Sawtooth Software

TECHNICAL PAPER SERIES

The CPM System for Composite Product Mapping

530 W. Fir St.
Sequim, WA 98382
(360) 681-2300
www.sawtoothsoftware.com

The CPM System for Composite Product Mapping

Sequim, WA 98382 USA
http://www.sawtoothsoftware.com

Introduction

The CPM System for Composite Product Mapping is a method for analyzing data about respondent perceptions and preferences to produce a "product map." A product map is a graphic representation of the ways people perceive products in terms of underlying attributes, as well as an aid in understanding their preferences.

We use the term "products," but this can refer to any collection of objects, concepts, or other ideas that can be described in terms of characteristics or attributes. Examples include:

Consumer products (toothpaste, lawnmowers, or frozen dinners)
Industrial products (airliners, road graders, or nuclear power plants)
Institutions (hospitals, corporations, or restaurants)
Services (HMOs, air traffic control systems, Internet service providers)
People (political candidates, TV anchorpersons, product spokespersons)

CPM uses data about people's **perceptions** (ratings of products on attributes) and their **preferences** (choices among products). Suppose we represent a city convention bureau, and want to find out what travelers think of us, compared to other cities. We might ask people to rate us and our competition on several attributes, such as:

Quality cultural entertainment
Safe and efficient public transportation
Reputation for safety
Activities for children
Historical landmarks
Shopping variety
Informality

We could compute the average rating for each city on each attribute. If we arranged those in a table we could look across the rows or down the columns to learn about our relative strengths and weaknesses. This can be a useful way to examine data, but there are some difficulties:

- If there are many products or many attributes, the table can become very large and hard to comprehend.

- Not all attributes are equally important. How much attention should we pay to each?

- Since many attributes may be very similar, or perhaps redundant, should we combine them in some way?

- Although we can learn whether we score higher or lower than other cities on a particular attribute, how can we tell what the "best" level is on an attribute such as "informality," where more may not necessarily be better?

It can be very helpful to condense a large and potentially confusing mass of data into a meaningful "picture." That is what CPM does.

CPM produces a "map" of the "product space." It shows how the products are perceived by respondents. Each product occupies a specific point in the space. Products that are similar are close together, while those that are very different are far apart. There can be as many dimensions as there are fundamental ways in which the products are seen to differ, but frequently a simple two-dimensional space will account for most of the information.

The attributes are represented as "arrows" radiating from the center of the space. Attributes that are closely related are represented by nearly parallel arrows, while those that are unrelated are represented by arrows at right angles. The relative ratings of the products on each attribute are communicated by the relationships between product points and attribute arrows. This graphic display makes it possible to visualize brand image information for rapid intuitive understanding. Researchers have found such product maps to be useful for understanding product images, customer desires, and the ways products might be modified to become more successful. Maps have proven to be valuable tools in the development of marketing strategy, as well as in communicating strategic ideas.

The Basic Idea Underlying CPM

Perceptual data have often been used to create product maps. The most common perceptual data are ratings of products on descriptive attributes. Mapping methods based on discriminant analysis or principal component analysis find positions of products in a space of few dimensions, as well as directions for attribute vectors. Product positions and attribute directions are chosen so that products' projections on each attribute vector are related to their ratings on that attribute.

Maps based on perceptions are usually easy to interpret, and good at conveying insights about product images. But, they are sometimes unsuccessful at predicting preferences. This may be because they tend to focus on differences among products that are easy to see, but which may be less important in accounting for preferences.

Preference data have also been used to create product maps. With such maps, each individual is assumed to have an ideal direction or an ideal point in the product space. Locations for each product, as well as for each individual's ideal point or ideal direction, are chosen so that products' projections on each respondent's ideal direction, or nearness to each respondent's ideal point, are related to that individual's stated preferences.

Maps based on preferences are naturally better at accounting for preference than maps based on perceptions. However, sometimes their dimensions are hard to interpret. For example, consider cups of coffee with differences in temperature ranging from boiling to tepid. Everyone would probably prefer some middle temperature, rejecting both extremes. But if no perceptual information is available to establish their differences on the underlying temperature scale, the most extreme cups may be close together in a preference-based map, because their only recognized property is that they are both rejected by everyone.

There is another problem with maps of both types. Nearly all product maps are based on aggregate data combining preferences or perceptions of many individuals. However, individuals often disagree about products. If that is true, no aggregate map can hope to capture the perceptions or preferences of individuals very precisely.

The CPM System for Composite Product Mapping seeks to overcome these difficulties, by producing maps that use both perceptual ***and*** preference data. There are two composite models, a "vector model" that assumes each individual has an ideal **direction** in the product space, and an "ideal point model" that assumes each individual has an ideal **point** in the space.

CPM uses perceptual data consisting of ratings of products on attributes. The preference data may be either paired-comparison preference proportions or conjoint part worths for the products rated. When the preference data are paired comparisons, these are the same kinds of data as used by APM, a perceptual mapping product released by Sawtooth Software in 1985.

CPM's two composite models share several characteristics:

- Every respondent is assumed to have a unique perceptual space, determined by his/her own ratings of products on attributes.

- Each dimension in the individual's space is a weighted combination of his/her ratings of products on attributes.

- However, the weights defining the dimensions are required to be ***identical*** for all respondents.

- Those dimensional weights are determined by optimizing the fit (over all individuals) between actual preferences and the preferences inferred from the individual perceptual spaces.

The overall product map is just the average of the individual maps. Since it is also a weighted combination of average product ratings, it is truly a ***perceptual*** map, although its dimensions are chosen so as to best account for preferences. In this way we produce maps that are firmly grounded in descriptive attributes, but which best account for individual preferences.

Both models require the estimation of two sets of parameters. One set consists of **dimensional weights**, identical for all individuals, to be applied to attribute ratings to obtain products' positions in each individual's space. The other parameters are unique for each individual: either individual **importance weights** in the case of the vector model, or individual **ideal point coordinates** in the case of the ideal point model.

We have tested composite methods with several data sets, with two results:

- When there are attributes on which products differ but which are unimportant in forming preferences, composite maps are better than conventional perceptual maps at accounting for preferences.

- When all attributes are involved in forming preferences, composite maps look very much like maps based on perceptual data alone.

Together, these results suggest that composite methods can provide insurance against unfortunate choices of attributes. Also, since composite maps provide a relatively tight linkage between perceptions and preferences, they may be used to forecast relative demand for new or modified products. All in all, there seems to be no downside to using composite mapping methods, and the benefit of possibly improved interpretation and prediction can be great.

The Relationship of CPM and APM

CPM may be regarded as a successor to APM. The APM System for Adaptive Perceptual Mapping, introduced by Sawtooth Software in 1985, had three main parts: a **questionnaire** module for constructing and administering computer-assisted interviews, a **mapping** module that makes discriminant-based perceptual maps, and a **simulation** module for estimating the result of modifying a product's levels on attributes. Since its introduction in APM, discriminant-based mapping has proved preeminently useful. In addition to its two composite models, CPM also includes the capability of making discriminant-based maps like those of APM.

In the years since APM's introduction, it has become apparent that conjoint analysis provides a more powerful method for "what if" simulations than mapping-based simulators. Rather than a simulator that tries to predict the result of changing a product on specific attributes (which is better left to conjoint analysis) CPM provides a method for estimating the density of demand at each point in the product space.

Data Requirements for CPM

Composite product maps require two kinds of data from each respondent: ratings of products on attributes, and preference data for the same products. CPM was designed with the recognition that researchers are usually interested in many more products and attributes than can be considered in detail by any single respondent. The typical CPM interview has these steps:

- Determine how familiar the respondent is with each product, and select a subset of familiar products.
- Determine the relative importance to the respondent of each attribute, and select a subset of important attributes.
- Have the respondent rate the selected products on the selected attributes.
- For the products rated, present a number of paired comparison preference questions. The respondent is asked to allocate 100 points of preference between the products in each pair. The product pairs should be chosen carefully, and we suggest below how to do so. If conjoint data are available for the same respondents, CPM can use conjoint part worths instead of paired comparison preferences.

The first two steps are optional. If the total numbers of attributes and/or products are small enough to be considered by each respondent, it is better to have each respondent rate every product on every attribute. However, it is more likely that you will have more attributes and products, in which case you must consider how many products each respondent should rate, and on how many attributes.

CPM does much of its analysis at the level of the individual respondent, and it requires product ratings and preference data from each individual. CPM is ***not*** able to work from aggregate tables of product means on attributes. Its ability to account for preferences is aided by having each respondent consider several products. As an absolute lower limit, each respondent ***must*** rate at least two products. For product maps with only two dimensions, each respondent should rate at least three products, and it is better if each respondent rates 5 products. If you intend to produce maps with more than two dimensions, it would be best to have each respondent rate even more products.

If a respondent does not provide ratings for a particular attribute, then missing data are estimated by assuming the respondent would have rated all products as equal on that attribute. Although a random pattern of such missing data is not severely damaging to the analysis, it is better if the number of attributes is small enough that all respondents can provide ratings using all of them.

Less is lost by having each respondent consider only a subset of products. Since much of the analysis is done separately for each respondent, it is not important that all respondents rate all products, and it is better to have knowledgeable ratings on fewer products than questionable ratings on more of them.

To define upper limits of products and attributes, CPM comes in large and small sizes. The small size is limited to 10 products and 15 attributes, the second size is limited to 30 products and 50 attributes, and the large size permits a maximum of 90 products and 90 attributes.

