The first conference on

Three-way methods **I**n **C**hemistry

TRIC

(A meeting of Psychometrics and Chemometrics)

August 22-25, 1993

Epe, The Netherlands

Organising and Scientific Committee

Age Smilde, Chair
Laboratory for Analytical Chemistry, University of Amsterdam, The Netherlands

Pieter Kroonenberg
Department of Education, Leiden University, The Netherlands

Jos ten Berge
Department of Psychology, University of Groningen, The Netherlands

Paul Geladi
Research Group for Chemometrics, University of Umeå, Sweden

Sponsors

The First conference on **Th**Ree-way analysis **I**n **C**hemistry is proud to acknowledge the sponsorship of the following organisations:

Royal Netherlands Chemical Society (KNCV)

Groningen Centre for Drug Research (GCDR), University Centre for Pharmacy, Rijksuniversiteit Groningen, Groningen, The Netherlands

Laboratory for Chemometrics, University of Washington, Seattle, Washington, USA

DSM, Geleen, The Netherlands

Shell, Den Haag, The Netherlands

Auspices

The first **TRIC** is held under the auspices of

Vereniging Voor Statistiek en Operationele Research (VVS)

Vereniging voor Ordinatie en Classificatie (VOC)

Werkgroep Overleg Bedrijfsleven Analytische Chemie (WOBAC)

Conference Programme

Sunday 22 August
Registration

Monday 23 August

Morning

Rank of three-way arrays

Donald S. Burdick (Duke University)
Trilinear models from the viewpoint of tensor geometry

Roberto Rocci (University of Rome "La Sapienza") &
Jos M.F. ten Berge* (University of Groningen)
A simplification of a result by Zellini with applications to INDSCAL and the rank of three-way arrays

Alain Franc (CEMAGREF, Clermont Ferrand)
Some rank inequalities for three-way arrays

Joachim von zur Gathen (University of Toronto)
Complexity and rank of three-dimensional arrays

Early afternoon

Informal discussions on three-way rank with Prof. Von zur Gathen

Poster sessions - Software demonstrations

Late afternoon

Sue E. Leurgans* (Rush-Presbyterian-St. Luke's Medical Center) &
Robert T. Ross (Ohio State University)
Multilinear models for energy transfer: Examples and identifiability

Tuesday 24 August

Morning

Algorithms

Alain Franc (CEMAGREF, Clermont Ferand)
Computation of PARAFAC and TUCKER models with metrics and linear constraints

9.45 Bjørn K. Alsberg (University of Bergen)
10:30 *Three-dimensional B-spline compression prior to third-order PCA*

11:00 Ben C. Mitchell (Duke University)
The effect of noise on the resolution of three-way mixtures

11:45 Pieter M. Kroonenberg (Leiden University)
A survey of algorithms for three-way component models

Afternoon

Application of three-way methods

13:30 Richard A. Harshman (University of Western Ontario)
14:30 *Generalizations, extensions, and structured special-purpose modifications of PARAFAC with possible applications in chemistry*

14:30
14:45 Philip K. Hopke (Clarkson University)
Application of TPCA and DTD to airborne particle composition data

15:30 Romà Tauler (University of Barcelona)
Unfolding solutions using multivariate curve resolution in analytical chemistry

Wednesday 25 August

Morning

Diagonal core matrices in Tucker's three-way PCA

René Henrion (Humboldt University of Berlin)
Derivation of bounds for body diagonalization of core matrices in three-way PCA

Roberto Rocci (University of Rome "La Sapienza")
Three-mode factor analysis with binary core and orthonormality constraints

Henk A.L. Kiers (University of Groningen)
Uniqueness of TUCKALS models with zero-constraints on the core

Age K. Smilde (University of Amsterdam)
Second order calibration of a reaction-based chemical sensor using a restricted Tucker model

Afternoon

Applications of three-way methods

Karl S. Booksh (University of Washington)
Recent advances in rank annihilation methods

Abstracts Lectures

Trilinear Models from the Viewpoint of Tensor Geometry

Donald S. Burdick

Tensor algebra provides a conceptual framework into which the methodology for analyzing three-way data fits very nicely. In particular, the geometrical perspective provided by tensor products of vector spaces leads to insights that make three-way methodology much less cumbersome. Unfortunately, much of what has been written in tensor algebra is oriented to physics applications and contains material not especially relevant to trilinear data analysis. This paper will describe the fundamental concepts of tensor algebra associated with tensor products of vector spaces and linear transformations. Emphasis will be given to those concepts which are most relevant to trilinear data analysis.

A SIMPLIFICATION OF A RESULT BY ZELLINI, WITH APPLICATIONS TO INDSCAL AND RANK OF SYMMETRIC THREE-WAY ARRAYS

Roberto Rocci
University "La Sapienza", Rome

Jos M.F. Ten Berge
University of Groningen

Zellini has shown how to decompose an arbitrary symmetric matrix of order n×n as a linear combination of $\frac{1}{2}n(n+1)$ fixed rank one matrices, thus constructing an explicit tensor basis for the set of symmetric n×n matrices. Zellini's decomposition is based on properties of persymmetric matrices.

In the present paper, a simplified proof is given, by showing that a symmetric matrix can also be decomposed in terms of $\frac{1}{2}n(n+1)$ fixed binary matrices of rank one. The decomposition has direct implications for the maximal number of dimensions needed to perfectly fit the INDSCAL model and for the maximal rank of a set of symmetric matrices. In fact, it implies that taking $r=\frac{1}{2}n(n+1)$ dimensions in INDSCAL guarantees the existence of a perfectly fitting solution. Although this value is sharp when the number k of symmetric matrices involved is greater or equal to r, it may overestimate the number of INDSCAL dimensions needed when k<r.

Some rank inequalities on 3-Way arrays

Alain Franc
CEMAGREF
Domaine de Laluas
F63 200 RIOM
FRANCE

Determination of the rank of a 3-way array is a difficult task.

We first present a concise definition of rank, which can easely be extended to N-way arrays, following KRUSKAL's definition.

We then recall main results on arrays of small dimension, whose elements are real, according to litterature.

Through algebraic computations, we then propose some new upper bounds for rank on 3-way arrays, and lower bounds of any dimension. Main idea is to study the minimal number of rank 1 matrices needed to span a vector subspace of a given dimension in the vector space of matrices. The procedure we propose leads to an effective decomposition of a given array as the sum of r arrays of rank 1. If this task can be performed for any matrix, then r is an upper bound of the rank of any 3-way array of given diemension.

We finally reacall that rank decomposition highly depends on the field upon which arrays are build : the problem is much simpler on the complex field, because it can be related fairly easely to the number of solutions of a system of polynomials of several variables, the computation of which is straightforward on the complex field, but much more difficult on the real field.

Complexity and rank of three-dimensional arrays

Joachim von zur Gathen
Department of Computer Science
University of Toronto
Toronto, Ontario, Canada M5S 1A4
gathen@theory.toronto.edu
Phone: (416) 978-6024
Fax: (416) 978-1931

Computational complexity theory investigates the resources required to solve problems by computer. After discussing the general impact of this theory on computing, this survey talk will look more closely at bilinear complexity and general insights into computation with three-dimensional arrays.

Computation of PARAFAC and TUCKER models with metrics and linear constraints on N-ways arrays

Alain Franc

CEMAGREF

Domaine de laluas

F63 200 RIOM

FRANCE

A tool allowing computation of PARAFAC, TUCKER and STATIS models on N-ways arrays is presented.

