Springer Series in Statistics

Ingwer Borg
Patrick J.F. Groenen

Modern Multidimensional Scaling

Theory and Applications

Second Edition

With 176 Illustrations

Springer

Ingwer Borg
ZUMA
P.O. Box 122155
D-68072 Mannheim
Germany
borg@zuma-mannheim.de

Patrick J.F. Groenen
Erasmus University
Econometric Institute
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
groenen@few.eur.nl

Library of Congress Control Number: 2005924955

ISBN-10: 0-387-25150-2
ISBN-13: 978-0387-25150-9

Printed on acid-free paper.

Printed in the United States of America. (MVY)

9 8 7 6 5 4 3 2 1

springeronline.com

To Leslie,
Handan,
Sezen, and
Martti.

Preface

Multidimensional scaling (MDS) is a technique for the analysis of similarity or dissimilarity data on a set of objects. Such data may be intercorrelations of test items, ratings of similarity on political candidates, or trade indices for a set of countries. MDS attempts to model such data as distances among points in a geometric space. The main reason for doing this is that one wants a graphical display of the structure of the data, one that is much easier to understand than an array of numbers and, moreover, one that displays the essential information in the data, smoothing out noise.

There are numerous varieties of MDS. Some facets for distinguishing among them are the particular type of geometry into which one wants to map the data, the mapping function, the algorithms used to find an optimal data representation, the treatment of statistical error in the models, or the possibility to represent not just one but several similarity matrices at the same time. Other facets relate to the different purposes for which MDS has been used, to various ways of looking at or "interpreting" an MDS representation, or to differences in the data required for the particular models.

In this book, we give a fairly comprehensive presentation of MDS. For the reader with applied interests only, the first six chapters of Part I should be sufficient. They explain the basic notions of ordinary MDS, with an emphasis on how MDS can be helpful in answering substantive questions. Later parts deal with various special models in a more mathematical way and with particular issues that are important in particular applications of MDS. Finally, the appendix on major MDS computer programs helps the reader to choose a program and to run a job.

Contents of the Chapters

The book contains twenty-four chapters, divided into five parts. In Part I, we have six chapters:

- Chapter 1 is an introduction to MDS that explains the four purposes of MDS: MDS as a technique for data explorations, MDS as a method for testing structural hypotheses, MDS as a methodology for the discovery of psychological dimensions hidden in the data, and, finally, MDS as a model of mental arithmetic that explains how similarity judgments are generated. Depending on the particular field of interest, researchers have typically concentrated on just one of these purposes.

- Chapter 2 shows how MDS solutions can be constructed—in simple cases—by purely geometric means, that is, with ruler and compass. Although, in practice, one would almost always use a computer program for finding an MDS solution, this purely geometric approach makes some of the fundamental notions of MDS much clearer than to immediately look at everything in terms of algebraic formulas and computations. It shows, moreover, that the geometric model comes first, and coordinate systems, coordinates, and formulas come later.

- Chapter 3 introduces coordinates and distinguishes different MDS models by the particular functions one chooses for mapping data into distances. Relating data to distances in a particular way also leads to the question of measuring misfit. The Stress index is introduced. An extensive discussion follows on how to evaluate this index in practice.

- Chapter 4 discusses three real-life applications of MDS. The examples are fairly complex but do not require much substantive background. They serve to show the reader some of the trade-off decisions that have to be made when dealing with real data and also some of the most important ways of interpreting an MDS solution.

- Chapter 5 deals with a particular class of MDS applications where the emphasis lies on establishing or testing correspondences of regions in MDS space to classifications of the represented objects in terms of some content facets. It is asked whether objects classified as belonging to type $X, Y, Z, \ldots$ can be discriminated in MDS space such that they lie in different regions. A variety of regional patterns that often arise in practice is discussed and illustrated.

- Chapter 6 describes how to collect similarity or dissimilarity data. Four approaches are distinguished: direct similarity judgments and how to possibly reduce the labor to collect them; deriving similarity measures from the usual cases-by-variables data; converting nonsimilarity measures into similarity measures; and some similarity measures defined for co-occurrence data.