Types of Data: Product Ratings

The perceptual data can be of almost any type. We speak of them as “ratings,” but they need not come from rating scales. Numerical estimates in pounds, miles, or hours will work equally well. For example, one might ask questions like any of the following:

How well does this statement describe this brand?
How much would you agree with this description of this brand?
How well does this description apply to this brand?
How much would this political candidate agree with this statement?
How many pounds do you think this computer would weigh?
How many minutes are you kept on hold when you call for customer support?
How many miles per gallon would this car deliver?

The data may be collected using rating scales, magnitude estimation, or any other means. The data can be integer, or may include decimal places.

We suggest that a respondent rate products on each attribute before moving on to the next attribute. CPM is concerned with differences among products, which are measured most accurately when they are considered together. If rating scales are used in a computer-assisted interview, it is useful to let the respondent see his/her ratings of previous products as each subsequent product is considered. Also, it is best if both products and attributes are presented to each respondent in randomized order. However, we recommend that the order of products and attributes be consistent within each interview.

Types of Data: Preference Information

The most common type of preference data for CPM is a set of paired-comparison preference judgements, where the respondent allocates 100 points of preference among the products in each pair. Each respondent must answer at least paired preference question if you plan to make composite product maps, and the upper limit for a respondent is 100, more than any respondent should ever receive. The pairs are usually chosen uniquely for each respondent.

The CPM System can also use conjoint part worths rather than paired-comparison preference judgements. CPM can directly read conjoint utilities from ASCII files, including those generated by Sawtooth Software's ACA, CVA, CBC/HB and ICE systems.

Geometric Models of Perception and Preference

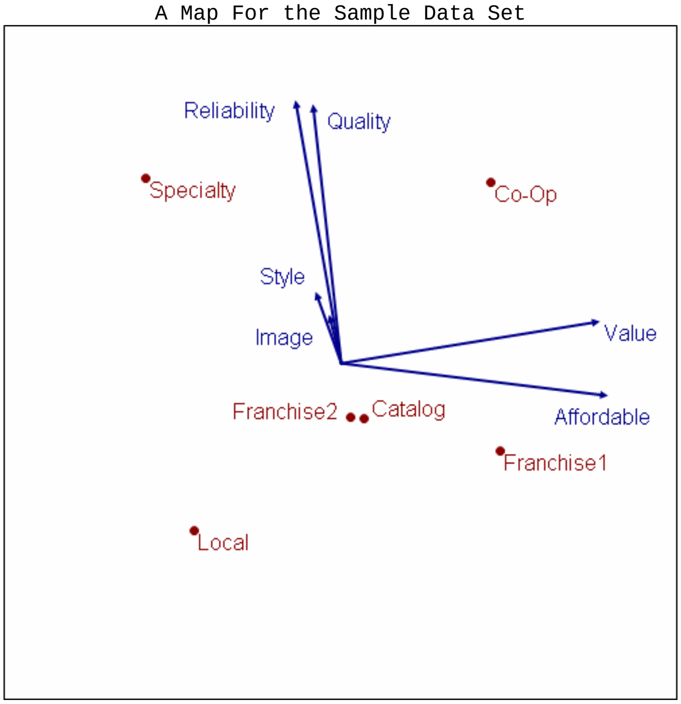
There are many kinds of product maps. To be clear about what CPM ***does***, we first mention two kinds of maps that CPM ***does not*** make:

> **Similarities-based mapping**, sometimes called "multidimensional scaling," is of considerable historical importance. The data for such maps are judgements of overall similarity of pairs of products, rather than ratings of products on attributes. Similarities-based mapping was used in early marketing research applications but has become rare in recent years. One strong advantage of similarities-based mapping is that the researcher doesn't need to know ahead of time what attributes are important. One disadvantage is that similarities judgements seem less useful in general than ratings of products on attributes. Since product ratings are normally collected in many marketing research studies, data for ratings-based mapping are often available at no extra cost.
>
> **Correspondence analysis**, a popular mapping method originally developed as a tool for dealing with frequencies rather than ratings. One type of data for which correspondence analysis would be appropriate is a products x attributes table of frequencies, such as number of times each attribute is judged as applying to each product. Sometimes correspondence analysis is also applied to tables of means. A strong advantage of correspondence analysis is that it can be used on aggregate data, and does not require that data be available for individual respondents. One disadvantage is that there is some disagreement among experts about how its maps should be interpreted. It is clear that one may make inferences about relationships among products and about relationships among attributes. However, experts are divided on the appropriateness of making inferences about how products are related to attributes, which is a critical question in many mapping studies.

CPM makes three kinds of maps: **discriminant** maps, **composite vector** maps, and **composite ideal point** maps. Each type of map is described in detail in this technical paper. Although there are important differences among these methods, they share many fundamentals, which we now describe.

CPM maps use a "point and vector" approach. Products are represented as points in space and attributes are represented as vectors (directions) in that space. For illustration, Figure 1 shows a map made from the Sample data set included in the CPM System. The product labels are upper case and the attribute labels are mixed upper and lower case. (A word of caution: these data were produced artificially, just for illustration. No substantive conclusions should be drawn from them.)

Figure 1
A Map For the Sample Data Set



Products are close together in space if they are rated similarly on attributes, and farther apart the more differently they are rated. We can deduce that Franchise2 and Catalog are seen to be similar. Franchise1 and Franchise2 are seen to be less similar. Specialty is seen as relatively dissimilar to all other products, as are Co-Op and Local.

If two attributes are positively correlated, in the sense that products with high or low ratings on one tend to have correspondingly high or low ratings on another, those attribute vectors tend to point in the same direction. If two attributes are negatively correlated, then their vectors tend to point in opposite directions. If two attributes are uncorrelated, their vectors tend to be at right angles to one another. It can be seen that Quality and Reliability are highly correlated, as are Value and Affordability. Also, those two pairs of vectors, being nearly at right angles, are essentially uncorrelated with each other.

If our space had as many dimensions as the number of attributes, or one less than the number of products, then all attribute vectors would be of unit length. However, sometimes an attribute is not involved in a particular two-dimensional space. The lengths of the vectors are important indicators of their salience in describing that space. For example, in this space one dimension seems to be identified mostly by Quality and Reliability, and the other seems to be identified mostly by Value and Affordability. Notice that Style and Image have very short vectors. If the products do differ systematically on those attributes, it must be in a different dimension.

Relationships between products and attributes are deduced from the "projections" of products on attribute vectors. Imagine rotating this figure slightly so the Quality vector were to become exactly vertical. Then the relative perceptions of the products on the Quality attribute would be indicated by their relative heights, with Specialty slightly ahead of Co-Op and Local having the perception of least quality. Similarly, Co-Op appears to have the strongest position with respect to Value and Affordability, closely followed by Franchise1, with Specialty and Local seen as lowest on Value and Affordability. Notice that we pay no attention to how ***close*** a point is to a vector, but rather how far the product is from the center of the space in the ***direction*** the vector is pointing.

It may be obvious that all of the properties we have just described would be equally true if we were to turn the page upside-down, so that Quality and Reliability pointed downward. Just as clearly, we could look at a mirror image of the diagram, in which the horizontal axis has been reversed, and the relationships between products and attributes would be unchanged. We are thus free to reflect either axis, changing all the plusses to minuses and vice versa. We are also free to rotate the entire configuration any desired number of degrees, either clockwise or counter-clockwise, without changing its meaning in any way.

However, when people look at product maps, they often wonder which are the favorable directions. This question is answered most easily if we adopt a convention of rotating our maps so that, to whatever extent possible, the upper-right quadrant is the one in which preferred products are located. We have done that in our example. The Co-Op product has the highest preference scores of the products in our example, which might be inferred from the fact that it alone has favorable projections on all four of the attributes strongly involved in this space: Quality, Reliability, Value, and Affordability.

Although the geometric notions we have described are equally applicable to all three types of maps produced by CPM, those maps differ considerably in how their dimensions are chosen. We now discuss each type in more detail.

Discriminant Maps: The Underlying Idea

Given that one has decided to collect perceptual data by having respondents rate products on attributes, then multiple discriminant analysis is the most powerful method available for producing a map from those ratings. By "powerful," we refer to its ability to compress the largest amount of perceptual information into a map of fewest dimensions. Perceptual maps are aids to visual comprehension, and it is vastly preferable to have only two, or at most three, dimensions. The interpretation of higher-dimensional spaces, difficult for everyone, can be almost impossible for many individuals.

To gain an intuitive feeling for how multiple discriminant analysis works, it is first useful to understand the statistical concept of an F ratio. Suppose respondents have rated several products on a particular attribute. The variation among those ratings can be thought of as being of two types:

> The differences ***between*** products, revealed in the differences between average ratings for different products. The more these averages vary from one another, the greater the overall difference between products on this attribute.
>
> The differences ***within*** products, revealed in the differences among respondents' ratings of the same product.

The variation within products is often considered to indicate the level of random "noise," and can provide a yardstick with which to measure the amount of variation between products. A statistic frequently used in this context is the "F ratio," a ratio of the variance between ratings of different products to the variance of the ratings within products. An attribute would have a higher F ratio either if its product averages were more different from one another, or if there were more agreement among respondents rating the same product.

If we had product ratings on many attributes, we could compute the F ratio for each attribute. The attribute with the largest F ratio would be the one on which products are seen to differ most dramatically. If we had to choose the single attribute on which the products were most different, that would be the attribute with the highest F ratio.

Sometimes it is possible to increase efficiency by combining measures. Suppose we had two attributes which were concerned with similar aspects of products (such as Reliability and Quality, or Affordability and Value). Even though each attribute had a sizeable F ratio on its own, we might be able to distinguish even more dramatically between products by creating a new variable obtained by averaging two others. We might expect the random error in ratings of the two attributes to cancel one another at least partially, leaving us with a more precise way of differentiating the products, indicated by a higher F ratio.