In fact, computation of simple PARAFAC model (no metrics, no constraints) has been known since the works of CARROLL & CHANG (1970) and HARSHMAN (1970). But the algorithms based on alternatuive least square optimisation are known to converge towards the expected results, but sometimes with long computation time. Taking into acount linear constraints (such as is CADELINC models by KRUSKALL & Al) leads to other optimisation problems. Taking into account distances measurements on spaces has not yet been studied.

We present an isometric transformation of datas and /or a projective transformation on N-way arrays whichs reduces PARAFAC models computation with metrics and constraints (PWMC) to simple PARAFAC model. Hence, any improvement on simple PARAFAC optimisation process can automatically be transfered to PWMC.

But difficulty to improve ALS algorithm of PARAFAC models is then emphasized, through comparison of three techniques : ALS, gradient methods and simulated annealing.

Same techniques are then presented for TUCKER models, whose solution has been known since KROONENBERG and De LEEUW work in 1980 for the 3-way case, and KAPTEYN & Al. and POLIT works in 1986 for the N-way case. An isometry and projection can reduce TUCKER optimisation method with constraints and metrics to the simple case.

Application of the same techniques on 2-way arrays, a well known case, leads to the duality scheme for metrics and ACP-VI for projection, enhancing deep algebraic unity of these problems.

3D B-spline compression prior to third order PCA

Bjørn K. Alsberg * Olav M. Kvalheim

Department of Chemistry
University of Bergen
Allegt. 41
N-5007 Bergen. Norway
e-mail: alsberg@saturn.kj.uib.no
* Corresponding author

Abstract

Third order PCA using Kroonenbergs algorithm is very computer demanding. In addition to more efficient third order PCA algorithms either powerful computers or compression must be used to cope with the rapidly increasing data size. Preliminary results from our laboratory suggest that using the coefficient tensor $\mathbf{C}$ instead of the larger original tensor $\mathbf{X}$ in a third order PCA algorithm reduced the number of FLOPS needed considerably. Similar analyses on second order data sets have been shown to be effective.

A tensor $\mathbf{X}$ with dimensions $[40 \times 40 \times 40]$ was analysed using a third order PCA routine made in MATLAB. It was difficult to use a larger tensor without getting memory problems. A tensor $\mathbf{C}$ with dimensions $[7 \times 7 \times 7]$ was constructed by using B-spline compression. Analysis of the original tensor required over a thousands times more FLOPS than the analysis of the smaller compressed representation. Visual inspection of the latent variables suggested that the results from the $\mathbf{C}$ tensor was very similar to the results from the $\mathbf{X}$ tensor.

The Effects of Noise on the Resolution of Three-Way Linear Mixtures

Ben Mitchell and Don Burdick

The alternating least squares PARAFAC algorithm is a useful tool for resolving trilinear three-way data arrays. However, there is a need for better understanding of its convergence propeties for noisy data. For example, it occasionally becomes bogged down for many iterations in the vicinity of a poor quality resolution before moving on to a much superior optimum fit.
This paper presents the results of a simulation study designed to investigate the behavior of PARAFAC when the data array contains noise. We find that when PARAFAC bogs down, the resolution tends to exhibit a two-factor degeneracy. We find that including a test for two-factor degeneracies is helpful in avoiding the trap of a premature termination at an inferior resolution.
Also considered is the problem of rank estimation. It is found that when two of the signal modes correspond to excitation and emission spectra of the sort encountered in fluorescence spectroscopy, a test for high frequency noise components based on the Fast Fourier Transform does a good job of estimating rank.

