Optimality Properties of Nway PLS and its relation to PARAFAC

Mohamed HANAFI[1],

Samia OUERTANI[1], Julien BOCCARD[2], Gérard MAZEROLLES[3], Serge RUDAZ[2]

[1] ONIRIS, Unité de Sensométrie et de Chimiométrie, Nantes, France

[2] School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland

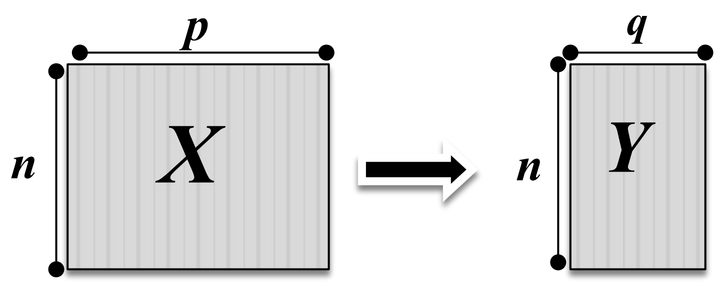
[3] 'INRA, UMR 1083, F-34060 Montpellier, France

Summary

- Introduction
- Trilinear PLS
- Motivation
- Results
- Conclusions and perspectives.

PLS regression and its variants

- When **X** is a matrix and **Y** is a vector or a matrix, a very popular method to build this regression model is Partial Least Squares Regression (PLSR) or its variants.



- Many papers have discussed this method from geometrical, mathematical and statistical point of view. Several reviews illustrating the interest of PLS regression in various applications.

Martens, H,, Naes, T., (1989) Multivariate Calibration (2nd edn), Vol. 1. Wiley: Chichester
Bhupimder, S., Dayal and McGregor, J.F., (1997). Improved PLS algorithms. Journal of Chemometrics, 11, 73-85.
[Phatak, A., de Jong, S., (1997). The geometry of partial-least squares. Journal of Chemometrics, 11, 311–338.
Helland, I. (1988). On the structure of partial-least squares regression. Commun. Stat.—Simul. 17, 581–607.

Compact form of PLS regression

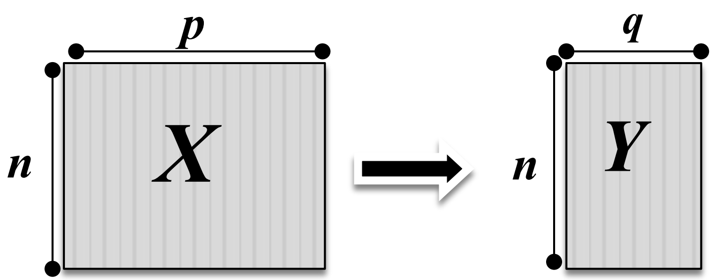
$$\underset{\|w\|=1,\mathbf{t}_{\mathbf{X}}^{(h)}=\mathbf{X}^{(h)}\mathbf{w}}{Max}\sum_{j=1}^{h}\operatorname{cov}^2\left(\mathbf{t}_{\mathbf{X}}^{(h)},\mathbf{y}_j^{(h)}\right)$$

$$\begin{cases}\mathbf{Y}^{(1)} = \mathbf{Y} \\ \mathbf{Y}^{(h+1)} = \mathbf{Y}^{(h)} - \dfrac{\mathbf{t}_{\mathbf{X}}^{(h)}\mathbf{t}_{\mathbf{X}}^{(h)T}}{\left\|\mathbf{t}_{\mathbf{X}}^{(h)}\right\|^2}\mathbf{Y}^{(h)}\end{cases}$$

$$\begin{cases}\mathbf{X}^{(1)} = \mathbf{X} \\ \mathbf{X}^{(h+1)} = \mathbf{X}^{(h)} - \dfrac{\mathbf{t}_{\mathbf{X}}^{(h)}\circ\mathbf{t}_{\mathbf{X}}^{(h)T}}{\left\|\mathbf{t}_{\mathbf{X}}^{(h)}\right\|^2}\mathbf{X}^{(h)}\end{cases}$$

