Some new methods for three-mode factor analysis and multi-set factor analysis
Lam, Tam

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 7

Conclusion

In this concluding chapter we classify the types of models discussed in this thesis, provide a brief discussion of model choice, and consider options for future research.

The component and factor models for three-mode and multi-set data (new and existing) that have been discussed in this thesis can be arranged in a $2 \times 2 \times 3$ classification array. The first mode of the array represents the three-mode/multi-set dichotomy. The second mode indicates whether the model is rotationally unique or not. The third mode represents the type of model: direct-fitting, indirect-fitting, or a factor model. Direct-fitting models are fitted on the observed data itself, which is given by $X_{(N \times JK)}$ for three-way data, and $X_k$, $k = 1, \ldots, K$, for multi-set data. Indirect-fitting models are fitted to the observed cross-products of the data, which we divide by the number of observations $N$ to make a comparison with the factor models. Thus, the cross-products are given by $\Sigma = N^{-1}X_{(N \times JK)}^T X_{(N \times JK)}$ for three-way data, and $\Sigma_k = N^{-1}X_k^T X_k$, $k = 1, \ldots, K$ for multi-set data. For the factor models, we assume mean-zero data and fit the model to the above cross-products, which are now covariance
matrices. Additionally, we assume a unique part that is uncorrelated with the factors for the common part.

Table 7.1: Classification of three-mode and multi-set component and factor models discussed in this thesis.

<table>
<thead>
<tr>
<th>three-mode</th>
<th>( \mathbf{X}_{(N \times JK)} = \mathbf{F} (\mathbf{C} \otimes \mathbf{B})^T )</th>
</tr>
</thead>
<tbody>
<tr>
<td>rotationally unique</td>
<td>( \mathbf{\Sigma} = (\mathbf{C} \odot \mathbf{B}) \mathbf{\Phi} (\mathbf{C} \odot \mathbf{B})^T + \mathbf{U} )</td>
</tr>
<tr>
<td>three-mode</td>
<td>( \mathbf{X}_{(N \times JK)} = \mathbf{F} \mathbf{G} (\mathbf{C} \otimes \mathbf{B})^T )</td>
</tr>
<tr>
<td>rotationally nonunique</td>
<td>( \mathbf{\Sigma} = (\mathbf{C} \otimes \mathbf{B}) \mathbf{G}^T \mathbf{\Phi} \mathbf{G} (\mathbf{C} \otimes \mathbf{B})^T + \mathbf{U} )</td>
</tr>
<tr>
<td>multi-set</td>
<td>( \mathbf{X}_k = \mathbf{F}_k \mathbf{C}_k \mathbf{B}^T, \quad \mathbf{F}_k^T \mathbf{F}_k = \mathbf{F}_l^T \mathbf{F}_l )</td>
</tr>
<tr>
<td>rotationally unique</td>
<td>( \mathbf{\Sigma}_k = \mathbf{B} \mathbf{C}_k \mathbf{\Phi} \mathbf{C}_k \mathbf{B}^T + \mathbf{U}_k )</td>
</tr>
<tr>
<td>multi-set</td>
<td>( \mathbf{X}<em>k = \mathbf{F}<em>k \left( \sum</em>{q=1}^{Q} c</em>{kq} \mathbf{G}_q \right) \mathbf{B}^T, \quad \mathbf{F}_k^T \mathbf{F}_k = \mathbf{F}_l^T \mathbf{F}_l )</td>
</tr>
<tr>
<td>rotationally nonunique</td>
<td>( \mathbf{\Sigma}<em>k = \mathbf{B} \left( \sum</em>{q=1}^{Q} c_{kq} \mathbf{G}<em>q \right)^T \mathbf{\Phi} \left( \sum</em>{q=1}^{Q} c_{kq} \mathbf{G}_q \right) \mathbf{B}^T + \mathbf{U}_k )</td>
</tr>
</tbody>
</table>

In Table 7.1 above, the models are given for each combination of the first two modes of classification. In our classification, rotationally unique three-mode models are of CP form, rotationally nonunique three-mode models are of the Tucker form, rotationally unique multi-set models are of Parafac2 form, and rotationally nonunique multi-set models are of multi-set Tucker form. The last model type may be new in the literature and is not discussed in this thesis. The three-mode factor models in Table 7.1 are easy to extend to more than three modes. The indirect and factor multi-set models are obtained from the corresponding three-mode models by taking the \( k \)th \( J \times J \) diagonal block of the \( JK \times JK \) matrix \( \mathbf{\Sigma} \). Note that Table 7.1 is not meant as a classification of all possible or existing models of these types. For example, there also exist other
types of multi-set component models (Timmerman & Kiers, 2003).