THE APPLICATION OF DIRECT TRILINEAR DECOMPOSITION (DTD) TO REAL AIRBORNE PARTICLE COMPOSITION DATA

P.K. Hopke and N. Gao
Department of Chemistry
Clarkson University
Potsdam, New York 13699-5810

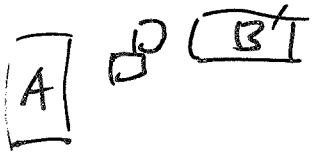
Abstract

Air quality data are typically obtained from samples taken at a number of sites at various times and which are then analyzed for a number of species. Thus, a 3-way data set is obtained that represents the spatial and temporal variations in the composition of the atmosphere. Specifically airborne particle samples are taken and analyzed for multiple elements to provide a suite of elemental concentrations over the space and time frames of the sampling network. Previously we have presented the results of a methodological study using the Direct Trilinear Decomposition (DTD). Using simulated data, DTD proved to be a quantitative model with potentially high power to identify particle sources as well as transformations in composition during transport. We have now applied this approach to regional scale data from the 6 sites in the Acid Precipitation in Ontario Study (APIOS) network covering southern Ontario and to data from8 sites for which samples were obtained during the 1987 Southern California Air Quality Study (SCAQS). The results of these studies will be presented.

Unfolding solutions using multivariate curve resolution in Analytical Chemistry

Romà Tauler, Department of Analytical Chemistry, University of Barcelona, c/ Diagonal 647, 08028 Barcelona, Spain

Three way data matrices are common in modern Analytical Chemistry. The particular structure of the data depends on the analytical method and also on the chemical problem under study. Highly structured data arrangements are obtained for instance in many hyphenated chromatography techniques and also in fluorescence spectroscopy. However even in these cases, small departures from a complete trilinear structure are common. Specially interesting are those cases where one of the orders of the data can change in a non-linear way between slices. In those situations unique solutions are not guarantee and constrained iterative methods are needed. Multivariate curve resolution is a factor analysis based technique which has traditionally applied to the study of two-way data matrices. Degenerate solutions owing to the rotational and intensity ambiguities can only be partly resolved using constraints based in a previous knowledge of the characteristics of the data and eventually solved. Examples of these constraints are non-negativity, closure, continuity, unimodality and other shape characteristics. Extremely helpful is the knowledge of how the rank is distributed along the data matrix and how to detect regions where the rank is very small and eventually the unity. In rank one regions the rotational ambiguity is completely solved and only persists the intensity or scale ambiguity. In the present communication these ideas are extended to the case of three-way data matrices. For these higher order structures additional constraints can be added which can contribute to find out the solutions physically meaningful.



Body Diagonalization of Core Matrices in Three-way Principal Components Analysis

R. Henrion
Humboldt-University of Berlin

Institute of Applied Mathematics
Unter den Linden 6, 1086 Berlin, Germany
FAX: 095-49-30-20932238 Email: henrion@mathematik.hu-berlin.dbp.de

In contrast to conventional PCA a direct superposition and joint interpretation of loading plots is not possible in three-way PCA since there may be data variance which is described by unequal components of different modes. The contributions to variance of all possible combinations of components are described in the core matrix. Body diagonalization, which is achieved by appropriate rotation of component matrices, is an essential tool for simplifying the core matrix structure. The maximum degree of body diagonality which may be obtained from such transformations is analyzed both from the mathematical and simulation viewpoints. It is shown that, at least in the average case, high degrees can be expected which makes the procedure reasonable for many practical applications. Furthermore simulation as well as theoretical derivation show that the success of body diagonality depends on the so called polarity of the core array. This dependence on polarity is explicitely derived for lower and upper bounds of single realizations and of expected values of the maximum degree of body diagonality. The methodology is illustrated by a three-way data example from environmental chemistry (heavy metal exposure of soils with different cultures of fodder plants at different locations of East Germany).

THREE-MODE FACTOR ANALYSIS WITH BINARY CORE AND ORTHONORMALITY CONSTRAINTS

Roberto Rocci
University "La Sapienza", Rome

A constrained version of Three-mode Factor Analysis model is considered in order to make its interpretation easier. The constraints are obtained by fixing some elements of the core to zero (binary core constraints) and requiring orthonormal factor loadings. The aim is to reduce the number of core elements in the model preserving the interpretation adopted by Kroonenberg in Three-mode Principal Components Analysis.