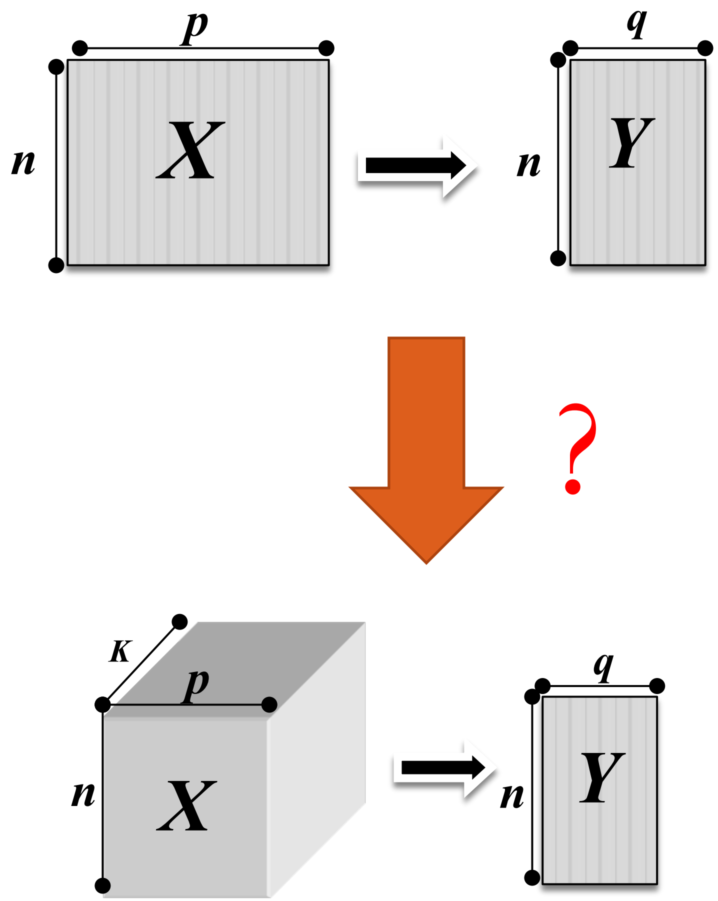
$$\mathbf{X} = \sum_{l=1}^{H}\mathbf{t}_{\mathbf{X}}^{(l)}\,\mathbf{a}_{\mathbf{X}}^{(l)T} + \mathbf{R}_{\mathbf{X}}^{(H)}$$

$$\mathbf{Y} = \sum_{l=1}^{H}\mathbf{t}_{\mathbf{X}}^{(l)}\,\mathbf{b}_{\mathbf{X}}^{(l)T} + \mathbf{R}_{\mathbf{Y}}^{(H)} \Rightarrow \mathbf{Y} = \mathbf{X}\mathbf{B}^{(H)} + \mathbf{R}_{\mathbf{Y}}^{(H)}$$



Martens, H,, Naes, T., (1989) Multivariate Calibration (2nd edn), Vol. 1. Wiley: Chichester
Bhupimder, S., Dayal and McGregor, J.F., (1997). Improved PLS algorithms. Journal of Chemometrics, 11, 73-85.
[Phatak, A., de Jong, S., (1997). The geometry of partial-least squares. Journal of Chemometrics, 11, 311–338.
Helland, I. (1988). On the structure of partial-least squares regression. Commun. Stat.—Simul. 17, 581–607.

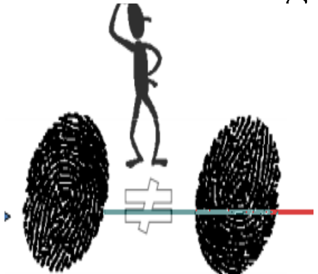
Data Analysis Problem statement : N way regression



