JOURNAL OF CHEMOMETRICS
J. Chemometrics 2003; **17**: 379–388
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.809

Shifted factor analysis—Part II: Algorithms

Sungjin Hong* and Richard A. Harshman
University of Western Ontario, London, Ontario N6A 5C2, Canada

Received 12 November 2001; Revised 31 March 2003; Accepted 16 May 2003

We previously proposed a family of models that deal with the problem of factor position shift in sequential data. We conjectured that the added information provided by fitting the shifts would make the model parameters identifiable, even for two-way data. We now derive methods of parameter estimation and give the results of experiments with synthetic data. The alternating least squares (ALS) approach is not fully suitable for estimation, because factor position shifts destroy the multilinearity of the latent structure. Therefore an alternative 'quasi-ALS' approach is developed, some of its practical and theoretical properties are dealt with and several versions of the quasi-ALS algorithm are described in detail. These procedures are quite computation-intensive, but analysis of synthetic data demonstrates that the algorithms can recover shifting latent factor structure and, in the situations tested, are robust against high error levels. The results of these experiments also provide strong empirical support for our conjecture that the two-way shifted factor model has unique solutions in at least some circumstances. Copyright © 2003 John Wiley & Sons, Ltd.

KEYWORDS: shifted factor analysis; latent variable position shift; lag; multilinearity; principal component analysis; bilinear model; quasi-ALS; uniqueness; identifiability

1. INTRODUCTION

In Part I [1] we described a method of 'shifted factor analysis' (SFA) that might be useful in analyzing sequentially organized data such as chemical spectra, time series or digitized images. This paper will develop estimation algorithms for the two-way model and describe results of applying them to synthetic data. We begin by briefly reviewing the non-standard formulation and notation needed to state the model.

The bilinear factor analysis model is typically written as a matrix product, such as

$$\mathbf{X} = \mathbf{A}\mathbf{B}' + \mathbf{E} \tag{1}$$

where $\mathbf{X}$ is an $I \times J$ matrix of data, $\mathbf{A}$ is an $I \times R$ factor loading matrix giving weights of each of the R factors for each of the I rows of $\mathbf{X}$, $\mathbf{B}$ is a $J \times R$ factor loading matrix giving weights of the R factors on the J columns of $\mathbf{X}$, and $\mathbf{E}$ is an $I \times J$ matrix of errors or residuals. As pointed out in Part I [1], the shifted factor generalization of the bilinear model cannot be written in this full-matrix form. Instead, we must either (a) write it in representative vector form and apply a 'shift operator' or (b) write it using array notation [2] and represent shifting by constants added to the subscripts.

Suppose we rewrite the bilinear model in terms of an arbitrary, representative column $\mathbf{x}_j$ taken from $\mathbf{X}$. This gives

$$\mathbf{x}_j = \mathbf{A}\mathbf{b}_j + \mathbf{e}_j \tag{2}$$

*Correspondence to: S. Hong, Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada.
E-mail: shong2@uwo.ca

where $\mathbf{A}$ is as in (1) and $\mathbf{b}_j$ is row j of $\mathbf{B}$ written as a column vector. Now consider the case where $\mathbf{X}$ contains sequential data, with each column representing a different time series or spectrum. The shifted factor representation of these data can be written for column j as

$$\mathbf{x}_j = \mathcal{J}_{\mathbf{s}_j}(\mathbf{A})\mathbf{b}_j + \mathbf{e}_j \tag{3}$$

where $\mathcal{J}_{\mathbf{s}_j}(\cdot)$ shifts the columns of $\mathbf{A}$ by the amounts contained in vector $\mathbf{s}_j$, which is row j in the $J \times R$ shift size matrix $\mathbf{S}$. See Part I [1] for a full description of the shift operator. Hereafter we use the shorthand form of the shifting operator for brevity unless ambiguity arises, e.g., $\mathcal{J}_j(\mathbf{A})$ instead of $\mathcal{J}_{\mathbf{s}_j}(\mathbf{A})$.

In array index notation (AIN) [2] the bilinear model (1) can be written in full-matrix terms as

$$x_{\mathrm{IJ}} = a_{\mathrm{IR}} b_{\mathrm{JR}} + e_{\mathrm{IJ}} \tag{4}$$

The shifted factor generalization is

$$x_{\mathrm{IJ}} = a_{[\mathrm{I}+\mathrm{s}_{\mathrm{jr}}]\mathrm{R}} b_{\mathrm{JR}} + e_{\mathrm{IJ}} \tag{5}$$

which differs only by the addition of s_{jr} to the index I, where s_{jr} is an element from an implied matrix of shift values. (For comparison with (2) and (3), the 'representative vector' versions would be written in array notation as $x_{\mathrm{Ij}} = a_{\mathrm{IR}} b_{\mathrm{jR}} + e_{\mathrm{Ij}}$ and $x_{\mathrm{Ij}} = a_{[\mathrm{I}+\mathrm{s}_{\mathrm{jr}}]\mathrm{R}} b_{\mathrm{jR}} + e_{\mathrm{Ij}}$.)

Methods of parameter estimation for this shifted factor model can be derived and expressed using either approach (a) or (b). For simplicity, we will only use the more familiar matrix approach in the following sections (refer to Reference [3] for the AIN equivalents of all equations in the following sections).

2. PARAMETER ESTIMATION FOR SFA: QUASI-ALS

Unusual difficulties are encountered when one tries to develop a procedure for estimating the parameters of the shifted factor model. The alternating least squares (ALS) approach is not fully suitable, because factor position shifts destroy the multilinearity of the latent structure. Therefore we have developed an alternative 'quasi-ALS' approach.

Our approach resembles the more conventional ALS method in that: (a) the parameters for each mode are estimated separately and are conditional on the parameter values for the other mode; (b) estimates for each parameter set are obtained sequentially and cyclically until convergence is reached; and (c) for any mode in which it is appropriate, a standard ALS regression-like estimation is used. However, factor position shifts make a (multi)linear approach inappropriate. Here, two other techniques are employed: (d) to estimate weights, we employ a regression using modified arrays from which interfering systematic variance has been (approximately) 'partialled out'; and (e) to estimate shifts, we use a straightforward search for a maximum fit by testing each of a set of feasible alternatives. Considered together, we refer to this collection of strategies as a 'quasi-ALS' approach.

