HUDM5124: Introduction to Multidimensional Scaling, Clustering, and Related Methods

Session 7:

- Spatial unfolding models for preference (2-way 2-mode) data

Dominance and Preference Data

Example of dominance data
(item scores for each subject):

Subject	A	B	C	D	E
S01	1	1	1	1	0
S02	1	1	0	0	0
S03	1	0	0	0	0
Etc.					

Typical model: Item Response Theory

Example of preference data
(rank ordering from each subject):

Subject	A	B	C	D	E
S01	2	1	3	4	5
S02	5	4	3	2	1
S03	5	3	1	2	4
Etc.					

Typical model: Unfolding
→ preference is thought of as a type of proximity relation

The unidimensional unfolding model

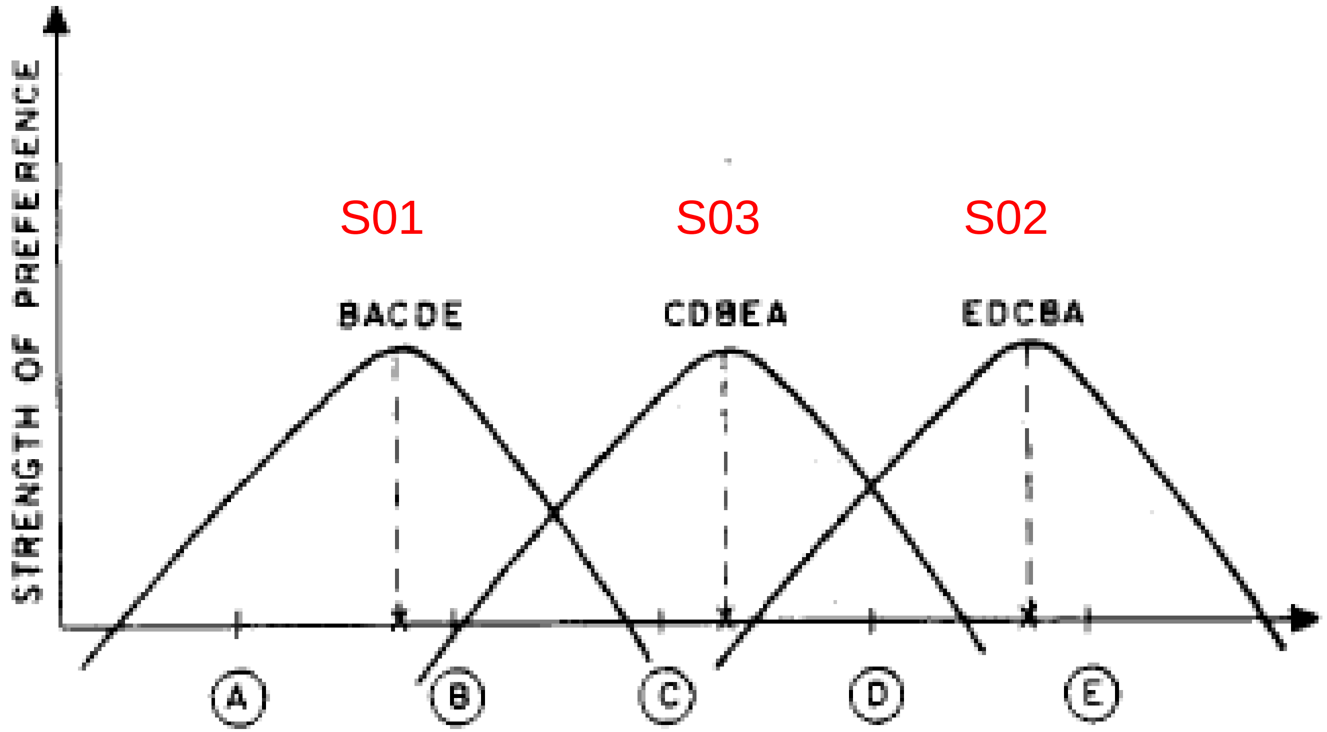


Fig. 3. Unidimensional unfolding model illustrated. Distance of stimulus from subject's «ideal point» (his optimal value on the unidimensional scale) is assumed to define (inversely) the preference scale for that subject.

Carroll (1980) – techniques for <u>multidimensional</u> analysis of preference data

- Typical data = preference ratings (rankings) of m stimuli by n individuals
- Goal: explain the individuals' ratings in terms of multiple latent dimensions

Spatial unfolding: the problem

GIVEN: a set of proximities between two sets of conceptual "objects" (usually n subjects x m stimuli). This gives rise to a rectangular N x M matrix of proximities, that are assumed to be linearly (or merely monotonically) related to distances in a psychological space.

ESTIMATE:

1) The locations of n "subject" points and m "stimulus" points in a joint space (of dimensionality R). These locations are summarized by an (n + m) x R matrix, X

Of course, the "correct" dimensionality R must be determined as well.

GOAL: a "joint space" that plots the locations of both stimulus points (e.g. breakfast foods) and subject points (respondents)

Data for unfolding (e.g., preference rankings) relates stimuli to subjects (sources):

- Preferences etc. can be thought of as proximities BETWEEN two sets of points
- Thus it can be thought of as an "off-diagonal" segment of a "super" proximity matrix of size (n+m) x (n+m), with rows (cols) corresponding to stimuli AND subjects

This means we can use regular MDS algorithms with missing data capabilities to fit the unfolding model (theoretically at least)

S =

	a	b	c	d	e	f	s1	s2	s3	s4
a	--									
b		--								
c			--							
d				--						
e					--					
f						--				
s1	1	4	2	5	6	3	--			
s2	6	3	4	1	2	5		--		
s3	3	5	6	1	2	4			--	
s4	2	3	1	4	6	5				--

Two basic spatial (geometric) models for unfolding:

1) Vector model:
   stimuli = points in R-dim space
   subjects (sources) = vectors (directions)

1) Ideal point model:
   stimuli = points in R-dim space
   subjects (sources) = points in R-dim space

The vector model for preferences

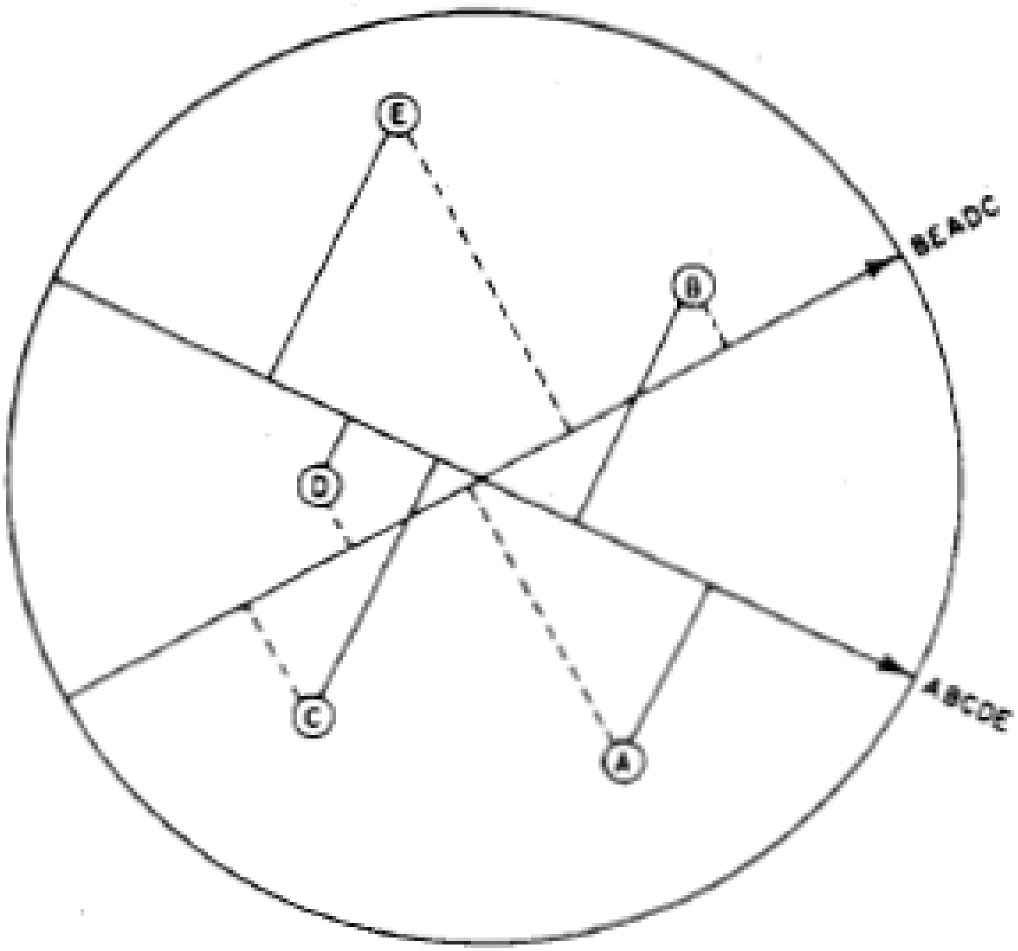


Fig. 1. Vector model for preference illustrated. Projections of stimulus points onto a subject's vector are assumed to define preference scale for that subject.

