbf{Y}$  no subespaço gerado pelos  $m$  primeiros componentes principais.

Portanto, se  $\mathbf{X}\hat{\beta}_{\text{OLS}} = P_{\text{Im}\mathbf{X}}\mathbf{Y} \Rightarrow \mathbf{X}(\sum a_i \gamma_i) = \sum a_i \xi_i$ , projetar  $P_{\text{Im}\mathbf{X}}\mathbf{Y}$  em  $W_m$  é simplesmente tomar as  $m$  primeiras coordenadas,  $P_{W_m}\mathbf{Y} = \sum_{i=1}^m a_i \xi_i$  e  $\mathbf{X}\hat{\beta}_{\text{PCR}} = P_{W_m}\mathbf{Y}$  implica que  $\hat{\beta}_{\text{PCR}} = \sum_{i=1}^m a_i \gamma_i$  e assim,  $\hat{\beta}_{\text{PCR}}$  é a projeção de  $\hat{\beta}_{\text{OLS}}$  no subespaço gerado por  $\{\gamma_1, \gamma_2, \dots, \gamma_m\}$ .

A matriz de variâncias e covariâncias de  $\hat{\beta}_{\text{PCR}}$  é:

$$\begin{aligned} \text{cov}(\hat{\beta}_{\text{PCR}}) &= \text{cov}(\mathbf{\Gamma}_m \mathbf{\Gamma}'_m \hat{\beta}_{\text{OLS}}) \\ &= \mathbf{\Gamma}_m \mathbf{\Gamma}'_m \text{cov}(\hat{\beta}_{\text{OLS}}) \mathbf{\Gamma}_m \mathbf{\Gamma}'_m \\ &= \sigma^2 \mathbf{\Gamma}_m \mathbf{\Gamma}'_m (\mathbf{X}'\mathbf{X})^{-1} \mathbf{\Gamma}_m \mathbf{\Gamma}'_m. \end{aligned}$$

Matricialmente, tem-se:

$$\begin{aligned}
 & [\gamma_1, \gamma_2, \dots, \gamma_m] \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_m \end{bmatrix} (\mathbf{X}'\mathbf{X})^{-1} [\gamma_1, \gamma_2, \dots, \gamma_m] \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_m \end{bmatrix} \\
 &= [\gamma_1, \gamma_2, \dots, \gamma_m] \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_m \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1} \gamma_1, \frac{1}{\lambda_2} \gamma_2, \dots, \frac{1}{\lambda_m} \gamma_m \end{bmatrix} \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_m \end{bmatrix} \\
 &= [\gamma_1, \gamma_2, \dots, \gamma_m] \begin{bmatrix} 1/\lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1/\lambda_m \end{bmatrix} \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_m \end{bmatrix} \\
 &= \begin{bmatrix} \gamma_1/\lambda_1, \gamma_2/\lambda_2, \dots, \gamma_m/\lambda_m \end{bmatrix} \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_m \end{bmatrix}.
 \end{aligned}$$

Essa matriz possui  $\gamma_i$  como autovetores, pois,

$$\begin{aligned}
 & \left( \begin{bmatrix} \gamma_1/\lambda_1, \gamma_2/\lambda_2, \dots, \gamma_m/\lambda_m \end{bmatrix} \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_m \end{bmatrix} \right) (\gamma_i) \\
 &= \begin{bmatrix} \gamma_1/\lambda_1, \gamma_2/\lambda_2, \dots, \gamma_m/\lambda_m \end{bmatrix} \left( \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_m \end{bmatrix} (\gamma_i) \right) \\
 &= \begin{bmatrix} \gamma_1/\lambda_1, \gamma_2/\lambda_2, \dots, \gamma_m/\lambda_m \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \frac{1}{\lambda_i} \gamma_i
 \end{aligned}$$

e, portanto, a variância total de  $\hat{\beta}_{\text{PCR}}$  é  $\sum_{i=1}^m \frac{1}{\lambda_i}$ .

Para o caso em que  $\text{cov}(\mathbf{Y}) = \Sigma$ , o estimador de menor variância é dado pelo

estimador de Gauss-Markov (PEREIRA, 2013):

$$\begin{aligned}\hat{\beta}_{\text{GM}} &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y} \\ \mathbf{X}\hat{\beta}_{\text{GM}} &= \mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}.\end{aligned}$$

Assim,  $\hat{\mathbf{Y}}$  é dado pela projeção de  $\mathbf{Y}$ , em que  $\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}$  é o projetor.

No caso usual, em que  $\text{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$ , as projeções nos subespaços  $W_m$  que definem  $\hat{\beta}_{\text{PCR},m}$  são ortogonais, e quando  $m = p$ ,  $\hat{\beta}_{\text{PCR}} = \hat{\beta}_{\text{OLS}}$ . Portanto, no caso  $\text{cov}(\mathbf{Y}) = \Sigma$ , é natural utilizar a mesma projeção. O projetor  $\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}$  é idempotente, mas não é simétrico, logo, não é um projetor ortogonal. No entanto, em relação ao produto interno  $\langle\langle\mathbf{x},\mathbf{y}\rangle\rangle = \mathbf{x}'\Sigma^{-1}\mathbf{y}$ , este projetor é simétrico. Assim, o projetor  $\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}$  é ortogonal em relação ao produto interno  $\langle\langle\mathbf{x},\mathbf{y}\rangle\rangle = \mathbf{x}'\Sigma^{-1}\mathbf{y}$ .

Deseja-se então, maximizar a variância da projeção de  $\mathbf{Y}$  na direção de  $\mathbf{X}\beta$ , isto é, maximizar  $\text{var}(\langle\langle\mathbf{X}\beta,\mathbf{Y}\rangle\rangle)$ , restrito a  $\beta'\beta = 1$ :

$$\begin{aligned}\text{var}((\mathbf{X}\beta)'\Sigma^{-1}\mathbf{Y}) &= \text{var}(\beta'\mathbf{X}'\Sigma^{-1}\mathbf{Y}) \\ &= \beta'\mathbf{X}'\Sigma^{-1}\text{cov}(\mathbf{Y})\Sigma^{-1}\mathbf{X}\beta \\ &= \beta'\mathbf{X}'\Sigma^{-1}\Sigma\Sigma^{-1}\mathbf{X}\beta \\ &= \beta'\mathbf{X}'\Sigma^{-1}\mathbf{X}\beta.\end{aligned}$$

Sejam  $\gamma_i$  autovetores de  $\mathbf{X}'\Sigma^{-1}\mathbf{X}$ :

$$\mathbf{X}'\Sigma^{-1}\mathbf{X}(\gamma_i) = \lambda_i\gamma_i \text{ e } \gamma_i'\gamma_i = 1.$$

Os vetores  $\mathbf{X}\gamma_i$  são ortogonais, pois

$$\begin{aligned}\langle\langle\mathbf{X}\gamma_i,\mathbf{X}\gamma_j\rangle\rangle &= (\mathbf{X}\gamma_i)'\Sigma^{-1}\mathbf{X}\gamma_j = \gamma_i'\mathbf{X}'\Sigma^{-1}\mathbf{X}\gamma_j \\ &= \gamma_i'(\lambda_j\gamma_j) = \lambda_j\gamma_i'\gamma_j = 0.\end{aligned}$$

Definindo o subespaço  $W_m = \text{span}\{\mathbf{X}\gamma_1,\mathbf{X}\gamma_2,\mathbf{X}\gamma_3,\dots,\mathbf{X}\gamma_m\}$ , tem-se

$$\mathbf{X}\hat{\beta}_{\text{PCR},m} = P_{W_m}\mathbf{Y},$$

em que  $P_{W_m}$  é a projeção ortogonal, em relação ao produto interno  $\langle\langle,\rangle\rangle$ , no subespaço gerado por  $\mathbf{X}\gamma_1,\mathbf{X}\gamma_2,\mathbf{X}\gamma_3,\dots,\mathbf{X}\gamma_m$ .

Da mesma forma, como feito para o caso em que  $\text{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$ , obtém-se:

$$P_{W_m} = \mathbf{\Xi}_m(\mathbf{\Xi}_m'\Sigma^{-1}\mathbf{\Xi}_m)^{-1}\mathbf{\Xi}_m'\Sigma^{-1}.$$

Como  $\mathbf{\Xi}_m = \mathbf{X}\Gamma_m$ , segue que

$$\begin{aligned}\mathbf{X}\hat{\beta}_{\text{PCR},m} &= \mathbf{X}\Gamma_m(\Gamma_m'\mathbf{X}'\Sigma^{-1}\mathbf{X}\Gamma_m)^{-1}\Gamma_m'\mathbf{X}'\Sigma^{-1}\mathbf{Y} \\ &= \mathbf{X}\Gamma_m(\Gamma_m'\mathbf{X}'\Sigma^{-1}\mathbf{X}\Gamma_m)^{-1}\Gamma_m'(\mathbf{X}'\Sigma^{-1}\mathbf{X})(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y} \\ &= \mathbf{X}\Gamma_m(\Gamma_m'\mathbf{X}'\Sigma^{-1}\mathbf{X}\Gamma_m)^{-1}\Gamma_m'(\mathbf{X}'\Sigma^{-1}\mathbf{X})\hat{\beta}_{\text{GM}}.\end{aligned}$$

Logo,

$$\hat{\beta}_{\text{PCR},m} = \Gamma_m \left( \Gamma_m' \mathbf{X}' \Sigma^{-1} \mathbf{X} \Gamma_m \right)^{-1} \Gamma_m' (\mathbf{X}' \Sigma^{-1} \mathbf{X}) \hat{\beta}_{\text{GM}}.$$

Observe que, se  $m = p$ ,  $\Gamma_p$  é uma matriz  $p \times p$  ortogonal. Assim:

$$\begin{aligned} \hat{\beta}_{\text{PCR},p} &= \Gamma_p \left( \Gamma_p' \mathbf{X}' \Sigma^{-1} \mathbf{X} \Gamma_p \right)^{-1} \Gamma_p' (\mathbf{X}' \Sigma^{-1} \mathbf{X}) \hat{\beta}_{\text{GM}} \\ &= \Gamma_p (\Gamma_p)^{-1} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} (\Gamma_p')^{-1} \Gamma_p' (\mathbf{X}' \Sigma^{-1} \mathbf{X}) \hat{\beta}_{\text{GM}} \\ &= \hat{\beta}_{\text{GM}}. \end{aligned}$$

## 6 Regressão por quadrados mínimos parciais: uma abordagem geométrica

### 6.1 PLS populacional

Esta seção tem como referência o artigo Helland (1990).

Seja uma situação em que se quer prever uma variável aleatória  $y$  a partir de  $p$  variáveis aleatórias  $x_1, \dots, x_p$ . Será considerada a situação restritiva em que as esperanças das covariáveis e da variável resposta são nulas. Tal restrição estará em concordância com o uso posterior do algoritmo PLS amostral, em que se pode centrar as variáveis na média.

$$E \begin{bmatrix} \mathbf{x} \\ y \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix} \quad \text{e} \quad \text{cov} \begin{bmatrix} \mathbf{x} \\ y \end{bmatrix} = \begin{bmatrix} \Sigma & \sigma \\ \sigma' & \sigma_y^2 \end{bmatrix},$$

em que,  $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ ,  $\sigma = \text{cov}(\mathbf{x}, y) = \begin{bmatrix} \text{cov}(x_1, y) \\ \vdots \\ \text{cov}(x_p, y) \end{bmatrix}$  e  $\Sigma = \text{cov}(\mathbf{x})$ .

Uma pergunta interessante: Qual é o vetor  $\boldsymbol{\eta} \in \mathbb{R}^p$  que define a direção de maior covariância com a variável  $y$ ? Isto é: se  $\mathbf{Z} = \begin{bmatrix} \mathbf{x} \\ y \end{bmatrix}$ , deseja-se o vetor  $\boldsymbol{\eta}$  em  $\mathbb{R}^p$ ,  $\|\boldsymbol{\eta}\| = 1$ , que, ao se projetar o vetor aleatório  $\mathbf{Z}$  na direção de  $\boldsymbol{\eta}$ , fornece a maior covariância com  $\mathbf{e}_{p+1} \cdot \mathbf{Z} = y$ . Tem-se, então, o problema de maximização  $\max_{\|\boldsymbol{\eta}\|=1} \text{cov}(\mathbf{e}_{p+1} \cdot \mathbf{Z}, \boldsymbol{\eta}' \cdot \mathbf{Z})$ , em que  $\boldsymbol{\eta}^* = \begin{bmatrix} \boldsymbol{\eta} \\ 0 \end{bmatrix}$ . Utilizando propriedades de vetores aleatórios, tem-se:

$$\begin{aligned} \text{cov} \left( \mathbf{e}_{p+1} \cdot \mathbf{Z}, \begin{bmatrix} \boldsymbol{\eta} \\ 0 \end{bmatrix} \cdot \mathbf{Z} \right) &= \left\langle \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}, \begin{bmatrix} \Sigma & \sigma \\ \sigma' & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ 0 \end{bmatrix} \right\rangle \\ &= \left\langle \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}, \begin{bmatrix} \Sigma \boldsymbol{\eta} \\ \sigma' \boldsymbol{\eta} \end{bmatrix} \right\rangle = \sigma' \boldsymbol{\eta} = \|\boldsymbol{\eta}\| \|\sigma\| \cos \theta = \|\sigma\| \cos \theta \end{aligned}$$

Para se ter covariância máxima, o ângulo  $\theta$  deve ser zero ( $\cos 0 = 1$ ). Logo,  $\boldsymbol{\eta}$  deve estar na mesma direção de  $\sigma$ . Assim, a direção que fornece a covariância

máxima é dada pelo próprio vetor de covariâncias  $\sigma$ . A projeção de  $\mathbf{Z}$  em  $\sigma$  é  $\sigma' \mathbf{Z} = \sigma' \mathbf{x}$ .

A ideia é, como uma primeira aproximação, substituir a variável aleatória de interesse  $y$  por um múltiplo do vetor de variáveis aleatórias preditoras  $\sigma' \mathbf{x}$ . Usando como critério de predição o Erro Quadrático Médio (EQM), devemos minimizar a expressão  $E \left[ (y - \alpha (\sigma' \mathbf{x}))^2 \right]$ :

$$\begin{aligned} & E \left[ (y - \alpha (\sigma' \mathbf{x}))^2 \right] \\ &= E \left[ y^2 \right] - 2\alpha E \left[ y (\sigma' \mathbf{x}) \right] + \alpha^2 E \left[ (\sigma' \mathbf{x})^2 \right] \\ &= \sigma_y^2 - 2\alpha \text{cov} (y, \sigma' \mathbf{x}) + \alpha^2 \text{var} (\sigma' \mathbf{x}) \end{aligned}$$

Derivando e igualando a zero, obtém-se:

$$\begin{aligned} -2\text{cov} (y, \sigma' \mathbf{x}) + 2\alpha \text{var} (\sigma' \mathbf{x}) &= 0 \\ \alpha &= \frac{\text{cov}(y, \sigma' \mathbf{x})}{\text{var}(\sigma' \mathbf{x})} \end{aligned}$$

Note que, a escala da variável  $y$  é diferente das escalas das variáveis  $x_i$ , e com esta ponderação pela  $\text{var} (\sigma' \mathbf{x})$ , este problema fica minimizado.