Multiple discriminant analysis goes one step further. It finds that ***optimal*** weighted combination of ***all*** the attributes which would produce the highest F ratio of between-product to within-product variation. That weighted combination of attributes becomes the first dimension of the perceptual map. It contains the most information possible for any single dimension, in terms of accounting for perceived differences among products, as measured by an F ratio.

Then a second weighted combination of all attributes is found which has the next highest F ratio, subject to the constraint that this combination be uncorrelated with the first. The absence of correlation means that we can express the two dimensions graphically as being at right angles.

Two dimensions may be all we need, but we could continue to determine additional optimal weighted combinations of attributes, each uncorrelated with those before, which could serve as the bases of a higher-dimensional space. If we were to continue in this way, we could find as many dimensions as there were attributes, or one fewer than the number of products, whichever is smaller. If we were to compute as many dimensions as possible and note the F ratio for each, then the sum of those F ratios could be taken as a measure of the total amount of information in the data about how products are seen to differ. Each dimension's F ratio can be divided by the sum of all the F ratios to obtain the percentage of information accounted for by each dimension.

We can compute each product's average score on each dimension, which are used to plot the positions of products in the space. The averages for all products are zero on each attribute, and also on each dimension. Geometrically, this means that the "center of gravity" of all the product points lies at the center of the space. Products with vague, undifferentiated images, or those about which respondents disagree, tend to lie near the center of the space. However, a position near the center does not necessarily mean that a product has a blurry image. Those products may be universally agreed to have "medium" characteristics.

We can also compute the correlation coefficient of each attribute with that weighted combination of attributes that constitutes each dimension. We plot each attribute as a vector in the space, emanating from the center of the space to a point that has as its coordinates the correlations of the attribute with the dimensions. This means that an attribute that is highly correlated with a dimension appears on the map as an arrow pointing nearly in the same direction as that dimension, and can be used to identify the dimension. (For users of APM we note that the CPM's attribute correlations are computed differently from APM's. APM computes correlations using covariance both between and within products. CPM bases the correlations on the product means, so it only uses between-product covariance, and the correlations are generally higher.)

The length of an attribute vector in a two-dimensional space is equal to the square root of the sum of its squared correlations with the dimensions. This can never be greater than one, and the relative length of an attribute vector in any two-space is an indicator of the extent to which that attribute is "accounted for" by those two dimensions.

When attributes and products are plotted in the space in this way, there is a useful relationship between them. Those products with highest averages on an attribute are farthest from the center of the space in the direction of that vector, and those products with lowest (most negative) averages are farthest from the center of the space in the opposite direction. If the vector were completely accounted for by a particular two-space (therefore having length of one), the positions of the products would line up perfectly along that vector according to the product's average ratings on the attribute.

Discriminant Maps: Advantages and Disadvantages

It is possible with discriminant analysis to produce maps that are very good at portraying the perceptual differences between products, but whose dimensions are not very good at showing differences in preference. An attribute that is an excellent discriminator among products for a dimension that has no bearing on preference can badly mislead discriminant analysis. Nonetheless, we do not wish to demean discriminant analysis. It remains a powerful way of analyzing perceptual data. It does an excellent job of reducing a large number of attributes to a small number of dimensions. If one had the luxury of carrying along extra dimensions, the resulting map would contain the useful dimensions as well as any useless ones.

The main advantage of discriminant analysis as a mapping technique is efficiency. For any number of dimensions, it will produce the space that accounts for the greatest amount of information about how products differ. If data are available for individual respondents (as opposed to having just a summary table of means) and mapping is to be done only with perceptual data (without preference data), then discriminant analysis is an excellent technique.

The two main disadvantages of discriminant analysis as ordinarily used for mapping are that it requires individual data (as opposed to aggregate data) and that it focuses exclusively on perceptual differentiation among products, rather than including preference data as well. If you have only aggregate data, then some other method must be used, such as correspondence analysis. If you have both perceptual and preference data, then you might consider a method that pays attention to both types of data, such as the composite methods to be described next.

Composite Vector Maps: The Underlying Idea

There are two popular conceptual approaches for integrating perceptions and preferences. One is the concept of an "ideal point." The other approach is that of an "ideal direction," where we assume that for each individual there is some direction in the space for which "farther is better." This concept seems appropriate for product categories in which the principal attributes are of the "more is better" type, such as efficiency, economy, or quality. It makes less sense for attributes that can offer "too much of a good thing," such as strength of flavor, temperature of a beverage, or softness of an automobile seat. The ideal point approach seems to make more sense for such attributes.

Models that entail ideal directions are called "vector models," because each individual is considered to have a vector, or direction in space, governing his/her preferences for products. Products farther from the center of the space in that direction should be preferred to others, and the product farthest from the center in the opposite direction should be liked least.

We repeat the description of the composite models provided earlier:

- Every respondent is assumed to have a unique perceptual space, determined by his/her own ratings of products on attributes.
- Each dimension in the individual's space is a weighted combination of his/her ratings of products on attributes.
- However, the weights defining the dimensions are required to be ***identical*** for all respondents.
- Those dimensional weights are determined by optimizing the fit (over all individuals) between actual preferences and the preferences inferred from the individual perceptual spaces.

To summarize each individual's actual preferences we derive a set of values rather like conjoint utilities. Large positive values indicate that a product is likely to be preferred and large negative values indicate that a product is likely not to be preferred. Although these numbers are not strictly conjoint utilities they have many similar properties, and we refer to them as utilities.

Suppose we have already developed a set of weights (the transformation matrix described in the previous section) that may be applied to product ratings to derive product coordinates in a space. Suppose also that we have applied such weights to an individual's product ratings, and as a result have a personal perceptual space for that individual. Suppose further that we have already determined the individual's ideal direction. We are interested in measuring how well the hypothetical preferences inferred from the perceptual space correspond to the actual preferences summarized by the utilities.

We estimate the individual's "hypothetical" preferences by drawing a line through the center of the personal perceptual space in that direction, and noting the projections of the product points on that line. (This would be equivalent to rotating the space so that the ideal direction were vertical and then measuring the relative height of each product.)

Since we have utilities (derived entirely from preference data), and also hypothetical preferences (derived entirely from a personal perceptual space), we may inquire how well one set of preferences fits the other. If the fit is very good for every respondent, then we have already found a perceptual space that accounts well for everyone's preferences. If not, then we may improve our estimate of the weights used to define each individual's perceptual space.

Composite vector maps are computed using alternating least squares, with the following iterative procedure:

- An initial estimate is made of attribute weights (the "transformation matrix") to transform attribute ratings into spatial coordinates. In a "normal start" this is obtained from a principal component analysis of the aggregated ratings.

- Using those weights, a personal perceptual space is developed for each respondent. A measure of goodness of fit that we call the average r-square for "capture" is accumulated over all individuals. This measures how well the best possible individual predictions, which are based on complete ratings data, are approximated when considering only the subspace determined by the perceptual map.

- While these computations are done for each individual, information is also accumulated that can be used to find a modified set of weights that will provide an improved r-square for capture in the next iteration.

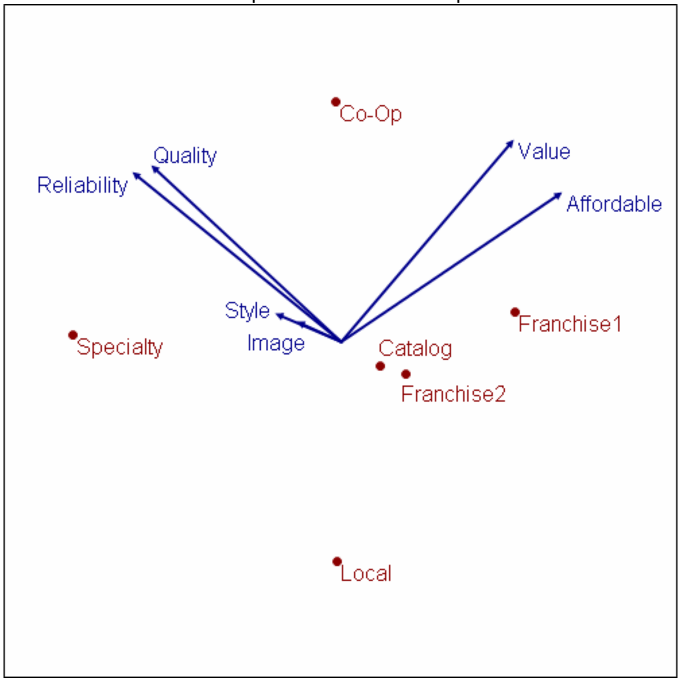
- The iterative process alternates between estimating individual importance values and estimating aggregate attribute weights, at each stage improving the r-squared for capture. Iteration is continued until improvements become smaller than a pre-specified convergence limit, or until a limit on the number of iterations is reached.

The average r-square for capture increases with each iteration, quickly at first and then with smaller and smaller improvements. The process converges rapidly to a global maximum.

The average r-square for "fit" measures how well the hypothetical preferences from the individual perceptual maps are able to predict individual utilities obtained from preference data. This value is reported at the end of the iterative computation, together with the number of violations of pairwise order, comparing hypothetical preferences based on the perceptual space with preferences inferred from the utilities.