Example of vector model – Wish nations data

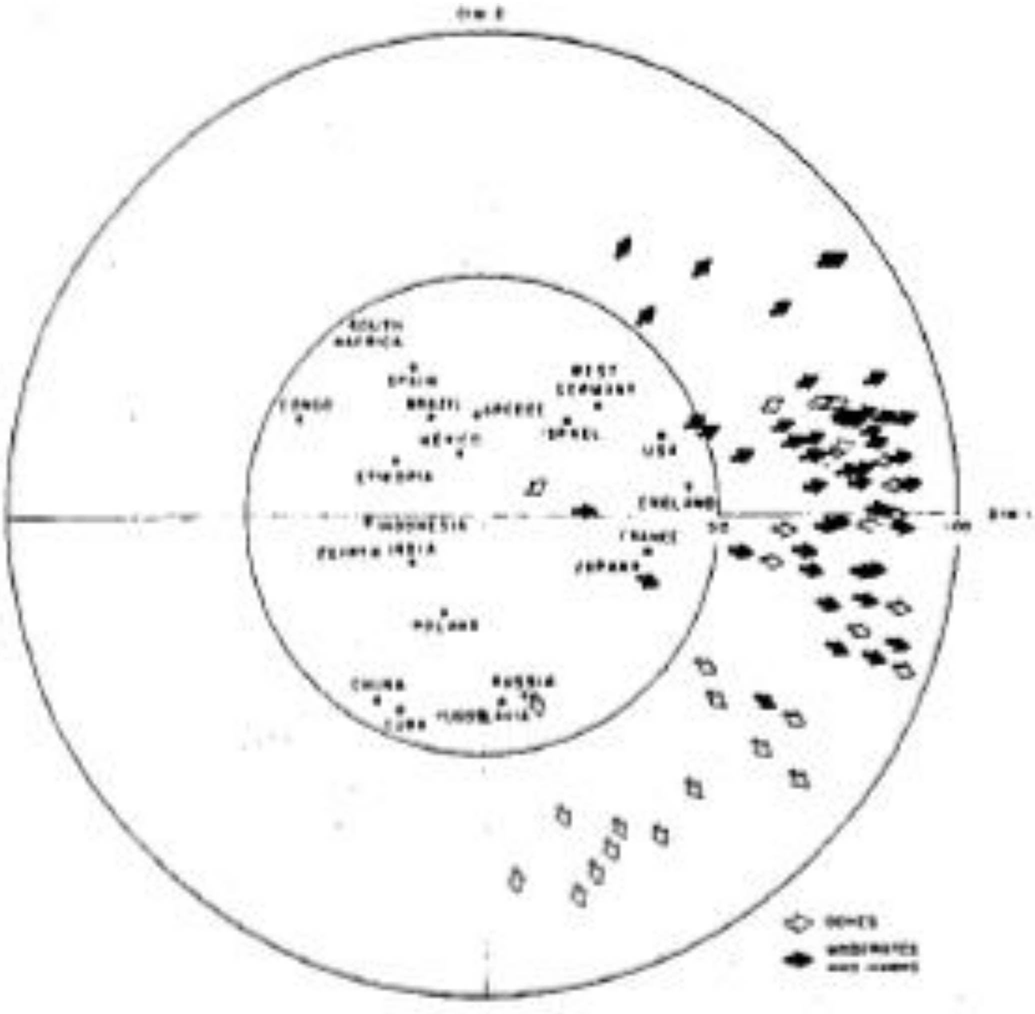


Fig. 2. Two-dimensional solution from Wish's MDPREF analysis of «Similarity to Ideal» judgments of 75 subjects on 21 nations.

The unidimensional unfolding model

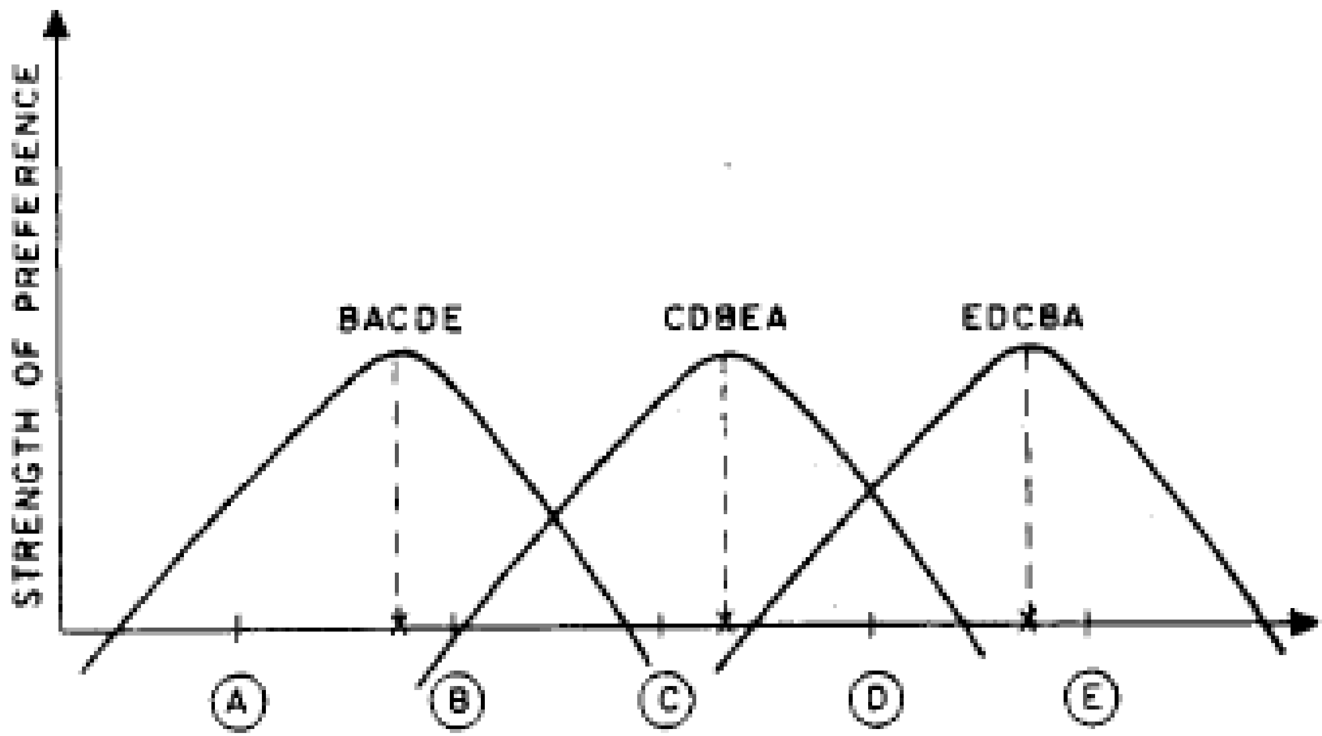


Fig.3. Unidimensional unfolding model illustrated. Distance of stimulus from subject's «ideal point» (his optimal value on the unidimensional scale) is assumed to define (inversely) the preference scale for that subject.

Multidimensional unfolding – the ideal point model

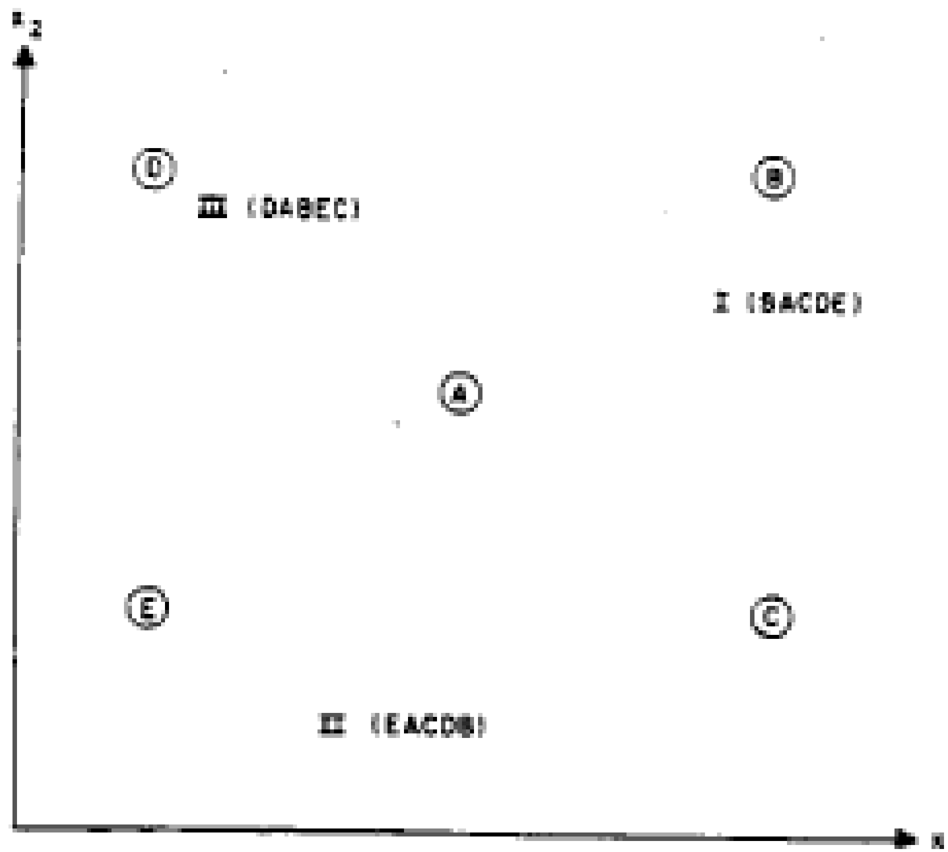


Fig. 4. Multidimensional unfolding model illustrated. Distance from subject's «ideal point» (inversely) defines his preference scale. Isopreference contours are concentric circles centered at «ideal points». They would be spheres or hyperspheres in higher dimensions.

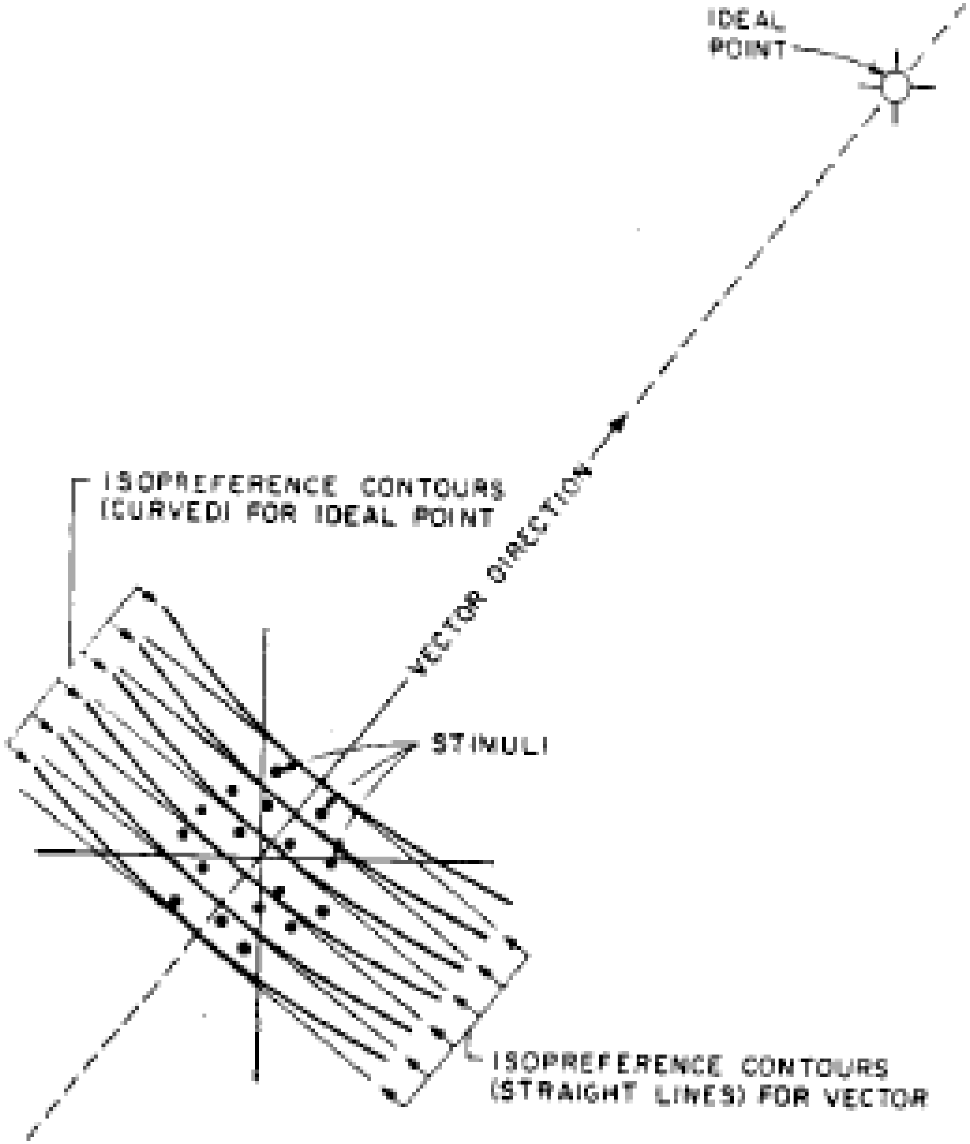


Fig.5. Isopreference contours (straight lines) for vector and (circles) for ideal point a large distance from centroid of stimuli in direction of vector. As ideal point moves farther and farther out in this same direction, circular contours approach (in the vicinity of the fixed stimuli) the straight line contours for the vector.