Assim, a variável que prediz  $y$  com menor erro quadrático médio é a variável  $\frac{\text{cov}(y, \sigma' \mathbf{x})}{\text{var}(\sigma' \mathbf{x})} \sigma' \mathbf{x}$ .

Denominando  $q_1 = \text{cov} \left( y, \frac{\sigma' \mathbf{x}}{\text{var}(\sigma' \mathbf{x})} \right)$ , tem-se:

$$\begin{aligned} & \text{cov} \left( y, \frac{\sigma' \mathbf{x}}{\text{var}(\sigma' \mathbf{x})} \right) \\ &= \text{cov} \left( \mathbf{e}_{p+1} \cdot \begin{bmatrix} \mathbf{x} \\ y \end{bmatrix}, \begin{bmatrix} \frac{\sigma}{\text{var}(\sigma' \mathbf{x})} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x} \\ y \end{bmatrix} \right) \\ &= \left\langle \mathbf{e}_{p+1}, \begin{bmatrix} \Sigma & \sigma \\ \sigma' & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \frac{\sigma}{\text{var}(\sigma' \mathbf{x})} \\ 0 \end{bmatrix} \right\rangle \\ &= \frac{\sigma' \sigma}{\text{var}(\sigma' \mathbf{x})} = \frac{\sigma' \sigma}{\sigma' \text{cov}(\mathbf{x}) \sigma} = \frac{\sigma' \sigma}{\sigma' \Sigma \sigma} \end{aligned}$$

Assim, a parte de  $y$  que não foi explicada por  $\frac{\text{cov}(y, \sigma' \mathbf{x})}{\text{var}(\sigma' \mathbf{x})} \sigma' \mathbf{x}$ , denominada por  $f_1$ , é dada por:

$$f_1 = y - \frac{\text{cov} (y, \sigma' \mathbf{x})}{\text{var} (\sigma' \mathbf{x})} \sigma' \mathbf{x} = y - \frac{\sigma' \sigma}{\text{var} (\sigma' \mathbf{x})} \sigma' \mathbf{x} = y - \frac{\sigma' \sigma}{\sigma' \Sigma \sigma} \sigma' \mathbf{x}.$$

No sentido de se obter uma sequência de variáveis preditoras, considera-se como primeira variável preditora de  $y$ , apenas a variável  $\frac{\sigma' \mathbf{x}}{\text{var}(\sigma' \mathbf{x})}$ . Assim, o vetor de correlações  $\sigma = \text{cov} (\mathbf{x}, y)$ , renomeado por  $\mathbf{w}_1$ , pode ser substituído pelo vetor  $\text{cov} \left( \mathbf{x}, \frac{\sigma' \mathbf{x}}{\text{var}(\sigma' \mathbf{x})} \right)$ . Denomina-se este vetor por:

$$\mathbf{p}_1 = \frac{\text{cov} (\mathbf{x}, \sigma' \mathbf{x})}{\text{var} (\sigma' \mathbf{x})} = \begin{bmatrix} \frac{\text{cov} (x_1, \sigma' \mathbf{x})}{\text{var}(\sigma' \mathbf{x})} \\ \vdots \\ \frac{\text{cov} (x_p, \sigma' \mathbf{x})}{\text{var}(\sigma' \mathbf{x})} \end{bmatrix}.$$



A  $i$ -ésima coordenada do vetor  $\mathbf{p}_1$  pode ser obtida fazendo-se:

$$\begin{aligned} x_i &= \mathbf{e}_i \cdot \mathbf{Z} \\ \boldsymbol{\sigma}'\mathbf{x} &= (\boldsymbol{\sigma}, 0) \cdot \mathbf{Z} \\ \text{cov}(x_i, \boldsymbol{\sigma}'\mathbf{x}) &= \text{cov}(\mathbf{e}_i \cdot \mathbf{Z}, \boldsymbol{\sigma} \cdot \mathbf{Z}) \\ &= \left\langle \mathbf{e}_i, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}' & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma} \\ 0 \end{bmatrix} \right\rangle = \left\langle \mathbf{e}_i, \begin{bmatrix} \boldsymbol{\Sigma}\boldsymbol{\sigma} \\ \boldsymbol{\sigma}'\boldsymbol{\sigma} \end{bmatrix} \right\rangle \\ &= \sum_{j=1}^p \text{cov}(x_i, x_j) \sigma_j = \sum_{j=1}^p \text{cov}(x_i, x_j) \text{cov}(x_j, y) \\ \Rightarrow \text{cov}\left(x_i, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) &= \frac{1}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})} \sum_{j=1}^p \text{cov}(x_i, x_j) \text{cov}(x_j, y). \end{aligned}$$

Ou ainda,

$$\begin{aligned} \text{cov}\left(x_i, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) &= \text{cov}\left(x_i, \sum_{j=1}^p x_j \frac{\text{cov}(x_j, y)}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) \\ &= \sum_{j=1}^p \text{cov}\left(x_i, x_j \frac{\text{cov}(x_j, y)}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) = \sum_{j=1}^p \frac{\text{cov}(x_j, y)}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})} \text{cov}(x_i, x_j) \\ &= \sum_{j=1}^p \frac{\text{cov}(x_i, x_j)}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})} \sigma_j = \frac{1}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})} (\boldsymbol{\Sigma}\boldsymbol{\sigma})_i. \end{aligned}$$

Logo,  $\mathbf{p}_1 = \frac{\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})}$ .

A componente de  $\mathbf{x}$  perpendicular ao vetor  $\boldsymbol{\sigma}$  não foi utilizada na predição de  $y$ . Deve-se então, encontrar  $\alpha$  tal que  $\mathbf{x} - \alpha\mathbf{p}_1$  seja perpendicular à  $\boldsymbol{\sigma}$  :

$$\begin{aligned} \langle \boldsymbol{\sigma}, \mathbf{x} - \alpha\mathbf{p}_1 \rangle &= \langle \boldsymbol{\sigma}, \mathbf{x} \rangle - \alpha \langle \boldsymbol{\sigma}, \mathbf{p}_1 \rangle \\ \Rightarrow \langle \boldsymbol{\sigma}, \mathbf{x} \rangle - \alpha \left\langle \boldsymbol{\sigma}, \frac{\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})} \right\rangle &= 0 \\ \Rightarrow \langle \boldsymbol{\sigma}, \mathbf{x} \rangle - \alpha \frac{\boldsymbol{\sigma}'\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})} &= 0 \\ \Rightarrow \boldsymbol{\sigma}'\mathbf{x} - \alpha &= 0 \\ \Rightarrow \alpha &= \boldsymbol{\sigma}'\mathbf{x}. \end{aligned}$$

Desta forma, denominando  $t_1 = \boldsymbol{\sigma}'\mathbf{x}$ , tem-se o vetor  $\mathbf{e}_1 = \mathbf{x} - t_1\mathbf{p}_1$ ,  $\mathbf{e}_1 \perp \boldsymbol{\sigma}$ . Note que, a preditora de  $y$  pode ser reescrita como  $\frac{t_1}{\text{var}(t_1)} e$ ,

$$\mathbf{e}_1 = \mathbf{x} - (\boldsymbol{\sigma}'\mathbf{x}) \frac{\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\text{var}(\boldsymbol{\sigma}'\mathbf{x})}.$$

O processo agora é repetido substituindo-se  $y$  por sua parte não explicada  $f_1$  e  $\mathbf{x}$  por sua componente ainda não utilizada  $\mathbf{e}_1$ . A direção de maior covariância entre

$f_1$  e os vetores aleatórios no subespaço perpendicular a  $\sigma$ , é dada pelo vetor  $w_2$  :

$$\begin{aligned} w_2 &= \text{cov}(\mathbf{e}_1, f_1) = \text{cov}\left(\mathbf{e}_{1,y} - \frac{\sigma' \sigma}{\text{var}(\sigma' \mathbf{x})} \sigma' \mathbf{x}\right) \\ &= \text{cov}(\mathbf{e}_{1,y}) - \frac{\sigma' \sigma}{\text{var}(\sigma' \mathbf{x})} \text{cov}(\mathbf{e}_1, \sigma' \mathbf{x}) \stackrel{\mathbf{e}_1 \perp \sigma}{=} \text{cov}(\mathbf{e}_{1,y}) \\ &= \text{cov}(\mathbf{x} - \mathbf{p}_1 t_1, y) = \text{cov}(\mathbf{x}, y) - \text{cov}(\mathbf{p}_1 t_1, y) \\ &= \sigma - \mathbf{p}_1 \text{cov}(\sigma' \mathbf{x}, y) = \sigma - \frac{\Sigma \sigma}{\sigma' \Sigma \sigma} \sigma' \sigma. \end{aligned}$$

Definindo  $t_2 = \mathbf{e}_1' w_2$ , a variável preditora de  $f_1$  será  $\frac{t_2}{\text{var}(t_2)}$ . Calculando-se a covariância entre  $f_1$  e seu preditor  $\frac{t_2}{\text{var}(t_2)}$ , temos a grandeza  $q_2$ , dada por

$$q_2 = \frac{\text{cov}(f_1, t_2)}{\text{var}(t_2)}.$$

Desta forma, a parte não explicada de  $f_1$  será dada por  $f_2 = f_1 - q_2 t_2$ .

Define-se agora, o vetor dado pela covariância entre o vetor aleatório  $\mathbf{e}_1$  e a variável preditora de  $f_1$ ,  $\frac{t_2}{\text{var}(t_2)}$ , denominado  $\mathbf{p}_2$  :

$$\mathbf{p}_2 = \frac{\text{cov}(\mathbf{e}_1, t_2)}{\text{var}(t_2)}.$$

Novamente, a componente do vetor  $\mathbf{e}_1$  que não foi utilizada para explicar  $f_1$  será dada pelo vetor  $\mathbf{e}_2 = \mathbf{e}_1 - \mathbf{p}_2 t_2$ .

Resumindo:

Para  $a = 1$ :

$$\begin{aligned} w_1 &= \text{cov}(\mathbf{x}, y) = \sigma \\ t_1 &= \mathbf{x}' w_1 = \sigma' \mathbf{x} \\ \mathbf{p}_1 &= \text{cov}\left(\mathbf{x}, \frac{t_1}{\text{var}(t_1)}\right) = \text{cov}\left(\mathbf{x}, \frac{\sigma' \mathbf{x}}{\text{var}(\sigma' \mathbf{x})}\right) = \frac{\Sigma \sigma}{\sigma' \Sigma \sigma} \\ q_1 &= \text{cov}\left(y, \frac{t_1}{\text{var}(t_1)}\right) = \text{cov}\left(y, \frac{\sigma' \mathbf{x}}{\text{var}(\sigma' \mathbf{x})}\right) = \frac{\sigma' \sigma}{\sigma' \Sigma \sigma} \\ \mathbf{e}_1 &= \mathbf{x} - \mathbf{p}_1 t_1 = \mathbf{x} - \frac{\Sigma \sigma}{\sigma' \Sigma \sigma} \sigma' \mathbf{x} \\ f_1 &= y - q_1 t_1 = y - \frac{\sigma' \sigma}{\sigma' \Sigma \sigma} \sigma' \mathbf{x} \end{aligned}$$

Para  $a = 2$ :

$$\begin{aligned}
 t_2 &= \mathbf{e}_1' \mathbf{w}_2 = \mathbf{x}' \left( \boldsymbol{\sigma} - \frac{\boldsymbol{\sigma}' \boldsymbol{\sigma}}{\boldsymbol{\sigma}' \boldsymbol{\Sigma} \boldsymbol{\sigma}} \boldsymbol{\Sigma} \boldsymbol{\sigma} \right) = \mathbf{x}' \boldsymbol{\sigma} - \frac{\boldsymbol{\sigma}' \boldsymbol{\sigma}}{\boldsymbol{\sigma}' \boldsymbol{\Sigma} \boldsymbol{\sigma}} \mathbf{x}' \boldsymbol{\Sigma} \boldsymbol{\sigma} \\
 \text{var}(t_2) &= \left( \boldsymbol{\sigma} - \frac{\|\boldsymbol{\sigma}\|^2}{\boldsymbol{\sigma}' \boldsymbol{\Sigma} \boldsymbol{\sigma}} \boldsymbol{\Sigma} \boldsymbol{\sigma} \right)' \boldsymbol{\Sigma} \left( \boldsymbol{\sigma} - \frac{\|\boldsymbol{\sigma}\|^2}{\boldsymbol{\sigma}' \boldsymbol{\Sigma} \boldsymbol{\sigma}} \boldsymbol{\Sigma} \boldsymbol{\sigma} \right) \\
 \mathbf{w}_2 &= \text{cov}(\mathbf{e}_1, f_1) = \boldsymbol{\sigma} - \frac{\boldsymbol{\Sigma} \boldsymbol{\sigma}}{\boldsymbol{\sigma}' \boldsymbol{\Sigma} \boldsymbol{\sigma}} \boldsymbol{\sigma}' \boldsymbol{\sigma} \\
 \mathbf{p}_2 &= \frac{\text{cov}(\mathbf{e}_1, t_2)}{\text{var}(t_2)} = \frac{\text{cov}(\mathbf{x}, t_2)}{\text{var}(t_2)} \\
 &= \text{cov} \left( \mathbf{x}, \mathbf{x}' \boldsymbol{\sigma} - \frac{\boldsymbol{\sigma}' \boldsymbol{\sigma}}{\boldsymbol{\sigma}' \boldsymbol{\Sigma} \boldsymbol{\sigma}} \mathbf{x}' \boldsymbol{\Sigma} \boldsymbol{\sigma} \right) = \boldsymbol{\Sigma} \boldsymbol{\sigma} - \frac{\boldsymbol{\sigma}' \boldsymbol{\sigma}}{\boldsymbol{\sigma}' \boldsymbol{\Sigma} \boldsymbol{\sigma}} \boldsymbol{\Sigma}^2 \boldsymbol{\sigma} \\
 q_2 &= \frac{\text{cov}(f_1, t_2)}{\text{var}(t_2)} \\
 \mathbf{e}_2 &= \mathbf{e}_1 - \mathbf{p}_2 t_2 \\
 f_2 &= f_1 - q_2 t_2
 \end{aligned}$$

Um algoritmo para a predição da variável aleatória de interesse  $y$ , em termos das covariáveis fica, então, definido.

