GRAPHICAL METHODS FOR CIRCULAR AND CYLINDRICAL DATA

Christine M. Anderson

IIQP Research Report
RR-93-08

September 1993

GRAPHICAL METHODS FOR CIRCULAR AND CYLINDRICAL DATA

Christine M. Anderson[1]
Department of Statistics and Actuarial Science
University of Western Ontario, London, Ontario, Canada

September 1993

ABSTRACT

Clear and concise presentation of a directional response can be helpful to gain better understanding and insight into the characteristics of the data. Several new graphical techniques for circular data are described, including an estimated density, the circular boxplot and an interaction plot. For cylindrical data, a combined summary of directional and linear components is illustrated.

Key words: Directional data, circular boxplot, interaction plot, graphical methods.

---

[1] Christine M. Anderson is an Assistant Professor in the Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario. The author wishes to thank C.F.J. Wu and R.W. Oldford for their input into the development of these graphical methods. This work is part of a Ph.D. Thesis from the Department of Statistics and Actuarial Science, University of Waterloo, under the supervision of C.F. Jeff Wu. Research supported by the National Sciences and Engineering Research Council of Canada, General Motors of Canada and the Manufacturing Research Corporation of Ontario.

1 Introduction

In industrial settings, the outcomes of experiments sometimes yield a response which is not the usual continuous linear response (such as a length or weight), but rather a measurement that can best be described by a direction or angle. For example in the automotive industry, a number of rotating parts (such as wheels, brake rotors and flywheels) are produced and need to be precisely balanced. Quantifying the imbalance results in two measurements: the angle at which an adjustment needs to be made and the size of the corrective weight required. Another of other applications for angular data also exist, including the assessment of proximity to a target. If a hole or marking needs to be placed on a production item, the accuracy of the placement is sometimes best measured by the usual x- and y-coordinate system, while in other cases the results of the study are better described by the angle in which the target was missed, and by how much it was missed. Finally, the location of surface defects on circular parts can be summarized by an angular measure.

To gain better understanding of processes which yield a directional response, graphical techniques can be invaluable for concisely highlighting key features of the data. In this paper, some of the existing methods for displaying circular data are shown, such as the dot diagram, rose diagram and circular histogram. In addition, new data summaries are described. The first graphical method is a smoothed estimate of the distribution function based on the frequency of observations in different sectors of the circle. The second method is a circular analog for the boxplot, or box-and-whiskers plot, is described. Finally, if the interaction between two or more factors in a factorial experiment is of interest, a plot can highlight the additive or non-additive nature of factors on a response.

In addition, a new data summary is described for situations where the response from an experiment yields a bivariate response: a directional component and a linear component indicating the magnitude of response. The first two examples listed above (balancing rotating parts and proximity to a target) are both of this type. In addition, this type of response is not restricted to the industrial setting. For example in biology, wind direction and velocity are commonly studied jointly. Batschelet (1981, p.191) gives a number of other examples.

Hence a number of directional applications exist where graphical tools can help display the data to better visualize the process.

2 Circular Graphical Methods

Both Mardia (1972) and Batschelet (1981) describe a number of techniques for graphically displaying directional data. The dot diagram, in Figure 1, illustrates how data measured with limited precision can be displayed. Frequently, circular data are only measured to the nearest 5°, 10° or 15°, and hence can usefully be displayed by presenting the number of observations at each observable location. The dots can be placed inside (see Batschelet 1981, p.23) or outside the circle, with the latter having the advantage of being able to accommodate larger data sets.

Figures 2, 3 and 4 illustrate the same industrial data set involving the location of flywheel imbalances. The study to determine which factors influence the location of part imbalances was a 2^4 full factorial experiment with 10 replicates per cell. The figures summarize the 160 observations that resulted. For a more complete quantitative analysis of this data set see Anderson (1993). The rose diagram (Mardia 1972, p.5), shown in Figure 2, is constructed by dividing the circle into equal sections and making the length of each "petal" proportional to the number of data points in that section. It is generally a poor data summary, since although the length of each petal is proportional to the observed frequency, the area is not. This leads to a distorted emphasis on higher frequency cells. An alternative is to make the total area of each petal proportional to the observed frequency, but as Cleveland (1985, pp.243-245) notes that the human eye introduces bias to area comparisons according to Steven's Law.