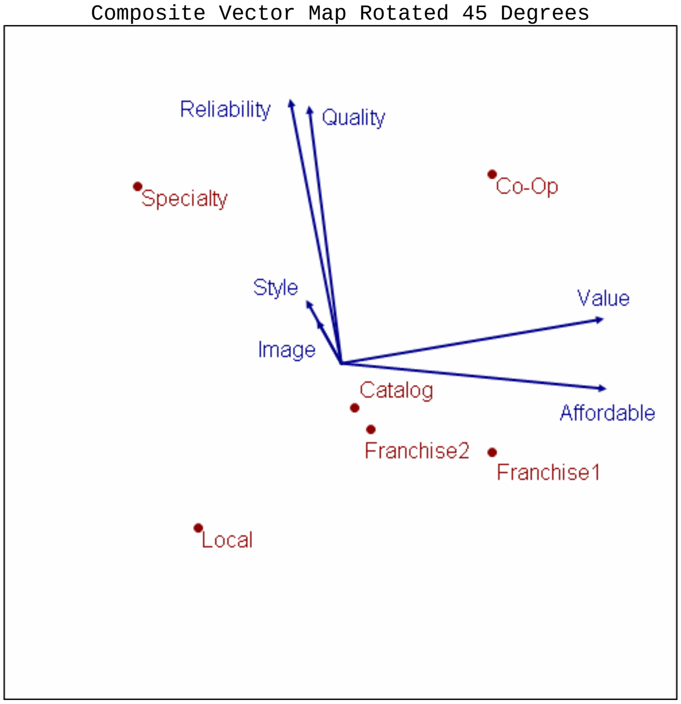
Figure 2 shows the graphic representation of this space. Notice that the vectors for Quality, Reliability, Value, and Affordability are all quite long, and that the vectors for Style and Image are short.

Figure 2
Composite Vector Map



This map could be interpreted; we could call the vertical direction "overall desirability" and the horizontal direction "quality vs. price." But it would be much simpler if we were to rotate all the points and vectors 45 degrees in a clockwise direction, as shown in Figure 3.

Figure 3
Composite Vector Map Rotated 45 Degrees



This map has the same relationships between products and attributes, but it is much simpler to interpret. The horizontal axis is concerned with Value and Affordability and the vertical axis is concerned with Style and Image. Before rotation the average ideal direction was straight up, and now it is at 45 degrees, in the middle of the upper right quadrant. Also, the Co-Op product, which has strongest average preference values, is squarely in the same quadrant.

Composite Vector Maps: Advantages and Disadvantages

The main advantage of composite vector maps is that they select those dimensions that are most critical in the formation of preferences. Composite vector maps often look very much like discriminant maps. For example, if we were to plot discriminant dimensions 2 vs. 3 for the Sample data set, the map would look very much like the composite vector map. But in the case where a discriminant dimension turns out to be unimportant in forming preferences (as with the Sample data set), composite vector maps provide insurance against being misled by an unfortunate choice of attributes.

Additionally, while not quite as quick as discriminant maps, composite vector maps can be obtained more quickly, and the iterative algorithm converges more dependably, than for composite ideal point maps.

Composite vector maps do have a disadvantage compared to discriminant maps. They provide less information about similarities among products than discriminant maps do. Since products are likely to differ more on some attributes than others, and discriminant maps are influenced only by perceptual data, products in a discriminant map are likely to be more spread out on one dimension than another. In contrast, the results for composite vector maps are not dependent on the scaling or even the independence of dimensions. We arbitrarily scale dimensions so that the sum of squares of product coordinates for each dimension is equal to one. This means that there will always be the same amount of visual variation among products in each dimension, which can obscure true differences in perceptual variation among products.

Composite Ideal Point Maps: The Underlying Idea

The composite ideal point approach is like the composite vector approach, with one important difference. Rather than assuming each individual to have an ideal direction, we assume he/she has an ideal point. If we knew that point, we would predict preferences by measuring the distances between products' locations and the respondent's ideal point in his/her personal perceptual space.

Composite ideal point maps are quite different from the discriminant map with explicit ideal points described above. In the discriminant approach we ask the individual for explicit ratings of an ideal product, and then we place that product in the space in the same way that we place actual products in the space. By contrast, the composite ideal point algorithm estimates the individual's ideal point by comparing preferences with perceptions.

We estimate the position of the ideal point for each respondent by minimizing an objective function based on the distances from the ideal point to each product. To compute this objective function we start with the utility values for each respondent (described in the previous section on the composite vector model), subjecting them to a transformation often used with conjoint utilities to estimate shares of preference. That is, we exponentiate each value and then divide by their sum so the resulting values sum to 1.00. We call the resulting numbers "preference weights." They are identical to shares of preference in conjoint analysis when a logit model is used ("Model 2" in Sawtooth Software conjoint simulators).

To illustrate, we compute preference weights for an individual who has rated three products and who has utilities as shown below:

Product	Utility	exp(utility)	Pref. Weight
A	1.0	2.718	.687
B	-.3	.741	.187
C	-.7	.497	.126
	Sum	3.956	1.000

To compute our objective function for a respondent, we compute the squared distance of each product from the ideal point, weight that squared distance by the preference weight for that product, and sum the results. Suppose the three products and ideal point have the coordinates shown below. We compute the objective function for this respondent as follows:

Product	Dimensions I	II	Squared Dist.	Pref. Weight	Squared Dist. x Pref. Weight
A	1.225	0.000	.294	.687	.202
B	0.000	1.225	1.794	.187	.336
C	-1.225	-1.225	5.352	.126	.672
Ideal	.688	.078			Sum =1.212

The sum of 1.212 is the weighted sum of squared distances that we wish to minimize for this respondent. Notice what minimizing this sum accomplishes. The largest weight is for the preferred product, and the smallest for the least preferred product. Since we weight squared distances from the ideal by these preference weights, the sum will be minimized if the distance to the preferred product is small, with distance to the least preferred product being relatively unimportant.

Given these product locations and preferences, the ideal point shown above is optimal (within rounding). If it is moved even slightly in any direction and the sum of weighted squared distances is recomputed, they will be somewhat higher.

It is very easy to find the ideal point. The coordinate of the optimal ideal point for each dimension is obtained by multiplying the product coordinates for that dimension by the respective preference weights and summing. For example, the coordinate of the optimal ideal point on the first dimension is:

```
1.225*.687  +  0.000*.187 – 1.225*.126 = .688
```

And the coordinate of the optimal ideal point on the second dimension is:

```
 .000*.687  +  1.225*.187 – 1.225*.126 = .078
```

Notice that on the first dimension the optimal ideal point is only slightly closer to the preferred product, A, than to the second choice, B. This is because there is a substantial amount of preference weight for products B and C, and an ideal point closer to B and C helps to diminish their squared distances. However, we may wish to have ideal points be closer to the preferred product. This can be done easily, simply by rescaling the utilities so as to be larger. For example, if we multiply all the utilities by 2, the preference weight for the preferred product becomes larger with respect to the others, and as a result the ideal point is estimated to be closer to the preferred product. Below are estimates of ideal point locations for this respondent as a result of scaling the utilities by progressively larger factors:

Scale Factor	Ideal Point Estimate I	II
1	.688	.076
2	1.069	.045
4	1.216	.005
8	1.225	.000

Notice that as we scale the utilities by larger positive factors, the estimated ideal point moves closer to product A. If we use a scale factor of 8, the estimated ideal point is equal to location of the preferred location to within 3 decimals. With scaling this large, nearly all the preference weight is on the preferred product, a situation similar to what we call a “first choice model” in conjoint simulations.

This presents an opportunity but also a puzzle. The opportunity is that by choice of scale factor we can vary the amount of attention paid to the first choice product vs. less-preferred products. The puzzle is what scale factor to use. If we wished to have the ideal point coincide with the location of the preferred product, we could achieve that without the bother of an optimization process. It seems that the best choice of the scale factor might depend on various properties of the data set. For that reason, we give the user the capability of choosing a scale factor.

The size of the scale factor seems to have relatively little impact on the appearance of the map, but can affect the accuracy of individual preference predictions. One of the outputs of all CPM mapping modules is a count of the proportion of incorrect order relations between actual and estimated pairwise preferences. For some data sets it seems that larger scale factors produce fewer order violations, and for others intermediate scale factors seem to work better. We suggest that you try two or three different values and see which works best with your data set. The default scale factor is 10.

An interesting property of ideal points estimated this way is "convexity." That is to say, they always lie within the region defined by the individual's perceived product locations. This is because each ideal point is a weighted average of the product locations, where the weights are positive and sum to one.

Composite ideal point maps are computed with the following iterative procedure:

- An initial estimate is made of attribute weights (the "transformation matrix") to transform attribute ratings into spatial coordinates. In a "normal start" this is obtained from a principal component analysis of the aggregated ratings.

- Using those attribute weights, a personal perceptual space is developed for each respondent. Within that space, the optimal ideal point is found. The objective function, the sum of weighted squared distances, is computed for that respondent.

- While these computations are done for each individual, information is also accumulated that can be used to find a modified set of attribute weights that will provide an improved average value of the objective function.

- The ideal points are re-estimated for each individual, using improved perceptual spaces, and information is again accumulated that is used to find a still better set of attribute weights for use in the next iteration.

- Iteration is continued until improvements in the objective function become smaller than a pre-specified convergence limit, or until a limit on the number of iterations is reached.

