Special Issue

JOURNAL OF CHEMOMETRICS

Received: 17 August 2008, Revised: 18 October 2008, Accepted: 29 October 2008, Published online in Wiley InterScience: 2009

(www.interscience.wiley.com) DOI: 10.1002/cem.1206

Modeling multi-way data with linearly dependent loadings†

Rasmus Bro[a]*, Richard A. Harshman[b], Nicholas D. Sidiropoulos[c] and Margaret E. Lundy[d]

A generalization/specialization of the PARAFAC model is developed that improves its properties when applied to multi-way problems involving linearly dependent factors. This model is called PARALIND (PARAllel profiles with LINear Dependences). Linear dependences can arise when the empirical sources of variation being modeled by factors are causally or logically linked during data generation, or circumstantially linked during data collection. For example, this can occur in a chemical context when end products are related to the precursor or in a psychological context when a single stimulus generates two incompatible feelings at once. For such cases, the most theoretically appropriate PARAFAC model has loading vectors that are linearly dependent in at least one mode, and when collinear, are nonunique in the others. However, standard PARAFAC analysis of fallible data will have neither of these features. Instead, latent linear dependences become high surface correlations and any latent nonuniqueness is replaced by a meaningless surface-level 'unique orientation' that optimally fits the particular random noise in that sample. To avoid these problems, any set of components that in theory should be rank deficient are re-expressed in PARALIND as a product of two matrices, one that explicitly represents their dependency relationships and another, with fewer columns, that captures their patterns of variation. To demonstrate the approach, we apply it first to fluorescence spectroscopy (excitation-emission matrices, EEM) data in which concentration values for two analytes covary exactly, and then to flow injection analysis (FIA) data in which subsets of columns are logically constrained to sum to a constant, but differently in each of two modes. In the PARAFAC solutions of the EEM data, all factors are 'unique' but this is only meaningful for two of the factors that are also unique at the latent level. In contrast, the PARALIND solutions directly display the extent and nature of partial nonuniqueness present at the latent level by exhibiting a corresponding partial uniqueness in their recovered loadings. For the FIA data, PARALIND constraints restore latent uniqueness to the concentration estimates. Comparison of the solutions shows that PARALIND more accurately recovers latent structure, presumably because it uses fewer parameters and hence fits less error.

Keywords: PARAFAC; PARATUCK2; constrained ALS estimation; uniqueness; linear dependence; equality constraints; multi-mode factor analysis

1. INTRODUCTION

Many methods have been proposed for multivariate curve resolution and more generally for factor or component modeling of (multi-way) data such as constrained Tucker3 [1], PARAFAC [2,3], INDSCAL/CANDECOMP [4], window factor analysis [5], and multivariate curve resolution [6]. In this paper, we propose a class of methods (which we call PARALIND for PARAllel profiles with LINear Dependences) that are specially designed to deal with complications arising from linearly dependent factors. They do this by preserving the component dependences (and thus any partial uniqueness properties) that characterize the latent structure of the data being analyzed.

When it is known on theoretical or other grounds that the latent PARAFAC structure involves linearly dependent factors, use of PARALIND to incorporate this information into the analysis will 'strengthen' the results in at least two ways: it will improve the probable accuracy of the factor estimates by improving the data-degrees-of-freedom per parameter estimated; and it will improve the theoretical appropriateness of the factor estimates by making their properties correspond more closely to those implied by the scientific model of the data generating process. Note that both benefits are obtained *whether or not the latent structure has a unique solution*.

* *Correspondence to: R. Bro, Department of Food Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark.*
E-mail: rb@life.ku.dk

a *R. Bro*
Department of Food Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

b *R. A. Harshman*
Psychology Department, University of Western Ontario, London, Ontario, Canada N6A 5C2

c *N. D. Sidiropoulos*
Department of Electronic and Computer Engineering, University Campus—Kounoupidiana, 731 00 Chania—Crete, Greece

d *M. E. Lundy*
Psychology Department, University of Western Ontario, London, Ontario, Canada N6A 5C2

† *This paper is a result of over 10 years of cross-disciplinary collaboration. Starting from the first draft back in 1998, this has been work in seemingly perpetual progress, as new insights and new interpretations were coming up every few years. When Richard Harshman passed away, the paper was just about to take a new direction due to some interesting experimental results and insights that Richard obtained. Unfortunately, that part was never finalized. Still, this paper is largely a result of Richards persistence and eagerness to keep digging deeper and deeper into the details to gain a complete understanding of observed results. His signature, thinking process and way of writing are imprinted on this paper.*

This work has implications for many different areas in which multi-way analysis is employed. For example, linear dependences arise naturally in handling certain types of radiative energy transfer in fluorescence spectroscopy, as described, e.g. by Ross and Leurgans [7]. In signal processing/communications applications, data with PARALIND structure arise, for example, in certain localization problems due to what is known as specular multipath interference, where identical signals appear with different delays (see, e.g.Reference [8]). In this case, by capitalizing on additional problem-specific structure, the associated PARALIND solution can be shown to be unique [9,10]. In some statistical applications (e.g. in Psychology or Environmental Science) data may be so scarce and/or expensive to collect that investigators are forced to analyze three-way arrays for which it is known in advance that certain of the factors will not show the independent variation over the 'third mode' required for full uniqueness.

1.1. Rank deficiency

With three-way data, it becomes possible for patterns generated by the underlying sources of variation to have independent effects in two modes yet nonetheless be linearly dependent in a third mode. When such linear dependences exist in the *latent* factor structure, the most appropriate PARAFAC solution would show these same dependences in the *recovered* factors. Such a solution could be called rank deficient in the sense that the component matrices for one (or even several) mode(s) would have less than full column rank. However, the solution obtained by standard PARAFAC will never have this property because data noise causes the estimated loadings for collinear latent factors to become linearly independent (though usually still quite correlated).

Such theoretical or latent rank deficiency is a common phenomenon in some types of data, as known from literature [11,12] and demonstrated in the example section. The reason can be either mathematical (e.g. preprocessing steps such as centering), physical (e.g. some analytes appear in identical proportions throughout an experiment) or combinations thereof (e.g. centering data that suffer from closure) [13,14].

To introduce and explain the modeling issues involved, we can start with a simple two-way example. Consider a chemical reaction $M+N \rightarrow P$ for which spectral data are obtained over time. If M and N have the same initial concentration, their subsequent concentration profiles will be identical (perfectly correlated over time). Yet, there are two different spectra (one for M and one for N). This illustrates how a given set of underlying phenomena can produce variation in the spectral domain and in the concentration domains that have different ranks. Some authors view this as showing that there is only one phenomenon evolving in the concentration/time domain—the reaction of M and N [11,15,16]. However, this perspective forces one to consider the spectral domain and time domain at different 'levels of description'—one for chemicals, the other for phenomena. One could view both modes at the same (chemical) level of description by allowing the rank-2 concentration domain to be described in terms of three concentration curves (say, length I vectors $\tilde{\mathbf{a}}_1$ and $\tilde{\mathbf{a}}_2$ and $\tilde{\mathbf{a}}_3$) but making explicit the fact that two of them are linked and have identical shapes. We will use the $\sim$ to signify the actually perceived profiles in order to be able to distinguish them from the 'un-tilded' profiles that are the ones estimated. Only two distinct concentration profiles are then needed (say, vectors $\tilde{\mathbf{a}}_1$ and $\tilde{\mathbf{a}}_2$), since the concentration profiles of analyte M and N, ($\tilde{\mathbf{a}}_2$ and $\tilde{\mathbf{a}}_3$**)** will be identical. Letting

$$\mathbf{A} = [\tilde{\mathbf{a}}_1\, \tilde{\mathbf{a}}_2] \tag{1}$$

and letting the $J \times 3$ matrix **B** hold the spectra of P, M and N, then a model of the measured spectra held in **X** ($I \times J$) can be written

$$\mathbf{X} = \mathbf{AHB}^{\mathrm{T}} \tag{2}$$

where the matrix **H** is a two by three binary matrix which, in this case, reads

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \tag{3}$$

1.2. Improving recovery of latent structure by explicit representation of dependences

In the above example, the number of parameters to be estimated in Mode A is reduced by 33% without introducing any new ones elsewhere (since the 1s in **H** are not estimated). As a result, there is less fitting of error, which improves recovery of the latent structure. Recovery is also improved because the factor dependency at the latent level (known to exist on theoretical or other grounds) is explicitly maintained in the estimated factor loadings. These advantages will be demonstrated below when PARALIND is applied to real data.

The 'dependency matrix' **H** will sometimes be referred to as an 'interaction matrix' because of its similarity to the core matrix in the Tucker3 model [17]. From a Tucker perspective, the rank deficiency can be interpreted as arising from 'interactions' between factors in a particular column of **A** and factors in a different column of **B**. Chemically though, it is arguably more natural to think of the rank deficiency as simply a (linear) dependency between distinct factors. Then, the physical interpretation is still that there are three (concentration) profiles, one for each analyte; the model simply specifies that two of these profiles are identical, through the relation

$$\tilde{\mathbf{A}} = \mathbf{AH} = [\tilde{\mathbf{a}}_1 \quad \tilde{\mathbf{a}}_2 \quad \tilde{\mathbf{a}}_2] \tag{4}$$

Thus, there are alternative ways one might look at the rank deficiency. In one view, the factors in **A** represent *two phenomena in the first mode that have mathematical 'interactions' with three in the second mode*, as defined through **H**. In the other view, a more chemical interpretation is taken. The factors in $\tilde{\mathbf{A}}$ represent *three phenomena* in the first mode (which are the same three that are represented by factors in the other modes), while the product of **A** and **H** describes *the structure of patterns they produce* in terms of a set of distinct patterns **A** and their linearly dependent relations to the observed patterns. Introducing the linear dependences in an explicit, well-defined way is a key feature of our approach to developing a model for handling rank deficiency.

1.3. Three-way solutions required

It is not, in general, possible to fit such a linearly dependent model for a two-way problem. Because of the intrinsic equality of matrix row-rank and column-rank, only two spectral profiles can

be determined. That is, the rank of a dataset such as that described above will be two (plus small factors due to random noise and other errors, which we ignore here) and hence only two independent factors can be found for the row-mode as well as for the column-mode. Usually, ways to circumvent rank deficiency in two-way problems are based on either rotations of the solution using prior knowledge of the data or data augmentation [11] where the closure and/or other linear dependences are 'broken' by adding samples of different constitution. However, by extending models such as that in Equation (2) to three-way data, it will become possible to handle the rank deficiency directly and in a rational way.

Before proceeding further, we note that essentially the same approach, though in less general form, has been proposed earlier by several authors. One version was introduced in Bro [18] and Smilde *et al*. [14], where it was developed as a restricted PARATUCK2 model (PARATUCK2 is described in References [3,19]). It was also independently proposed by Harshman [20] and referred to as 'Structured PARAFAC'. Finally, Reference [8] independently proposed a PARALIND-type model for handling certain problems in signal analysis.

The PARALIND family of models is related to both of the larger general families of multi-way models: the Tucker and PARAFAC families. For example, just as PARAFAC can be viewed as a constrained case of Tucker3, so can PARALIND [1]. However, in this paper PARALIND's derivation and its distinct approach to substantive scientific (e.g. chemical) interpretation is as a constrained PARAFAC model. It is also useful to view the various constrained PARAFAC models described here as members of an identifiable model (sub)family, because they all share certain distinctive uniqueness properties and practical statistical benefits.

1.4. Outline of paper

The rest of this paper is organized as follows. In Section 2, the basic PARALIND approach that we have conceptually introduced in two-way terms will be developed more rigorously by means of explicit three-way formulations. In Section 3, the approach is demonstrated more concretely by developing two specific PARALIND models, one for a flow injection dataset and one for a fluorescence dataset. The actual results of applying the models are discussed in Section 4. An algorithm for fitting the PARALIND model is developed in Appendix A and some specific proofs of uniqueness for the application are given in Appendix B.