2.1. B and S estimation

Unlike ALS, the SFA estimation procedure differs from one mode to another. First one estimates **B** and **S** in an interlocking fashion, deriving the R factor weights and R shift values at each level of mode B (the shifting mode) before proceeding to the next level. For, say, level j, conditionally optimal shift values are determined by trying a number of combinations of values for $\mathbf{s}_j$ and, for each set of trial values, obtaining regression estimates for the corresponding $\mathbf{b}_j$ factor weights to find the one with the best fit. Of course, the estimates of both **B** and **S** are conditional on all the other model parameter values, which are held fixed. One might consider the estimation of **B** and **S** as a subiterative optimization within one of the quasi-ALS steps. $\mathbf{b}_j$ can be estimated as

$$\hat{\mathbf{b}}_j = \left[\mathcal{S}_j(\mathbf{A})\right]^{+}\mathbf{x}_j \tag{6}$$

where $(\cdot)^{+}$ represents the Moore–Penrose generalized inverse.

It is implied in (6) that the shift vector $\mathbf{s}_j$ is provided. Unfortunately, the shift sizes cannot be estimated directly by regression and so the best $\mathbf{s}_j$ must be found by some other optimization method. At this stage of the quasi-ALS estimation we assume integral values of shift. It is thus possible to find the conditional global optimum by means of an exhaustive search. The first stage in this (initial) algorithm uses (6) to compute and evaluate trial values of $\mathbf{b}_j$ for all combinations of allowed integer shift values, a method which we refer to as 'Exhaustive Integer Search' (EIS). For example, when it is assumed that there are three factors and that the real shift is no greater than 5 in either direction, there are 11 possible positions for each factor (11 being the number of integers from -5 to 5) and hence 11^3 alternative combinations of integer shifts to try for column j. Recall that the sign of shift values represents forward or backward shifting. At the completion of the trials for column j, one replaces $\mathbf{s}_j$ with the best combination of shifts, and $\mathbf{b}_j$ with the resulting regression estimate that minimizes the sum of squares of error for column j (i.e. the $\hat{\mathbf{b}}_j$ due to (6)). When this has been done for all columns, the updated shift and factor weight estimates are then used in estimating the other mode loadings. The maximum allowed shift must be determined externally, based on the knowledge about a given data set and the nature of the sequential factors of interest. Since the number of combinations to try increases multiplicatively with the maximum shift allowed, there are limits to what is feasible.

2.2. A estimation

In the estimation of **A** it is necessary to undo the factor shifting in the data so that the alignment of a given factor is consistent across the columns of **X** and between **X** and **A**. However, the realignment of columns in the data must be done independently for each factor so that the 'unshifting' (i.e. lining up across columns) applies only to the contribution of a certain factor to the data. This is because the pattern of shifts (across J columns) for each factor is allowed to be independent of the pattern for the other factors. Thus it is necessary to separate the data into two latent parts before unshifting a certain factor's contribution to the data: one part is due to factor r and the other is due to the remaining factors and residuals. The factor r part in column $\mathbf{x}_j$ is estimated by

$$\hat{\mathbf{x}}_{j(r)} = \mathbf{x}_j - \mathcal{S}_j\left(\mathbf{A}_{(\sim r)}\right)\mathbf{b}_{j(\sim r)} \tag{7}$$

where the parenthesized subscripts indicate which factors are involved either in the isolated part of the data or in a parameter set to be used for estimation. For example, the parenthesized $\sim r$ indicates that column r is replaced with zeros so that the contribution of factor r is removed. Once $\hat{\mathbf{x}}_{j(r)}$ is isolated from each data column $\mathbf{x}_j$, one can then reverse-shift $\hat{\mathbf{x}}_{j(r)}$ to undo the factor shifting that occurred in the factor r part. This can be written as an inverse shifting for all columns in the factor r part at once as

$$\widetilde{\mathbf{X}}_{(r)} = \mathcal{S}_r^{-1}\left(\mathbf{X}_{(r)}\right) \tag{8}$$

where $\mathcal{S}_r^{-1}(\cdot)$ represents an inverse shifting operator (i.e. $\mathcal{S}_{\mathbf{s}_r}^{-1}(\cdot) = \mathcal{S}_{-\mathbf{s}_r}(\cdot)$; refer to Part I [1] for the properties of the shifting operator). It is used here to 'unshift' the isolated factor r part of the data. The shift for each of the J columns in the isolated factor r part is given by the values in $\mathbf{s}_r$, the rth *column* of the shift matrix **S**. The tilde on top of the left hand $\mathbf{X}_{(r)}$ indicates that the J columns in $\mathbf{X}_{(r)}$ are properly unshifted for factor r. Once the factor isolation and the following unshifting are done, $\mathbf{a}_r$ can be estimated by regression as

$$\hat{\mathbf{a}}_r = \widetilde{\mathbf{X}}_{(r)}\left(\mathbf{b}_r'\right)^{+} \tag{9}$$

Note that the isolation of the variance that is due to factor r is only approximate. The isolation is particularly approximate at the beginning stage of iterations, since all other parameter estimates are yet very rough, and improves as the other parameters are better estimated. The least squares property of (9) would hold if Equation (9) described a regression based on an observed **X** matrix. Since $\widetilde{\mathbf{X}}_{(r)}$ is a modified **X** adjusted by imperfectly subtracting variance due to all factors except r, the regression is only approximately

least squares in terms of residuals from the original **X**. As a result, the estimation of **A** is not true least squares, but in later iterations comes close to it.

2.3. WLS for the ends of factor profiles

When one shifts forward the location (i.e. increases the subscript value) of elements in the columns of **A** or **X**, some elements at the beginning of the shifted column become 'empty'; there are no elements preceding them that can be shifted into the newly opened positions. The question then becomes: how should one fill these cells? One choice is to fill them with zeros. However, when their previous values are substantially greater than zero in absolute value, zero may be an inappropriate estimate. This could produce substantial error in data fit at the ends and indirectly lead to distorted values for the intermediate locations as well. To handle this problem more gracefully, a weighted least squares (WLS) algorithm has been developed for estimation of the sequential mode parameters.

Given an $I \times J$ data matrix (I being the number of levels in the sequential mode) and a predetermined maximum integer shift of $s_{\max}$, the number of levels in the sequential mode is increased to $I + 2s_{\max}$ by adding $s_{\max}$ levels to both ends before the actual estimation starts. These locations are referred to as 'beyond-window' levels. These added levels are then treated as missing data. Of course, **A** is also augmented so that it contains the weights for the beyond-window levels as well as for the original window levels. The estimation of **A** then 'ignores' those locations by means of a weighted least squares version of (9) in which these locations are given a weight of zero in the fixed parameter set.