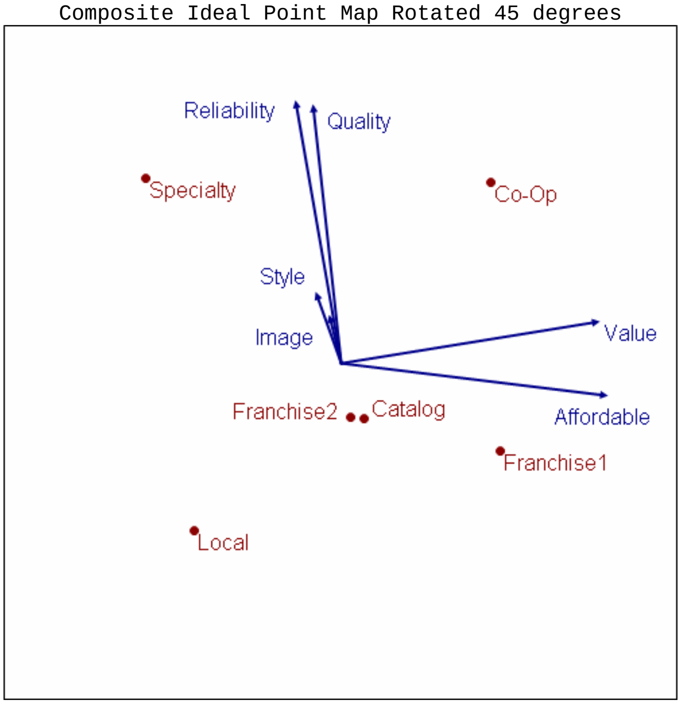
The average objective function decreases with progressively smaller improvements. Although the global maximum is not always reached, solutions obtained usually appear to be quite similar, regardless of starting point.

Minimizing the sum of weighted squared distances seems to work very well, but one trick is required to make it do so. Without constraints of some kind, the algorithm would lead to a solution where all brands and ideal points were at the same location in space, so that all distances would be zero. Accordingly, the dimensions are uncorrelated and we place the same constraint on these maps as we did on composite vector maps: that the sum of squared product coordinates in each dimension must be equal to the number of products.

This iterative process is not quite so well behaved as that of the composite vector model. More iterations are usually required, and occasionally an iteration fails to improve and the procedure terminates, even though the best solution has not been found. For this reason we advocate computing composite ideal point maps from different random starting points, which we have made easy to do.

Interestingly, the two composite maps are nearly identical. That will become even more evident when we look at the graphic display.

Figure 4
Composite Ideal Point Map Rotated 45 degrees



This map is very similar to that for the composite vector model, suggesting that with this data set it doesn't make much difference which model is used.

Composite Ideal Point Maps: Advantages and Disadvantages

Both composite methods have an advantage over discriminant analysis: they are more sure to capture the perceptual dimensions important in determining preferences. They also share a disadvantage: not being based solely on perceptual data, they contain less information about perceived similarities and differences among products. Also, the arbitrary scaling of their dimensions (so that the sum of squared product coordinates is equal to the number of products) may obscure information about product perceptions.

As to differences between the composite vector and ideal point models, the vector model has the advantage of better convergence and the assurance of global optima.

However, the ideal point model is in some ways more general. The vector model may be considered equivalent to an ideal point model in which all the ideal points are restricted to lie in a circle far outside the region containing the products. Our ideal point model restricts the ideal point for any individual to lie within the region defined by his/her product perceptions. The two composite models have also produced quite similar maps for several other data sets we have examined. But we have also found data sets for which they have much less overlap. It seems likely that the ideal point model will be more effective with attributes for which intermediate levels are preferred.

Analyzing Density of Demand: Previous Approaches

Given a product map, it seems reasonable to wonder about relative preferences for products at different locations, or to ask what would happen if a product were introduced at a particular point on the map or repositioned in a certain way. These questions seem particularly reasonable with CPM, since its maps are based on both perceptual and preference information.

CPM's predecessor, APM, offered a simulation module to address such "what if" questions. A perceptual space was estimated for each respondent, containing his/her perceived product locations and explicit ideal point. An analysis was made of the relationships between preferences and distances of products from the ideal point, leading to predictions about how any product's preference should change if it were repositioned. Results were accumulated over respondents to estimate shares of preference for modified products, much in the spirit of conjoint simulators.

However, the APM manual listed several potential problems with this approach. We quote from the APM manual about one of those problems:

> "It would be nice if we could produce an advertising campaign to move a product in a particular direction, and have that be the only effect. However, the map takes into account correlations between attributes. Suppose Price and Performance are highly correlated at present. If we decide to offer high performance at a low price, we may modify this relationship. We may not only move our product, but also change directions of attribute vectors, and perhaps even create an additional dimension."

APM hoped to reduce the severity of this problem by observing that simulations were done at the level of the individual respondent, where attributes should be less strongly correlated than in aggregate. But the problem remains with us today.

To elaborate, our estimation of how respondents should behave depends on current relationships between perceptions and perceptions. The perceptual space is obtained by transforming product's scores on many attributes into scores on a small number of dimensions. If a product changes in a way inconsistent with that transformation, then it will no longer remain in the same space.

For example, suppose we have a perceptual space for which one dimension is based on Price and Performance and those attributes have a strong negative correlation. We have little information about how to model the effect of a change in Performance without a corresponding change in Price. If respondents have previously assumed Price to be a negative indicator of Performance, then changing one without the other could add a new dimension to the space. At the least, modifying a product to have high Performance and low Price would change the correlation between those attributes and the directions of their vectors.

To use the current map as a basis for predicting effects of changes, we must assume that if a product is changed on one attribute, then it should also be changed on other attributes correlated with that one. But if we recognize that limitation, then we are not really in a position to infer the effects of changes on single attributes, only on dimensions. With benefit of hindsight it seems that predicting shares of preference for products modified one attribute at a time is too ambitious a goal for mapping-based analyses. We probably should confine ourselves to analyzing positioning questions in terms of movements on spatial dimensions rather than movements on individual attributes.

Also, in the years since APM's introduction, it has become apparent that conjoint analysis provides a more powerful method for "what if" simulations than mapping-based simulators. Rather than a simulator that tries to predict the result of changing a product on specific attributes (which is better left to conjoint analysis) CPM provides a method for estimating the density of demand at each point in the product space.

Density of Demand: The Basic Idea

With all of CPM's mapping methods it is possible to construct a personal perceptual map for each respondent. The aggregate map is just the average of the individual maps. Since the aggregate map is obtained by applying a set of attribute weights to average ratings, each respondent's personal map is obtained by applying the same attribute weights to his/her own product ratings.

For the composite methods, each respondent's personal map also includes an estimated ideal point or ideal direction. If the data set included explicit ideal product ratings, then each respondent's personal discriminant map likewise includes an estimated ideal point.

CPM estimates the demand for a new product throughout the product space by doing an analysis for each respondent and aggregating the results. The individual analysis consists of introducing a new product at each of many grid points and seeing whether such a product would be closer than any existing product to the respondent's ideal point, or whether it would be "farther out" than any exiting product in the respondent's ideal direction. The computation that is done uses a "first choice" simulation rather than a "share of preference" simulation. It is important to note that CPM's analysis is for hypothetical products ***at each point in the existing space*** rather than for hypothetical products with any conceivable pattern of changes in attributes.

We make two assumptions in doing the density analysis.

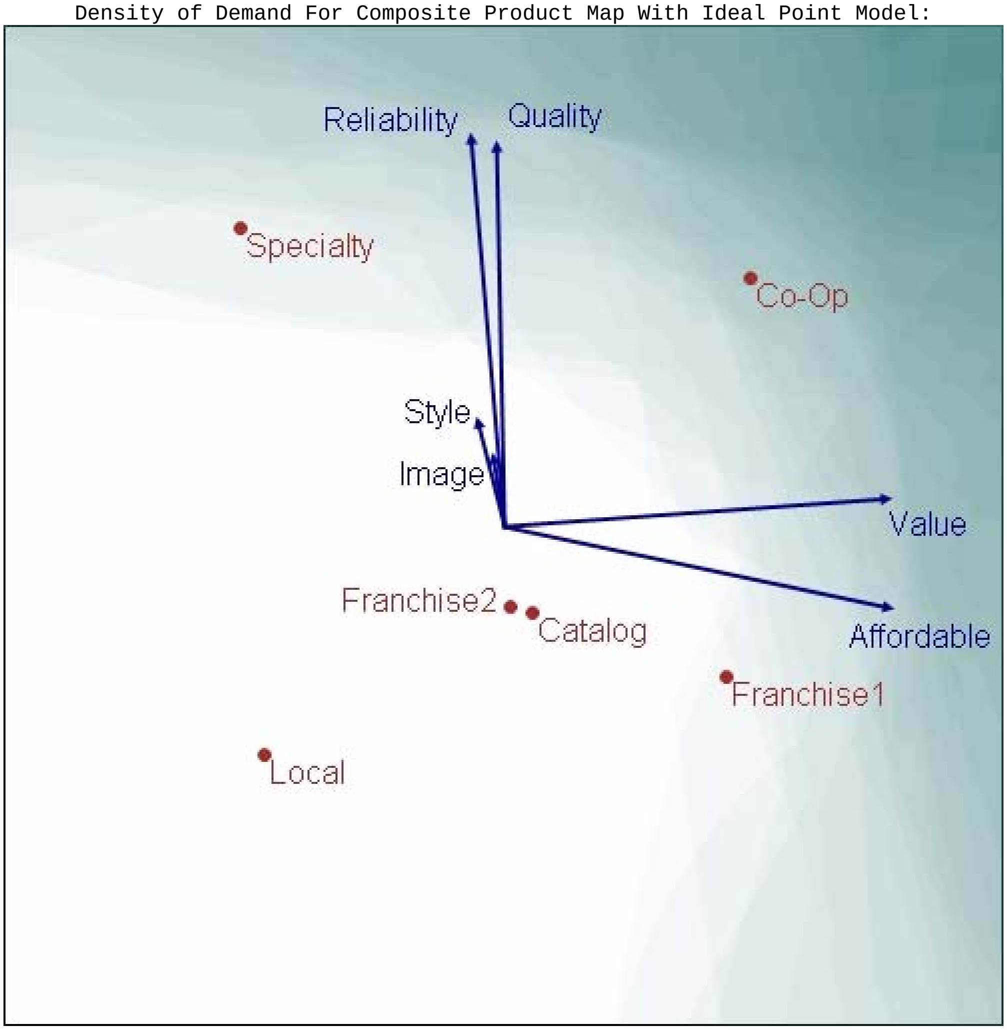
- The first assumption counters problems that arise because people use rating scales differently from one another. We assume that differences among individuals in their mean ratings are "noise" and should be removed from the data. We do this by zero-centering each individual's product ratings for each attribute.

