Chimiométrie XIX - 2019

ComDim-ICA
Multiblock Independent Components Analysis

D. N. Rutledge
UMR GENIAL, INRA, AgroParisTech, Université Paris-Saclay Paris/ France
rutledge@agroparistech.fr
L. Schmidtke
National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga / Australia
LSchmidtke@csu.edu.au

ComDim

- ComDim or Common Components and Specific Weights Analysis (CCSWA) is an exploratory multi-block data analysis method
- Simultaneous analysis of several data tables with different variables describing the same samples

- Determines a common space for all blocks
- Each block has a specific contribution (*salience*) to the definition of each dimension of the common space

- Originally developed in sensometrics
- Has been applied to the fusion of analytical data

E.M. Qannari, I. Wakeling, P. Courcoux, H.J.H. Macfie,
Defining the underlying sensory dimensions, Food Quality and Preference, 11 (2000) 151-154

V. Cariou, D. Jouan-Rimbaud Bouveresse E.M. Qannari, D.N. Rutledge
"ComDim methods for the analysis of multiblock data in a data fusion perspective" in
Data Fusion Methodology and Applications (Data Handling in Science and Technology), (ed. Marina Cocchi)
Elsevier Science Publishers, Amsterdam, 2018

Original ComDim algorithm

Start with *p* matrices $\mathbf{X_i}$ of size $n \times k_i$ (i = 1 to *p*)

Each $\mathbf{X_i}$ column-centered and scaled by dividing by matrix norm :

$$\mathbf{Xs_i}$$

For each $\mathbf{Xs_i}$, an $n \times n$ scalar product matrix $\mathbf{W_i}$ can be computed as :

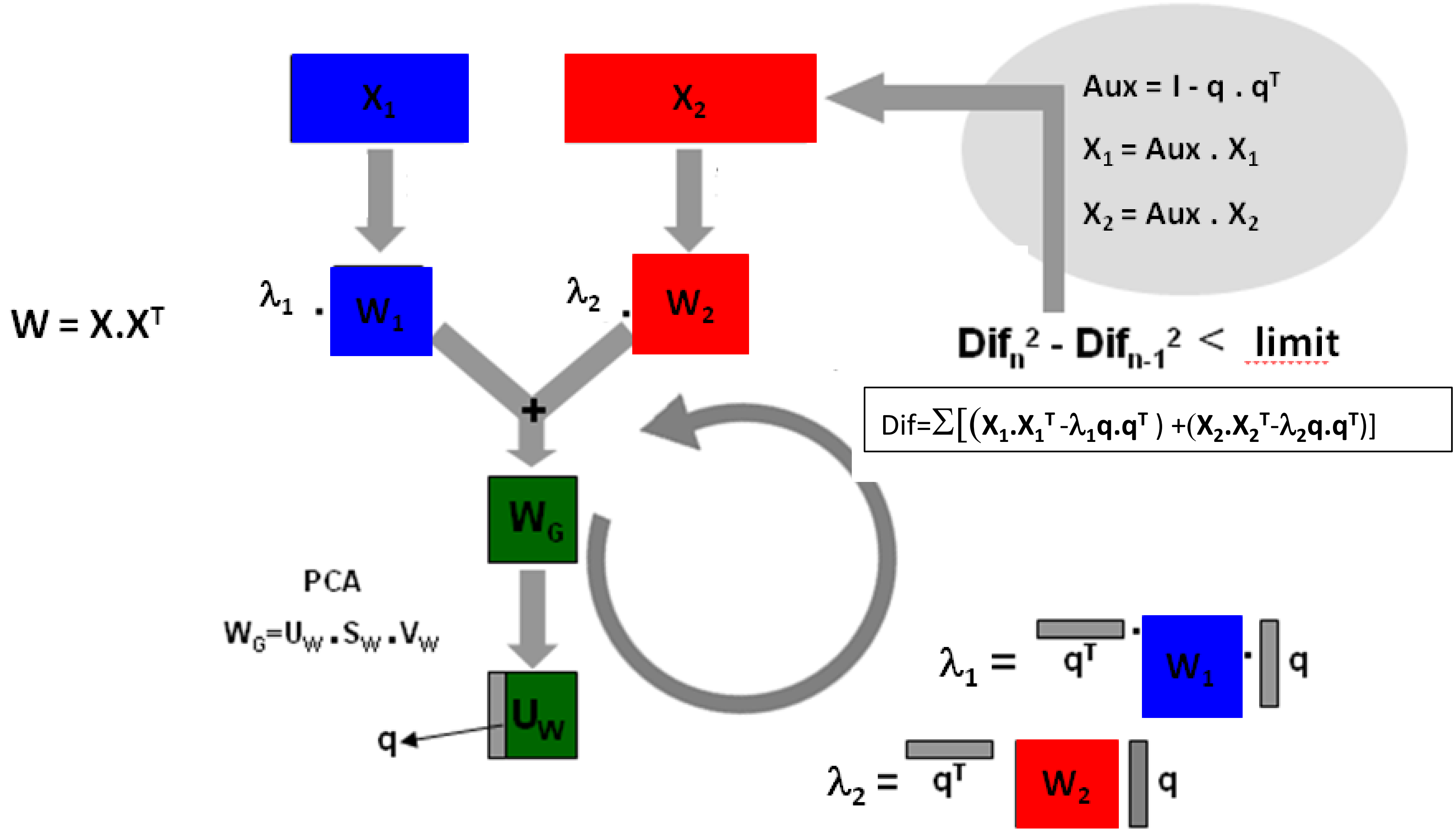
$$\mathbf{W_i} = \mathbf{Xs_i} \bullet \mathbf{Xs_i}^{T}$$

