MULTIDIMENSIONAL UNFOLDING

JAN DE LEEUW

ABSTRACT. This is an entry for The Encyclopedia of Statistics in Behavioral Science, to be published by Wiley in 2005.

The unfolding model is a geometric model for preference and choice. It locates individuals and alternatives as points in a joint space, and it says that an individual will pick the alternative in the choice set closest to its ideal point. Unfolding originated in the work of Coombs [4] and his students. It is perhaps the dominant model in both **scaling of preferential choice** and **attitude scaling**.

The multidimensional unfolding technique computes solutions to the equations of unfolding model. It can be defined as **multidimensional scaling** of off-diagonal matrices. This means the data are dissimilarities between n row objects and m column objects, collected in an $n \times m$ matrix Δ. An important example is preference data, where δ_{ij} indicates, for instance, how much individual i dislikes object j. In unfolding we have many of the same

Date: April 3, 2004.
Key words and phrases. fitting distances, multidimensional scaling, unfolding, choice models.

distinctions as in general multidimensional scaling: there is unidimensional and multidimensional unfolding, metric and nonmetric unfolding, and there are many possible choices of loss functions that can be minimized.

First we will look at (metric) unfolding as defining the system of equations $\delta_{ij} = d_{ij}(X, Y)$, where X is the $n \times p$ configuration matrix of row-points, Y is the $m \times p$ configuration matrix of column points, and

$$d_{ij}(X, Y) = \sqrt{\sum_{s=1}^{p}(x_{is} - y_{js})^2}.$$

Clearly an equivalent system of algebraic equations is $\delta_{ij}^2 = d_{ij}^2(X, Y)$, and this system expands to

$$\delta_{ij}^2 = \sum_{s=1}^{p} x_{is}^2 + \sum_{s=1}^{p} y_{js}^2 - 2\sum_{s=1}^{p} x_{is}y_{js}.$$

We can rewrite this in matrix form as $\Delta^{(2)} = ae_m' + e_nb' - 2XY'$, where a and b contain the row and column sums of squares, and where e is used for a vector with all elements equal to one. If we define the centering operators $J_n = I_n - e_ne_n'/n$ and $J_m = I_m - e_me_m'/m$, then we see that doubly centering the matrix of squared dissimilarities gives the basic result

$$H = -\frac{1}{2}J_n\Delta^{(2)}J_m = \tilde{X}\tilde{Y}',$$

where $\tilde{X} = J_nX$ and $\tilde{Y} = J_mY$ are centered versions of X and Y. For our system of equations to be solvable, it is necessary that $\textbf{rank}(H) \leq p$. Solving the system, or finding an approximate solution by using the singular value decomposition, gives us already an idea about X and Y, except that we do not know the relative location and orientation of the two points clouds.

More precisely, if $H = PQ'$ is is full rank decomposition of H, then the solutions X and Y of our system of equations $\delta^2_{ij} = d^2_{ij}(X, Y)$ can be written in the form

$$X = (P + e_n\alpha')T,$$

$$Y = (Q + e_m\beta')(T')^{-1},$$

which leaves us with only the $p(p + 2)$ unknowns in α, β, and T still to be determined. By using the fact that the solution is invariant under translation and rotation we can actually reduce this to $\frac{1}{2}p(p + 3)$ parameters. One way to find these additional parameters is given in [10].

Instead of trying to find an exact solution, if one actually exists, by algebraic means, we can also define a multidimensional unfolding loss function and minimize it. In the most basic and classical form, we have the Stress loss function

$$\sigma(X, Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}(\delta_{ij} - d_{ij}(X, Y))^2$$

This is identical to an ordinary multidimensional scaling problems where the diagonal (row-row and column-column) weights are zero. Or, to put it differently, in unfolding the dissimilarities between different row objects and different column objects are missing. Thus any multidimensional scaling program that can handle weights and missing data can be used to minimize this loss function. Details are in [7] or [1, Part III]. One can also consider measuring loss using SStress, the sum of squared differences between the squared dissimilarities and squared distances. This has been considered in [11, 6].

Area	Plot Code
Social Psychology	SOC
Educational and Developmental Psychology	EDU
Clinical Psychology	CLI
Mathematical Psychology and Psychological Statistics	MAT
Experimental Psychology	EXP
Cultural Psychology and Psychology of Religion	CUL
Industrial Psychology	IND
Test Construction and Validation	TST
Physiological and Animal Psychology	PHY

TABLE 1. Nine Psychology Areas

We use an example from Roskam [9, p. 152]. The Department of Psychology at the University of Nijmegen has, or had, 9 different areas of research and teaching. Each of the 39 psychologists working in the department ranked all 9 areas in order of relevance for their work. The areas are given in Table 1. We apply metric unfolding, in two dimensions, and find the solution in Figure 1.