- The second assumption regards differences in amount of variability among products. Because the aggregate map is the average of the individual maps, there is "canceling out" in computing those averages, and as a result products are usually closer together in the aggregate map than they are in most individual maps. To compensate for this we re-scale each individual map so that its variability among products is similar to variability for the aggregate map.

Since we want to estimate demand throughout the space, we must limit the region investigated – otherwise we would have an infinity of points to consider. The Density of Demand analysis provided by the Plot module solves this problem by scaling product coordinates so that, for the product farthest from the center, that distance is 0.8. To facilitate plotting, the attribute vectors are likewise reduced so as to have maximum lengths of 0.8. We explore demand for points that lie within the range of plus and minus one for each dimension. This means that we explore the square centered on zero and bounded by the largest product coordinate, plus an additional 25% of that distance in each direction.

Following is the display for density of demand for the Sample composite ideal point map, after rotation.

Figure 5
Density of Demand For Composite Product Map With Ideal Point Model:



Up to thirty regions are displayed on the screen, each having a degree of density of background color or shading, with intensity representing relative demand. Notice that the Co-Op is in the upper right quadrant. The region of highest demand is also toward that corner of the plot.

Appendix A
Discriminant Maps

In this appendix the usual multiple discriminant solution will be obtained as the solution to the problem of finding the linear combination of variables that maximizes the F ratio of between-to-within variance.

Assume that objects have been described on several attributes by many persons. (We speak of objects rather than products to avoid ambiguity about "product," which we must use to indicate the result of multiplication.) It is not necessary for anyone to describe all objects, but we assume that everyone has described two or more. We pre-process each respondent's data by subtracting his/her mean for each attribute from all ratings on that attribute, so that each individual has attribute means of zero.

Let N = the total number of ratings on each attribute
R = the number of respondents
k = the number of objects rated
p = the number of variables
n_i = the number of ratings of the ith object
$\mathbf{X}$ = the (N x p) data matrix, partitioned as follows:

$$\mathbf{X} = \begin{vmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \mathbf{\ldots} \\ \mathbf{X_k} \end{vmatrix}$$

where $\mathbf{X_i}$ is a submatrix of order n_I x p containing ratings of the ith object.

Consider the matrix $\mathbf{X_B}$ obtained from $\mathbf{X}$ by replacing every row of $\mathbf{X_i}$ with the row of column means of that submatrix, and let $\mathbf{X_W} = \mathbf{X} - \mathbf{X_B}$.

Let $\mathbf{T} = \mathbf{X'X}$
$\mathbf{B} = \mathbf{X_B'X_B}$
$\mathbf{W} = \mathbf{X_W'X_W}$

Then, since $\mathbf{X_B'X_W}$ is zero by construction, $\mathbf{T} = \mathbf{B} + \mathbf{W}$.

$\mathbf{T}$ is a p x p matrix of total sums of squares and cross products, $\mathbf{B}$ is a p x p matrix of sums of squares and cross products between objects, and $\mathbf{W}$ is a p x p matrix of sums of squares and cross products within objects.

The univariate F ratio for the jth variable can be obtained from the diagonal elements of $\mathbf{B}$ and $\mathbf{W}$ by dividing each by its appropriate degrees of freedom, and then forming the ratio of those quotients. The degrees of freedom between products are $k - 1$, and the degrees of freedom within products are $N - k - R$ (Subtraction of R is necessary because we set each respondent's means at zero).

$$F_j = [\mathbf{b_{jj}} / (k - 1)] / [\mathbf{w_{jj}} / (N - k - R)] \qquad (1)$$

Now consider a (p x 1) vector of weights $\mathbf{V}$, which are to be applied to the variables to get a linear combination of them. Let

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\ \mathbf{V} \\ \mathbf{Y_B} &= \mathbf{X_B}\mathbf{V} \\ \mathbf{Y_W} &= \mathbf{X_W}\mathbf{V} \end{aligned} \qquad (2)$$

Where $\mathbf{Y}$, $\mathbf{Y_B}$, and $\mathbf{Y_W}$ are vectors of order N. Their entries are values on a derived variable obtained by weighting the columns of the respective matrices by elements of $\mathbf{V}$ and summing.

The between-objects to within-objects F ratio for this new variable is given by the expression

$$F_Y = [\mathbf{Y_B'Y_B} / (k - 1)] / [\mathbf{Y_W'Y_W} / (N - k - R)]$$

$$= [(N - k - R) / (k - 1)] \ \mathbf{V'BV / V'WV}$$

$$= [(N - k - R) / (k - 1)] \lambda \qquad (3)$$

$$\text{where} \quad \lambda = \mathbf{V'BV / V'WV} \qquad (4)$$

We want that particular set of weights **V** that will maximize λ . The solution is obtained by solving for the partial derivatives of λ with respect to **V** and equating them to zero. This results in the matrix equation:

$$(\mathbf{B} - \lambda \mathbf{W})\mathbf{V} = \mathbf{0}. \qquad (5)$$

If **W** in nonsingular we may write

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda \mathbf{I}) \mathbf{V} = \mathbf{0}. \qquad (6)$$

This is the well-known Eigen equation that is involved in multiple discriminant analysis. The desired weights are elements of the right characteristic vector of $\mathbf{W}^{-1}\mathbf{B}$ corresponding to the largest root, which is related to the F ratio by equation (3).

To maximize F_Y we choose as weights elements of the characteristic vector corresponding to the largest root. The second derived variable providing the highest F ratio among all derived variables while uncorrelated with the first (the second discriminant dimension) is obtained by using as weights elements of the characteristic vector associated with the second largest root, etc.

In practice, we do not solve equation (6) directly for several reasons: $\mathbf{W}^{-1}\mathbf{B}$ is not symmetric, so its characteristic roots and vectors are not as easy to obtain as those of related symmetric matrices that can alternatively be analyzed, and **W** may be singular, in which case $\mathbf{W}^{-1}$ will not exist. Also, as in multiple regression analysis, more stable and useful solutions can often be obtained by dealing with a subset of variables or with factor scores. The following development was first suggested to the author by Ledyard Tucker.

Since **W** = **T** – **B**, we may substitute in equation (5), obtaining

$$[\mathbf{B} - \lambda (\mathbf{T} - \mathbf{B})]\mathbf{V} = \mathbf{0}$$

or

$$(\mathbf{B} - \alpha\mathbf{T})\mathbf{V} = \mathbf{0} \qquad (7)$$

and

$$(\mathbf{T}^{-1}\mathbf{B} - \alpha\mathbf{I})\mathbf{V} = \mathbf{0}.$$

where

$$\alpha = \lambda / (1 + \lambda)$$

and

$$\lambda = \alpha / (1 - \alpha). \qquad (8)$$

The right characteristic vectors of $\mathbf{T}^{-1}\mathbf{B}$ and $\mathbf{W}^{-1}\mathbf{B}$ are therefore identical, and their characteristic roots are related as in (8).

Rather than compute the inverse of **T**, we approximate it with its principal components.

We use the Eckart-Young decomposition (Johnson, 1961) of **X**,

$$\mathbf{X} = \mathbf{P}\,\Delta\,\mathbf{Q'}. \qquad (9)$$

where Δ is a diagonal matrix of positive square roots of the characteristic roots of **T** arranged in order of descending magnitude, **Q** is a matrix whose columns are the corresponding characteristic vectors of **T**, and **P'P** = **Q'Q** = **I**.

It is well known that using the first *r* columns of **P**, **Q**, and Δ results in the least square approximation to **X** of rank *r*. In what follows, assume that we use only those *r* components, performing the discriminant analysis on values in the matrix **P** of principal component factor scores.

The corresponding lower-rank approximation to **T** is obtained by the product $\mathbf{Q}\,\Delta^2\,\mathbf{Q'}$
Substituting that approximation for **T** in (7), and recognizing that **Q'Q** = **I**, we get

$$(\mathbf{B} - \alpha\,\mathbf{Q}\,\Delta^2\,\mathbf{Q'})\mathbf{V} = \mathbf{0}$$

or

$$(\Delta^{-1}\,\mathbf{Q'B}\,\mathbf{Q}\,\Delta^{-1} - \alpha\,\mathbf{I})\,\Delta\mathbf{Q'}\,\mathbf{V} = \mathbf{0}. \qquad (10)$$

Letting

$$\mathbf{S} = \Delta^{-1}\,\mathbf{Q'B}\,\mathbf{Q}\,\Delta^{-1} \qquad (11)$$

and

$$\beta = \Delta\mathbf{Q'}\,\mathbf{V} \qquad (12)$$

we now have the smaller system of equations,

$$(\mathbf{S} - \alpha\,\mathbf{I})\,\beta = \mathbf{0}. \qquad (13)$$

in which **S** is symmetric. The solution is obtained quickly and with good numerical stability.