$\mathbf{W_i}$ reflect the dispersion of the ***samples*** in the space of that table

Each $\mathbf{W_i}$ is multiplied by a scalar, λ_i (initially all set to 1)

At each iteration, a sum of the *p* weighted $\mathbf{W_i}$ matrices is computed, resulting in a global $\mathbf{W_G}$ matrix

Original ComDim algorithm



Original ComDim algorithm

$$\mathbf{W}_k = \sum_{dim=1}^{r} \lambda_{dim}^{(k)} \mathbf{q}_{dim} \mathbf{q}'_{dim} \quad + \quad \mathbf{E}_k$$

- Sequential determination of :
  - Global scores of individuals on each CC: $\mathbf{q}_{dim}$
  - Saliences of tables : λ^{k}_{dim}
  - Loadings of variables : $\mathbf{u}^{k}_{dim}$
  - Local scores of individuals for each table : $\mathbf{t}^{k}_{dim}$
  - Sum of saliences of all tables for each CC
  - Sum of saliences of all CCs for each table
  - Variance extracted by each CC
  - …

New ComDim algorithm
(*proposed by M. Hanafi*)

Start with p matrices $\mathbf{X_i}$ of size $n \times k_i$ (i = 1 to p)

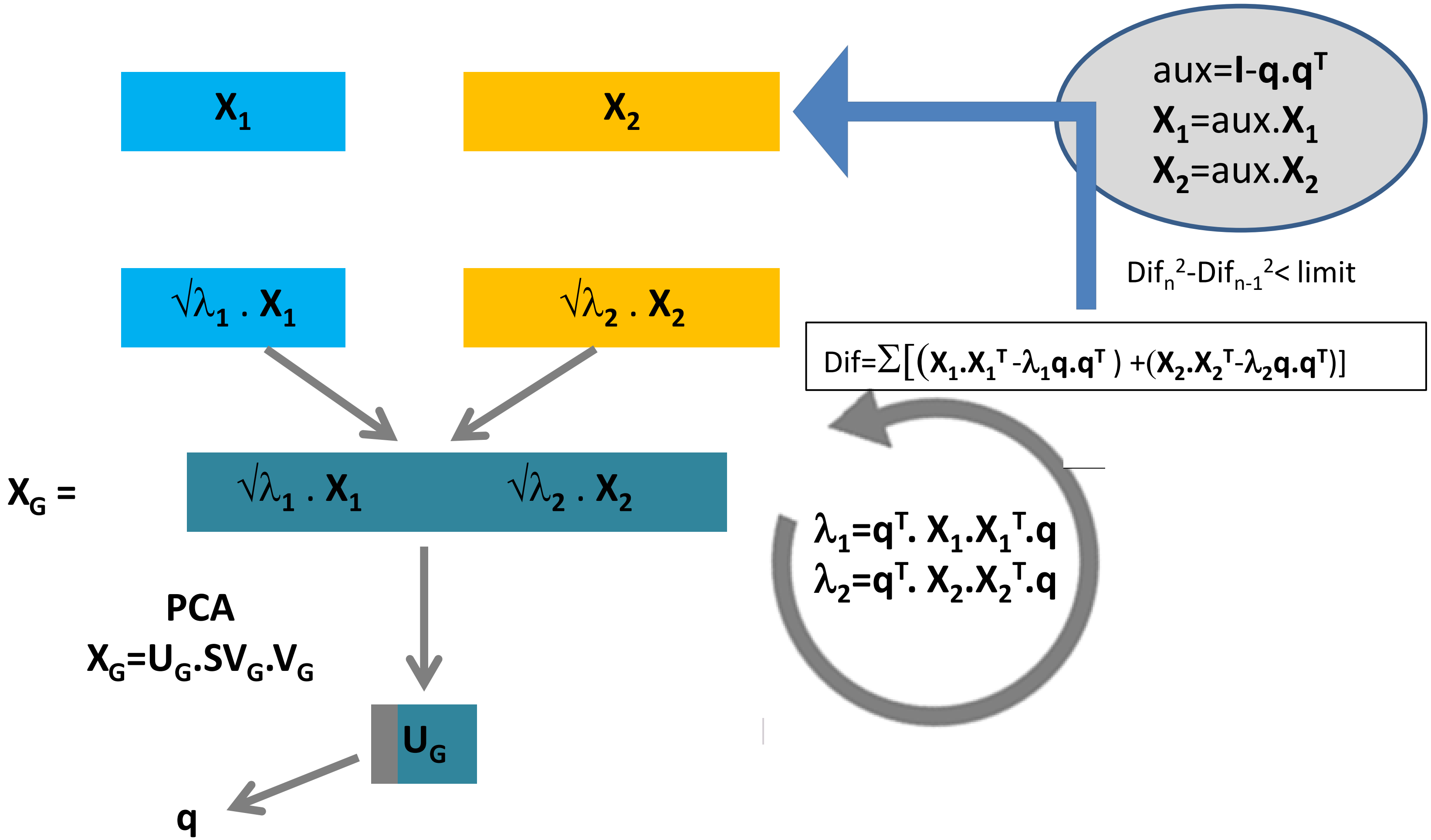
Each $\mathbf{X_i}$ column-centered and scaled by dividing by matrix norm to give $\mathbf{Xs_i}$