O algoritmo para o PLS populacional (HELLAND, 1990) pode ser apresentado como:

1. Defina os valores iniciais para  $\mathbf{x}$  residual ( $\mathbf{e}_a$ ) e  $y$  residual ( $f_a$ ) :

$$\begin{aligned}
 \mathbf{e}_0 &= \mathbf{x} - \boldsymbol{\mu}_x \\
 f_0 &= y - \mu_y
 \end{aligned}$$

Para  $a = 1, 2, \dots$ , execute os passos (2) – (4) a seguir:

2. Introduza escores  $t$  como combinações lineares do  $\mathbf{x}$  residual a partir do último passo; para fazer os scores mais relacionados com  $y$ , use covariâncias com  $y$  residual como coeficientes/pesos ( $\mathbf{w}$ ) :

$$\begin{aligned}
 \mathbf{w}_a &= \text{cov}(\mathbf{e}_{a-1}, f_{a-1}) \\
 t_a &= \mathbf{e}_{a-1}' \mathbf{w}_a
 \end{aligned}$$

3. Determine  $\mathbf{x}$  loadings ( $\mathbf{p}_a$ ) e  $y$  loading ( $q_a$ ) por quadrados mínimos:

$$\begin{aligned}
 \mathbf{p}_a &= \frac{\text{cov}(\mathbf{e}_{a-1}, t_a)}{\text{var}(t_a)} \\
 q_a &= \frac{\text{cov}(f_{a-1}, t_a)}{\text{var}(t_a)}
 \end{aligned}$$



3. O vetor aleatório  $\mathbf{e}_a$  é não correlacionado com a variável  $t_a$ , pois  $t_a$  é obtida por uma projeção em  $\mathbf{w}_a$  e o vetor  $\mathbf{e}_a$  é perpendicular ao vetor  $\mathbf{w}_a$ . De fato, temos mais do que isso: como  $\mathbf{e}_a$  é perpendicular a  $\mathbf{w}_a$ , por construção, e  $\mathbf{w}_a$  é perpendicular a  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{a-1}$ , temos que  $\mathbf{e}_a$  é perpendicular a todos eles,  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_a$ , implicando então, que  $\mathbf{e}_a$  é não correlacionado com as variáveis  $t_1, t_2, \dots, t_a$ . Da mesma forma, este resultado segue da construção geométrica se  $\Sigma = \mathbf{I}$ .

Estas três propriedades (Propriedade.i a Propriedade.iii) podem ser demonstradas algebricamente (HOSKULDSSON, 1988).

As propriedades 2 e 3 permitem as seguintes simplificações no algoritmo:

#### Propriedade.i

$$\begin{aligned} \mathbf{w}_a &= \text{cov}(\mathbf{e}_{a-1}, f_{a-1}) = \text{cov}(\mathbf{e}_{a-1}, f_{a-2} - q_{a-1}t_{a-1}) \\ &= \text{cov}(\mathbf{e}_{a-1}, f_{a-2}) - q_{a-1}\text{cov}(\mathbf{e}_{a-1}, t_{a-1}) \\ &\stackrel{\text{P.3}}{=} \text{cov}(\mathbf{e}_{a-1}, f_{a-2}) = \text{cov}(\mathbf{e}_{a-1}, f_{a-3} - q_{a-2}t_{a-2}) \\ &= \text{cov}(\mathbf{e}_{a-1}, f_{a-3}) - q_{a-2}\text{cov}(\mathbf{e}_{a-1}, t_{a-2}) \\ &\stackrel{\text{P.3}}{=} \text{cov}(\mathbf{e}_{a-1}, f_{a-3}) = \dots = \text{cov}(\mathbf{e}_{a-1}, y) \end{aligned}$$

$$\text{Logo, } \mathbf{w}_a = \text{cov}(\mathbf{e}_{a-1}, f_{a-1}) = \text{cov}(\mathbf{e}_{a-1}, y).$$

#### Propriedade.ii

$$\begin{aligned} \text{cov}(\mathbf{e}_{a-1}, t_a) &= \text{cov}(\mathbf{e}_{a-2} - \mathbf{p}_{a-1}t_{a-1}, t_a) \\ &= \text{cov}(\mathbf{e}_{a-2}, t_a) - \text{cov}(t_{a-1}, t_a) \mathbf{p}_{a-1} \\ &\stackrel{\text{P.2}}{=} \text{cov}(\mathbf{e}_{a-2}, t_a) = \text{cov}(\mathbf{e}_{a-3} - \mathbf{p}_{a-2}t_{a-2}, t_a) \\ &= \text{cov}(\mathbf{e}_{a-3}, t_a) - \text{cov}(t_{a-2}, t_a) \mathbf{p}_{a-2} \\ &\stackrel{\text{P.2}}{=} \text{cov}(\mathbf{e}_{a-3}, t_a) = \dots = \text{cov}(\mathbf{x}, t_a) \end{aligned}$$

$$\text{Logo, } \mathbf{p}_a = \frac{\text{cov}(\mathbf{e}_{a-1}, t_a)}{\text{var}(t_a)} = \frac{\text{cov}(\mathbf{x}, t_a)}{\text{var}(t_a)}.$$

#### Propriedade.iii

$$\begin{aligned} \text{cov}(f_{a-1}, t_a) &= \text{cov}(f_{a-2} - q_{a-1}t_{a-1}, t_a) \\ &= \text{cov}(f_{a-2}, t_a) - q_{a-1}\text{cov}(t_{a-1}, t_a) \\ &\stackrel{\text{P.2}}{=} \text{cov}(f_{a-2}, t_a) = \text{cov}(f_{a-3} - q_{a-2}t_{a-2}, t_a) \\ &= \text{cov}(f_{a-3}, t_a) - q_{a-2}\text{cov}(t_{a-2}, t_a) \\ &\stackrel{\text{P.2}}{=} \text{cov}(f_{a-3}, t_a) = \dots = \text{cov}(y, t_a) \end{aligned}$$

$$\text{Logo, } q_a = \frac{\text{cov}(f_{a-1}, t_a)}{\text{var}(t_a)} = \frac{\text{cov}(y, t_a)}{\text{var}(t_a)}.$$

Portanto,  $\mathbf{e}_{a-1}$  pode ser substituído por  $\mathbf{x}$  na definição de  $\mathbf{p}_a$ , e  $f_{a-1}$  pode ser substituído por  $y$  nas definições de  $\mathbf{w}_a$  e  $q_a$ , nos passos 2 e 3 do algoritmo.

Assim, tem-se que o algoritmo fica da forma:

$$\begin{aligned} \mathbf{e}_0 &= \mathbf{x} - \boldsymbol{\mu}_x \\ f_0 &= y - \mu_y \\ \text{Para } a &= 1, 2, 3, \dots \\ t_a &= \mathbf{e}_{a-1}' \mathbf{w}_a \\ \mathbf{w}_a &= \text{cov}(\mathbf{e}_{a-1}, y) \\ \mathbf{p}_a &= \frac{\text{cov}(\mathbf{x}, t_a)}{\text{var}(t_a)} \\ q_a &= \frac{\text{cov}(y, t_a)}{\text{var}(t_a)} \\ \mathbf{e}_a &= \mathbf{e}_{a-1} - \mathbf{p}_a t_a \\ f_a &= f_{a-1} - q_a t_a. \end{aligned}$$

Apesar do algoritmo nesta versão ser mais simples, como o objetivo é a motivação para se obter o algoritmo para o caso amostral, a primeira versão é a mais útil.

No passo A do algoritmo, obtemos a representação:

$$\begin{aligned} &\mathbf{p}_1 t_1 + \mathbf{p}_2 t_2 + \mathbf{p}_3 t_3 + \dots + \mathbf{p}_A t_A \\ &= (\mathbf{e}_0 - \mathbf{e}_1) + (\mathbf{e}_1 - \mathbf{e}_2) + (\mathbf{e}_2 - \mathbf{e}_3) + \dots + (\mathbf{e}_{A-1} - \mathbf{e}_A) \\ &= \mathbf{e}_0 - \mathbf{e}_A = (\mathbf{x} - \boldsymbol{\mu}_x) - \mathbf{e}_A, \end{aligned}$$

logo,  $\mathbf{x} = \boldsymbol{\mu}_x + \mathbf{p}_1 t_1 + \dots + \mathbf{p}_A t_A + \mathbf{e}_A$ .

$$\begin{aligned} &q_1 t_1 + q_2 t_2 + q_3 t_3 + \dots + q_A t_A \\ &= (f_0 - f_1) + (f_1 - f_2) + (f_2 - f_3) + \dots + (f_{A-1} - f_A) \\ &= f_0 - f_A = (y - \mu_y) - f_A, \end{aligned}$$

logo,  $y = \mu_y + q_1 t_1 + \dots + q_A t_A + f_A$ .

Temos então, um preditor para a variável resposta  $y$ , dado por

$$\hat{y}_{A, \text{PLS}} = \mu_y + q_1 t_1 + \dots + q_A t_A.$$

Este preditor é, de fato, um preditor linear, pois utilizando a segunda versão do algoritmo, tem-se:

$$\hat{y}_{A, \text{PLS}} = \mu_y + \boldsymbol{\beta}'_{A, \text{PLS}} (\mathbf{x} - \boldsymbol{\mu}_x)$$

em que (HELLAND, 1990),

$$\boldsymbol{\beta}_{A, \text{PLS}} = \mathbf{W}_A (\mathbf{W}_A' \boldsymbol{\Sigma} \mathbf{W}_A)^{-1} \mathbf{W}_A' \boldsymbol{\sigma}.$$

Explicitando estas matrizes:

$$\begin{aligned} \mathbf{W}_A &= [\mathbf{w}_1, \dots, \mathbf{w}_A] \\ \Sigma(\mathbf{W}_A) &= \Sigma[\mathbf{w}_1, \dots, \mathbf{w}_A] = [\Sigma\mathbf{w}_1, \dots, \Sigma\mathbf{w}_A] \\ \mathbf{W}_A' \Sigma \mathbf{W}_A &= \begin{bmatrix} \mathbf{w}_1' \\ \vdots \\ \mathbf{w}_A' \end{bmatrix} [\Sigma\mathbf{w}_1, \dots, \Sigma\mathbf{w}_A] = \begin{bmatrix} \mathbf{w}_1' \Sigma \mathbf{w}_1 & \cdots & \mathbf{w}_1' \Sigma \mathbf{w}_A \\ \vdots & \ddots & \vdots \\ \mathbf{w}_A' \Sigma \mathbf{w}_1 & \cdots & \mathbf{w}_A' \Sigma \mathbf{w}_A \end{bmatrix}. \end{aligned}$$

*Casos particulares:*

**1º Caso:** Se  $\mathbf{w}_i' \Sigma \mathbf{w}_j = 0, \forall i \neq j$ , temos uma matriz diagonal e poderemos calcular a inversa:

$$\begin{aligned} (\mathbf{W}_A' \Sigma \mathbf{W}_A)^{-1} &= \begin{bmatrix} \frac{1}{\mathbf{w}_1' \Sigma \mathbf{w}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\mathbf{w}_2' \Sigma \mathbf{w}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mathbf{w}_A' \Sigma \mathbf{w}_A} \end{bmatrix} \\ \mathbf{W}_A' \sigma &= \begin{bmatrix} \mathbf{w}_1' \\ \vdots \\ \mathbf{w}_A' \end{bmatrix} \sigma = \begin{bmatrix} \mathbf{w}_1' \sigma \\ \vdots \\ \mathbf{w}_A' \sigma \end{bmatrix} \\ (\mathbf{W}_A' \Sigma \mathbf{W}_A)^{-1} \mathbf{W}_A' \sigma &= \begin{bmatrix} \frac{\mathbf{w}_1' \sigma}{\mathbf{w}_1' \Sigma \mathbf{w}_1} \\ \vdots \\ \frac{\mathbf{w}_A' \sigma}{\mathbf{w}_A' \Sigma \mathbf{w}_A} \end{bmatrix} \\ \mathbf{W}_A (\mathbf{W}_A' \Sigma \mathbf{W}_A)^{-1} \mathbf{W}_A' \sigma &= [\mathbf{w}_1, \dots, \mathbf{w}_A] \begin{bmatrix} \frac{\mathbf{w}_1' \sigma}{\mathbf{w}_1' \Sigma \mathbf{w}_1} \\ \vdots \\ \frac{\mathbf{w}_A' \sigma}{\mathbf{w}_A' \Sigma \mathbf{w}_A} \end{bmatrix} \\ &= \left[ \left( \frac{\mathbf{w}_1' \sigma}{\mathbf{w}_1' \Sigma \mathbf{w}_1} \right) \mathbf{w}_1 + \cdots + \left( \frac{\mathbf{w}_A' \sigma}{\mathbf{w}_A' \Sigma \mathbf{w}_A} \right) \mathbf{w}_A \right]. \end{aligned}$$

**2º Caso:** Se o número de passos A do algoritmo é igual a n,  $A = n$ , a matriz  $\mathbf{W}_A$  é quadrada e inversível, assim,

$$\begin{aligned} \beta_{A,PLS} &= \mathbf{W}_A (\mathbf{W}_A' \Sigma \mathbf{W}_A)^{-1} \mathbf{W}_A' \sigma, \\ \beta_{n,PLS} &= \mathbf{W}_A \mathbf{W}_A^{-1} \Sigma^{-1} (\mathbf{W}_A')^{-1} \mathbf{W}_A' \sigma = \Sigma^{-1} \sigma, \end{aligned}$$

que é o mesmo vetor de coeficientes do melhor preditor linear dado pelo teorema de Gauss-Markov ( $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} = \Sigma^{-1} \sigma$ ), no caso em que  $\begin{bmatrix} \mathbf{x} \\ y \end{bmatrix}$  tem distribuição Normal.

3º Caso: O preditor para apenas 1 passo, fica da forma:

$$\hat{Y}_{A,PLS} = \mu_y + q_1 t_1 = \mu_y + \frac{\|\sigma\|^2}{\sigma' \Sigma \sigma} \mathbf{x}' \sigma = \mu_y + \frac{\|\sigma\|^2}{\sigma' \Sigma \sigma} \sigma' (\mathbf{x} - \mu_x).$$

## 6.2 PLS amostral

Esta seção tem como referência o artigo Hoskuldsson (1988).