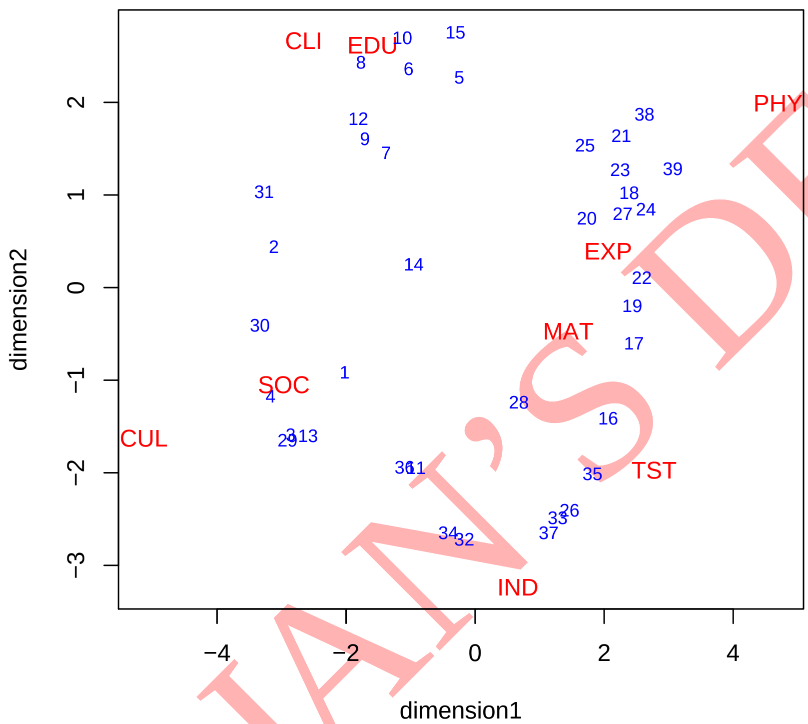

FIGURE 1. Metric Unfolding Roskam Data

In this analysis we used the rank orders, more precisely the numbers 0 to 8. Thus, for good fit, first choices should coincide with ideal points. The grouping of the 9 areas in the solution is quite natural.

In this case, and in many other cases, the problems we are analyzing suggest that we really are interested in nonmetric unfolding. It is difficult to think of actual applications of metric unfolding, except perhaps in the life and physical sciences. This does not mean that metric unfolding is uninteresting. Most nonmetric unfolding algorithms solve metric unfolding subproblems, and one can often make a case for metric unfolding as a robust form to solve nonmetric unfolding problems.

The original techniques proposed by Coombs [4] were purely nonmetric and did not even lead to metric representations. In preference analysis, the protypical area of application, we often only have ranking information. Each individual ranks a number of candidates, or food samples, or investment opportunities. The ranking information is row-conditional, which means we cannot compare the ranks given by individual i to the ranks given by individual k. The order is defined only within rows. Metric data are generally unconditional, because we can compare numbers both within and between rows. Because of the paucity of information (only rank order, only row-conditional, only off-diagonal) the usual Kruskal approach to nonmetric unfolding often leads to degenerate solutions, even after clever renormalization and partitioning of the loss function [8]. In Figure 2 we give the solution minimizing

$$\sigma(X,Y,\Delta)=\sum_{i=1}^{n}\frac{\sum_{j=1}^{m}w_{ij}(\delta_{ij}-d_{ij}(X,Y))^2}{\sum_{j=1}^{m}w_{ij}(\delta_{ij}-\delta_{i\star})^2}$$

over X and Y and over those Δ whose rows are monotone with the ranks given by the psychologists. Thus there is a separate **monotone regression** computed for each of the 39 rows.

