

“Classic” Multidimensional Unfolding and the Vector Model for Unfolding

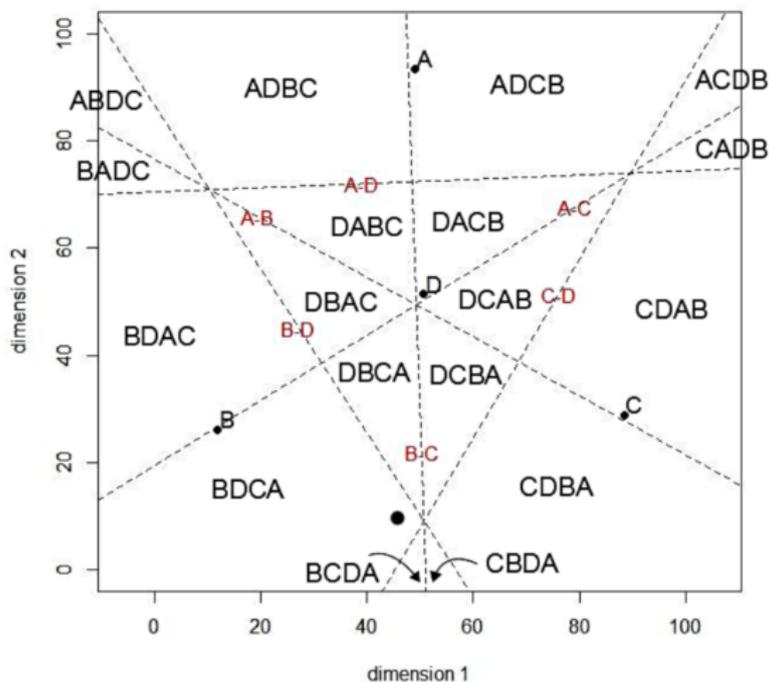
Analyzing Survey Research Data
GSERM, Summer 2019
Prof. Adam M. Enders

Refresher on Unfolding

- Scaling model that assumes “single-peaked” item response functions, rather than monotonic ones, like with the cumulative scaling model
- Represents **proximities** between the n rows and k columns of a **rectangular** data matrix as distances between points along a single continuum
- Objectives:
 - ▶ Represent row and column objects in some low-dimensional space (two dimensions, ideally!)
 - ▶ Proximity between row and each column object should represent, to the best extent possible, the preferences/(dis)similarities from the original data matrix

Refresher on Unfolding

Imagine a napkin!



Moving to the Vector Model of Unfolding...

- Think back to the summated rating model
- In that case, we collapsed across the columns of an n by k data matrix to create a unidimensional measure of some latent construct
- In doing this, we made several assumptions, one of which is that the items are “parallel measures” – they measure the same thing in roughly the same way
- What happens with a similar dataset, but the full profile of the rows across columns matters?
- Example: ranking a set of stimuli according to importance or attractiveness
 - ▶ In this case, we want to preserve the data within rows/across columns
 - ▶ Rankings where each cell value is a rank number are not parallel measures
 - ▶ Summing across the columns would result in scale scores identical for all rows

The Vector Model of Unfolding

- So far, we've represented row objects as points in variable space (everything up to SVD)
- Also represented variables as both points (MDS, cumulative scaling, unfolding) and vectors (factor analysis, biplot)
- The Vector Model represents row objects as vectors emanative from the centroid, and column stimuli as points
 - ▶ Sometimes referred to as MDPREF, for MultiDimensional PReference Scaling
- Such a representation of the data allows for geometric interpretations of the relationships between pairs of row objects and between row objects and column objects
- The column points and row vectors are arranged so that *to the greatest extent possible*, row vectors point toward higher ranked column points, and away from lower ranked column objects

Example

Hypothetical dataset with two individuals (row objects) and three generic stimuli (column objects)

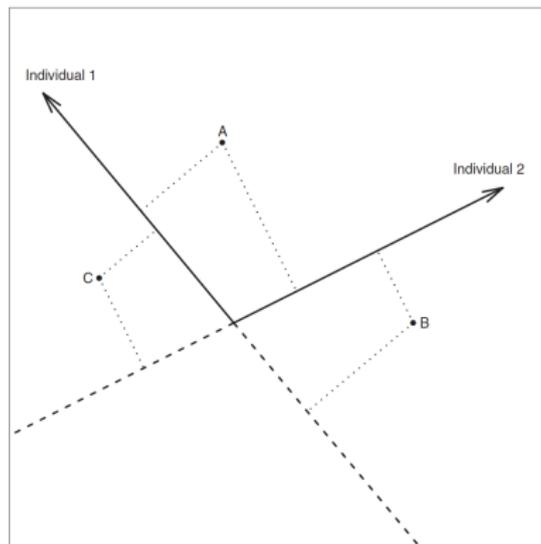
Respondents asked to rank order the stimuli, where larger numerical values correspond to more preferred stimuli

