RR 90-01

SOME POSSIBILITIES FOR THE ANALYSIS OF DYNAMIC THREE-WAY DATA

Peter Verboon

Willem J. Heiser

Department of Data Theory

University of Leiden

SOME POSSIBILITIES FOR THE ANALYSIS OF DYNAMIC THREE-WAY DATA

Peter Verboon

Willem Heiser

In this paper we will discuss the possibility of some models for the extension of the two-way DYNAMALS method. For each of these models a loss function is presented and it is shown how the parameters of these functions can be found. One of the models, the so-called points of view approach, will be discussed in more detail. Besides the technical problems corresponding with the proposed models, we will also discuss some data-analytic aspects.

1. Introduction

Consider the following problem. We have a datamatrix of order T by m. The rows are dependent observations; for the moment let's assume they are consecutive time points. The columns of the datamatrix are variables which can be divided into two sets, **X** and **Y**. The set **X** contains m_1 so-called input variables and **Y** contains m_2 output variables, with $m_1+m_2=m$. We assume that the relations between the input and output variables are linear and that they are channeled by p latent variables, also called the p-dimensional state-space of the system. These latent variables are represented by the p-dimensional matrix **Z**. Furthermore we assume a stationary time process. See Figure 1.

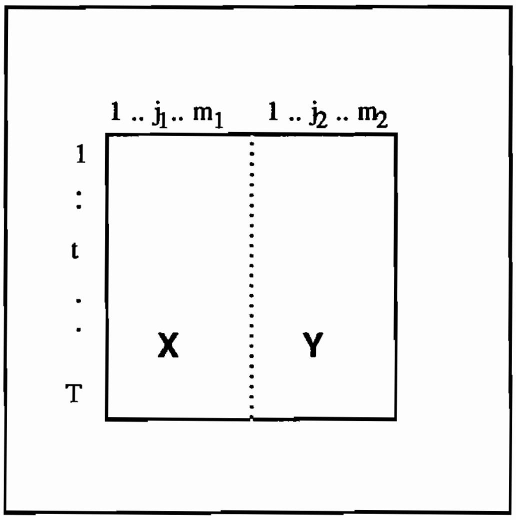

Figure 1. Two-way datamatrix

To analyze such a linear dynamic system in the context of exploratory data analysis we can make use of the DYNAMALS method (De Leeuw & Bijleveld, 1988; Bijleveld, 1989). DYNAMALS considers the following well-known state-space model:

measurement equation $\mathbf{Y} = \mathbf{Z}\,\mathbf{H}'$ (1.1)

system equation $\mathbf{Z} = \mathbf{B}\,\mathbf{Z}\,\mathbf{F}' + \mathbf{X}\,\mathbf{G}'$ (1.2)

The matrices **G** (p by m_1) and **H** (m_2 by p) contain weights that indicate the importance of the input and the output variables, respectively. The matrix **F** (p by p) is called the transition matrix and shows the extent of the time dependency. And finally the **B** (T by T) matrix defines the relations between the rows of the datamatrix. This **B** matrix is called shift or contiguity matrix and in the current program it can be defined by the user.

In DYNAMALS a least squares loss function is minimized, based on (1.1) and (1.2).

$$\sigma(\mathbf{F},\mathbf{G},\mathbf{H},\mathbf{Z}) = \omega^2\ \mathrm{SSQ}\ (\mathbf{Z} - \mathbf{B}\,\mathbf{Z}\,\mathbf{F}' - \mathbf{X}\,\mathbf{G}') + \mathrm{SSQ}\ (\mathbf{Y} - \mathbf{Z}\,\mathbf{H}') \qquad (1.3)$$

The parameter ω^2 can be chosen by the user to control the relative importance of the input or output variables. The algorithm to solve this problem is iterative. Each iteration consists of three distinct steps. In the first step the parameter matrices **F**, **G** and **H** are computed, in the next step **Z** is computed and finally the variables are optimally scaled (for this last part, cf. Gifi, 1990).