An Alternating Least Squares algorithm is given in order to solve the related minimisation problem. It is built on the fact that the optimal core does not depend on the choice of the binary core constraints when the factor loadings are fixed.

We consider also three different kinds of binary core constraints following different purposes. First, we show how it is possible to impose them without loss of fit. Second, we demonstrate how the interpretation can be improved using a particular set of constraints. They are chosen so that the factor scores matrices are orthogonal and have the core entries and the factor loadings as singular values and singular vectors, respectively. Finally, we focus on the particular case where only two factors for each mode are considered. By means of binary core constraints we relate this model with the PARAFAC model. In detail, we can compute the rank two approximation of the array of data, if it exists. When such approximation fails, i.e. it is not possible to find the best rank two approximation of the array, then it can be shown that the algorithm finds an array of rank three which can be arbitrarily approximated with another of rank two.

Analysis of Three-way Data by Constrained Tuckals Methods

Henk A.L. Kiers
University of Groningen

Address: Department of Psychology, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands; e-mail: h.a.l.kiers@ppsw.rug.nl

Two popular models for representing three-way data are the PARAFAC model and the TUCKALS3 model. The PARAFAC model describes three-way data in a simple way by adding triple products of coefficients. For each dimension there is a unique set of coefficients for each of the three modes. The TUCKALS model is an extension of the PARAFAC model that allows for interactions of components corresponding to different modes and different dimensions. In full generality, this model is much more complex and has many more parameters than the PARAFAC model. Moreover, its components are not unique. However, TUCKALS3 also has certain advantages over PARAFAC. A first advantage is that it fits the data at least as well as a PARAFAC model of the same dimensionality, and usually better. A second advantage is that TUCKALS3 is not burdened with one of the unfavorable properties of PARAFAC: that of finding so called degenerate solutions for certain data sets. Here, it will be proposed to use variants of TUCKALS3 that share some of the good features of both methods. Specificaly, it is proposed to constrain the TUCKALS3 model in such a way that it reduces to a PARAFAC model supplemented with a few (rather than all) interaction terms. It will be demonstrated that such constraints can often be made without seriously affecting the fit of the model. In fact, it will be shown that a considerable number of interactions can always be dropped without loss of fit. In addition, it will be illustrated that models with only a few interaction terms tend to have the same uniqueness properties as PARAFAC. It will be explained how uniqueness can be verified empirically, and some techniques will be described for proving or disproving uniqueness theoretically. Apart from constraints in order to drop interaction terms, other constraints like nonnegativity or equality of parameters will be discussed too. It will be explained how an algorithm for handling all the resulting constrained variants of TUCKALS3 can be constructed. Several of the above issues will be illustrated on exemplary models and an example data set.

Second order calibration of a reaction-based chemical sensor using a restricted Tucker model

Age K. Smilde[1] and Bruce R. Kowalski[2]

[1] University of Amsterdam, Laboratory for Analytical Chemistry, Nieuwe Achtergracht 166, 1018 WV Amsterdam

[2] Center for Process Analytical Chemistry, University of Washington, BG-10, Seattle, WA 98195

For the *in situ* analysis of the chlorinated hydrocarbons chloroform, trichloroethylene and 1,1,1-trichloroethane a chemical sensor is developed. This chemical sensor is based on the Fujiwara reaction. In this reaction the chlorinated hydrocarbons form intermediate and final products in reacting with pyridine and a base. The spectral properties of these intermediate and final species can be used to quantitate the chlorinated hydrocarbons. The hydrocarbons permeate through a membrane in the chemical sensor. Then the Fujiwara reagent is added and the reaction starts at time zero. At regularly spaced time points a UV/VIS spectrum is taken. After a certain cut off point in time the resulting analytical data can be described as a matrix of absorbances dimensioned number of time points versus number of wavelengths. Since the data collected can be represented in a matrix (a second order tensor) the chemical sensor is called a second order sensor.