	Stimuli:		
	A	B	C
Individual 1:	3	1	2
Individual 2:	2	3	1

Example, cont'd

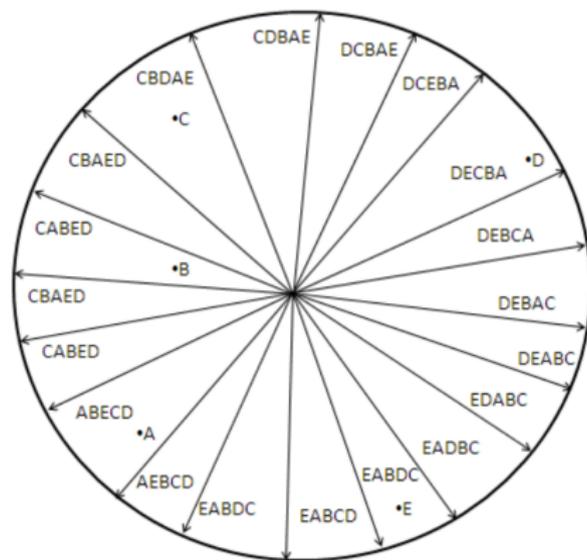
Objective is to locate column objects as points and row objects as vectors in a shared space

Drawing a perpendicular line from the column points to the row vectors should reproduce, as perfectly as possible, the original rank ordering in the dataset



Example, cont'd

Expanded to more column and row objects:



Each vector reproduces the preferences/rank orders from the original data matrix

The Model

- Assume some matrix \mathbf{X} contains n observations' scores for k stimuli
- Let the stimulus set, S_1, S_2, \dots, S_k , be represented by k points in m -dimensional space
- Point coordinates are collected into the $k \times m$ matrix, Φ
- The observations, O_1, O_2, \dots, O_n , are represented by n unit-length vectors in the same m -space
- All of these vectors emanate from the origin, and their terminal points are collected into the $n \times m$ matrix, Γ

The Model, cont'd

- Algebraically, the vector γ_i contains the terminal points of observation i 's vector within the m -space
- The projections from the stimulus points onto i 's vector are obtained by taking the scalar products of the stimulus points and the observation vector:

$$\hat{x}_i = \Phi \gamma_i$$

- We assume the dimensionality, m , should be relatively small relative to the dimensionality of the original data matrix
- If that's true, then we need to introduce an error term:

$$x_i = f(\hat{x}_i) + \epsilon_i = f(\Phi \gamma_i) + \epsilon_i$$

- ϵ_i is an error vector

The Model, cont'd

- If the data are assumed to be interval level:

$$x_i = \alpha + \beta \hat{x}_i + \epsilon_i$$

- If the data are ordinal level:

$$x_i = f(\hat{x}_i) + \epsilon_i$$

- Here, f is a monotonic function

Estimation: Metric Model (Interval/Ratio Data)

- First, we're going to consider how to estimate a metric model, which assumes interval or ratio level data
- Estimation is very easy: basically just a singular value decomposition
- Steps:
 1. Standardize rows of \mathbf{X} to 0 mean and unit variance, \mathbf{X}_{std}
 2. Perform SVD of $\mathbf{X}_{std} = \mathbf{UDV}'$
 - \mathbf{U} in an n by q matrix of left singular vectors
 - \mathbf{D} is the q -order diagonal matrix of singular values
 - \mathbf{V} is the k by q matrix of right singular vectors
 3. Determine appropriate dimensionality, m , by considering model fit in a number of dimensions
 - $R^2 = tr(\mathbf{D}_m^2)/tr(\mathbf{D}^2)$, where tr is the matrix trace, or sum of diagonal elements
 4. After determining m , $\Phi = \mathbf{V}_m$, and $\Gamma = \mathbf{U}_m \mathbf{D}_m$