The vector model is a special case of the ideal-point model (obtained when ideal points move out to an infinite distance from the origin)

“internal” vs. “external” unfolding

- “Internal” unfolding analysis: both stimulus and subject points fit by one set of proximity data
- “External” unfolding analysis: stimulus configuration fit by “external” data (e.g. sim. ratings among stimuli)

Data for unfolding (e.g., preference rankings) relates stimuli to subjects (sources):

- Preferences etc. can be thought of as proximities BETWEEN two sets of points
- Thus it can be thought of as an “off-diagonal” segment of a “super” proximity matrix of size (n+m) x (n+m), with rows (cols) corresponding to stimuli AND subjects

This means we can use regular MDS algorithms with missing data capabilities to fit the unfolding model (theoretically at least)

S =

	a	b	c	d	e	f	s1	s2	s3	s4
a	--									
b		--								
c			--							
d				--						
e					--					
f						--				
s1	1	4	2	5	6	3	--			
s2	6	3	4	1	2	5		--		
s3	3	5	6	1	2	4			--	
s4	2	3	1	4	6	5				--

Commonly used computer software for fitting the spatial unfolding model:

program	Author	source	type
KYST	Carroll & Pruzansky	NETLIB	internal
ALSCAL	Young	SPSS, SAS	internal
PREFSCAL	Leiden group	SPSS	internal
MDPREF	Carroll, Pruzansky	NETLIB	(+dominance)
PREFMAP	Carroll, Meulman, Heiser	NETLIB	external
smacof	Mair, de Leeuw, Groenen	R	internal

Unfolding in SPSS: Prefscal

```
TITLE    RUN PREFSCAL unfolding on Green & Rao (1972) breakfast data.
DATA LIST FREE
 /1 subj  TP BT EM JD CT BM HR TM BJ TG CB DP GD CC CM.
BEGIN DATA
1 13 12 7 3 5 4 8 11 10 15 2 1 6 9 14
2 15 11 6 3 10 5 14 8 9 12 7 1 4 2 13
3 15 10 12 14 3 2 9 8 7 11 1 6 4 5 13
4 6 14 11 3 7 8 12 10 9 15 4 1 2 5 13
PREFSCAL VARIABLES=TP BT EM JD CT BM HR TM BJ TG CB DP GD CC CM
/INITIAL=CLASSICAL(TRIANGLE)
/CONDITION=ROW
/TRANSFORMATION=LINEAR
/PROXIMITIES=DISSIMILARITIES
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001) MAXITER(5000)
/PENALTY=LAMBDA(0.5) OMEGA(1.0)
/PRINT=HISTORY MEASURES COMMON
/PLOT=COMMON.
```

Unfolding in R

(ref: Mair, de Leeuw & Lienbacher, 2009)

Rectangular SMACOF (Unfolding)

Rectangular $n_1 \times n_2$ preference matrix Δ.

- Stress becomes

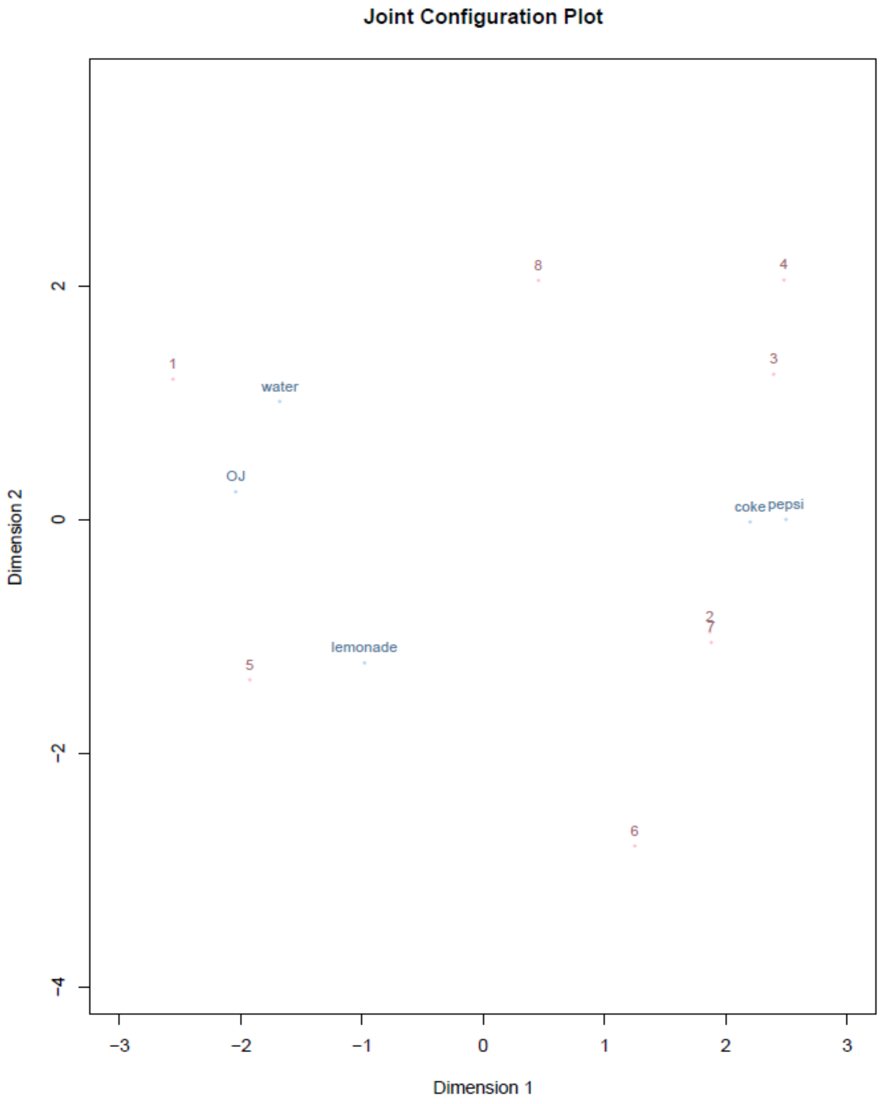
$$\sigma(X) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij}(\delta_{ij} - d_{ij}(X_1, X_2))^2 \rightarrow \text{min!}$$