The circular histogram in Figure 3, is a superior alternative to the rose diagram in that the length and area of each bar is proportional to the number of observed values in that section of the circle. It may be helpful to add several concentric circles around the histogram to give an indication of the number of observations in each sector (see Mardia 1972, p.4). Both the rose diagram and the circular histogram differ from the dot diagram in that they group the data into user selected sectors for presentation, while the dot diagram is presenting the actual observed values from limited precision data.

If there is interest in examining an estimate of the density function, one can be constructed by smoothing the results from the circular histogram using a running median on the cell frequencies. Twicing, a process which smoothes the residuals of the first running median and adds them to the original estimate, is used to further smooth the estimated density (Tukey, 1977, p.526). Alternately, kernel smoothers, such as the box, triangular or normal (Hastie and Tibshirani, 1990, Ch. 3) could be used to obtain other density estimates. The selection of the range of the data is arbitrary. Common choices are 0° to 360°, or -180° to 180°. However, the selection of the range does not alter the inherent connection between the two end points. Hence for our density estimate, we extend the range of the data to include additional cells on each end to take this association into account. For example using the range 0° to 360°, if the cell sizes used are 10° increments, we would extend the range to $(-30°, 390°)$ in order to have a continuous estimate of the density at 0°. Since this method does not guarantee a density estimate of the correct size an adjustment is performed to preserve the consistency of the estimate. We calculate the area under the curve and multiply our estimate by a constant factor to ensure the density has the right size. The line in Figure 3 illustrates this method. This density estimate can be useful both in isolation or in combination with the circular histogram to gain a better understanding of the overall shape of the distribution.

However, the previous graphical methods are not ideal for making comparisons among different observed samples as it is hard to discern the overall attributes of each data set. A new graphical data summary suitable for unimodal data, developed as an analogous summary to the box-and-whisker plot commonly used for linear data, is the circular boxplot shown in Figure 4. The data used in this diagram is again the automotive flywheel data described previously. The circular boxplot provides a five-quantile summary of the data, which makes quick comparisons of location, dispersion and symmetry possible. In the centre of the shaded wedge, is a radius marking the median direction. For small sample situations where the median, as previously defined is not unique, the average direction of the candidates for median is taken, giving a reasonable approximation to the middle of the data. The shaded region encompasses the central 50% of the data, 25% on either side of the median. Finally, the curved "whiskers" map out the central 90% of the data. In this way, the eye is able to determine quickly the overall direction of the data, a measure of its spread, as well as a feeling for the skewness of the distribution. In addition to the graphical summary of the data, the angles, measured in degrees, for each of the five quantiles are reported. In Figure 4, the range selected was -180° to 180°.

A final and more specialized graphical method for circular data is the directional interaction plot. It is useful for summarizing the influence of two factors from a factorial experiment on an angular response. It is analogous to the interaction plot commonly used for continuous linear data, where the notion of no interaction is well understood to involve additivity of two or more main effects. Graphically, this is interpreted as parallel lines for a fixed level of one factor as we move from low to high of another factor, as shown in Figure 5. To translate this notion of additivity to directional data, we consider a two-way experiment with two levels for each of the factors, A and B. We assume that changing the levels of either of the factors will result in an angular rotation on the circle (clockwise being negative, counter-clockwise as positive). Here we define no interaction effect to occur when the effect of the two factors is additive in angular rotations for all combinations of the levels of the two factors. Figure 6 illustrates this situation for the 2^2 factorial case. In this plot, the four radial lines describe the circular averages (frequently called the mean direction in the literature) of the groups considered for the interaction plot, namely A low B low, A low B high, A high B low, and A high B high. The arcs outside the circle summarize the rotation from changing the level of one factor while the other factor is held constant first at its low level and then at the high level. If no interaction is present, the plot will show that the magnitude and angular direction of rotation will be the same. Figure 7 gives an example of a non-zero interaction where the size and direction of the rotations are different for the two levels of factor A. The interaction plot is helpful for gaining a better understanding of the simultaneous effect of several factors on a circular response.