The choice of model obviously depends on the data type. For three-mode data, we can choose between three-mode models or multi-set models. It may seem obvious to prefer three-mode models in this case, but multi-set models also have advantages when fitted to three-mode data (De Roover, Timmerman, Van Mechelen, & Ceulemans, 2013). For the choice between a rotationally unique or nonunique model, we recommend to try both. In our classification, the rotationally unique models (CP-based) are a special case of rotationally nonunique models (Tucker-based). After rotating, the Tucker-based solutions can be compared to solutions from CP-based models. As we have seen in Chapter 5, for one three-mode dataset the CP-based factor model may be more suitable, while for another three-mode dataset the Tucker-based factor model yields a clearer interpretation. Naturally, the Tucker-based models involve more elaborate choices regarding numbers of components and rotation methods.

The choice between a component model (direct or indirect) or a factor model is a fundamental one. Component models aim at a best-fitting summary of the data as it is observed. The components are found as linear combinations of the observed variables. Factor models aim at taking into account measurement error and aspects measured by single observed variables only, by assuming a unique part for each observed variable. A summary of the common part of the observed variables is then sought in terms of the common factors, where the latter are usually not linear combinations of the observed variables. For two-way data, a lively discussion of PCA versus (exploratory) common factor analysis (EFA) can be found in Velicer and Jackson (1990). PCA can be favored because of its computational simplicity and manifest component approach, whereas EFA is computationally more difficult and features latent factors. PCA and EFA are said to yield similar estimated loadings in simulation studies. However, the
PCA loadings are difficult to compare for different samples, since no unique parts have been specified. Apart from the results from simulation studies, the same arguments apply to the three-mode and muti-set component and factor models in Table 7.1.

The factor models in this thesis are more difficult to estimate than the corresponding component models. We have proposed a two-step estimation procedure for the factor models, where we first use MRFA to estimate the unique variances. Next, the covariance matrix (or matrices) of the common part(s) is decomposed and the corresponding direct-fitting component model is used to estimate the factor loadings. Using MRFA has as advantage that percentages of explained common variance can be computed, overall and for each observed variable separately. A disadvantage may be that MRFA imposes an additional constraint ($\Sigma - U \geq 0$) on the factor solution, which is assumed to be true in the population but may not hold in the sample at hand. Also, our use of MRFA is not optimal since the form of the model for the common part is not taken into account when estimating the unique variances. However, despite these potential limitations, the simulation studies in Chapter 4 and Chapter 6 have shown that our rotationally unique factor models and estimation methods are successful in retrieving the true underlying factor loadings.

Finally, we indicate some topics for future research. As in Chapters 4 and 6, a simulation study could be done to assess the performance of the estimation method for the Tucker3 three-mode factor model in Chapter 5. This involves choosing a rotation method to rotate the true and estimated solutions to simple structure. Also, it would be interesting to apply our approach to the multi-set Tucker3 factor model (the final model in Table 7.1). For this, we propose an analogous estimation procedure as for the multi-set Parafac2 factor model in section 6.3.2. In step 3 of this procedure, the direct-fitting multi-set Tucker3
model should be fitted. This can be done analogous to the direct Parafac2 algorithm in section 6.2.1, by replacing the fitting of CP in step 2 of this procedure by the fitting of Tucker3. As for the other factor models, a simulation study can be conducted and the model should be fitted to multi-set data. For example, it would be interesting to fit the multi-set Tucker3 factor model to the SPPC dataset in Chapter 6 and compare the solution(s) to the solution of the multi-set Parafac2 factor model.

Another topic for future research is to conduct simulation studies to compare the performance of the factor models in Table 7.1 to that of the corresponding indirect component models, i.e., without a unique part. As mentioned above, of PCA and EFA it is said that they yield similar loading estimates. Would this be true for the three-mode and multi-set component and factor models as well? Or would the suboptimal two-step approach hamper the performance of the estimation procedures of the factor models, as compared to fitting the corresponding component model? Or would the performance rather be hampered by the use of MRFA instead of a less restrictive MINRES approach? These are interesting research directions to pursue. In our opinion, however, the choice between a component model and a factor model remains a fundamental one that should not be based solely on estimation accuracy.
Bibliography


Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young de-


Domanov, I., & De Lathauwer, L. (2013b). On the Uniqueness of the Canonical Polyadic Decomposition of Third-Order Tensors-Part II: Uniqueness of the


and three-way Dedicom. *Computational Statistics & Data Analysis, 16*(1), 103–118. doi: http://dx.doi.org/10.1016/0167-9473(93)90247-Q


Pearson, K. (1901). On lines and planes of closest fit to systems of points in


Summary

This thesis discussed methods for Exploratory Component and Factor Analysis of multi-way and multi-set data. In particular, new methods for multi-way and multi-set factor analysis were proposed and demonstrated. Following is a summary of the contents of each chapter of the thesis.