2. THEORY

2.1. Dependences in three-way arrays

Assume that a three-way data array $\underline{\mathbf{X}}$ ($I \times J \times K$) is given for which an S-component PARAFAC model holds. For a typical slice, say, level k of the third mode, the standard 'typical slice' way to write this is

$$\mathbf{X}_k = \tilde{\mathbf{A}}\mathbf{D}_k\mathbf{B}^{\mathrm{T}} + \mathbf{E}_k \qquad k = 1, \ldots, K \tag{5}$$

In the above, the first mode loading matrix is denoted by $\tilde{\mathbf{A}}$ ($I \times S$) to distinguish it from another loading matrix to be introduced shortly. The PARAFAC model may also be written in matricized form [18] as[1]

$$\mathbf{X}^{(I\times JK)} = \tilde{\mathbf{A}}(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}} + \mathbf{E} \tag{6}$$

where the operator $\odot$ is the Khatri-Rao product [21], which is equivalent to a columnwise Kronecker product.

Now consider an example involving factor dependency. Assume that, say, a four-component model is valid, but that the loading matrix $\tilde{\mathbf{A}}$ is rank-deficient (has rank 3) because $\tilde{\mathbf{a}}_3$ and $\tilde{\mathbf{a}}_4$ are identical; we assume that $\mathbf{B}$ and $\mathbf{C}$ have full column rank (in this example, rank 4). In terms of our hypothetical chemical interpretation, this could represent a dataset involving four chemical analytes that have different emission (Mode B) and excitation (Mode C) spectra, but the same time profiles (Mode A) for the third and fourth analytes.

This is a *three-way* dataset, and the rank of a three-way array can exceed the rank of the fiber (e.g. column) spaces of its individual modes [see. e.g. Reference [22]]. The rank of this array will be four (before addition of random error), requiring a four factor PARAFAC solution. With error free data, the solution would recover the first two factors exactly and uniquely (provided they showed the requisite independent variation in all three modes), and it would also reveal that two of the factors have identical loading patterns in Mode A, which would alert us to a uniqueness problem for these two factors. As will be explained in the discussion of partial uniqueness below, the proportionality of factors three and four in Mode A will prevent PARAFAC from uniquely resolving these two factors in Modes B and C. However, the other two factors will still be uniquely recovered—provided, of course, they show the independent variation in all three modes generally needed for uniqueness.

At this point, it seems that the analysis problem may be at least partly solved. There is unfortunate loss of information due to the mixing of Mode B profiles of factors three and four, and also their Mode C profiles, which is due to the nonuniqueness, but at least we are able to obtain a four factor solution which has the same level of description in all modes—as was our goal earlier in the two way example but which was infeasible there because of rank problems. However, because of the rank deficiency, there is a subtler analysis problem that remains, one that (to our knowledge) has previously been overlooked.

The problem arises from the disturbance of the linear dependences by the (inevitable) presence of error and random noise in the data. The linear dependences of the latent source patterns will not be reflected in the recovered loadings (though factors may appear highly correlated) and the nonuniqueness will be lost. Even when the theoretical model of the generating process call for linear dependences, it cannot be enforced by standard PARAFAC. Put in another way, unless PARAFAC is 'informed' about the need for dependency (via PARALIND), noise will inevitably lead to actual PARAFAC solutions which are not rank-deficient. The factor matrices that should physically be rank-deficient will obtain full rank and non-sensible uniqueness because PARAFAC will also fit some of the noise part of the data. This solution will be technically unique, since there will normally always be one set of loadings which happens to best fit the systematic variation *plus* the random error variation better than any other. However, the resulting position of the estimates of $\tilde{\mathbf{a}}_3$

[1] *For further discussion of these two kinds of representation, as well as Bro's superscript notation for unfolding, and related issues, see Harshman and Hong [44].*

and $\tilde{\mathbf{a}}_4$ axes will be specific to the error that perturbed this particular set of observations, and will not replicate in a new sample of data. This phenomenon is what we call 'surface' (surface-structure-determined) rather than 'deep' (deep-structure-determined) uniqueness. It is important to minimize 'surface' uniqueness and PARALIND provides an explicit way to do this.

2.2. PARALIND: a new family of PARAFAC models that explicitly represent rank deficiency

By introducing a new matrix, **H**, which is called a *dependency matrix* (from a PARALIND perspective) or an *interaction matrix* (from a Tucker perspective), rank deficiency can be explicitly incorporated into the three-way model in a concise and parsimonious way. If the rank of $\tilde{\mathbf{A}}$ is R ($\leq S$) then it holds that $\tilde{\mathbf{A}}$ may be expressed

$$\tilde{\mathbf{A}} = \mathbf{AH} \tag{7}$$

where **A** is an $I \times R$ matrix and **H** is and $R \times S$ matrix, and the model for the data becomes

$$\mathbf{X}_k = \mathbf{AHD}_k\mathbf{B}^{\mathrm{T}} \tag{8}$$

If there are four different components in the above example then $S = 4$. The situation with $\tilde{\mathbf{a}}_3 = \tilde{\mathbf{a}}_4$ will yield

$$\mathbf{A} = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,] \tag{9}$$

and

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{10}$$

It directly follows that

$$\tilde{\mathbf{A}} = \mathbf{AH} = [\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3\ \mathbf{a}_3] = [\tilde{\mathbf{a}}_1\ \tilde{\mathbf{a}}_2\ \tilde{\mathbf{a}}_3\ \tilde{\mathbf{a}}_4] \tag{11}$$

and thus the third and fourth column of $\tilde{\mathbf{A}}$ are forced to be identical. This example of a dependency or interaction matrix is quite simple. As a second example, consider a chemical process in which factor 4 values depend on the joint contribution of two other factors; the loading matrix could then have the form

$$\tilde{\mathbf{A}} = \mathbf{AH} = [\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3\ (\mathbf{a}_1 + \mathbf{a}_2)] = [\tilde{\mathbf{a}}_1\ \tilde{\mathbf{a}}_2\ \tilde{\mathbf{a}}_3\ \tilde{\mathbf{a}}_4] \tag{12}$$

with the associated **H** matrix

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{13}$$

More complicated relations can, of course, be envisioned allowing for multiple dependences involving additional factors, such as

$$\begin{aligned} \tilde{\mathbf{A}} = \mathbf{AH} &= [\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3\ \mathbf{a}_4\ (\mathbf{a}_1 + \mathbf{a}_2)\ (\mathbf{a}_1 + \mathbf{a}_3 + \mathbf{a}_4)] \\ &= [\tilde{\mathbf{a}}_1\ \tilde{\mathbf{a}}_2\ \tilde{\mathbf{a}}_3\ \tilde{\mathbf{a}}_4\ \tilde{\mathbf{a}}_5\ \tilde{\mathbf{a}}_6] \end{aligned} \tag{14}$$

These last two examples are provided not just for added clarity, but because they have different properties. They are examples of structures that include dependences yet retain uniqueness of decomposition. *(For succinctness we will use 'uniqueness' without the qualifier 'essential' but ask the reader to always recognize that there is a harmless indeterminacy of column order and relatively benign indeterminacy of column scaling in all the multiway factor models.)*

The PARALIND principle can also be applied in more complicated situations where there are linear dependences in more than one mode, and it can be extended to higher-way arrays; many of these extensions are relatively straightforward, but would take us beyond the scope of the current paper.

The **H** matrix can also become more general, and used in more complex ways. Intermediate situations may be envisioned where, for example, it is known that certain interactions are non-existent whereas the strength of the remaining ones is to be estimated. In some exploratory applications, the dependency matrix **H** need not even be predefined. This matrix, which defines the pattern and strength of the interactions, may also be estimated from the data if no prior knowledge is available. The approach would then be more similar to the PARATUCK2 model introduced by Harshman and Lundy [3,19]. This is the approach to dependences taken by Bro [18] and Smilde *et al.* [14].

2.3. Effects of linear dependences on uniqueness

As demonstrated above, constructing three-way models with linear dependences can make it possible to use distinct factors to *directly* represent theoretically distinct—but not linearly independent—sources of variation, even in cases where two-way models cannot. However, there are sometimes limitations on the information that can be recovered: certain kinds of linear dependences will limit or reduce the uniqueness of the obtained solutions. Because of the importance of uniqueness in the scientific use and interpretation of PARAFAC and PARALIND, the relation between different loading-vector dependences and different levels of uniqueness needs to be carefully considered. In particular, we need to consider a phenomenon called 'partial uniqueness' that can result from some kinds of linear dependences.

We discuss the phenomenon in terms of PARAFAC. However, we believe that the PARAFAC and PARALIND models will have the same properties 'at the latent level', i.e. if one decomposes an *error free* array into the correct number of components. This is because the PARALIND model would also be a best-fitting PARAFAC model, and so the constraints imposed by PARALIND would be 'inactive'. It is only with *fallible* data (i.e. any 'real' data) that the constraints become active and the difference becomes important; this will be discussed in Section 2.5.

The PARAFAC model is neutral with respect to linear independence or dependence among the columns of factor loadings in any mode; it is simply defined as a sum of n-way outer products which approximates a given array, with no special restrictions placed on the vectors of factor loadings used to generate these outer products. In many cases, linearly dependent loading vectors (e.g. columns of **A**, the set of vectors $\{\mathbf{a}_1, \mathbf{a}_2, \ldots\}$) will be required, as when the rank of the array exceeds the number of levels in any (or even all) of its modes (e.g. see References [22,23]). In this case, the factor-loading matrix necessarily involves linear dependences since it will have more columns than rows. (Even linear independence of the rank-1 outer products, the $\mathbf{f}_r = \mathbf{a}_r \otimes \mathbf{b}_r \otimes \cdots$, is not required by the model definition; more outer products than needed to exactly reproduce a given array could constitute a 'legal' PARAFAC model, and it might even be meaningful in certain contexts, e.g. if additional observations could require the added dimensions).

In some cases, linear dependences among factor loading vectors will affect the uniqueness of the decomposition, while in other cases, it will not. Solutions with multi-factor dependences, as in Examples 2 and 3 (Equations (12) and (14)) can retain standard PARAFAC uniqueness properties: loadings of all factors in all modes would still be determined up to column order and scaling—providing, of course, the factors otherwise meet the usual requirement of adequate distinct variation in all three modes. However, solutions with simple factor collinearity (as in our three-way Example 1, above) will lose uniqueness of the factors that are collinear.

A precise description of the nonuniqueness in Example 1 would be as follows: any nonsingular linear transformation **T** could be applied to the Mode B loadings for factors 3 and 4, e.g.

$$\left[\widehat{\mathbf{b}}_3\ \widehat{\mathbf{b}}_4\right] = [\mathbf{b}_3\ \mathbf{b}_4]\mathbf{T} \tag{15}$$

where **T** is a nonsingular 2×2 matrix, so long as the inverse transformation is applied to Mode C (i.e. $\left[\widehat{\mathbf{c}}_3\ \widehat{\mathbf{c}}_4\right] = [\mathbf{c}_3\mathbf{c}_4]\mathbf{T}^{-1}$); the Mode A loadings remain the same (i.e. two copies of the shared collinear pattern of variation). The remaining factors will be unique in all three modes, just as they would in a PARAFAC solution with no dependences. This is an example of the phenomenon referred to as 'partial uniqueness', in which only some factors (in this example, factors 1 and 2) are uniquely determined in all modes and others have rotational freedom.

2.4. Rules governing partial uniqueness

Because PARALIND is specifically performed when some factors are expected to be linearly dependent in some modes, a general understanding of the impact of dependences on uniqueness and partial uniqueness is desirable. This section reviews some relevant aspects of what is known at this point, including some unpublished results from our previous and ongoing study of uniqueness and partial uniqueness. In Section 2.5 we return to the question of PARALIND itself, and the advantages of PARALIND over PARAFAC for cases involving latent dependences, with or without uniqueness.