Each $\mathbf{Xs_i}$ is multiplied by a scalar, $\sqrt{\lambda_i}$ (initially all set to 1)

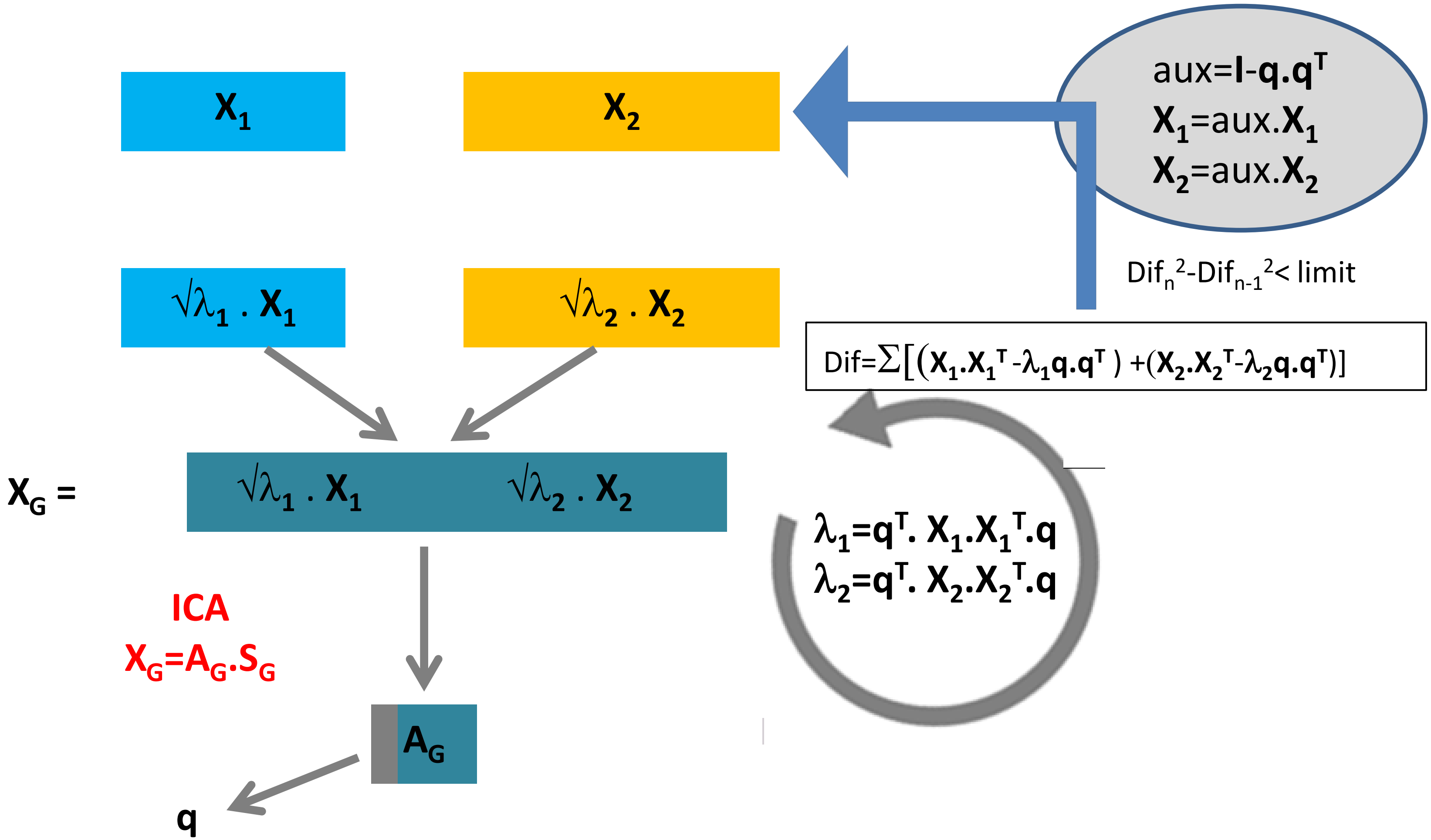
At each iteration, the p weighted $\mathbf{Xs_i}$ matrices are concatenated column-wise , resulting in a global $\mathbf{X_G}$ matrix

$$\mathbf{X_G} = [\sqrt{\lambda_1}\mathbf{Xs_1}\ ,\ \sqrt{\lambda_2}\mathbf{Xs_2}\ ,\ \sqrt{\lambda_3}\mathbf{Xs_3}\ ,\ \ldots\ ,\ \sqrt{\lambda_p}\mathbf{Xs_p}\ ]$$

New ComDim algorithm



Multi-Block ICA !