Part II discusses technical aspects of MDS:

- Chapter 7 builds some matrix algebra background for later chapters. Eigendecompositions and singular value decompositions, in particular, are essential tools for solving many of the technical problems in MDS. These tools are put to work immediately for constructing a coordinate matrix from a distance matrix, and for principal axes rotations.

- Chapter 8 concentrates on algorithms for optimally solving MDS problems. To that end, basic notions of differentiation of functions and, in particular, of matrix traces are introduced. Then, the majorization method for minimizing a function is explained and applied to solve the MDS problem. This algorithm, known as the Smacof algorithm, is presented in detail.

- Chapter 9 generalizes the approach of Chapter 8 by allowing for transformations of the dissimilarity data. First, ordinal transformations are discussed, both by monotone regression and rank-images. Then, monotone spline and power transformations are considered in some detail.

- Chapter 10 focuses on confirmatory MDS, where external constraints are enforced onto the MDS solution. These constraints typically are derived from a substantive theory about the data, and it is then tested to what extent this theory is compatible with the data. Two types of constraints are discussed: those imposed on the coordinates and those on the distances of the MDS solution.

- Chapter 11 considers some varieties of indices that assess the goodness of an MDS representation (such as different forms of Stress and the alienation coefficient) and shows some of their relations. Also, we discuss using weights on the dissimilarities and show their effects on MDS solutions.

- Chapter 12 is devoted to one of the first models used for MDS, Classical Scaling. This form of MDS attempts to transform given dissimilarity data into scalar products for which an optimal Euclidean distance representation can be found algebraically without an iterative algorithm.

- Chapter 13 discusses some technical problems that may occur in MDS applications. MDS solutions may degenerate, that is, they become almost perfect in terms of the fit criterion but, nevertheless, do not represent the data in the desired sense. Another important problem is how to avoid local minimum solutions in iterative procedures. Various conditions and solutions for both problems are presented and discussed.

Part III is devoted to unfolding:

- Chapter 14 is concerned with unfolding, a special case of MDS. In unfolding, one usually has preference data from different individuals for a set of objects. Such data are represented by distances between two sets of points that represent individuals and objects, respectively. The model is psychologically interesting but poses a number of difficult technical problems when transformations are allowed on the data.

- Chapter 15 describes a variety of approaches designed to overcome the problem of degenerate solutions in unfolding. We discuss how to replace missing data with reasonable values, how to make the transformation that maps the data into the distances of the model more rigid, and how to properly adjust the loss function to avoid degeneracies.

- Chapter 16 introduces a number of special models for unfolding such as external unfolding, the vector model of unfolding, individual-differences unfolding with weighted dimensions and anti-ideal points, and a metric unfolding model that builds on scale values constructed within a particular (BTL) choice theory.

Part IV treats the geometry of MDS as a substantive model:

- Chapter 17 concentrates on one particular tradition of MDS where the MDS space is equated with the notion of a "psychological" space. Here, the formula by which we compute distances from point coordinates is taken as a model of the mental arithmetic that generates judgments of dissimilarity. Some varieties of such models (in particular, the Minkowski distance family) and their implications are investigated in some detail.

- Chapter 18 studies a particular function on pairs of multi-valued objects or vectors, scalar products. Scalar products have attractive properties. For example, one can easily find an MDS space that explains them. Hence, various attempts were made in the psychological literature to generate similarity judgments that can be directly interpreted as scalar products (rather than distance-like values).

- Chapter 19 concentrates on the most important distance function in practice, the Euclidean distance. It is asked what properties must hold for dissimilarities so that they can be interpreted as distances or even as Euclidean distances. We also discuss what transformations map such dissimilarities into Euclidean distances. A further question is how to find a linear transformation that leads to approximate Euclidean distances in a small dimensionality.

Part V discusses some techniques and models that are closely associated with MDS:

- Chapter 20 treats Procrustean problems. Given one particular configuration or target, $\mathbf{X}$, it is asked how one can fit another configuration, $\mathbf{Y}$, to it without destroying meaningful properties of $\mathbf{Y}$. Procrustean solutions are important in practice because they serve to eliminate irrelevant—and often misleading—differences between different MDS solutions.