The phenomenon of partial uniqueness has been of interest ever since the original PARAFAC monograph (References [2], p. 35–44; [24]), and interest has continued to this day (e.g. see Reference [25]). Some mathematical grounds for partial uniqueness were noted briefly in Reference [24], but the full implications were not fully developed. To illuminate the issue from a different direction, we present in Appendix B an analysis of partial uniqueness based on eigen-decomposition[2]. It is a constructive proof, based on a two-slice approach as in generalized rank annihilation [26–28] and ESPRIT [10,29,30].As noted earlier, we believe that the uniqueness and partial uniqueness properties of a PARALIND model should follow directly from the uniqueness and partial uniqueness properties of the corresponding PARAFAC model in the case of an *error free* array in which the rank-one patterns have the linear dependences specified in the PARALIND **H** matrices.

2.4.1. *When two modes have full column rank*

The case where factor-loading matrices in at least two modes have full column rank is better understood than cases where only one or no modes have this property. Fortunately, this corresponds to many, possibly most, 'real world' three-way cases. In these cases, the property governing uniqueness turns out to be remarkably simple: all that matters is whether or not factor loading vectors are proportional (i.e. collinear) in that third mode (see, e.g. References [24] or [31] for direct discussion of this point).

To aid our intuition, we can translate our algebraic discussion of partial uniqueness into a simple geometric description of partial uniqueness, formulated in terms of 'subspaces'. This can be done as follows: Let A and B designate the two modes with factor loading matrices of full column rank. If the Mode C loading matrix has no proportional columns, all factors in the solution are uniquely determined. If, however, there is a set of Mode C loading vectors that are proportional to one another, it identifies a nonunique set of factors. The Mode A loading vectors for those factors span a *transformable subspace* of the Mode A factor space, and their Mode B vectors span the *dual transformable subspace* in the Mode B factor space. The factors in either subspace can be given a non-singular linear transformation, so long as the inverse transformation is applied to the factors in the other. Any factor not in that set (i.e. whose Mode C loading vector is not proportional to those for members of that set) has A and B loading vectors that lie outside of their (respective) transformable subspaces. In the combinations of loading vectors appearing as factors in alternative solutions, these vectors lying 'outside' that subspace will never be combined with any of those belonging to factors 'inside' that subspace.

Similarly, if there is a second set of factors with loading vectors proportional to one another in C, but not proportional to those in the first set, these vectors span a second transformable subspace in each mode. While factors in both subspaces will be nonunique within their respective subspaces, those in one subspace will never be linearly combined with those in the other in any solution. The subspaces remain distinct and are uniquely determined *as subspaces*.

It follows that a factor which is nonproportional to all others in C will be in its own distinct 'subspace' and will remain 'separate' from the other factors across all alternative solutions. That factor will be distinct in all decompositions and hence uniquely determined. A factor in any decomposition that is nonproportional to some but not all of the other factors will be 'incompletely resolved' in alternative decompositions, reliably distinct from all and only those factors to which it is nonproportional.

In summary, then, two factors of a given initial solution that lie in different transformable subspaces will be 'separated' from one another in not only that solution but in all solutions, but will be 'incompletely resolved' relative to the total factor set, since they will not be reliably separated from other factors in their subspace. (Although stated here for the three-mode case, these results carry over to arrays of all orders, provided there are two modes of the array in which the factor loading vectors have full column rank.)

2.4.2. *When only one or no modes have full column rank*

The uniqueness properties of factor structures where dependences exist in two or even all three modes of a three-way array are not as well understood. There are some general uniqueness results that have been obtained for restricted Tucker3 models [16] which could be relevant for some PARALIND models. Also, there are empirical approaches for testing uniqueness that have been suggested for multilinear models [32] and these apply for PARALIND models as well. The uniqueness properties of a complicated PARALIND model could often be inferred with

[2] *A similar application of an eigendecomposition approach can be found in 10 Berge [25].*

reasonably high probability from empirical results of experiments applying PARAFAC to error free synthetic datasets.

However, we do not have a full characterization of uniqueness conditions for the case in which only one mode or no modes have loading matrices with full column rank.

We have been investigating such cases both mathematically and empirically, and there are several principles which we can tentatively offer to provide some guidelines; these convey a general impression of the kind of restrictions and relations that appear to exist.

(1) If a factor has loading vectors that are linearly independent of all other factors in at least three modes, that factor will be uniquely determined. (2) If two factors have collinear loadings in one mode of a three-mode array (or *N*-2 modes of an *N*-way array), they will not be uniquely determined; their loadings will not be 'resolved' or separable in the two non-collinear modes; in some circumstances, they may also become mixed with other factors in the collinear mode. (3) If two factors have collinear loadings in two modes of a three-mode array (or *N*-1 modes of an *N*-way array), they will collapse into a single factor.

We also conjecture that there are general principles similar to the following: (4) If a factor is a linear combination of certain other factors in two modes, and these dependences have the same form in both modes, it will not be separated from those other factors in some modes; (5) if no factor has linearly dependent loadings in two modes, and no mode has collinear loading vectors, the solution should be unique. (6) If a subset of columns satisfies the Kruskal criterion (has a sum of *k*-ranks of $2R+2$, where R is the number of factors involved in the subset), then that subset will be uniquely determined. More generally, it appears likely that uniqueness and partial uniqueness properties of subsets of factors—with respect to their separation from one another within the subset can be obtained by applying the same criteria to that subset that one would apply if they constituted a separate dataset and solution. Note, however, that these conjectures all assume that the factor structure in question, here often referred to as 'the solution', has the *minimum* number of factors sufficient to provide the resulting array product: the factor set or subset does not have redundant components. Note, however, that these conjectures all assume that the factor structure(s) in question, here often referred to as 'the solution', has (have) the minimum number of factors sufficient to generate the corresponding array product(s): the number of factors is equal to the resulting array rank.

It also seems, from numerical results, that when linear dependences involve 'enough' factors (have adequately high k-rank), they need not interfere with uniqueness. A factor may have loading vectors that are linearly dependent on other factors in any or all modes, and yet still be uniquely determined, if the dependences are 'adequately complex'. For example, this will happen (i) if each linearly dependent vector of loadings includes contributions from a vector that is not involved in other linear dependences in any mode or (ii) if it includes contributions from enough other vectors—for example, this will happen if each linearly dependent vector of loadings includes contributions from a vector that is not involved in other linear dependences in any mode, or if it includes contributions from enough other vectors.

The earliest example of unique solutions with linear dependences among factors in all modes is from Reference [2] (subsequently noted by Kruskal and others), in which a unique 10 factor solution was obtained from an $8 \times 8 \times 8$ dataset. However, we now know that this did not come close to the upper bound for this kind of case. Some recent empirical work indicates that much higher-rank unique solutions are possible, even in cases where one or more modes have low k-ranks (e.g. *k*-rank of 2 in one mode) by carefully constructing the factor loading matrices (e.g. so as to avoid the same linear combination in the same columns across modes). We have built $8 \times 8 \times 8$ arrays with PARAFAC structure, in which the dependences reduce the *k*-rank in each mode to two, (and so the Kruskal criterion only assures unique determination of two factors), and yet it was possible to uniquely recover more than 16 factors! Even higher numbers of unique factors can be obtained if one does not impose any added dependences above those implicit in the more-columns-than-rows nature of the factor loading matrices. For the $8 \times 8 \times 8$ case, recent unpublished experiments suggest that the maximum rank of a unique solution is 23, which is just one short of what is believed to be (and numerically seems to be) the 'typical [maximum] rank' of $8 \times 8 \times 8$ arrays. So far, the 24-factor PARAFAC structures constructed for $8 \times 8 \times 8$ arrays have been found to have multiple perfect fitting solutions, and thus appear to be nonunique. However, these arrays *do* appear to have a true rank of 24, since they have not been perfectly fit in any attempts at finding 23-factor solutions. In contrast, $8 \times 8 \times 8$ arrays constructed with 25 factors can apparently be perfectly fit with 24, suggesting that 24 is the maximum rank of $8 \times 8 \times 8$ arrays.

2.5. How PARALIND improves the characteristics of the solution

The advantages of PARALIND are not related to uniqueness. They have to do with better recovery of latent structure and more appropriate characteristics of the solution. These result from a 'better informed' PARAFAC model, one that incorporates information about known or expected dependences in the latent structure.

As noted earlier, there is a mathematical equivalence between PARALIND and PARAFAC decompositions of the same error-free array, and so we can use existing knowledge about PARAFAC uniqueness when considering the 'latent structure' properties of PARALIND. This equivalence does not hold when the models are fit to fallible (error contaminated) data, however. Here, the special properties of PARALIND can be a distinct advantage, and these will be the focus of this section.

In PARALIND, linear dependences are incorporated and fit in such a way that the theoretical *dependences* among PARAFAC loadings hypothesized for the latent structure also hold among the *factor loading estimates*. This is one of the senses in which PARALIND can be viewed as constrained PARAFAC. A consequence of this is that when dependences introduce *partial uniqueness* of latent model parameters, these are reflected in corresponding partial uniqueness of their *fitted parameter estimates*.

For example, consider the factor structure discussed earlier whose dependency relations are given in Equation (10) and Mode A loadings are given in Equation (11). As was pointed out in Section 2.3, the collinearity of the last two columns of Mode A loadings introduces a partial nonuniqueness of representation: the decomposition will still be exact for any alternative solution in which a nonsingular linear transformation $\mathbf{T}$ is applied to factors 3 and 4 (i.e. $\left[\widehat{\mathbf{b}}_3\ \widehat{\mathbf{b}}_4\right] = [\mathbf{b}_3\mathbf{b}_4]\mathbf{T}$), so long as this is compensated for by an inverse transformation in Mode C. The additional point to be made here is that the PARALIND factor loading *estimates*

obtained by fitting this model to fallible data exhibit the corresponding nonuniqueness $[\hat{\hat{\mathbf{b}}}_3 \hat{\hat{\mathbf{b}}}_4] = [\hat{\mathbf{b}}_3 \hat{\mathbf{b}}_4]\mathbf{T}$, where the ^ operator means the estimate. The alternative versions of the loadings estimates or fitted values produce exactly the same approximation of the data and consequently, have the same residuals and overall fit.

With regular PARAFAC, on the other hand, this does not happen. PARAFAC loadings give a *better* fit value for the specific dataset than the PARALIND solution does, because it better 'capitalizes on' fitting the random error variance. Some of the improvement in fit to the random error is due to distinct perturbations introduced into the factor loading vectors $\tilde{\mathbf{a}}_3$ and $\tilde{\mathbf{a}}_4$, which as a result are no longer strictly collinear. In this respect, the fitted approximation to the data is less theoretically appropriate or correct, since linear dependence between these two loading vectors is to be expected because of the processes being modeled.

Also, for PARAFAC, linear transformations of loadings in Modes B and C do not produce alternative factor loading estimates in Mode B and C that have identical fit (i.e. $\text{fit}([\hat{\hat{\mathbf{b}}}_3 \hat{\hat{\mathbf{b}}}_4]) \neq \text{fit}([\hat{\mathbf{b}}_3 \hat{\mathbf{b}}_4])$), because PARAFAC's $\hat{\mathbf{b}}_3$ and $\hat{\mathbf{b}}_4$ incorporate perturbations which make these loading vectors independent, and which (in conjunction with perturbations incorporated into $\tilde{\mathbf{a}}_3$ and $\tilde{\mathbf{a}}_4$) maximally fit the random error specific to that dataset.