The WLS procedure for **A** must estimate one element in $\mathbf{a}_r$ at a time, since whether or not a location in any column of $\mathbf{X}_{(r)}$ is empty will differ from one row to the next. The estimates are obtained as

$$\hat{a}_{ir} = \tilde{\mathbf{x}}'_{i(r)} \left(\mathbf{b}'_r * \mathbf{w}'_{i(r)} \right)^{+} \tag{10}$$

where '$*$' represents the Hadamard (i.e. element-wise) product and $\tilde{\mathbf{x}}_{i(r)}$ is row i in $\widetilde{\mathbf{X}}_{(r)}$, the matrix of isolated and realigned (or 'unshifted') factor r contributions to the data. $\mathbf{w}_{i(r)}$ is a J-element weighting vector of 1 or 0 for row i of factor r. If a cell in $\tilde{\mathbf{x}}_{i(r)}$ is empty (and hence temporarily filled with zero) after unshifting, the corresponding value in the weighting vector becomes 0; otherwise, it is 1. This ensures that the elements in the fixed parameter set (i.e. $\mathbf{b}_r$) corresponding to the empty cells are not used in the a_{ir} estimation. It is noted that this can happen only at the 'beyond-window' levels, i.e. in the first and last $s_{\max}$ rows of the $(I + 2s_{\max}) \times J$ augmented (and isolated) data $\widetilde{\mathbf{X}}_{(r)}$. When all values are zero in $\tilde{\mathbf{x}}_{i(r)}$, the previous a_{ir} remains, since there is no valid element at level i in the data for a_{ir}.

In the mode B estimation, only the original window levels in the augmented **A** are used after properly shifting the columns of **A**. After all parameters are updated once in this way, one replaces the 'missing' values of the beyond-window levels in the augmented **X** with the values predicted by (3), this time with the augmented **A**. The imputed values for the beyond-window levels will provide the least squares property for the first and last $s_{\max}$ levels when they are included in the window after unshifting. See Reference [4] for the least squares property of the imputed values in the ALS estimation.

2.4. Fractional shift estimation

Although it seems natural to shift a sequential factor profile by an integer number of levels along the sequential mode, such an integer limitation may not be valid for real data. Obviously, the measurement unit is somewhat arbitrary; it depends on the refinement of the measuring tool used, the smoothness of the sequential factors of interest, conventions, etc. Consequently, it may be useful to generalize the shift estimation method to one that allows for a fractional number of shifting units. This might become particularly valuable when sequential factor profiles rise or fall steeply from one level to the next, such as happens with sharp-peaked and complex multi-peaked spectra. A method that we currently use to implement a fractional shift algorithm is discussed below.

Once all parameters have converged in the EIS (Exhaustive Integer Search) stage, our current quasi-ALS procedure shifts to a new estimation algorithm which searches fractional shifts to minimize further the residual sum of squares. This algorithm uses a shift estimation procedure based on a 'Fractional Line Search' (FLS). It involves both a bracketing procedure and semi-exhaustive trials. Given all converged factor loadings and integer shifts from EIS, FLS starts with an initial trial interval of 0.5. For example, given an integer shift estimate of 3 for s_{jr} from EIS, FLS tries shifts of 2.5, 3 and 3.5. If either 2.5 or 3.5 results in a smaller residual sum of squares for level j than does 3, one replaces s_{jr} with it and cuts the interval in half. FLS then tries 2.25, 2.5 and 2.75 in the next iteration if 2.5 had a better fit, or 3.25, 3.5 and 3.75 if 3.5 did. Otherwise, both the bracketing intervals of fractional shifts and the shift estimates remain the same. FLS tries all combinations of the three evenly spaced fractional values for each factor in order to avoid any order effect among factors. Thus there will be always 3^R combinations of fractional shifts for level j in FLS when R is the number of factors to extract. See References [5,6] for further details on the bracketing procedure.

2.5. Modified procedures to accelerate estimation

In addition to the basic estimation procedures just described, several techniques to speed up the procedure have been developed. These include two-stage EIS, application of exhaustive search only every n iterations, identification of conditions allowing early termination of the shift search, rational initialization of shifts, and, in certain cases, a method of reducing the number of modes in the data. These are described in the Appendix.

3. DATA ANALYSIS

We now study the properties of the shifted factor model (3), and the quasi-ALS estimation procedure described above, by using them to analyze error-free and fallible synthetic data. This is to be followed by an application to real data (the same data as in Reference [7]), but since this involves a four-way data set and model, it will be covered in Part III [8].

The initial analyses are performed on error-free synthetic data to empirically confirm the conjectured uniqueness

properties of SFA models. Next we explore the performance of SFA with fallible data. Finally we go beyond the domain covered by current mathematical analysis: instead of using two-way data sets in which shifted factor profiles are weighted by mode B factor weights as defined in (3), we test the shifting *per se* by fitting model (3) to a two-way shifted data set in which the mode B factor weights are all unity. This 'pure shift' case, with no variation in mode B weights, is of particular interest with regard to the contribution of shifts to the uniqueness properties of the two-way SFA model.

3.1. Weighted factor case

3.1.1. *Previous results*

The first computational tests of uniqueness were carried out by Hong [5]. His methods and results have not been published elsewhere, but they contribute significantly to the evidence for uniqueness, so we briefly summarize them here.

Because the initial algorithm and computer were relatively slow, these tests were restricted to the two-factor case, and only four random starting points were used in the analyses of each data set. Mode A, the sequential 'shifted' mode, contained 100 or 60 levels and mode B, the 'shifting' mode, contained 20 levels. Three different sets of mode A factor profiles, shown in Figure 1, were created to simulate electrophysiological (evoked response) data. The shape of the 'handmade' factors is relatively arbitrary; the other two were created from sections of sine functions. The mode B factor weights were randomly sampled from a uniform distribution between 0 and 1. The 20×2 matrix $\mathbf{S}$ of shift sizes contained integers randomly and uniformly sampled from the range from -3 to 3.

For all three error-free data sets, Hong found that the true loadings were perfectly recovered by all perfect-fitting solutions (four of four starts for the 'handmade' data, three for the 'hump' and one for the 'sine' data). The other solutions were considered local optima. There was never a perfect-fitting solution that was different from the true solution used to generate the data. When the tests were repeated with 5%, 10% and 25% error added to the data sets, the recovered solutions were consistent with true loadings that had been corrupted by the noise in the data.

Hong subsequently performed 10 additional analyses of these data sets (somewhat truncated) using a faster algorithm with an unconstrained nonlinear optimization routine, FMINU (in the MATLAB Optimization Toolbox, ver. 1.0, which uses a quasi-Newton method). Again all perfect-fitting solutions matched the true loadings (10 for the 'handmade' data, six for the 'hump' and four for the 'sine' data). The other solutions were local optima or cases of very slow convergence. As will be noted in the following subsections, it seems that curves that change very gradually presented a more difficult problem to the algorithm, probably because there was little change in the fit when only modest shifts were applied. The reader is referred to Reference [5] for a more complete description and discussion.

