

# Hyperdimensional Data Analysis Using Parallel Coordinates<sup>1</sup>

Edward J. Wegman  
Center for Computational Statistics  
George Mason University  
Fairfax, VA 22030

**Abstract.** This paper presents the basic results for using the parallel coordinate representation as a high dimensional data analysis tool. Several alternatives are reviewed. The basic algorithm for parallel coordinates is laid out and a discussion of its properties as a projective transformation are shown. The several of the duality results are discussed along with their interpretations as data analysis tools. A discussion of permutations of the parallel coordinate axes is given and some examples are given. Some extensions of the parallel coordinate idea are given. The paper closes with a discussion of implementation and some of our experiences are relayed.

---

<sup>1</sup>This research was supported by the Air Force Office of Scientific Research under grant number AFOSR-87-0179, by the Army Research Office under contract number DAAL03-87-K-0087 and by the National Science Foundation under grant number DMS-8701931.

# Hyperdimensional Data Analysis Using Parallel Coordinates

**1. Introduction.** The classic scatter diagram is a fundamental tool in the construction of a model for data. It allows the eye to detect such structures in data as linear or nonlinear features, clustering, outliers and the like. Unfortunately, scatter diagrams do not generalize readily beyond three dimensions. For this reason, the problem of visually representing multivariate data is a difficult, largely unsolved one. The principal difficulty, of course, is the fact that while a data vector may be arbitrarily high dimensional, say  $n$ , Cartesian scatter plots may only easily be done in two dimensions and, with computer graphics and more effort, in three dimensions. Alternative multidimensional representations have been proposed by several authors including Chernoff (1973), Fienberg (1979), Cleveland and McGill (1984) and Carr et al. (1986).

An important technique based on the use of motion is the computer-based kinematic display yielding the illusion of three dimensional scatter diagrams. This technique was pioneered by Friedman and Tukey (1973) and is now available in commercial software packages (Donohoe's MacSpin and Velleman's Data Desk). Coupled with easy data manipulation, the kinematic display techniques have spawned the exploitation of such methods as projection pursuit (Friedman and Tukey, 1974) and the grand tour (Asimov, 1985). Clearly, projection-based techniques lead to important insights concerning data. Nonetheless, one must be cautious in making inferences about high dimensional data structures based on projection methods alone.

The reason for this cautionary note is that normal 2- and 3-dimensional geometric intuition fails for hyperdimensional geometry very rapidly with increase in dimension. An easy illustration involves the general intersection of two 2-planes in Euclidean 4-space. Our normal three-dimensional intuition is that the general intersection of two planes is a line. However, 4-space can be constructed analytically as the Cartesian product of two orthogonal 2-planes. It is immediate from that observation that the general intersection of two 2-planes in 4-space is a point. The next two examples are designed to illustrate the potential pitfalls of projection-based methods. We refer to Kendall (1961) for a general treatment of  $n$ -dimensional geometry.

**Example 1.1 Diagonals in Hyperspace.** Consider the general ray-diagonal in  $n$ -dimensional space, that is, the vector passing through  $(d_1, d_2, \dots, d_n)$  originating at  $(0, 0, \dots, 0)$ . Here  $d_j = \pm 1$ . Choose any diagonal and fix it. For purposes of discussion consider the ray diagonal through  $(1, 1, \dots, 1)$ . It is easy to see for  $n = 2$  that the angle  $\theta_2$  between the ray diagonal and any of the coordinate axes is characterized by  $\cos \theta_2 = 1/\sqrt{2}$ . A simple inductive argument leads to  $\cos \theta_n = 1/\sqrt{n}$  in the general case. For  $n \rightarrow \infty$ ,  $\cos \theta_n \rightarrow 0$  so that  $\theta_n \rightarrow \pi/2$ . Thus the ray diagonals are nearly orthogonal to the coordinate axes for  $n$  reasonably large. A simple computation shows that there are  $2^n$  ray diagonals or  $2^{n-1}$  diagonals in  $n$ -dimensional space. From a data analytic perspective, the implication of this computation is that data structures lying near a diagonal in hyperspace will be mapped nearly into the origin in every lower dimensional projection while a similar data structure located near a coordinate axis will not be. Thus the intuition gained from a scatter diagram in two or three dimensions is highly dependent on the coordinate axis system chosen. It is relatively easy to miss real data structures by simply examining lower dimensional projections.

**Example 1.2 Hypervolume of a Thin Shell.** The area of a circle of radius  $r$  is  $c_2 r^2$ , the volume of a sphere of radius  $r$  is  $c_3 r^3$  and, in general, the hypervolume (content) of a hypersphere of radius  $r$  is  $c_n r^n$ . In general then the hypervolume of a thin shell is  $c_n [r^n - (r - \epsilon)^n]$  and the hypervolume of the thin shell relative to the hypervolume of the  $n$ -sphere is  $1 - (1 - \epsilon/r)^n$ . Since  $1 - \epsilon/r < 1$ , the relative hypervolume converges to 1 as  $n \rightarrow \infty$ . Loosely speaking, most of the hypervolume is close to the  $(n - 1)$ -dimensional hypersurface of the  $n$ -sphere.

Consider then a probability measure uniform in the volume of an  $n$ -dimensional hypersphere. Consider further a random sample of observations drawn at random according to this measure. If we were capable of visualizing  $n$ -dimensional scatter diagrams, we would notice that most of the observations would lie close to the  $(n - 1)$ -hypersurface of the  $n$ -sphere. If on the other hand, we project observations onto a 2-plane, we would have a scatter diagram with a circular cross-section. The most intense concentration of observations would be near the center, i.e. the distribution would appear unimodal. Indeed, it is easy to show the marginal density would be

unimodal. (In the case of two dimensions, the density on the 1-plane would be  $f(x) = \pi^{-1}\sqrt{1-x^2}$ .) Thus we have the curious state of affairs that in the n-dimensional scatter diagram, most of the observations would lie near the boundaries of the n-sphere, but in the 2-dimensional projection, most of the observations would lie near the center. We contend that the two-dimensional projection may convey the wrong intuition.

The point of these examples is that exploratory data analysis based on projection techniques including 2- and 3-dimensional scatter diagrams is potentially misleading. It would be highly desirable to have a simultaneous representation of all coordinates of a data vector especially if the representation treated all components in a similar manner. The cause of the failure of the standard Cartesian coordinate representation is the requirement for orthogonal coordinate axes. In a 3-dimensional world, it is difficult to represent more than three orthogonal coordinate axes. We propose to give up the orthogonality requirement and replace the standard Cartesian axes with a set of n parallel axes.

**2. Parallel Coordinates.** We propose as a multivariate data analysis tool the following representation. In place of a scheme trying to preserve orthogonality of the n-dimensional coordinate axes, draw them as parallel. A vector  $(x_1, x_2, \dots, x_n)$  is plotted by plotting  $x_1$  on axis 1,  $x_2$  on axis 2 and so on through  $x_n$  on axis n. The points plotted in this manner are joined by a broken line. Figure 2.1 illustrates two points (one solid, one dashed) plotted in parallel coordinate representation. In this illustration, the two points agree in the fourth coordinate. The principal advantage of this plotting device is clear. Each vector  $(x_1, x_2, \dots, x_n)$  is represented in a planar diagram so that each vector component has essentially the same representation.

The parallel coordinates proposal has its roots in a number of sources. Griffen (1958) considers a 2-dimensional parallel coordinate type device as a method for graphically computing the Kendall tau correlation coefficient. Hartigan (1975) describes the "profiles algorithm" which he describes as "histograms on each variable connected between variables by identifying cases." Although he does not recommend drawing all profiles, a profile diagram with all profiles plotted is a parallel coordinate plot. There is however far more mathematical structure, particularly high dimensional structure, to the parallel coordinate diagram than Hartigan exploits. Inselberg (1985) originated the parallel coordinate representation as a device for computational geometry. His 1985 paper is the culmination of a series of technical reports dating from 1981. Finally we note that Diaconis and Friedman (1983) discuss the so-called M and N plots. Their special case of a 1 and 1 plot is a parallel coordinate plot in two dimensions. Indeed, the 1 and 1 plot is sometimes called a before-and-after plot and has a much older history. Also related is the Andrews (1972) plot which can be viewed as a Fourier series interpolation of the points on the parallel coordinate axes. This interpolation preserves some least squares properties through Parseval's relationship, but does not enjoy the statistical interpretations of parallel coordinates due to the projective duality. The fundamental theme of this paper is that the transformation from Cartesian coordinates to parallel coordinates is a highly structured mathematical transformation, hence, maps mathematical objects into mathematical objects. Certain of these can be given highly useful statistical interpretations so that this representation becomes a highly useful data analysis tool.