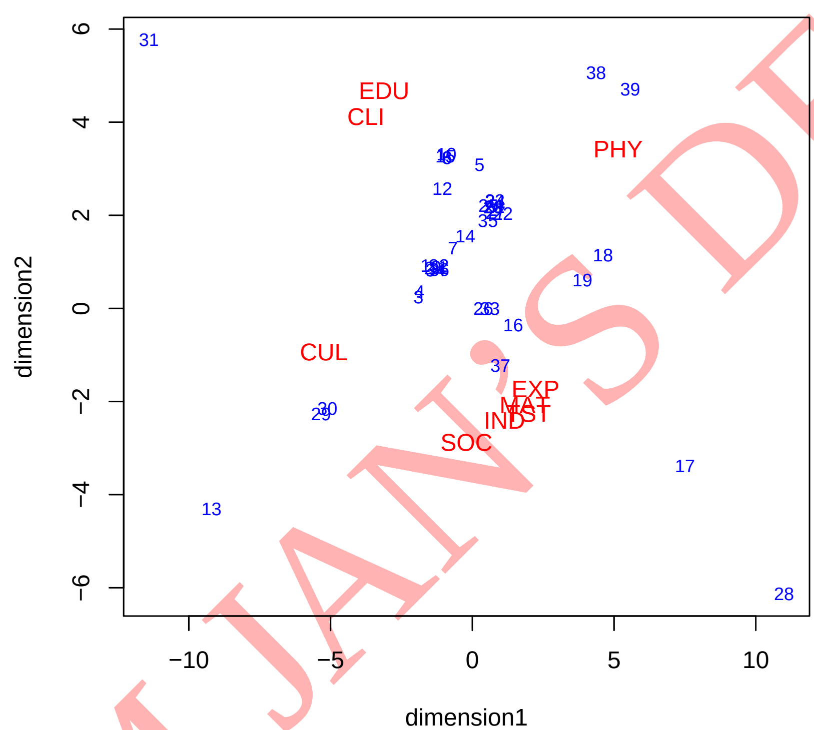

FIGURE 2. Nonmetric Unfolding Roskam Data

The solution is roughly the same as the metric one, but there is more clustering and clumping in the plot, and this makes the visual representation much less clear. It is quite possible that continuing to iterate to higher precision will lead to even more degeneracy. More recently Busing et al. [2] have adapted the Kruskal approach to nonmetric unfolding by penalizing for the flatness of the **monotone regression** function.

One would expect even more problems when the data are not even rank orders but just binary choices. Suppose n individuals have to choose one alternative from a set of m alternatives. The data can be coded as an *indicator matrix*, which is an $n \times m$ binary matrix with exactly one unit element in each row. The unfolding model says there are n points x_i and m points y_j in $\mathbb{R}^p$ such that, if individual i picks alternative j, then $\|x_i - y_j\| \leq \|x_i - y_\ell\|$ for all $\ell = 1, \ldots, m$. More concisely, we use the m points y_j to draw a *Voronoi diagram*. This is illustrated in Figure 3 for six points in the plane.

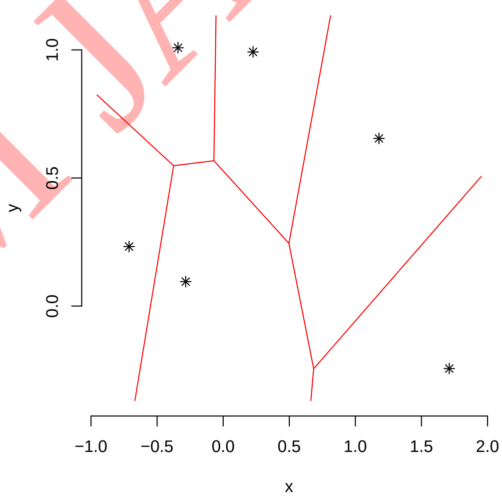

FIGURE 3. A Voronoi Diagram

There is one Voronoi cell for each the y_j, and the cell (which can be bounded on unbounded) contains exactly those points which are closer to y_j than to any of the other y_ℓ. The unfolding model says that individuals are in the Voronoi cells of the objects they pick. This clearly leaves room for a lot of indeterminacy in the actual placement of the points.

The situation becomes more favorable if we have more than one indicator matrix, that is if each individual makes more than one choice. There is a Voronoi diagram for each choice and individuals must be in the Voronoi cells of the object they choose for each of the diagrams. Superimposing the diagrams creates smaller and smaller regions that each individual must be in, and the unfolding model requires the intersection of the Voronoi cells determined by the choices of any individual to be nonempty.

It is perhaps simplest to apply this idea to binary choices. The Voronoi cells in this case are half spaces defined by hyperplanes dividing $\mathbb{R}^n$ in two parts. All individuals choosing the first of the two alternatives must be on one side of the hyperplane, all others must be on the other side. There is a hyperplane for each choice.

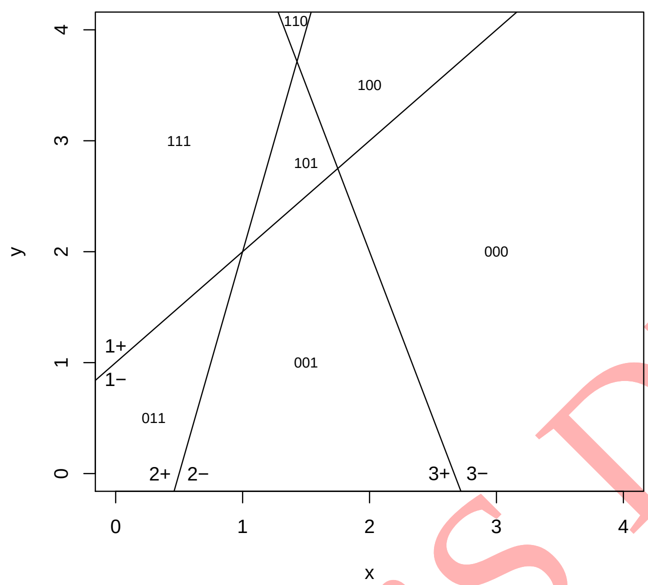

FIGURE 4. Unfolding Binary Data

This is the nonmetric factor analysis model studied first by Coombs and Kao [5]. It is illustrated in Figure 4.