3 A Cylindrical Plot

Frequently, a circular response, θ_i, has a corresponding linear attribute, x_i, as in the case of wind direction and velocity, or direction and size of the imbalance of a rotating part. Such bivariate responses are sometimes called cylindrical data since we can think of the data lying on a cylinder with radius one and the length of the cylinder corresponding to the linear measurement. In these situations it is sometimes desirable to determine the relationship connecting these two responses. However, if we attempt to plot the data in cylindrical form, we encounter problems in that the data would need to be displayed in three dimensions. The plot in Figure 8 shows how this problem can be circumvented and both of the attributes can be combined in a single plot. The inner circle represents the zero value of the linear component, and hence, the radial length from the inner circle to the data point is proportional to the linear response. The relative size of the inner circle can be adjusted to enhance the appearance of the graph. If the inner circle is reduced to radius zero, the plot simply becomes a scatter of plot of $x_i \cos\theta_i$ versus $x_i \sin\theta_i$. However, the key aspect of the data is the connection between the observed data and the origin. If the inner circle's radius is reduced to zero, there is nothing to emphasize the origin, and the human eye is naturally drawn to the centre of the cloud of points. In this case, the connection between the linear and angular components is difficult to identify. Hence the inner circle should be large enough to allow the nature of the data to be illustrated, but small enough that the data is not obscured by its prominence. This author has found that setting the radius of the inner circle equal to one third to one sixth of the total range of the linear data gives a satisfactory graph. The direction is incorporated into the plot by plotting the points along the ray corresponding to the vector through the origin in the direction of the response. Concentric circles are added to give scale to the linear measurement. In Figure 8, the units of the linear measurement, flywheel imbalance, are pound-inches. An alternative, to using equally spaced circles to indicate the scale, is to plot three circles with radii selected to reflect the lower and higher quartiles as well as the median of the linear data. Labels for the quartiles and their magnitude will further enhance the plot. This choices provides some additional information about the spread of this component.

In some situations, actual meaning can be given to the coordinates obtained by the plot. For example, if we consider the balancing of rotating parts, we can think of either placing a variable sized weight on the circumference of the circle, or a unit sized weight at different distances away from the centre of the part. In this situation, since the force exerted by a weight is proportional to either the distance from the radius or the size of the weight, the graph can be interpreted in two possible ways. In practice, the correction is made by attaching different sized weights to the circumference of the part. If a relationship exists between the radius and some function of the weight, it is best to plot the function of the weight. For example, if the square of the linear measure is physically related to the radius, then the plot should reflect this relation.

4 Conclusions

Graphical tools for studying an angular response can be valuable for gaining a better understanding of the underlying process and the nature of the data. The smoothed estimate of the density gives a visual way of examining the overall shape of the observed spread of the data. The circular boxplot provides a quick five-data point summary of the location, dispersion and skewness of the data. The more specialized directional interaction plot can illustrate the influence of more that one factor on a circular response from a factorial experiment. Finally, the cylindrical data plot allows a two-dimensional summary of the relationship between the angular measure and its associated linear component. These graphical plots can assist in the clear presentation of directional data.

References

[1] ANDERSON, C.M. (1993), "Location and Dispersion Analyses for Factorial Experiments with Directional Data," Ph.D. Thesis, Department of Statistics and Actuarial Science, University of Waterloo.

[2] BATSCHELET, E. (1981), *Circular Statistics in Biology,* London: Academic Press.

[3] BEST, D.J. and FISHER, N.I. (1979), "Efficient Simulation of the von Mises Distribution," *Appl. Statist.*, 28, 152-157.

[4] CLEVELAND, W.S. (1985), *The Elements of Graphing Data,* Monterey: Wadsworth Advanced Books and Software.

[5] HASTIE, T.J. & TIBSHIRANI, R.J. (1990), *Generalized Additive Models,* London: Chapman and Hall.

[6] MARDIA, K.V. (1972), *Statistics of Directional Data,* London: Academic Press.

[7] TUKEY, J.W. (1977), *Exploratory Data Analysis,* Reading: Addison-Wesley Publishing Company.