**3. Parallel and Non-Orthogonal Coordinate Geometry.** The parallel coordinate representation enjoys some elegant duality properties with the usual Cartesian orthogonal coordinate representation. Consider a line  $\mathcal{L}$  in the Cartesian coordinate plane given by  $\mathcal{L}: y = mx + b$  and consider two points lying on that line, say  $(a, ma + b)$  and  $(c, mc + b)$ . For simplicity of computation we consider the xy Cartesian axes mapped into the xy parallel axes as described in Figure 3.1a. We superimpose a Cartesian coordinate axes t,u on the xy parallel axes so that the y parallel axis has the equation  $u = 1$ . The point  $(a, ma + b)$  in the xy Cartesian system maps into the line joining  $(a, 0)$  to  $(ma + b, 1)$  in the tu coordinate axes. Similarly,  $(c, mc + b)$  maps into the line joining  $(c, 0)$  to  $(mc + b, 1)$ . It is a straightforward computation to show that these two lines intersect at a point (in the tu plane) given by  $\bar{\mathcal{L}}: (b(1-m)^{-1}, (1-m)^{-1})$ . Notice that this point in the parallel coordinate plot depends only on m and b the parameters of the original line in the Cartesian plot. Thus  $\bar{\mathcal{L}}$  is the dual of  $\mathcal{L}$  and we have the interesting duality result that points in Cartesian coordinates map into lines in parallel coordinates while lines in Cartesian coordinates map into points in parallel coordinates.

For  $0 < (1-m)^{-1} < 1$ , m is negative and the intersection occurs between the parallel coordinate axes. For  $m = -1$ , the intersection is exactly midway. A ready statistical interpretation can be given. For highly negatively correlated pairs, the dual line segments in parallel coordinates will tend to cross near a single point

between the two parallel coordinate axes. The scale of one of the variables may be transformed in such a way that the intersection occurs midway between the two parallel coordinate axes in which case the slope of the linear relationship is negative one.

In the case that  $(1 - m)^{-1} < 0$  or  $(1 - m)^{-1} > 1$ ,  $m$  is positive and the intersection occurs external to the region between the two parallel axes. In the special case  $m = 1$ , this formulation breaks down. However, it is clear that the point pairs are  $(a, a + b)$  and  $(c, c + b)$ . The dual lines to these points are the lines in parallel coordinate space with slope  $b^{-1}$  and intercepts  $-ab^{-1}$  and  $-cb^{-1}$  respectively. Thus the duals of these lines in parallel coordinate space are parallel lines with slope  $b^{-1}$ . We thus append the ideal points to the parallel coordinate plane to obtain a projective plane. These parallel lines intersect at the ideal point in direction  $b^{-1}$ . In the statistical setting, we have the following interpretation. For highly positively correlated data, we will tend to have lines not intersecting between the parallel coordinate axes. By suitable linear rescaling of one of the variables, the lines may be made approximately parallel in direction with slope  $b^{-1}$ . In this case the slope of the linear relationship between the rescaled variables is one. See Figures 3.2 for an illustration of a sequence of correlations ranging from large positive to large negative. Of course, nonlinear relationships will not respond to simple linear rescaling. However, by suitable nonlinear transformations, it should be possible to transform to linearity.