The datamatrix shown in Figure 1 is a two-way matrix: time points by variables measured in one system. This means that we are basically analyzing so-called $N = 1$ designs. However, if we have more systems or subjects to analyze the data will look like the format shown in Figure 2. We then cannot use DYNAMALS as described above without adaptions and therefore we will need a form of *three-way DYNAMALS.*

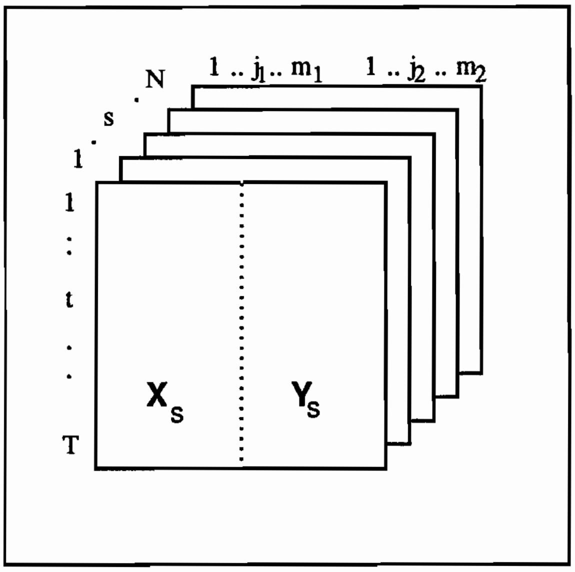

Figure 2. Three-way datamatrix

The aim of this paper is to study several extensions of the two-way case to accomodate the third way. We will present five possibilities, discuss their data-analytic merits and try to solve most of their technical problems. The models that we propose have one important feature in common: they allow the variables to be transformed with respect to their measurement level. This feature is called optimal scaling (Gifi, 1990).

The units of the third way will be referred to as subjects; however, it should be kept in mind that these subjects could be any independent data source: human beings, animals, rivers, crossroads etc..

In the remainder of this paper we will refer to the analysis of these three-way data by the acronym STADS, which stands for simultaneous three-way analysis of dynamic systems. Hence, with the name DYNAMALS we now refer to the two-way problem.

2. First extension: unweighted averaging

As a first extension of DYNAMALS we propose a very simple and straightforward possibility. We reduce our three-way datamatrix to a two-way matrix by averaging over the N subjects. Next this average matrix is analyzed by the ordinary DYNAMALS program. This approach is radical in the sense of data reduction, for we estimate for all subjects only one latent variable (**Z**), one transition matrix (**F**) and one pair of weights matrices (**G**,**H**). Furthermore, it needs very little extra effort compared to DYNAMALS.

However, a serious drawback of this approach is the fact that all information about the differences between subjects is left unmodelled. It is clear that information about these differences can sometimes be very interesting, if not a primary goal for the data analysis. A slight improvement of this approach might be to add one extra step to the method. Using as target values the optimal quantified variables, which are the result of the DYNAMALS analysis, we can perform a *post-hoc* optimal scaling for each subject separately. Inspection of the transformation plots may show some interesting differences between the subjects. If for a particular variable all plots look rather similar, we may conclude that with respect to this variable the subjects form a homogeneous group.

3. Second extension: the "points of view" approach

Let's assume we have found as an initial step the solution from the first extension as described before. We now introduce two extra vectors (**w**, **v**) with weight parameters for all the output and input variables, respectively. To find these weights we propose to minimize the following loss function:

$$\sigma(\mathbf{v},\mathbf{w},\mathbf{F},\mathbf{G},\mathbf{H},\mathbf{Z}) = \omega^2\Sigma_s SSQ(\mathbf{Z} - \mathbf{BZF}' - v_s\mathbf{X}_s\mathbf{G}') + \Sigma_s SSQ(w_s\mathbf{Y}_s - \mathbf{ZH}') \qquad (3.1)$$

where v_s and w_s represent the individual subject weights. We will call this the *points of view* approach, because the idea is rather similar to a technique with the same name in the context of multidimensional scaling with individual differences (Tucker and Messick, 1963, see also Meulman and Verboon, 1989).