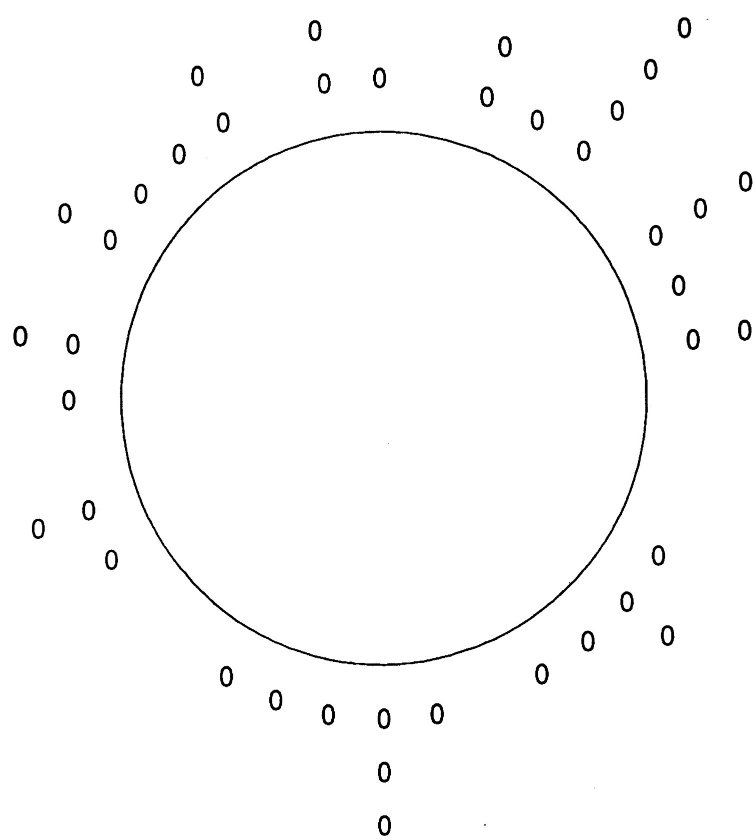

Figure 1: Dot Diagram

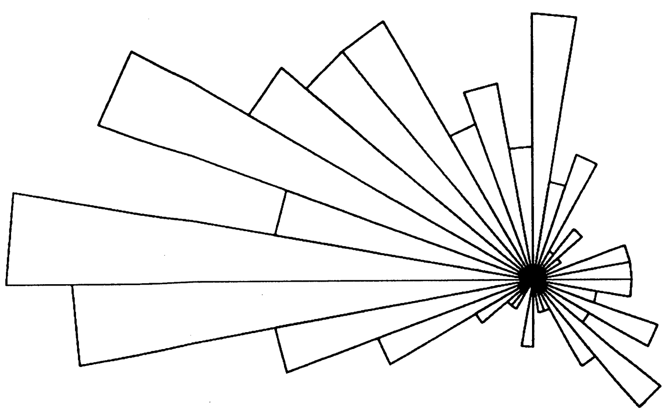

Figure 2: Rose Diagram

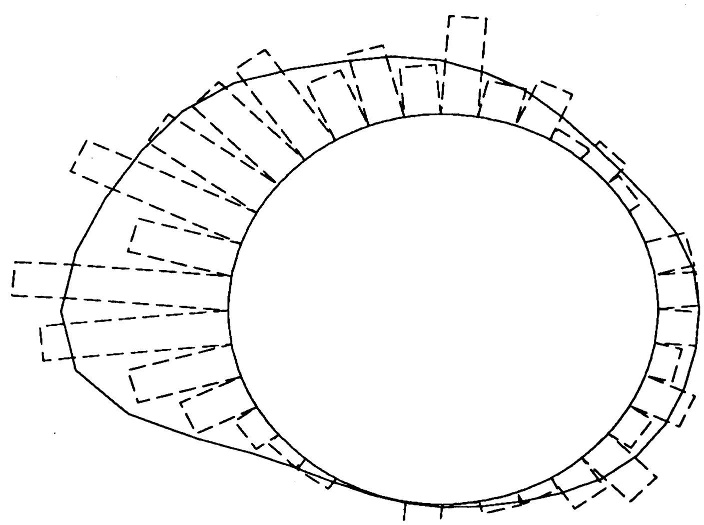