The star diagrams discussed in Fienberg (1979) can be extended in a natural way by computations similar to above. The basic idea is to consider an  $n$ -dimensional plot as follows. Draw  $n$  radial lines from a common center point at equal angles ( $\theta = 2\pi/n$ ) and label each according to one of the variables,  $x_i$ . Mark each axis proportionally to the size of the variable,  $x_i$ , and join markings on adjacent axes by a straight line segment. This is, of course, similar to the parallel coordinate paradigm in that points map into line segments except that for star diagrams the coordinate axes intersect. Although not commonly done, it is possible to plot all data points on the same diagram. One sector of such a diagram is illustrated in Figure 3.1b. If points  $(a, ma + b)$  and  $(c, mc + b)$  lie on a straight line,  $\mathcal{L}$ , in Cartesian coordinates, we have already seen that the parallel coordinate dual of  $\mathcal{L}$ ,  $\bar{\mathcal{L}}$ , is a point which depends only on the parameters  $m$  and  $b$  determining  $\mathcal{L}$ . A natural question is whether or not such a corresponding phenomena happens in the star diagram. It is not hard to see that the point  $(a, ma + b)$  is mapped into the line segment joining  $(a, 0)$  to  $((ma + b)\cos\theta, (ma + b)\sin\theta)$  where the latter points are given in the  $tu$  coordinate system. Similarly,  $(c, mc + b)$  is mapped into the line segment joining  $(c, 0)$  to  $((mc + b)\cos\theta, (mc + b)\sin\theta)$ . The line segments joining these points respectively have equations in the  $tu$  coordinate system

$$u = \frac{(ma+b)\sin\theta(t-a)}{(ma+b)\cos\theta-a}$$

and

$$u = \frac{(mc+b)\sin\theta(t-c)}{(mc+b)\cos\theta-c}.$$

Simultaneous solution of these two equations does not yield solutions independent of  $a$  and  $c$ , hence the locus of intersection points is not degenerate. In fact since points on the  $y$  axis are projectively related (but not by a central perspectivity) to points on the  $x$  axis (since they are linear transforms  $x \rightarrow mx + b$ ) and the two axes are not coincident, by elementary projective geometry the locus of line segments is a conic (see p. 142 of Fishback, 1962). If the points are related by a central perspectivity, then the conic is degenerate. This is true in the special case we showed earlier for parallel coordinates. Notice that in the special case  $b = 0$ , the lines given above become

$$u = \frac{m \sin\theta (t-a)}{m \cos\theta - 1}$$

and

$$u = \frac{m \sin\theta (t-c)}{m \cos\theta - 1}.$$

Thus the two line segments have common slope independent of  $a$  or  $c$ , but intercept which is dependent on  $a$  and  $c$ . Thus the lines do have a common intersection at an ideal point whose direction depends only on  $m$  and  $\theta$ . The star diagrams have mathematical roots in projective geometry in common with parallel coordinates, but because of the lack of parallelism, they do not share the line-point duality properties which yield useful statistical interpretations.

**4. Natural Homogeneous Coordinates and Conics.** The point-line, line-point duality seen in the transformation from Cartesian to parallel coordinates extends to conic sections. To see this consider both the  $xy$  plane and the  $tu$  plane to be augmented by suitable ideal points so that we may regard both as projective planes. The representation of points in parallel coordinates is thus a transformation from one projective plane to another. Computation is simplified by an analytic representation. However, the usual coordinate pair,  $(x, y)$ , is not sufficient to represent ideal points. We plan to represent points in the projective plane by triples  $(x, y, z)$ . Consider two distinct parallel lines having equations

$$ax + by + cz = 0 \text{ and } ax + by + c/z = 0.$$

Simultaneous solution yields  $(c - c/z)z = 0$  so that  $z = 0$  describes ideal points. The representation of points in the projective plane is by triples,  $(x, y, z)$ , which are called natural homogeneous coordinates. If  $z = 1$ , the resulting equation is  $ax + by + c = 0$  and so  $(x, y, 1)$  is the natural representation of a point  $(x, y)$  in Cartesian coordinates lying on  $ax + by + c = 0$ . Notice that if  $(px, py, p)$  is any multiple of  $(x, y, 1)$  on  $ax + by + c = 0$ , we have

$$apx + bpy + cp = p(ax + by + c) = p \cdot 0 = 0.$$

Thus the triple  $(px, py, p)$  equally well represents the Cartesian point  $(x, y)$  lying on  $ax + by + c = 0$  so that the representation in natural homogeneous coordinates is not unique. However, if  $p$  is not 1 or 0, we can simply re-scale the natural homogeneous triple to have a 1 for the  $z$ -component and thus read off the Cartesian coordinates directly. If the  $z$  component is zero, we know immediately that we have an ideal point.