Sejam duas matrizes  $\mathbf{X}_{n \times p}$  e  $\mathbf{Y}_{n \times k}$ , centradas na média, em que p é o número de covariáveis, k é o número de variáveis respostas e n é o número de observações (repetições):

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix}.$$

A matriz de covariâncias amostrais, a menos de uma constante, é dada por  $\mathbf{X}'\mathbf{Y}$ :

$$\begin{aligned} \mathbf{X}'\mathbf{Y} &= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix} \\ &= \begin{bmatrix} \text{côv}(\mathbf{x}_1, \mathbf{y}_1) & \text{côv}(\mathbf{x}_1, \mathbf{y}_2) & \cdots & \text{côv}(\mathbf{x}_1, \mathbf{y}_k) \\ \text{côv}(\mathbf{x}_2, \mathbf{y}_1) & \text{côv}(\mathbf{x}_2, \mathbf{y}_2) & \cdots & \text{côv}(\mathbf{x}_2, \mathbf{y}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{côv}(\mathbf{x}_p, \mathbf{y}_1) & \text{côv}(\mathbf{x}_p, \mathbf{y}_2) & \cdots & \text{côv}(\mathbf{x}_p, \mathbf{y}_k) \end{bmatrix}. \end{aligned}$$

Observe que as colunas da matriz  $\mathbf{X}'\mathbf{Y}$ , a menos de uma constante, estimam os vetores  $\sigma_j = \text{cov}(\mathbf{X}, \mathbf{Y}_j)$  com  $j = 1, \dots, k$ .

### 6.2.1 Componentes com covariância máxima

Considere dois componentes  $\mathbf{f}$  e  $\mathbf{g}$  definidos pelas colunas das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ :

$$\mathbf{f} = \mathbf{X}\mathbf{d}, \quad \|\mathbf{d}\| = 1$$

$$\mathbf{g} = \mathbf{Y}\mathbf{e}, \quad \|\mathbf{e}\| = 1.$$

Note que,

$$\mathbf{f} = \mathbf{X}\mathbf{d} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_p \end{bmatrix} = d_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} + \cdots + d_p \begin{bmatrix} x_{1p} \\ x_{2p} \\ \vdots \\ x_{np} \end{bmatrix},$$



logo,  $\mathbf{f} = \mathbf{X}\mathbf{d}$  é uma combinação linear das colunas de  $\mathbf{X}$ , isto é, é um vetor definido como uma média ponderada das realizações de todas as covariáveis. Da mesma forma,  $\mathbf{g} = \mathbf{Y}\mathbf{e}$  é uma combinação linear das colunas de  $\mathbf{Y}$ .

A covariância amostral entre os dois componentes é dada por:

$$\text{cov}(\mathbf{f}, \mathbf{g}) = \frac{\mathbf{f}'\mathbf{g}}{n}.$$

Uma boa escolha para  $\mathbf{f}$  e  $\mathbf{g}$  será aquela que fornecer covariância máxima. Aqui, a covariância deve ser máxima em valor absoluto, visto que a covariância pode ser positiva ou negativa.

Para obter os vetores que realizam essa maximização da covariância, é necessário fazer a decomposição em valores singulares da matriz de covariâncias amostrais  $\mathbf{X}'\mathbf{Y}$ . Para tal, considere  $\mathbf{X}'\mathbf{Y}$  como uma transformação linear de  $\mathbb{R}^k$  em  $\mathbb{R}^p$ ,  $k < p$ . Existe uma base  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$  em  $\mathbb{R}^k$ , e uma base  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_p\}$  em  $\mathbb{R}^p$ , ambas ortonormais, tais que  $\mathbf{X}'\mathbf{Y}(\mathbf{v}_i) = a_i\mathbf{z}_i$ , sendo a ordem dos vetores nas bases definida por  $a_1^2 \geq a_2^2 \geq \dots \geq a_k^2$ , conforme mostra a Figura 12. Desta forma,

tem-se a igualdade de matrizes  $\mathbf{X}'\mathbf{Y} = \sum_{i=1}^k a_i\mathbf{z}_i\mathbf{v}_i'$ , pois

$$\left( \sum_{i=1}^k a_i\mathbf{z}_i\mathbf{v}_i' \right) (\mathbf{v}_j) = \sum_{i=1}^k (a_i\mathbf{z}_i\mathbf{v}_i') (\mathbf{v}_j) = \sum_{i=1}^k (a_i\mathbf{z}_i) (\mathbf{v}_i'\mathbf{v}_j) = a_j\mathbf{z}_j = \mathbf{X}'\mathbf{Y}(\mathbf{v}_j).$$

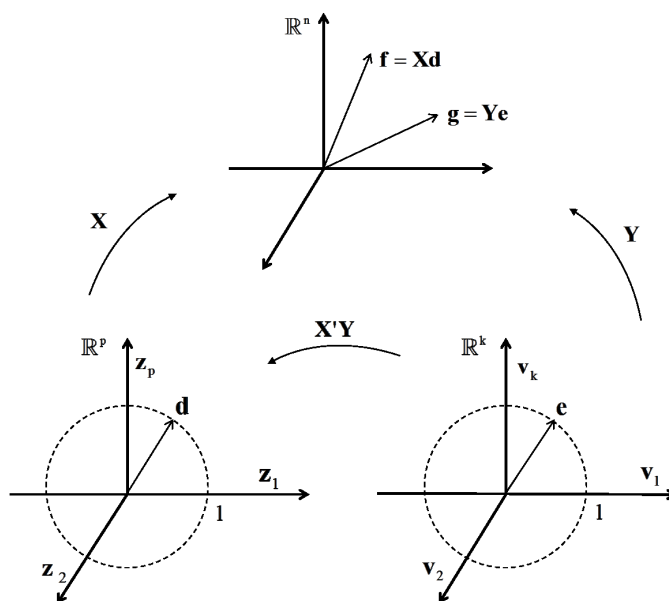


Figura 12 - Representação de  $\mathbf{X}'\mathbf{Y}$  como uma transformação linear de  $\mathbb{R}^k$  em  $\mathbb{R}^p$ .

O problema da maximização,  $\max(\mathbf{f}'\mathbf{g})$ , sujeito a  $\|\mathbf{d}\| = \|\mathbf{e}\| = 1$ , pode ser colocado da forma:

$$\begin{aligned}\max(\mathbf{f}'\mathbf{g}) &= \max((\mathbf{X}\mathbf{d})'(\mathbf{Y}\mathbf{e})) = \max(\mathbf{d}'\mathbf{X}'\mathbf{Y}\mathbf{e}) \\ &= \max(\mathbf{d}'(\mathbf{X}'\mathbf{Y})\mathbf{e}) = \max\left(\mathbf{d}'\left(\sum_{i=1}^k a_i \mathbf{z}_i \mathbf{v}_i'\right)\mathbf{e}\right) \\ &= \max\left(\sum_{i=1}^k a_i (\mathbf{d}'(\mathbf{z}_i \mathbf{v}_i'))\mathbf{e}\right) = \max\left(\sum_{i=1}^k a_i (\mathbf{d}'\mathbf{z}_i)(\mathbf{v}_i'\mathbf{e})\right).\end{aligned}$$

Assim, os vetores  $\mathbf{d}$  e  $\mathbf{e}$  que realizam essa maximização são os vetores  $\mathbf{d} = \mathbf{z}_1$  e  $\mathbf{e} = \mathbf{v}_1$ , pois,

$$\left(\sum_{i=1}^k a_i (\mathbf{z}_1' \mathbf{z}_i) (\mathbf{v}_1' \mathbf{v}_i)\right) = a_1.$$

**Prova:** Objetiva-se maximizar (ou minimizar) a expressão:

$$\begin{aligned}\max((\mathbf{X}\mathbf{d})'(\mathbf{Y}\mathbf{e})) &= \max(\mathbf{d}'(\mathbf{X}'\mathbf{Y})\mathbf{e}) \\ &= \max\left(\mathbf{d}'\left(\sum_{i=1}^k a_i \mathbf{z}_i \mathbf{v}_i'\right)\mathbf{e}\right) \\ &= \max\left(\sum_{j=1}^k d_j \mathbf{z}_j\right)' \left(\sum_{i=1}^k a_i \mathbf{z}_i \mathbf{v}_i'\right) \left(\sum_{s=1}^k e_s \mathbf{v}_s\right) \\ &= \max \sum_{j,i,s} d_j a_i e_s \mathbf{z}_j' (\mathbf{z}_i \mathbf{v}_i') \mathbf{v}_s \\ &= \max \sum_{j,i,s} d_j a_i e_s (\mathbf{z}_j' \mathbf{z}_i) (\mathbf{v}_i' \mathbf{v}_s) \\ &= \max \sum_i d_i a_i e_i (\mathbf{z}_i' \mathbf{z}_i) (\mathbf{v}_i' \mathbf{v}_i) \\ &= \max_{\|\mathbf{d}\|=\|\mathbf{e}\|=1} \sum_i d_i a_i e_i\end{aligned}$$

Assim, obtém-se a função lagrangiana com duas restrições,  $\|\mathbf{d}\| = \|\mathbf{e}\| = 1$ :

$$H(d_1, \dots, d_k, e_1, \dots, e_k, \lambda, \alpha) = \sum d_i a_i e_i + \lambda \left(\sum d_i^2 - 1\right) + \alpha \left(\sum e_i^2 - 1\right).$$

De onde segue o sistema de equações:

$$\begin{cases} \frac{\partial H}{\partial d_j} = a_j e_j + 2\lambda d_j = 0 \\ \frac{\partial H}{\partial e_1} = a_1 d_1 + 2\alpha e_1 = 0 \\ \sum d_i^2 = 1 \\ \sum e_i^2 = 1 \end{cases}$$

Multiplicando a 1ª equação por  $d_j$  e a 2ª equação por  $e_j$ , tem-se:

$$\begin{aligned} d_j a_j e_j + 2\lambda d_j^2 &= 0 \\ e_j a_j d_j + 2\alpha e_j^2 &= 0 \end{aligned}$$

Igualando as duas equações, segue que:

$$\lambda d_j^2 = \alpha e_j^2 \Rightarrow \lambda \sum d_j^2 = \alpha \sum e_j^2 \Rightarrow \lambda = \alpha \Rightarrow d_j = |e_j|.$$

Assim, volta-se ao problema da maximização, mas agora com apenas uma restrição:

$$\max_{\|\mathbf{d}\|=1} \sum a_i d_i^2.$$

Dada a função lagrangiana:

$$H(d_1, \dots, d_k, \lambda) = \sum a_i d_i^2 + \lambda \left( \sum d_i^2 - 1 \right),$$

segue que,

$$\frac{\partial H}{\partial d_j} = 2a_j d_j + 2\lambda d_j = 0.$$

Logo,  $a_j d_j = \lambda d_j$ .

Supondo todos os  $a_i$  diferentes entre si,  $a_i \neq a_j$  para  $i \neq j$ , a solução consiste em todos os  $d_j$  serem zero, exceto um deles que deve ser 1, pois  $\|\mathbf{d}\| = 1$ . Assim, o máximo ocorre para  $i = 1$ , pois  $a_1$  é maior que os outros. Logo,

$$\max_{\|\mathbf{d}\|=\|\mathbf{e}\|=1} (\mathbf{d}'(\mathbf{X}'\mathbf{Y})\mathbf{e}) = a_1. \quad \blacksquare$$

Portanto, foram obtidos os vetores de correlação máxima  $\mathbf{f} = \mathbf{X}\mathbf{z}_1$  e  $\mathbf{g} = \mathbf{Y}\mathbf{v}_1$ . O algoritmo PLS é justamente um processo iterativo para se obter outros pares de componentes que também maximizam a covariância com mais restrições (em ordem decrescente de magnitude).

## 6.2.2 O algoritmo PLS multivariado

No sentido de melhor explicar os passos do algoritmo, seu desenvolvimento será paralelamente relacionado ao algoritmo PLS Populacional, anteriormente descrito. Com o objetivo de simplificar a notação, todo vetor da forma  $\frac{\mathbf{a}}{\mathbf{a}'\mathbf{a}}$  receberá o mesmo nome  $\mathbf{a}$ .

O algoritmo se inicia da seguinte forma: tome o vetor  $\mathbf{u}$  como a primeira coluna da matriz de respostas  $\mathbf{Y}$ , isto é,  $\mathbf{y}_1$ . Faça  $\frac{\mathbf{u}}{\mathbf{u}'\mathbf{u}}$  e seja  $\mathbf{w} = \frac{\mathbf{X}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}$ . Logo,  $\mathbf{w}$  é uma combinação linear das linhas da matriz  $\mathbf{X}$ , ou seja, é um vetor de covariâncias amostrais entre a primeira variável resposta e as covariáveis:

$$\begin{aligned} \mathbf{w} = \mathbf{X}'\mathbf{u} &= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{bmatrix} \\ &= y_{11} \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{bmatrix} + \cdots + y_{n1} \begin{bmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{np} \end{bmatrix} = \begin{bmatrix} \hat{\text{cov}}(\mathbf{x}_1, \mathbf{y}_1) \\ \hat{\text{cov}}(\mathbf{x}_2, \mathbf{y}_1) \\ \vdots \\ \hat{\text{cov}}(\mathbf{x}_p, \mathbf{y}_1) \end{bmatrix}. \end{aligned}$$

Observe que  $\mathbf{w} = \frac{\mathbf{X}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}$  é o equivalente amostral do vetor  $\mathbf{w}_1 = \boldsymbol{\sigma} = \text{cov}(\mathbf{x}, y)$ , no algoritmo PLS Populacional Univariado, com  $y = y_1$ .

Normalize  $\mathbf{w}$  e seja  $\mathbf{t} = \mathbf{X}\mathbf{w}$ :

$$\mathbf{t} = \mathbf{X}\mathbf{w} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \hat{\text{cov}}(\mathbf{x}_1, \mathbf{y}_1) \\ \hat{\text{cov}}(\mathbf{x}_2, \mathbf{y}_1) \\ \vdots \\ \hat{\text{cov}}(\mathbf{x}_p, \mathbf{y}_1) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^p x_{1i} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \sum_{i=1}^p x_{2i} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \vdots \\ \sum_{i=1}^p x_{ni} \hat{\text{cov}}(x_i, \mathbf{y}_1) \end{bmatrix}.$$

Veja que, a  $i$ -ésima entrada de  $\mathbf{t}$  corresponde a uma estimativa da variável  $t = \boldsymbol{\sigma}'\mathbf{x} = (\text{cov}(\mathbf{x}, y))'\mathbf{x}$ , com  $y = y_1$ .

Em seguida, faça  $\mathbf{c} = \frac{\mathbf{Y}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$  :

$$\mathbf{c} = \mathbf{Y}'\mathbf{t} = \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{n1} \\ y_{12} & y_{22} & \cdots & y_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1k} & y_{2k} & \cdots & y_{nk} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^p x_{1i} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \sum_{i=1}^p x_{2i} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \vdots \\ \sum_{i=1}^p x_{ni} \hat{\text{cov}}(x_i, \mathbf{y}_1) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^n \sum_{i=1}^p y_{j1} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \sum_{j=1}^n \sum_{i=1}^p y_{j2} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \vdots \\ \sum_{j=1}^n \sum_{i=1}^p y_{jk} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \end{bmatrix}.$$

Fazendo correspondência com o PLS populacional, a  $i$ -ésima entrada do vetor  $\mathbf{c}$  é uma estimativa do parâmetro  $q_1 = \text{cov}(y, \boldsymbol{\sigma}'\mathbf{x}) = \text{cov}(y, t_1)$ , com  $y = y_1$ .