- Chapter 21 looks at generalized Procrustes analysis, where one wants to fit several configurations to a target or to each other. We also consider extensions where further fitting parameters are admitted that do not preserve the configurations' shapes but that have some meaning in terms of individual differences (e.g., different dimensional weights).

- Chapter 22 focuses on the question of how we can scale a set of K dissimilarity matrices into only one MDS solution and explain the differences among the K data sets by different weights on the dimensions of the "group space" of all K data sets. One algorithm for solving this problem, INDSCAL, is considered in some detail. Some algebraic properties of such models are also investigated.

- Chapter 23 concentrates on asymmetric proximities. They require special considerations or models. We show that asymmetric data can always be decomposed in a symmetric and a skew-symmetric part. Some models for visualizing asymmetry only study the skew-symmetric part and others try to represent both parts at the same time. We discuss several models such as Gower's decomposition for skew-symmetry, a model that represents the skew-symmetries as force vectors in an MDS solution of the the symmetries, unfolding, the slide-vector model, a hill-climbing model, and the radius-distance model.

- Chapter 24 focuses on two methods that are closely related to MDS: principal component analysis and correspondence analysis. We present their formal properties, show some applications to empirical data sets, and discuss how they are related to MDS.

In the Appendix, we cover two issues:

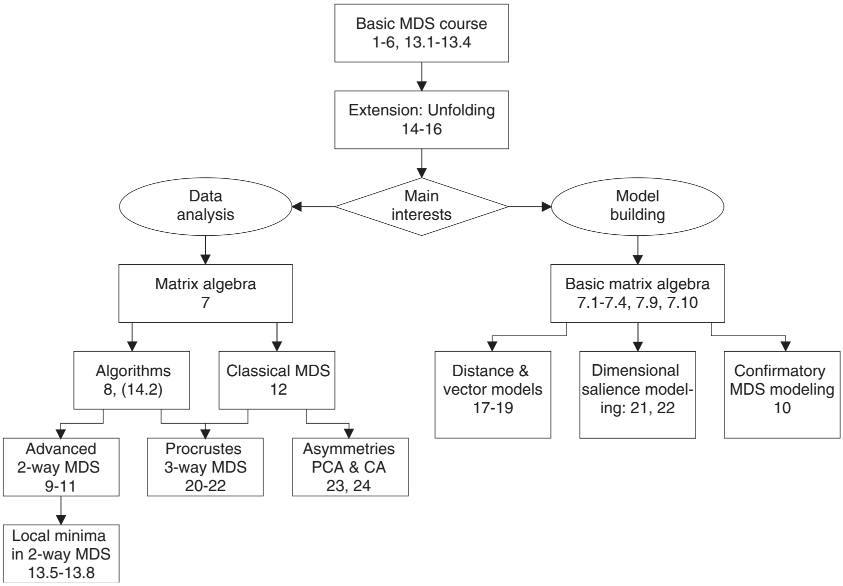

FIGURE 1. Some suggestions for reading this book.

- Appendix A describes in some detail the major computer programs available today for doing MDS. The programs selected are Ggvis, Permap, the MDS modules in SAS, SPSS (Proxscal and Alscal), Statistica, and Systat, as well as the standalone programs Newmdsx©, Fssa, and the classics Kyst, Minissa, and Multiscale.

- Appendix B contains a summary of the notation used throughout this book.

How to Read This Book

Beginners in MDS should first study Chapters 1 through 6. These chapters make up a complete introductory course into MDS that assumes only elementary knowledge of descriptive statistics. This course should be supplemented by reading Sections 13.1–13.4 because they cover, in the same nontechnical way, two technical problems (degenerate solutions, local minima) of which every MDS user should be aware.

The basic course can be extended by adding Chapters 14 to 16, if technical sections are skipped. These chapters add the idea of unfolding and discuss some variants of this model.

After mastering the fundamentals, the reader may either read on sequentially or first consider his or her primary interests. If these interests are primarily in the psychology of similarity and choice, then the reader should move to the chapters on the right-hand side in Figure 1. That is, after reviewing some basic matrix algebra, the reader should move on to one of the topics of particular substantive interest. The most natural place to start is Chapter 17, which focuses directly on different attempts to model similarity by distance functions; to Chapter 18 which is concerned with how to assess scalar products empirically; and to Chapter 19 which studies some of the basic issues involved in modeling proximities in geometric models. Then, the essential ideas of of Chapters 21 and 22 are interesting candidates for further study. Also, the substantively relevant material in Chapter 10 should be of particular interest.