3.1.2. *New results*

Synthetic data sets of two sizes were constructed. The 'narrow' size was 60×15 (i.e. 60 sequential levels in mode A and 15

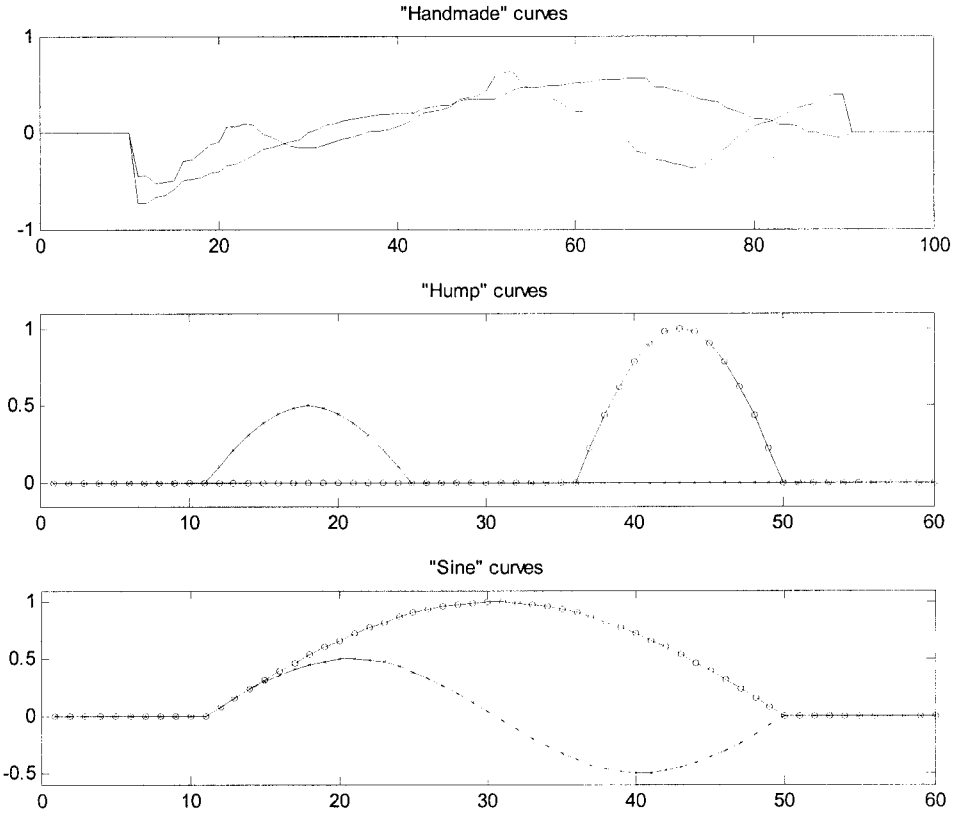

Figure 1. Sequential factor profiles used in the initial experiment: horizontal axis gives sequence position (i.e. subscript in factor matrix **A**); vertical axis shows factor loading size. For clarity, one profile of the 'hump' and 'sine' curves is marked with 'x' and the other with '○'.

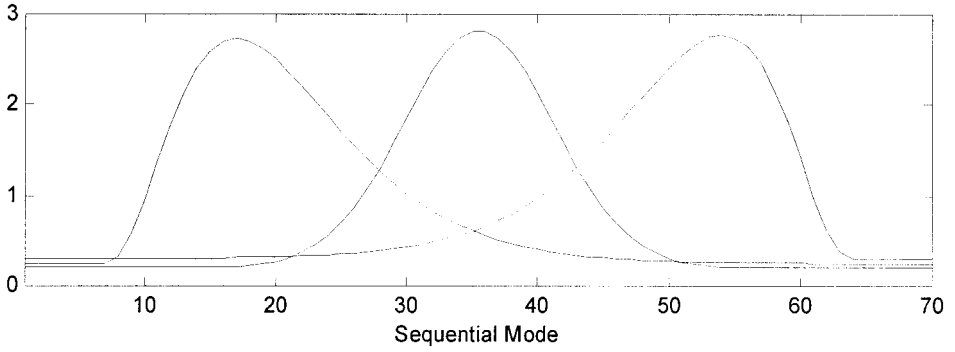

Figure 2. Sequential factor profiles. The middle curve is the normal density function of 70 evenly spaced standard deviates from -6 to 6. For the skewed curves the chi-square density function of 70 evenly spaced chi-square values from 0.5 to 35 with d$f = 7$ is first obtained and then shifted left by 7 levels with zeros filling the empty levels. The resulting profiles are standardized such that each has unit mean square. Then a constant of 0.2 or 0.25 is added to these standardized curves to produce, respectively, the symmetrical (middle) and positively skewed (left) curves. The negatively skewed curve (right) is the same as the positively skewed curve except that the order of levels is reversed and the added constant is 0.3.

levels of mode B) and the 'wide' size was 60×75. The loadings used to generate these data were obtained as follows. Three vectors of 70 elements were generated to provide the latent sequential mode factor loading profiles. The profiles are unimodal and slowly changing (this reduces the effects of small shift changes during iteration and thus makes the data more 'challenging' for the algorithm). The shape of the three curves was defined by using a normal density function and a chi-square density function. In order to ensure that the resulting mixture of curves was not 'selective' at some levels (i.e. contained levels in which all but one of the curves have near zero values), a different positive constant was added to each of the curves. Figure 2 illustrates the resulting sequential profiles*. Since shifts of up to ± 5 levels were allowed, we used 70-element latent profiles when generating the 60-element observed data profiles under different conditions of shifting. This resulted in 60 levels of complete data for the sequential mode (mode A). Two sets of the mode B factor loadings (15×3 and 75×3) were sampled randomly from a uniform distribution of real numbers bounded by 0 and 1 exclusively. Two shift parameter matrices of the same sizes were also randomly sampled from a uniform distribution of integers, but in this case bounded by -5 and 5 inclusively. Then the three sequential profiles, the two sets of mode B weights, and shifts were combined to generate error-free 60×15 and 60×75 shifted data sets with latent structure specified by the two-way SFA model (3).

To produce fallible data, normally distributed random errors were added to the error-free shifted data; this resulted in some negative observed values. (While we realize that in many situations noise would be truncated because the data are non-negative, this is not always the case. Using error which was not bounded by zero was considered a conservative choice, since it should make it harder to recover the true profiles.) The error proportion in the fallible data is, on average, 14.88% of the total mean square and 42.82% of the total variance. This large difference reflects the fact that the true parts have, on average, a total mean of 1.74 standard units, while the expected mean in the error part is zero. The fitting of the large mean values in the true part does not contribute substantially to the estimation of model parameter profiles; hence the proportion of error *variance* better reflects the 'difficulty' of the problem—the information loss or signal contamination due to error. By this reasoning, the fallible data sets can be considered difficult cases to fit, since almost half of the total variance is due to error.