The prototype here is roll call data [3]. If 100 US senators vote on 20 issues, then the unfolding model says that (for a representation in the plane) there are 100 points and 20 lines, such that each issue-line separates the "aye" and the "nay" voters for that issue. Unfolding in this case can be done by **correspondence analysis**, or by maximum likelihood logit or probit techniques. We give an example, using 20 issues selected by Americans for Democratic Action, and the 2000 US Senate.

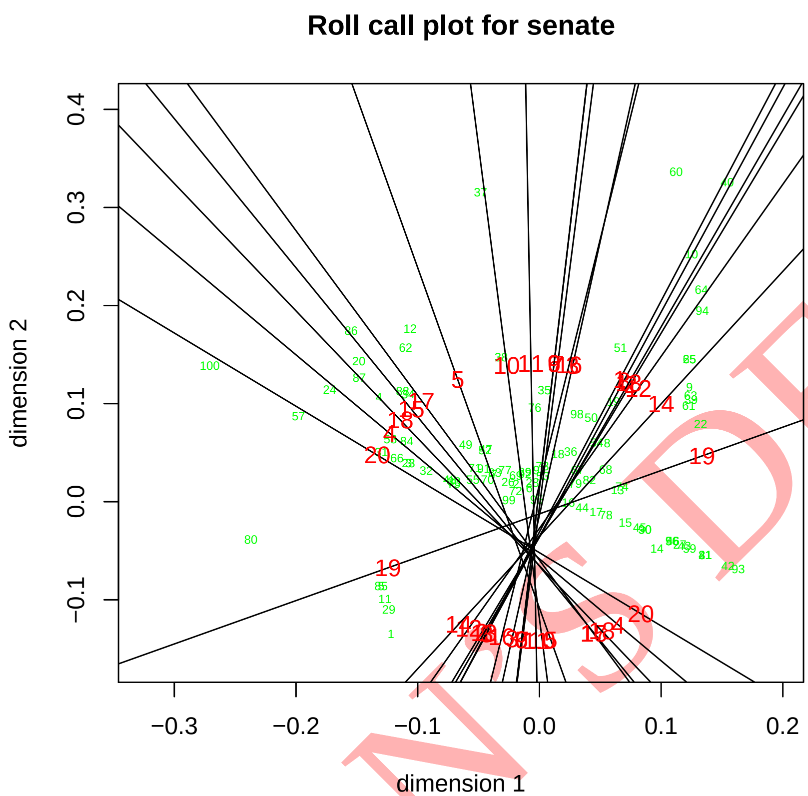

FIGURE 5. The 2000 US Senate

REFERENCES

[1] I. Borg and P.J.F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, New York, 1997.

[2] F.M.T.A. Busing, P.J.F. Groenen, and W.J. Heiser. Avoiding Degeneracy in Multidimensional Unfolding by Penalizing on the Coefficient of Variation. *Psychometrika*, 2004.

[3] J. Clinton, S. Jackman, and D. Rivers. The Statistical Analysis of Roll Call Data. *American Political Science Review*, (in press).

[4] C. H. Coombs. *A Theory of Data*. Wiley, 1964.

[5] C.H. Coombs and R.C. Kao. Nonmetric Factor Analysis. Engineering Research Bulletin 38, Engineering Research Institute, University of Michigan, Ann Arbor, 1955.

[6] M.J. Greenacre and M.W. Browne. An Efficient Alternating Least-Squares Algorithm to Perform Multidimensional Unfolding. *Psychometrika*, 51:241–250, 1986.

[7] W.J. Heiser. *Unfolding Analysis of Proximity Data*. PhD thesis, University of Leiden, 1981.

[8] J.B. Kruskal and J.D. Carroll. Geometrical Models and Badness of Fit Functions. In P.R. Krishnaiah, editor, *Multivariate Analysis, Volume II*, pages 639–671. North Holland Publishing Company, 1969.

[9] E.E.CH.I. Roskam. *Metric Analysis of Ordinal Data in Psychology*. PhD thesis, University of Leiden, 1968.

[10] P.H. Schönemann. On Metric Multidimensional Unfolding. *Psychometrika*, 35:349–366, 1970.

[11] Y. Takane, F.W. Young, and J. De Leeuw. Nonmetric Individual Differences in Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 42:7–67,

1977.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu