4. The CP and Tucker3 models with smoothness constraints

4.1. Introduction

In Chapter 3, the rationale for applying constraints to models for longitudinal data was introduced. In this context, imposing smoothness is a flexible type of constraint. In the present chapter, a method for applying smoothness constraints, possibly combined with monotonicity constraints, to the Tucker3 and CP models will be presented. The Tucker3 and CP models were introduced in Section 2.3. The smoothness constraints can easily be used in the Tucker2 and Tucker1 models as well, but this is not explicitly discussed here.

In the Tucker3 or CP models, smoothness constraints can be applied in two ways. One could smooth the raw data, that is the various univariate series per subject and variable. In doing so, one aims at (partly) eliminating measurement error from the data. Subsequently, an unconstrained Tucker3 or CP analysis can be performed on the smoothened data. Alternatively, one could constrain the occasion components in the Tucker3 or CP models to be smooth and thus fit a constrained model to the raw data. In this case, it is hoped that less measurement error is fitted by the model. In the case of growth data, it might be useful to combine smoothness constraints with monotonicity constraints. Fortunately, the question of which of the two approaches to take is trivialized in an important class of cases. As will be shown, fitting the smooth descriptions of the observed data by the unconstrained Tucker3 or CP models is equivalent to fitting the Tucker3 or CP models with smoothness constraints in a particular class of cases. The usefulness of smoothing in the Tucker3 and CP models will be examined under different conditions in a simulation study. An empirical example illustrates the use of smoothness constraints within the Tucker3 model.

4.2. The choice of a smoother

A smoother is used to describe a response measurement as a smooth function of one or more predictor measurements (Hastie & Tibshirani, 1990), usually by so-called local averaging. In our applications, the predictor will simply be the time point at which each measurement is made, to be denoted as \( t_k = t_1, \ldots, t_K \). Local averaging aims at averaging the observed measurements associated with predictor values close to each other (i.e., in each other’s neighborhood). The different types of smoothers differ mainly in their method of averaging. The size of the neighborhood influences the smoothness, and the accuracy: a large neighborhood leads to an estimate with low variance (i.e., high smoothness) but high potential bias (i.e., low accuracy), whereas
the opposite holds for small neighborhoods: there is a trade-off between bias and variance, as discussed by Hastie and Tibshirani (1990).

Hastie and Tibshirani (1990) and Ramsay and Silverman (1997) offer overviews of different smoothers and their properties. Polynomial regression splines, which form a class of smoothers that is computationally convenient, will be used here. Polynomial regression splines are constructed from different polynomial pieces, which are joined at certain predictor values, the knots. B-splines are a popular type of polynomial splines (De Boor, 1978). B-splines can easily be used for smoothing the data before analysis (see Alsberg & Kvalheim (1993) for an example involving three-way data). Monotonicity restrictions on the solutions can be useful in certain longitudinal applications. They can be imposed by using I-splines (Ramsay, 1988).

B-splines (Basis splines) are non-negative basis functions. The degree \( d \) of a B-spline is the degree of the polynomial pieces on which it is based, and any degree of polynomial can be chosen. Each B-spline is determined fully by its degree and by its knot sequence. The knots are positioned in the domain \( t_1 \) through \( t_K \) of the predictor. Given the degree and the location of the knots, B-splines can be computed by a recursive formula (De Boor, 1978). If they are of equal degree and they are positioned equidistantly, the basis functions are equal in size and shape. An example of a set of seven third degree B-splines, evaluated on the interval 1 to 2 is presented in Figure 4.1.

![Figure 4.1. Example of seven non-zero third degree B-splines, evaluated on the interval 1 to 2.](image)

In Figure 4.1, the five equidistantly placed knots are indicated by tick marks outside the x-axis. In general, the polynomial pieces join at \( d \) inner knots, and at these joining points, the derivative up to order \( d-1 \) is continuous. The number of non-zero B-splines \( N \) on the domain \( t_1 \) through \( t_K \) is equal to the total number of knots on this
domain plus the degree of the polynomial minus one. In the example, this boils down to $5+3-1=7$ non-zero B-splines in Figure 4.1. A B-spline is positive on a domain spanned by $d+2$ knots; everywhere else it is zero. The B-spline that is zero everywhere outside the interval 1 to 2 in Figure 4.1 is indicated by the bold line.

Usually, a set of response measurements collected in $y$ ($K \times 1$) is to be approximated by a linear combination of the B-splines that are evaluated in the values of the predictor $t$. Let $B^*$ denote a $K \times N$ B-spline matrix, in which the $n^{th}$ column contains the values of the $n^{th}$ B-spline that is evaluated for all values of the predictor $t$ ($K \times 1$), and where $K \geq N$; let $w$ ($N \times 1$) denote the vector with weights for the $N$ B-splines, and $\hat{y}$ the vector with estimated response measurements, which is called the smooth in the sequel, then

$$\hat{y} = B^*w. \quad (4.1)$$

Since B-splines are always non-negative, the estimated response measurement $\hat{y}$ can be restricted to be non-negative by restricting the weights $w$ to be non-negative.

I-splines (Integrated splines; Ramsay, 1988) are monotonically increasing basis functions. They are based on integrated M-splines, which are proportional to B-splines. Because M-splines are non-negative everywhere, the integrated M-splines are a natural basis for monotone splines. Since bases for I-splines are monotonically increasing, a non-negativity constraint on the set of coefficients of the I-splines leads to monotonically non-decreasing estimated response variables.

The use of B-splines and I-splines requires the selection of the ‘smoothing parameters’, that is the degree of the splines and the number and the position of the knots. The degree of the spline is commonly fixed. For B-splines, a popular choice is a third degree B-spline (Hastie & Tibshirani, 1990, p. 22); smoothers based on higher degree splines tend to oscillate wildly (Van Rijckevorsel, 1988). Ramsay (1988) claims that low (e.g., second) degree I-splines generally suffice. The number and location of knots influences the smooth: more knots in a region lead to a greater flexibility of estimation in that region, whereas fewer knots lead to a greater rigidity in that region. Using too many knots leads to overfitting, and using too few knots leads to underfitting. As to choosing the position of the knots, a simple approach is to position them uniformly over the domain $t_1$ through $t_K$. Another approach is to place them at appropriate quantiles of the predictor variable.

The smoothing parameters can be selected by subjective comparisons of several estimated response variables and the observed variable. Automatic selection methods for the smoothing parameters are also available (see Hastie & Tibshirani, 1990, pp. 42-52). Although the usefulness of these methods is debatable, they can be helpful in deciding on the number of knots. A commonly used procedure is cross-validation by means of the leave-one-out approach. Hastie and Tibshirani (1990, pp. 46-48) showed that the cross-validation sum of squares for linear smoothers can be computed by
where \( \lambda \) denotes the smoothing parameters (i.e., the degree, and the number and the positions of the knots), and \( S(\lambda)_{kk} \) are the diagonal elements of the projection-matrix \( S(\lambda) \), which relates \( \hat{y} \) to \( y \). To select the smoothing parameters, one may search for those that minimize \( CV(\lambda) \).

### 4.3. How to smooth in the Tucker3 model and CP model?

Smoothers are generally used to describe observed scores by a function. As discussed above, smoothing in the Tucker3 or CP models can be performed by smoothing the raw data before analysis by the unconstrained Tucker3 or CP model, or by constraining the component scores in the Tucker3 or CP models to be smooth and thus fitting a constrained model to the raw data. The latter approach is used in functional PCA by Ramsay and Silverman (1997, Ch. 7), by applying a roughness penalty to prevent the roughness of the estimated principal components from being too large. We prefer a different approach, which has the advantage that smoothing the raw data and smoothing the components lead to the same estimated model parameters. We propose imposing a smoothness constraint on the occasion component matrix \( C \) by constraining \( C \) \((K \times R)\) so that it can be written as \( B^sU \), for a B-spline matrix \( B^s \) \((K \times N)\) and a particular weight matrix \( U \) \((N \times R)\), and where \( N \geq R \). As a result, the Tucker3 and CP models with smoothness constraints on the occasion component matrix can be written as

\[
X_c = B^s U G_c (B^t \otimes A) + E_c,
\]

where \( X_c \) denotes the \( K \times IJ \) matricized data array \( X \), \( B^s \) \((K \times N, K \geq N)\) a B-spline matrix, \( U \) \((N \times R, N \geq R)\) a weight matrix, \( A \) \((I \times P)\) and \( B \) \((J \times Q)\) component matrices, \( G_c \) \((R \times PQ)\) the supermatrix containing the frontal slices of the core array \( G \) \((P \times Q \times R)\), and \( E_c \) \((K \times IJ)\) the matricized error array \( E \); in the case of the CP model with smoothness constraints the core array is fixed at superidentity. The B-spline bases are computed using ‘time’ as predictor. Note that the same basis is used for all components. In fact, formula (4.1) is used repeatedly for \( r=1,\ldots,R \) as \( c_r = B^s u_r \). If monotonicity restrictions are required, it is proposed that the B-splines basis matrix be replaced by an I-splines basis matrix, and that non-negativity constraints be imposed on the weights.

The Tucker3 and CP models with smoothness constraints, like their unconstrained counterparts, are fitted to data by minimizing the sum of squared residuals. Now, it will be shown that restricting the component matrix \( C \) to be in the column space of the B-spline matrix \( B^s \) in the Tucker3 or CP models is equivalent to analyzing the projection of the data matrix \( X_c \) on \( B^s \) by the unrestricted Tucker3 or CP
4.3. How to smooth in the Tucker3 model and CP model?

model, which in turn comes down to a Tucker3 or CP analysis of the B-spline smoothed data. To show this, we replace \( \mathbf{B}^s \) by the QR-factorization \( \mathbf{B}^s = \mathbf{Q} \mathbf{R} \), with \( \mathbf{Q} \) column-wise orthonormal, and \( \mathbf{R} \) a square upper triangular matrix. Note that since \( \mathbf{B}^s \) is of full column rank, \( \mathbf{R} \) is non-singular. Then, the function to be minimized is

\[
f_1(U, A, B, G_c) = \| \mathbf{X}_c - \mathbf{Q} \mathbf{R} \mathbf{G}_c (\mathbf{B}' \otimes \mathbf{A}') \|^2, \tag{4.4}
\]

with \( \mathbf{X}_c \) \((K \times IJ)\) the matricized data array \( \mathbf{X} \), and \( \mathbf{G}_c \) \((R \times PQ)\) the matricized core in the Tucker3 model, or the matricized superidentity array in the case of the CP model. As already noted by Carroll, Pruzansky and Kruskal (1980, p. 7), minimization of (4.4) is equivalent to minimizing

\[
f_2(\tilde{U}, A, B, G_c) = \| \mathbf{Q}' \mathbf{X}_c - \tilde{\mathbf{U}} \mathbf{G}_c (\mathbf{B}' \otimes \mathbf{A}') \|^2, \tag{4.5}
\]

with \( \tilde{\mathbf{U}} \) written for \( \mathbf{R} \mathbf{U} \).

It will now be shown that minimizing (4.5) is equivalent to analyzing the projection of \( \mathbf{X} \) on the B-spline matrix \( \mathbf{B}^s \) by the unrestricted Tucker3 or CP model. This in turn comes down to smoothing the data matrix \( \mathbf{X}_c \) by means of B-splines before Tucker3 or CP analysis, which is achieved by minimizing

\[
f_3(\mathbf{W}) = \| \mathbf{X}_c - \mathbf{B}^s \mathbf{W} \|^2. \tag{4.6}
\]

The optimal weights \( \mathbf{W} \) are given by \( (\mathbf{B}^s' \mathbf{B}^s)^{-1} \mathbf{B}^s \mathbf{X}_c \), hence the smooth of \( \mathbf{X}_c \) is

\[
\hat{\mathbf{X}}_c = \mathbf{B}' (\mathbf{B}^s' \mathbf{B}^s)^{-1} \mathbf{B}^s \mathbf{X}_c, \tag{4.7}
\]

the projection of \( \mathbf{X}_c \) on \( \mathbf{B}^s \). Analyzing this projection by Tucker3 or CP comes down to minimizing

\[
f_4(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G}_c) = \| \hat{\mathbf{X}}_c - \mathbf{C} \mathbf{G}_c (\mathbf{B}' \otimes \mathbf{A}') \|^2 = \| \mathbf{B}^s (\mathbf{B}^s' \mathbf{B}^s)^{-1} \mathbf{B}^s' \mathbf{X}_c - \mathbf{C} \mathbf{G}_c (\mathbf{B}' \otimes \mathbf{A}') \|^2. \tag{4.7}
\]

Let \( \mathbf{B}^s \) be replaced by the QR-factorization as \( \mathbf{B}^s = \mathbf{QR} \). Note that \( \mathbf{R} \) is nonsingular. Minimization of (4.7) comes down to minimizing

\[
f_5(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G}_c) = \| \mathbf{Q} \mathbf{Q}' \mathbf{X}_c - \mathbf{C} \mathbf{G}_c (\mathbf{B}' \otimes \mathbf{A}') \|^2. \tag{4.8}
\]

The optimal \( \mathbf{C} \) will be in the column space of \( \mathbf{Q} \), hence \( \mathbf{C} \) can be written as \( \mathbf{Q} \tilde{\mathbf{C}} \), and minimizing (4.8) is equivalent to minimizing
4. The CP and Tucker3 models with smoothness constraints

\[
f_5(A, B, \tilde{C}, G) = \left\| QQ' X_c - Q \tilde{C} G_c (B' \otimes A') \right\|^2
\]

\[
= \left\| Q (Q' X_c - \tilde{C} G_c (B' \otimes A')) \right\|^2
\]

\[
= \left\| Q' X_c - \tilde{C} G_c (B' \otimes A') \right\|^2,
\]

(Kiers & Harshman, 1997). Clearly, minimizing (4.9) is equivalent to minimizing (4.5); the solutions for \( A, B, \) and \( G_c \) of (4.5), and of (4.9) are equivalent; the solution for \( \tilde{U} \) in (4.5) is equivalent to that for \( \tilde{C} \) in (4.9). Because \( \tilde{U} \) leads to \( C \) by \( C= B' U = Q R U = Q \tilde{U} \), and \( \tilde{C} \) leads to \( C \) by \( C = Q \tilde{C} \), we see that both methods give the same solution for \( C \) as well. It has thus been shown that analyzing the original data by means of a smooth constrained Tucker3 or CP models is equivalent to analyzing smoothed data by the unconstrained Tucker3 or CP model, as long as smoothness is defined in terms of unrestricted linear combinations of B-splines.

The matrix \( Q' X_c \) in (4.9) is a compressed version of the smooth matrix \( B^t \left( B^s, B^s \right)^{-1} B^s' X_c \) (see Kiers & Harshman, 1997). Since the matrix \( Q' X_c \) is (much) smaller than the matrix \( \hat{X}_c \), minimization of (4.9) over \( A, B, \tilde{C}, \) and \( G_c \) can be considerably faster than minimization of (4.7) over \( A, B, C, \) and \( G_c \). On the other hand, the use of (unconstrained) splines on the data rather than on the components may be easier to handle, since standard software can be used to obtain the smooths, and subsequently analyze them by the Tucker3 or the CP model.

By constraining the B-spline or I-spline weights, one can impose a constraint on the solution in addition to the smoothness constraint. If I-spline weights are restricted to non-negativity, the smooth is monotonically increasing (and non-negative as well). The smooth can be restricted to be non-negative by requiring the B-splines weights to be non-negative. The problem of finding non-negative weights for the splines can be solved by treating the problem as a non-negative least squares problem (Lawson & Hanson; 1974, pp. 158-164; Bro & De Jong, 1997). Note that if spline weights are constrained, imposing a spline basis on a component matrix will have a different effect than imposing a spline basis on the data matrix. If a spline basis is imposed on a component matrix with constrained weights, we have to minimize (4.4) over \( U, A, B, \) and \( G_c, \) subject to appropriate constraints. If a spline basis is imposed on the data matrix with constrained weights \( W, \) (4.6) has to be minimized, subject to the appropriate constraints, before analyzing the restricted projection by Tucker3 or CP.

It is expected that applying smoothness constraints will help recover the underlying true curves when applying the Tucker3 or CP models to data with smooth underlying structure. In fitting the unconstrained CP model, local minima are often encountered when the underlying curves are smooth. In this case the associated component matrix often has a high degree of multicollinearity, which may cause the model to be ill-defined, and thus lead to local minimum problems. It is expected that
4.3. How to smooth in the Tucker3 model and CP model?

the algorithm to fit CP will land in a local minimum less frequently if smoothness constraints are used, especially in the case of high multicollinearity of the component matrices.

4.4. Comparing constrained with unconstrained CP and Tucker3 models

To test the usefulness of smoothing in the CP and Tucker3 models and compare it with the unrestricted CP and Tucker3 models, respectively, we performed a simulation study on the basis of 480 data sets for the CP model, and 960 data sets for the Tucker3 model. The algorithms were programmed in MATLAB5 (1998), and the analyses were carried out on a Pentium 333Mhz 32 Mb RAM personal computer in a Windows 95 environment.

4.4.1. Construction of the data for the simulation study

CP data for the simulation study

For the CP simulation study, 480 data sets were constructed with known CP structure with smooth components in one mode \( (C) \), and various data sizes, numbers of components, degrees of multicollinearity in \( A, B, \) and \( C \), and error levels. The data matrices \( X_c (K\timesIJ) \) were constructed according to

\[
X_c = C_o H (B_o \otimes A_o) + \epsilon N_c, \tag{4.10}
\]

where \( A_o (I\timesQ) \), \( B_o (J\timesQ) \), and \( C_o (K\timesQ) \) are ‘true’ component matrices for the respective three modes, \( H (Q\timesQ^2) \) is the matrix version of the superdiagonal three-way array \( \tilde{H} (Q\timesQ\timesQ) \), \( \epsilon \) is a scalar, and \( N_c (K\timesIJ) \) denotes the matrix expression of the three-way error array \( N (I\timesJ\timesK) \).

The data sizes \( I, J, K \) were 10, 10, 20; 10, 10, 50; 10, 50, 20 and 10, 50, 50. The numbers of components were two and four. The elements of the matrices \( A_o \) and \( B_o \) were drawn randomly from the uniform \([0,1]\) distribution (mild multicollinearity condition), and from the uniform \([0.5,1.5]\) distribution (severe multicollinearity condition). To ensure smoothness of the components of \( C_o \), every component of \( C_o \) followed a smooth function evaluated at \( K \) equidistant points (to be denoted by \( t_1, ..., t_K \)) in every condition. Half of the components of \( C_o \) followed an exponential function, and half of the components followed a logistic function, both of which are often employed as growth curves (Browne, 1993). The parameters were varied so that in the mild multicollinearity condition of \( C_o \) the condition numbers for two and four components of \( C_o \) were two and six, respectively, whereas in the severe multicollinearity condition the condition numbers were six and 42. The values of \( N_c \) were drawn randomly from the standard normal distribution and multiplied by a scalar \( \epsilon \) chosen so that the expected percentages of error in \( X \) were 2\%, 26\%, or 50\% (in terms of sums of squares). The number of replications was five. The design was
fully crossed, leading to a total of four (data sizes) × two (numbers of components) ×
two (degrees of multicollinearity of \( A_o \) and \( B_o \)) × two (degrees of multicollinearity of
\( C_o \)) × three (error levels) × five (replications) = 480 matrices.

**Tucker3 data for the simulation study**

For the Tucker3 simulation study, 960 data sets were constructed with known
Tucker3 model structure with smooth components in one mode (\( C \)), and various data
sizes, numbers of components, degrees of multicollinearity in the core, and error
levels. The data matrices were constructed as

\[
X_c = C_o G_o (B_o \otimes A_o) + e N_c, \tag{4.11}
\]

where \( A_o (I \times P) \), \( B_o (J \times Q) \), and \( C_o (K \times R) \) are ‘true’ component matrices for the three
modes, \( G_o (R \times PQ) \) is the matricized version of the three-way core array \( G_o \) (the
subscript ‘\( c \)’ is omitted for notational simplicity), \( e \) is a scalar, and \( N_c (K \times IJ) \)
denotes the matrix expression of the three-way error array \( N \).

The sizes of the data array \( X_{I,J,K} \) were 10,10,20; 10,20,20; 10,10,50; 10,50,20;
10,20,50; 30,20,20; 10,50,50 and 30,20,50. The numbers of components \( P,Q,R \) for the
three modes were 2,2,2; 2,4,2; 2,2,4 and 4,4,4. The component matrices \( A_o, B_o, \) and
\( C_o \) were chosen column-wise orthonormal. The components of the smooth \( C_o \)
followed the same functions as in the CP simulation study, but now the orthonormal
bases of the matrices used in the CP simulation study were used. The matrices \( A_o \) and
\( B_o \) were obtained by taking the orthonormal bases of a matrix with equal size as \( A_o \)
and \( B_o \) with elements drawn randomly from the uniform \([0,1]\) distribution. These
choices do not place severe limitations on the simulation study, since the component
matrices in the Tucker3 solution, and hence any set of ‘true’ component matrices of a
Tucker3 model in a simulation study, can be transformed to orthonormality, provided
that this transformation is compensated in the core. However, transformation of a
multicollinear true component matrix to orthonormality and compensation for this in
the core array would lead to a multicollinear core. For example, suppose we have a
matrix \( C \) and \( G \), where \( \text{cond}(C) = 100 \), and \( G \) is row-wise orthonormal so that \( \text{cond}(G) \)
is 1, where \( \text{cond}(\ ) \) means the condition number. Orthonormalization of \( C \) into \( \tilde{C} \), and
compensation for the orthonormalization in \( G \) by transforming \( G \) into \( \tilde{G} \) results in
\( \text{cond}(\tilde{C}) = 1 \). This can be achieved by taking the QR-decomposition of \( C = QR \),
defining \( \tilde{C} = QCR^{-1} \), and \( \tilde{G} = RG \), and, as a result, \( \text{cond}(\tilde{G}) = \text{cond}(RG) = 100 \).
Therefore, to represent a reasonable range of possible data matrices, the degree of
multicollinearity of the core is varied in this study. The elements of \( G_o \) were drawn
randomly from the uniform \([0,1]\) distribution in the low multicollinearity condition,
and from the uniform \([0.5,1.5]\) distribution in the high multicollinearity condition. The
error level was varied in the same way as in the CP simulation study, that is the
expected percentages of error sum of squares of \( X \) were 2%, 26%, and 50%. The
number of replications in each condition was five. The design was fully crossed, leading to a total of eight (data sizes) × four (numbers of components) × two (degrees of multicollinearity of $G_o$) × three (error levels) × five (replications) = 960 matrices.

4.4.2. Analyses of simulation data

The simulated data sets $X_c$ were all analyzed by one unconstrained CP or Tucker3 analysis, and by two CP or Tucker3 analyses with smoothness constraints. Specifically, in the analyses with smoothness constraints, the estimated component matrix $C$ was restricted to be in the column space of a set of B-splines $B_s$ of degree three. The knots were equidistantly placed on the time interval $t_1,...,t_K$, with a knot at $t_1$ and one at $t_K$. The CP or Tucker3 analyses with smoothness constraints were performed on the compressed data array (see (4.9)) instead of the full data array to reduce computation time. In one of the analyses with smoothness constraints, the numbers of knots were chosen so that the sum of the cross-validation sum of squares, $CV(\lambda)$, see (4.2), over columns of $X_c$ was minimized. That is, for a fixed number of knots, the $CV(\lambda)$ was computed for each column of $X_c$, and then the sum of the $CV(\lambda)$’s obtained in this way was computed. The sum of the $CV(\lambda)$’s was computed successively for solutions based on 2,3,...,$t_K$ knots, and the number of knots that goes with the minimal sum of $CV(\lambda)$’s was chosen. The CP and Tucker3 analyses with these restrictions are referred to as CP-Bs(CV) and Tucker3-Bs(CV), respectively. In the other analysis with smoothness constraints, the number of knots of the B-splines was fixed at three. A small number of knots was chosen, since it is known that using too many knots leads to overfitting. However, the number three was somewhat arbitrary. The CP and Tucker3 analysis with this restriction are referred to as CP-Bs(3) and Tucker3-Bs(3), respectively. The estimates of CP were obtained by means of the CP algorithm of Harshman (1970), and Carroll and Chang (1970). Each Tucker3 analysis was performed using the efficient algorithm by Andersson and Bro (1998). In each analysis, the CP and Tucker3 algorithms were run from five different starts, one started rationally and four randomly, to reduce the chance of missing the global minimum. The rationally started runs were started with the parameters resulting from Tucker’s Method I (Tucker, 1966a). The convergence criterion was set at $10^{-6}$.

4.4.3. Criteria of interest

The main interest in this study was to determine how well the original component matrices (and core matrix in the case of the Tucker3 model) were recovered by each of the methods. Since the CP analysis yields unique estimates of the component matrices (up to permutation and scaling), while the solution of the Tucker3 analysis is not uniquely defined, different comparison criteria are used for the CP and the Tucker3 analyses.
CP analyses: criteria of interest

In the analysis of the CP data, a comparison of the estimated component matrices \( \hat{A}, \hat{B}, \) and \( \hat{C} \) and the original component matrices \( A_o, B_o, \) and \( C_o \) has to take into account possible permutations, rescalings, and sign reversions of the estimated component matrices. Following Kiers (1998a), and Mitchell and Burdick (1994), we compared the CP solutions by computing the cosines between the tensor products \( a_r^o \otimes b_r^o \otimes c_r^o, \ r=1,...,R, \) for the original component matrices and \( \hat{a}_r \otimes \hat{b}_r \otimes \hat{c}_r, \ r=1,...,R, \) for the estimated component matrices, where the subscript \( r \) denotes the \( r^{th} \) column of the matrix at hand. Given a data array \( \mathbf{X} \) that is represented by a set of \( R \) tensor products of components, which are collected in component matrices \( A, B, \) and \( C, \) other sets of component matrices that yield the same representation of \( \mathbf{X} \) are composed of the same such tensor products, although possibly in a different order. Therefore, a useful comparison measure of the original and the estimated component matrices is the mean of the \( R \) cosines between the tensor products of the original components and the tensor products of the estimated components, with the latter tensor products ordered so that they lead to the highest mean of cosines. The cosines are known as Tucker’s coefficient of congruence (Tucker, 1951). The coefficient of congruence between two columns \( x \) and \( y, \) \( \phi_{xy}, \) is defined as the normalized inner product between the columns \( x \) and \( y, \) namely as

\[
\phi_{xy} = \frac{\mathbf{x}' \mathbf{y}}{\sqrt{\mathbf{x}' \mathbf{x} \mathbf{y}' \mathbf{y}}}. \tag{4.12}
\]

One rationally started and four randomly started runs of the CP analysis were carried out. The runs which led to a sub-optimal solution (defined here as a solution with a function value higher than 1.001 times the fit of the optimal solution, out of the five runs) were counted to get an impression of the sensitivity to local minima of the constrained and unconstrained analyses.

Tucker3 analyses: criteria of interest

To investigate how well the original matrices of the Tucker3 model are recovered, two aspects are of importance, namely the recovery of the column spaces of the component matrices, and the recovery of the weights of the interactions of the components. The column spaces of the component matrices are compared as follows: A comparison of the component matrices \( \hat{A}, \hat{B}, \) and \( \hat{C} \) as estimated by Tucker3 to the underlying \( A_o, B_o, \) and \( C_o \) has to take into account the fact that \( \hat{A}, \hat{B}, \) and \( \hat{C} \) can be transformed without loss of fit, provided that such transformations are compensated in the core. Therefore, the estimates of \( \hat{A}, \hat{B}, \) and \( \hat{C} \) are transformed towards \( A_o, B_o, \) and \( C_o, \) respectively, by postmultiplying \( \hat{A}, \hat{B}, \) and \( \hat{C} \) by the
matrices $S$, $T$, and $V$, respectively. The transformation matrices $S$, $T$, and $V$ are found by minimizing the Euclidean distance between the original component matrices $A_o$, $B_o$, and $C_o$ and the transformed component matrices $\hat{A}_S$, $\hat{B}_T$, and $\hat{C}_V$, respectively. The transformations are compensated in the estimated core matrix $\tilde{G}$ by computing the transformed core array $\tilde{G} = V^{-1} \hat{G} ((T^*)^{-1} \otimes (S^*)^{-1})$. The component matrices $\hat{A}_S$, $\hat{B}_T$, and $\hat{C}_V$ are compared to the original component matrices $A_o$, $B_o$, and $C_o$ by computing the Proportion of Agreement ($PA_A$, $PA_B$, and $PA_C$, respectively) as

$$ PA_A = 1 - \frac{\|A_o - \hat{A}_S\|^2}{\|A_o\|^2} ; \quad PA_B = 1 - \frac{\|B_o - \hat{B}_T\|^2}{\|B_o\|^2} ; \quad PA_C = 1 - \frac{\|C_o - \hat{C}_V\|^2}{\|C_o\|^2}. $$

To be able to compare the Proportion of Agreement of the component matrices simultaneously, the average of $PA_A$, $PA_B$, and $PA_C$, denoted as $PA_{ABC}$, is used as the measure of agreement between the original and the estimated component matrices.

The recovery of the weights of the interactions of the components is examined by comparing the transformed core matrix $\tilde{G}$ to the original core matrix $G_o$ by computing the Proportion of Agreement ($PA_G$) as

$$ PA_G = 1 - \frac{\|G_o - \tilde{G}\|^2}{\|G_o\|^2}. $$

Note that the transformed component matrices are optimally transformed towards the original component matrices, whereas the associated core matrix is not optimally transformed towards the original core matrix. Hence, it can be expected that the $PA_G$ is smaller than the $PA_{ABC}$ in the case of a Tucker3 solution deviating from the original matrices.

4.4.4. Results of the simulation studies

Results of the CP simulation study

The original component matrices and the estimated component matrices, as obtained by unconstrained CP analysis (CP) and CP with smoothness constraints (CP-Bs(CV) and CP-Bs(3)), are compared by inspecting the $\phi$-values. The $\phi$-values have a negatively skewed distribution over the replications within each condition. The median $\phi$-values of the three analysis methods are plotted overall as well as per main condition in Figure 4.2.
The CP and Tucker3 models with smoothness constraints

Figure 4.2. Median $\phi$-values of CP, CP-Bs(CV) and CP-Bs(3) per condition. ‘Multic.’ denotes ‘multicollinearity’.

The following observations can be made in Figure 4.2. The median coefficient of congruence of the constrained CP solutions is larger than the median coefficient of congruence of the unconstrained CP solutions, whereas virtually no difference was found between the median coefficients of congruence of CP-Bs(CV) and CP-Bs(3). Furthermore, the difference between the unconstrained and the constrained CP solutions gets clearly larger with increasing condition numbers of $C_o$, and with increasing error level, and varies in a more complicated manner with data size (see Figure 4.2).

A repeated measurement ANOVA was performed to test whether the observed effects of type of analysis and of the interactions of analysis method with the various manipulated factors could be distinguished from random fluctuations. To correct for the deviation from normality for the repeated measurement ANOVA, the $\phi$-values were transformed into $\tilde{\phi} = \log(\phi/(1-\phi))$ before analysis, where the two observed negative $\phi$-values were excluded from the analysis. The transformation of negatively skewed $\phi$-values on the interval $[0,1]$ results in approximately normally distributed $\tilde{\phi}$-values on the interval $[-\infty, \infty]$. The effects which were described in the previous paragraph, were all found to be significant at $\alpha=0.001$ in the repeated measurement ANOVA of the $\tilde{\phi}$-values.
In addition to the coefficients of congruence of the three analysis methods in the different conditions, the number of cases in which the unconstrained CP leads to a ‘good’ solution, and the constrained CP to a ‘bad’ solution is of interest. On the basis of inspection of a number of plots of original and estimated components and the accompanying coefficient of congruence, solutions with a coefficient of congruence smaller than 0.75 were considered to be bad. The resulting frequencies according to this criterion are presented in Table 4.1.

Table 4.1. Frequencies of good ($\varphi \geq 0.75$) and bad ($\varphi < 0.75$) solutions per analysis method (CP with CP-Bs(CV) and CP with CP-Bs(3)).

<table>
<thead>
<tr>
<th></th>
<th>CP-Bs(CV)</th>
<th>CP-Bs(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>good</td>
<td>bad</td>
</tr>
<tr>
<td>CP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>good</td>
<td>348</td>
<td>3</td>
</tr>
<tr>
<td>bad</td>
<td>43</td>
<td>86</td>
</tr>
</tbody>
</table>

In a large number of cases, both the constrained and the unconstrained CP model lead to a good solution. The most important finding is that if the unconstrained CP leads to a bad solution, the constrained CP model leads to a good solution in about 33% of the cases. Furthermore, it is rarely found that the unconstrained CP model leads to a good solution and the constrained CP model to a bad solution. The proportion of bad solutions, as well as the differences between the constrained and unconstrained CP model increases with error level, and condition number of $C_0$. Thus, on the basis of these results we can conclude that, if there is a smooth underlying structure, B-spline constrained CP is helpful in a fair number of cases, and that there is very little risk in replacing unconstrained CP by CP with smoothness constraints. Moreover, the choice for the number of knots does not seem crucial.

Differences between the constrained and unconstrained CP analyses in sensitivity to local minima were also studied. The constrained CP analyses led to a sub-optimal solution a little less frequently (both 0.10 out of five starts on average) than the unconstrained analyses (0.16 on average). No difference in average local minima has been found between the rationally and the randomly started runs. The number of local minima increased with increasing error level, whereas no substantial interaction between any other of the manipulated factors and type of analysis was found.

Results of the Tucker3 simulation study

The original component matrices and the original core matrix were compared to the estimated component matrices and the estimated core matrix by means of the Proportion of Agreement of the component matrices and the core matrix, the $\text{PA}_{ABC}$, which is based on the average of the expressions in (4.13), and the $\text{PA}_{G}$, (4.14), respectively. The average $\text{PA}_{ABC}$ values per analysis method give a good impression of the condition effects, and they are plotted per condition in Figure 4.3.
4. The CP and Tucker3 models with smoothness constraints

Figure 4.3. Mean PAABC of Tucker3, Tucker3-Bs(CV) and Tucker3-Bs(3) per condition. ‘Multic.’ denotes ‘multicollinearity’.

As can be seen in Figure 4.3, the PAABC of Tucker3-Bs(CV) is generally higher than that of Tucker3 and Tucker3-Bs(3), whereas almost no difference was found between the PAABC of Tucker3 and Tucker3-Bs(3), over all conditions. The difference in PAABC between the three methods of analysis increases with increasing core size, error percentage, and degree of multicollinearity of the core, and varies with data size. The gain of the smooth Tucker3 over the unconstrained Tucker3 is largest in the case of a relatively large size of the smooth mode, and relatively small sizes of the non-smooth modes, for example, data size 10,10,50. If the size of the smooth mode is smaller than the size of one of the non-smooth modes, the performance of Tucker3 is better than the smooth Tucker3’s (e.g., data sizes 30,20,20 and 10,50,20). Tucker3 clearly outperforms Tucker3-Bs(3) in the case of low error level (2%) and low multicollinearity of the core, whereas Tucker3-Bs(CV) performs best of the three. In high error level and high multicollinearity conditions, Tucker3-Bs(3) performs better than Tucker3, but Tucker3-Bs(CV) gives best recovery of the component matrices. This finding suggests that the smoothness restricted Tucker3 is sensitive to the choice of number of knots, and that in ‘easy conditions’ an unconstrained Tucker3 model performs even better than a smoothness constrained Tucker3 model with a non-optimal number of knots.

A repeated measurement ANOVA was performed to test whether the observed effects of type of analysis and of the interactions of analysis method with the various
manipulated factors could be distinguished from random fluctuations. For the repeated measurement ANOVA, the PA\textsubscript{ABC}-values were transformed to correct for the observed heterogeneity of variances for the groups by computing $\tilde{P}\tilde{A}_{ABC} = \arcsin(\text{PA}_{ABC})^{\frac{1}{2}}$ (Stevens, 1992). The effects that were explicitly described in the previous paragraph, were found to be significant at $\alpha=0.001$ in the repeated measurement ANOVA of the $\tilde{P}\tilde{A}_{ABC}$-values.

The estimated component matrices $\hat{A}$, $\hat{B}$, and $\hat{C}$ are optimally transformed to the original component matrices $A_o$, $B_o$, and $C_o$, whereas the transformation of the estimated core matrix $\hat{G}$ is so that the transformations of the original component matrices are compensated. Therefore, a non-optimal recovery will be expressed in a low PA\textsubscript{G} value, and possibly in a low PA\textsubscript{ABC} value. The PA\textsubscript{G} values appear highly negatively skewed, with some extremely low values, hence the median PA\textsubscript{G} values give a better insight into the condition effects than the mean PA\textsubscript{G} values. The median PA\textsubscript{G} values per condition appeared to be high (>0.985), and they hardly differ from each other, neither between type of analysis nor between conditions. The extremely low values all occurred in the ‘more difficult’ conditions, namely large core size, small data size, high condition number of the core $G_o$, and high error level. The Tucker3 analysis showed more extremely low PA\textsubscript{G} values than the Tucker3-Bs(CV) and Tucker3-Bs(3) analyses, as is indicated by, for example, the percentages of the cases with PA\textsubscript{G} values lower than 0.5 of 4.7%, 1.1% and 1.3%, respectively.

A second way of comparing the achievement of the three methods of analysis is to inspect the number of cases that are recovered well by the different methods. On the basis of inspection of a number of original and estimated components, cores and associated PA\textsubscript{ABC} and PA\textsubscript{G} solutions with a PA\textsubscript{ABC} or a PA\textsubscript{G} smaller than 0.9 were considered to be bad. The resulting frequencies of good and bad solutions are presented in Table 4.2. It can be seen in this table that if Tucker3 leads to a bad solution, Tucker3-Bs(CV) leads to a good solution in 44% of the cases. In only 1% of the cases, the Tucker3-Bs(CV) is bad, while the Tucker3 solution is good. According to the frequencies in Table 4.2, Tucker3-Bs(3) performs almost as well as Tucker3-Bs(CV).

<table>
<thead>
<tr>
<th></th>
<th>Tucker3-Bs(CV)</th>
<th>Tucker3-Bs(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>good</td>
<td>bad</td>
</tr>
<tr>
<td>Tucker3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>good</td>
<td>811</td>
<td>7</td>
</tr>
<tr>
<td>bad</td>
<td>63</td>
<td>79</td>
</tr>
</tbody>
</table>

Although the number of solutions that were reasonably recovered by Tucker3-Bs(3) does not deviate much from the number of reasonable recoveries using Tucker3-Bs(CV), the smoothing technique is sensitive to the choice for the number of knots, as
indicated by the better recovery of the underlying component structure by Tucker3-
Bs(CV) than of Tucker3-Bs(3). Thus, we can conclude on the basis of these results,
that if there is a smooth underlying structure, a smoothness constrained Tucker3
model is helpful in a reasonable number of cases, and that, conversely, there is very
little risk in using smoothness constrained instead of unconstrained Tucker3.
Tucker3-Bs(CV), the method with optimal knot selection, performed best, and is
therefore preferable to Tucker3-Bs(3).

4.5. Empirical example: Learning to read study (I)
In this section, an empirical example is presented to illustrate the use of smoothness
constraints in the Tucker3 model, and the use of monotonicity constraints on the data
before analyzing the data by the Tucker3 model. The ‘Learning to read study’ (Jansen
& Bus, 1982; Bus & Kroonenberg, 1982) investigates the learning process of reading.
Seven pupils were tested weekly (except for holidays) on 37 occasions using five
different tests intended to measure different aspects of reading ability. The primary
research questions focused on whether the performance of the pupils per test and over
tests was equal over time. This data set has been analyzed using a Tucker3 model by
Kroonenberg (1983). However, mainly because our preprocessing procedure is
different from the one applied by Kroonenberg, the results are not directly
comparable.

The data elements \( x_{ijk} \), where \( i = 1, \ldots, 7 \) denote the subjects, \( j = 1, \ldots, 5 \) the variables,
and \( k = 1, \ldots, 37 \) the occasions, were collected in the data array \( X \). Since the tests have
different score ranges, the data array \( X \) was rescaled so that the scores of all tests
ranged from 0 to 1. In this way, all the differences in variation were maintained in the
data, while the test scores were comparable. The data have a meaningful zero point,
as a score of zero on a test indicates that the pupil has not mastered the corresponding
ability at all. We consider the data to be approximately ratio scale. We therefore did
not center the data. The rescaled scores were collected in the data array \( Y \) and first
analyzed by the unconstrained Tucker3 model.

The scores are viewed as evaluations of growth curves, which are assumed to
follow some smooth curves in the course of time. Therefore, in the second analysis,
the components of the occasion mode \( (C) \) are constrained to be smooth, and this
analysis will be referred to as ‘T3-Bs’. A smoothness constrained Tucker3 model is
fitted to \( Y_c \) by minimizing (4.9), which is equivalent to minimizing (4.4). The degree
of the B-spline was fixed at three. The knots were placed equidistantly, and the
number of knots was chosen so that the sum of cross-validation sum of squares
\( (C V(\lambda); \text{see (4.2)}) \) of the columns of \( Y_c \) was minimized, by computing the sum of
\( C V(\lambda) \)'s related to B-splines with 2, 3, ..., 10 knots, and choosing the number of knots
that goes with the minimal sum of \( C V(\lambda) \)'s.

Because the data pertain to learning data, it seems reasonable to assume that the
true scores per variable per subject are non-decreasing on subsequent occasions, that
is that the reading ability of the child never decreases in the course of time. To model non-decreasing true scores, a smoothed data matrix $\tilde{Y}_e = B^i W$ is obtained by minimizing $\left\| Y_e - B^i W \right\|^2$ over $W$, where $B^i$ is a fixed I-spline matrix and $W$ the matrix of weights for the I-splines that are restricted to non-negativity. As a result, $\tilde{Y}_e$ is restricted in the sense that $\tilde{y}_{ijk} \leq \tilde{y}_{ijk(k+1)}$ for all $i=1,...,I; j=1,...,J$, and $k=1,...,K-1$.

An unconstrained Tucker3 analysis is applied to the smoothed data. This third analysis of the ‘Learning to read data’ will be referred to as ‘T3-Bi’. The degree of the I-spline matrix was fixed at two. The number of knots was selected by subjective comparison of the observed variables and several estimated response variables.

In all three analyses, the numbers of components were chosen to be (2,1,2) for the subject, variable and occasion modes, respectively. As will be discussed below, the models with this number of components fits the data well, and is interpretable, parsimonious, and stable.

The fit of the unconstrained Tucker3 model applied to $Y$ is 96.26%. The estimated core matrix $G_a$ of the model positioned in principal axes orientation was diagonal. The core matrix was transformed to identity (which can always be done in case $P=QR$ (Murakami, Ten Berge & Kiers, 1998, p. 256)), and this rescaling was compensated in the subject component matrix. The columns of the component matrices were rescaled so that the solution was easy to interpret: the maximum values of the second subject component, of the variable component, and of the first occasion component were rescaled to 1, and these rescalings were compensated in the first subject component and the second occasion component. The component matrices for the subjects ($A$) and the variables ($B$) of the unconstrained Tucker3 analysis of $Y$ are presented in Tables 4.4 and 4.5, respectively. The occasion component scores are plotted in Figure 4.4.

### Table 4.3. Subject component scores of the unconstrained Tucker3 solution.

<table>
<thead>
<tr>
<th>A (subjects)</th>
<th>1$^\text{st}$ component</th>
<th>2$^\text{nd}$ component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.06</td>
<td>-0.42</td>
</tr>
<tr>
<td>2</td>
<td>0.96</td>
<td>-0.30</td>
</tr>
<tr>
<td>3</td>
<td>0.99</td>
<td>-0.38</td>
</tr>
<tr>
<td>4</td>
<td>1.28</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>1.16</td>
<td>0.19</td>
</tr>
<tr>
<td>6</td>
<td>1.09</td>
<td>-0.01</td>
</tr>
<tr>
<td>7</td>
<td>0.89</td>
<td>-0.42</td>
</tr>
</tbody>
</table>
Table 4.4. Variable component scores of the unconstrained Tucker3 solution.

<table>
<thead>
<tr>
<th>B (variables)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter Knowledge</td>
<td>0.91</td>
</tr>
<tr>
<td>Regular Orthographic Short Words</td>
<td>1.00</td>
</tr>
<tr>
<td>Regular Orthographic Long Words</td>
<td>0.87</td>
</tr>
<tr>
<td>Regular Orthographic Long and Short Words within Context</td>
<td>0.99</td>
</tr>
<tr>
<td>Irregular Orthographic Long and Short Words</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Figure 4.4. Occasion component scores resulting from the unconstrained Tucker3 model; c(1) denotes the scores on the first component, c(2) the scores on the second component.

To interpret the component scores, we start with the occasion components. In Figure 4.4, it can be seen that the scores on the first component (c(1) in the figure) gradually increase from week 3 to week 20, and then levels off to an asymptote of one. The scores on the second component show a steady increase to week 10, a steady decrease from week 10 to 20, and then level off to slightly below zero. We would interpret the first component as indicating general performance level, and the second component as approximately reflecting learning rate. The latter component is not entirely interpretable as learning rate, because of the negative component scores, which are due to the estimated model parameters, and do not indicate that the performance decreases in the end.

The core matrix is identity, implying that the first component of the subject component matrix (A) is only related to the first component of the occasion component matrix (C), and the same holds for the second components of A and C.

Now, the subject component matrices can be interpreted. Recall that the general performance level of a subject (thus apart from specific variable effects) is a weighted sum of the two occasion components, which reflect general performance level and learning rate. A relatively high score on the first component means that the subject
concerned performs above the general performance level. A relatively high score on the second component implies that the subject shows a relatively fast growth in the learning. Hence, for example, Subject 4 performs by far best, as (s)he has a high performance level and a high learning rate. Subject 4 is followed at some distance by Subject 5, as (s)he has second position for performance level, and learning rate. Subjects 1 and 6 show approximately the same weighting for general performance level, but Subject 1 has a lower weight for the second component. Hence, their asymptote scores are more or less equal, but Subject 1 develops much more slowly than Subject 6. The performance order between Subjects 1 to 3 is somewhat difficult to see at once, as the performance level of a subject is a weighted sum of the two occasion components, and the weights are close to each other. One could plot the weighted sum of the occasion component scores for the three subjects (hence, rows 1, 2 and 3 of $\text{AG}_aC'$ in this particular case), and this would reveal that Subject 1 performs best of the three, and Subject 2 worst.

As there is only one variable component, the relative sizes of the variable component scores denote the difficulties of the items. The variable ‘Irregular Orthographic Long and Short words’ is by far the most difficult variable, as indicated by the lowest variable component score. Hence, the scores on ‘Irregular Orthographic Long and Short Words’ develop slowly in the course of time, compared to the other variables. The variable component scores of the ‘Regular Orthographic Short Words’ and the ‘Regular Orthographic Long and Short Words within Context’ are the highest variable component scores, showing that these scores develop fastest in the course of time. The variable component score of the ‘Letter Knowledge’ is slightly larger, and thus develops slightly faster, than the ‘Regular Orthographic Long Words’.

The stability of the model just discussed was investigated via a split-half analysis, following the guidelines by Kiers and Van Mechelen (2001). That is, the data were split in two halves over the subject mode, resulting in one data set of three subjects, and one of four subjects, to be denoted as $Y_1$ and $Y_2$. A Tucker3 analysis was performed for each of the two data sets. The solutions for $B$ and $C$ for each of the data sets were optimally transformed (in the least squares sense) to the solutions for $B$ and $C$ of the full data set (as presented in Table 4.4 and Figure 4.4, respectively). The two transformed occasion component matrices obtained in this way were compared by computing the coefficients of congruence (Tucker, 1951; see also Section 4.4.3) between the columns of the matrices. The two variable component matrices were compared analogously. The subject component matrices were compared as follows: The two solutions for $A$ for each of the splits were collected in one matrix $A_{12}$, in which the rows pertain to the same subjects as in $A$ of the full data set (as presented in Table 4.3). The matrix $A_{12}$ was regressed on $A$, and the resulting transformed component matrix was compared to $A$ by computing the coefficients of congruence between the two columns of the matrices. For each of the splits, the transformations of $A$, $B$ and $C$ were compensated in the core array. The separate split-half core arrays and the full data set core array were compared by computing the mean absolute difference between the split-half core array and the full data set core array.
4. The CP and Tucker3 models with smoothness constraints

The split-half procedure was repeated for every possible combination of the seven subjects split into two groups of three and four subjects, resulting in 35 split-half analyses. The mean coefficients of congruence for the subject component matrices over the 35 analyses was 1.000 and 0.997. The mean coefficient of congruence for the variable components was 0.998. The mean coefficient of congruence for the occasion components was 0.996 and 0.713 for the first and second occasion component, respectively. This implies that the stability of the subject components, the variable component and the first occasion component is high, whereas the stability of the second occasion component is moderate. The mean absolute difference between the split-half core array and the full data set core array averaged over the 35 analyses was 0.000. On the basis of these results, we conclude that the current Tucker3 model is sufficiently stable, given the small sample size at hand. We now turn to the results of the smoothness constrained analyses (T3-Bs and the T3-Bi), which will be discussed successively.

In the smooth Tucker3 analysis with B-splines, the T3-Bs, an unconstrained Tucker3 analysis was performed on the smoothed data array. The number of knots as indicated by the (minimal) cross-validation sum of squares was five. (The CV values for two through seven knots were 0.0096, 0.0089, 0.0089, 0.0087, 0.0089, 0.0097, respectively.) The fit of this constrained model to the data array $\mathbf{Y}$ was 96.18%, which is only 0.08% less than the fit of unconstrained Tucker3 model.