From (9), $\mathbf{Q}\,\Delta^{-1}$ is the matrix that transforms **X** into **P**, and we have seen that β is the set of weights that transforms **P** into a discriminant dimension. Therefore, **V** may be obtained by the product

$$\mathbf{V} = \mathbf{Q}\,\Delta^{-1}\,\beta. \qquad (14)$$

If all the non-null principal components of **T** are used, then the elements of **V** computed with (14) will be the same as those computed by (6), and the associated F ratio will be the same as in (3). Otherwise the result will contain information about discrimination in the space spanned by the first *r* principal components. In either case, the appropriate F ratio is obtained by the formula

$$F_Y = [(N - k - R) / (k - 1)]\,\alpha / (1 - \alpha). \qquad (15)$$

Since discriminant weights are determined only to within multiplicative constants, their scaling is arbitrary. In mapping, however, it is desirable that their magnitudes relate meaningfully to their discrimination among objects. This is accomplished by scaling them so that the within-objects sum of squares for each discriminant dimension is equal to the same constant. One way of doing this conveniently is by scaling the solution so that the sum of squares between objects for each discriminant dimension is equal to its F ratio. We further scale all discriminant weights uniformly so that the sum of squared coordinates of objects on the first dimension is equal to the number of products.

Appendix B
Composite Product Mapping

The vector model and the ideal point model both use the following definitions:

Let there be ***N*** respondents, ***n*** products, ***p*** attributes, ***d*** dimensions. For the i-th individual, let:

$\mathbf{X_i}$ = an (n x p) matrix of product ratings, with column sums of zero, and $\mathbf{X} = {}^{1}/_{N} \sum \mathbf{X_i}$

$\mathbf{Y_i}$ = an (n) vector of product preference values. APM provides constant-sum preference information obtained by having respondents divide 100 points among members of each of several product pairs. We construct $\mathbf{Y_i}$ by taking logits of those preference percentages, and awarding half the logit value to the winning product in that pair, and penalizing the losing product with half of the logit value. Each value of $\mathbf{Y_i}$ is divided by the count of the number of values accumulated, resulting in $\mathbf{Y_i}$ values similar to conjoint utilities that sum to zero for each individual.

$\mathbf{T}$ = a (p x d) matrix of attribute weights used to transform attribute ratings into dimensional coordinates. $\mathbf{T}$ is common to all respondents. We want to find a $\mathbf{T}$ that permits the best fit to all respondents' preferences. We start with an approximation obtained from a principal components analysis of attribute ratings, and then improve it iteratively.

$\mathbf{C_i}$ = an (n x d) matrix giving the configuration of products for the i-th individual.

$$\mathbf{C_i} = \mathbf{X_i}\,\mathbf{T} \qquad (1)$$

The Vector Model

In the vector model each individual is thought of as having an ideal direction in the space, represented by a vector, and should prefer products according to their projections onto that vector.

Let $\mathbf{W_i}$ = a (d) vector of importance weights to be applied to columns of the individual's configuration to best predict that individual's preferences. Our basic individual preference equation is

$$\mathbf{C_i}\ \mathbf{W_i} - \mathbf{Y_i} = \mathbf{E_i} \qquad (2)$$

where $\mathbf{E_i}$ is a vector of errors of fit. Equation 2 says that the individual's configuration of products in space $\mathbf{C_i}$ is weighted by the elements of $\mathbf{W_i}$ to get a prediction of $\mathbf{Y_i}$.

Substituting from (1) into (2), we get

$$\mathbf{X_i}\ \mathbf{T}\,\mathbf{W_i} = \mathbf{Y_i} + \mathbf{E_i} \qquad (3)$$

If we knew $\mathbf{T}$, we could solve for $\mathbf{W_i}$, using ordinary least squares. We start with an initial approximation of $\mathbf{T}$ and improve it in subsequent iterations. After estimating a $\mathbf{W_i}$ for each individual, we then combine information from all individuals to find a $\mathbf{T}$ that fits individuals better on average. By alternating between re-estimation of the **W**'s and **T**, we eventually find estimates for the **W**'s and **T** that best fit the data.

An initial estimate of $\mathbf{T}$ is obtained either from random numbers or from the principal components of $\mathbf{X}$, the sum of all individuals' $\mathbf{X_i}$ matrices. In each iteration, we solve for weights for each individual using ordinary least squares, minimizing the sum of squared errors in $\mathbf{E_i}$:

$$\hat{\mathbf{W}}_\mathbf{i} = (\mathbf{T'}\,\mathbf{X_i'}\mathbf{X}\mathbf{T})^{-1}\mathbf{T'}\,\mathbf{X_i'}\mathbf{Y_i} \qquad (4)$$

We also record the r-square for each individual as a measure of how well the model fits that individual's preferences.

To produce an improved estimate of **T**, we could use the partial derivatives of the sum of squared errors with respect to **T**. For the i-th individual, those partials are given in equation (5):

$$\frac{\partial(\mathbf{E}_i'\mathbf{E}_i)}{\partial\mathbf{T}'} = 2\mathbf{X}_i'\mathbf{X}_i\mathbf{T}\mathbf{W}_i\mathbf{W}_i' - 2\mathbf{X}_i'\mathbf{Y}_i\mathbf{W}_i' \qquad (5)$$

The partial derivatives of the total sum of squared errors involves summing equation (5) over all respondents. Setting the partial derivatives of that sum to zero yields an expression for **T** that minimizes the sum of squared errors.

$$\sum_i^N \mathbf{X}_i'\mathbf{X}_i\mathbf{T}\mathbf{W}_i\mathbf{W}_i' - \mathbf{X}_i'\mathbf{Y}_i\mathbf{W}_i' = \mathbf{0} \qquad (6)$$

However, this equation appears to be intractable. In each term of the sum, **T** is premultiplied by $\mathbf{X}_i'\mathbf{X}_i$ and postmultiplied by $\mathbf{W}_i\mathbf{W}_i'$, so it is not clear how to solve for **T**.

One possibility would be separately to sum the products $\mathbf{X}_i'\mathbf{X}_i$, $\mathbf{W}_i\mathbf{W}_i'$, and $\mathbf{X}_i'\mathbf{Y}_i\mathbf{W}_i'$ and then to premultiply the sum of $\mathbf{X}_i'\mathbf{Y}_i\mathbf{W}_i'$ by the inverse of the sum of $\mathbf{X}_i'\mathbf{X}_i$ and postmultiply by the inverse of the sum of $\mathbf{W}_i\mathbf{W}_i'$. We have tried that, but find that the sum of squared errors obtained with that approximation does not decrease monotonically from iteration to iteration.

However, we have had success with a slightly different procedure. In equation (3) the weights $\mathbf{W}_i$ are applied to the product $\mathbf{X}_i\,\mathbf{T}$ to predict $\mathbf{Y}_i$, but it is also true that the weights $\mathbf{T}\mathbf{W}_i$ are applied to the matrix of attribute ratings, $\mathbf{X}_i$, to predict $\mathbf{Y}_i$. We may estimate weights ($\mathbf{T}\,\mathbf{W}_i$) by premultiplying equation (3) by $(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'$:

$$\mathbf{T}\,\mathbf{W}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{Y}_i + (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{E}_i \qquad (7)$$

The first term on the right hand side of (7) is the estimate of regression weights that would be obtained in trying to predict $\mathbf{Y}_i$ from the individual's ***entire*** set of attribute ratings, $\mathbf{X}_i$. The second term on the right hand side is an error term that we hope is small. If it were zero, then equation (7) would state that the individual's weights for predicting his/her preferences in his/her entire attribute space would be expressible as a weighted combination of the columns of **T**. If that were true, then there would be no loss of predictive ability from using a subspace of small dimensionality common to all respondents. Our method of estimation improves **T** in each iteration so as to minimize the sum of squares of the last term in equation (7).

To show this more clearly we define

$$\mathbf{V}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{Y}_i \qquad (8)$$

$$\mathbf{F}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{E}_i \qquad (9)$$

$\mathbf{V}_i$ is the vector of weights that would best predict $\mathbf{Y}_i$ from $\mathbf{X}_i$. Consider a new regression computation for each individual, fitting $\mathbf{V}_i$ as a weighted combination of the columns of **T.** Equation (10) is obtained by substituting from (8) and (9) into (7). Since the errors are different, (**F** rather than **E**), the estimated coefficients will be different as well. Call these coefficients $\mathbf{U}_i$ (rather than $\mathbf{W}_i$).

$$\mathbf{T\,U_i} = \mathbf{V_i} + \mathbf{F_i} \qquad (10)$$

The OLS solution for $\mathbf{U_i}$ is:

$$\mathbf{U_i} = (\mathbf{T'T})^{-1}\,\mathbf{T'\,V_i} \qquad (11)$$

Assemble the $\mathbf{U_i}$ = and $\mathbf{V_i}$ vectors as columns of matrices **U** and **V**:

$$\mathbf{U} = (\mathbf{U_1}\,,\,\mathbf{U_2},\ldots\mathbf{U_N})$$

$$\mathbf{V} = (\mathbf{V_1}\,,\,\mathbf{V_2},\ldots\mathbf{V_N})$$

Then the OLS estimate of **T** that best fits equation (10) for all individuals is:

$$\hat{\mathbf{T}} = \mathbf{VU'(UU')^{-1}} \qquad (12)$$

Each iteration consists of two steps. During the first step, individual r-squares are computed using the regression indicated in equation (3) to measure the goodness of fit to each individual's preferences with the current estimate of **T**. A second r-square value is also computed for each individual using equation (11), to indicate how successfully the individual's preference structure is captured by the **T** matrix. The $\mathbf{U_i}$ and $\mathbf{V_i}$ vectors are also saved for each individual, as determined in equations (8) and (11).

