L04. Outer Product Analysis (OPA) as a method to study the relations between sets of variables measured on the same individuals

D. Rutledge
Laboratoire de Chimie Analytique, AgroParisTech, Paris, France

Let **X** and **Y** be two groups of variables containing *p* and *q* variables respectively, measured on the same *N* individuals.

The analysis of the relations between the two tables **X** and **Y** is often carried out by Tucker analysis [1]. This analysis is based on the singular values decomposition of the matrix of covariances between the two groups of variables (*1/n* $\mathbf{X}^T\mathbf{Y}$). This decomposition of the covariance matrix corresponds to the successive search of couples of variables ($\mathbf{T}_H = \mathbf{X}A_H$ $\mathbf{U}_H = \mathbf{Y}B_H$) such that the covariance between $\mathbf{T}_H$ et $\mathbf{U}_H$ is maximum, under the constraints that the axes A_H are orthogonal as are the axes B_H.

Recently, Barros *et al*. [2] proposed an alternative procedure based on the analysis of a three-way table generated by the outer product between two tables **X** and **Y**. The three entries of this cube are respectively the individuals, the variables of **X** and the variables of **Y**. The analysis this three-way table can be carried out by unfolding followed by standard multivariate methods such as *PCA*, *ICA*, *PLS* etc., or directly by multiway methods such as *PARAFAC*.

Compared to the often-used Tucker analysis for the study of the relations between two tables **X** and **Y**, the first interesting aspect of OPA, is that the Tucker method is in fact *a compromise analysis (uniform average) of the cube of outer products between* **X** *and* **Y**. The matrix of covariance between the two groups of variables X and Y is equal to the average of the cube of outer products along the direction of the individuals *I* (averages by columns).

In an explicit way:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i = \frac{1}{n}\mathbf{X}^T\mathbf{Y}$$

The second interesting aspect is that with OPA it is possible to study relations among several tables, **W**, **X** and **Y**, by calculating multiple Outer Products to generate a hyper-cube.

The most interesting property of OPA is that it is possible to not only unfold the outer products and apply classical multivariate methods, but it is also possible to retain the cubic structure and apply multiway methods such as *PARAFAC* which are already well-known for the analysis of cubic data.

In this presentation, Outer Product Analysis will be compared to other techniques such as Tucker [1], Generalised 2D-correlation spectroscopy [3], and PLS2 regression [4] using examples taken from many fields including Time-Domain NMR, Mid- and Near Infrared spectroscopy, X-ray diffraction and NMR spectroscopy.

References

1. LR Tucker, 1958, *Psychometrika*, **23**, 111 -136.
7. A.S. Barros, M. Safar, M.F. Devaux, P. Robert, D. Bertrand, D.N. Rutledge, 1997, *Applied Spectroscopy*, **51** (9) 1384-1393
8. I Noda, A.E. Dowrey, C. Marcott, G.M. Story, Y. Osaki, 2000, *Applied Spectroscopy*, **54**, 236A-248A.
9. R. Lew and S. T. Balke, 1993, *Applied Spectroscopy*, **47**, 1747-1750