- Judge $n_1 \times p$ configuration matrix
- Object $n_2 \times p$ configuration matrix

An example:
rated preferences for 5 types of beverage by 8 participants

subj	**coke**	**pepsi**	**lemonade**	**water**	**OJ**
1	5	5	3	1	1
2	1	1	3	5	3
3	1	1	5	3	5
4	2	2	5	4	5
5	4	5	1	2	2
6	3	3	3	5	4
7	1	1	2	4	5
8	3	3	3	3	3

```
# unfolding example
> beverage <- read.delim("C:/Users/user1/ex_rect_data.txt", sep=" ",
                header=TRUE)
> beverage
> dim(beverage)
> bev <- smacofRect(beverage[,2:6])
> bev
> plot(bev, xlim = c(-3, 3), joint = TRUE, asp = 1)
```



- Call: smacofRect(delta = beverage[, 2:6])
- Model: Rectangular smacof
- Number of subjects: 8
- Number of objects: 5
- Stress-1 value: 0.133
- Number of iterations: 273

Some practical issues:

1. **Choosing the dimensionality**. The problem of determining the dimensionality R of the solution space must be faced. Often the dimensionality is selected *a priori* based on theory or *post-hoc* based on interpretability or on perception of an "elbow" in fit function (RSQ for SINDSCAL; normalized raw stress for PREFSCAL).

2. **Degenerate solutions**. We are fitting *(n+m)R* parameters from *nm* data points. Depending on the values of *n,m,R*, this may be problematic and lead to poorly determined or indeterminate solution, leading to degenerate solution configurations. Degeneracy may show up as clumps of points, or all points of one type (e.g. subjects) on one side of space, or points distributed in a ring, etc.

POSSIBLE FIXES:

-use a metric instead of nonmetric analysis

-specify matrix-conditional rather than row-conditional (?)

-try to ensure diversity of individual preferences

-use rational start (?)

-make sure to use Stress formula 2, not 1 (if using KYST)

-use PREFSCAL (see Busing et al., 2005)

Avoiding trivial /degenerate solutions in unfolding

(ref: Modern MDS)

TABLE 15.3. Comparison of approaches aimed at avoiding trivial solutions.

	Un-conditional	Row-conditional	Transformation	Trivial Solution Excluded	Quality
Adjusting Data					
Ratio-ordinal	+	+	Ordinal	Yes	+
Interval-ordinal	+	+	Ordinal	No	+/−
Augmenting within-persons block	+	−	All for between-sets ratio for within-sets	No	+/−
Augmenting both within-sets	+	+	All for between-sets ratio for within-sets	Yes	+
Adjusting the Transformation					
Ratio transformation	+	+	Ratio	Yes	+
Approach of Kim et al. (1999)	+	+	Ratio	Yes	+
Smoothed monotone regression	+	+	Restricted ordinal	Yes	+
Adjusting the Loss Function					
Stress-2	+	+	All	No	−
Weighting approach by DeSarbo	+	+	All	No	−
Penalizing the intercept	+	+	Interval	Yes	+
Penalized Stress by PREFSCAL	+	+	All	Yes	+