Notice that we could equally well consider the triples  $(a, b, c)$  as natural homogeneous coordinates of a line. Thus, triples can either represent points or lines reiterating the fundamental duality between points and planes in the projective plane. Recall now that the line  $\mathcal{L}: y = mx + b$  mapped into the point  $\bar{\mathcal{L}}: (b(1 - m)^{-1}, (1 - m)^{-1})$  in parallel coordinates. In natural homogeneous coordinates,  $\mathcal{L}$  is represented by the triple  $(m, -1, b)$  and the point  $\bar{\mathcal{L}}$  by the triple  $(b(1 - m)^{-1}, (1 - m)^{-1}, 1)$  or equivalently by  $(b, 1, 1 - m)$ . The latter yields the appropriate ideal point when  $m = 1$ . A straightforward computation shows for

$$A = \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & -1 \\ 1 & 0 & 0 \end{bmatrix}$$

that  $t = xA$  or  $(b, 1, 1 - m) = (m, -1, b)A$ . Thus the transformation from lines in orthogonal coordinates to points in parallel coordinates is a particularly simple projective transformation with the rather nice computational property of having only adds and subtracts.

Similarly, a point  $(x_1, x_2, 1)$  expressed in natural homogeneous coordinates maps into the line represented by  $(1, x_1 - x_2, -x_1)$  in natural homogeneous coordinates. Another straightforward computation shows that the linear transformation given by  $t = xB$  or  $(1, x_1 - x_2, -x_1) = (x_1, x_2, 1)B$  where

$$B = \begin{bmatrix} 0 & 1 & -1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

describes the projective transformation of points in Cartesian coordinates to lines in parallel coordinates. Because these are nonsingular linear transformations, hence projective transformations, it follows from the elementary theory of projective geometry that conics are mapped into conics. This is straightforward to see since an elementary quadratic form in the original space, say  $xCx'$  where  $x'$  denotes  $x$  transpose, represents the general conic. Clearly then since  $t = xB$ ,  $B$  nonsingular, we have  $x = tB^{-1}$ , so that  $tB^{-1}C(B^{-1})t'$  is a quadratic form in the image space. An instructive computation involves computing the image of an ellipse  $ax^2 + by^2 - cz^2 = 0$  with  $a, b, c > 0$ . The image in the parallel coordinate space is  $c(t+u)^2 - bu^2 = av^2$ , a general hyperbolic form.

It should be noted that the solution to this equation is not a locus of points, but the natural homogeneous coordinates of a locus of lines, a line conic. The envelope of this line conic is a point conic. In the case of this computation, the point conic in the original Cartesian coordinate plane is an ellipse, the image in the parallel coordinate plane is as we have just seen a line hyperbola with a point hyperbola as envelope. This fact is illustrated in Figure 4.1 and, as it turns out, has an important statistical interpretation. A fuller discussion of projective transformations of conics is given by Dimsdale (1984). Inselberg (1985) generalizes this notion into parallel coordinates resulting in what he calls hstars. A general reference to natural homogeneous coordinates can be found in Fishback (1962).

We mentioned the duality between points and lines and conics and conics. It is worthwhile to point out two other nice dualities. Rotations in Cartesian coordinates become translations in parallel coordinates and vice versa. Perhaps more interesting from a statistical point of view is that points of inflection in Cartesian space become cusps in parallel coordinate space and vice versa. Thus the relatively hard-to-detect inflection point property of a function becomes the notably more easy to detect cusp in the parallel coordinate representation. Inselberg (1985) discusses these properties in detail. It is well worth noting that the natural homogeneous coordinate representation is a standard device in computer graphics because a nonlinear transformation (rotation in two space) can be represented as a  $3 \times 3$  linear transformation of natural homogeneous coordinates. See for example Plastock and Kalley (1986).