Side Note: Utility of the SVD

- Weller and Romney (1980) point out that many scaling methodologies involve
 1. A transformation, or pre-processing, of data (e.g., row standardization, double-centering)
 2. A singular value decomposition of the transformed dataset
 3. Some post-SVD transformation of eigenvectors
- Examples:
 - ▶ PCA: row standardize, SVD, row scores = $\mathbf{U}_m \mathbf{D}_m$
 - ▶ MDS: double-center, SVD, stimuli coordinates = $\mathbf{U}_m \mathbf{D}_m$
 - ▶ Vector Model: row standardize, SVD, row coordinates = $\mathbf{U}_m \mathbf{D}_m$

Estimation: Nonmetric Model (Ordinal Data)

- To do a estimate a nonmetric vector model, we just need to iteratively optimally scale the row and column coordinates to be monotonic functions of the original dataset
- Just like we did the first week with ALSOS
- Steps:
 1. Specify m , the dimensionality of the space, initialize R^2 to zero, and initialize \mathbf{X}^* by setting it equal to the original \mathbf{X} matrix. Standardize within rows of \mathbf{X}^* to obtain \mathbf{X}_{std}^*
 2. Perform SVD on \mathbf{X}_{std}^* to obtain \mathbf{U}_m , \mathbf{D}_m , and \mathbf{V}_m . Use these matrices to calculate current estimates of $\hat{\mathbf{X}}$ and R^2
 3. If the current R^2 is larger than the previous value, then continue. If R^2 has not changed from the previous iteration, go to step 6

Estimation: Nonmetric Model (Ordinal Data)

- Steps, cont'd:
 4. For $i = 1, 2, \dots, n$, use Kruskal's monotonic regression (1964) to find a new estimate of \hat{x}_i^* that are maximally correlated with the current model-based predicted values but always weakly monotonic to the original data
 5. Return to step 2 and carry out another iteration of the estimation procedure on the new version of \mathbf{X}^* that was obtained in step 4
 6. When R^2 converges, construct the Φ and Γ matrices from the singular vectors and values, and use the final R^2 as the goodness-of-fit
- Row and column coordinates are calculated exactly as they were before, except the model will likely fit better due to the weaker assumption of (weak) monotonicity

Example: Jacoby 2014

Respondents were confronted with “triads” of the following issues, from which rankings were constructed

TABLE 1. Values and Definitions Shown to Respondents in Post-Election Wave of the 2006 Cooperative Congressional Election Study. In the Actual Administration of the Survey, the Order of the Values was Randomized Across Respondents

FREEDOM, that is the widest liberty possible for everyone to act and think as they consider most appropriate.

EQUALITY, that is making sure that everyone has the same chance to get ahead in life.

ECONOMIC SECURITY, that is making sure that everyone has a steady job, a decent income, and a reasonable standard of living.

MORALITY, that is people living according to the rules that most people agree constitute decent human behavior.

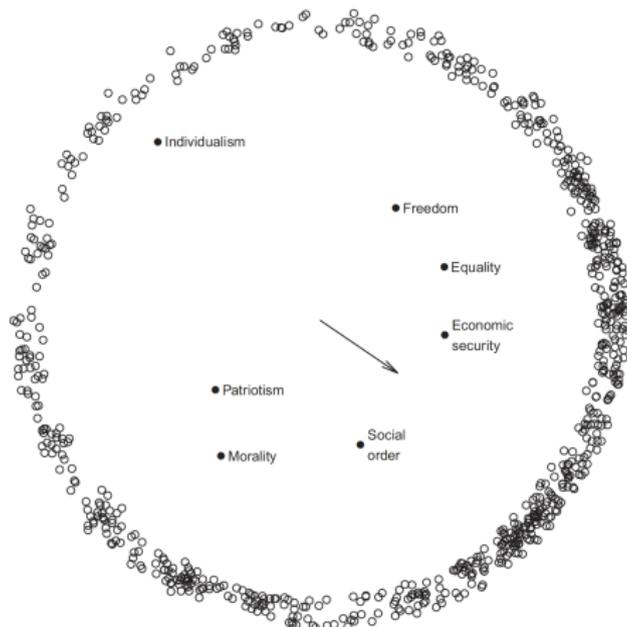
INDIVIDUALISM, that is everyone getting ahead in life on their own, without extra help from government or other groups.

SOCIAL ORDER, that is being able to live without fear, in a safe, peaceful society where the laws are respected and enforced.