Model (3) was fit to each of the resulting four data sets using the quasi-ALS procedure. Ten different random starts were applied in order to increase the probability of obtaining the global optimal solution. Convergence was evaluated using the changes in loading parameters rather than in fit, in order to ensure that the parameter estimates themselves had become stabilized. The optimization of each of EIS1, EIS2 and FLS was stopped either when all parameters had changed by less than 0.001% from one iteration to the next, or when this criterion was not met in 1000 iterations. Factor loadings were constrained to be non-negative by using the fast non-negativity-constrained least squares (FNNLS) algorithm due to Bro and de Jong [9]. This constraint is appropriate because all true parameters are sampled from a population of positive values as described above, and realistic because typical chemical spectral data are non-negative*.

The top part of Table I summarizes SFA recovery correlations for the shifted data generated by using the profiles given

*Initially we tried another set of sequential profiles that are smoother than those illustrated here and which will be used for the three-way shifted data in Part III [8]. However, the resulting two-way error-free shift data were too difficult for the current two-way quasi-ALS algorithm to perfectly recover the parameters.

*We also performed a small experiment without the non-negativity constraint (using the 60×75 two-way data and the three-way shifted data to be discussed in Part III [8]) to check uniqueness (see following paragraphs) and to see if there is a meaningful difference in fitting time between the constrained and unconstrained SFA. The time saving that usually results from dropping non-negativity was not observed here. This is because the SFA algorithm cannot take advantage of simultaneous estimation of all levels which normally would become possible. Also, since we used Bro and de Jong's fast non-negativity-constrained algorithm, the time cost of applying the constraint was considerably reduced. More importantly, keeping the constraint improved the algorithm's ability to avoid problematic paths and local optima, thereby reducing the total number of iterations required to obtain reasonable recovery.

Table I. Recovery correlations and model fits (R^2) for weighted two-way shifted data

	15 mode B levels		75 mode B levels	
	Error-free	Fallible[a]	Error-free	Fallible[b]
Parameter recovery (*r*)				
SFA				
A	0.9423	0.8396	1.0000	0.9598
B	0.9775	0.9343	1.0000	0.9285
S	0.8123	0.6393	0.9846	0.7271
S ∗ **B**	0.9337	0.7594	1.0000	0.8511
PCA[c]				
Mode A agreement maximized:				
A	0.7434	0.6249	0.8504	0.8130
B	0.9874	0.9274	0.8838	0.8438
Mode B agreement maximized:				
A	0.7790	0.6699	0.8868	0.8380
B	0.9566	0.8426	0.8379	0.8046
Fit to data (R^2)				
SFA	0.9990	0.7132	1.0000	0.6706
PCA	0.9662	0.6463	0.8610	0.5447

[a]Error proportion is 14.77% of total mean square and 44.50% of total variance; mean of the true part is 1.7754 standard units.
[b]Error proportion is 14.99% of total mean square and 41.13% of total variance; mean of the true part is 1.7019 standard units.
[c]Components in one mode are 'rotated' (linearly transformed) to maximize agreement with the true structure while a compensatory transformation is applied to those in the other mode.

in Figure 2. The results reported are from the final step in the quasi-ALS estimation, FLS. All recovery correlations are averaged over the three factors within each solution for brevity. For the error-free shifted data with 75 levels in mode B the quasi-ALS procedure correctly recovers all model parameters except for a few shift values. Recovery (within 1000 iterations) is not as good when there are only 15 levels. In no case, however, was perfect fit obtained with other than the values in **A** and **B** used to synthesize the data. This supports the conjecture that model (3) has a unique solution.* It also shows that the quasi-ALS procedure provides a way to recover the unique valid solution so long as there are sufficient sources of independent shifts (i.e. enough mode B levels).

Even without the non-negativity constraint, the sequential factors were perfectly recovered from the wide two-way data except for about 5 levels at each end (presumably owing to algorithm limitations). This means that two-way shifted factors are uniquely identifiable in a broader range (i.e. the range of real numbers instead of only positive numbers), although it may be impossible to fully recover the latent factors near end levels of the sequential mode.

Recovery is always worse for the shift values than it is for the factor weights. When both the estimated and true shifts are weighted by the corresponding mode B weights, though, the recovery significantly improves; the recovery correlation for **S**∗**B** (Hadamard product of **S** and **B**) becomes almost perfect. The marked difference between **S** and **S** ∗ **B** is due to divergence of only one or two shift estimates for each factor, but the effect is large because these estimates deviate substantially from the true shifts and hence substantially reduce the recovery correlation for **S**. However, inspection reveals that these highly incorrect shift estimates are at levels for which the mode B weight for the factor concerned is very small. Thus the occurrence of these highly deviant values is understandable: a factor with a very low weight at some level of mode B makes very little contribution to the data at that level, and so highly deviant shifts do not substantially reduce the overall fit. When both the missed shifts and their true values are downweighted by the corresponding mode B weights, the recovery correlation improves considerably.

Quasi-ALS recovers **A** and **B** from the 75-level fallible shifted data better than might be expected, given that more than 40% of the total variance is due to error. This suggests that the SFA modeling is robust against random error. Plotted in Figure 3 are the best recovered sequential profiles from the 75-level cases. The right plot shows the good recovery by SFA of the sequential factors, even with a substantial amount of error.

For comparison, the shifted data sets were also fit by the principal component model. However, the standard two-way PCA solution places axes in positions which are mathematically convenient but usually do not reflect the position of data-generating factors. Because of this, the associated recovery correlations would all be low and uninformative. To make the correlations more meaningful, the components were 'rotated' (i.e. linearly transformed by regression) to maximum agreement with the true parameters. Two rotations were done, one maximizing the agreement between the estimated **A** and true **A** (with a compensatory rotation of **B**), and another one maximizing the agreement between the estimated **B** and true **B** (with compensatory rotation of **A**). The PCA model fit and recovery correlations for each data set are based on one solution, because the two-way PCA provides a non-iterative closed form solution, given a rotation criterion.