The PARAFAC model is full rank in all modes and has standard PARAFAC uniqueness, but the unique factors are oriented arbitrarily by the fitting of the random error. This phenomenon has been observed repeatedly with real data, and demonstrated with synthetic data involving collinear true structure plus random error. (Sometimes, there are several different, arbitrarily rotated, solutions which appear from different random starting positions, all highly collinear, with very similar but non-identical residuals and overall fit; this is one way that it is sometimes possible to detect latent-level nonuniqueness.)

In either case, the superior fit value of the 'optimal' PARAFAC solution is misleading. The loading perturbations optimally fit the specific random error in the given sample or dataset, but these perturbed loadings will not be the best fit for a new sample or experimental dataset, since the new one will of course have different random error. In fact, on average, the 'poorer fitting' PARALIND solution has better cross-validation, because it uses fewer parameters and so absorbs less error.

More important, however, is the fact that PARALIND, by using fewer parameters, provides better recovery of latent structure. This can easily be demonstrated with synthetic data, but we will do even better and demonstrate it with the real data of our example applications. A second advantage is that by blocking 'surface level' uniqueness of the parameter estimates (i.e. uniqueness due to error-fitting, rather than uniqueness that reflects latent structure), the analyst is able to assess the nature and range of nonuniqueness directly, by comparing solutions obtained from different starting points.

3. APPLICATIONS

3.1. Fluorescence excitation-emission data

As our first example, we report the results of an experiment set up specifically to test PARALIND. Twenty three different chemical samples were created, containing different amounts of the four fluorophores cathecol, hydroquinone, indole, and L-tryptophane. Flourescence excitation-emission matrices (EEMs) were then obtained for each sample (and this analysis was repeated four more times to allow later assessment of reliability). To simulate circumstances where it is not possible to get independent concentration profiles for all analytes, the concentrations of cathecol and hydroquinone were made identical (Table I). Typical landscapes for these samples are shown in Figure 1. To avoid problems from scatter, an EEM of pure solvent (water) was subtracted and emission at slightly (10 nm) below and above the excitation frequency was set to missing [18,33,34]. Each sample was measured in five replicates and hence for each sample five EEMs are available.

3.1.1. *Background for the experiment*

It is well known that EEMs can be well approximated by PARAFAC (unless concentrations are too high, leading to inner filter effects). This is because all parts of the flourescence spectrum for a given compound show the appropriate proportional variation (i.e. with excitation frequency, emission frequency, and concentration) to make its pattern of 'latent contributions' consistent with the contributions of a single PARAFAC factor [7,18,31,35]. In the EEM data used here, two factors will have identical loading profiles in the Concentration Mode (here, Mode A), which we know from Section 2.4.1 will make these factors nonunique with respect to one another. While this was set up deliberately for our experiment, such unfortunate dependence could arise realistically in various ways, e.g. in standard addition or for biochemical reasons.

Table I. Concentrations of analytes in the fluorescence dataset (scale is arbitrary)

Sample	Cathecol	Hydroquinone	Indole	Tryptophane
1	15	15	0	10
2	0.5	0.5	0.5	10
3	20	20	15	15
4	15	15	0.5	20
5	0.5	0.5	10	20
6	0.5	0.5	15	20
7	10	10	0	10
8	0	0	0.5	0
9	15	15	10	10
10	0	0	0	0.5
11	15	15	15	0
12	10	10	15	0
13	0.5	0.5	20	20
14	10	10	0.5	15
15	0	0	10	0
16	0	0	15	15
17	20	20	0.5	0.5
18	10	10	10	0.5
19	20	20	10	20
20	20	20	0	0
21	20	20	20	0.5
22	0	0	20	10
23	10	10	20	15

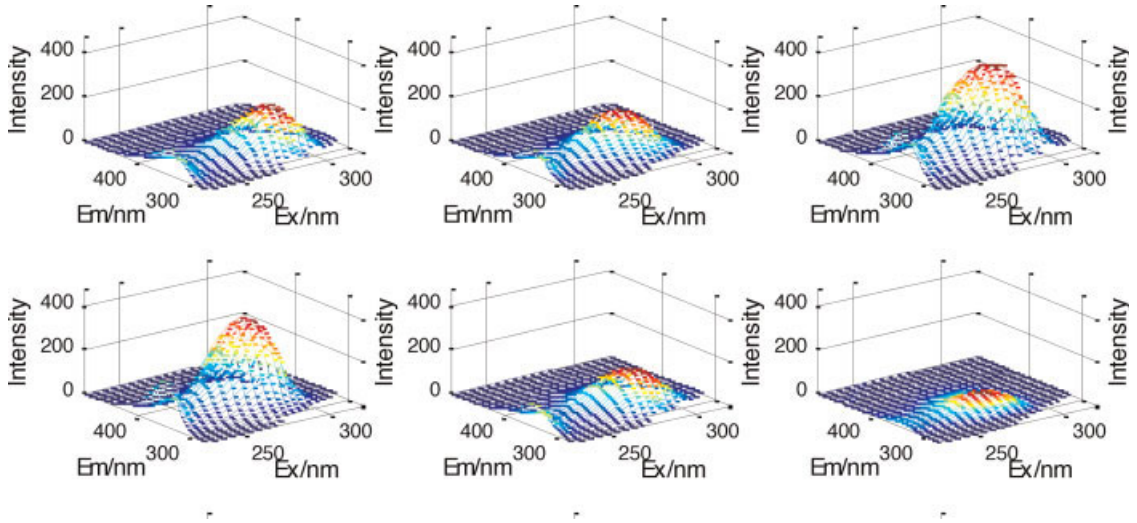

Figure 1. Plots of the fluorescence excitation-emission data of the first six samples. This figure is available in color online at www.interscience.wiley.com/journal/cem

Because of the dependency in Mode A, these data should be suitable to test whether the theoretical advantages of a PARALIND analysis, compared to an application of unconstrained PARAFAC, can be demonstrated with real data. As stated in Section 2.5, these advantages should include less error-fitting due to a better degrees-of-freedom ratio, and restoration of exact collinearity and consequent nonuniqueness in the parameter estimates for components that have these properties at the latent level, thus facilitating empirical assessment of the range and nature of nonuniqueness in the recovered information.

3.1.2. *Constructing the model*

The appropriate PARALIND model for these data would use a Mode A dependency matrix of the form

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{16}$$

thus making the fact that there are two collinear loading patterns in that mode a part of the model. (Of course, the model does not directly specify which two analytes are dependent, but the maximization of fit ensures that the appropriate loading patterns end up being assigned to the appropriate pairs of columns.) This dependency relationship is the same as the one for the first example discussed in Section 2.2.

Four-factor PARAFAC and PARALIND models, with and without nonnegativity constraints, will be applied to these data.

3.1.3. *Anticipated uniqueness properties*

The latent structure of the EEM data means that theoretically, without nonnegativity constraints, both PARAFAC and PARALIND models are only partially unique at the latent level. The rules discussed in Section 2.4.1 allow us to predict that two of the Mode B components would be unique while the other two components, the ones that in Mode B correspond to the factors collinear in A, will only be determined up to a general linear transformation of the plane that they span. In addition, we know that the arbitrary linear transformation of a column pair in Mode B by which one solution differs from another will be matched by the inverse transformation applied to the corresponding plane or column pair in Mode C. We can also deduce that all the loading vectors in Mode A will be 'unique', with two columns identical. (It is perhaps a matter of mathematical aesthetics whether the two identical Mode A columns are considered to be 'really' unique or simply subject to an indeterminacy that is invisible because the columns are identical.)

While both models are partially nonunique at the latent-structure level, we expect only the PARALIND model to remain partially nonunique at the 'surface level (i.e. when fitting the noise-perturbed observed data). In contrast, the PARAFAC solution will become fully unique, due to fitting of non-collinear perturbations of the otherwise collinear patterns in the data. Such uniqueness would be misleading, however, because it would not depend on the latent structure of the data, which is what we want to fit, but rather on characteristics of random noise (and sometimes on nonrandom but irrelevant distortions of measurement) that happen to be present in the data. Thus, we expect the PARALIND solution to more accurately recover the latent structure.

3.2. Flow injection analysis (FIA)

Our second application will be a dataset generated by FIA originating from Nørgaard and Ridder [36].

3.2.1. *Background*

This is an example of linear dependences among factors that arise unavoidably from the logical or causal structure of a measurement technique. In cases of this kind, the dependences are unavoidable but at least they have a mathematical structure that can be anticipated and thus they permit us to construct an appropriately constrained analysis model. In this dataset, the dependences are more extensive and complex than those in our first example, and so they constitute a more challenging case for PARAFAC, one that calls for an even more extensive application of the enhancements provided by applying the principles of PARALIND.

The Nørgaard and Ridder dataset has been used in at least two other demonstrations of proposed three-way methods of analysis. By applying PARALIND to these data, we make it possible for the reader to directly compare our approach and results to those in these other recent proposals. The most closely related model was that described in Reference [14], an paper that is not only useful in itself but also useful in this context because it presents comparisons with other alternatives for second order calibration. It should be particularly interesting and instructive, however, for the reader to compare the approach we take below to that taken by Kiers and Smilde [1]. They used a constrained Tucker3 model, and this led to an interpretation in the more dynamic terms often characteristic of models that use a core array. Their article also includes a proof of (partial) uniqueness which is relevant to our analysis because the structural restrictions assumed by the two approaches are formally equivalent. We would recommend that the reader compare these articles and methods to the one we propose here, since, unfortunately, limitations of space and scope, and the subtlety and complexity of some of the required discussion, make it infeasible for us to do an adequate job in this paper.

3.2.2. *The FIA method*

In FIA, a chemical sample is introduced into a transparent channel through which a carrier fluid is flowing. As it moves downstream, the sample disperses into the carrier (or reagent) and possibly undergoes a chemical change. These changes are typically monitored by spectral detection. For each sample, a 'landscape' is obtained, showing the absorption spectrum at each of the measurement times; conversely, this could be described as a landscape showing the time profile of changes in light absorbance at each of the measured wavelengths.

It is characteristic of FIA that there is no physical separation of the analytes. The carrier fluid is set up to control the progress of the reactions; in the example considered here, this will be done by means of a pH gradient. For this reason, it often happens that the chemical transformations and/or exchanges that help to separate, identify and quantify the separate constituents are causally linked to one another in ways that introduce linear dependences into the resulting measurements, as happens in this dataset.

3.2.3. *The FIA data*

In this example, three analytes were measured, 2HBA, 3HBA, and 4HBA (2-, 3-, and 4-hydroxy-benzaldehyde, respectively). All three analytes have different absorption spectra depending on whether they are in their acidic or basic form, and this fact plays an important role in the procedure. Depending on the pK_a of a given analyte, it will appear in its acidic and basic form in different proportions, and hence will show different relative strength of its acidic vs. basic spectrum in different parts of the injected sample 'plug' and hence at different measurement times. The first part of the sample plug, i.e. the earliest measurements of a sample, will be dominated by deprotonated (basic) analytes while the end of the sample plug and the later measurements will be dominated by protonated (acidic) analytes. For any one time point, the FIA measurements can be held in an $I \times J$ matrix called $\mathbf{X}_k$ where I is the number of samples, J is the number of wavelengths, and k indicates the specific pH/time selected. Our analysis used every other level of the original Modes B and C values, resulting in a dataset where $I = 12$ and $J = 50$ and there are $K = 45$ matrices.