Let's derive the least squares solution for the weights, starting with the output weights. To find these weights we need to set the partial derivatives with respect to w_s equal to zero, which yields the following solution for the w_s:

$$w_s = \frac{\mathrm{tr}\ \mathbf{Y}_s'\mathbf{ZH}'}{\mathrm{tr}\ \mathbf{Y}_s'\mathbf{Y}_s}. \qquad (3.2)$$

Next we may write

$$w_s\mathbf{Y}_s = \mathbf{Y}^+ + (w_s\mathbf{Y}_s - \mathbf{Y}^+), \qquad \text{with } \mathbf{Y}^+ = N^{-1}\Sigma_s w_s\mathbf{Y}_s. \qquad (3.3)$$

Substitution of (3.3) into (3.1) and working out the result will give for the second part of loss function (3.1):

$$\Sigma_s SSQ(w_s\mathbf{Y}_s - \mathbf{ZH}') = N\ SSQ(\mathbf{Y}^+ - \mathbf{ZH}') + \Sigma_s(w_s\mathbf{Y}_s - \mathbf{Y}^+) \qquad (3.4)$$

It follows that the parameter matrices **Z** and **H** can be computed in relation to $\mathbf{Y}^+$ instead of in relation to the subject dependent matrices $\mathbf{Y}_s$. So at least the computation of **H** causes no additional problems compared to DYNAMALS. Now let's examine the input weights which are a little more complicated to compute. Setting the partial derivatives with respect to v_s to zero, yields the following result for the weights v_s:

$$v_s = \frac{\text{tr } \mathbf{Z'X_sG'} - \text{tr } \mathbf{FZ'B'X_sG'}}{\text{tr } \mathbf{GX_s'X_sG'}}. \qquad (3.5)$$

Furthermore, we write

$$v_s\mathbf{X}_s = \mathbf{X}^+ + (v_s\mathbf{X}_s - \mathbf{X}^+), \qquad \text{with } \mathbf{X}^+ = N^{-1}\Sigma_s v_s \mathbf{X}_s. \qquad (3.6)$$

By substituting this in (3.5) all double product terms vanish and very much like the second part, the first part of (3.1) can be written as

$$\omega^2\Sigma_s \text{SSQ}(\mathbf{Z} - \mathbf{BZF'} - v_s\mathbf{X}_s\mathbf{G'}) = \omega^2 N \text{ SSQ}(\mathbf{Z} - \mathbf{BZF'} - \mathbf{X}^+\mathbf{G'}) +$$

$$\omega^2 \text{ SSQ } \Sigma_s(v_s\mathbf{X}_s - \mathbf{X}^+)\mathbf{G'}. \qquad (3.7)$$

Now from (3.7) and (3.4) it follows that $\mathbf{H}, \mathbf{F}$ and $\mathbf{Z}$ can be computed similarly to DYNAMALS, because they only depend on the average matrices of the variables. However, the computation of $\mathbf{G}$ will give some extra problems, since this matrix cannot be eliminated from the subject dependent part and thus it is also present in the right most part of (3.7).

Since in STADS we cannot use the concatenation of $\mathbf{BZ}$ with $\mathbf{X}^+$ to compute $\mathbf{F}$ and $\mathbf{G}$ in one step as is done in Bijleveld and De Leeuw (1988), we also have to derive an explicit solution for $\mathbf{F}$. From (3.7) it is easy to derive the conditionally optimal least squares estimates for $\mathbf{F}$ and $\mathbf{G}$, as follows:

$$\mathbf{F} = (\mathbf{Z'B'BZ})^{-1}(\mathbf{Z'BZ} - \mathbf{GX^{+'}BZ}), \qquad (3.8)$$

$$\mathbf{G'} = N\,(\Sigma_s v_s^2 \mathbf{X}_s'\mathbf{X}_s)^{-1}\mathbf{X}^{+'}(\mathbf{Z} - \mathbf{BZF'}). \qquad (3.9)$$

We could also assume equal weights for input and output variables ($v_s = w_s$). The solution for these restricted weights is then given by

$$u_s = \frac{\omega^2(\text{tr } \mathbf{Z'X_sG'} - \text{tr } \mathbf{FZ'B'X_sG'}) + \text{tr } \mathbf{Y_s'ZH'}}{\omega^2(\text{tr } \mathbf{GX_s'X_sG'}) + \text{tr } \mathbf{Y_s'Y_s}}. \qquad (3.10)$$

The computations for the parameter matrices obviously do not change when we have equal weights. Furthermore we can see from (3.10) that if ω goes to infinity, u_s goes to v_s and if ω is zero u_s becomes w_s. This is what should be expected.

At this point the loss function in (3.1) is only defined for just one point of view. However, we could define one or more new vector of weights. In general this would would enable us to find *r* different points of view. In order to find these new weights vectors we will rewrite the loss function. For both the input and output weights corresponding with the first point of view they are as follows:

$$\sigma(\mathbf{v}_1) = \omega^2 \Sigma_t \, \mathrm{SSQ}(\mathbf{z}_t - \mathbf{FZ'b}_t - \mathbf{GX}_t'\mathbf{v}_1). \tag{3.11a}$$

$$\sigma(\mathbf{w}_1) = \Sigma_t \, \mathrm{SSQ}(\mathbf{Y}_t'\mathbf{w}_1 - \mathbf{Hz}_t). \tag{3.11b}$$

The vector $\mathbf{b}_t$ is the t^{th} row of the matrix $\mathbf{B}$. Note that the summation is now over the T time points. The vectors $\mathbf{v}_1$ and $\mathbf{w}_1$ have length N and represent the weights for the first point of view. Setting the partial derivatives with respect to the weights to 0, yields the following solution for the weights, when the problem is considered in this alternative form:

$$\mathbf{w}_1 = (\Sigma_t \, \mathbf{Y}_t\mathbf{Y}_t')^{-1}(\Sigma_t\mathbf{Y}_t\mathbf{Hz}_t), \tag{3.12a}$$

$$\mathbf{v}_1 = (\Sigma_t \, \mathbf{X}_t\mathbf{G}'\mathbf{GX}_t')^{-1}(\Sigma_t\mathbf{X}_t\mathbf{G}'\mathbf{z}_t - \Sigma_t\mathbf{X}_t\mathbf{G}'\mathbf{FZ}'\mathbf{b}_t). \tag{3.12b}$$

So instead of using repeatedly (3.2) and (3.5) we could also use (3.12) to compute the weights for one point of view. Now for the second point of view we have to compute new weight vectors, but these vectors should not be correlated with the first ones. In other words we have to impose the restrictions $\mathbf{w}_1'\mathbf{w}_2 = 0$ and $\mathbf{v}_1'\mathbf{v}_2 = 0$. The computation of this second point of view is done by means of the anti-projection matrix $\mathbf{J}_1 = \mathbf{I} - \frac{\mathbf{w}_1\mathbf{w}_1'}{\mathbf{w}_1'\mathbf{w}_1}$. The restriction of orthogonality can now be written in the form $\mathbf{w}_2 = \mathbf{J}_1\mathbf{w}_2$. This yields for the second weight vector the following loss function:

$$\sigma(\mathbf{w}_2) = \Sigma_t \, \mathrm{SSQ}(\mathbf{Y}_t'\mathbf{J}_1\mathbf{w}_2 - \mathbf{Hz}_t). \tag{3.13}$$

The procedure described above is a successive solution to the STADS problem. In the solution described above the weights are not restricted to be positive. From a data analysis point of view, however, positive weights are possibly preferable (c.f. Meulman and Verboon, 1989).

There is also a simultaneous approach possible. If we propose to normalize the weights, then for the k^{th} point of view we obtain $\mathbf{w}_k'\mathbf{w}_k = 1$ and $\mathbf{v}_k'\mathbf{v}_k = 1$. With these restrictions we actually have a so-called asymmetric orthogonal Procrustes problem (Ten Berge and Knol, 1984). To see this more clearly we are going to write the loss functions in a still different form. At first for the output weights:

$$\sigma(\mathbf{W}) = \mathrm{SSQ}(\mathbf{Y}^*\mathbf{W} - \mathbf{Q}^*) \qquad \text{with } \mathbf{W'W} = \mathbf{I}. \tag{3.14a}$$

We arrive at formulation (3.14a) as follows: first we join the r vectors $\mathbf{w}_k$ in a matrix $\mathbf{W}$, which changes (3.11b) into

$$\sigma(\mathbf{W}) = \Sigma_t \, \mathrm{SSQ}(\mathbf{Y}_t'\mathbf{W} - \mathbf{Hz}_t\mathbf{e}'), \tag{3.14b}$$

where **e** is the unit vector of length r. This vector repeats the vector $\mathbf{Hz}_t$. The notation is simplified by defining

$$\mathbf{Q}_t = \mathbf{Hz}_t\mathbf{e}', \tag{3.14c}$$

which is of the order m_2 by r, this yields

$$\sigma(\mathbf{W}) = \Sigma_t \, SSQ(\mathbf{Y}_t'\mathbf{W} - \mathbf{Q}_t). \tag{3.14d}$$

Finally we define the concatenated matrices $\mathbf{Y}^* = (\mathbf{Y}_1,...,\mathbf{Y}_t,...,\mathbf{Y}_T)'$ of the order N by Tm_2 and $\mathbf{Q}^* = (\mathbf{Q}_1,...,\mathbf{Q}_t,...,\mathbf{Q}_T)$ of the order Tm_2 by r, which brings us to (3.14a) again. This is a standard problem, which has been solved by Green and Gower (1979).

For the input weights a similar procedure is followed. From (3.11a) we first define the difference vector $\mathbf{u}_t$ as

$$\mathbf{u}_t = \mathbf{z}_t - \mathbf{FZ}'\mathbf{b}_t. \tag{3.15a}$$

Next we join the r weights vectors $\mathbf{v}_k$ in the matrix $\mathbf{V}$, which yields:

$$\sigma(\mathbf{V}) = \Sigma_t \, SSQ(\mathbf{U}_t - \mathbf{GX}_t'\mathbf{V}), \tag{3.15b}$$

with $\mathbf{U}_t = \mathbf{u}_t\mathbf{e}'$. Finally the summation over t is replaced by vertically concatenated matrices. This gives:

$$\sigma(\mathbf{V}) = SSQ(\mathbf{U}^* - \mathbf{A}^*\mathbf{V}), \qquad \text{with } \mathbf{V}'\mathbf{V} = \mathbf{I} \tag{3.15c}$$

where $\mathbf{U}^* = (\mathbf{U}_1,...,\mathbf{U}_t,...,\mathbf{U}_T)'$ and $\mathbf{A}^* = (\mathbf{GX}_1',...,\mathbf{GX}_t',...,\mathbf{GX}_T')'$. The problem described by (3.15c) is a asymmetric Procrustes problem too. So for both input and output weights we have to solve an asymmetric Procrustes problem, defined by (3.14a) and (3.15c) respectively.

A necessary condition for the minimum of (3.14a) and (3.15c) is that the matrices $\mathbf{V}'\mathbf{A}^{*\prime}\mathbf{U}^{*\prime}$ and $\mathbf{W}'\mathbf{Y}^{*\prime}\mathbf{Q}^{*\prime}$ are symmetric and positive semidefinite (Ten Berge and Nevels, 1984).

Let's give some final remarks on the points of view approach. Firstly it reduces the number of subjects to a small number of points of view and a complete DYNAMALS solution is given for each point of view. So if we choose to have for instance only one point of view, a very interpretable solution arises, since we obtain a one dimensional weight vector that shows how much each subjects contributes to the solution.

4. Third extension: generalized canonical correlation approach

An idea that is slightly related to the former one considers another loss function. Again our starting point is the solution found by averaging. But now instead of loss function (3.1) we propose the following loss:

$$\sigma(\mathbf{F},\mathbf{G},\mathbf{H},\mathbf{Z},\mathbf{A}) = \omega^2\Sigma_s SSQ(\mathbf{Z} - \mathbf{BZF}' - \mathbf{X}_s\mathbf{G}') + \Sigma_s SSQ(\mathbf{Y}_s\mathbf{A}_s - \mathbf{Z}), \tag{4.1}$$

with $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$ and where $\mathbf{A}_s$ is a matrix of order m_2 by p, defined for each subject. We will refer to the entries of this matrix as component loadings. When we compare the present loss function to (3.1) we notice that the matrix **H** has been left out and its role is more less taken over by **A**. This function is very similar to the so called *omnibus* loss function introduced by Van Buuren(1990). An important difference is the presence of input variables $\mathbf{X}_s$ in (4.1), which makes it more general.

