# Mathematical properties of Consensus PCA

**J.J. Jansen, H.F.M. Boelens, A.K. Smilde**
Process Analysis and Chemometrics, Department of Chemical Engineering,
University of Amsterdam, The Netherlands
*website:* http://www-its.chem.uva.nl/ ✉ jjansen@science.uva.nl

## Introduction

A multiblock extension for PCA has been developed by Wold et al.[1] called Consensus PCA (CPCA). Westerhuis et al.[2] modified this method. Westerhuis' CPCA version is on the super level equivalent to ordinary PCA[2,3].
CPCA generates several models of the data in each block. The properties of these models in terms of orthogonality and splitting of variance are determined. Knowledge of these properties gives insight into the interpretation of the obtained block scores and loadings.

## Algorithm

The algorithm developed by Westerhuis et al. is given by:

(1) $\mathbf{X}^{(1)} = \mathbf{T}_{sup}\mathbf{P}_{sup}^{T} + \mathbf{X}^{(R+1)}$

(2) $\mathbf{T}_{sup} = \left[\mathbf{t}_{sup,1} \ \cdots \ \mathbf{t}_{sup,R}\right]; \mathbf{P}_{sup}^{T} = \left[\mathbf{P}_{sup,1}^{T} \ \cdots \ \mathbf{P}_{sup,B}^{T}\right];$
$\mathbf{P}_{sup,b} = \left[\mathbf{p}_{sup,b1} \ \cdots \ \mathbf{p}_{sup,bR}\right]$

(3) for every dimension $r$, for $r$ = 1...$R$

a) $\mathbf{p}_{br} = \dfrac{\mathbf{p}_{sup,br}}{\left\|\mathbf{p}_{sup,br}\right\|}$

b) $\mathbf{t}_{br} = \mathbf{X}_b^{(r)}\mathbf{p}_{br}$

c) $\mathbf{T}_r = \left[\mathbf{t}_{1r} \ \cdots \ \mathbf{t}_{Br}\right]$

d) $\mathbf{w}_r = \dfrac{\mathbf{T}_r^{T}\mathbf{t}_{sup,r}}{\left\|\mathbf{T}_r^{T}\mathbf{t}_{sup,r}\right\|}$

e) Deflate: $\mathbf{X}_b^{(r+1)} = \mathbf{X}_b^{(r)} - \mathbf{t}_{sup,r}\mathbf{p}_{sup,br}^{T}$

end

## Models

### Model A1: Regular PCA

$\mathbf{X}^{(1)} = \mathbf{T}_{sup}\mathbf{P}_{sup}^{T} + \mathbf{X}^{(R+1)}$

| Symbol | Dimensions | | Symbol | |
|---|---|---|---|---|
| $\mathbf{X}^{(1)}$ | $I \times \sum_{b=1}^{B} J_b$ | Raw Data Matrix | $1 \ldots b \ldots B$ | Block Index |
| $\mathbf{T}_{sup}$ | $I \times R$ | Super Scores | $1 \ldots j_b \ldots J_b$ | Variable index |
| $\mathbf{P}_{sup}$ | $\sum_{b=1}^{B} J_b \times R$ | Super Loadings | $I$ | Number of samples |
| $\mathbf{X}^{(R+1)}$ | $I \times \sum_{b=1}^{B} J_b$ | Residuals of the $R$-component **Model A1** | $1 \ldots r \ldots R$ | PC index |

*Properties of* **Model A1**
Variance can be split in PCA. Orthogonality properties in PCA are already well known.

$$\left\|\mathbf{X}^{(1)}\right\|^2 = \left\|\mathbf{t}_{sup,1}\mathbf{p}_{sup,1}^{T}\right\|^2 + \ldots + \left\|\mathbf{t}_{sup,R}\mathbf{p}_{sup,R}^{T}\right\|^2 + \left\|\mathbf{X}^{(R+1)}\right\|^2$$

### Model A2: Model on super level for each block

$\mathbf{X}_b^{(1)} = \mathbf{T}_{sup}\mathbf{P}_{sup,b}^{T} + \mathbf{X}_b^{(R+1)}$

| Symbol | Dimensions | |
|---|---|---|
| $\mathbf{X}_b^{(1)}$ | $I \times J_b$ | Raw data matrix for block $b$ |
| $\mathbf{P}_{sup,b}$ | $J_b \times R$ | Super Loadings for block $b$ |
| $\mathbf{X}_b^{(R+1)}$ | $I \times J_b$ | Residuals of the $R$-component **Model A2** |

*Properties of* **Model A2**

$\mathbf{P}_{sup,b}^{T} = \left(\mathbf{T}_{sup}^{T}\mathbf{T}_{sup}\right)\mathbf{T}_{sup}^{T}\mathbf{X}_b^{(1)}$ ; $\mathbf{P}_{sup,b}^{T}\mathbf{P}_{sup,b} \neq 0$ ;
$\mathbf{X}_b^{(R+1)}\mathbf{P}_{sup,b} \neq 0$ ; $\mathbf{T}_{sup}^{T}\mathbf{X}_b^{(R+1)} = 0$