3.2.4. *The model with single mode dependency*

Since there are three analytes, there are only three different variations in their concentrations. Thus, in the sample mode, a three-component decomposition would seem appropriate. However, each analyte exists in two forms (acidic and basic) each of which generates a separate spectrum to resolve. To accommodate these seemingly conflicting requirements, the PARAFAC-like form of the model in Equation (5) would need six components. The PARALIND form, corresponding to Equation (8), would have three components in **A** and an **H** matrix that expressed the sample equivalence of the concentrations for the acid-base spectral pairs, as follows:

$$\mathbf{H}_A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \tag{17}$$

The corresponding $K \times 6$ **C** matrix gives the weights/fractions for the six spectra (corresponding to the two forms of the three analytes) at each of the K time points/pH values measured. Thus, the kth row of **C** holds the six weights for time point k. The first element in this row is simply the amount of the first analyte in the basic form and the second is the amount in acidic form. For every k, we can construct a 6×6 diagonal matrix $\mathbf{D}_k$ which contains the kth row of **C** on its diagonal. With this we can describe the data at time point k. In the general case, this model can be written (in 'typical slice' form) as

$$\begin{aligned} \mathbf{X}_k &= \mathbf{A}\mathbf{H}\mathbf{D}_k\mathbf{B}^{\mathrm{T}} \\ &= \mathbf{A} \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 1 & & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix} \begin{bmatrix} c_{k1} & 0 & \cdots & 0 \\ 0 & c_{k2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & c_{kJ} \end{bmatrix} \mathbf{B}^{\mathrm{T}} \end{aligned} \tag{18}$$

This model in Equation (18) demonstrates how the use of a distinct **H** and **C** (and thus distinct $\mathbf{D}_k$) allows the qualitative and quantitative relationships between **A** and **B** to be expressed separately.

In this version, the model is another case in which two modes have full column rank and so its partial uniqueness properties can be deduced from the principles discussed in Section 2.4.1 and detailed in Appendix B. For our particular six-factor analysis, we expect that: (i) Mode A will be unique but will consist of three pairs of identical columns and (ii) the corresponding column pairs in Modes B and C each reflect a separate compound for which the acidic and basic forms are not uniquely resolved.

3.2.5. *The model with dependency in two modes*

While one of the models that we will fit is the one discussed above, there is actually an additional set of linear dependences in these data, and these must be included in our full model of the FIA dataset.

It happens that all analytes measured here have the same dilution profile, and so they have identically shaped *total* (acidic + basic) time-profiles. Incorporation of this into the PARALIND model requires a second **H** matrix. The resulting

model can be written as

$$\mathbf{X}_j = \mathbf{A}\mathbf{H}_\mathrm{A}\langle \mathbf{b}_j \rangle \mathbf{H}_\mathrm{C}^\mathrm{T}\mathbf{C}^\mathrm{T} \tag{19}$$

with a (transposed) right hand dependency matrix

$$\mathbf{H}_\mathrm{C}^\mathrm{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

and a Mode C loading matrix **C** consisting of four columns, the first three representing the concentration of, say, the odd numbered column substances, plus a fourth which represents the total concentration for each pair at each level of C, (i.e. at each time).

3.2.6. *Uniqueness properties*

When the added dependency in Mode C is considered, the FIA data become an example of those cases where at most one mode has full column rank, discussed in Section 2.4.2. Unfortunately, our current knowledge does not let us predict with confidence which aspects of the doubly-dependent PARALIND analysis will be unique and which will not.

As we will report in Section 4, the, models of flow injection data subject to constraints such as those found here can be expected to provide means for quantitative analysis by PARALIND, and possibly also with PARAFAC with more difficulty, but neither model can be used directly for qualitative analysis. That is, the models are useful for identifying how much of the compound is present, but not what the compound is. This situation can be improved with the addition of nonnegativity and additional constraints as shown earlier [18].

Even with the uniqueness limitations that are intrinsic to these data, we will see that, compared to PARAFAC, the PARALIND model improves the accuracy of recovery for those parameter estimates that can be uniquely determined. And because it restores to the estimated parameter values the partial non-uniqueness that holds at the latent level, it allows better assessment of the degree to which parameter estimates can be used as approximate measurements, and, more generally, increases the theoretical appropriateness of the solution.

4. RESULTS

4.1. Fluorescence data

Theoretically, the fluorescence EEM data should be a relatively straightforward application of PARALIND, since there is factor dependency in only one mode, and this involves only two factors. However, real data can sometimes show additional complications not anticipated by the theory. This is demonstrated by this case, in which an important illustration of misleading uniqueness revealed itself.

4.1.1. *Assessing uniqueness and identification of interpretable parameters*

Although the theoretical expectations are eventually confirmed, this is not initially apparent in the results of the PARAFAC analysis, which would seem to indicate that the the unconstrained PARAFAC model is doing a surprisingly good job (Figure 2, left). All 20 fitted solutions are superimposed and are seen to be virtually identical. Not only do the four refitted models of each dataset give the same results as expected but also the five different replicate datasets are apparently also (almost) identical. If the data behave according to theory, then there should be only three different variations in the sample mode, and as a result fitting four components should lead to some non-identifiability in the spectral modes when fitting different datasets. This is not seen. Apparently, the chemical analytes interact: different amounts of one analyte affect the signal of the other analytes to a small, but sufficiently reproducible degree, to produce a four-component solution that is consistent over the replicates. However, this solution is misleading; it does *not* provide estimates that correspond (at least linearly) with the pure analyte profiles. For example, one emission mode profile is systematically below zero at approximately 300 nm, as is the corresponding excitation profile.

The solution can be dissected into two parts, and the misleading results identified. By elimination of the recognizable and approximately correct profiles, we can identify the profile with the impossible negative weights as one of the two corresponding to collinear factors in the sample mode. However, the impossible loadings are not literally the result of 'nonuniqueness' at the surface level, since the solutions replicate from different random starting positions. In fact, although there is theoretically 'nonuniqueness' at the latent structural level, this is not apparent from these data because the impossible curve is replicated in all five different datasets.

It is possible to get an indication of the problem by looking at the values of the concentration-mode loadings or 'scores' that were obtained in the PARAFAC solution. Figure 3 shows a scatter plot of scores for factor three vs. those for four from a PARAFAC analysis. The plot clearly indicates that the two factors have scores that are approximately collinear, yet there appears to be some subtle bifurcation in which, for example, samples 21 and 18 are distinguished from samples 4, 20, and 22.

A comparison of these results with the unconstrained PARALIND solution is quite instructive. At first sight, the PARALIND results seem more difficult to interpret. However, looking more carefully at Figure 2 (right), reveals that in the spectral modes, two of the components are uniquely identified whereas the two remaining ones vary from replicate to replicate and from refit to refit. This is even more evident in Figure 4 which shows the PARALIND emission mode loadings, split in two parts. Results for components corresponding to cathecol and hydroquinone are shown to the right and those for components corresponding to indole and tryptophane are shown to the left. As can be seen, only the two left-most components are uniquely determined whereas the two rightmost ones are definitely not. In different runs, different (rotated) solutions are obtained. This result is in exact accordance with theory and thus immediately leads to the (verifiably correct) interpretation that all concentration-mode scores are uniquely recovered as are the loadings defining spectra for two of the four compounds in the other two modes. On the other hand, the two components in the spectral modes that correspond to factors that are collinear in the concentration mode are not uniquely identified. (Further careful analysis revealed, however, that the nonunique alternative solutions consist solely of different linear combinations of the two nonunique spectra, with nothing from the other unique ones

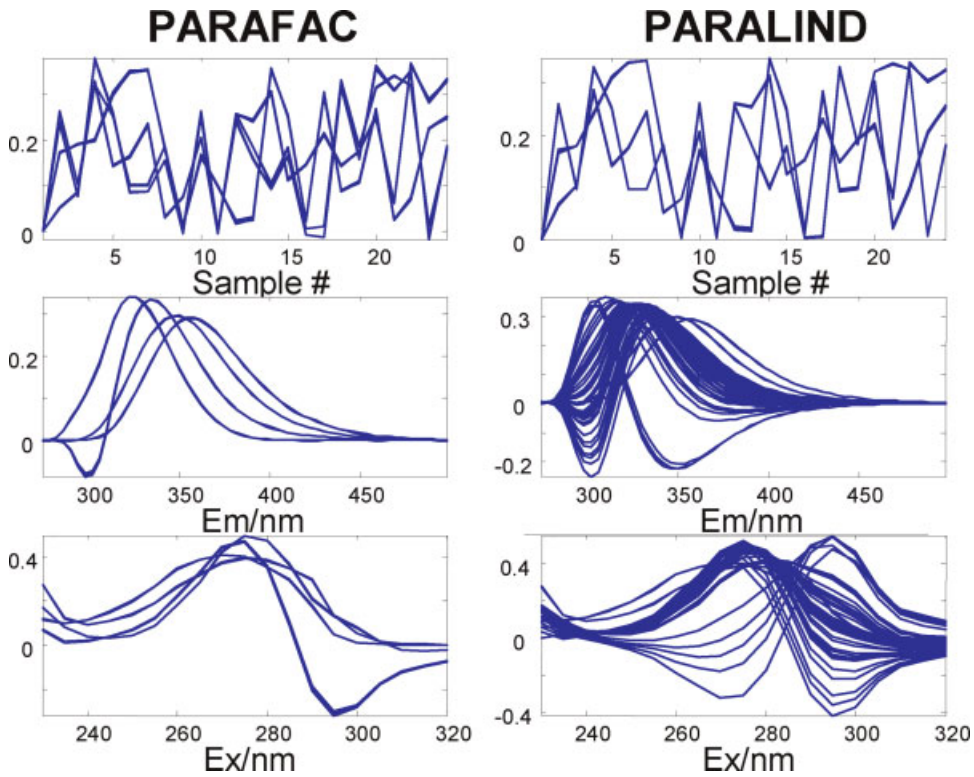

Figure 2. Component matrices from four repeated runs on each of five replicate datasets using unconstrained PARAFAC (left) and unconstrained PARALIND (right). Top plot holds the scores, middle plots emission loadings and bottom plots the excitation loadings. The similarity but incorrectness of the twenty repeated runs of PARAFAC illustrates the problem of apparent uniqueness of PARAFAC. This figure is available in color online at www.interscience.wiley.com/journal/cem

mixed in). Hence, in the PARALIND solution, the only parameters that are uniquely determined are those that are unique at the latent level, and hence those whose unique solutions are meaningful. In contrast, the PARAFAC solution is also unique for the other parameters but this is misleading and not fully meaningful. While we were expecting to demonstrate that these surface-unique spectra varied randomly across replicates, revealing the 'meaninglessness', we instead demonstrated that a deeper illusion of uniqueness was produced by some kind of nonlinear interaction that produced nonmeaningful results that were in fact consistent across replicates. In some sense, this further demonstrates a strength of the nonuniqueness built into the PARALIND model.

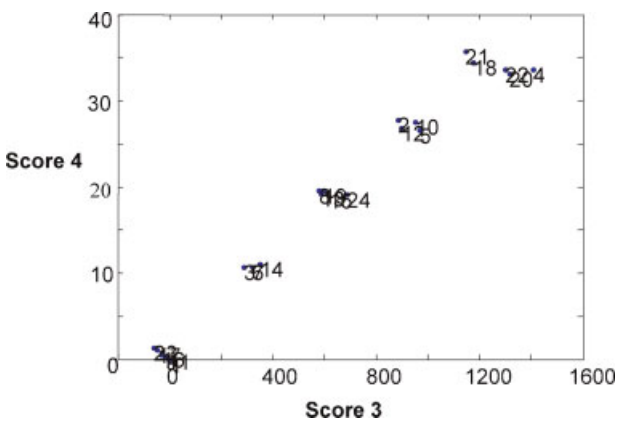

Figure 3. Scatter plot of scores for factor three and four from the PARAFAC solution. This figure is available in color online at www.interscience.wiley.com/journal/cem

As stated in Section 2.5, the fit of PARAFAC will be better than PARALIND regardless of which model is appropriate. This is also verified as expected. An unconstrained PARAFAC model fits 99.98% of the variation whereas PARALIND fits 99.93%. For the nonnegativity-constrained models, PARAFAC fits 99.91% and PARALIND 99.83%. Hence, minor difference in fit values do not warrant any conclusions in themselves about which model would be preferred.