PATRIOTISM, that is looking beyond our own personal interests and doing things that honor, respect, and protect our nation as a whole.

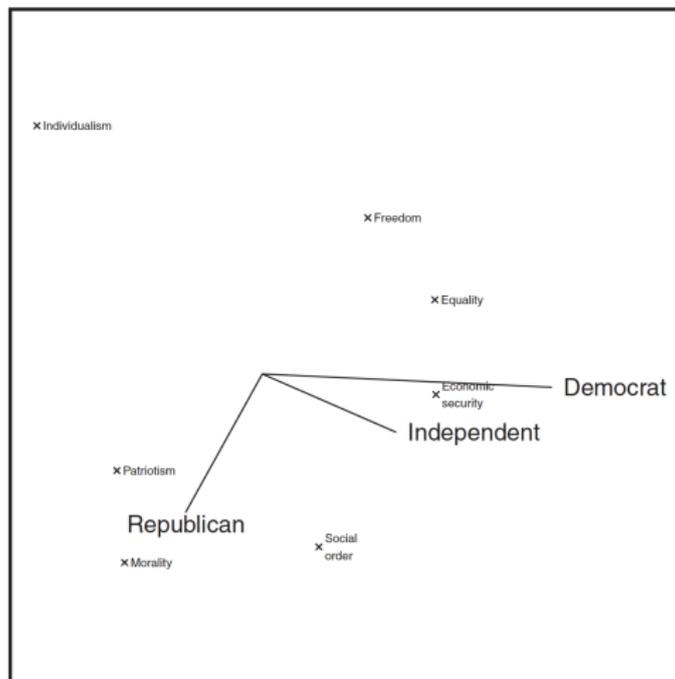
Example: Jacoby 2014

Vector from dataset of rank-ordered value preferences



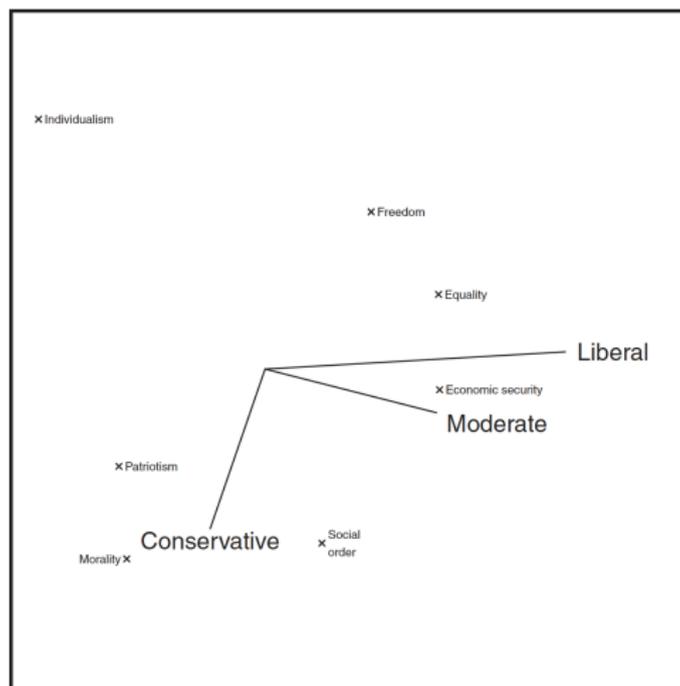
Example: Jacoby 2014, cont'd

Plotting average vectors for partisan self-identifiers



Example: Jacoby 2014, cont'd

Plotting average vectors for ideological self-identifiers



Vectors and Angles

- Angles between vectors have same interpretation as with the biplot or factor analysis
- Correlation between vectors = cosine of the angle between the vectors:

$$\cos\theta_{xy} = \frac{x \cdot y}{|x||y|}$$

- The length of $x = |x| = \sqrt{x \cdot x} = \sqrt{x_1x_1 + x_2x_2 \dots + x_mx_m}$
- Can use the vector coordinations to calculate average vector, or average vector for groups
 - ▶ Direction of mean vectors will correspond to average preference profile
 - ▶ Length of average vector will correspond to homogeneity of group preferences, where longer mean vectors denote greater homogeneity of preferences
- Can use Analysis of Angular Variation (ANAVA), the vector variant of ANOVA, to test whether group vectors are significantly different