Targest application

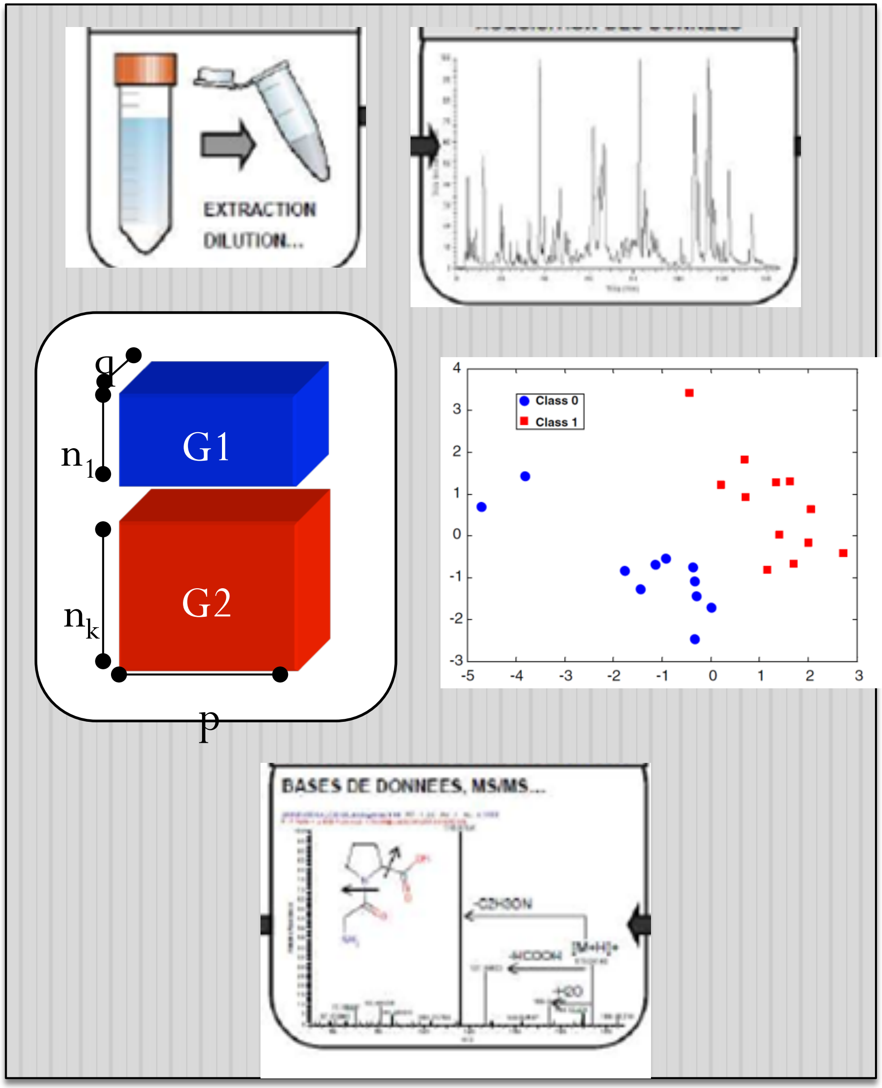
Metabomic approach

- Overall characterization of fluids and / or food in order to characterize the biological impact of chemical contaminants on the scale of the body, or the body of the cell, ultimately with the aim of highlighting the biological exposure biomarkers signing.



Boccard, J., Veuthey, J.-L, Rudaz, S. (2010).. J. Sep. Sci. 2010, 33, 1–15.
Boccard J., Badoud, F., Grata, E., Ouertani, S., Hanafi, M., Mazerolles, G., Lantéri, P., Veuthey, J.-L., Rudaz, S., (2012. Spectra Analyse, 284, 48-54.
Dyrby, M., Petersen, M., Whittaker, AK., Lambert, L., Norgaard, L., Bro, R., Engelsen, SB. (2005)..Anal. Chim. Acta, 531, 209–216.
Rubingh, CM., Bijlsma, S., Jellema, RH., Overkamp, KM., van der Werf, MJ., Smilde, AK. (2009). J. Proteome Res., 8, 4319–4327.

Metabolomic analysis

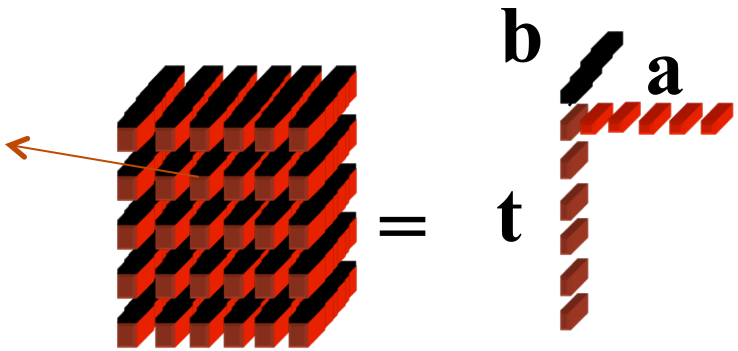


Notations and definitions (1)

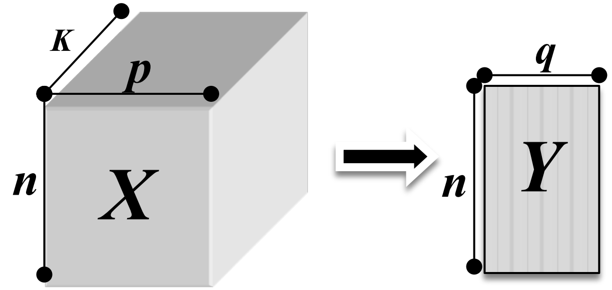
- **Outer product of vectors.** The outer product of the vectors **t,a and b** is a three 3 way array **P** . The elements of **P** are expressed as $P_{ijk} = t_i a_j b_k$

- **Three way data with rank one.** A three way data is of rank 1 if and only if it can be written as the outer product of 3 vectors. $\mathbf{P} = \mathbf{t} \circ \mathbf{a} \circ \mathbf{b} = \lfloor t_i a_j b_k \rfloor$

$$P_{ijk} = t_i a_j b_k$$



Trilinear PLS



- Introduced by Bro [1996] Bro, R. (1996). Journal of Chemometrics, 10, 47–61.
- Few studies around Nway PLS (Smilde and al. [6]). Smilde, A.K. (1997). Journal of Chemometrics, 11, 367–377.

- Trilinear PLS Model

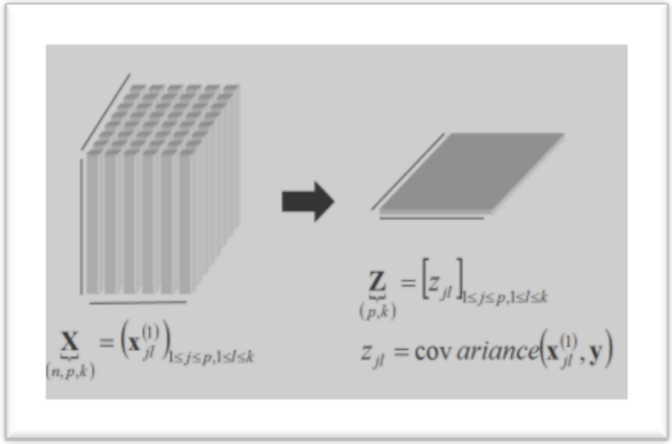
$$\mathbf{X} = \sum_{l=1}^{H} \mathbf{t}_{\mathbf{X}}^{(l)} \circ \mathbf{a}_{\mathbf{X}}^{(l)} \circ \mathbf{b}_{\mathbf{X}}^{(l)} + \mathbf{R}_{\mathbf{X}}^{(H)}$$

$$\mathbf{Y} = \mathbf{T}_{\mathbf{X}}^{(H)} \mathbf{B}^{(H)} + \mathbf{R}_{\mathbf{Y}}^{(H)}$$