In Chapter 2 we introduced the concepts of multi-way and multi-set component and factor models whose comprehension was crucial to fully understand all the subsequent chapters of this thesis. These concepts concern two-way decompositions (e.g., Singular Valued Decomposition, Principal Component Analysis), three-way decompositions (e.g., Candecomp/Parafac, Tucker3), multi-set component and factor analysis (e.g., Parafac2, Simultaneous Component Analysis models, Multi-set Parafac2 common factor model). Additionally, we discussed the uniqueness properties of both Candecomp/Parafac and Tucker3 models, and algorithms to fit these models.

Chapter 3 was devoted to multi-way exploratory analysis of 4-way BJW data in which the subjects are asked how strongly they believe that a number of 6 actors (Nature, God, Human Institutions, Other People, Yourself, and Chance) bring about justice in the world for themselves or other people. The 4 modes of the dataset are: 345 subjects, 7 items, 6 actors, and 2 perspectives (Yourself and Others). In this chapter, we included the results of preliminary analysis of correlations and t-tests, and of PCA on the BJW dataset. Moreover, we also fitted the Tucker3 and Tucker4 models to this dataset. Compared to PCA on unfolding data, the Tucker3 and Tucker4 solutions are almost the same. They also have simple structures and easy interpretation. However, they are a more parsimonious summary than PCA solutions of 2-way unfoldings of the dataset. The Tucker4 solution is the most parsimonious summary because it is
the solution of the complete dataset. The Tucker4 rotation was discussed at the end of this chapter.

Chapters 4, 5, 6 proposed and demonstrated new methods for three-mode and multi-set factor analysis using MRFA to estimate unique variances, and using CP, Tucker3 (for three-mode factor analysis), and Parafac2 (for multi-set factor analysis) to fit the covariance matrix of common part.

In Chapter 4, we did a simulation study that showed that three-mode factor analysis using CP performed very well in retrieving underlying factors when the data was randomly sampled from a normal distribution with a covariance matrix satisfying the CP covariance model. The solution of this model is unique up to permutation and scaling due to uniqueness properties of CP.

In Chapter 5, we extended the CP-based method in Chapter 4 to three-mode factor analysis using MRFA and Tucker3. Since solutions of this method are not unique due to non-uniqueness of Tucker3, a rotation method is needed. This extension was demonstrated by means of two applications to datasets in literature. The results of these two applications showed that one factor model is not always more appropriate than the other. In this chapter, we used the Joint Orthomax rotation in (Kiers, 1998a) to obtained simple structure of Tucker3 matrices.

In Chapter 6, we proposed and demonstrated an exploratory multi-set factor model with common covariance part of indirect Parafac2 form. The simulation study showed that our relatively simple procedure for the multi-set Parafac2 factor model performs very well in retrieving underlying factors when the data is randomly sampled with true covariance matrices satisfying the indirect Parafac2 model. The solution of this method is unique due to uniqueness properties of indirect Parafac2.

For our methods in Chapters 4, 5, 6, one can compute the percentage of
explained common variance which is usually not possible for other methods of three-mode or multi-set factor analysis. The reason for that is we use MRFA to estimate the unique variances. Moreover, the algorithms that we proposed are simple and easy to run. Our solutions are easy to interpret and our models are parsimonious.

In Chapter 7, we provided a classification of the three-mode and multi-set component and factor models featured in the thesis. Analogous to Tucker3-based three-mode factor model of Chapter 5, a Tucker3-based multi-set factor model exists that can be estimated analogous to the Parafac2 multi-set factor model of Chapter 6. This would be an interesting topic for future research.
Samenvatting (Summary in Dutch)

Dit proefschrift bespreekt exploratieve methoden voor componenten- en factoranalyse van meer-weg en meer-set data. In het bijzonder worden nieuwe methoden voor meer-weg en meer-set factoranalyse voorgesteld en gedemonstreerd. Nu volgt een samenvatting van de inhoud per hoofdstuk.

In hoofdstuk 2 bespreken we de achterliggende wiskundige concepten van meer-weg en meer-set componenten- en factormodellen. Deze zijn nodig om de volgende hoofdstukken volledig te kunnen begrijpen. Achtereenvolgens bespreken we twee-weg ontbindingen (e.g., de singuliere waarden ontbinding, principale componenten), drie-weg ontbindingen (e.g., Candecomp/Parafac (CP), Tucker3), meer-set componenten- en factormodellen (e.g., Parafac2, Simultaneous Component Analysis, meer-set Parafac2 common factor model). Verder bespreken we de uniciteitseigenschappen van zowel Candecomp/Parafac als Tucker3 en algoritmen om deze modellen te fitten.

In hoofdstuk 3 wordt een exploratieve analyse gedaan van vier-weg data afkomstig van een Belief in a Just World vragenlijst. De personen werd gevraagd hoe sterk zij geloven dat een aantal actoren (de natuur, god, menselijke instituties, andere mensen, jijzelf en het toeval) zorgt voor rechtvaardigheid in de wereld voor hemzelf of voor anderen. De vier ‘wegen’ in de dataset zijn: 345 personen, 7 items, 6 actoren en 2 perspectieven (voor jezelf en voor anderen). Eerst kijken we naar correlaties van items en actoren, naar t-toetsen om significante verschillen in scores op te sporen en naar principale componentenanalyse van twee-weg vormen van de dataset. Daarna fitten we Tucker3 en Tucker4 modellen. Vergeleken met de principale componentenanalyse, leveren de Tucker3 en Tucker4 modellen bijna hetzelfde resultaat. Ook hebben ze eenvoudig te interpreteren componenten. Echter, de Tucker3 en Tucker4 modellen bevatten veel
minder parameters dan de principale componentenanalyse. De Tucker4 oplossing bevat de minste parameters en wordt gefit op de complete vier-weg dataset. Aan het eind van het hoofdstuk bespreken we ook hoe de Tucker4 oplossing naar een eenvoudige vorm geroteerd kan worden.

In hoofdstukken 4, 5 en 6 introduceren en demonstreren we nieuwe modellen voor drie-weg en meer-set factoranalyse. We gebruiken Minimum Rank Factor Analysis (MRFA) om de unieke varianties te schatten en CP of Tucker3 (voor drie-weg factoranalyse) en Parafac2 (voor meer-set factoranalyse) voor het schatten van het gemeenschappelijke deel van de geobserveerde covarianties.

In hoofdstuk 4 doen we een simulatiestudie die laat zien dat drie-weg factoranalyse met MRFA en CP de onderliggende factoren goed schat wanneer de data normaal verdeeld is en de theoretische covariantiematrix aan het CP covariantiemodel voldoet. De oplossing van dit model is uniek tot op permutatie en schaling vanwege de uniciteitseigenschappen van CP.

In hoofdstuk 5 breiden we het CP covariantiemodel van hoofdstuk 4 uit naar drie-weg factoranalyse met MRFA en Tucker3. Omdat oplossingen van dit model niet uniek zijn (door de niet-uniciteit van Tucker3), is een rotatiemethode nodig. De uitbreiding naar Tucker3 wordt gedemonstreerd bij twee datasets uit de literatuur. De conclusie is dat het ene model niet altijd te prefereren is boven het andere model. Met andere woorden, beide modellen hebben bestaansrecht. In dit hoofdstuk gebruiken we de Joint Orthomax rotatiemethode van Kiers (1998a) voor de Tucker3 oplossingen.

In hoofdstuk 6 introduceren en demonstreren we een exploratief meer-set factormodel waarbij het gemeenschappelijke deel van de geobserveerde covarianties gemodelleerd wordt met behulp van het indirekte Parafac2 model. Een simulatiestudie laat zien dat onze relatief eenvoudige schattingmethode de onderliggende factoren goed schat wanneer de data normaal verdeeld is en de the-
oretische covariantiematrices aan het Parafac2 covariantiemodel voldoen. De oplossing van dit factormodel is uniek vanwege de uniciteitseigenschappen van het indirekte Parafac2 model.

In onze factormodellen uit hoofdstukken 4, 5 en 6 kan een percentage verklaarde gemeenschappelijke variantie worden berekend, hetgeen voor andere modellen voor drie-weg of meer-set factoranalyse meestal niet mogelijk is. De reden hiervoor is dat we gebruik maken van MRFA voor het schatten van de unieke varianties. Verder zijn de algoritmen voor het schatten van onze factormodellen relatief eenvoudig en makkelijk te runnen. Onze oplossingen zijn makkelijk te interpreteren en onze modellen bevatten relatief weinig parameters.

In hoofdstuk 7 maken we een classificatie van de drie-weg en meer-set componenten- en factormodellen uit dit proefschrift. Analoog aan het Tucker3 covariantiemodel van hoofdstuk 5, bestaat er een Tucker3 meer-set covariantiemodel dat geschat kan worden analoog aan het Parafac2 meer-set covariantiemodel van hoofdstuk 6. Dit zou een interessant onderwerp kunnen zijn voor verder onderzoek.
Curriculum Vitae

Full name: LAM Thi Thanh Tam
Full name with accent in Vietnamese: Lâm Thị Thanh Tâm

LAM Thi Thanh Tam was born on 14 July 1980 in Ba To, Quang Ngai, Vietnam. She received her B.Sc. and M.Sc. degrees in Mathematics from Quy Nhon University, Vietnam, in 2002 and 2006, respectively. She has been a lecturer at the Department of Mathematics, Quy Nhon University, Vietnam since September 2002.

Since February 2011 she has been a Ph.D. student at Psychometrics and Statistics Group, Heymans Institute for Psychological Research, the Faculty of Behavioural and Social Sciences, the University of Groningen. Her Ph.D. project was funded by NWO (Dutch Scientific Research Organization) under the supervision of drs. A.W. (Alwin) Stegeman.

Her Ph.D. research is in the field of Psychometrics and Statistics. Her main research interests are focused on multi-way and multi-set factor analysis.
Acknowledgements

The research presented in this thesis had been performed within Psychometrics and Statistics Group, Heymans Institute for Psychological Research, the Faculty of Behavioural and Social Sciences, the University of Groningen, the Netherlands in the past four years. This work has finished under the help, advice and contribution of my supervisor, my colleagues, and my friends. I would like to take an opportunity to thank them.

First and foremost, I am grateful and would like to thank Professor Rob Meijer and Associate Professor Alwin Stegeman for accepting me as their Ph.D. student at the Psychometrics and Statistics group and for supervising me for the past four years. As my promotor, Rob has given me tremendous freedom to work on my own, but he was always willing to give me advice and support whenever I needed. I truly feel so thankful him for everything he did for me. I would also like to express a deepest thanks to my supervisor, Alwin. He was everything I needed as a Ph.D. student, academically, and spiritually. I appreciate all his contribution of time, scientific ideas, many efforts, and kind heart to make my Ph.D. experience productive and stimulating. How much I would like to thank him for his thorough guidance, constructive discussion, positive remarks, and encouragement.

Next, I gratefully acknowledge the financial support of Dutch Scientific Research Organization (NWO) and the University of Groningen. Special thanks go to the secretaries of the Psychometrics and Statistics group for their kind support. Also, I am deeply grateful to the Department of Mathematics, Quy Nhon University, Vietnam for accepting my absence from work at there for a long period of time.

For this dissertation, I would also like to thank my reading committee mem-
bers for their time, interest, and constructive comments.

I would like to thank my colleagues with whom I shared so many things and from whom I received a lot of helps. My special thanks first go to my roommates Tanja Krone and Mariska Barendse for their helps, nice daily discussions, and for sharing the office. Many thanks also go to the colleagues and professors in the group of Psychometrics and Statistics. I would like to thank them for all interesting discussions and research presentations during our group seminars over the past years.

It is my pleasure to thank Katherine Stroebe of the Social Psychology Department, and my promotor Rob Meijer for providing us with their data. And I would also like to thank professor Henk A. L. Kiers for his helpful Matlab programs.

I would like to thank all my Vietnamese friends in Groningen who helped me, shared abroad living experience, and made my life in the Netherlands being less boring with unforgettable memories. Special thanks go to my dearest friends: Hao, Loan, Bay, Thang, Anh Nguyen, Tinh, Anh Pham, Hoa, Cuong-Huong for being by my side, encouraging, and supporting me.

Last but most of all, my greatest thanks must go to my dear family. I first would like to thank my parents, my brother and sister for their endless love and raising me. I would also like to thank my dear husband Quang Thuan and my dearest son Minh Khoi for all their love, always being my side, encouraging, support, and giving me power throughout these years.