4.1.2. Recovery of latent information from the EEM data

The quantitative results for different analyses of the EEM data are given below. We discuss the results obtained using PARAFAC with nonnegativity constraints as well as PARALIND with nonnegativity constraints. For assessment of accuracy of recovery of information in the concentration mode, estimates are computed based on regression. To quantify the accuracy of the concentration estimates, the R^2 (fraction of variance explained) and RMSEC (the Root of the Mean Squared Error of Calibration) are given for each analyte. These values are based on a univariate regression of the appropriate analyte concentration on the associated score (done, as is usual, 'through the origin'—without

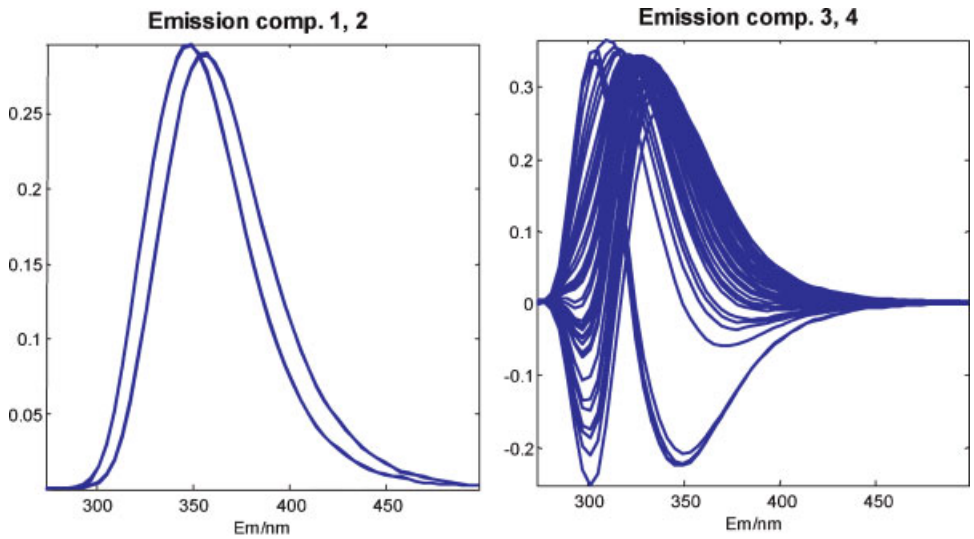

Figure 4. Emission mode loadings from the unconstrained PARALIND models in Figure 2. Left, the 20 estimates of the first two components are shown and to the right components three and four are shown. This figure is available in color online at www.interscience.wiley.com/journal/cem

a constant or intercept term). The RMSEC is computed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{I} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{I} (y_i - \bar{y})^2}$$

PARALIND provides systematically better results than PARAFAC (Table II). The predictions are shown in Figure 5. It is also worth noting that the two collinear analytes (catechol and hydroquinone) are much better predicted than the remaining two. The reason for this is most likely the highly similar spectral properties of indole and tryptophane.

4.2. Flow injection data

In order to explore the usefulness of alternative PARALIND models of the FIA dataset, three different versions were tested: (i) linear dependency in the sample mode; (ii) linear dependency in both sample and time modes; and (iii) dependences in both modes plus nonnegativity constraints on the parameters. All were compared to a baseline analysis that used a standard PARAFAC model with six components.

To perform these comparisons, the quality of the resulting solutions is evaluated by looking at how well the parameter estimates recover the known concentrations.

4.2.1. Recovery of latent information on concentrations

These values, computed are shown in Table III. As can be seen, the predictions are fair except for 3HBA. The reason for this may be that the similarity of all time profiles causes multi-collinearity makes it difficult to distinguish 2- and 3HBA. The most important trend in the results is that the predictions improve by adding more constraints. The more constrained the model is, the less degrees-of-freedom are used for fitting the data, but still the prediction consistently improves except some minor deviations. Thus, the quantitative results confirm that adding the chemical information specifically into the structure of the model is beneficial and actually takes the predictions from mediocre in the case of PARAFAC to close to excellent for the double PARALIND with nonnegativity. It is found that the predicted uniqueness

Table II. Calibration results for one of the five replicate datasets

	Catechol	Hydroquinone	Indole	Tryptophane
PARAFAC nonneg	0.45 (86)	0.12 (86)	0.11 (41)	0.18 (33)
PARALIND nonneg	0.36 (89)	0.09 (89)	0.09 (49)	0.11 (59)

For each analyte the RMSEC ($\times 10^6$) is given and R^2 (%) is given in parentheses. The results are from a univariate regression of the reference using as independent variable, the scores of an unconstrained model.

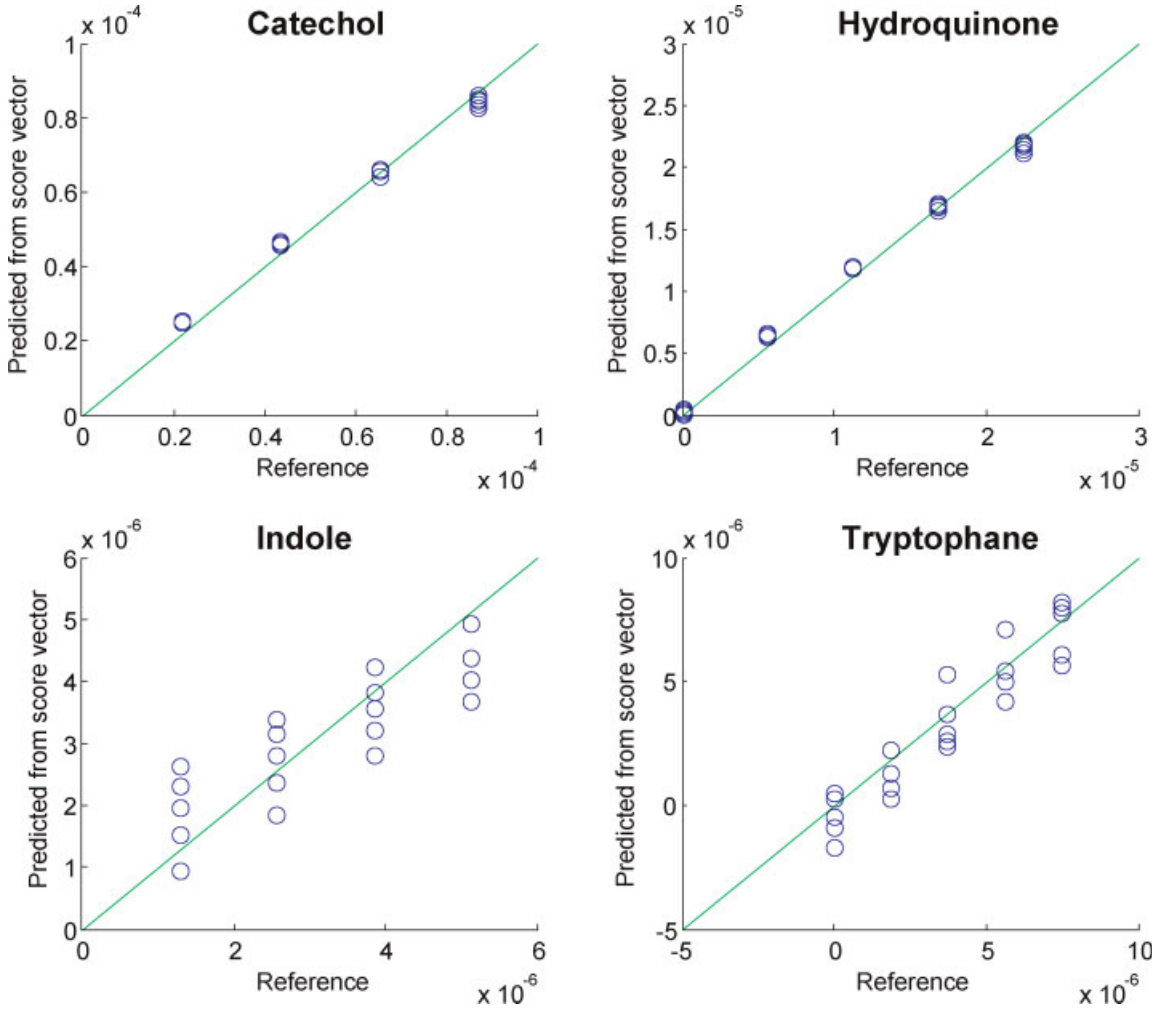

Figure 5. Predicted versus reference concentrations for nonnegativity constrained PARALIND. Each plot shows the actual concentration on the *x*-axis and the prediction from a univariate regression of the appropriate score vector on the concentration. This figure is available in color online at www.interscience.wiley.com/journal/cem

properties (Section 2.4) are confirmed in the empirical results. For example, the Mode A loadings in single PARALIND are unique and the Modes B and C parameters are not unique. However, for each analyte, the two loading vectors e.g. in Mode B are found to vary but always stay in the same subspace. More detailed analysis of the qualitative recovery of the spectral data confirms the predicted uniqueness properties but these results can be found elsewhere [18].

The actual predictions are shown in Figure 6 visualizing the quantitative results given in Table III. The plot clearly shows the significant improvements that are gained by incorporating sound *a priori* knowledge into the algebraic structure.

Table III. Calibration results

	2HBA	3HBA	4HBA	Average R^2
PARAFAC	51 (0)	24 (66)	13 (49)	38
Single PARALIND	15 (62)	30 (28)	3 (88)	59
Double PARALIND	21 (56)	18 (76)	2 (91)	74
Double PARALIND + nonnegativity	2 (95)	7 (77)	1 (97)	90

For each analyte the RMSEC ($\times 1000$) is given and R^2 (%) is given in parentheses. The results are from a univariate regression of the reference using as independent variable, the scores of an unconstrained model.