The weighted unfolding model

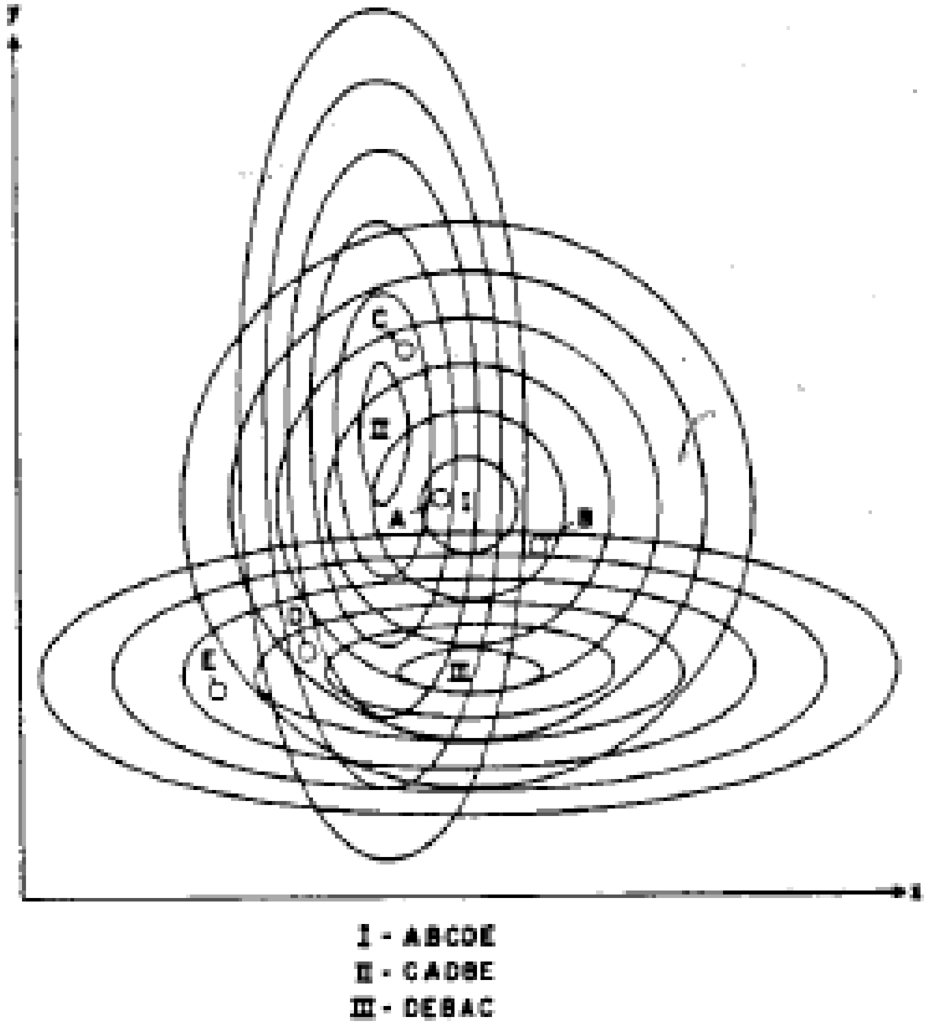


Fig. 6. Illustration of multidimensional unfolding model with differential weights. Subject I weights x_1 and x_2 equally, II weights x_1 more than x_2, while III weights x_2 more than x_1. Iso-preference contours are ellipses with axes parallel to coordinate axes, and lengths of axes proportional to reciprocal of square root of weights. In higher dimensions these would be ellipsoids or hyperellipsoids. Generalized Euclidean distance from «ideal point» defines preference. Order implied for the three hypothetical subjects is shown at bottom.

In the weighted unfolding model, negative weights are possible

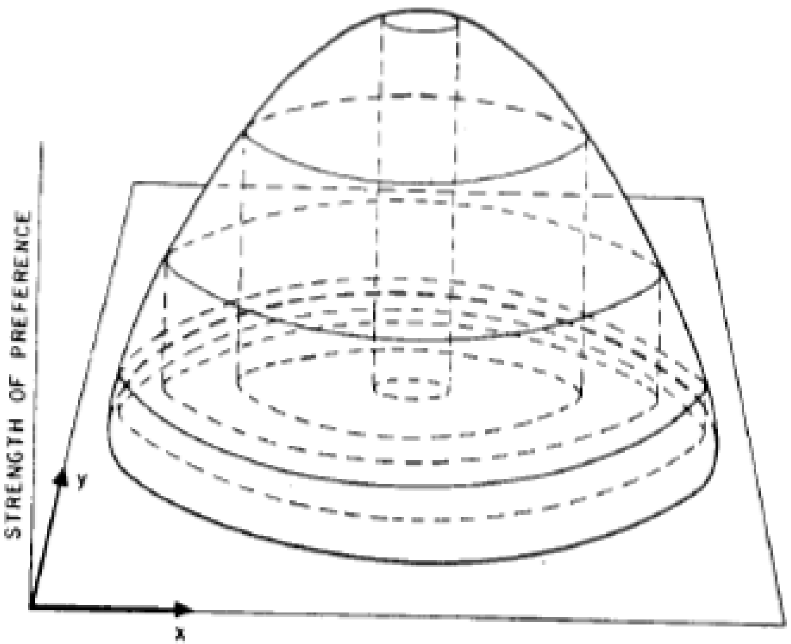


Fig. 8. Typical preference function for two-dimensional unfolding model with both weights positive. Horizontal slices projected into x_1–x_2 plane define elliptical isopreference contours.

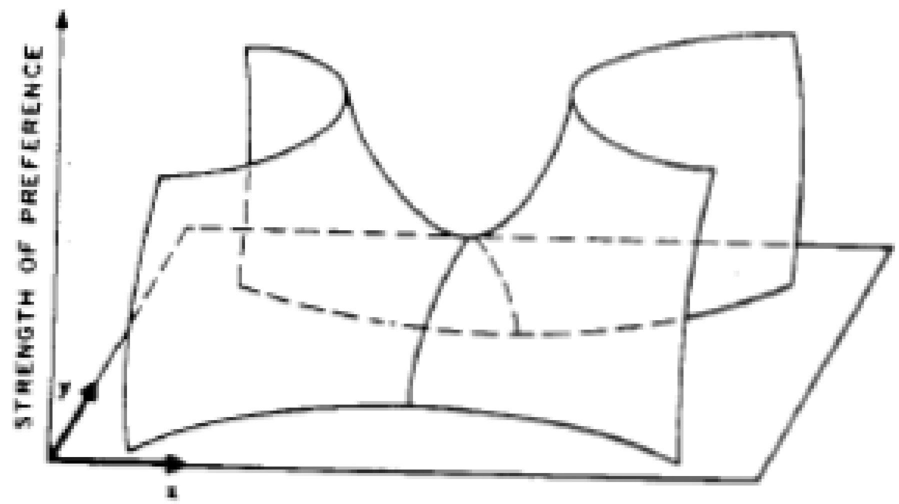


Fig. 9. Saddle-shaped preference function when one dimension (x_1) has negative and other (x_2) has positive weight.