The PCA recovery correlations are presented in the middle part of Table I. As expected, the recovery of **A** is better when PCA components are linearly transformed to maximise fit of **A** to the true **A** than when transformed to maximise fit of **B** to the true **B** and *vice versa* for the recovery of **B**. With more levels in mode B the recovery of **A** improves while the recovery of **B** worsens. The best recovery of **B** by PCA (after rotation to maximal agreement with the true) occurs with the 15-level error-free data. Here the fit is better than the corresponding SFA recovery of **B**. This confirms that 15 levels are not enough for the quasi-ALS procedure to recover model parameters well in the current study.* However, the sufficiency of the number of levels of the shifting mode is conditional on other things, such as steepness of sequential profiles, imbedded shift magnitude, independence of the shifting pattern, etc., and so these results must be interpreted as given those conditions chosen here.

*Technically, the solution is 'essentially unique', i.e. unique up to permutation and rescaling of factors; this does not normally affect interpretation, and so this distinction is overlooked for simplicity in this paper.

*In order to reduce the possibility that the imperfect solution from the error-free 15-level data is caused by a premature stop of iteration, we applied extremely stringent stopping criteria: a maximum iteration of 10 000 in each stage and a maximum percentage change in parameter estimates of 10^{-10}. While providing some improvement, these criteria still did not result in a perfect solution from the 15-level data.

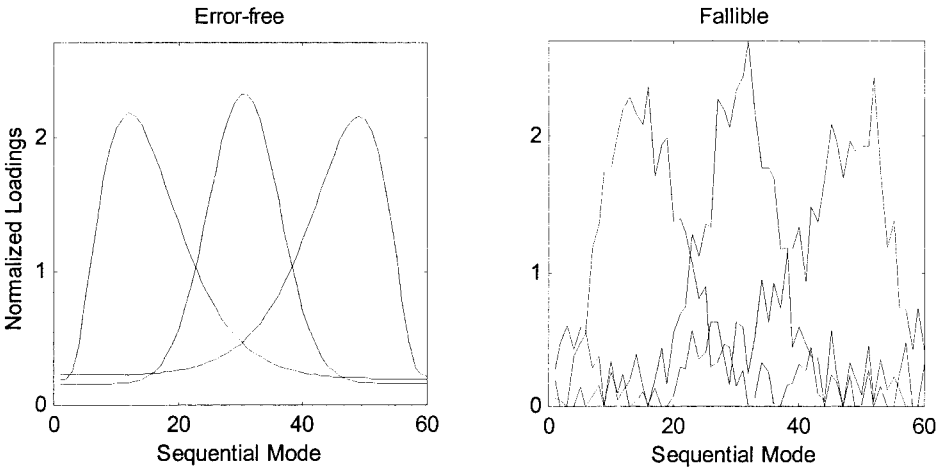

Figure 3. Recovered sequential factors from error-free (left) and fallible (right) 75-level shifted data. Loadings are normalized so that each factor has unit mean square.

The bottom part of Table I shows the overall fit to the data obtained by each procedure. The fit values for the shifted factor model are better than the PCA values; this is to be expected, since the SFA model incorporates and thus 'corrects for' the effects of the shifts. In the error-free case the SFA fit is essentially perfect; in the fallible case it approximates the proportion of true variance in the data, except that it is inflated somewhat by the fitting of error. While the PCA model seems to do very well in the 15-level error-free case, this is because of substantial fitting of error. The lower fit obtained with the 75-level data sets provides a more realistic (though still inflated) indication of fit, where there is less leeway to fit error. Note that both fallible and error-free shifted data are fallible in PCA, since both the random error and the independent shifting are not accountable for by the standard bilinear model.

The FLS algorithm was developed for shifted data in which the measurement unit of the sequential mode is not small enough for EIS to precisely adjust for a position offset. However, even when the true shift values are integers (as they are here), FLS almost always improves the parameter estimates as well as model fit unless EIS has already provided a perfect solution. Inspection of the solutions reveals that FLS shifts the incorrect position of latent profiles one unit toward the true positions.

3.2. Pure shift case

Given that the incorporation of shift information into the factor model allows a unique two-way solution to be obtained (when the shifts vary independently), the question now becomes: what is the strength of the information provided by the variation in the data resulting from the factor shifts—to what degree might it determine the factors recovered by the two-way SFA model, independently of information provided by the data variation resulting from factor size changes across levels of mode B? To explore this, we created a synthetic shifted data set using the same true **A** and **S** parameters as for the 75-level error-free data, but for which all true **B** weights are unity. Thus the mode B variation in the resulting 60×75 error-free 'pure shift' data is purely due to the independent shifting. Put another way, the data matrix would have rank one if not for the effects of factor shifts. (Because of the shifts, it has full rank.)

In one experiment a restricted SFA model was fit to the data; this model 'knew' that all factor variation was due to shifting. That is, when this restricted version of the SFA model (3) was fit the mode B factor weights were constrained to be unity. Other than this, the fitting procedure was the same a non-negativity constraint on the mode A weights, and the stopping criterion of maximum parameter change less than 0.001% or a maximum of 1000 iterations in each of EIS1, EIS2 and FLS. The result was that **A** and **S** were perfectly recovered (to four decimals). Since the question here was theoretical, no fallible data experiments were conducted.

In the second and more stringent test we removed the equality constraints on the mode B weights; thus we fit the same model that we did in the weighted factor case. Provided the two-stage EIS and the same fitting procedure, no solution was perfect out of 10 random starts; on average, model fit (R^2) was 0.9947 and parameter recovery was 0.6317 for the sequential factors and 0.7987 for shifts.

We speculated that this difficulty in attaining perfect fit could be due to the nature of the algorithm. For one thing, the two-stage EIS is more likely to get trapped in a local optimum. Consequently, we fit the same model given the same fitting procedure except that the two-stage EIS was replaced with the single-stage EIS, in which all combinations of integer shifts are tested in each iteration. Two solutions out of the 10 random starts were nearly perfect; on average, R^2 was 0.999998 and recovery was 0.9995 for the sequential factors and perfect for shifts. The imperfect recovery of the sequential factors in these solutions seemed to be a result of insufficient iteration (1000 iterations in each of EIS and FLS). Since shifts were already perfectly recovered and almost certainly would not change in subsequent iterations given the closeness of the recovery of the other parameters, we sped up the subsequent iterations by skipping this (by far the most time consuming) part of the estimation procedure. We adopted a more stringent stopping criterion of 10 000 iterations or maximum parameter change less than 10^{-10}% and resumed the tests. This time we were able to obtain perfect recovery (to four decimals). These results confirm that independent shifting by itself can (at least sometimes)

provide sufficient information to determine the model parameters.