In an early attempt to use the chemical sensor for *in situ* analysis of chlorinated hydrocarbons only the absorbances at one wavelength were used. This is zero order calibration and the approach has the disadvantage of zero order calibration: the problem of unknown interferences in a new sample (the background problem) is not solved. Using the complete spectrum for the calibration makes the calibration method of first order. Still, the background problem is not solved, but the advantage of using first order instead of zero order is that a background problem can be recognized.

In order to solve the background problem second order calibration has to be used. The chemical sensor produces second order data: a matrix of responses is obtained for each

measured sample. A response matrix of a sample containing only the analyte of interest is called a pure analyte response matrix. If the rank of such a response matrix is one, then generalized rank annihilation (GRAM) can be used to quantitate this analyte in a new mixture. Despite of the possible unknown interferences in the new mixture quantitation of the analyte of interest is possible. This property is called the second order advantage. If the rank of a pure analyte response is more than one then the second order advantage is not necesssarily retained. The method nonbilinear rank annihilation (NBRA) has been proposed as an extention of GRAM to deal with the case of higher than rank one systems.

A complete new way to solve the problem of second order calibration with responses higher than rank one is presented. Since the pure analyte responses generated by the chemical sensor explained above have ranks higher than one, the chemical sensor is used as an application to illustrate the method. The calibration method is based on a new development in three-way analysis. The Tucker model is a way to generalize the singular value decomposition to three-way matrices. The Tucker decomposition of a three-way array is a product of three loading matrices and a (small) core-array. This core-array is a three-way array containing elements representing the magnitude of importance of the latent variables and of interactions between those latent variables. A new class of three-way methods is obtained by setting core-array elements to zero. With this class of models the above mentioned second order calibration for higher than rank one cases can be tackled. The results for the application of this new class of calibration techniques to the chemical sensor data will be presented.

Recent Advances in Rank Annihilation Methods

Karl S. Booksh and Bruce R. Kowalski
Center for Process Analytical Chemistry
University of Washington
Seattle, WA USA

Psychologists calibrate multiway data using iterative least-squares based procedures. Chemists, by way of the mathematical ideas of tensors and orders, solve these calibration problems by obtaining eigenvector based solutions. These higher order data analysis tools developed in chemistry include the Generalized Rank Annihilation Method (GRAM)[1] and Trilinear Decomposition (TLD)[2].

GRAM and TLD are used to determine the concentration of analytes present in each sample and to estimate the pure analyte response in each order, or mode, of analysis. GRAM is useful when a calibration set consists of only two samples, a standard and a mixture. The mixture can contain compounds not included in the sample. TLD is the extension of GRAM to three or more samples.

These two methods assume the trilinear model $\mathbf{R} = \mathbf{C}\mathbf{X}\mathbf{Y}^T$, where the instrument response, R, is the weighted sum of the outer products of the pure component responses in each order, X and Y. For trilinear data, the diagonal weighting matrix, C, contains the concentrations of each component present in the sample.

GRAM and TLD rely on the eigenvector based solutions to estimate the pure component responses in each order. The quantitative information in GRAM is eigenvalue based, while TLD utilizes least-squares fitting of the estimated pure component responses to predict component concentrations.

Li, et. al.[3] have demonstrated cases in which GRAM predicts pure component responses that have significant imaginary components in each order with non-trilinear data. Other cases exist where the analyte of interest does not respond linearly with concentration. In both cases, TLD does not provide reliable results. The algorithm published by Sanchez and Kowalski[2] has been modified and extended to become more robust to these and other potential pitfalls. These advances are applied to the non-trilinear calibration of second order (2 way) optical sensors.

1 Wilson, B., E. Sanchez, and B. R. Kowalski, *J. Chemo.*, 3, 493 (1989).
2 Sanchez, E. and B. R. Kowlaski, *J. Chemo.*, 4, 29 (1990).
3 Li, S., J. Hamilton, and P. Gemperline, *Anal. Chem.*, 64, 599 (1992).