Just as in the unconstrained Tucker3 model, the estimated core matrix of the solution in principal axes orientation was diagonal, and was rescaled to identity. The component score matrices of the T3-Bs model were rescaled in the same way as was done with the unconstrained Tucker3 model. The estimated component matrices $\mathbf{A}$ and $\mathbf{B}$ of T3-Bs are compared to the solutions of the unconstrained Tucker3 model by computing the coefficient of congruence between the pairs of components concerned. This coefficient was large (>0.999) for all pairs, and therefore the solutions of $\mathbf{A}$ and $\mathbf{B}$ for T3-Bs can be interpreted in the same way as the corresponding solutions for the unconstrained Tucker3. The component scores for the occasions for T3-Bs are plotted in Figure 4.5.
4.5. Empirical example: Learning to read study (I)

Figure 4.5. Occasion component scores of the T3-Bs model, where c(1) denotes the scores on the first component, c(2) the scores on the second component.

Not surprisingly, the component scores of the occasions follow more or less the same curve as in Figure 4.4. However, the wiggles have disappeared, and the overall trend in the component scores of the occasions is more clear.

It is interesting to investigate whether the stability of the occasion component matrices of the T3-Bs has been improved on the unconstrained Tucker3 model. Additionally, it is important to check whether the subject and variable component matrices, and the core array of the T3-Bs model have a high stability, just as their counterparts in the Tucker3 model. The stability of the T3-Bs model was investigated using the split-half procedure, as discussed before for the unconstrained Tucker3 model, where the procedure was applied to the smoothed data. The mean coefficients of congruence of the occasion component matrices were 0.998 and 0.792, which is higher than the coefficients of congruence found in the unconstrained analyses of 0.996 and 0.713, respectively. The mean coefficients of congruence for the subject component matrices and the variable component matrices were equal to the ones found for the unconstrained Tucker3 model (1.000 and 0.997 for the subject component matrices, and 0.998 for the variable component matrices). Also, the mean absolute difference between the split-half core array and the full data set core array averaged over the 35 analyses was 0.000. On the basis of this results, one can conclude that the stability of the second occasion component of the T3-Bs model has indeed been improved somewhat compared to the unconstrained counterpart. The subject and variable component matrices, the first occasion component and the core array of the T3-Bs model are highly stable, just as their unconstrained counterparts.

In the T3-Bi analysis, the subsequent scores per variable and per subject were restricted to be non-decreasing in the course of time before analysis. The number of knots for the second degree I-spline matrix was chosen to be seven, on the basis of subjective comparison of the observed variables and several estimated response variables. The I-splines were defined on the interval from week 0 to 50. The fit of the
resulting estimates of the Tucker3 model to the unconstrained data array $Y$ was 96.17%. Just as in the unconstrained Tucker3 and the Tucker3-Bs models, the core matrix of the solution in principal axis direction was diagonal. The estimated core and component matrices were rescaled in the same way as in the Tucker3 and Tucker3-Bs models. The estimated solutions for T3-Bi of $A$ and $B$ were compared to the associated solutions for the unconstrained T3 by the coefficient of congruence. The coefficients were high ($>.999$) for all pairs concerned, and $A$ and $B$ are interpreted in the same way as $A$ and $B$ of the unconstrained Tucker3. The occasion component scores for T3-Bi resemble the occasion component scores for T3-Bs closely, as indicated by the coefficients of congruence (1.000 and 0.999, respectively). Therefore, our interpretation of this model is identical to the interpretation of the T3-Bs model.
4.6. Discussion and conclusion

The results from the CP and Tucker3 simulation study demonstrate that, if smooth underlying components are present, applying smoothness constraints in the CP and Tucker3 models is generally useful for providing a better estimate of the (underlying) components of the CP and Tucker3 models (and the core of the Tucker3 model). The gain in estimation accuracy of constrained estimation is more salient in the case of larger numbers of components, high condition numbers of the component matrices and high error levels.

In the simulation study, the smoothness constraints were imposed by requiring that the smooth component matrix lies in the column space of a B-spline matrix. The performance of the constrained Tucker3 model appeared considerably better if the number of knots of the B-splines was optimized according to the cross-validation criterion compared to the fixed knots choice (of three knots). On the other hand, the performance of CP did not appear to be influenced by the method of choosing the number of knots. This finding suggests that the performance of the smoothness constrained Tucker3 model is more sensitive to the choice of the number of knots than the smoothness constrained CP model. This might be due to the intrinsic axis property of the CP model.

The empirical example demonstrates the use of smoothing in Tucker3 analysis, and the use of monotonicity and smoothness constraints on the data before a Tucker3 analysis. The interpretation of the component matrices of the variables and the subjects, and the core did not alter if a constrained Tucker3 model was used instead of an unconstrained Tucker3 model. The T3-Bi model, in which the analyzed data are constrained to be non-decreasing in the course of time, appears to be reasonable for the data concerned. However, adding the monotonicity constraint to the smoothness constraint did not alter the interpretation of the solution at all, and therefore the simpler T3-Bs model can be preferred here. The interpretation of the time component scores of the smooth constrained T3-Bs solution is clearer than the interpretation of the unconstrained Tucker3 solution, as it is hard to judge whether certain wiggles in the plot of the time component scores of the unconstrained Tucker3 model should be considered important. Additionally, the stability of the T3-Bs solution is higher than of the unconstrained Tucker3 solution. Therefore, in the case of (presumed) smooth components, it appears to be useful to use smoothness constraints on the Tucker3 and CP model.

The commonly used procedures to estimate the Tucker3 or CP models require all elements of the three-way data box to be observed. In the case of data with a smooth mode, the use of the proposed procedures for smoothing the data can be helpful in estimating missing data elements. In longitudinal data, this procedure can be particularly useful if all measurements take place in the same time span, but at different sets of time points for different variables and/or occasions, where the missing data can be assumed to be missing completely at random (Little & Rubin, 1987; see also Chapter 3). Note that the B-spline matrix $B^s (K\times N, K\geq N)$ is a matrix
with \( N \) B-splines which are evaluated in all \( K \) values of the predictor. In the present chapter, the \( K \) measurements of the predictor, which simply represents the time points in the case of longitudinal data, were assumed to be equal for all \( i (i=1,...,I) \), and \( j (j=1,...,J) \), thus the B-spline matrix \( B^s \) is equal for all \( i \) and all \( j \), and (4.4) can be used. However, if there are different measurements of the predictor for different \((i,j)\) combinations, a B-spline matrix \( B^s_{ij} \) must be defined for every combination of \( i \) and \( j \). Now, provided that \( B^s_{ij} \) is of full column rank, \( x_{ij} \) can be projected on \( B^s_{ij} \) by minimizing

\[
\min f_6(w_{ij}) = \|x_{ij} - B^s_{ij}w_{ij}\|^2.
\]

(4.15)

The weights \( w_{ij} \) can be used to estimate \( \hat{x}_{ij} \) on the same time points for all \( i \) and \( j \), namely by defining \( \hat{x}_{ij} = B^s_{ij}w_{ij} \). If the vectors \( \hat{x}_{ij} \) are collected in \( \hat{X} (K \times IJ) \), \( \hat{X} \) can be analyzed by unrestricted Tucker3 or CP procedures. Hence, the use of smoothness constraints in the Tucker3 and CP models is not only useful in improving the estimation accuracy, but also in dealing with data measured at unequal sets of time points.