Figure 3: Circular Histogram with Density Estimate

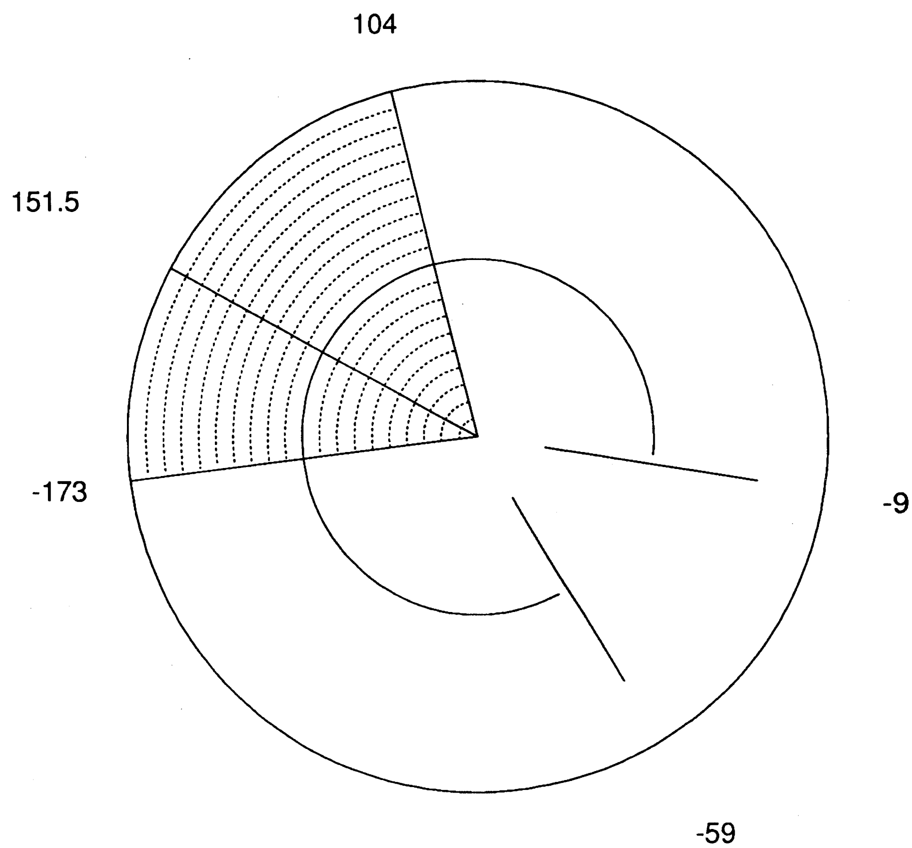

Figure 4: Circular Boxplot

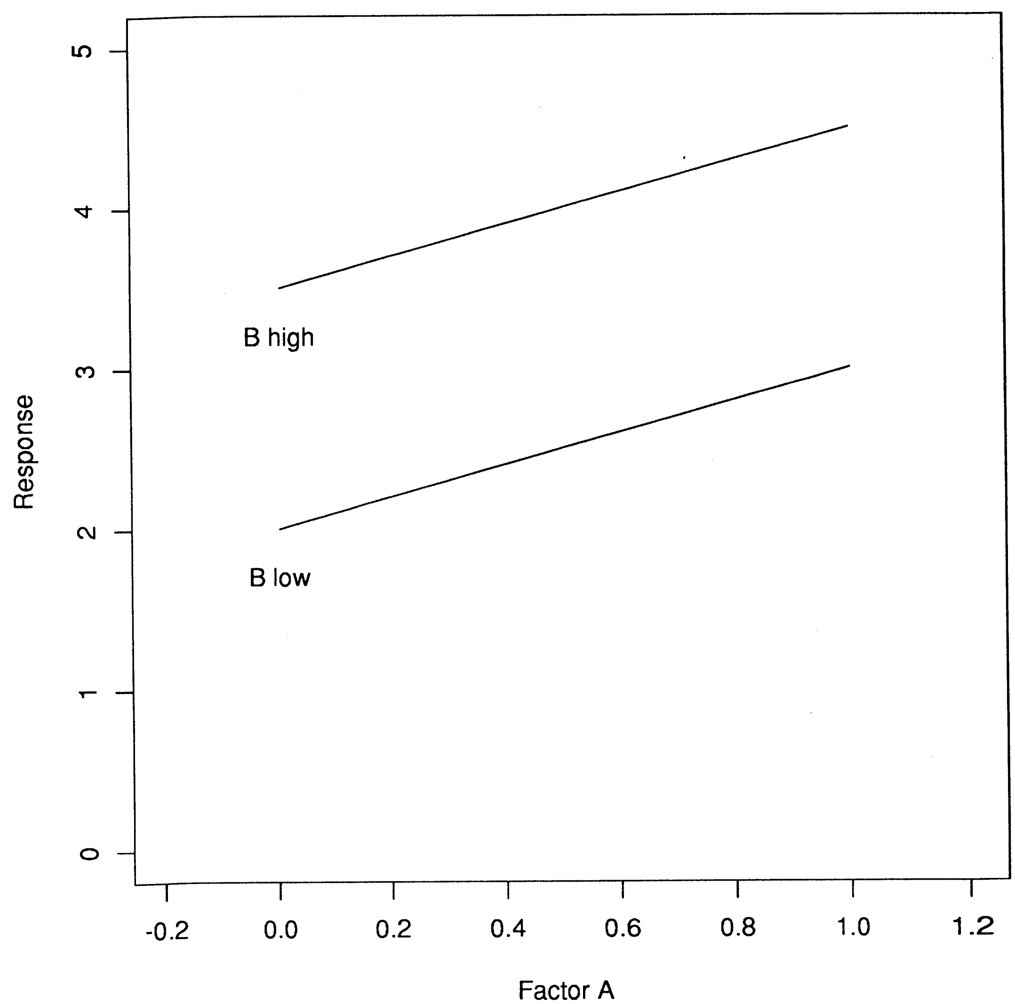