**5. Further Statistical Interpretations.** Since ellipses map into hyperbolas, we can have an easy template for diagnosing uncorrelated data pairs. Consider Figure 5.1a. With a completely uncorrelated data set, we would expect the 2-dimensional scatter diagram to fill substantially a circumscribing circle. As illustrated in Figure 5.1a, the parallel coordinate plot would approximate a figure with a hyperbolic envelope. As the correlation approaches negative one, the hyperbolic envelope would deepen so that in the limit we would have a pencil of lines, what we like to call the cross-over effect. As the correlation approaches positive one, the hyperbolic envelope would widen with fewer and fewer cross-overs so that in the limit we would have parallel lines. Thus correlation structure can easily be diagnosed from the parallel coordinate plot. As noted earlier, Griffen (1958) used this as a graphical device for computing the Kendall tau.

Griffen, in fact, attributes the graphical device to Holmes (1928) which predates Kendall's discussion. The computational formula is

$$r = 1 - \frac{4X}{n(n-1)}$$

where  $X$  is the number of intersections resulting by connecting the two rankings of each member by lines, one ranking having been put in natural order. While the original formulation was framed in terms of ranks for both  $x$  and  $y$  axes, it is clear that the number of crossings is invariant to any monotone increasing transformation of either  $x$  or  $y$ , the ranks being one such transformation. Because of this scale invariance, one would expect rank-based statistics to have an intimate relationship to parallel coordinates.

It is clear that if there is a perfect positive linear relationship with no crossings, then  $X = 0$  and  $r = 1$ . Similarly, if there is a perfect negative linear relationship, Figure 3.3b is appropriate and we have a pencil of lines. Since every line meets every other line, the number of intersections is  $\binom{n}{2}$  so that

$$r = 1 - \frac{4\binom{n}{2}}{n(n-1)} = -1.$$

It should be further noted that clustering is easily diagnosed using the parallel coordinate representation. See for example Figures 5.1b and 5.1c which illustrate separation in both x and y and separation only in the first coordinate.

So far we have focused primarily on pairwise parallel coordinate relationships. The idea however is that we can, so to speak, stack these diagrams and represent all n dimensions simultaneously. Figure 5.2 thus illustrates a 5-dimensional hypersphere plotted in parallel coordinates. A 5-dimensional ellipsoid would have a similar general shape but with hyperbolas of different depths. Of course, observations from a 5-dimensional Gaussian density with a covariance matrix of the form  $\sigma^2 I$  would have an approximately spherical 5-dimensional scatterplot, thus Figure 5.2 would be a diagnostic template for this situation.

Figure 5.3 is illustrative of some data structures one might see in a five-dimensional data set. First it should be noted that the plots along any given axis represent dot diagrams (a refinement of the histograms of Hartigan), hence convey graphically the one-dimensional marginal distributions. In this illustration, the first axis is meant to have an approximately normal distribution shape while axis two the shape of the negative of a  $\chi^2$ . As discussed above, the pairwise comparisons can be made. Figure 5.3 illustrates a number of instances of linear (both negative and positive), nonlinear and clustering situations. Indeed, it is clear that there is a 3-dimensional cluster along coordinates 3, 4 and 5.

Consider also the appearance of a mode in parallel coordinates. The mode is, intuitively speaking, the location of the most intense concentration of probability. Hence, in a sampling situation it will be the location of the most intense concentration of observations. Since observations are represented by broken line segments, the mode in parallel coordinates will be represented by the most intense bundle of broken line paths in the parallel coordinate diagram. Roughly speaking, we should look for the most intense flow through the diagram. In Figure 5.3, such a flow begins near the center of coordinate axis one and finishes on the left-hand side of axis five.

Figure 5.3 thus illustrates some data analysis features of the parallel coordinate representation including the ability to diagnose one-dimensional features (marginal densities), two-dimensional features (correlations and nonlinear structures), three-dimensional features (clustering) and a five-dimensional feature (the mode). In section 7 of this paper we consider a real data set which will be illustrative of some additional capabilities.

**6. Permutation of the Axes for Pairwise Comparisons.** In Figure 5.3, the parallel coordinate axes were ordered from 1 through 5. This allowed an easy pairwise comparison of 1 with 2, 2 with 3 and so on. The pairwise comparison of 1 with 3, 2 with 5 and so on was not easily done however since these axes were not adjacent. A natural question arises about the number of permutations required so that in some permutation every axis is adjacent to every other axis. Although there are n! permutations, many of these duplicate adjacencies. Actually far fewer permutations are required.