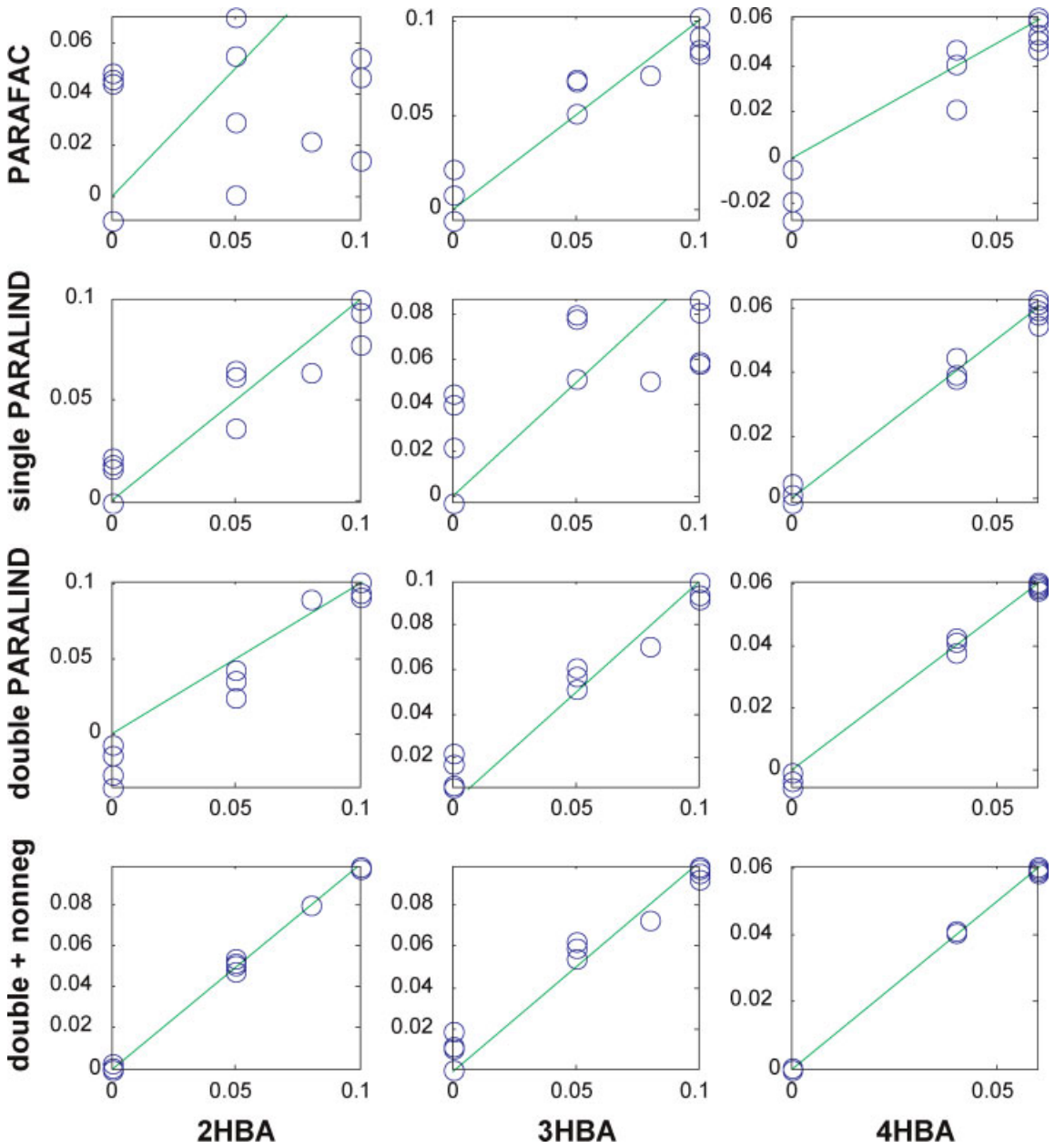

Figure 6. Predicted versus reference concentrations. Each plot shows the actual concentration on the *x*-axis and the prediction from a univariate regression of the appropriate score vector on the concentration. This figure is available in color online at www.interscience.wiley.com/journal/cem

5. CONCLUSION AND FURTHER PERSPECTIVES

Both theoretical and empirical results indicate that PARALIND strengthens a PARAFAC-type analysis in those cases where the process generating the data introduces linear dependences into the latent structure in one or more modes. PARALIND fits less error and better recovers the true latent structure, and the properties of the solution are more appropriate theoretically and because of the restored partial uniqueness allow better assessment and characterization of the range and nature of partial uniqueness that results from the dependences. PARALIND models transform implicit dependences into explicit representation of dependences in the model structure. An algorithm for fitting certain PARALIND models has been developed and is described in Appendix A.

Thus, PARALIND would seem to provide a more appropriate analysis method for certain multi-way problems that can be characterized by having component matrices with the same number of columns but different ranks in different modes. The method employs a constrained-PARAFAC perspective, which maintains the one-factor-per-chemical analyte scheme of interpretation, instead of introducing factors that represent interactions of processes or other levels of description. This may or may not be advantageous depending on one's scientific problem and analysis objectives, and perhaps also ones scientific style or aesthetic preferences, but in any case it serves as a way to strengthen PARAFAC for these kinds of cases, and hence provides a new tool in the analyst's toolbox.

6. MATERIALS AND METHODS

The type of PARALIND algorithm applicable for the FIA data has been implemented for imposing linear dependency in one mode. It is written in MATLAB v. 7.6 (MathWorks, Inc.) and is available from the Internet at http://www.models.life.ku.dk. To run the PARALIND algorithm it is advisable to have the *N*-way toolbox [37] which is available at the same site. A more general PARALIND version is available in PLS_Toolbox (Eigenvector Inc.) which was used in this paper. This version allows linear dependences in all modes and also allows higher-order models.

Acknowledgements

Lars Nørgaard and Carsten Ridder are thanked for letting us use their original data. The second author gratefully acknowledges financial support provided by NSERC (The Natural Science and Engineering Research Council of Canada). The third author has been supported by NSF/CAREER CCR-9733540.

REFERENCES

1. Kiers HAL, Smilde AK. Constrained three-mode factor analysis as a tool for parameter estimation with second-order instrumental data. *J. Chemom.* 1998; **12**: 125–147.
2. Harshman RA. Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multi-modal factor analysis. *UCLA Working Papers in Phonetics* 1970; **16**: 1–84.
3. Harshman RA, Lundy ME. PARAFAC: parallel factor analysis. *Comput. Stat. Data Anal.* 1994; **18**: 39–72.
4. Carroll JD, Chang J. Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika* 1970; **35**: 283–319.
5. Brereton RG, Gurden SP, Groves JA. Use of eigenvalues for determining the number of components in window factor analysis of spectroscopic and chromatographic data. *Chemom. Intell. Lab. Syst.* 1995; **27**: 73–87.
6. Izquierdo-Ridorsa A, Saurina J, Hernandez-Cassou S, Tauler R. Second-order multivariate curve resolution applied to rank-deficient data obtained from acid-base spectrophotometric titrations of mixtures of nucleic bases. *Chemom. Intell. Lab. Syst.* 1997; **38**: 183–196.
7. Ross RT, Leurgans SE. Component resolution using multilinear models. *Methods Enzymol.* 1995; **246**: 679–700.
8. Sidiropoulos ND, Dimic GZ. Blind multiuser detection in W-CDMA systems with large delay spread. *IEEE Signal Process. Lett.* 2001; **8**: 87–89.
9. Sidiropoulos ND, Liu XQ. Identifiability results for blind beamforming in incoherent multipath with small delay spread. *IEEE Trans. Signal Process.* 2001; **49**: 228–236.
10. van der Veen AJ. Algebraic methods for deterministic blind beamforming. *Proc. IEEE* 1998; **86**: 1987–2008.
11. Amrhein M, Srinivasan B, Bonvin D, Schumacher MM. On the rank deficiency and rank augmentation of the spectral measurement matrix. *Chemom. Intell. Lab. Syst.* 1996; **33**: 17–33.
12. Garrido M, Lazaro I, Larrechi MS, Rius FX. Multivariate resolution of rank-deficient near-infrared spectroscopy data from the reaction of curing epoxy resins using the rank augmentation strategy and multivariate curve resolution alternating least squares approach. *Anal. Chim. Acta* 2004; **515**: 65–73.
13. Amrhein M. Reaction and flow variants/invariants for the analysis of chemical reaction data. *PhD Thesis*, École Polytechnique Fédérale de Lausanne, 1998.
14. Smilde AK, Tauler R, Saurina J, Bro R. Calibration methods for complex second-order data. *Anal. Chim. Acta* 1999; **398**: 237–251.
15. Amrhein M, Srinivasan B, Bonvin D, Schumacher MM. Calibration of spectral reaction data. *Chemom. Intell. Lab. Syst.* 1999; **46**: 249–264.
16. Smilde AK, Tauler R, Henshaw JM, Burgess LW, Kowalski BR. Multicomponent determination of chlorinated hydrocarbons using a reaction-based chemical sensor. 3. Medium-rank second-order calibration with restricted Tucker models. *Anal. Chem.* 1994; **66**: 3345–3351.
17. Levin J. Three-mode factor analysis. *Psychol. Bull.* 1965; **64**: 442–452.
18. Bro R. Multi-way analysis in the food industry. models, algorithms, and applications. *PhD Thesis*, University of Amsterdam (NL), 1998. http://www.models.life.kvl.dk/research/theses/
19. Harshman RA, Lundy ME. Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and PARAFAC/CANDECOMP. *Psychometrika* 1996; **61**: 133–154.
20. Harshman RA. Generalizations, extensions and structured special-purpose modifications of PARAFAC with possible applications in chemistry. *Talk presented to TRIC (Three-way analysis methods in Chemistry) in Epe, Holland*, 1993.
21. Rao CR, Mitra S. *Generalized Inverse of Matrices and its Applications*. Wiley: New York, 1971.
22. Kruskal JB. Three-way arrays: rank and uniqueness of trilinear decompositions with applications to arithmetic complexity and statistics. *Linear Algebra Appl.* 1977; **18**: 95–138.
23. Kruskal JB. More factors than subjects, tests and treatments: an indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika* 1976; **41**: 281–293.
24. Harshman RA. Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics* 1972; **22**: 111–117.
25. ten Berge JMF. Partial uniqueness in CANDECOMP/PARAFAC. *J. Chemom.* 2004; **18**: 12–16.
26. Faber NM, Buydens LMC, Kateman G. Generalized rank annihilation method. I: derivation of eigenvalue problems. *J. Chemom.* 1994; **8**: 147–154.
27. Lorber A. Quantifying chemical composition from two-dimensional data arrays. *Anal. Chim. Acta* 1984; **164**: 293–297.
28. Sanchez E, Kowalski BR. Generalized rank annihilation factor analysis. *Anal. Chem.* 1986; **58**: 496–499.
29. Sidiropoulos ND, Giannakis GB, Bro R. Blind PARAFAC receivers for DS-CDMA systems. *IEEE Trans. Signal Process.* 2000; **48**: 810–823.
30. Swindlehurst A, Kailath T. Azimuth/elevation direction finding using regular array geometries. *IEEE Trans. Aerosp. Electron. Syst.* 1993; **29**: 145–156.
31. Leurgans SE, Ross RT. Multilinear models: applications in spectroscopy. *Stat. Sci.* 1992; **7**: 289–319.
32. Paatero P. The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. Comput. Graph. Stat.* 1999; **8**: 854–888.
33. Jiji RD, Booksh KS. Mitigation of Rayleigh and Raman spectral interferences in multiway calibration of excitation-emission matrix fluorescence spectra. *Anal. Chem.* 2000; **72**: 718–725.
34. Rinnan A, Andersen CM. Handling of first-order Rayleigh scatter in PARAFAC modelling of fluorescence excitation-emission data. *Chemom. Intell. Lab. Syst.* 2005; **76**: 91–99.
35. Bro R. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemom. Intell. Lab. Syst.* 1999; **46**: 133–147.
36. Nørgaard L, Ridder C. Rank annihilation factor analysis applied to flow injection analysis with photodiode-array detection. *Chemom. Intell. Lab. Syst.* 1994; **23**: 107–114.
37. Andersson CA, Bro R. The N-way Toolbox for MATLAB. *Chemom. Intell. Lab. Syst.* 2000; **52**: 1–4.
38. Bro R, Andersson CA. Improving the speed of multiway algorithms part II: compression. *Chemom. Intell. Lab. Syst.* 1998; **42**: 105–113.
39. Kiers HAL. A three-step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity. *J. Chemom.* 1998; **12**: 155–171.
40. Bro R, de Jong S. A fast non-negativity-constrained least squares algorithm. *J. Chemom.* 1997; **11**: 393–401.
41. Bro R, Sidiropoulos ND. Least squares algorithms under unimodality and non-negativity constraints. *J. Chemom.* 1998; **12**: 223–247.
42. Henderson HV, Searle SR. The vec-permuation matric, the vec operator and the Kronecker products: a review. *Linear and Multilinear Algebra* 1981; **9**: 271–288.
43. Bro R. PARAFAC. Tutorial and applications. *Chemom. Intell. Lab. Syst.* 1997; **38**: 149–171.
44. Harshman, RA, Hong, S. 'Stretch' vs. 'slice' methods for representing three-way structure via matrix notation. *J. Chemom.* 2002; **16**, 198–205.

APPENDIX A: AN ALS-TYPE PARALIND ALGORITHM

In this appendix, we present the basics of an algorithm for fitting PARALIND models. In practical implementations, the algorithm can be made more efficient by the use of line-search, compression and similar approaches [2,18,38,39]. As currently written, this algorithm only implements linear dependency in one mode. Extension to higher orders and linear dependency in more than one mode can be implemented along the same lines. For illustration of how to incorporate constraints into the algorithm the reader is referred to the literature [18,40,41].