4. DISCUSSION

The first and most fundamental finding of this paper is that factor shifting can be 'undone'. Position shifting of sequential factors *per se* does not irretrievably destroy or degrade otherwise strongly imbedded (bi)linearity. The second is that the two-way SFA model, which incorporates added structure to account for and describe the shifting is able to extract additional information from the data, and so can fully identify its parameter values, whereas the non-shifted bilinear equivalent is under-identified. The third point is that the quasi-ALS procedures presented here provide one way to fit the SFA model and are robust against high levels of error, probably because of the added information due to the shifting variation.

Estimation of SFA parameters has a few novel properties that do not occur with the standard bilinear (PCA) model. The shift parameters (collected in **S**), which determine the position of the sequential factors, have an unusual kind of linkage with the corresponding mode B weights. The magnitude of each mode B weight moderates the effect of the corresponding shift on the data. Thus it is important to interpret an estimated shift in conjunction with the corresponding mode B weight, as we did when using the 'weighted' shifts ($\mathbf{S} * \mathbf{B}$) in assessing the recovery of shifts. The shape of the sequential factors is another aspect to take into account. The overall impact of a given factor's shifts on the data also depends on the steepness and irregularity of its profile in the sequential factors; the steeper or more irregular a sequential factor, the stronger impact its shift has on the data [5].

It is the independent shifting that makes two-way SFA an 'intrinsic axis' [10] or rotationally unique model. In certain cases, however, independent shape of sequential factors can also become critical for a unique identification of shifted factors by the quasi-ALS procedure. For example, when two sequential factors are similar in shape and close to one another in position, they may get easily confused with one another in the SFA fitting process; this happens if they become aligned or cross over one another in position. (Of course, this requires that the distance between the two sequential events is less than the maximum shift prescribed in the fitting.)

For various reasons, sequential factors might change shape, as well as position, across levels of mode B. Since SFA requires sequential factors to be invariant in shape, it will fail if the shapes change substantially. Sequential factors might sometimes change shape because of measurement artifacts, but at other times because of a systematic and meaningful process. As noted in Part I [1], when the source of the shape change is not arbitrary, it would be useful to incorporate it into the model by introducing a few more parameters. For example, in the simplest form a parameter might control the width of events in the sequential mode. Such a generalized model might be, however, less stable than the SFA model, since the shift and shape parameters might be confounded with each other as well as with the normal factor weights. Another kind of provision for the shape change problem is the use of external knowledge about the data, as in the transformation of discrete PDS calibration into a CPDS model by means of known temperature variation [11]. Our understanding of the shape change problem is too limited, so it is premature for us to provide concrete discussion at this time.

We discovered a few interesting algorithmic properties of the quasi-ALS procedure. One might naturally think that FLS would be advantageous primarily when the true shifts have fractional values or, equivalently, when the measurement unit of the sequential mode is not refined enough to accurately estimate shifts and, consequently, factor weights as well. However, if the data are fallible, FLS has been found to consistently improve the parameter estimates (as well as the model fit) even when the true shifts are integers. This is probably because the fractional shifting allowed in FLS always results in better-optimized least squares estimates than does the integer shifting in EIS. Consequently, FLS tends to provide a better chance to avoid local minima, which may be more likely to occur in the quasi-ALS than in the standard ALS procedure. FLS is particularly essential when the two-stage EIS (see the Appendix) is used, which tries much fewer shift combinations in order to save time, and hence the solution is more likely to converge at a local minimum than it is in the single-stage EIS.

The imperfect variance isolation procedure in the sequential mode estimation makes the quasi-ALS procedure not true least squares. This undesirable property tends to make the resulting solution less reliable, more prone to local optima and slower to converge than the least squares solution, say, for Parafac1 or Parafac2. However, the worsening fit, if any, usually occurs in the middle phase of iteration, but hardly at all as the solution gets close to the global optimum. In most cases a more stringent criterion for convergence and a few more starting positions may be enough to ensure that the best quasi-ALS solution is as reliable as the standard ALS solution.

Acknowledgements

We thank Rasmus Bro for the fast non-negativity constrained least squares (FNNLS) algorithm; it has been essential for this study. We also thank Margaret E. Lundy for editing most sentences in the text; it made the manuscript readable and eliminated many ambiguities.

APPENDIX. ACCELERATION OF THE QUASI-ALS ALGORITHM

The exhaustive trials in EIS for the estimation of **S** and **B** are a bottleneck in the computing speed of the quasi-ALS algorithm as the data become large (e.g. the two-way 75-level shifted data used for the pure shift case and the chromatographic data analyzed in Part III [8]). For multiway cases it is usual that more than 95% of the total computation time is spent in inverting an unfolded version of the fixed parameter part in the mode B estimation (e.g. Khatri–Rao product [12] of **A** and **C** in a three-way case). For example, for the $60 \times 15 \times 10$ synthetic data set analyzed in Part III [8] (with three factors and 11 allowed shift positions), one iteration takes about 2 min in EIS and about 3 s in FLS. If 300 iterations

are needed for convergence, the EIS and FLS runs for this data set need respectively 10 h and 15 min per random start. With two speed-up provisions, the two-stage EIS and periodic shift search every 10th iteration, time of the 10 h EIS run was reduced to about 23 min.

In general, the increase in the computation time is approximately exponential to the number of factors and multiplicative to the maximum allowed shift and the number of levels in all modes. The number of levels in the data and the maximum shift are controllable to some degree with prudent consideration, but not the number of factors, in the sense that it is independent of a particular measurement situation or an analytic method for a given type of data. It is also noted that the number of iterations needed for convergence is a structural question. That is, it depends on many conditions in the latent level, such as error level, sequential profiles' shape, independence in shifting pattern (as well as in factor weights), etc., which largely vary from a certain type of data to another. In the following subsections a few provisions have been introduced so that EIS becomes practical even for large data sets.

A.1. Two stage EIS

EIS has been split into two successive stages: a less refined first stage and a further refining second stage. Stage 1 (EIS1) tries all combinations of only even-numbered shifts. Then stage 2 (EIS2) further refines the final shift estimates from EIS1 by allowing odd-numbered shifts.

For example, when three factors are allowed to be shifted at maximum by an absolute value of 5, EIS1 tries 5^3 combinations, using -4, -2, 0, 2 and 4 as shift values. Once all factor loadings have converged with the even-numbered shifts, EIS2 tries 3^3 combinations, using three successive integers around the final shift estimate from EIS1. For example, when the final estimate from EIS1 for s_{jr} is 2, EIS2 tries 1, 2 and 3.