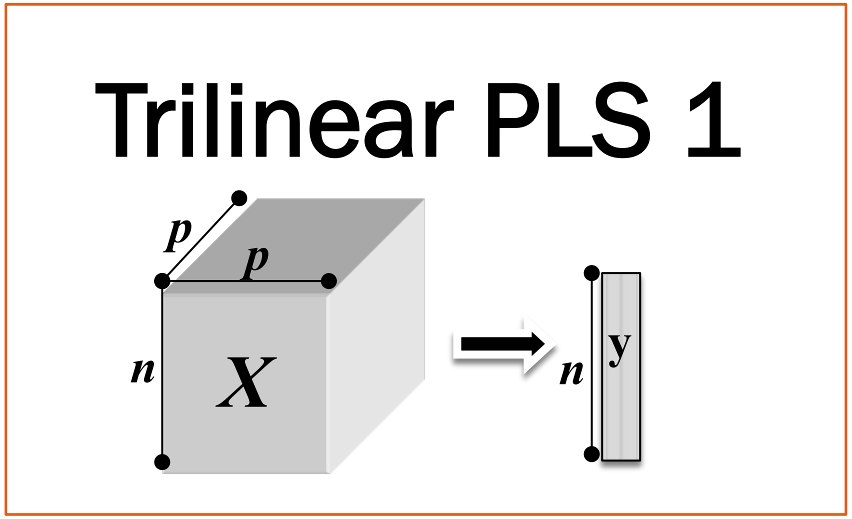
- Trilinear PLS algorithm(sequential)
  - Stage 1. Computation of Scores and loadings
  - Stage 2. Deflation

