Multi-way data

Background. There is considerable interest in multi-way data analysis as reflected by recent special issue on multi-way data analysis in Journal of Chemometrics[2], shows that multi-way analysis is used to approach many problems in applied sciences. From mathematical point of view there are two practical problems, when handling multi-way data. The first is the notation to use. Authors often use special symbols to simplify the notation, which sometimes makes the work difficult to read. The other problem is that we don't have geometric interpretations like we often have for matrices. Also, some magnitudes like e.g., the inverse or generalized inverse are not well defined. Here we shall try to keep notation and formulae simple.

The H-principle and associated algorithms provide with a natural framework to handle multi-way data. Thus, we shall show how we carry out the modeling in steps, where weight vectors are found that reflect the emphasis of analysis. The weight vectors generate loading vectors, which again produce the transformation or causal vectors as shown in previous section. The importance of this framework is due to that the standard techniques of linear regression analysis (e.g. influence diagnostics, bootstrapping) can be carried out in the same way as in a standard regression analysis. Furthermore, when the data reduce to two-way data, the methods the methods reduce to the respective type of regression analysis, if we have **X** and **Y**, and PCA, if we only have **X**.

Multi-way data are traditionally handled by analysis of variance techniques. These methods only specify the mean value structure of the data. It is often more informative to use loading/score vectors to study the variation in data. E.g., analysis may result in that there is a significant interaction, which may be difficult to utilize. Score vectors will show more precisely how the variation in data is that has this interaction.

Two-way data are given by two indices, $\mathbf{X}=(x_{ij})$. If the data are three-way data, we have three indices, $\mathbf{X}=(x_{ijk})$. The same or similar procedures can be applied to three way or multi-way data. The difference lies in that there are more possibilities in choosing the weight vectors. In Fig. 1 we show a schematic illustration of the possibilities of weight vectors in the case of three-way data. If there is one weigh vector w, the results will be a matrix $\mathbf{T}=(t_{jk})$ with $t_{ik} = \Sigma_j x_{ijk}\, w_j$. Similarly, if two weight vectors are given, $\mathbf{w}_1$ and $\mathbf{w}_2$, the results is a vector $\mathbf{t}=(t_i)$, $t_i=\Sigma_{jk}x_{ijk}w_{1j}w_{2k}$. In case of three weight vectors we get one number, $c = \Sigma_{ijk}x_{ijk}w_{1j}w_{2k}w_{3i}$. It will be convenient to use the multiplication $\times_i$ to indicate the mode or index of multiplication. Using this notation, see Ref [??], we can write $\mathbf{t}=\mathbf{X}\times_2\mathbf{w}_1\times_3\mathbf{w}_2$ instead of $t_i=\Sigma_{jk}x_{ijk}w_{1j}w_{2k}$. The size or dimension of the left hand side can be derived from the multiplication. Furthermore, it is standard to use the notation of Kronecker products. E.g., $\mathbf{w}_3\otimes\mathbf{w}_1\otimes\mathbf{w}_2$ is three way and equal $(w_{3i}w_{1j}w_{2k})$. We shall denote by $\mathbf{w}_3$ the weight vector associated with the first index in $\mathbf{X}$. The reason is that the first index is often the sample mode as illustrated in Fig. 2, where we do not optimise $\mathbf{w}_3$, but compute it as proportional to the score vector.

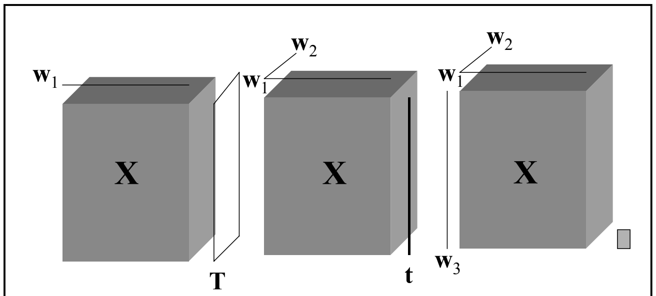

Figure 1. Schematic illustration of weight vectors for three-way data.

Weight vectors. The choice of weight vectors reflects the choice of analysis or method. Therefore, the subject of choosing weight vectors is a large one. Here we shall only consider three ways to determine the weight vectors. Note that there is equally many weight vectors as the modes. Here we consider only three way data, and thus there are three weight vectors at each iteration.

The first case is where the weight vectors, $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$, are optimized simultaneously. In this case we look at the constant c,

$$c= \Sigma_{ijk}\, x_{ijk}\, w_{1j}\, w_{2k}\, w_{3i} = \Sigma_i \left[\, \Sigma_{jk}\, x_{ijk}\, w_{1j}\, w_{2k} \,\right] w_{3i} = \mathbf{b}^T\mathbf{w}_3$$

The initial values of the weight vectors can be non-zero row/column in $\mathbf{X}$. Then this equation is iterated. $\mathbf{w}_3$ is computed as $\mathbf{w}_3=\mathbf{b}/|\mathbf{b}|$, and this value inserted in the equation. Then $\mathbf{w}_2$ is eliminated similarly and estimated as the new value of $\mathbf{b}$, $\mathbf{w}_2=\mathbf{b}/|\mathbf{b}|$. This is continued until the value of c does not change (less than 10^{-5}). This method is an extension of the power method of finding the eigen vector associated with the largest eigen value.

The second is where only one weight vector is found. In this case we find the optimal weights independently of the other weight vectors. Let us consider the weight vector $\mathbf{w}_1$. When multiplied by $\mathbf{X}$ it gives the matrix $\mathbf{T}$, with $t_{ik}=\Sigma_j (x_{ijk}\, w_{1j})$. Following the H-principle we want the covariance between $\mathbf{X}$ and $\mathbf{T}$ to be as large as possible, which is a vector $\mathbf{p}=(p_j)$,

$$p_j = \Sigma_{ik}\, x_{ijk}\, t_{ik} = \Sigma_{ik}\Sigma_n\, w_{1n}\, x_{ink}\, x_{ijk} = \Sigma_n\, w_{1n}\, [\Sigma_{ik}\, x_{jnk}\, x_{ijk}\,] = \Sigma_n\, w_{1n}\, c_{nj}$$

or $\mathbf{p}=\mathbf{C}\mathbf{w}_1$, where $\mathbf{C}=(c_{nj})$ and $c_{nj}=\Sigma_{ik}\, x_{jnk}\, x_{ijk}$, $(n,j=1,2,\ldots,J)$. Maximizing the size of $\mathbf{p}$, $|\mathbf{p}|^2$, is equivalent to finding the eigen vector corresponding to the largest eigen value to the eigen system

$$\mathbf{C}^T\mathbf{C}\mathbf{w}_1 = \mathbf{C}^2\mathbf{w}_1 = \lambda\, \mathbf{w}_1, \qquad |\mathbf{w}_1|=1.$$

The weight vector found in this way can be compared to the result of the previous procedure.

As a third method we shall look at the case of multi-way regression. The situation illustrated in Fig. 2. Following the H-principle we are looking for a vector $\mathbf{t}$, $\mathbf{t}=\mathbf{X}\times_2\mathbf{w}_1\times_3\mathbf{w}_2$, and another $\mathbf{u}$, $\mathbf{u}=\mathbf{Y}\times_2\mathbf{v}_1\times_3\mathbf{v}_2$, such that the covariance $(\mathbf{t}^T\mathbf{u})$ is as large as possible. The weight vectors are found by the same routine as described in the first method. We have

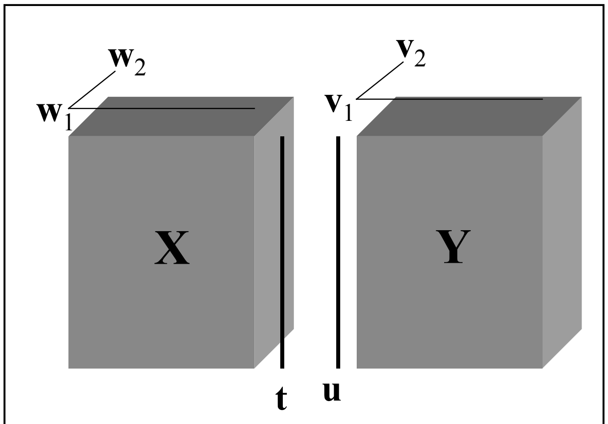

Figure 2. Schematic illustration of three-way regression analysis.

$$(\mathbf{t}^T\, \mathbf{u}) = \Sigma_i(\Sigma_{jk}\, x_{ijk}\, w_{1j}\, w_{2k})(\Sigma_{mn}\, y_{imn}\, v_{1m}\, v_{2n}) = \Sigma_{ijkmn}\;\, x_{ijk}\, y_{imn}\, w_{1j}\, w_{2k}\, v_{1m}\, v_{2n}$$

$$= \Sigma_n\, [\, \Sigma_{ijkm}\;\, x_{ijk}\, y_{imn}\, w_{1j}\, w_{2k}\, v_{1m}\,]\, v_{2n} = \mathbf{b}^T\mathbf{v}_2$$

This is iterated over the weight vectors until the covariance does not increase (within the limit of 10^{-5}).

It is possible that a weight vector is not optimised. In that case it is chosen as proportional to the corresponding loading vector.

Loading vectors. The loading vectors are computed for each mode. They are computed by multiplying **X** by all weight vectors except the one associated with that mode. Thus, for given weight vectors, $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$,

$\mathbf{p}_3=(p_{3i})=(\Sigma_{jk}\, x_{ijk}w_{1j}\, w_{2k})$, $\mathbf{p}_1=(p_{1j})=(\Sigma_{ik}\, x_{ijk}w_{3i}w_{2k})$,$\mathbf{p}_2=(p_{2k})=(\Sigma_{ij}\, x_{ijk}w_{3i}\, w_{1j})$.

If e.g., $\mathbf{w}_3$ has not been optimised, it is computed as $\mathbf{w}_3=\mathbf{p}_3/|\,\mathbf{p}_3|$.

Adjustment of X. When the loading vectors have been computed the data matrix **X** is adjusted as

$$\mathbf{X} \leftarrow \mathbf{X} - d^2\, \mathbf{p}_3 \otimes \mathbf{p}_1 \otimes \mathbf{p}_2,$$

where

$$d = 1/\Sigma_{ijk} x_{ijk} w_{1j} w_{2k} w_{3i}.$$

The modelling starts over using the adjusted data matrix **X**.

Causal vectors. Generalised inverse. The causal vectors, the **r**'s, are computed in the same way as specified in the two-way analysis, see [??]. The generalised inverse is now given by

$$\mathbf{X}^+ = \mathbf{r}_3 \otimes \mathbf{r}_1 \otimes \mathbf{r}_2 + \ldots\,.$$

When we multiply $\mathbf{X}^+$ by **X** we, must multiply them mode-wise to get the identity matrices of appropriate dimension.

Properties of the decomposition. There are some important properties of the decomposition that is derived by above procedure. Let $\mathbf{X}_{(1)}$ be the reduced matrix, $\mathbf{X}_{(1)}=\mathbf{X} - d^2\, \mathbf{p}_3 \otimes \mathbf{p}_1 \otimes \mathbf{p}_2$. If we multiply $\mathbf{X}_{(1)}$ by say, $\mathbf{w}_3$, $(\mathbf{X}_{(1)} \times_1 \mathbf{w}_3)$, we get a J times K matrix that is orthogonal to $\mathbf{w}_1$ and $\mathbf{w}_2$,

$$(\mathbf{X}_{(1)} \times_1 \mathbf{w}_3) \times_2 \mathbf{w}_1 = \mathbf{0}, \qquad (\mathbf{X}_{(1)} \times_1 \mathbf{w}_3) \times_3 \mathbf{w}_2 = \mathbf{0},$$

This follows from the equation

$$\mathbf{X}_{(1)} \times_1 \mathbf{w}_3 = \mathbf{X} \times_1 \mathbf{w}_3 - d\, \mathbf{p}_1\, \mathbf{p}_2^T,$$

the property $\mathbf{w}_1^T\mathbf{p}_1=\mathbf{w}_2^T\mathbf{p}_2= \mathbf{w}_3^T\mathbf{p}_3=1/d$ and the definition of the loading vectors. These equations are consistent with the two-way data in the sense that they reduce to the respective equations, if one of the mode collapses so that the data become two-way data. This result is an important aspect of the H-principle. Whatever weight vectors $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$ there have been found, the residual data are orthogonal to the weight vectors chosen as shown above. The only requirement is that the scaling constant, 1/d, is not the zero. In multi-way data analysis this orthogonality is the only thing that is generally valid. One can say that in multi-way analysis we leave the area of linearity and orthogonality.

Summary. We have presented an approach to handle multi-way data that is a natural extension of standard analysis, where (two-way) matrices are used. In fact, the associated algorithms are for loops over the modes, and reduce to normal two-way analysis in case of two-way data. It means that if one has worked with and is well acquainted with two-way analysis, the multi-way analysis is the same type of analysis, just having more modes to work with.

The formulae of regression coefficients, predicted values etc. follow the same formulae as given in the two-way case. We also have similar formulae for influence diagnostics, bootstrapping, etc. see Ref. [??].

The approach is consistent with the H-principle that the analysis is carried out in steps, where at each step we seek maximising the covariance. It has the important property that at each step we seek an optimal balance between fit and prediction and results at later step are orthogonal, in appropriate sense, to the weights that have used at the step.