Although the time gain by the two-stage EIS estimation becomes substantial when there are more than two factors, it tends to hit local minima slightly more often than does the single-stage EIS. This is because EIS2 and FLS cannot compensate for a shift estimate in EIS1 that misses its true shift value by more than 2 in absolute value. Let us suppose EIS1 estimates s_{jr} to be 2 when its true value is 5. The best approximation is then 3 in EIS2, since it allows only 1, 2 and 3 during all iterations. FLS can further approximate toward the true value by 1 at most, since the bracketing procedure described above cannot change a shift by more than 1 in either direction. Hence the final approximation is 4. It has been consistently observed with synthetic shifted data that FLS tends to compensate for such poor estimation of shifts by EIS1 even when the true shifts are integers.

A.2. Periodic application of shift search

Another provision for time saving is to apply the exhaustive trials only every few iterations, say 5 or 10, in both EIS1 and EIS2. Since the exhaustive trials are computationally very expensive, it suffices to apply the actual EIS procedure every 10th iteration. During all intermediate iterations, one updates only factor loadings, given a set of fixed shifts from the last actual EIS trials. This periodic application of the actual exhaustive trials has been confirmed to be useful with two-, three- and four-way error-free shifted data. It is strongly recommended that one applies the actual EIS procedure every iteration at least during the first 10 in both EIS1 and EIS2, in order to avoid any bad model search path arising from poor initial shift values. It is also noted that SFA becomes equivalent to Parafac1 until the EIS procedure has actually been applied, when all initial shift values are set to be zero.

A.3. Shift search termination

Our current quasi-ALS algorithm successively applies three stages: EIS1, EIS2 and then FLS. In our experience the shift value estimates do not change after a number of iterations in each of the three stages, usually the first third of the total iterations. Thus the shift combinations other than the best fitting one do not actually contribute to parameter improvement. That is, $(2s_{\max}+1)^R - 1$ trials in the EIS stages and $3^R - 1$ trials in FLS are in vain. One way to avoid these presumably unnecessary trials is simply to stop the search once the shift value estimates have not changed for, say, 50 successive iterations. Since the shift change in FLS is a fractional value and hence can be as small as a computation routine allows, one may set a non-zero criterion value in FLS, for example, to stop when the sum of squared shift changes is less than 10^{-5}.

It is possible that sometimes a few shift estimates change after 50 (or even 100) non-changing iterations. If so, the converged solution resulting from applying the stopping criterion can be considered as a type of local minimum. This termination strategy can result in three or four times as many solutions from different random starts as one might get without it. Thus we suggest trying a few more starting positions when one adopts a stopping criterion, in order to maximize the likelihood of finding the global minimum.

A.4. Rational shift starts

Sometimes one can get a reasonable approximation of shifts by using some data analytic techniques (e.g. cross-correlation maximization [13], time warping [14], etc.). Such shift estimates can be used as rational start values in order to reduce the computation time for the exhaustive trials in EIS. This involves a two-stage SFA, first optimizing the factor weights with shift values constrained at the rational start values, and then optimizing both factor weights and shifts, starting from the final estimates for factor weights given from the first stage. In this way the number of iterations needed in the second stage will be significantly smaller than in a single-stage SFA, so that one can save the computation time to be spent for the exhaustive shift trials in the early stage of iterations in the normal EIS.

In addition to time saving, the rational start for shift values might provide a better chance to avoid getting a locally optimal final solution. An example of the SFA with such shift value constraint is shown in Part III [8] as an assessment for local optimality of the SFA solution for four-way chromatographic data.

A.5. Mode reduction

In cases where one analyzes a four (or higher)-way shifted data array, it may be possible to save time in the SFA by

reducing the number of data modes. This mode reduction procedure requires that multilinearity is well imbedded and reliably estimable in those modes that are not affected by the independent shifting (e.g. the emission and excitation modes in the four-way chromatographic data analyzed in Part III [8]). Suppose we have an $I \times J \times K \times L$ four-way data set in which sequential factors (mode A) are shifted independently across levels of mode B. Multilinearity holds almost perfectly for modes C and D. Then one can obtain a good estimate of the mode C and D loadings by fitting Parafac1 to an $IJ \times K \times L$ unfolded three-way version of the data set in which the original modes A and B are confounded in the first mode. Up to the validity of the mode C and D loading estimates, one can reconstruct an idealized three-way shifted data set where the factor variation of either mode C or D is removed. The time gain by the mode reduction procedure is dramatic (at least proportional to the number of levels in the removed mode), but one must be careful in interpreting a solution from such reconstructed data, since all estimates are conditional on the validity of the factor loading estimates in the removed mode(s).

REFERENCES

1. Harshman RA, Hong S, Lundy ME. Shifted factor analysis. Part I. Models and properties. *J. Chemometrics* 2003; **17**: 363–378.
2. Harshman RA. An index formalism that generalizes the capabilities of matrix notation and algebra to n-way arrays. *J. Chemometrics* 2001; **15**: 689–714.
3. Harshman RA. Shifted factor analysis models and estimation expressed in array notation. *Department of Psychology Res. Bull. #762*, University of Western Ontario, London, ON, 2001. http://publish.uwo.ca/~harshman.
4. Kiers HAL. Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika* 1997; **62**: 251–266.
5. Hong S. Shifted factor analysis: a test of models and algorithms. *Master's Thesis*, University of Western Ontario, London, ON, 1997.
6. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipies in FORTRAN: the Art of Scientific Computing*. Cambridge University Press: New York, 1992; 387–448.
7. Bro R, Anderson CA, Kiers HAL. PARAFAC2—Part II: modeling chromatographic data with retention time shifts. *J. Chemometrics* 1999; **13**: 295–309.
8. Hong S, Harshman RA. Shifted factor analysis. Part III. *N*-way generalization and application. *J. Chemometrics* 2003; **17**: 389–399.
9. Bro R, de Jong S. A fast non-negativity constrained linear least squares algorithm for use in multi-way algorithms. *J. Chemometrics* 1997; **11**: 393–401.
10. Harshman RA, Lundy ME. The Parafac model for three-way factor analysis and multidimensional scaling. In *Research Methods for Multimode Data Analysis*, Law HG, Snyder Jr CW, Hattie JA, MacDonald RP (eds). Praeger: New York, 1984; 122–215. http://publish.uwo.ca/~harshman/lawch5.pdf.
11. Wulfert F, Kok WT, de Noord OE, Smilde AK. Correction of temperature induced spectral variation by continuous piece-wise direct standardization. *Anal. Chem.* 2000; **72**: 1639–1644.
12. Khatri CG, Rao CR. Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhya, Ser. A* 1968; **30**: 167–180.
13. Cattell RB. The structuring of change by P-technique and incremental R-technique. In *Problems in Measuring Change*, Harris CW (ed.). University of Wisconsin Press: Madison, WI, 1963; 167–198.
14. Nielsen NPV, Carstensen M, Smedsgaard J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr. A* 1998; **805**: 17–35.