Stage 1. Computation of scores and loadings

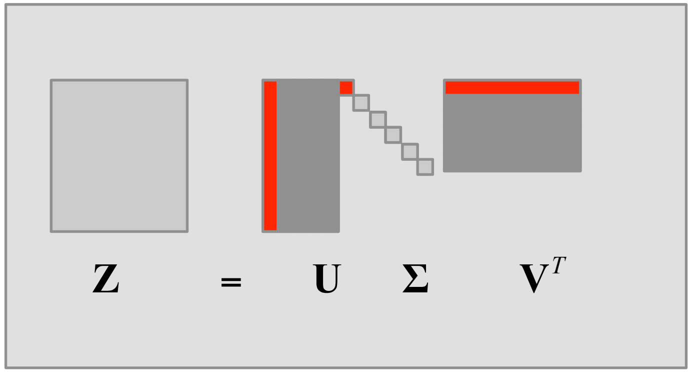
1



Covariance between X and y

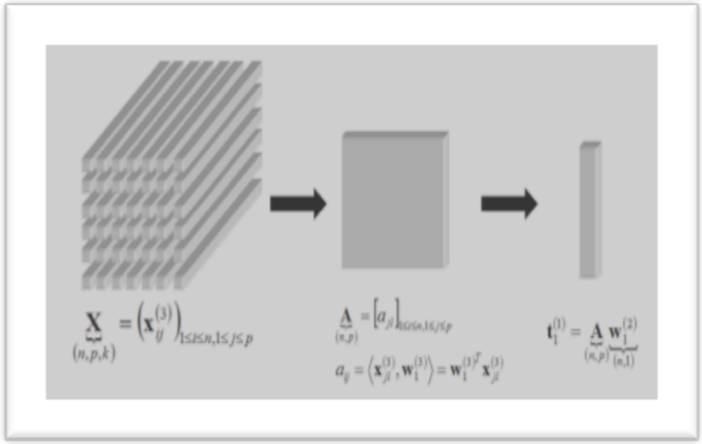


2



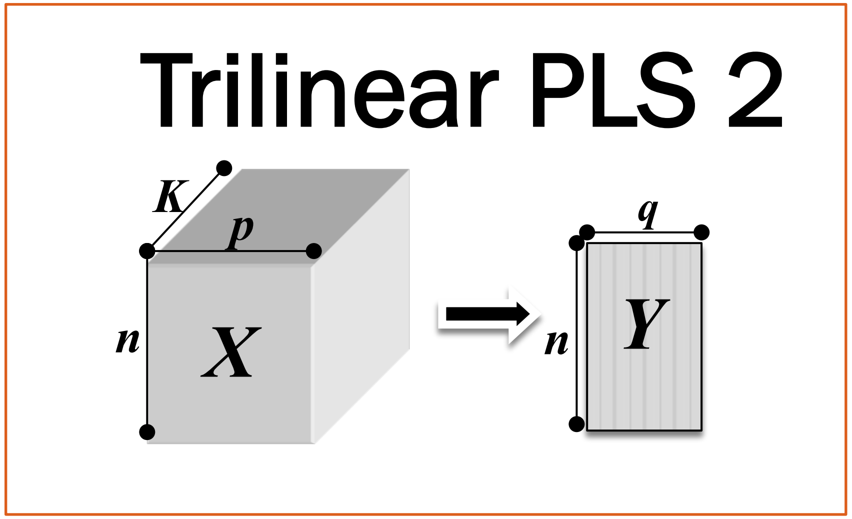
Singular Value Decomposition

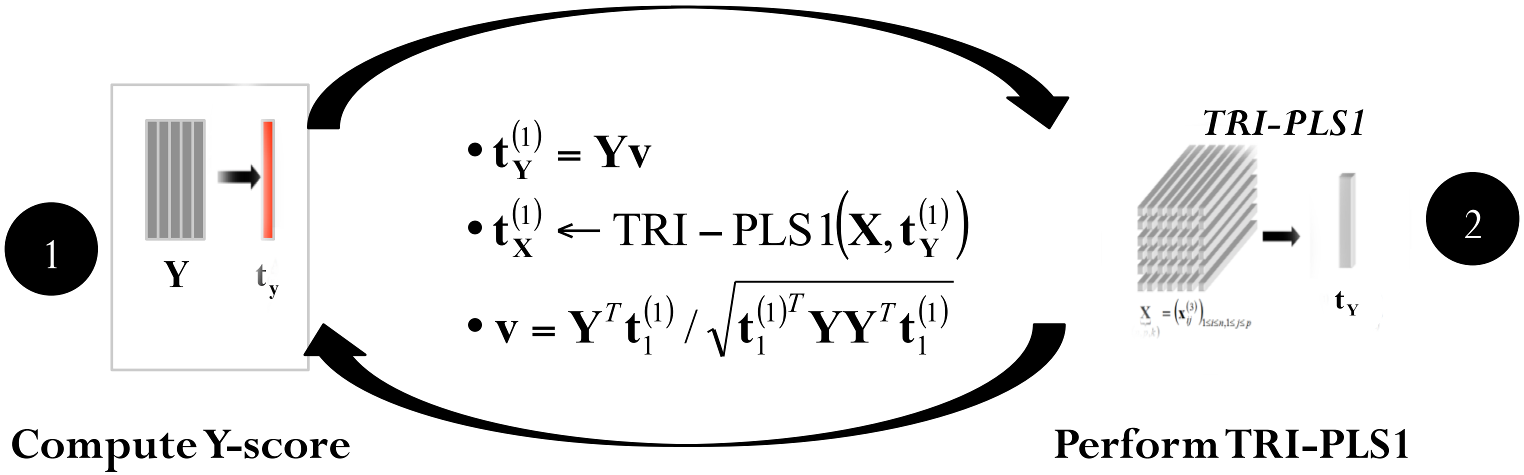
3



Compute X score

Stage 1. Computation of scores and loadings.





Trilinear PLS Algorithm : deflation stage

$$\begin{cases} \mathbf{X}^{(1)} = \mathbf{X} \\ \mathbf{X}^{(h+1)} = \mathbf{X}^{(h)} - \mathbf{t}_{\mathbf{X}}^{(h)} \circ \mathbf{a}_{\mathbf{X}}^{(h)} \circ \mathbf{b}_{\mathbf{X}}^{(h)} \end{cases}$$

$$\begin{cases} \mathbf{Y}^{(1)} = \mathbf{Y} \\ \mathbf{Y}^{(h+1)} = \mathbf{Y} - \mathbf{T}_{\mathbf{X}}^{(h)} \mathbf{B}^{(h+1)} \end{cases}$$

- Using MLR for deflating Y
- X-Scores are not Orthogonal.
- X two way data ➔ PLS2 (deflates on loadings) or Tucker Analysis (1958)

Motivation

	Convergence Issues		Optimality issues	
	Scores and loadings	Deflation	Scores and loadings)	Deflation
Trilinear PLS1	**No problem (not iterative)**	**No problem (not iterative)**	**Known**	**Known**
Trilinear PLS2	**Unknown**	**No problem (not iterative)**	**Unknown**	**Known**

- Main Lack of knowledge :
  - Convergence of the procedure (felt but not rigorously demonstrated).
  - no apparent optimization is provided to characterize the parameters (scores and loadings) of the method
  - may be trilinear PLS 2 is equivalent to existing methods

How to do ?

- Sequences generated by Tri-PLS2

TRI-PLS2

- $\mathbf{t}_{\mathbf{Y}}^{(1)} = \mathbf{Y}\mathbf{v}$
- $\mathbf{t}_{\mathbf{X}}^{(1)} \leftarrow \mathrm{TRI-PLS1}\left(\mathbf{X}, \mathbf{t}_{\mathbf{Y}}^{(1)}\right)$
- $\mathbf{v} = \mathbf{Y}^T \mathbf{t}_1^{(1)} / \sqrt{\quad)^T \mathbf{Y}\mathbf{Y}^T \mathbf{t}_1^{(1)}}$

$$\left(\mathbf{t}_{\mathbf{X}}^{(l)}, \mathbf{t}_{\mathbf{Y}}^{(l)}\right)_{l \geq 1}$$