Figure 5: Example of No Interaction for Linear Data

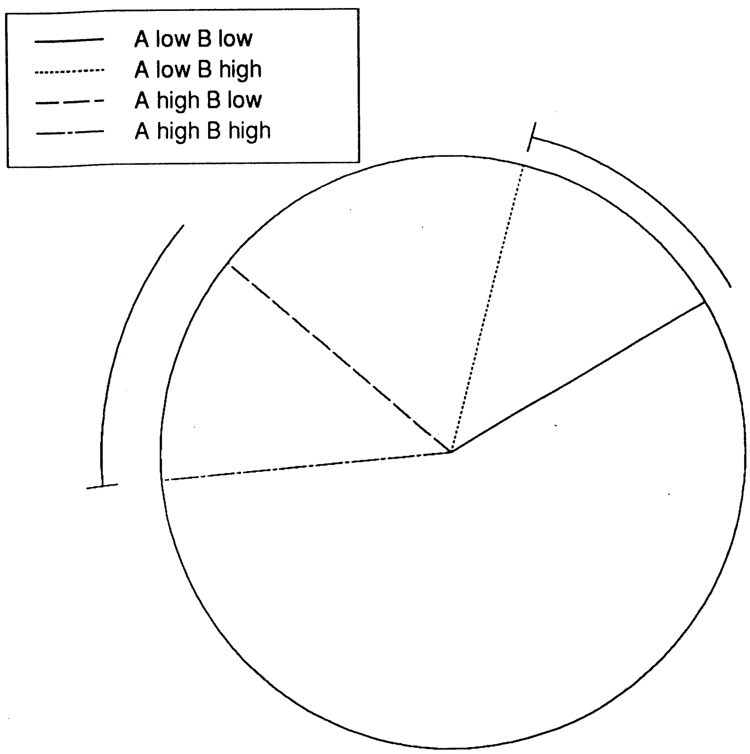

Figure 6: Zero Interaction for Directional Data