Let $\mathbf{X}$ be an $I \times JK$ matrix obtained by rearranging (matricizing) the $I \times J \times K$ array $\underline{\mathbf{X}}$, so the first J columns correspond to x_{ij1} ($i = 1,\ldots,I$, $j = 1,\ldots,J$, $k = 1$) and the last J columns correspond to x_{ijK}. A PARALIND model of left-dimension (column-dimension of $\mathbf{A}$) R and right-dimension (column dimension of $\mathbf{B}$ and $\mathbf{C}$) S is sought. Thus the solution to

$$\min \left\| \mathbf{X} - \mathbf{AH}(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}} \right\|_{\mathrm{F}}^{2} \tag{A1}$$

is sought.

The operator diag($\mathbf{X}$) means a column vector with the diagonal elements of $\mathbf{X}$. The operator vec is introduced for simplifying the formulae. vec($\mathbf{X}$) is the vector obtained by stringing out $\mathbf{X}$ columnwise to a column vector [42].

The operator '$*$' is the Hadamard (elementwise) product. The algorithm proceeds as follows after initialization of all loading matrices

1. $\mathrm{vec}\mathbf{H} = \left[(\mathbf{B}^{\mathrm{T}}\mathbf{B}) * (\mathbf{C}^{\mathrm{T}}\mathbf{C}) \otimes (\mathbf{A}^{\mathrm{T}}\mathbf{A}) \right]^{-1} \mathrm{vec}\left(\sum_{k=1}^{K} \mathbf{A}^{\mathrm{T}}\mathbf{X}_{\mathrm{k}}\mathbf{B}\mathbf{D}_{\mathrm{k}} \right)$
2. $\mathbf{A} = \left(\sum_{k=1}^{K} \mathbf{X}_{\mathrm{k}}\mathbf{B}\mathbf{D}_{\mathrm{k}}\mathbf{H}^{\mathrm{T}} \right) \left\{ \mathbf{H} \left[(\mathbf{B}^{\mathrm{T}}\mathbf{B}) * (\mathbf{C}^{\mathrm{T}}\mathbf{C}) \right] \mathbf{H}^{\mathrm{T}} \right\}^{-1}$
3. $\mathbf{B} = \left(\sum_{k=1}^{K} \mathbf{X}_{k}^{\mathrm{T}}\mathbf{A}\mathbf{H}\mathbf{D}_{\mathrm{k}} \right) \left\{ (\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{H}) * (\mathbf{C}^{\mathrm{T}}\mathbf{C}) \right\}^{-1}$
4. $\mathrm{diag}\mathbf{D}_{\mathrm{k}} = \left\{ (\mathbf{B}^{\mathrm{T}}\mathbf{B}) * (\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{H}) \right\}^{-1} \mathrm{diag}\left(\mathbf{A}\mathbf{H}^{\mathrm{T}}\mathbf{X}_{\mathrm{k}}\mathbf{B} \right), \quad k = 1,\ldots,K$
5. Go to Step 1 until relative change in fit is small.

The algorithm basically follows from extending traditional PARAFAC-ALS algorithms [2,4,43]. If the interaction matrix is fixed, the updating step for $\mathbf{H}$ is simply skipped. The main difference from ordinary PARAFAC comes from the need to develop an efficient update step for $\mathbf{H}$. The updating of $\mathbf{H}$ can be explained as follows. The model for any frontal slab of the array is

$$\begin{aligned} \mathbf{X}_{\mathrm{k}} &= \mathbf{A}\mathbf{H}\mathbf{D}_{\mathrm{k}}\mathbf{B}^{\mathrm{T}} \Rightarrow \\ \mathrm{vec}\mathbf{X}_{\mathrm{k}} &= \mathrm{vec}(\mathbf{A}\mathbf{H}\mathbf{D}_{\mathrm{k}}\mathbf{B}^{\mathrm{T}}) \Rightarrow \\ \mathrm{vec}\mathbf{X}_{\mathrm{k}} &= (\mathbf{B}\mathbf{D}_{\mathrm{k}} \otimes \mathbf{A})\mathrm{vec}\mathbf{H} \end{aligned} \tag{A2}$$

Considering all slabs simultaneously this leads to

$$\begin{aligned} \begin{bmatrix} \mathrm{vec}\mathbf{X}_1 \\ \mathrm{vec}\mathbf{X}_2 \\ \vdots \\ \mathrm{vec}\mathbf{X}_{\mathrm{K}} \end{bmatrix} &= \begin{bmatrix} \mathbf{B}\mathbf{D}_1 \otimes \mathbf{A} \\ \mathbf{B}\mathbf{D}_2 \otimes \mathbf{A} \\ \vdots \\ \mathbf{B}\mathbf{D}_{\mathrm{K}} \otimes \mathbf{A} \end{bmatrix} \mathrm{vec}\mathbf{H} \Rightarrow \\ \mathrm{vec}\mathbf{X}^{(\mathrm{I}\times\mathrm{JK})} &= [(\mathbf{C} \odot \mathbf{B}) \otimes \mathbf{A}]\mathrm{vec}\mathbf{H} \end{aligned} \tag{A3}$$

from which the update of $\mathbf{H}$ in Step 1 immediately follows taking into account that

$$(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}}(\mathbf{C} \odot \mathbf{B}) = (\mathbf{B}^{\mathrm{T}}\mathbf{B}) * (\mathbf{C}^{\mathrm{T}}\mathbf{C}) \tag{A4}$$

and using the rules for Kronecker and Khatri-Rao products [18].

APPENDIX B: PROOF OF PARTIAL UNIQUENESS OF FIA-TYPE PARALIND MODEL WITHOUT LINEAR DEPENDENCE IN THE THIRD MODE

Consider the recasting of the PARALIND model as a PARAFAC model. Let

$$\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1\ \tilde{\mathbf{a}}_2\ \tilde{\mathbf{a}}_3\ \tilde{\mathbf{a}}_4\ \tilde{\mathbf{a}}_5\ \tilde{\mathbf{a}}_6] = [\mathbf{a}_1\ \mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_2\ \mathbf{a}_3\ \mathbf{a}_3] \tag{B1}$$

Then the FIA model

$$\mathbf{X}_k = \mathbf{A}\mathbf{H}\mathbf{D}_k\mathbf{B}^{\mathrm{T}}, \quad k = 1,\ldots,K \tag{B2}$$

with the special structure of $\mathbf{H}$ is equivalent to

$$\mathbf{X}_k = \tilde{\mathbf{A}}\mathbf{D}_k\mathbf{B}^{\mathrm{T}}. \quad k = 1,\ldots,K \tag{B3}$$

Rather than representing the model in terms of the frontal slabs consider the model represented by the horizontal slabs

$$\mathbf{X}_i = \mathbf{B}\mathbf{D}_i\mathbf{C}^{\mathrm{T}},\ i = 1,\ldots,I \tag{B4}$$

where $\mathbf{D}_i$ is to be seen as an operator $\mathbf{D}_i(\tilde{\mathbf{A}})$ which outputs a diagonal matrix with the ith row of $\tilde{\mathbf{A}}$ in its diagonal [29]. For the basic FIA model, it follows that the diagonal elements of $\mathbf{D}_i$ are pairwise identical because of the structure in $\tilde{\mathbf{A}}$. Consider the generalized eigenvalue problem representation of PARAFAC as exemplified in, e.g. GRAM [26–28] and ESPRIT [10,29,30]. Two horizontal slabs of size $J \times K$ are given for which all factors contribute with a non-zero amount and where the initial concentrations of the three analytes are distinct. These slabs follow the model

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{B}\mathbf{C}^{\mathrm{T}} \\ \mathbf{X}_2 &= \mathbf{B}\mathbf{D}\mathbf{C}^{\mathrm{T}} \end{aligned} \tag{B5}$$

where $\mathbf{D}$ has multiple nonzero elements, and $\mathbf{B}$ and $\mathbf{C}$ are tall, F columns, and full rank. Let $\mathbf{U}$ be the first F left singular vectors of the $2J \times K$ matrix $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$. Then it is true that the span of $\mathbf{U}$ is equal to the span of $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ and since $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{B} \\ \mathbf{B}\mathbf{D} \end{bmatrix} \mathbf{C}^{\mathrm{T}}$ and $\mathbf{C}$ is full rank, it follows that the span of $\mathbf{U}$ is equal to the span of $\begin{bmatrix} \mathbf{B} \\ \mathbf{B}\mathbf{D} \end{bmatrix}$. Hence

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{B}\mathbf{T} \\ \mathbf{B}\mathbf{D}\mathbf{T} \end{bmatrix} \tag{B6}$$

where $\mathbf{T}$ is square full rank. Define $\mathbf{G} = \mathbf{T}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}$ and define

$$\begin{aligned} \mathbf{R}_1 &= \mathbf{U}_1{}^{\mathrm{T}}\mathbf{U}_1 = \mathbf{T}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{T} = \mathbf{G}\mathbf{T} \\ \mathbf{R}_2 &= \mathbf{U}_1{}^{\mathrm{T}}\mathbf{U}_2 = \mathbf{T}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{D}\mathbf{T} = \mathbf{G}\mathbf{D}\mathbf{T} \end{aligned} \tag{B7}$$

Notice, that $\mathbf{R}_1$ and $\mathbf{R}_2$ are full rank and square. It follows that

$$\mathbf{G}^{-1}\mathbf{R}_1 = \mathbf{D}^{-1}\mathbf{G}^{-1}\mathbf{R}_2 \tag{B8}$$

or, equivalently

$$\mathbf{G}^{-1} = \mathbf{D}^{-1}\mathbf{G}^{-1}\mathbf{R}_2\mathbf{R}_1^{-1} \Leftrightarrow (\mathbf{R}_2\mathbf{R}_1^{-1})^{\mathrm{T}}(\mathbf{G}^{-1})^{\mathrm{T}} = (\mathbf{G}^{-1})^{\mathrm{T}}\mathbf{D} \qquad \text{(B9)}$$

Hence $(\mathbf{G}^{-1})^{\mathrm{T}}$ and $\mathbf{D}$ are regular eigenvectors and corresponding eigenvalues of the matrix $(\mathbf{R}_2\mathbf{R}_1^{-1})^{\mathrm{T}}$. Eigenvalues including multiplicities are uniquely determined whereas the span of eigenvectors corresponding to a given eigenvalue is uniquely determined from the corresponding eigenvectors. Thus, $\mathbf{D}$ is unique, whereas $(\mathbf{G}^{-1})^{\mathrm{T}}$ is recovered up to the stated ambiguity. So we transpose it and get $\mathbf{G}^{-1}$ up to the said ambiguity. Then from $\mathbf{R}_1 = \mathbf{GT}$ we get $\mathbf{T}$ as $(\mathbf{G}^{-1})\mathbf{R}_1$; then from $\mathbf{U}_1 = \mathbf{BT}$ we get $\mathbf{B}$ as $\mathbf{U}_1\mathbf{T}^{-1}$; then from $\mathbf{X}_1 = \mathbf{BC}^{\mathrm{T}}$ we get $\mathbf{C}$ as $\mathbf{B}^{+}\mathbf{X}_1$, all up to the said ambiguity, and the proof is complete.

It thus follows that for the FIA data $\mathbf{A}$, span([$\mathbf{c}_1$ $\mathbf{c}_2$]), span([$\mathbf{c}_3$ $\mathbf{c}_4$]), span([$\mathbf{c}_5$ $\mathbf{c}_6$]), span([$\mathbf{b}_1$ $\mathbf{b}_2$]), span([$\mathbf{b}_3$ $\mathbf{b}_4$]), span([$\mathbf{b}_5$ $\mathbf{b}_6$]) are unique.