A student whose primary interest is data analysis should first study the matrix algebra in Chapter 7 in somewhat more detail to prepare for Chapter 12 (classical MDS). From Chapter 12, one can proceed to Chapter 23 (asymmetric models) and Chapter 24 (PCA and correspondence analysis) or to Chapters 20–22 (Procrustean methods, three-way models, individual-differences models). A different or additional route in Figure 1 is to turn to Chapter 8 (algorithms) after having studied Chapter 7. The discussion of how to find optimal transformations of the proximities (as in ordinal MDS) in Chapter 9 can be read, to a large extent, without knowledge of Chapter 8. Knowing how to solve MDS problems numerically is, however, a prerequisite for studying a number of advanced issues in Chapter 10 (confirmatory MDS and how to do it) and Chapter 11 (fit measures). From Chapter 11, one should proceed to the technical sections of Chapter 13, which discuss local minima problems.

History of the Book

One could say that the present book is the third edition of a book on multidimensional scaling. The book appeared in German in 1981 under the name *Anwendungsorientierte Multidimensionale Skalierung* by Ingwer Borg (Heidelberg, Germany: Springer). This book served as a basis for an English version. It was called, somewhat cryptically, *Multidimensional Similarity Structure Analysis*. Authored by Ingwer Borg and the late Jim Lingoes, it appeared in 1987 (New York: Springer). As the copies of this book sold out, a revised reprint was considered to bring the book up to date, but then this revision led to a complete overhaul and substantial additions, in particular on the algorithmic side. We have changed the order of presentation, excluded or shortened some material, and included recent developments in the area of MDS. To reflect these changes, we have added "Modern" to the book's title. We also replaced the term "Similarity Structure Analysis" by the better-known term "Multidimensional Scaling". Proponents of SSA may feel that this is an unfortunate regression in terminology, but the term MDS is simply much better known in general. In any case, the shift from SSA to MDS does not imply a change of perspective. We still consider all aspects of MDS representations as potentially interesting, not just "dimensions." The present book is the second revised edition of *Modern Multidimensional Scaling*.

Preface to the Second edition

The second edition of *Modern Multidimensional Scaling* differs from the first edition on several aspects. The changes have increased the number of pages from 471 to 611 pages and the number of figures from 116 to 176. Two new chapters were added to the book. The first new chapter is devoted to the problem of how to avoid degeneracies in unfolding. New developments in this area are covered and several solutions are presented. One of these solutions, the PrefScal program, is scheduled to become available soon in SPSS.

The other new chapter is an expansion of a section on asymmetric models into a full chapter. There, we discuss several models for visualizing asymmetry and skew-symmetry in MDS. Some of these models are new and others are known in the literature.

In addition, we have updated, extended, and added several sections in existing chapters. Some of these additions reflect new insights from the literature; others are aimed at clarifying existing material. The appendix on MDS software contains the description of four new MDS programs.

Also, exercises have been added to each chapter. They should help the reader to better learn MDS by, first of all, actually doing MDS on empirical data sets, or by rethinking the various issues within a particular scientific context. The exercises differ, of course, with respect to their level. Some emphasize more practical skills such as actually using one or another MDS computer program; others are more demanding and have no simple right-or-wrong answers. These exercises make the book easier to use in a course on MDS. All data in the book are available on the Internet at

http://www.springeronline.com/0-387-25150-2.

Acknowledgment

There are several people we would like to thank for their comments and suggestions on this text. Their inputs certainly have been beneficial for the quality of this book. In particular, we would like to thank Frank Busing, Katrijn van Deun, Luc Delbeke, and Akinori Okada for their constructive feedback on parts of the manuscript. We are also grateful to Joost van Rosmalen for his careful reading and his remarks on the entire manuscript.

Ingwer Borg, Patrick J.F. Groenen, May, 2005, Mannheim and Rotterdam

Contents