Independent Components Analysis

Aims to extract the unknown source signals mixed together in unknown proportions in the observed signals that form the rows of the data matrix.

ICs or Source Signals : analogous to PCA Loadings

Proportions : analogous to PCA Scores

D. Jouan-Rimbaud Bouveresse, D.N.Rutledge
"Independent Components Analysis: Theory And Applications" *in*
Resolving Spectral Mixtures, (ed. C. Ruckebusch)
Elsevier Science Publishers, Amsterdam, 2017, pp. 225-278

Independent Components Analysis (ICA)

Data matrix $\longrightarrow$ **a set of observed signals**, where :

- each observed sensor signal, $\mathbf{x_i}$, is the weighted sum of pure source signals, $\mathbf{s_j}$
- the weighting coefficients, a_{ij}, are proportions of the source signals, $\mathbf{s_j}$

$$\mathbf{x_1 = a_{11} * s_1 + a_{12} * s_2}$$
$$\mathbf{x_2 = a_{21} * s_1 + a_{22} * s_2}$$
...
$$\mathbf{x_n = a_{n1} * s_1 + a_{n2} * s_2}$$

$\Longrightarrow$

In matrix notation :

$$\mathbf{X = A * S}$$

Independent Components Analysis
ICA looks for "*meaningful*" vectors

Hypotheses :

1) No reason for the variations in one pure signal to depend *in any way* on
the variations in another pure signal

Pure source signals should therefore be « independent »

2) The measured signals being combinations of several independent sources,
they should be *more gaussian* than the sources
(Central Limit Theorem)

JADE
(Joint Approximate Diagonalization of Eigenmatrices)

- Developed by Cardoso and Souloumiac in 1993

- A *blind source separation* method to extract independent non-Gaussian sources from signal mixtures with Gaussian noise

- Based on the construction of a fourth-order *cumulant* array from the data

- Matlab function freely downloadable from

  http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m

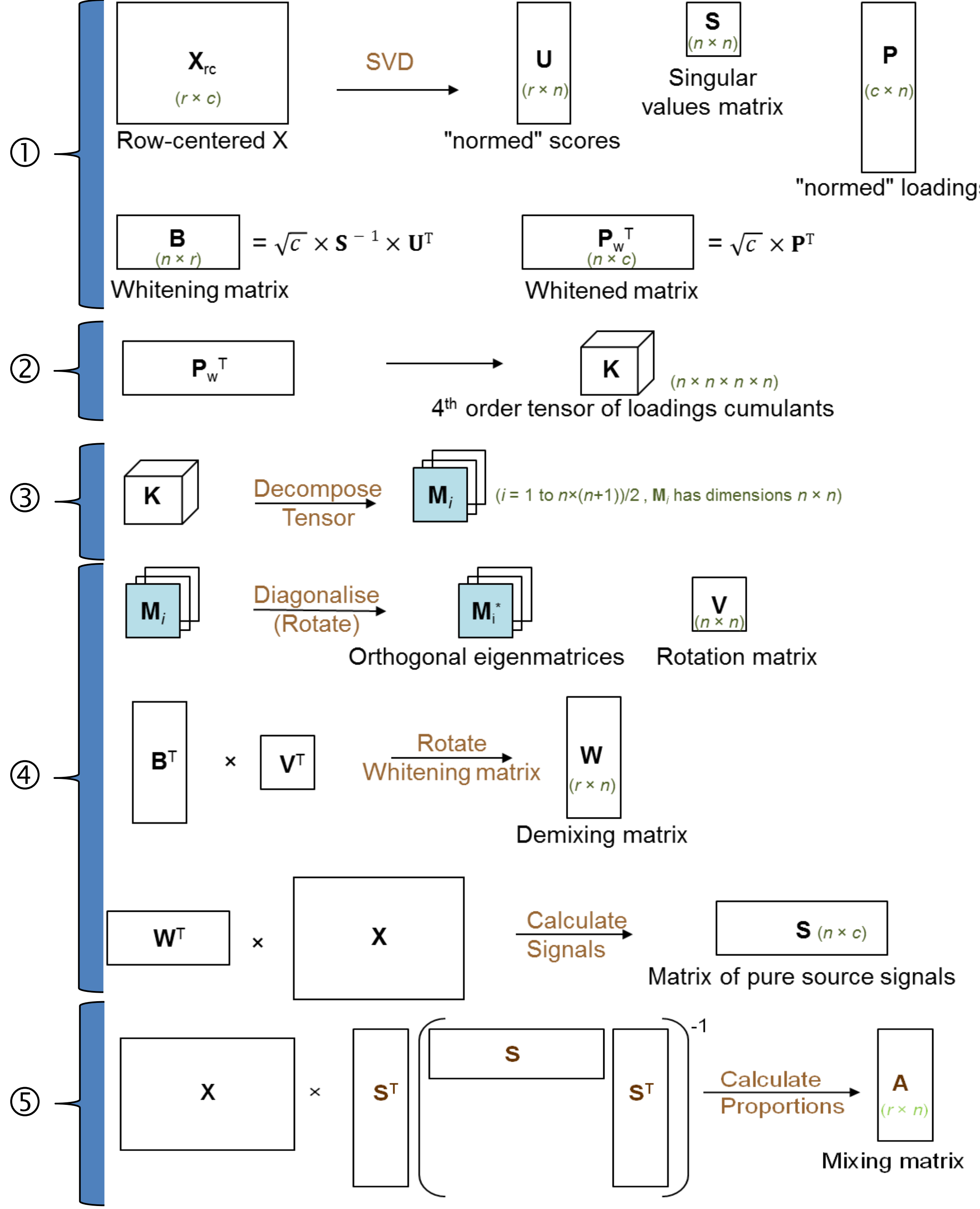
Cardoso, J-F. and Souloumiac, A.
Blind beamforming for non-Gaussian signals.
IEE proceedings-F, (1993). 140 (6) 362-370