In the second step, the $\mathbf{U_i}$ and $\mathbf{V_i}$ vectors are assembled and used as in equation (12) to re-estimate **T**.

Because the same **T** is used for all individuals, the aggregate product configuration is obtained by averaging the individual $\mathbf{C_i}$ matrices. Since the goodness of fit to individual data is not affected by a linear transformation of the columns of **T**, we also adjust **T** in each iteration so that the columns of the aggregate configuration are orthogonal and each has sum of squares of *n*.

As iterations progress, the sum of goodness-of-fit r-squares from the regression of equation (3) tends to increase, but it is not required to do so and it often fluctuates in later iterations. However, the sum of "preference structure capture" r-squares from the regression of equation (10) increases monotonically, and iterations are terminated when the increases fall to less than a small positive value.

There are some numerical problems that must be overcome. Some respondents produce little or no useful information in their attribute ratings, and there may be more attributes than products. In either case the inverse of $\mathbf{X_i'X_i}$ will not exist. Therefore all matrices to be inverted first have a small positive amount added to their diagonal elements. This "ridge regression" trick remedies both problems.

The individual weights $\mathbf{W_i}$ are saved in a file, together with the goodness-of-fit r-square value for that individual. Coordinates of a map are provided, showing the average positions of products and attribute vectors in a space. The product configuration is just the average of individuals' configurations. The attribute vector positions are indicated by correlations between average product attribute ratings and product coordinates on each dimension.

The Ideal Point Model

For convenience, we repeat definitions stated above. For the i-th individual, let:

$\mathbf{X_i}$ = an (n x p) matrix of product ratings, with column sums of zero, and $\mathbf{X} = {}^{1}/_{N} \sum \mathbf{X_i}$

$\mathbf{Y_i}$ = an (*n*) vector of product preference values. APM provides constant-sum preference information obtained by having respondents divide 100 points among members of each of several product pairs. We construct $\mathbf{Y_i}$ by taking logits of those preference percentages, and awarding half the logit value to the winning product in that pair, and penalizing the losing product with half of the logit value. Each value of $\mathbf{Y_i}$ is divided by the count of the number of values accumulated, resulting in $\mathbf{Y_i}$ values similar to conjoint utilities that sum to zero for each individual.

$\mathbf{T}$ = a (*p* x *d*) matrix of weights used to transform attribute ratings into dimensional coordinates. $\mathbf{T}$ is common to all respondents. We want to find a $\mathbf{T}$ that permits the best fit to all respondent's preferences. We start with an approximation obtained from a principal components analysis of attribute ratings, and then improve it iteratively.

$\mathbf{C_i}$ = an (*n* x *d*) matrix giving the configuration of products for the i-th individual.

$$\mathbf{C_i} = \mathbf{X_i}\,\mathbf{T} \qquad (1)$$

We start by exponentiating and then percentaging each individual's $\mathbf{Y_i}$ values to get a set of positive values that sum to unity, similar to "shares of preference" in conjoint analysis.
Call this vector of preference information $\mathbf{R_i}$.

We assume each individual to have an "ideal point" in his/her perceptual space, defined as the row vector $\mathbf{P_i'}$. The squared distance from each product to that ideal point is obtained by subtracting $\mathbf{P_i'}$ from each row of $\mathbf{C_i}$ and then summing the squared differences. Call the vector of squared distances $\mathbf{\Delta_i^2}$.

We expect that the distances should be small for products most preferred, and larger for products less preferred. We can express this desired relationship between preferences and distances in terms of the sum over products of the individual's preference weights times his/her squared distances, which we wish were small. Call this value for the ith individual $\theta_\mathbf{i}$.

$$\theta_\mathbf{i} = \mathbf{R_i'}\,\mathbf{\Delta_i^2} \qquad (13)$$

If $\theta_\mathbf{i}$ is small, then for the ith individual the products with large preferences must have small distances from the ideal. Our goal is to find an ideal point for each respondent ($\mathbf{P_i'}$) and a matrix of weights common to all respondent ($\mathbf{T}$) that minimize the sum of the θ values for all respondents:

$$\theta = \sum \theta_\mathbf{i} \qquad (14)$$

To estimate the ideal point for the ith respondent, we differentiate $\theta_\mathbf{i}$ with respect to $\mathbf{P_i'}$ and set the result to zero. Observing that the sum of the $\mathbf{R}$'s is unity, we get the equation:

$$\mathbf{P_i'} = \mathbf{R_i'}\,\mathbf{C_i} \qquad (15)$$

The estimate of the individual's ideal point ($\mathbf{P_i'}$) that minimizes $\theta_\mathbf{i}$ is simply the weighted average of the rows of his/her matrix of perceived product locations, where the weights are the $\mathbf{R_i}$ values. Recall that the $\mathbf{R_i}$ values are positive and sum to unity. If the respondent has such extreme preference for one product that its value is unity and the rest are all zero, then the ideal point will be estimated to be coincident with that product's location. If the respondent is indifferent among products so that all $\mathbf{R_i}$ values are equal, then the ideal point will be estimated to be at the center of the perceptual space. No matter what the respondent's preferences, ideal points estimated this way will always lie within the convex hull of that respondent's perceived product locations.

Given an estimate of the ideal points for each individual, an improved estimate of $\mathbf{T}$ can be obtained as follows. Let $\mathbf{D_i}$ be a diagonal matrix whose diagonal elements are corresponding elements of $\mathbf{R_i}$. Then, differentiating θ with respect to $\mathbf{T}$ and setting the partial derivatives to zero and summing over respondents gives the equation:

$$\sum \mathbf{X_i' D_i X_i \; T} = \sum \mathbf{X_i' R_i P_i'} \qquad (16)$$

It would seem that one way to estimate T would be to cumulate the two sums

$$\mathbf{A} = \sum \mathbf{X_i' D_i X_i} \qquad (17)$$

and

$$\mathbf{B} = \sum \mathbf{X_i' R_i P_i'} \qquad (18)$$

so that

$$\mathbf{A\;T = B}$$

and then simply estimate **T** as $\mathbf{A}^{-1}\mathbf{B}$.

However, the problem with this approach is that θ is minimized trivially by a **T** of zero, and an iterative process that estimates **T** in this way eventually converges to a **T** of zero. To avoid that, it is necessary to impose constraints on **T**. We choose to make columns of the overall configuration **XT** orthogonal, and for each column to have sum of squares equal to the number of products, ***n***. (Recall that **X** is the average of the $\mathbf{X_i}$ matrices.)

This is done with a symmetric matrix of Lagrange multipliers, following Schonemann (1965). We differentiate the sum of $\theta + \phi$ with respect to **T**, where

$$\phi = \mathrm{trace}(\mathbf{S\;T'\;X'\;X\;T}) \qquad (19)$$

with **S** an unknown symmetric matrix.

Setting the sum of partial derivatives to zero yields:

$$\mathbf{AT = B + X'X\;T\;S} \qquad (20)$$

Premultiplying by **T’** and recalling that **T’X’X T** is equal the identity matrix, we get:

$$\mathbf{S = T'(\,AT - B)} \qquad (21)$$

Premultiplying (20) by the inverse of **A**,

$$\mathbf{T = A^{-1}B + A^{-1}\,X'X\;T\;S} \qquad (22)$$

Equation (20) cannot be solved explicitly for **T**, but does submit to an iterative solution consisting of the following steps:

1) Use the value of **T** from the previous iteration or some initial value to compute **S** as in equation (21). Early estimates of **S** will not be symmetric, so force symmetry by averaging corresponding elements above and below the diagonal.

2) Obtain the product $\mathbf{A^{-1}\,X'X\;T\;S}$, using current estimates of **T** and **S**.

3) Determine a scalar α by which to multiply the product obtained in step 2) so that $\mathbf{XA^{-1}B} + \alpha\mathbf{XA^{-1}\,X'X\;T\;S}$ has sum of squares equal to $\boldsymbol{n\,d}$.

4) Use the sum of terms: $\mathbf{A}^{-1}\mathbf{B} + \alpha\mathbf{A}^{-1}\,\mathbf{X'X\;T\;S}$ as an interim estimate of **T** to obtain an interim (not necessarily orthogonal) estimate of the configuration of products in space, **XT**.

5) Find the matrix with orthogonal columns and column sums of squares equal to ***n*** which is closest in the least squares sense to **XT**, using the procedure of Johnson (1966), as well as the right-hand transformation matrix that performs that orthogonalization.

6) Finally, postmultiply the interim estimate of **T** from step 4) by the transformation matrix determined in step 5) to get the estimate of **T** for the current iteration.

Repeat steps 1-6 until estimates of **T** stabilize.

References

Coombs, Clyde H. (1964) *A Theory of Data*. New York: Wiley.

Johnson, Richard M. (1961) "On A Theorem Stated by Eckart and Young," Psychometrika, 28, 259-63.

Johnson, Richard M. (1966) "The Minimal Transformation to Orthonormality," Psychometrika, 31, March, 61-66.

Schonemann, Peter H. (1965) "On the Formal Differentiation of Traces and Determinants," Researach Memorandum No. 27, The Psychometric Laboratory, University of North Carolina.

Tucker, Ledyard R and Samuel Messick (1964) "An Individual Differences Model of Multi-Dimensional Scaling," *Psychometrika*, 333-367.