The following normalization for **A**, the three-way weights matrix of order m_2 by p by N seems to be the most interesting:

$$\Sigma_s a_{j2is}^{\ \ 2} = 1. \tag{4.2}$$

With these a_{jrs} we can make interesting plots now, for example let's assume p=2. Then for each variable we can make a two dimensional plot with the subjects represented as vectors in this two dimensional space. On the other hand we can plot the variables in the dimensional space for each separate subject. Both types of plots may show interesting interaction effects. Furthermore we are now able to define goodness of fit measures for the subjects, such as

$$\lambda_{rs} = \Sigma_{j2} a_{j2is}^{\ \ 2} / m_2, \tag{4.3}$$

or an overall fit measure like

$$\lambda_s = \Sigma_i \Sigma_{j2} a_{j2is}^{\ \ 2} / m_2 p. \tag{4.4}$$

For both measures the following inequality holds: $0 \leq \lambda \leq 1$.

For each subject the unrestricted solution for $\mathbf{A}_s$ is given by the well known product

$$\mathbf{A}_s = (\mathbf{Y}_s'\mathbf{Y}_s)^{-1}\mathbf{Y}_s'\mathbf{Z} \tag{4.5}$$

The form of the problem defined in (4.1) looks rather similar to the OVERALS problem (Van der Burg et al.,1988; Gifi, 1990). The second part of (4.1) is basically the OVERALS loss. The first part defines restrictions on the computation of **Z** and adds a new set of variables to the problem. So there are two important differences; first there is the computation of **Z,** which in the present context is computed in a 'DYNAMALS-like' way, while in the OVERALS problem it is just a weighted mean of the different sets (i.e. the $\mathbf{Y}_S$). A second difference is the specific normalization of the component loadings.

The loss function in (4.1) has been changed with respect to the DYNAMALS loss function from a *join* loss function to a so called *meet* loss function (cf. Gifi, 1990; Meulman, 1986). Because this approach is conceptually quite different from DYNAMALS we have chosen not to discuss this generalized canonical correlation extension any further, but only mention that it could be an interesting possibility, especially because it is closely related to other techniques.

5. Fourth Extension: the individual latent variables approach

In the former sections we have discussed three alternatives with respect to the STADS problem. These alternatives had one feature in common: conceptually they all consisted of an ordinary two-way DYNAMALS step after which some kind of weights are computed for the variables. So the techniques we described before consist more or less of a shell around the DYNAMALS without changing anything in this old part. The techniques to come are different in that respect.

Consider once again the basic equations in DYNAMALS as formulated in (1.1) and (1.2). Now suppose we want to estimate a latent state for each subject separately and not one general latent state for all subjects. However, the parameter matrices **F**,**G** and **H**,which characterize the dynamics of the system, are assumed to be common across subjects. This changes the basic equations in

measurement equation $\quad \mathbf{Y}_S = \mathbf{Z}_S \mathbf{H}', \qquad (5.1)$

system equation $\quad \mathbf{Z}_S = \mathbf{B}\ \mathbf{Z}_S\ \mathbf{F}' + \mathbf{X}_S\ \mathbf{G}'. \qquad (5.2)$

The solutions for the parameter matrices **F**, **G** and **H** are straight-forward and given by

$$\mathbf{H} = N^{-1}\Sigma_S(\mathbf{Y}_S'\mathbf{Z}_S), \qquad (5.3)$$

$$\mathbf{F}' vc \mathbf{G}' = \Sigma_S(\mathbf{U}_S'\mathbf{U}_S)^{-1}\Sigma_S\mathbf{U}_S'\mathbf{Z}_S, \qquad (5.4)$$

where *vc* stands for vertical concatenation. The matrix $\mathbf{U}_S$ is defined as

$$\mathbf{U}_S = \mathbf{B}\mathbf{Z}_S' \,||\, \mathbf{X}_S, \qquad (5.5)$$

where || stands for horizontal concatenation.

To find the $\mathbf{Z}_S$ we have to solve n independent majorization problems analogous to the computation of **Z** in the DYNAMALS algorithm (see De Leeuw and Bijleveld, 1988). Since we require the restriction $\mathbf{Z}_S'\mathbf{Z}_S = \mathbf{I}$, this amounts to solving a Procrustes problem for each subject in each iteration step.

6. Fifth extension: the individual time-process approach

For this final approach to the STADS problem we assume that each subject obtains its own transition matrix $\mathbf{F}_S$. So the extent of the time-, c.q. space-dependency is allowed to vary over the different systems. However, it is also assumed that there is only one matrix with latent variables which should describe the latent state of all subjects. So we can write down the following model:

measurement equation $\mathbf{Y}_S = \mathbf{Z}\ \mathbf{H}'$, (6.1)

system equation $\mathbf{Z} = \mathbf{B}\ \mathbf{Z}\ \mathbf{F}_S' + \mathbf{X}_S\mathbf{G}'$. (6.2)

A solution for $\mathbf{H}$ is given by

$$\mathbf{H} = N^{-1}(\Sigma_S \mathbf{Y}_S')\mathbf{Z}_S, \tag{6.3}$$

To solve for $\mathbf{F}_S$ and $\mathbf{G}$ we cannot use an equation like the one in (5.4) because $\mathbf{F}_S$ is indexed with a subject number and $\mathbf{G}$ is estimated for all subjects together. However, we could also estimate individual $\mathbf{G}$ matrices, i.e. add an index s to $\mathbf{G}$. Now the solution for $\mathbf{F}_S$ and $\mathbf{G}_S$ is given by

$$\mathbf{F}_S' \nu c \mathbf{G}_S' = (\mathbf{U}_S'\mathbf{U}_S)^{-1}\mathbf{U}_S'\mathbf{Z}, \tag{6.4}$$

with the matrix $\mathbf{U}_S$ defined as

$$\mathbf{U}_S = \mathbf{B}\mathbf{Z}' \| \mathbf{X}_S, \tag{6.5}$$

If we want to keep the model defined in (6.1) and (6.2) the matrix $\mathbf{G}$ can be computed as the centroïd of the $\mathbf{G}_S$. To find the matrix $\mathbf{Z}$ with the latent variables we follow the DYNAMALS approach (De Leeuw and Bijleveld, 1988), which brings us to the following problem:

$$\sigma(\mathbf{Z}) = \gamma\ \Sigma_S\ \mathrm{SSQ}\ (\mathbf{Z} - (\mathbf{Z}^O + \mathbf{S}_S), \tag{6.6}$$

which should be solved in each iteration step. The matrix $\mathbf{Z}^O$ is the best estimate of $\mathbf{Z}$ so far; it has been found in the previous iteration step. The matrix $\mathbf{S}_S$ is defined similar to the one in De Leeuw and Bijleveld (1988, p9). It is easy to verify that this problem is a Procrustes problem with the solution for $\mathbf{Z}$ given by the well known product of the left and right singular vectors of the matrix sum $\Sigma_S\ (\mathbf{Z}^O + \mathbf{S}_S)$.

To enhance the interpretability of (6.2) we could replace $\mathbf{F}_S'$ by $\mathbf{F}'\mathbf{W}_S$, where $\mathbf{W}_S$ is a diagonal weights matrix, which shows how each subject fits to the general time process represented by $\mathbf{F}$. This can be done by an extra step after the algorithm for the problem defined by (6.1) and (6.2) has converged. We want to minimize the following loss function:

$$\sigma(\mathbf{F},\mathbf{W}_s) = \Sigma_s \mathrm{SSQ}\ (\mathbf{F}_s' - \mathbf{F}'\mathbf{W}_s). \tag{6.7}$$

The loss can be partitioned into a summation over columns, as follows:

$$\sigma(\mathbf{F},\mathbf{W}_s) = \Sigma_j \Sigma_s \mathrm{SSQ}\ (\mathbf{f}_{js} - \mathbf{f}_j w_{js}). \tag{6.8}$$

Now the problem can be written in the form:

$$\sigma(\mathbf{F},\mathbf{W}_s) = \Sigma_j \mathrm{SSQ}\ (\mathbf{F}_j - \mathbf{f}_j \mathbf{w}_j'), \tag{6.9}$$

with the matrix $\mathbf{F}_j$ defined as $(\mathbf{f}_{j1},...,\mathbf{f}_{js},...,\mathbf{f}_{jN})'$, which is a matrix of order p by N and the vector $\mathbf{w}_j'$ as $(\mathbf{w}_{j1}, .. \mathbf{w}_{js} .., \mathbf{w}_{jN})'$. With the proper normalization the solution of (6.9) amounts to solving p independent rank-one decomposition problems, which are in fact similar to p PCA solution in one dimension on the matrix $\mathbf{F}_j$. The solutions for the different dimensions are independent as can be seen in (6.9).