Normalize  $\mathbf{c}$  e tome  $\mathbf{u} = \mathbf{Y}\mathbf{c}$  :

$$\mathbf{u} = \mathbf{Y}\mathbf{c} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix} \begin{bmatrix} \sum_{j=1}^n \sum_{i=1}^p y_{j1} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \sum_{j=1}^n \sum_{i=1}^p y_{j2} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \vdots \\ \sum_{j=1}^n \sum_{i=1}^p y_{jk} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{l=1}^k \sum_{j=1}^n \sum_{i=1}^p y_{1l} y_{j1} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \sum_{l=1}^k \sum_{j=1}^n \sum_{i=1}^p y_{2l} y_{j1} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \\ \vdots \\ \sum_{l=1}^k \sum_{j=1}^n \sum_{i=1}^p y_{nl} y_{j1} x_{ji} \hat{\text{cov}}(x_i, \mathbf{y}_1) \end{bmatrix}.$$

Este vetor não tem uma interpretação em termos do PLS populacional.

Nesta etapa, o algoritmo define um *loop* obtendo seqüências de vetores  $\mathbf{u}_n = \mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{u}_{n-1}$ ,  $\mathbf{c}_n = \mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{c}_{n-1}$ ,  $\mathbf{t}_n = \mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{t}_{n-1}$  e  $\mathbf{w}_n = \mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{w}_{n-1}$ . Estas seqüências são convergentes em razão de uma série de propriedades de álgebra linear, relativas ao diagrama da Figura 13.

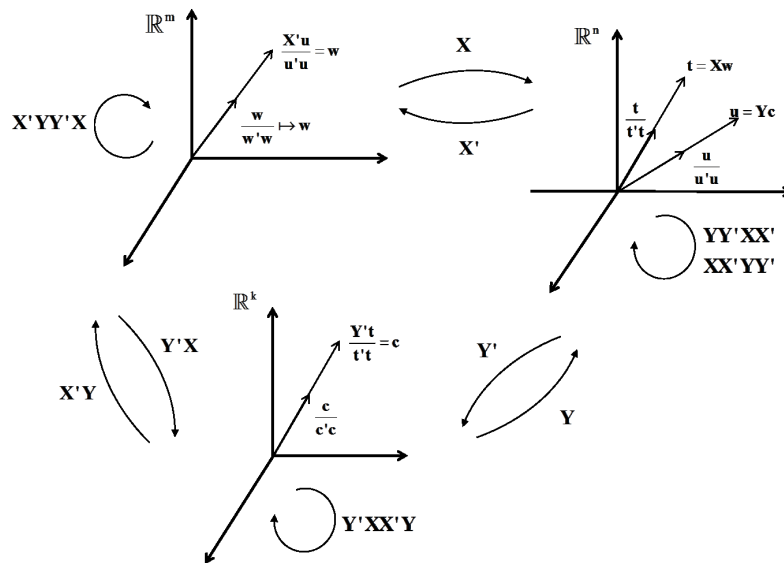


Figura 13 - Representação geométrica das seqüências de vetores  $\mathbf{u}_n = \mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{u}_{n-1}$ ,  $\mathbf{c}_n = \mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{c}_{n-1}$ ,  $\mathbf{t}_n = \mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{t}_{n-1}$  e  $\mathbf{w}_n = \mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{w}_{n-1}$

Observe que:

$$\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X} = \left( \sum_{i=1}^k a_i \mathbf{z}_i \mathbf{v}_i' \right) \left( \sum_{j=1}^k a_j \mathbf{v}_j \mathbf{z}_j' \right) = \sum_{i=1}^k \sum_{j=1}^k a_i a_j \mathbf{z}_i \mathbf{v}_i' \mathbf{v}_j \mathbf{z}_j' = \sum_{i=1}^k a_i^2 \mathbf{z}_i \mathbf{z}_i'.$$

A partir desta decomposição, os autovalores de  $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$  são  $a_1^2 \geq a_2^2 \geq \dots \geq a_k^2$ , com respectivos autovetores  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k$ , (observe que  $k < p$ ), pois:

$$\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}(\mathbf{z}_j) = \sum_{i=1}^k a_i^2 \mathbf{z}_i \mathbf{z}_i' \mathbf{z}_j = a_j^2 \mathbf{z}_j.$$

Em relação às matrizes  $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$ ,  $\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'$ ,  $\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}$  e  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'$ , da Figura13, tem-se:

- As matrizes  $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$ ,  $\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'$ ,  $\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}$  e  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'$  possuem os mesmos autovalores, pois:

se  $\mathbf{r}$  é um autovetor da matriz  $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$ , então

$$\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'(\mathbf{X}\mathbf{r}) = \mathbf{X}(\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X})\mathbf{r} = \mathbf{X}(\lambda\mathbf{r}) = \lambda(\mathbf{X}\mathbf{r}).$$

Se  $\mathbf{r}$  é um autovetor da matriz  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'$ , então

$$\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{r}) = \mathbf{Y}'(\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}')\mathbf{r} = \mathbf{Y}'(\lambda\mathbf{r}) = \lambda(\mathbf{Y}'\mathbf{r}).$$

Se  $\mathbf{r}$  é um autovetor da matriz  $\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}$ , então

$$\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'(\mathbf{Y}\mathbf{r}) = \mathbf{Y}(\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y})\mathbf{r} = \mathbf{Y}(\lambda\mathbf{r}) = \lambda(\mathbf{Y}\mathbf{r}).$$

Portanto, o maior autovalor para todas as matrizes é dado por  $a_1^2$ .

- Aplicando o método das potências para o cálculo do autovetor relativo ao maior autovalor às sequências  $\mathbf{u}_n = \mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{u}_{n-1}$ ,  $\mathbf{c}_n = \mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{c}_{n-1}$ ,  $\mathbf{t}_n = \mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{t}_{n-1}$ , e  $\mathbf{w}_n = \mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{w}_{n-1}$ , obtém-se os vetores limites das sequências  $\mathbf{u} = \lim_{n \rightarrow \infty} \mathbf{u}_n$ ,  $\mathbf{c} = \lim_{n \rightarrow \infty} \mathbf{c}_n$ ,  $\mathbf{t} = \lim_{n \rightarrow \infty} \mathbf{t}_n$  e  $\mathbf{w} = \lim_{n \rightarrow \infty} \mathbf{w}_n$ . Assim:

$$\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{u} = a_1^2\mathbf{u}$$

$$\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{c} = a_1^2\mathbf{c}$$

$$\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{t} = a_1^2\mathbf{t}$$

$$\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{w} = a_1^2\mathbf{w}.$$

Ocorrendo a convergência pela aplicação de algum critério de parada, ficam definidos os vetores  $\mathbf{w}$  e  $\mathbf{c}$ , relativos ao maior autovalor  $a_1^2$ . Pelo problema de maximização das componentes com maior covariância, os vetores relativos ao maior autovalor são dados explicitamente a partir da decomposição em valores singulares de  $\mathbf{X}'\mathbf{Y}$ , isto é,  $\mathbf{w} = \mathbf{z}_1$  e  $\mathbf{c} = \mathbf{v}_1$ . Portanto, as componentes com maior covariância são dadas por  $\mathbf{t} = \mathbf{X}\mathbf{w}$  e  $\mathbf{u} = \mathbf{Y}\mathbf{c}$ .

O próximo passo do algoritmo é definido por: primeiramente se calcula a covariância amostral entre as componentes  $\mathbf{t} = \mathbf{X}\mathbf{w}$  e  $\mathbf{u} = \mathbf{Y}\mathbf{c}$ , dada por  $b = \frac{\mathbf{u}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$ . Como o vetor  $\mathbf{t}$  é a combinação linear das colunas de  $\mathbf{X}$  que possui maior covariância com uma combinação linear das colunas de  $\mathbf{Y}$ , a ideia agora é obter uma nova matriz  $\mathbf{X}$  de tal forma que as colunas desta nova matriz  $\mathbf{X}$  sejam ortogonais ao vetor  $\mathbf{t}$ . A projeção ortogonal no subespaço ortogonal ao vetor  $\mathbf{t}$  é dada por  $\mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}$ . De fato,  $\mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}$  é simétrico e idempotente, pois:

$$\left(\mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}\right) \left(\mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}\right) = \mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}} + \frac{\mathbf{t}\mathbf{t}'\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}} = \mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}.$$

Portanto, a nova matriz  $\mathbf{X}$  é dada por

$$\left(\mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}\right) \mathbf{X}.$$

Assim,

$$\left(\mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}\right) \mathbf{X} = \mathbf{X} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}\mathbf{X} = \mathbf{X} - \mathbf{t} \frac{\mathbf{t}'\mathbf{X}}{\mathbf{t}'\mathbf{t}} = \mathbf{X} - \mathbf{t}\mathbf{p}',$$

em que  $\mathbf{p} = \frac{\mathbf{X}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$ .

O vetor  $\mathbf{p}$  representa o vetor de covariâncias amostrais entre as covariáveis dadas pelas colunas de  $\mathbf{X}$  e o vetor de estimativas das variáveis preditoras  $\mathbf{t}$ , e a

nova matriz  $\tilde{\mathbf{X}}$  representa as componentes das covariáveis que não foram utilizadas na predição, sendo denominada matriz residual e que será utilizada na próxima iteração do algoritmo.

Uma construção semelhante é feita para a matriz  $\mathbf{Y}$ . O melhor preditor das variáveis respostas  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_k \end{bmatrix}$  é o vetor  $\text{cov}(\mathbf{t}, \mathbf{y}) = \begin{bmatrix} \text{cov}(\mathbf{t}, \mathbf{y}_1) \\ \text{cov}(\mathbf{t}, \mathbf{y}_2) \\ \vdots \\ \text{cov}(\mathbf{t}, \mathbf{y}_k) \end{bmatrix}$ , que é estimado pelo vetor  $\mathbf{c}$ , em que,

$$\mathbf{c}_{k \times 1} = \frac{\mathbf{Y}'\mathbf{t}}{\mathbf{t}'\mathbf{t}} = \begin{bmatrix} \mathbf{Y}'_1 \\ \mathbf{Y}'_2 \\ \vdots \\ \mathbf{Y}'_k \end{bmatrix}_{k \times n} \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} \frac{\mathbf{Y}'_1 \mathbf{t}}{\mathbf{t}'\mathbf{t}} \\ \frac{\mathbf{Y}'_2 \mathbf{t}}{\mathbf{t}'\mathbf{t}} \\ \vdots \\ \frac{\mathbf{Y}'_k \mathbf{t}}{\mathbf{t}'\mathbf{t}} \end{bmatrix}_{k \times 1}.$$

Segue que,

$$\mathbf{t}\mathbf{c}' = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}_{n \times 1} \begin{bmatrix} \frac{\mathbf{Y}'_1 \mathbf{t}}{\mathbf{t}'\mathbf{t}}, \frac{\mathbf{Y}'_2 \mathbf{t}}{\mathbf{t}'\mathbf{t}}, \dots, \frac{\mathbf{Y}'_k \mathbf{t}}{\mathbf{t}'\mathbf{t}} \end{bmatrix}_{1 \times k} = \begin{bmatrix} \frac{\mathbf{Y}'_1 \mathbf{t}}{\mathbf{t}'\mathbf{t}} \mathbf{t}, \frac{\mathbf{Y}'_2 \mathbf{t}}{\mathbf{t}'\mathbf{t}} \mathbf{t}, \dots, \frac{\mathbf{Y}'_k \mathbf{t}}{\mathbf{t}'\mathbf{t}} \mathbf{t} \end{bmatrix}_{n \times k}.$$

As colunas da matriz  $\mathbf{t}\mathbf{c}'$  representam a parte que o vetor  $\mathbf{t}$  explica de  $\mathbf{Y}$ . Tem-se, portanto, duas fontes de explicação da variável  $\mathbf{Y}$ : a parte de  $\mathbf{t}$  que explica  $\mathbf{u}$ , dada por  $\mathbf{b} = \frac{\mathbf{u}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$ , que é uma combinação linear das colunas de  $\mathbf{Y}$ , e a parte de  $\mathbf{t}$  que explica  $\mathbf{Y}$ , dada por  $\frac{\mathbf{t}'\mathbf{Y}_i}{\mathbf{t}'\mathbf{t}}$ . Portanto, a quantidade de informação que o vetor  $\mathbf{t}$  tem do vetor  $\mathbf{u}$  e dos vetores colunas da matriz  $\mathbf{Y}$  é obtida pelo produto  $\left(\frac{\mathbf{u}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\right) \left(\frac{\mathbf{t}'\mathbf{Y}_i}{\mathbf{t}'\mathbf{t}}\right) = \mathbf{b} \frac{\mathbf{t}'\mathbf{Y}_i}{\mathbf{t}'\mathbf{t}}$ .

Assim, ponderando a quantidade de informação pelo produto, a parte não explicada de  $\mathbf{Y}$  é a nova matriz  $\tilde{\mathbf{Y}}$ , dada pela diferença

$$\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{b}\mathbf{t} \left(\frac{\mathbf{t}'\mathbf{Y}'}{\mathbf{t}'\mathbf{t}}\right) = \mathbf{Y} - \mathbf{b}\mathbf{t}\mathbf{c}'.$$

Obtidas as novas matrizes  $\tilde{\mathbf{X}}$  e  $\tilde{\mathbf{Y}}$ , o processo é repetido, ficando o algoritmo da forma:

Algoritmo para o PLS Amostral (HOSKULDSSON, 1988):

1. Seja  $\mathbf{u}$  a primeira coluna de  $\tilde{\mathbf{Y}}$
2.  $\mathbf{w} = \frac{\tilde{\mathbf{X}}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}$
3. Normalize  $\mathbf{w}$



4.  $\mathbf{t} = \mathbf{X}\mathbf{w}$
5.  $\mathbf{c} = \frac{\mathbf{Y}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$
6. Normalize  $\mathbf{c}$
7.  $\mathbf{u} = \mathbf{Y}\mathbf{c}$
8. Se houver convergência vá para o passo 9, caso contrário, volte ao passo 2
9.  $\mathbf{X}$ -loadings:  $\mathbf{p} = \frac{\mathbf{X}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$
10.  $\mathbf{Y}$ -loadings:  $\mathbf{q} = \frac{\mathbf{Y}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}$
11. Regressão ( $\mathbf{u}$  sobre  $\mathbf{t}$ ):  $b = \frac{\mathbf{u}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$
12. Matrizes Residuais:  $\mathbf{X} \leftarrow \mathbf{X} - \mathbf{t}\mathbf{p}'$  e  $\mathbf{Y} \leftarrow \mathbf{Y} - \mathbf{b}\mathbf{t}\mathbf{c}'$

Realizadas A iterações deste algoritmo, obtêm-se os vetores:

$$\begin{aligned} &\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_A, \\ &\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A, \\ &\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_A, \\ &\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_A, \\ &\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_A. \end{aligned}$$

É importante observar que:

- i. Da construção do algoritmo, tem-se que a sequência dos vetores de componentes  $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A$ , formam uma sequência de vetores ortogonais.
- ii. Os vetores  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_A$ , são vetores mutuamente ortogonais.

