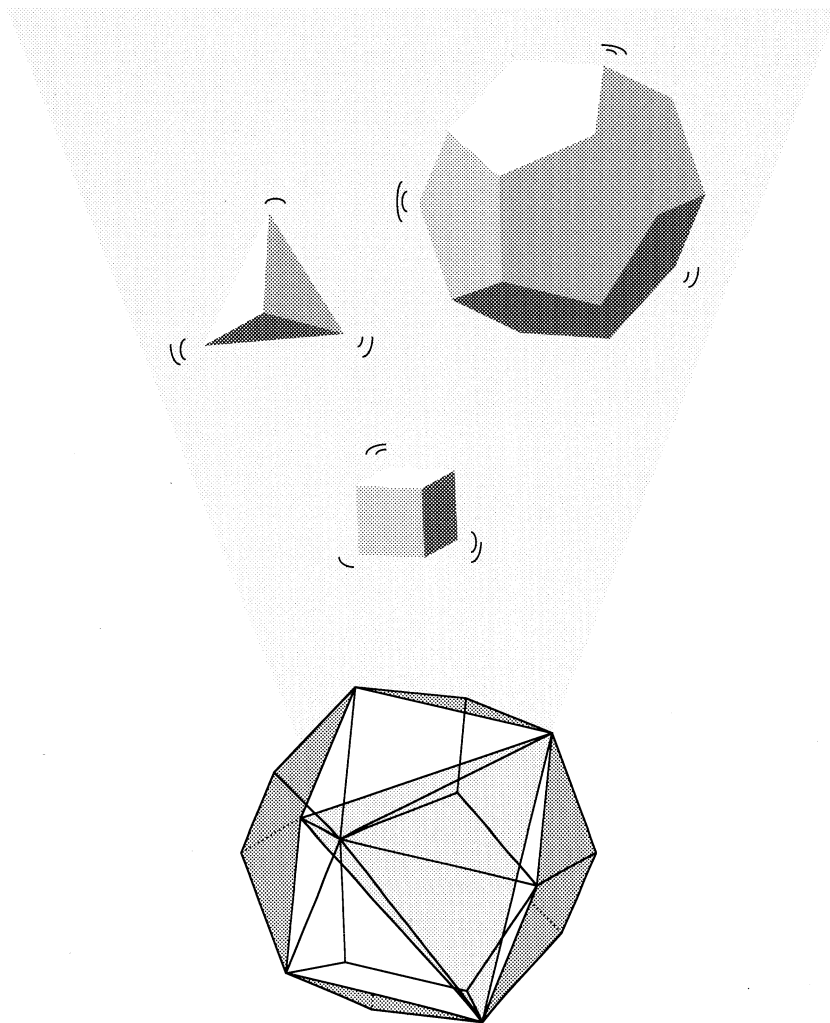


# MATCHING CONFIGURATIONS

Jacques J.F. Commandeur



**DSWO PRESS**



## MATCHING CONFIGURATIONS

## M&T SERIES 19

### Editorial Board:

Prof. dr. J. M. F. ten Berge  
Prof. dr. W.J. Heiser  
Prof. dr. L.J.Th. van der Kamp  
Prof. dr. J. de Leeuw

Technical Editor: L. Delvaux

Available from DSWO Press  
Leiden University  
P.O.Box 9555  
2300 RB Leiden  
The Netherlands  
Tel. (071) 273795 / 273794

### Earlier publications in this series:

Jacqueline Meulman, Homogeneity analysis of incomplete data.

M&T series 1, 1982

Pieter M. Kroonenberg, Tree-mode principle component analysis.

M&T series 2, 1983, reprint 1989

Jan de Leeuw, Canonical analysis of categorical data.

M&T series 3, 1984

Ronald A. Visser, Analysis of longitudinal data in behavioural and social research

M&T series 4, 1985

John P. van de Geer, Introduction to linear multivariate data analysis.

M&T series 5, volume 1 & 2, 1986

Jacqueline Meulman, A distance approach to nonlinear multivariate analysis.

M&T series 6, 1986

Jan de Leeuw, Willem Heiser, Jacqueline Meulman, Frank Critchley (editors), Multidimensional data analysis.

M&T series 7, 1987

Peter G.M. van der Heijden, Correspondence analysis of longitudinal categorical data.

M&T series 8, 1987

Jan van Rijckevorsel, The application of fuzzy coding and horseshoes in multiple correspondence analysis.

M&T series 9, 1987

Abby Israëls, Eigenvalue techniques for qualitative data.