D.N. Rutledge, D. Jouan-Rimbaud Bouveresse,
Independent Components Analysis with the JADE algorithm
Trends in Analytical Chemistry, 50, (2013) 22–32

D.N. Rutledge, D. Jouan-Rimbaud Bouveresse,
Corrigendum to “Independent Components Analysis with the JADE algorithm”
Trends in Analytical Chemistry, 67, (2015) 220

The JADE algorithm:

a multi-step procedure

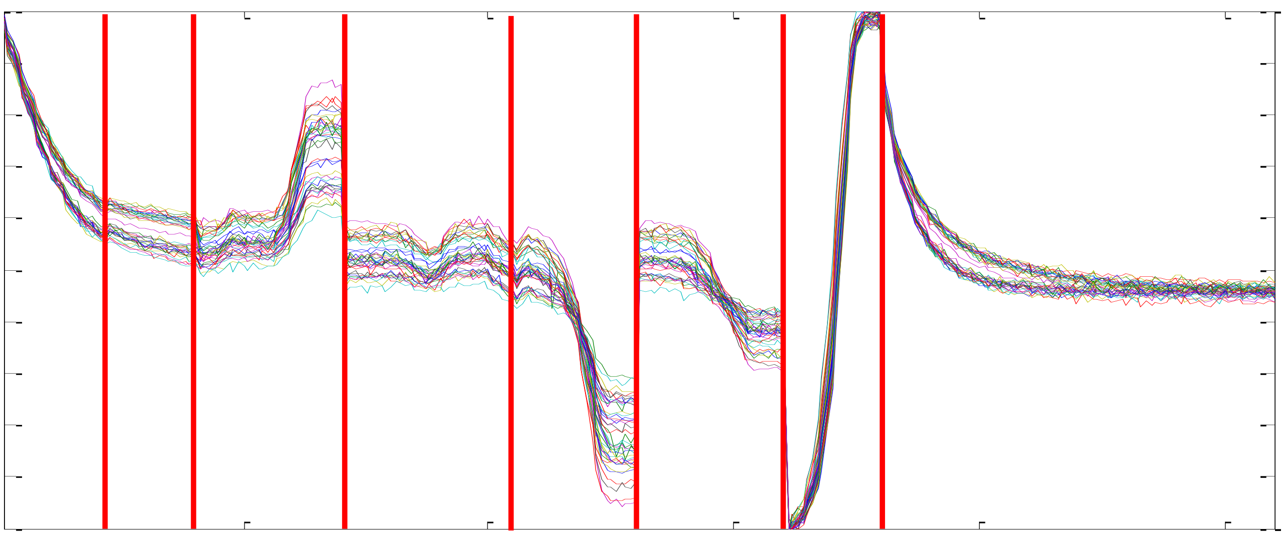


Comparison of ICA and ComDim with ComDim-ICA

Application to TD-NMR Lignin-Starch data

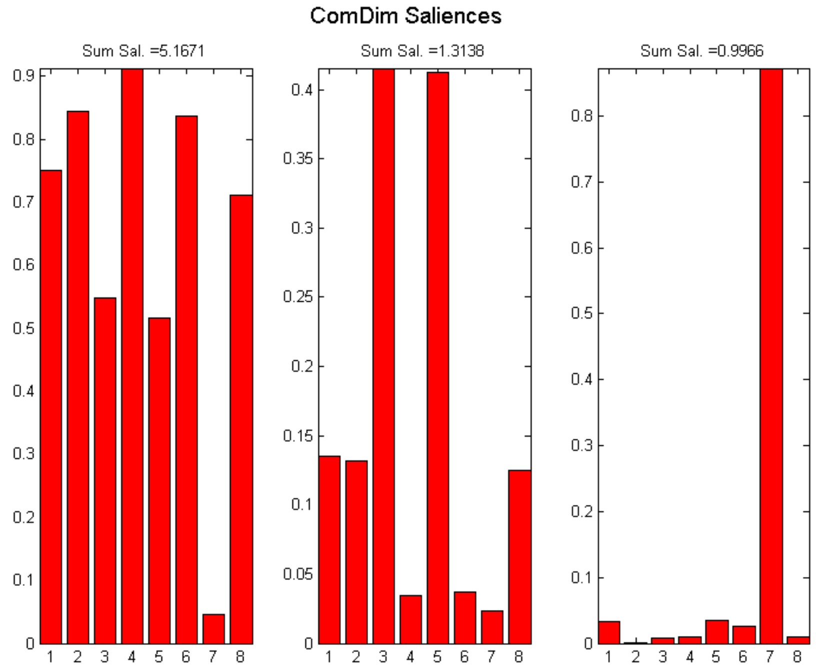
20 samples in triplicate, with different characteristics :