Abstracts Posters

Expressing n'th order data arrays in diagram notation

Bjørn K. Alsberg * Olav M. Kvalheim

Department of Chemistry
University of Bergen
Allegt. 41
N-5007 Bergen. Norway
e-mail: alsberg@saturn.kj.uib.no
* Corresponding author

Abstract

The traditional matrix notation is difficult to use in higher order tensor algebra. In order to accomplish this unfolding and application of the Kronecker product is necessary. For tensors of order higher than three it is difficult to apply the more visual notation which depict third order data matrices as cubes and matrices as planar sheets.

A *supplemental* notation which easily can be applied to any higher order data array is here suggested which is based on a very simplified version of the *Feynman diagrams*, originally invented by Richard P. Feynman in order to simplify a special set of complex tensor algebra equations in theoretical physics. For our purposes the efficient book-keeping function of indices provided by diagrams is important. Each data array is depicted as a circular object with "hands" where each "hand" signify an index. Matrices will have two such "arms", third order data arrays will have three and so on. A link between one or several "arms" of object(s) (can be the object itself) means that a summation (or in general an integration) over the current index is intended. Diagram equations can be manipulated and used in ways analogous to manipulation of matrices. The diagrams have the following advantages over other more traditional tensor notation:

- The equations are intuitive and easier to understand for non-mathematically inclined people.
- There is a one-to-one equivalence to summation formulas and indices
- Symmetries and patterns in tensor equations are easily observed in the notation
- Enables manipulation of tensors of any order without too much trouble

Applying the least squares approach of factorization to solving three-way problems

Pentti Paatero and Unma Tapper,
Dept. of Physics, university of Helsinki

During the last two years, a new approach (PMF) on the factor analytic problem has been developed, based on a least squares formulation. In the first part of the poster, a short summary is given of the Positive Matrix Factorization approach:

The approach is based on the concept of individual error estimates for all the entries in the data matrix. It is shown that the customary approaches (R-mode and Q-mode) are equivalent to such least squares problems where non-realistic error estimates are applied. In "hard" sciences, physics, chemistry, etc. there usually are realistic error estimates available, in contrast to such soft sciences as psychology or biology.

The 2-way model is described by the equation

$$X_{ij} = \sum_{r=1}^{p} G_{ir} F_{rj} + E_{ij}$$

which is solved by minimizing the F-norm of the matrix of quotients

$$E_{ij} / S_{ij}$$

constrained by the inequalities $G_{ir} >= 0$, $F_{rj} >= 0$.

Here X is the observed data matrix, S is the matrix of error estimates of X, and G and F are the unknown "factors" of X.

Theoretical questions are mentioned: uniqueness of solution, presence or non-presence of free rotations in the solution, comparison with standard methods (PCA, SVD). Practical applications of the method (in environmental problems, aerosol physics, astronomy) have been extremely successful. One practical merit of the method is that missing values and outliers can be handled by increasing their error estimates.

In the second part of the poster, the application of PMF approach to 3-way problems is discussed. There are different ways for extending the 2-way approach to higher dimensions. The following will be used:

$$X_{ijk} = \sum_{r=1}^{p} G_{ir} F_{jr} W_{kr} + E_{ikj}$$

where G, F, and W are the unknown factor matrices, other notation is the same as for the 2-way model. This model would describe some chemical instruments, e.g. a combined

chromatograph-optical spectrometer applied on a batch of similar samples which are assumed to consist of same basic constituents. Then the three axes would be: sample #, retention time, and optical wavelength.

The question of free rotations is discussed: if two components differ in all three dimensions, then there are no rotations between them. But if they are similar in one dimension, then the results from 2-way modelling apply; there may or may not be free rotations, depending on the shapes of the factors.