M&T series 10, 1987

Eeke van der Burg, Nonlinear canonical correlation and some related techniques.

M&T series 11, 1988

Kees van Montfort, Estimating in structural models with non-normal distributed variables: some alternative approaches.

M&T series 12, 1989

Jan T. A. Koster, Mathematical aspects of multiple correspondence analysis for ordinal variables.

M&T series 13, 1989

Catrien C. J. H. Bijleveld, Exploratory linear dynamic systems analysis.

M&T series 14, 1989

Henk A. L. Kiers, Three-way methods for the analysis of qualitative and quantitative two-way data.

M&T series 15, 1989

Stef van Buuren, Optimal Scaling of Time Series.

M&T Series 16, 1990

Rien van der Leeden, Reduced rank regression with structured residuals.

M&T Series 17, 1990.

Rian van Blokland-Vogelzang, Unfolding and group consensus ranking for individual preferences.

M&T Series 18, 1991.

Jacques J.F. Commandeur, Matching Configurations.

M&T Series 19, 1991



MATCHING CONFIGURATIONS

JACQUES J.F. COMMANDEUR

*DEPARTMENT OF PSYCHOMETRICS AND RESEARCH  
METHODOLOGY  
LEIDEN UNIVERSITY*

1991 DSWO Press, Leiden University

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Commandeur, Jacques J.F.

Matching Configurations / Jacques J.F. Commandeur. — Leiden: DSWO Press, Leiden University. — Ill. — M&T series; 19)

Also published as thesis Leiden, 1991. — With ref. —With summary in Dutch.

ISBN 90-6695-052-8

SISO 517 UDC 51-7

Subject headings: matching / alternating least squares.

© 1991 DSWO Press, Leiden University and Jacques J.F. Commandeur

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the publisher.

This book was written on a Macintosh Plus and SE/30, using the Laser Author and MacWrite II wordprocessors. Proofs and final prints were produced on a LaserWriter II NTX. The text was set in Times.

Special symbols were taken from the Symbol and Bartholomew fonts. Graphics were created with MacDraw, MacDraw II and Superpaint. Programming and computations were done in APL.68000.

Cover drawing: Jacques J.F. Commandeur

Printed by 'Reprodienst Faculteit Sociale Wetenschappen', Rijksuniversiteit Leiden and by printing office 'Karstens drukkers b.v.' Leiden'

ISBN 90-6695-052-8



*To Spip*



## Preface

This book is intended for anyone interested in the subject of comparing two or more configurations. The book starts out from the PINDIS framework, and discusses five models for comparing configurations containing the coordinates of the same objects in a number of dimensions. With these models one can investigate the similarity of configurations, as well as the existence of a number of systematic patterns in the differences between configurations. The book introduces improved estimation procedures of the corresponding model parameters, and provides a generalized framework allowing one to compare configurations even when they contain information about different numbers of objects.

Many people helped in the realization of this book. I first would like to mention all the people referenced at the end of the book. They, and many others, have erected the edifice upon which I could continue to build. On a more personal note, I am greatly indebted to Wim van der Kloot. He taught me the basics of data analysis, aroused my interest in and enthusiasm for this branch of science, and put me on the track of the subject of matching configurations. Both Wim van der Kloot and Leo van der Kamp gave me the opportunity and time to develop the algorithms presented in this book, supported me, scientifically and otherwise, throughout the project, and gave many helpful comments on all earlier drafts of the manuscript. John van de Geer was instrumental in shedding a new light on my view of data analysis. Due to his geometrical approach, as well as to his eagle-eye, I became aware of an inherent order in the previously bewildering chaos of data analysis methods. I thank him for his comments on a previous version of the book. The contribution of Jos ten Berge to the present work has been invaluable. He suggested many important improvements and was always quick as lightning in answering any questions I had. I learned a lot from his extensive knowledge of the algorithmic aspects of data analysis. Most importantly, he taught me the lazy man's way to optimization, as is witnessed by the complete absence of calculus in the present study. Ab Mooijaart, Willem Heiser, and Pieter Kroonenberg gave helpful comments on earlier versions of the manuscript. Countless times Wim den Brinker solved problems I had with computer software, A Programming Language in particular. Job van de Raad and Ali Flohr were always

prepared to help me, and Lutgart Delvaux gave very useful advice concerning the layout of this book.

Renée Verdegaal, Peter Verboon, and Frits Meijerink stood by me through thick and thin. I thank them for their kindness and friendship. With Rien van der Leeden I continuously shared a røøm for no less than eight years. Among lots of other things, I thank him for the fact that, although he had to put up with my growing demands for peace and quiet in the røøm, this did not affect our friendship in the least.