**Prova:**

Do algoritmo, sabe-se que:

$$\begin{aligned} \mathbf{X}_2 &= \left( \mathbf{I} - \frac{\mathbf{t}_1\mathbf{t}'_1}{\mathbf{t}'_1\mathbf{t}_1} \right) \mathbf{X}_1 \\ \mathbf{X}_3 &= \left( \mathbf{I} - \frac{\mathbf{t}_2\mathbf{t}'_2}{\mathbf{t}'_2\mathbf{t}_2} \right) \mathbf{X}_2 = \left( \mathbf{I} - \frac{\mathbf{t}_2\mathbf{t}'_2}{\mathbf{t}'_2\mathbf{t}_2} \right) \left( \mathbf{I} - \frac{\mathbf{t}_1\mathbf{t}'_1}{\mathbf{t}'_1\mathbf{t}_1} \right) \mathbf{X}_1 \\ &\vdots \\ \mathbf{X}_j &= \left( \mathbf{I} - \frac{\mathbf{t}_{j-1}\mathbf{t}'_{j-1}}{\mathbf{t}'_{j-1}\mathbf{t}_{j-1}} \right) \mathbf{X}_{j-1} \\ &= \left( \mathbf{I} - \frac{\mathbf{t}_{j-1}\mathbf{t}'_{j-1}}{\mathbf{t}'_{j-1}\mathbf{t}_{j-1}} \right) \left( \mathbf{I} - \frac{\mathbf{t}_{j-2}\mathbf{t}'_{j-2}}{\mathbf{t}'_{j-2}\mathbf{t}_{j-2}} \right) \dots \left( \mathbf{I} - \frac{\mathbf{t}_1\mathbf{t}'_1}{\mathbf{t}'_1\mathbf{t}_1} \right) \mathbf{X}_1. \end{aligned}$$

Assim, para  $i < j$ , tem-se:

$$\mathbf{X}_j = \mathbf{Z} \left( \mathbf{I} - \frac{\mathbf{t}_i \mathbf{t}_i'}{\mathbf{t}_i' \mathbf{t}_i} \right) \mathbf{X}_i,$$

em que  $\mathbf{Z}$  é o produto de matrizes  $\left( \mathbf{I} - \frac{\mathbf{t}_k \mathbf{t}_k'}{\mathbf{t}_k' \mathbf{t}_k} \right)$ ,  $k = i + 1, \dots, j$ .

De onde segue que, para  $i < j$ :

$$\begin{aligned} \mathbf{X}_j \mathbf{w}_i &= \mathbf{Z} \left( \mathbf{X}_i - \frac{\mathbf{t}_i \mathbf{t}_i'}{\mathbf{t}_i' \mathbf{t}_i} \mathbf{X}_i \right) \mathbf{w}_i = \mathbf{Z} \left( \mathbf{X}_i \mathbf{w}_i - \frac{\mathbf{t}_i \mathbf{t}_i'}{\mathbf{t}_i' \mathbf{t}_i} \mathbf{X}_i \mathbf{w}_i \right) \\ &= \mathbf{Z} \left( \mathbf{t}_i - \frac{\mathbf{t}_i \mathbf{t}_i'}{\mathbf{t}_i' \mathbf{t}_i} \mathbf{t}_i \right) = \mathbf{Z} (\mathbf{t}_i - \mathbf{t}_i) = 0. \end{aligned}$$

Sabendo que  $\mathbf{X}_j' \mathbf{Y}_j \mathbf{Y}_j' \mathbf{X}_j \mathbf{w}_j = a_j \mathbf{w}_j$ , obtém-se:

$$\begin{aligned} \mathbf{w}_j' \mathbf{w}_i &= \left( \frac{\mathbf{X}_j' \mathbf{Y}_j \mathbf{Y}_j' \mathbf{X}_j \mathbf{w}_j}{a_j} \right)' \mathbf{w}_i = \frac{\mathbf{w}_j' \mathbf{X}_j' \mathbf{Y}_j \mathbf{Y}_j' \mathbf{X}_j \mathbf{w}_i}{a_j} \\ &= \frac{\mathbf{w}_j' \mathbf{X}_j' \mathbf{Y}_j \mathbf{Y}_j'}{a_j} (\mathbf{X}_j \mathbf{w}_i) = 0. \end{aligned}$$

■

iii. Os vetores  $\mathbf{w}_i$  são ortogonais aos vetores  $\mathbf{p}_j$ , para  $i < j$ .

**Prova:**

$$\mathbf{w}_i' \mathbf{p}_j = \mathbf{w}_i' \frac{\mathbf{X}_j' \mathbf{t}_j}{\mathbf{t}_j' \mathbf{t}_j} = \mathbf{w}_i' \mathbf{X}_j' \frac{\mathbf{t}_j}{\mathbf{t}_j' \mathbf{t}_j} = (\mathbf{X}_j \mathbf{w}_i)' \frac{\mathbf{t}_j}{\mathbf{t}_j' \mathbf{t}_j} = 0.$$

■

iv. A sequência de vetores  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_A$ , também é uma sequência de vetores ortogonais.

### 6.2.3 Regressão em PLS

Uma das questões mais importantes em análise de regressão é reduzir o número de variáveis independentes. Uma justificativa que torna vantajosa a seleção de variáveis é que a variância dos coeficientes de regressão aumenta quando novas variáveis são introduzidas no modelo (HOSKULDSSON, 1988). A regressão PLS pode ser vista como um bom método de análise de regressão porque os componentes são selecionados de modo que eles descrevem as variáveis respostas. A principal característica é que o método PLS é capaz de reduzir o número de covariáveis, para um número menor de componentes, durante o processo de estimação. Vamos considerar aqui, a regressão via método PLS.

Um modelo de regressão linear múltipla e multivariada pode ser escrito como

$$\mathbf{Y}_{n \times k} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times k} + \mathbf{e}_{n \times k},$$

em que  $\boldsymbol{\beta}$  é uma matriz de coeficientes de regressão,  $\mathbf{X}$  é a matriz de covariáveis e  $\mathbf{e}$  é a matriz de erros.

Sendo obtidos os vetores  $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A$ , em  $A$  iterações, a matriz do delineamento  $\mathbf{X}$  pode ser decomposta da forma

$$\mathbf{X}_{n \times p} = \sum_{i=1}^A \mathbf{t}_i \mathbf{p}_i' + \mathbf{X}_0,$$

onde  $\mathbf{X}_0$  é a matriz residual do processo.

Note que se o número de iterações  $A$  é igual ao posto de  $\mathbf{X}$ , então  $\mathbf{X}_0 = \mathbf{0}$ .

Desprezando a matriz residual  $\mathbf{X}_0$ , tem-se

$$\mathbf{X}_{n \times p} = \sum_{i=1}^A \mathbf{t}_i \mathbf{p}_i' = \mathbf{T}_{n \times A} \mathbf{P}'_{A \times p}$$

com  $\mathbf{T}_{n \times A} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A)$  em que  $\mathbf{t}_i$  é a  $i$ -ésima coluna de  $\mathbf{T}$  e  $\mathbf{P}_{p \times A} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_A)$  em que  $\mathbf{p}_i$  é a  $i$ -ésima coluna de  $\mathbf{P}$ .

Como os vetores  $\mathbf{p}_i$  são ortonormais, multiplica-se a igualdade  $\mathbf{X}_{n \times p} = \mathbf{T}_{n \times A} \mathbf{P}'_{A \times p}$  por  $\mathbf{P}_{p \times A}$ , obtendo

$$\mathbf{X}_{n \times p} \mathbf{P}_{p \times A} = \mathbf{T}_{n \times A} \mathbf{P}'_{A \times p} \mathbf{P}_{p \times A} = \mathbf{T}_{n \times A} \mathbf{I}_{A \times A} = \mathbf{T}_{n \times A}.$$

Observe que  $\mathbf{P}$  é uma inversa generalizada à direita de  $\mathbf{P}'$ . De fato, qualquer inversa generalizada poderia ser utilizada (JONG, 1993).

Assim,  $\mathbf{T} = \mathbf{X}\mathbf{P}$ , e a equação de regressão pode ser escrita como

$$\mathbf{Y}_{n \times k} = \mathbf{X}_{n \times p} \mathbf{P}_{p \times A} \mathbf{P}'_{A \times p} \boldsymbol{\beta}_{p \times k} + \mathbf{e}_{n \times k} = \mathbf{T}_{n \times A} (\boldsymbol{\beta}_{\text{PLS}, A})_{A \times k} + \mathbf{e}_{n \times k},$$

em que  $\mathbf{T}$  é a nova matriz de covariáveis, denominada matriz de componentes PLS, e  $\boldsymbol{\beta}_{\text{PLS}, A} = \mathbf{P}'\boldsymbol{\beta}$  é a nova matriz de parâmetros de regressão, denominada matriz de parâmetros PLS.

Com os novos parâmetros PLS, temos um problema de regressão usual, porém, agora, com o número de parâmetros muito menor que do problema original. O vetor de parâmetros,  $\boldsymbol{\beta}_{\text{PLS}, A}$ , é estimado pelo método dos quadrados mínimos, isto é,

$$\left( \hat{\boldsymbol{\beta}}_{\text{PLS}, A} \right)_{A \times k} = (\mathbf{T}'\mathbf{T})_{A \times A}^{-1} \mathbf{T}'_{A \times n} \mathbf{Y}_{n \times k}. \quad (2)$$

Obtidas as estimativas dos coeficientes de regressão, a equação de predição para novos valores  $\mathbf{x}_0$  das covariáveis, retornando às variáveis originais não centradas, fica da forma

$$\hat{\mathbf{y}} - \hat{\boldsymbol{\mu}}_{\mathbf{y}} = (\mathbf{t} - \hat{\boldsymbol{\mu}}_{\mathbf{t}})' \hat{\boldsymbol{\beta}}_{\text{PLS}, A} = (\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_{\mathbf{x}})' \mathbf{P} \hat{\boldsymbol{\beta}}_{\text{PLS}, A}. \quad (3)$$

### 6.3 Uma alternativa ao algoritmo PLS

Esta seção tem como referências os artigos Garthwaite (1994) e Phatak e Jong (1997), e usará a mesma notação. A descrição deste algoritmo é feita para apenas uma variável resposta,  $k = 1$ .

Como anteriormente exposto, a matriz de covariáveis  $\mathbf{X}_{n \times p}$ , é considerada como uma transformação do espaço dos parâmetros  $\mathbb{R}^p$  no espaço dos dados  $\mathbb{R}^n$ , e o vetor da variável resposta é  $\mathbf{Y}_{n \times 1}$ , conforme Figura 14.

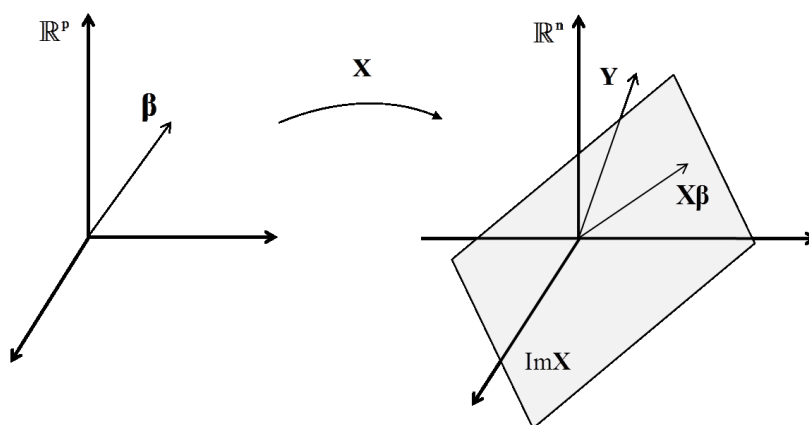


Figura 14 - Representação da matriz de covariáveis  $\mathbf{X}_{n \times p}$  como uma transformação do espaço de parâmetros  $\mathbb{R}^p$  no espaço dos dados  $\mathbb{R}^n$ .

A matriz  $\mathbf{X}$  aplicada ao vetor canônico  $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$  tem como imagem a  $i$ -ésima coluna de  $\mathbf{X}$ , denotada por  $\mathbf{X}_i$  (Figura 15).

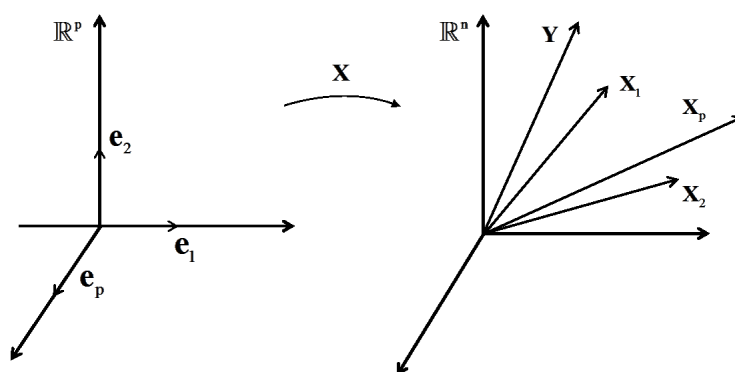


Figura 15 - Representação das colunas de  $\mathbf{X}$  como imagem dos vetores canônicos  $\mathbf{e}_i$ .

Como é usual, utilizam-se variáveis centradas na média:

$$\mathbf{U}_1 = \mathbf{Y} - \bar{\mathbf{Y}} \text{ e } \mathbf{V}_{1j} = \mathbf{X}_j - \bar{\mathbf{X}}_j \text{ para } j = 1, \dots, p,$$

sendo  $\bar{\mathbf{Y}}' = (\bar{y}, \bar{y}, \dots, \bar{y})'$ ,  $\bar{\mathbf{X}}_j = (\bar{x}_j, \bar{x}_j, \dots, \bar{x}_j)$  e  $\mathbf{1}' = (1, 1, \dots, 1)'$ , vetores linha  $n$ -dimensionais. Observe, como na Figura 16, que os vetores  $\mathbf{Y} - \bar{\mathbf{Y}}$  e  $\mathbf{X}_j - \bar{\mathbf{X}}_j$  são ortogonais ao subespaço gerado pelo vetor  $\mathbf{1}$ .