- 2 Shapes : Films / Cylinders
- 2 Moisture levels : stabilized in atmospheres at 33% / 75% H_2O
- 5 Lignin concentrations : 0%, 5%, 10%, 15%, 30%

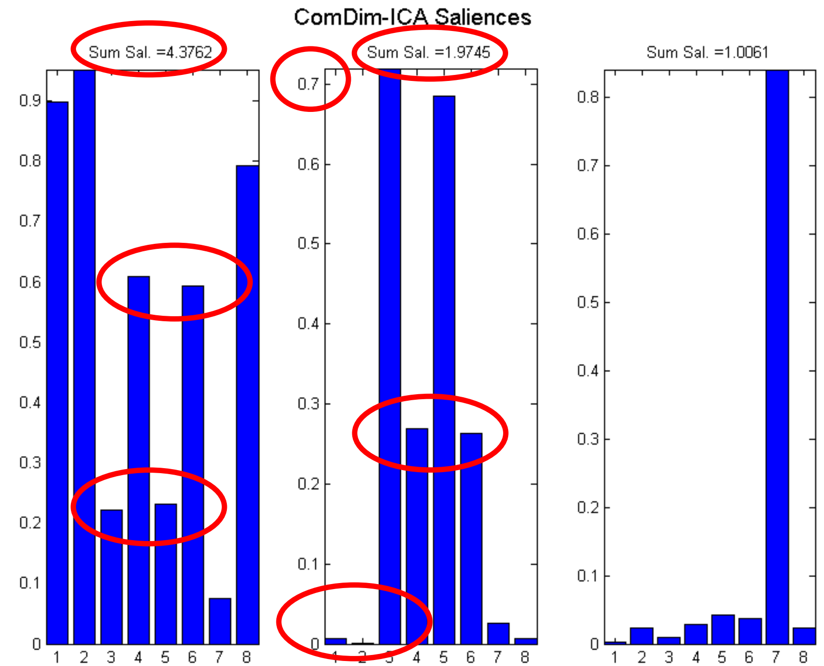


8 types of Time Domain-NMR signals

Comparison of ComDim with ComDim-ICA

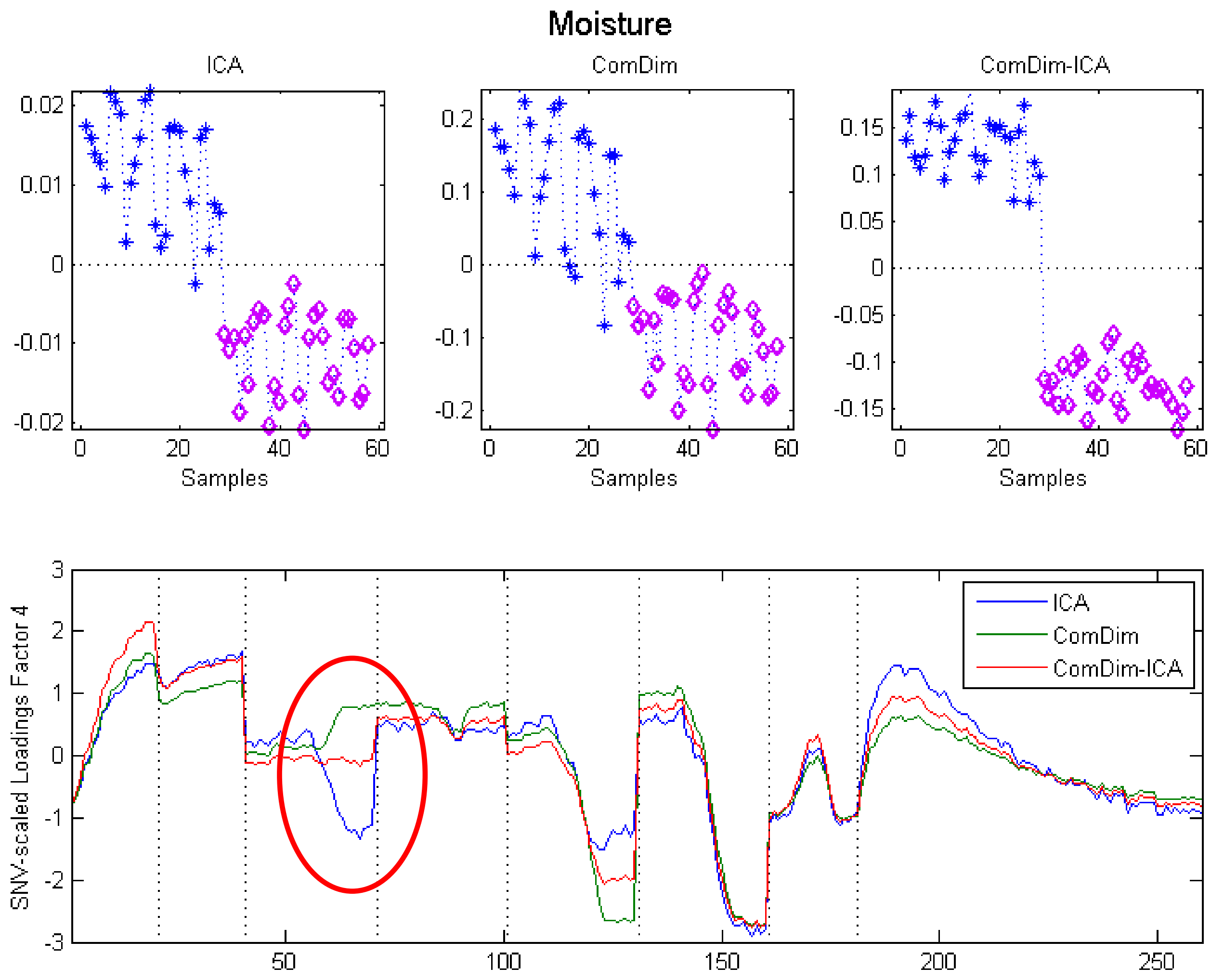