Variance can be split into the different PCs in **Model A2**, due to the fact that $\mathbf{T}_{sup}^{T}\mathbf{T}_{sup} = diag$ and the orthogonality of $\mathbf{T}_{sup}$ with the residuals.

$$\left\|\mathbf{X}_b^{(1)}\right\|^2 = \left\|\mathbf{t}_{sup,1}\mathbf{p}_{sup,b1}^{T}\right\|^2 + \ldots + \left\|\mathbf{t}_{sup,R}\mathbf{p}_{sup,bR}^{T}\right\|^2 + \left\|\mathbf{X}_b^{(R+1)}\right\|^2$$

### Model B: Model on block level for each block

$\mathbf{X}_b^{(1)} = \mathbf{T}_b\mathbf{P}_b^{T} + \mathbf{Z}_b^{(R+1)}$

| Symbol | Dimensions | |
|---|---|---|
| $\mathbf{T}_b$ | $I \times R$ | Block Scores |
| $\mathbf{P}_b$ | $J_b \times R$ | Block Loadings |
| $\mathbf{Z}_b^{(R+1)}$ | $I \times J_b$ | Residuals of the $R$-component **Model B** |

*Properties of* **Model B**

$\mathbf{t}_{br} = \mathbf{X}_b^{(r)}\mathbf{p}_{br}$ ; $\mathbf{T}_b^{T}\mathbf{T}_b \neq diag$ ; $\mathbf{P}_b^{T}\mathbf{P}_b \neq diag$ ; $\mathbf{T}_b^{T}\mathbf{Z}_b^{(R+1)} \neq 0$ ; $\mathbf{Z}_b^{(R+1)}\mathbf{P}_b \neq 0$

Variance can not be split in this model, due to the non-orthogonality of $\mathbf{T}_b$ and $\mathbf{P}_b$ and their non-orthogonality with the residuals. To be able to split variance on the block level, **Model C** is required.

### Model C: Updated model on block level for each block

When the block data $\mathbf{X}_b^{(1)}$ is projected on the block scores $\mathbf{T}_b$, updated block loadings are obtained. Calculation of $\tilde{\mathbf{P}}_b$ can only take place for all PCs simultaneously, due to the fact that $\mathbf{T}_b$ is not orthogonal.

$\mathbf{X}_b^{(1)} = \mathbf{T}_b\tilde{\mathbf{P}}_b^{T} + \tilde{\mathbf{Z}}_b^{(R+1)}$

| Symbol | Dimensions | |
|---|---|---|
| $\tilde{\mathbf{P}}_b$ | $J_b \times R$ | Updated Block Loadings |
| $\tilde{\mathbf{Z}}_b^{(R+1)}$ | $I \times J_b$ | Residuals of the $R$-component **Model C** |

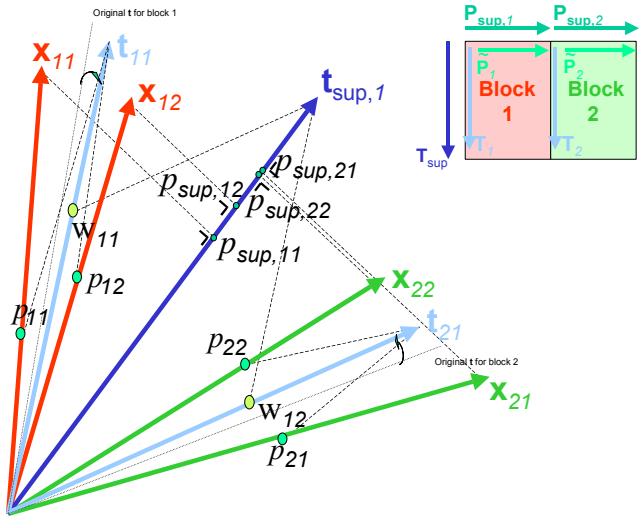
*Properties of* **Model C**

$\tilde{\mathbf{P}}_b^{T} = \left(\mathbf{T}_b^{T}\mathbf{T}_b\right)^{-1}\mathbf{T}_b^{T}\mathbf{X}_b^{(1)}$ ; $\mathbf{T}_b^{T}\tilde{\mathbf{Z}}_b^{(R+1)} = 0$

Variance can be split into explained and unexplained variance in **Model C**, due to the orthogonality of $\mathbf{T}_b$ with the residuals. It is not possible to split variance per component.

$$\left\|\mathbf{X}_b^{(1)}\right\|^2 = \left\|\mathbf{T}_b\tilde{\mathbf{P}}_b^{T}\right\|^2 + \left\|\tilde{\mathbf{Z}}_b^{(R+1)}\right\|^2$$

### Geometric Description




[1] S. Wold, S. Hellberg, T. Lundstedt, M. Sjostrom, H. Wold, *Proc. Symp. on PLS Model Building: Theory and Aplication*, Frankfurt am Main 1987

[2] Westerhuis JA, Kourti T and MacGregor JF, Analysis of multiblock and hierarchical PCA and PLS models, *Journal of Chemometrics*, **12**, (1998), 301-321.

[3] S.J. Qin, S. Valle, M.J. Piovoso, Unifying Multiblock Analysis, *Journal of Chemometrics*, **15**, 715-742