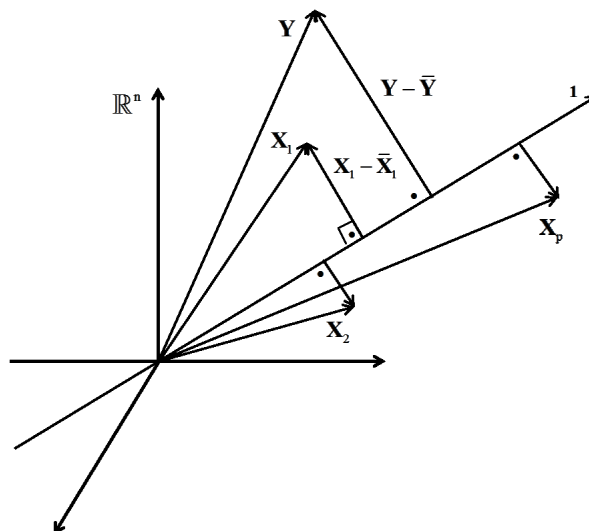


Figura 16 - Representação das variáveis centradas através da projeção ortogonal no subespaço gerado pelo vetor  $\mathbf{1}$ .

A ideia é regressir o vetor  $\mathbf{U}_1$  em cada um dos vetores  $\mathbf{V}_{1j}$  (Figuras 17 e 18). Tal procedimento é uma forma de medir o quanto de informação o vetor da covariável  $\mathbf{X}_j$  explica o vetor de dados, obtendo-se o vetor  $\mathbf{U}_{1j} = \frac{\mathbf{u}_1' \mathbf{v}_{1j}}{\mathbf{v}_{1j}' \mathbf{v}_{1j}} \mathbf{v}_{1j}$ .

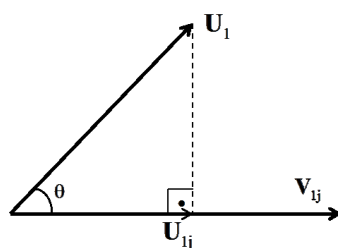


Figura 17 - Projeção ortogonal do vetor  $\mathbf{U}_1$  em cada um dos vetores  $\mathbf{V}_{1j}$ .

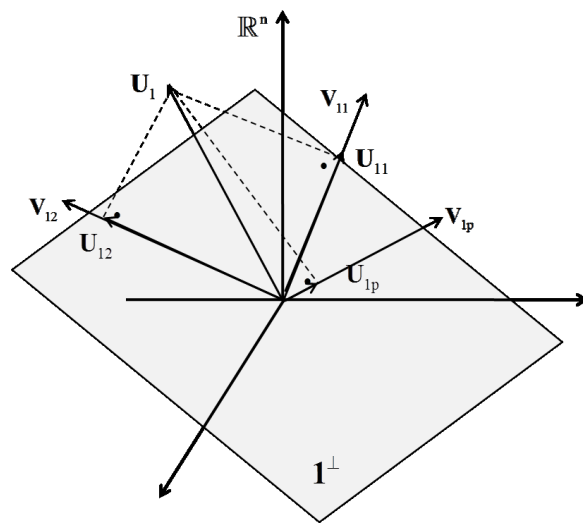


Figura 18 - Projeção ortogonal do vetor  $\mathbf{U}_1$  em cada um dos vetores  $\mathbf{V}_{1j}$ .

Uma maneira de se coletarem todas estas informações sobre o vetor de dados é através de uma média ponderada dos vetores  $\mathbf{U}_{1j}$ , utilizando pesos apropriadamente escolhidos,  $w_{1j} \geq 0$  e  $\sum_{j=1}^p w_{1j} = 1$ , como na Figura 19.

$$\mathbf{T}_1 = \sum_{j=1}^p w_{1j} \mathbf{U}_{1j}. \quad (4)$$

Uma escolha usual para os pesos é a média simples,  $w_{1j} = 1/p$ .

A quantidade de informação que  $\mathbf{T}_1$  contém do vetor de dados  $\mathbf{U}_1$  é obtida pela regressão do vetor  $\mathbf{U}_1$  em  $\mathbf{T}_1$ ,  $P_{\mathbf{T}_1} \mathbf{U}_1 = \frac{\mathbf{u}_1' \mathbf{t}_1}{\mathbf{t}_1' \mathbf{t}_1} \mathbf{t}_1$ . A parte não explicada é definida pelo vetor de resíduo  $\mathbf{U}_2 = \mathbf{U}_1 - P_{\mathbf{T}_1} \mathbf{U}_1$ . Novamente, realiza-se a regressão de  $\mathbf{V}_{1j}$  em  $\mathbf{T}_1$ ,  $P_{\mathbf{T}_1} \mathbf{V}_{1j} = \frac{\mathbf{v}_{1j}' \mathbf{t}_1}{\mathbf{t}_1' \mathbf{t}_1} \mathbf{t}_1$ , e toma-se os vetores de resíduos  $\mathbf{V}_{2j} = \mathbf{V}_{1j} - P_{\mathbf{T}_1} \mathbf{V}_{1j}$ , que são uma medida do quanto o vetor  $\mathbf{T}_1$  não explica as covariáveis, conforme Figura 20.

Procede-se à regressão do vetor  $\mathbf{U}_2$  nos vetores  $\mathbf{V}_{2j}$ , que são denotados por  $\mathbf{U}_{2j} = P_{\mathbf{V}_{2j}} \mathbf{U}_2 = b_{2j} \mathbf{V}_{2j} = \frac{\mathbf{U}_2' \mathbf{V}_{2j}}{\mathbf{V}_{2j}' \mathbf{V}_{2j}} \mathbf{V}_{2j}$  (Figura 21).

Utilizando pesos  $w_{2j}$ , encontramos o vetor  $\mathbf{T}_2$  fazendo a média ponderada dos vetores  $\mathbf{U}_{2j}$  (Figura 22).

Observe que, por construção, os vetores  $\mathbf{T}_1$  e  $\mathbf{T}_2$  são ortogonais (Figura 23).

O procedimento pode ser repetido para a construção de outros vetores  $\mathbf{T}_3, \mathbf{T}_4, \dots, \mathbf{T}_m$ .

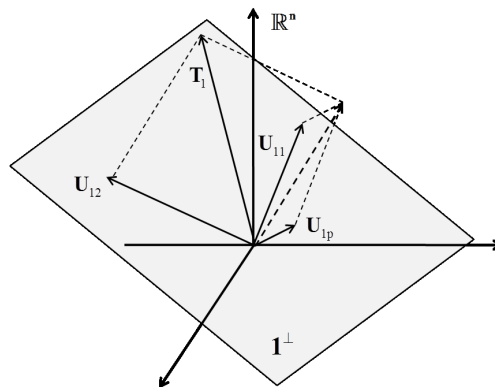


Figura 19 - Construção do vetor  $\mathbf{T}_1$  como média ponderada dos vetores  $\mathbf{U}_{1j}$ .

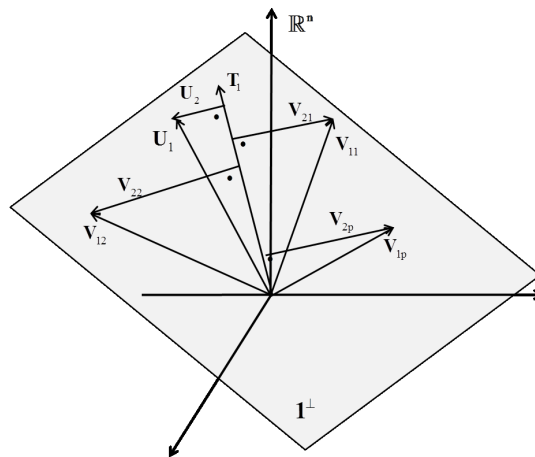


Figura 20 - Construção dos vetores  $\mathbf{U}_2$  e  $\mathbf{V}_{2j}$  através da projeção dos vetores  $\mathbf{U}_1$  e  $\mathbf{V}_{1j}$ , respectivamente, em  $\mathbf{T}_1$ .

Suponha que  $\mathbf{T}_i$  ( $i \geq 1$ ) tenha sido construído a partir das variáveis  $\mathbf{U}_i$  e  $\mathbf{V}_{ij}$  com  $j = 1, \dots, p$ . Para obter o componente  $\mathbf{T}_{i+1}$ , as variáveis  $\mathbf{V}_{(i+1)j}$  e  $\mathbf{U}_{(i+1)}$  devem ser determinadas. Com este propósito, é feita uma regressão entre  $\mathbf{V}_{ij}$  e  $\mathbf{T}_i$  e,  $\mathbf{V}_{(i+1)j}$  fica definido por:

$$\mathbf{V}_{(i+1)j} = \mathbf{V}_{ij} - \frac{\mathbf{v}'_{ij} \mathbf{t}_i}{\mathbf{t}'_i \mathbf{t}_i} \mathbf{t}_i,$$

sendo  $\mathbf{t}_i$  o vetor de valores de  $\mathbf{T}_i$ ,  $\mathbf{v}_{(i+1)j}$  os resíduos da regressão e  $\frac{\mathbf{v}'_{ij} \mathbf{t}_i}{\mathbf{t}'_i \mathbf{t}_i}$ , o coeficiente da regressão entre  $\mathbf{V}_{ij}$  e  $\mathbf{T}_i$ .

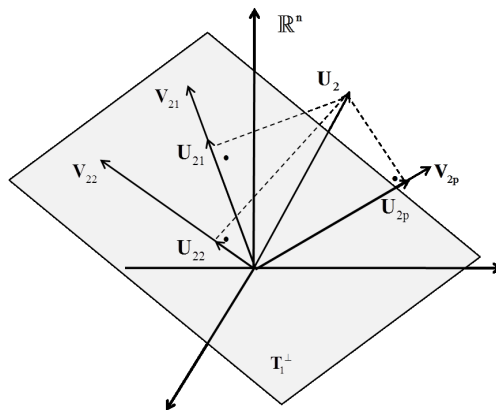


Figura 21 - Construção dos vetores  $U_{2j}$  como projeção do vetor  $U_2$  nos vetores  $V_{2j}$ .

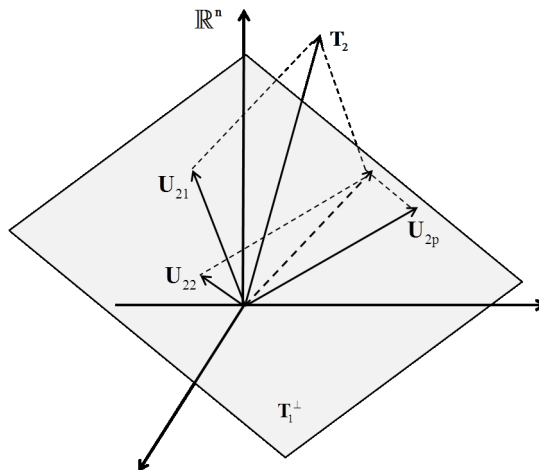


Figura 22 - Construção do vetor  $T_2$  como média ponderada dos vetores  $U_{2j}$

De forma análoga,  $U_{(i+1)} = U_i - \frac{u'_i t_i}{t'_i t_i} t_i$  e  $u_{(i+1)}$  são os resíduos da regressão entre  $U_i$  e  $T_i$ . Assim, a  $j$ -ésima regressão entre  $U_{(i+1)}$  e  $V_{(i+1)j}$  resulta num coeficiente de regressão dado por:

$$b_{(i+1)j} = \frac{u'_{(i+1)} v_{(i+1)j}}{v'_{(i+1)j} v_{(i+1)j}}$$



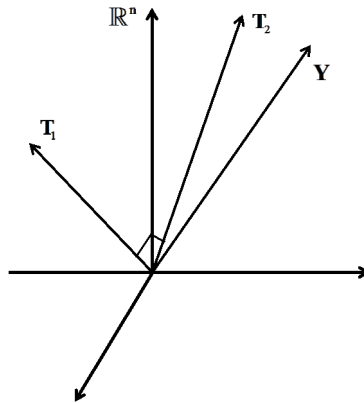


Figura 23 - Componentes ortogonais  $\mathbf{T}_1$  e  $\mathbf{T}_2$

Analogamente à equação 4, define-se  $\mathbf{T}_{(i+1)}$  como sendo:

$$\mathbf{T}_{(i+1)} = \sum_{j=1}^p w_{(i+1)j} b_{(i+1)j} \mathbf{V}_{(i+1)j}.$$

Os vetores  $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_m$  são ortogonais e, portanto, como vetores aleatórios, são não correlacionados (no caso em que  $\text{cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}$ ).

Construídos os vetores  $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_m$  fica definido na  $\text{Im}\mathbf{X}$ , um subespaço  $m$ -dimensional,  $\text{Im}\mathbf{T}$ . Projeta-se ortogonalmente o vetor  $\mathbf{Y}$  no subespaço  $\text{Im}\mathbf{T}$  obtendo-se  $\hat{\mathbf{Y}}_{\text{PLS}} = P_{\text{Im}\mathbf{T}} \mathbf{Y}$ . O processo de estimação do  $\hat{\boldsymbol{\beta}}_{\text{PLS}}$  é o usual utilizando o método dos quadrados mínimos, isto é,  $\hat{\boldsymbol{\beta}}_{\text{PLS}}$  é obtido como uma solução particular, bem determinada, das equações normais  $\mathbf{X} \hat{\boldsymbol{\beta}}_{\text{PLS}} = P_{\text{Im}\mathbf{T}} \mathbf{Y}$ . Uma vez obtida a estimativa  $\hat{\boldsymbol{\beta}}_{\text{PLS}}$ , a equação de predição fica da forma

$$\mathbf{y} - \bar{y} = \hat{\boldsymbol{\beta}}'_{\text{PLS}} \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}.$$

Para obter a expressão matricial das equações normais que definem o  $\hat{\boldsymbol{\beta}}_{\text{PLS}}$  e também obter uma relação linear entre  $\hat{\boldsymbol{\beta}}_{\text{PLS}}$  e  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ , foi utilizada uma formalização do método em termos de matrizes, isto é, transformações lineares.

Suponha que  $m$  componentes  $t_1, t_2, \dots, t_m$  tenham sido construídos. Considere a matriz  $\mathbf{T}_{n \times m}$ , cujas colunas são definidas pelos vetores  $\mathbf{t}_i$ ,  $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m]$ .  $\mathbf{T}$  define uma transformação linear do  $\mathbb{R}^m$  para o  $\mathbb{R}^n$ , tal que  $\mathbf{T}\mathbf{e}_i = \mathbf{t}_i$ , conforme Figura 24.

A imagem de  $\mathbf{T}$ ,  $\text{Im}\mathbf{T}$ , é um subespaço da imagem de  $\mathbf{X}$  gerado pelos vetores  $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m$ . Portanto, a projeção ortogonal sobre a  $\text{Im}\mathbf{T}$ , é dada pela matriz  $\mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'$ , e a projeção do vetor de dados  $\mathbf{Y}$  na  $\text{Im}\mathbf{T}$  é dada por (Figura 25):

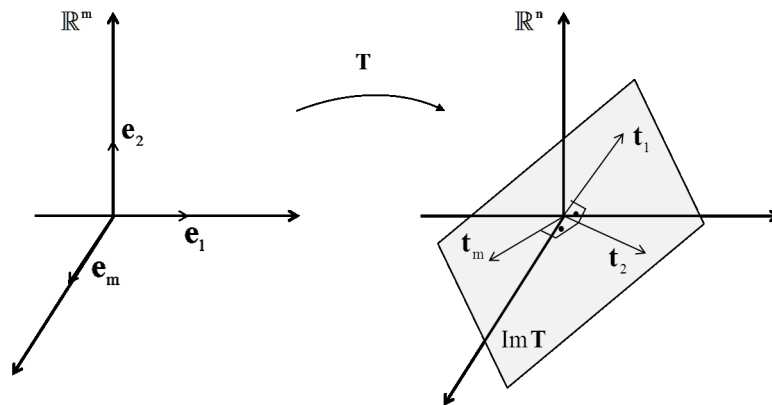


Figura 24 - Representação geométrica dos componentes como imagem da transformação linear  $\mathbf{T}$  do  $\mathbb{R}^m$  para o  $\mathbb{R}^n$ .

$$\hat{\mathbf{y}}_{\text{PLS}}^m = \mathbf{T}_m(\mathbf{T}_m' \mathbf{T}_m)^{-1} \mathbf{T}_m' \mathbf{y}. \quad (5)$$

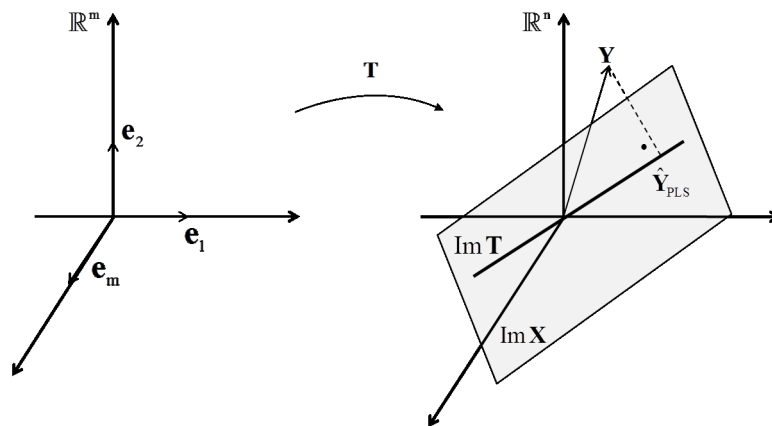


Figura 25 - Projeção do vetor de dados  $\mathbf{Y}$  no subespaço  $\text{Im T}$  gerado pelos componentes.

Seja o vetor  $\mathbf{r}_i$  definido como uma solução qualquer da equação  $\mathbf{X}\mathbf{r}_i = \mathbf{t}_i$ ,  $i=1, \dots, m$ . Da mesma forma, pode-se definir a matriz  $\mathbf{R}_{p \times m}$ , cujas colunas são os vetores  $\mathbf{r}_i$ ,  $i=1, \dots, m$ ,  $\mathbf{R}_m = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m]$ . Assim,  $\mathbf{R}\mathbf{e}_i = \mathbf{r}_i$ .

Como  $\mathbf{T}\mathbf{e}_i = \mathbf{t}_i$  e  $\mathbf{X}\mathbf{R}\mathbf{e}_i = \mathbf{t}_i$ , tem-se que  $\mathbf{T}\mathbf{e}_i = \mathbf{X}\mathbf{R}\mathbf{e}_i$ . Logo,  $\mathbf{T} = \mathbf{X}\mathbf{R}$  e obtém-se o diagrama comutativo da Figura 26.

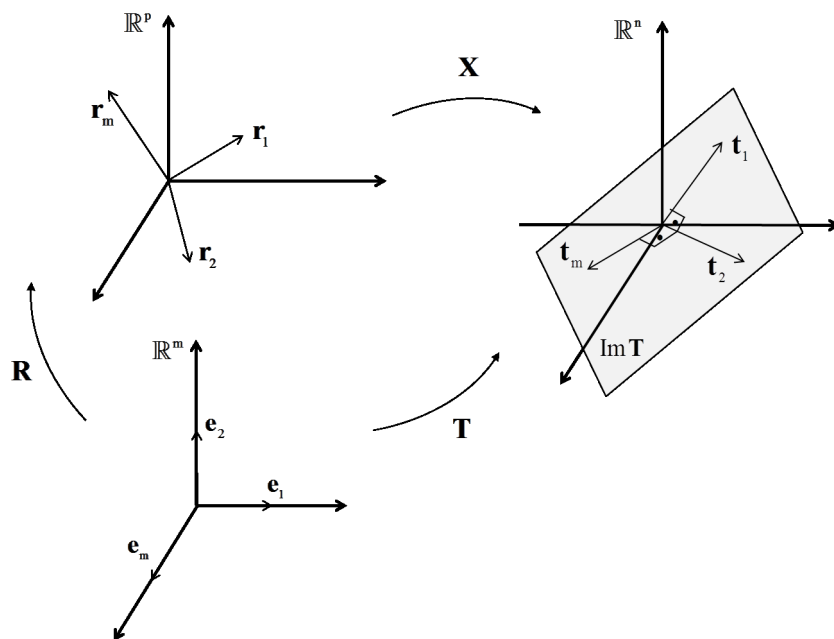


Figura 26 - Representação geométrica das transformações lineares  $\mathbf{T}$ ,  $\mathbf{R}$  e  $\mathbf{X}$ .

A matriz  $\mathbf{T}$  é claramente uma aplicação linear injetiva, pois, suas colunas são linearmente independentes. Assim, restrita à imagem de  $\mathbf{R}$ , segue que,  $\mathbf{X}$  também é uma aplicação injetiva. Vetorialmente,  $\text{Im}\mathbf{R} \cap \text{Ker}\mathbf{X} = \{\mathbf{0}\}$ .

Substituindo  $\mathbf{T}$  por  $\mathbf{XR}$  e  $\mathbf{X}'\mathbf{y}$  por  $\mathbf{X}'\mathbf{X}\hat{\beta}_{\text{OLS}}$ , na equação (5), obtemos:

$$\hat{\mathbf{y}}_{\text{PLS}}^m = \mathbf{XR}_m \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{R}_m' \mathbf{X}' \mathbf{X} \hat{\beta}_{\text{OLS}}.$$

Observe que a substituição de  $\mathbf{X}'\mathbf{y}$  por  $\mathbf{X}'\mathbf{X}\hat{\beta}_{\text{OLS}}$  é verdadeira em relação a qualquer  $\hat{\beta}_{\text{OLS}}$  escolhido.

Utilizando o fato de  $\mathbf{X}$  ser injetiva quando restrita à imagem de  $\mathbf{R}$ , temos que  $\hat{\beta}_{\text{PLS}}^m$  está bem definido e é dado por:

$$\hat{\beta}_{\text{PLS}}^m = \mathbf{R}_m \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{R}_m' \mathbf{X}' \mathbf{X} \hat{\beta}_{\text{OLS}}. \quad (6)$$

A escolha particular dos vetores  $\mathbf{r}_i$ , não afeta a estimativa de  $\hat{\beta}_{\text{PLS}}^m$ , pois se  $\mathbf{A}$  é uma transformação linear inversível,  $\mathbf{A} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ , a mudança de coordenadas

$\tilde{\mathbf{R}} = \mathbf{R}\mathbf{A}$  é tal que, substituindo na equação (6), é obtida a mesma estimativa:

$$\begin{aligned}\hat{\beta}_{\text{PLS}}^m &= \tilde{\mathbf{R}}_m \left( \tilde{\mathbf{R}}_m' \mathbf{X}' \mathbf{X} \tilde{\mathbf{R}}_m \right)^{-1} \tilde{\mathbf{R}}_m' \mathbf{X}' \mathbf{X} \hat{\beta}_{\text{OLS}} \\ &= \mathbf{R}_m \mathbf{A} \left( \mathbf{A}' \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \mathbf{A} \right)^{-1} \mathbf{A}' \mathbf{R}_m' \mathbf{X}' \mathbf{X} \hat{\beta}_{\text{OLS}} \\ &= \mathbf{R}_m \mathbf{A} (\mathbf{A}')^{-1} \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} (\mathbf{A})^{-1} \mathbf{A}' \mathbf{R}_m' \mathbf{X}' \mathbf{X} \hat{\beta}_{\text{OLS}} \\ &= \mathbf{R}_m \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{R}_m' \mathbf{X}' \mathbf{X} \hat{\beta}_{\text{OLS}}.\end{aligned}$$

Este fato reflete que apenas o subespaço gerado pelos vetores  $\mathbf{r}_i$  é importante, e não uma escolha particular da base.

É possível obter uma relação linear entre  $\hat{\beta}_{\text{PLS}}$  e  $\hat{\beta}_{\text{OLS}}$  que possui uma interpretação geométrica muito interessante. Observe que

$$\begin{aligned}& \left( \mathbf{R}_m \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{R}_m' \mathbf{X}' \mathbf{X} \right) \left( \mathbf{R}_m \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{R}_m' \mathbf{X}' \mathbf{X} \right) \\ &= \mathbf{R}_m \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right) \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{R}_m' \mathbf{X}' \mathbf{X} \\ &= \mathbf{R}_m \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{I} \mathbf{R}_m' \mathbf{X}' \mathbf{X} \\ &= \mathbf{R}_m \left( \mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{R}_m' \mathbf{X}' \mathbf{X},\end{aligned}$$

logo, essa transformação é um projetor. Como ela não é simétrica, esse projetor não é um projetor ortogonal, e sim, um projetor oblíquo, portanto, como todo projetor, a projeção se dá ao longo do kernel (Figura 27).

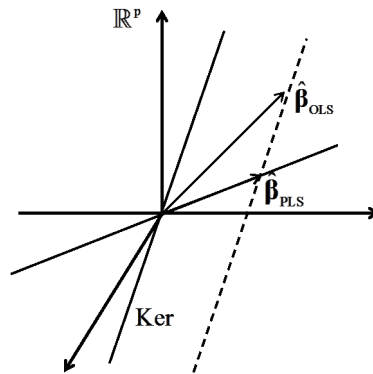


Figura 27 - Vetor  $\hat{\beta}_{\text{PLS}}$  como projeção oblíqua do vetor  $\hat{\beta}_{\text{OLS}}$ .

Como as matrizes  $\mathbf{R}$  e  $\mathbf{X}'\mathbf{X}$  são injetivas,  $\mathbf{R}_m' \mathbf{X}' \mathbf{X} \mathbf{R}_m$  é inversível e essas

transformações não têm kernel, portanto, o kernel da transformação

$$\mathbf{R}_m \left( \mathbf{R}'_m \mathbf{X}' \mathbf{X} \mathbf{R}_m \right)^{-1} \mathbf{R}'_m \mathbf{X}' \mathbf{X},$$

é exatamente o kernel de  $\mathbf{R}'_m$ . Note também que, o kernel de  $\mathbf{X}$  está contido no kernel de  $\mathbf{R}'_m \mathbf{X}' \mathbf{X}$ .

Um exemplo didático via simulação utilizando o software R (R Core Team, 2013) e uma aplicação a dados reais de seleção genômica em suínos são apresentados em Silveira (2014).

SILVEIRA, F. G.; CHAVES, L. M. Tutorial of the partial least square theory with a geometric approach. *Rev. Bras. Biom.*, São Paulo, v.32, n.1, p.59-104, 2014.

- **ABSTRACT:** *The geometric approach to the Partial Least Square (PLS) Method is natural and intuitive. It makes clear the similarities among the PLS, the Ordinary Least Square (OLS) and the Principal Components (PC) regression methods. We use the orthogonal projections to explain the step by step construction of the PLS algorithm, for sample and for population, as well as the PCR.*
- **KEYWORDS:** *Projections; random vectors; regression; principal components.*

## Referências

- AZEVEDO, C. F. et al. Regressão via componentes independentes aplicada à seleção genômica para características de carcaça em suínos. *Pesquisa Agropecuária Brasileira*, Brasília, v.48, n.6, p.619-626, jun. 2013.
- GARTHWAITE, P. H. An Interpretation of Partial Least Squares. *Journal of the American Statistical Association*, v.89, n.425, p.122-127, 1994.
- HELLAND, I. S. Partial Least Squares Regression and Statistical Models. *Scandinavian Journal of Statistics*, v.17, n.2, p.97-114, 1990.
- HOSKULDSSON, A. PLS Regression Methods. *Journal of Chemometrics*, v.2, p.211-228, 1988.
- HOTELLING, H. The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Mathematical and Statistical Psychology*, v.10, p.69-79, 1957.
- JONG, S. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, v.18, p.251-263, 1993.
- KENDALL, M. G. *A Course in Multivariate Analysis*. London: Griffin, 1957.
- OTTO, M. *Chemometrics*. Weinheim: Wiley, 1999. 328p.

PEREIRA, L. S. *Abordagem geométrica à teoria dos modelos de Gauss-Markov*. 2013. 130p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG, 2013. 328p.

PHATAK, A.; JONG, S. The geometry of partial least squares. *Journal of the Chemometrics*, v.11, p.311-338, 1997.

R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, 2013. Disponível em: <http://www.R-project.org>. Acesso em: 14 nov. 2013.

ROGGO, Y.; CHALUS, P.; MAURER, L.; LEMA-MARTINEZ, C.; EDMOND, A.; JENT, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, v.44, p.683-700, 2007.

SILVEIRA, F.G. *Abordagem geométrica do método dos quadrados mínimos parciais com uma aplicação a dados de seleção genômica*. 2014. 176 f. Tese (Doutorado em Estatística Aplicada e Biometria) - Universidade Federal de Lavras, Lavras, MG, 2014.

SINGH, I.; JUNEJA, P.; KAUR, B.; KUMAR, P. Pharmaceutical applications of chemometric techniques. *Analytical Chemistry*, v.2013, p.1-13, 2013.

WOLD, H. Path models with Latent Variables: The NIPALS Approach. In et al., H. B., editor, *Quantitative Sociology: International Perspectives on Mathematical and Statistical Model Building*, p.307-357. Academic Press, 1975.

Recebido em 12.02.2014.

Aprovado após revisão em 16.05.2014.