- Scores

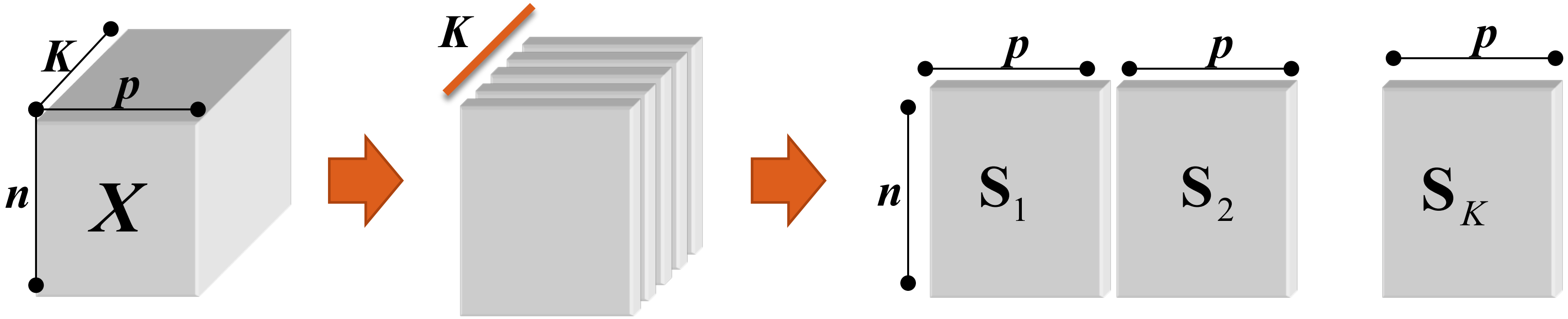
$$\left(\mathbf{a}_{\mathbf{Y}}^{(l)}, \mathbf{a}_{\mathbf{X}}^{(l)}; \mathbf{b}_{\mathbf{X}}^{(l)}\right)_{l \geq 0}$$

- Loadings

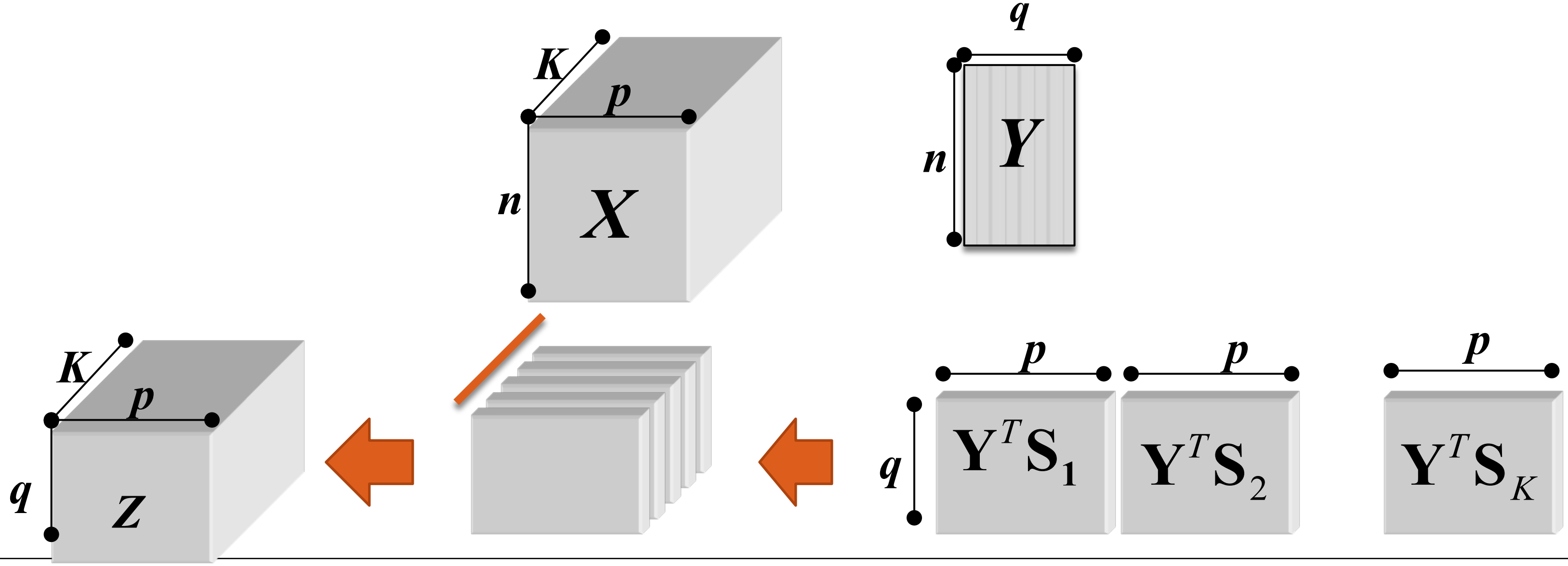
- Study TRI-PLS2 ←→ study the various sequences generated by the procedure

Notations and definitions (2)

- From three way data two multiblock (Sk)