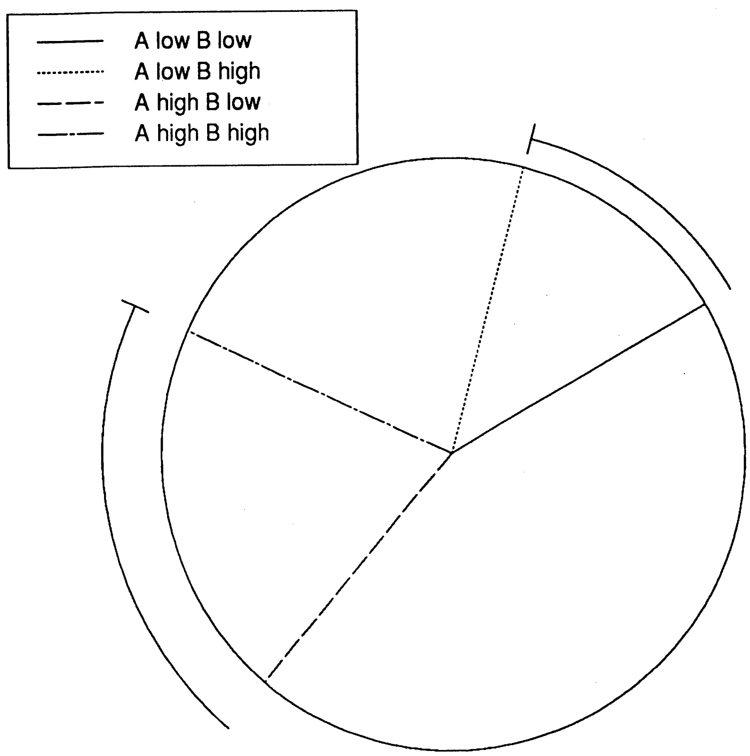

Figure 7: Non-Zero Interaction with Magnitude and Directional Differences

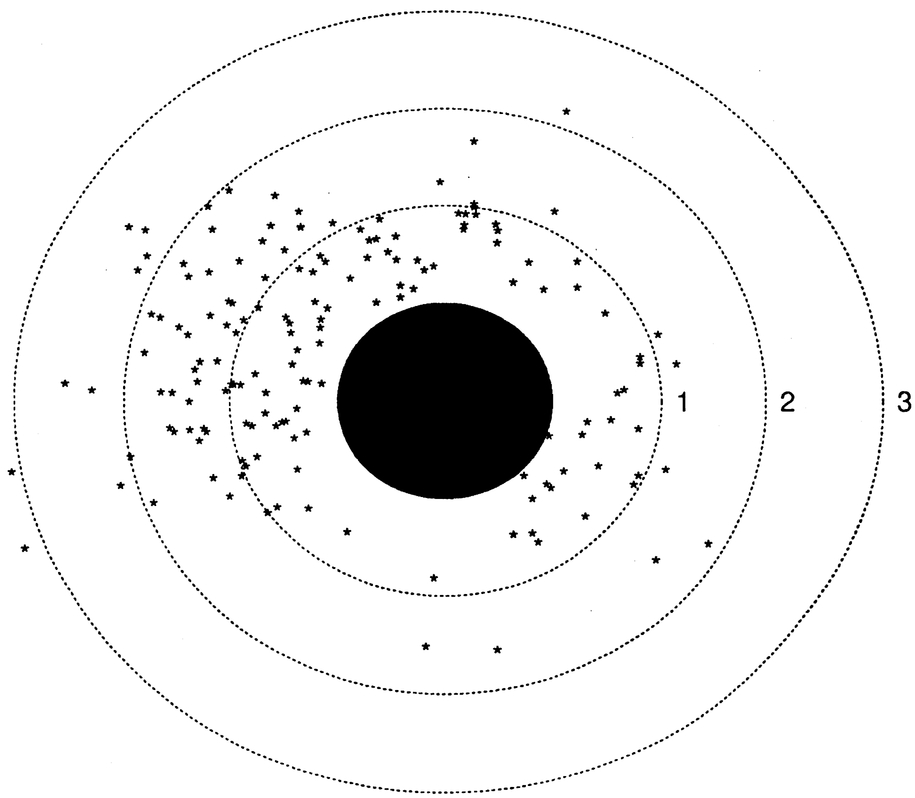

Figure 8: Cylindrical Plot