Application of weighted unfolding model: tea-tasting data

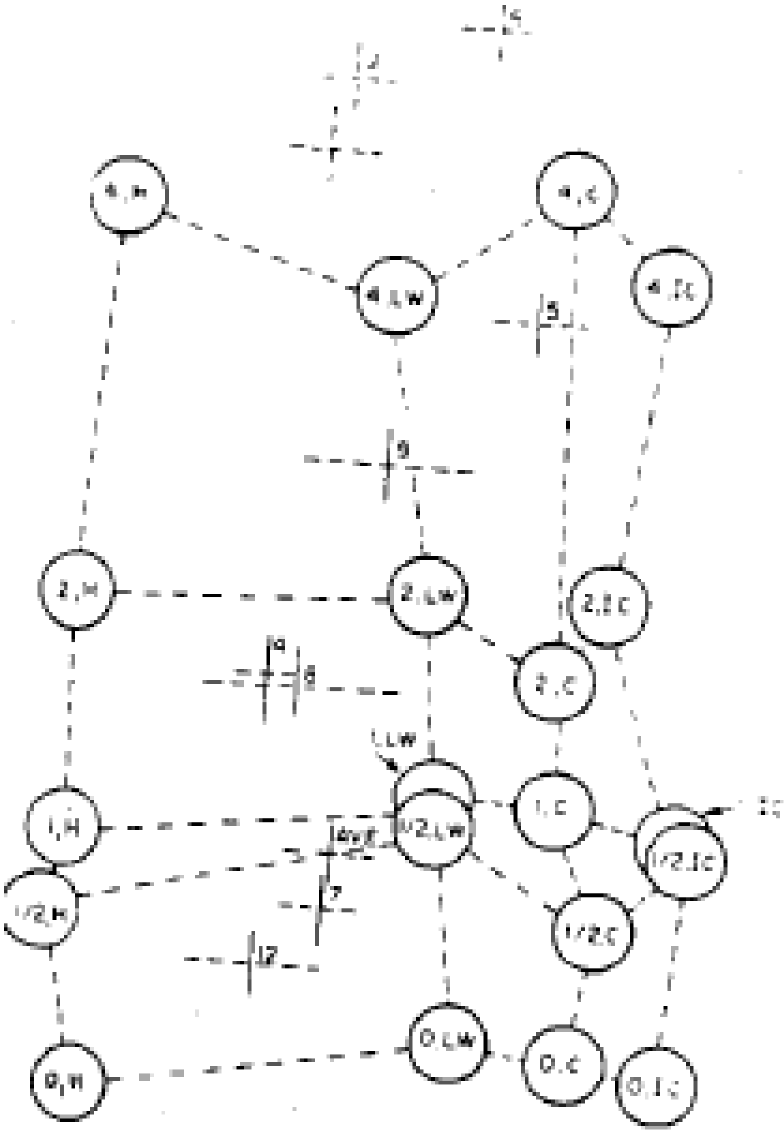


Fig. 11. Partial result of external analysis of «tea tasting» data in terms of weighted unfolding model. Stimulus coding: H = Hot; LW = Lukewarm; C = Cold; IC = Ice Cold. Number indicates number of teaspoons of sugar. Five «Steaming Hot» stimuli were left out of analysis. Not all subjects are shown. AVE = «Average Subject». Crossed lines for subjects interpreted as follows: length of line proportional to absolute value of weight for corresponding dimension; solid line = positive weight; dotted line = negative weight.

A generalized weighted unfolding model

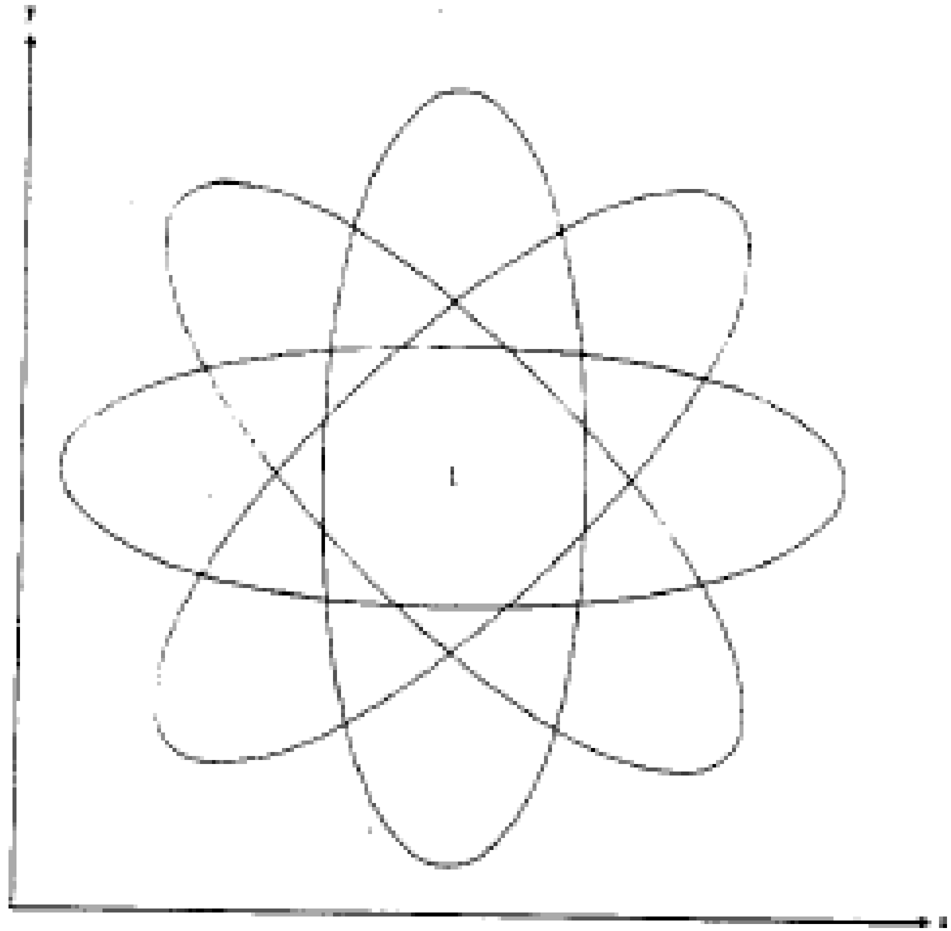


Fig. 7. Illustration of generalized unfolding model, in which differential rotations and differential patterns of weights are allowed. Shape of sample isopreference contours indicated, all (arbitrarily) about same «ideal point» and with same pattern of weights. Unlike previous model, isopreference ellipses need not be parallel to coordinate axes.