- Covariance tensor between X and Y



Results : monotony properties

$$\mathrm{cov}^2\left(\mathbf{t}_{\mathbf{X}}^{(l)},\mathbf{t}_{\mathbf{Y}}^{(l)}\right)\le \mathrm{cov}^2\left(\mathbf{t}_{\mathbf{X}}^{(l+1)},\mathbf{t}_{\mathbf{Y}}^{(l+1)}\right)$$

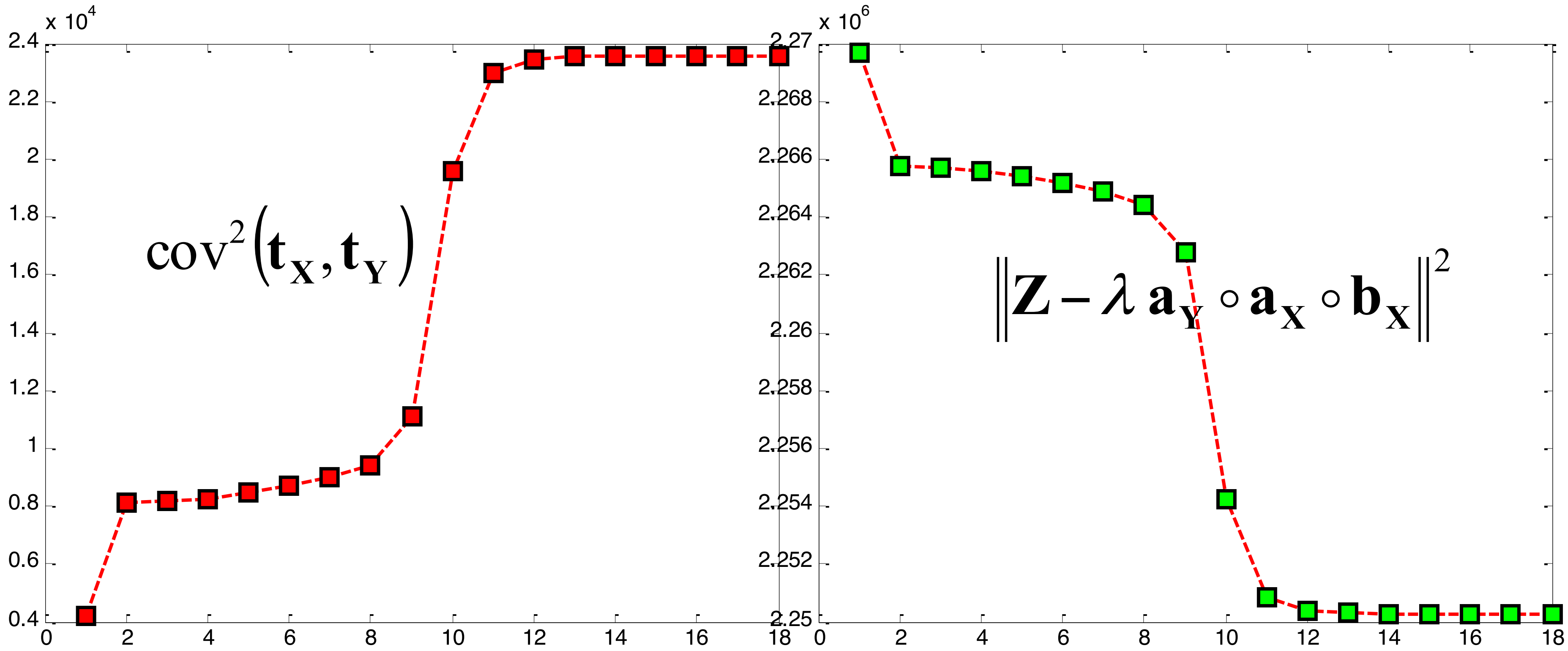
$$\left\|\mathbf{Z}-\underbrace{\lambda^{(l)}}_{\mathrm{cov}\left(\mathbf{t}_{\mathbf{X}}^{(l)},\mathbf{t}_{\mathbf{Y}}^{(l)}\right)}\mathbf{a}_{\mathbf{Y}}^{(l)}\circ\mathbf{a}_{\mathbf{X}}^{(l)}\circ\mathbf{b}_{\mathbf{X}}^{(l)}\right\|^2\ge\left\|\mathbf{Z}-\lambda^{(l+1)}\mathbf{a}_{\mathbf{Y}}^{(l+1)}\circ\mathbf{a}_{\mathbf{X}}^{(l+1)}\circ\mathbf{b}_{\mathbf{X}}^{(l+1)}\right\|^2$$

$$\lambda^{(l)}=\mathrm{cov}\left(\mathbf{t}_{\mathbf{X}}^{(l)},\mathbf{t}_{\mathbf{Y}}^{(l)}\right)$$

$$\sum_{k=1}^{K}\mathrm{cov}^2\left(\mathbf{t}_{\mathbf{S}_k}^{(l)},\mathbf{t}_{\mathbf{Y}}^{(l)}\right)\le\sum_{k=1}^{K}\mathrm{cov}^2\left(\mathbf{t}_{\mathbf{S}_k}^{(l+1)},\mathbf{t}_{\mathbf{Y}}^{(l+1)}\right)$$

$$\mathbf{t}_{\mathbf{S}_k}^{(l)}=\mathbf{S}_k\mathbf{a}_{\mathbf{X}}^{(l)}\left(k=1,2,..,K\right)$$

Monotony properties



Monotony properties of Trilinear PLS 2

- Show the monotony convergence of trilinear PLS2
- Assess the best solutions to be choosen when several starting vectors are used
- the various criteria can be used as stopping criterien (Implementation)
- to assess the correcteness of the procedure (Implementation)

Optimality Results for Trilinear PLS 2 (1)

Scores are solutions the following optimization problem :

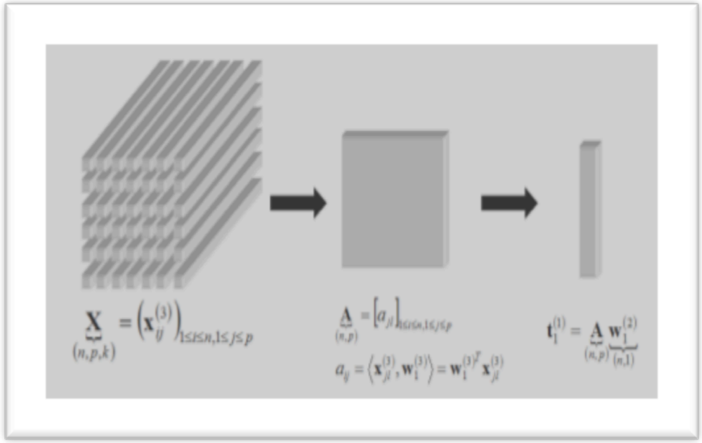
$$Maximize\ \mathrm{cov}^2\left(\mathbf{t_X}, \mathbf{t_Y}\right)$$

under the constraints

$$\underbrace{\mathbf{t_X}}_{(n,1)} = \underbrace{\mathbf{X}}_{(n,p,K)} \times_2 \underbrace{\mathbf{a_X}}_{(1,p)} \times_3 \underbrace{\mathbf{b_X}}_{(K,1)}$$

$$\underbrace{\mathbf{t_Y}}_{(n,1)} = \underbrace{\mathbf{Y}}_{(n,q)} \underbrace{\mathbf{a_Y}}_{(1,q)}$$

$$\left\|\mathbf{a_X}\right\| = \left\|\mathbf{b_X}\right\| = \left\|\mathbf{a_Y}\right\| = 1$$



$$\underbrace{\mathbf{t_X}}_{(n,1)} = \underbrace{\mathbf{X}}_{(n,n,m)} \times_2 \underbrace{\mathbf{a_X}}_{(1,p)} \times_3 \underbrace{\mathbf{b_X}}_{(,1,m)}$$

Optimality Results for Trilinear PLS 2 (2)

Loadings are solutions the following optimization problem :

$$\textit{Minimize} \quad \left\| \mathbf{Z} - \lambda\, \mathbf{a}_{\mathbf{Y}} \circ \mathbf{a}_{\mathbf{X}} \circ \mathbf{b}_{\mathbf{X}} \right\|^2$$

under the constraints

$$\left\|\mathbf{a}_{\mathbf{X}}\right\| = \left\|\mathbf{b}_{\mathbf{X}}\right\| = \left\|\mathbf{a}_{\mathbf{Y}}\right\| = 1$$

$$\lambda \in R$$

Trilinear PLS 2 = PARAFAC applied to covariance tensor between X and Y

All procedure which estimates parafac can be used to estimate loadings of Trilinear PLS2

Optimality Results for Trilinear PLS 2 (2)

Trilinear PLS 2

$$\textit{Maximize} \quad \sum_{k=1}^{K} \operatorname{cov}^2\left(\mathbf{S}_k \mathbf{a}_{\mathbf{X}}, \mathbf{Y}\mathbf{a}_{\mathbf{Y}}\right)$$

under the constraints

$$\left\|\mathbf{a}_{\mathbf{X}}\right\| = \left\|\mathbf{a}_{\mathbf{Y}}\right\| = 1$$

MB PLS

$$\textit{Maximize} \quad \sum_{k=1}^{K} \operatorname{cov}^2\left(\mathbf{S}_k \mathbf{a}_k, \mathbf{Y}\mathbf{a}_{\mathbf{Y}}\right)$$

under the constraints

$$\left\|\mathbf{a}_{\mathbf{Y}}\right\| = \left\|\mathbf{a}_k\right\| = 1 \quad \left(k = 1,2,\ldots,K\right)$$

Trilinear PLS 2 can be seen as a constrainted (equality of block loadings) of the well known Multiblock PLS regression

Conclusions and perspectives

- Monotony convergence of trilinear PLS2
- Optimality properties for parameters of Trilinear PLS2
- Connexion of Trilinear PLS2 with PARAFAC and MBPLS
- Application to discrimination

- Perspective 1 :
  - Focus on the deflation stage (scores are not orthogonal)
- Perspective 2 : Explore other tensor decompositions like Tucker 3