Algorithms are discussed but no "best" algorithm can be named. Examples are shown of synthetic and (hopefully) real test data.

Analytical 3-Way Data from Multiple Sites Air Monitoring, Mapped by Prim's Minimal Spanning Tree and Projected into a Kohonen Neural Network

Dietrich Wienke*, Ning Gao, Philip K. Hopke

wienke@sci.kun.nl, gaon@craft.clarkson.camp.edu,
hopkepk@craft.clarkson.camp.edu
Clarkson University, Department of Chemistry, Potsdam (NY 13699-5810), U.S.A.
* on leave: Catholic University of Nijmegen, Laboratory of Analytical Chemistry, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Air Quality Monitoring Networks of $i = 1..m$ sampling sites are used to monitor the multivariate concentrations of $j = 1..p$ species over the time $k = 1..n$. An analysis of the obtained m x p x n data block by Three Mode Factor Analysis (TMFA) and Direct Trilinear Decomposition (DTD) can provide the number of air pollution sources, their individual source emission profile and their individual source contribution for a given time k. However, DTD and TMFA do not provide any geographical information. An additional theoretical limitation can be that both methods require a synchronized operating sampling network. Prim's Minimal Spanning Tree projected into a Kohonen Neural Network, can map 3-way environmental analytical data from asynchron operating sampling networks. For this aim the $m * n$ p-dimensional sample vectors are arranged in random order in a two-dimensional data matrix of dimension N x p with $N = m * n$. The choice of a two-dimensional Kohonen Neural Network forces the N data vectors to self-organize as a qualitative correct, geographical map. However, the correct geographical directions but wrong distances between the extracted sources of air pollution can be further quantified by an additional projection of Prim's Minimal Spanning Tree into the Kohonen Map. The application of this hybrid algorithm, called 3MAP, to the Southern California Air Quality Study (SCAQS'87) provided a correct geographical pattern of several sources of air pollution. The number of sources and their averaged emission profile as found by 3MAP are in agreement with results provided by a DTD analysis of the three-dimensional data block.
A computer demonstration of 3MAP will be arranged.

Analyzing three-way skew-symmetric matrices

Berrie Zielman

Department of Data Theory

University of Leiden

PO Box 9555

2300 RB Leiden

The Netherlands

Email: Zielman@rulfsw.LeidenUniv.nl

Abstract

Tables where the rows and columns classify the same set of n objects occur frequently in the social sciences. The objects or stimuli can be anything; for example personality disorders, soft drinks, occupational categories and so on. These tables can be asymmetric, for example in social mobility it is unlikely that the number of sons from farmers who become professionals equal the number of sons from professionals who become farmer. Asymmetric tables Δ can be decomposed into a *symmetric* and a *skew-symmetric* part:

$$\Delta = S + A,$$

where the matrix S is symmetric; this matrix is obtained by averaging the elements across the diagonal. The matrix A describes the departures from symmetry, this matrix A has the property of *skew symmetry*: $a_{ij} = -a_{ji}$. The matrix S can be analyzed by a symmetric method like cluster analysis or multidimensional scaling. Although most examples of skew-symmetry analysis are found in the social sciences, some chemistry applications are also reported in the literature (Gower, 1980).

Gower (1977) proposed to analyze the matrix A by singular value decomposition. The singular value decomposition reflects the skew symmetric structure of the matrix A. Diagrams obtained by plotting pairs of the singular vectors can be interpreted in terms of areas and collinearities (Gower, 1977).

This paper develops methods for analyzing multiple skew-symmetric tables simultaneously. The models can be regarded as INDSCAL (Carroll and Chang, 1970) and IDIOSCAL analogues for skew-symmetric matrices. The proposed model yields a common space that can be interpreted in terms of areas of triangles formed by two points and the origin and a set of weights indicating the importance of pairs of the common space. Because we wish to retain the area interpretation in our three-way model we have to weigh pairs of dimensions simultaneously instead of single