Jacques Commandeur  
Leiden, december 1990

# Contents

<b>1 Introduction</b>	<b>1</b>
1.1 Preliminaries	2
1.2 The PINDIS models of Lingoes and Borg	4
1.3 Problems with PINDIS	8
1.4 MATCHALS as an alternative	10
1.5 Restrictions	11
1.6 Overview	12
<b>2 Generalized Procrustes analysis</b>	<b>13</b>
2.1 Introduction	14
2.2 The direct and the centroid approach: geometry and algebra	18
2.2.1 Overview of the solutions	23
2.3 The centroid approach	25
2.3.1 Translations	25
2.3.2 The configuration of centroids $\mathbf{Z}$	27
2.3.3 The elimination of $\mathbf{Z}$	30
2.3.4 Orthonormal transformations	31
2.3.5 Scaling constants	32
2.4 The algorithm	37
2.5 Analysis of variation in GPA with missing data	41
2.6 GPA with missing data and MAXNEAR, MAXBET and MAXDIFF	50
2.7 Illustrations	52
2.7.1 Regular solids	52
2.7.2 Personality traits	58

<b>3 The dimension weighting models</b>	<b>65</b>
3.1 Introduction	66
3.2 Geometry and algebra	67
3.3 The DIMFREE model	73
3.3.1 Introduction	73
3.3.2 Translations	74
3.3.3 Orthonormal transformations	75
3.3.4 Dimension weights	76
3.3.5 The centroid configuration $\mathbf{Y}$	77
3.3.6 The direct approach in DIMFREE	78
3.3.7 Uniqueness of the DIMFREE solution	79
3.3.8 The algorithm	81
3.4 The DIMIDIO model	83
3.4.1 Introduction	83
3.4.2 Translations	83
3.4.3 Reducing the estimation of $\mathbf{Q}$ , $\mathbf{S}$ , and $\mathbf{W}$ to one set of parameters	84
3.4.4 The regression weights $\mathbf{B}$	85
3.4.5 The centroid configuration $\mathbf{Y}$	86
3.4.6 Uniqueness of the DIMIDIO solution	87
3.4.7 The algorithm	88
3.5 Analysis of variation in the dimension weighting models	90
3.5.1 Introduction	90
3.5.2 Partitioning of the sums of squares in DIMFREE	90
3.5.3 Partitioning of the sums of squares in DIMIDIO	94
3.6 Illustrations	96
3.6.1 DIMFREE analysis of a constructed data set	96
3.6.2 DIMIDIO analysis of a constructed data set	104

<b>4</b>	<b>The stimulus weighting models</b>	<b>111</b>
4.1	Introduction	112
4.2	Geometry and algebra	113
4.3	The STIMFREE model	118
4.3.1	Introduction	118
4.3.2	Translations	118
4.3.3	Orthonormal transformations	119
4.3.4	Stimulus weights	120
4.3.5	The centroid configuration $\mathbf{Y}$	121
4.3.6	The direct approach in STIMFREE	122
4.3.7	Uniqueness of the STIMFREE solution	123
4.3.8	The algorithm	126
4.4	The STIMIDIO model	128
4.4.1	Introduction	128
4.4.2	Translations of the $\mathbf{X}_j$	128
4.4.3	Orthonormal transformations	129
4.4.4	Stimulus weights	130
4.4.5	Idiosyncratic translations of $\mathbf{Y}$	131
4.4.6	The centroid configuration $\mathbf{Y}$	131
4.4.7	Non-uniqueness of the STIMIDIO solution	132
4.4.8	The algorithm	133
4.5	Analysis of variation in the stimulus weighting models	135
4.5.1	Introduction	135
4.5.2	Partitioning of the sums of squares in STIMFREE	135
4.5.3	Partitioning of the sums of squares in STIMIDIO	139
4.6	Illustrations	140
4.6.1	STIMFREE analysis of a constructed data set	140
4.6.2	STIMIDIO analysis of a constructed data set	151

<b>5 Conclusions</b>	157
5.1 Introduction	158
5.2 The MATCHALS algorithms	158
5.3 Uniqueness properties of the MATCHALS models	159
5.4 Analyses of variation	160
5.5 Statistical aspects	161
5.6 Confirmatory versus exploratory analysis	164
<b>Appendix</b>	165
<b>References</b>	167
<b>Samenvatting</b> (Summary in Dutch)	169