7. Discussion

In the first three approaches there was one step in which the average of the variables was taken. Especially with nominal variables this averaging is a rather peculiar operation. However after an optimal scaling of these variables averaging is of course more meaningful. So in the "points of view" and generalized canonical correlation approach we do not have a real problem, because within these approaches the variables are optimally quantified in each step of the main loop.

In the first approach, however, we should introduce an extra step. After we have found the optimally scaled variables for each subject, a new average matrix is computed, now based on optimally scaled variables. This new matrix is analyzed by DYNAMALS. The process could be repeated until convergence.

The most simple way to do STADS is by concatenating all matrices to a two-way matrix of order (NxT) by m and by defining the shift matrix **B** in such a way that there is no dependence specified between the subjects. This solution yields one **F**, one **G** and one **H**, but N matrices **Z**. So we see that this simple approach is quite similar to the fourth extension. However, there are some important differences, which have to do with the normalization of the variables. When we analyze the concatenated matrix, we obtain the observed supervariables $\underline{\mathbf{x}}_i$ $\underline{\mathbf{y}}_i$ and the latent supervariables $\underline{\mathbf{z}}_i$, all having length (NxT). These supervariables are then normalized. This is different from the normalization that takes place in the fourth extension, where the variables of each subject are normalized separately. This difference is similar to what is called the *weak* and *strong* normalization in Gifi (1981). Further research is necessary to see whether these differences are substantial.

Another straightforward method to analyze this kind of three way data is to do N separate DYNAMALS analyses. We find parameter matrices and latent variables for each subject. This way we obtain very detailed information about the data, but it is difficult to discover relations between the subjects. So if the interest is mainly in the relations within each subject (system), this may be a sound way to follow. Note that this is just the opposite approach from the one we called the first extension in which all subjects were taken together.

8. References

Bijleveld, C.C.J.H. (1989). *Exploratory linear dynamic sytems analysis*. Leiden: DSWO-press.

De Leeuw, J. and Bijleveld, C.C.J.H. (1988). *Fitting longitudinal reduced rank regression models by alternating least squares*. RR-88-03. Leiden: Department of Data Theory.

Green, B.F. and Gower, J.C. (1979). *A problem with congruence*. Paper presented at the Annual Meeting of the Psychometric society, Monterey, California.

Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: Wiley.

Meulman, J.J. (1986). *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO-press.

Meulman, J.J. and Verboon, P. (1989). *Nonlinear Principal Component Analysis with distance restrictions*. Paper presented at the European Meeting of the Psychometric Society. Leuven, Belgium.

Ten Berge, J.M.F. and Knol, D.L. (1984). Orthogonal rotations to maximal agreement for two or more matrices of different column orders. *Psychometrika*, 49, 49-55.

Ten Berge, J.M.F. and Nevels, K. (1977). A general solution for Mosier's oblique Procrustes problem. *Psychometrika*, 42, 593-600.

Tucker, L.R. and Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika*, 28, 333-367.

Van Buuren, S. (1990). *Optimal scaling of time series*. Leiden: DSWO-press.

Van der Burg, E.,Verdegaal, R. and De Leeuw, J. (1988). Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 53, 177-197.

APPENDIX List of matrices and indices

$\mathbf{X}_s$	Set of input variables	$(T \times m_1)$
$\mathbf{Y}_s$	Set of output variables	$(T \times m_2)$
$\mathbf{Z}$	Set of latent variables	$(T \times p)$
$\mathbf{F}$	Transition matrix	$(p \times p)$
$\mathbf{G}$	parameter matrix	$(p \times m_1)$
$\mathbf{H}$	parameter matrix	$(m_2 \times p)$
$\mathbf{B}$	contiguity matrix	$(T \times T)$
$\mathbf{W}$ (section 3)	weights output part	$(N \times r)$
$\mathbf{V}$ (section 3)	weight input part	$(N \times r)$
$\mathbf{A}_s$ (section 4)	components loadings	$(m_2 \times p)$

m_1	number of input variables (j_1)
m_2	number of output variables (j_2)
m	total number of variables (j)
T	number of rows datamatrix (t)
N	number of subjects (s)
p	number of latent variables or dimensionality (i)
r (section 3)	number of points of view (k)