ComDim Saliences for CC1, CC2 & CC3
Convergence in 124 mS

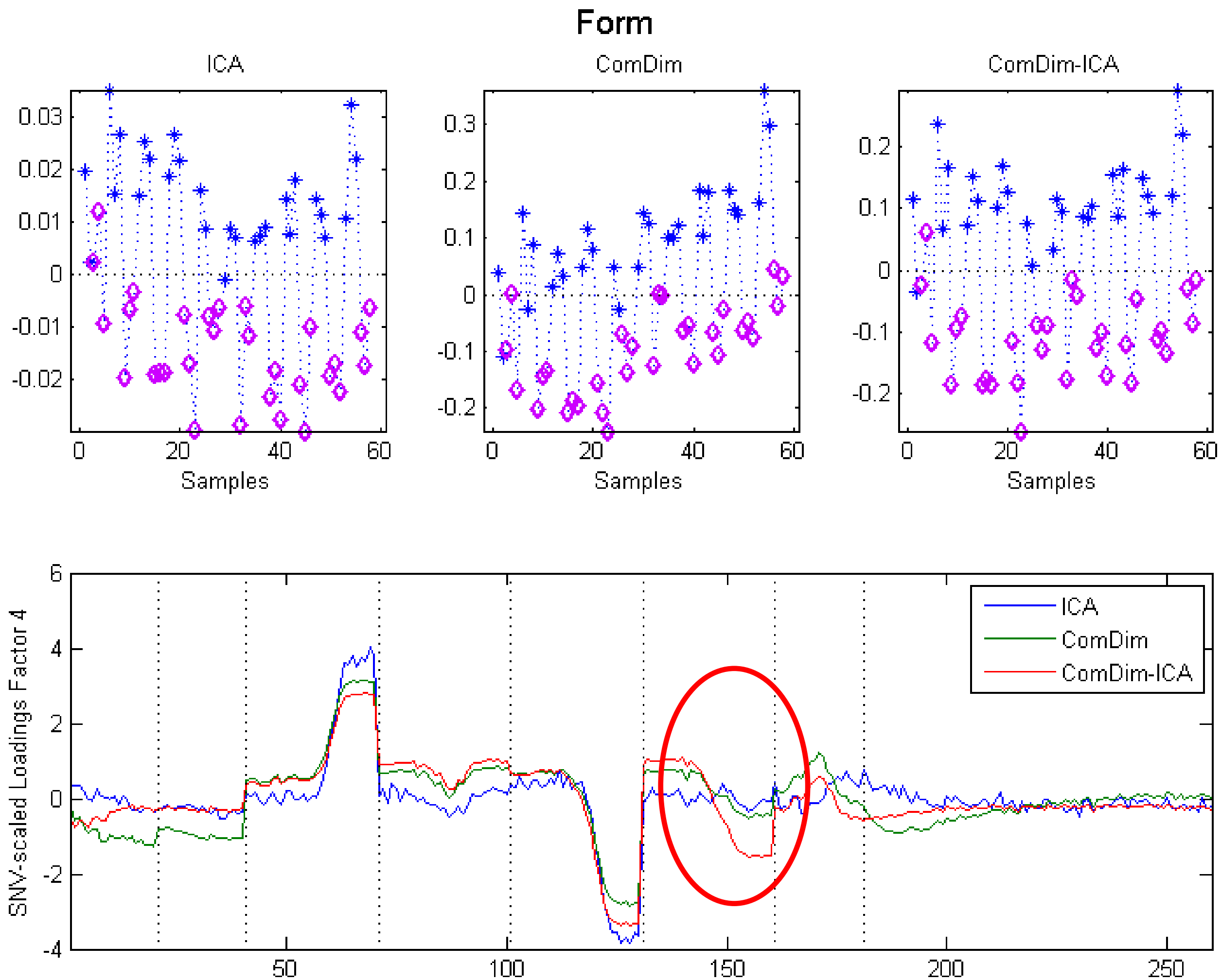


ComDim-ICA Saliences for CC1, CC2 & CC3
Convergence in 104 mS

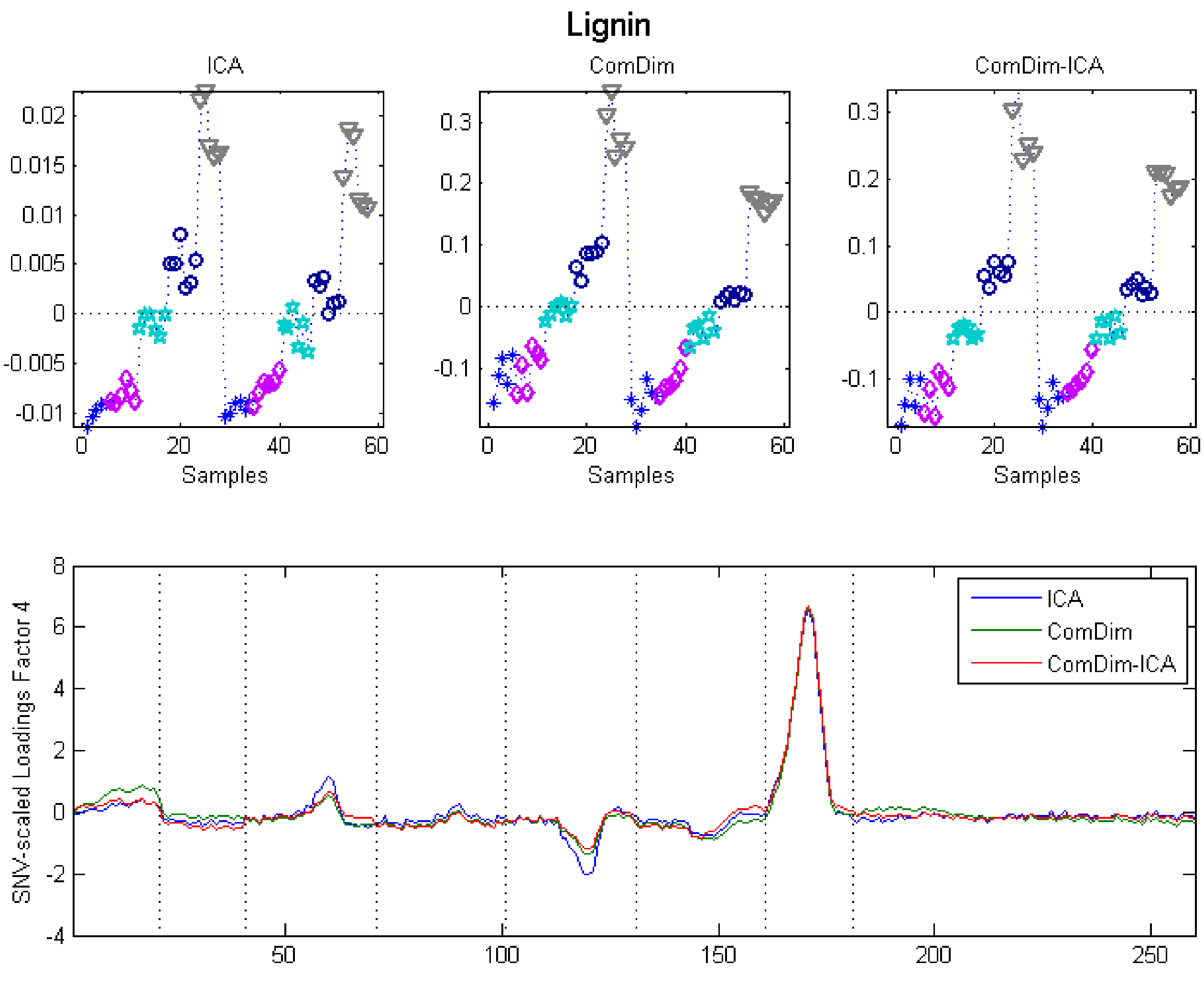
Comparison of ICA and ComDim with ComDim-ICA



Comparison of ICA and ComDim with ComDim-ICA



Comparison of ICA and ComDim with ComDim-ICA



Conclusion

ComDim-ICA

A non-supervised multi-block method
ICA on iteratively re-weighted concatenated data tables

- Better than *ICA on unweighted* concatenated data tables

- Better than ComDim
  (*PCA on iteratively re-weighted* concatenated data tables)

Thank you