Power of the Vector Model

- We reduced dimensionality
- Made a picture that is fairly easy to interpret
- Used some basic geometry to visually represent:
 - ▶ The distribution of preferences
 - ▶ Average preferences of particular groups, and within group variance of those preferences
 - ▶ Correspondence (correlation) between preferences of individuals and groups
- Created a parsimonious measure of individual preferences that can be used in subsequent statistical analyses

Example: Jacoby 2014, cont'd

Can use individual vector positions along the unit circle to do a “circular regression”

Conceptualize vector positions as angular deviations from some arbitrary, fixed vector position (say, 12:00)

TABLE 3. Circular Regression Showing the Effects of Social and Political Characteristics on Individual Preference Vector Orientations

	MLE Coefficient	Standard Error	Observed Probability
Respondent's Age	-0.008	0.002	0.000
Some College	0.148	0.051	0.002
College Graduate	0.429	0.065	0.000
Religious Commitment	-0.020	0.006	0.000
Party Identification	-0.109	0.016	0.000
Lib-Con Ideology	-0.140	0.026	0.000
Intercept (μ)	-0.951	0.044	0.000

Multidimensional Unfolding

- Very similar to classical multidimensional scaling
- Major difference is that we need distances between row and column objects that are not the same
- Can also be estimated using the SMACOF algorithm in the “smacof” package in R using the “smacofRect” function
- Minimizes a loss function just like classical MDS

$$\sum_{i=1}^n \sum_{j=1}^q (d_{ij} - \delta_{ij})^2$$

- This looks just like Stress, except our error (differences between distances and dissimilarities) is summed over all pairs of n and q , since rows and columns are different

Multidimensional Unfolding, cont'd

Δ matrix for a multidimensional unfolding analysis:

$$D = \begin{matrix} n \times q \\ \left[\begin{array}{cccc} \sum_{k=1}^s (x_{1k} - z_{1k})^2 & \sum_{k=1}^s (x_{1k} - z_{2k})^2 & \cdot & \cdot & \cdot & \sum_{k=1}^s (x_{1k} - z_{qk})^2 \\ \sum_{k=1}^s (x_{2k} - z_{1k})^2 & \sum_{k=1}^s (x_{2k} - z_{2k})^2 & \cdot & \cdot & \cdot & \sum_{k=1}^s (x_{2k} - z_{qk})^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sum_{k=1}^s (x_{nk} - z_{1k})^2 & \sum_{k=1}^s (x_{nk} - z_{2k})^2 & \cdot & \cdot & \cdot & \sum_{k=1}^s (x_{nk} - z_{qk})^2 \end{array} \right] \end{matrix}$$

Software computes this for you – can just submit a rectangular matrix of proximity data

Weighted MDS: The INDSCAL Model

- INDSCAL = “individual differences scaling” (sometimes called “three-way MDS”)
- Question it asks, answers:
 - ▶ What if we have multiple dissimilarities matrices that we can model?
 - ▶ Perhaps we have two experimental conditions (control + treatment), where (dis)similarity perceptions are hypothesized to be different?
 - ▶ Or, maybe we just think that two groups of people think about the same set of stimuli in a different way?
- Goal:
 1. Find a set of stimuli coordinates that satisfy the standard goals of simpler MDS procedures
 2. Find the matrix of weights that “stretch” and/or “shrink” the MDS dimensions for group – these weights correspond to the relative emphasis placed on the MDS dimensions

Weighted MDS: The INDSCAL Model, cont'd

- Uses a *weighted* Euclidean distance to model dissimilarities:

$$d_{ij}(X_k) = \sqrt{\sum_{s=1}^p c_{sk} (x_{is} - x_{js})^2}$$

- c_{sk} = “salience” or “importance” of dimension s for individual/group k
- c_{sk} is a diagonal matrix of weights – if we replace the diagonal weights with 1's, the model degenerates to classic MDS
- The groups/individuals k are the “third way” of the data
- Everything else is the same regarding assessing dimensionality, assessing fit, etc.

Options for Estimating MDS Models

1. Smacof, developed by de Leeuw and colleagues (lots of papers)
 - ▶ R package of same name, ordinal and interval input data
 - ▶ Won't fit unidimensional models, no dichotomous data
 - ▶ In a pinch, could fit 2-dimensional model to unidimensional data and just use first dimension coordinates
2. ALSCAL and PROXSCAL in SPSS
3. PROC MDS/PROC MDS in SAS
4. Can *basically* do vector unfolding model with nonlinear PCA (“catpca” in SPSS)