A construction for determining the permutations is represented in Figure 6.1. A graph is drawn with vertices representing coordinate axes, labeled clockwise 1 to n. Edges represent adjacencies, so that vertex one connected to vertex two by an edge means axis one is placed adjacent to axis two. To construct a minimal set of permutations that completes the graph is equivalent to finding a minimal set of orderings of the axes so that every possible adjacency is present. Figure 6.1b illustrates the basic zig-zag pattern used in the construction. This creates an ordering which in the example of Figure 6.1b is 1 2 7 3 6 4 5. For n even this general sequence can be written as 1, 2, n, 3, n - 1, 4, n - 2, ... , (n + 2)/2 and for n odd as 1, 2, n, 3, n - 1, 4, n - 2, ... , (n + 3)/2.

An even simpler formulation is

$$n_{k+1} = (n_k + (-1)^{k+1}k) \bmod n, k = 1, 2, \dots, n-1 \quad (6.1)$$

with  $n_1 = 1$ . Here it is understood that  $0 \bmod n = n \bmod n = n$ . This zig-zag pattern can be recursively applied to complete the graph. That is to say if we let  $n_k^{(1)} = n_k$ , we may define

$$n_k^{(j+1)} = (n_k^{(j)} + 1) \bmod n, j = 1, 2, \dots, \lfloor \frac{n-1}{2} \rfloor \quad (6.2)$$

where  $\lfloor \cdot \rfloor$  is the greatest integer function. For  $n$  even, it follows that this construction generates each edge in one and only one permutation. Thus  $n/2$  is the minimal number of permutations needed to assure that every edge appears in the graph or equivalently that every adjacency occurs in the parallel coordinate representation. For  $n$  odd, the result is not exactly the same. We will not have any duplication of adjacencies for  $j < \lfloor (n-1)/2 \rfloor$ . However,  $j < \lfloor (n-1)/2 \rfloor$  will not provide a unique graph. The case  $j = \lfloor (n-1)/2 \rfloor$  in equation (6.2) will complete the graph, but also create some redundancies. Nevertheless, it is clear that  $\lfloor (n+1)/2 \rfloor$  permutations are the minimal number needed to complete the graph and thus provide every adjacency in the parallel coordinate representation. Thus we have

**Theorem:** The minimal number of permutations of the  $n$  parallel coordinate axes needed to insure adjacency of every pair of axes is  $\lfloor (n+1)/2 \rfloor$ . These permutations may be constructed using formulas (6.1) and (6.2).

**7. An Auto Data Example.** Table 1 in the appendix consists of data on 86 1980 model year automobiles. It consists of thirteen simultaneous measurements including price, miles per gallon, manufacturer, reliability of the same model in 1978, reliability of the same model in 1979, gear ratio, cubic inch engine displacement, head room, turning radius, rear seatroom, length, trunk size and weight. Figure 7.1 is a parallel coordinate plot of this data. While rather overwhelming at first, some clear features emerge. First, and perhaps most striking, is the lack of any evidence at all of hyperbolic envelopes. Thus any discussion of this data that assumes normality is clearly suspect. The categorical nature of the reliability and manufacturer data is also clear. The headroom measurement also shows marked discretization, the data being rounded to the nearest  $\frac{1}{2}$  inch. Several other categories show quantization effects including turning radius and trunk space.

In order to focus a bit we consider a 5-dimensional subset of the data, the measured variables being price, miles per gallon, gear ratio, weight and cubic inch displacement. For  $n = 5$ ,  $\lfloor (n+1)/2 \rfloor = 3$  presentations are needed to present all pairwise permutations. Figures 7.2, 7.3 and 7.4 are these three presentations. In Figure 7.2, perhaps the most striking feature is the cross-over effect evident in the relationship between gear ratio and weight. This suggests a negative correlation. Indeed, this is reasonable since a heavy car would tend to have a large engine providing considerable torque thus requiring a lower gear ratio. Conversely, a light car would tend to have a small engine providing small amounts of torque thus requiring a higher gear ratio.

Consider as well the relationship between weight and cubic inch displacement. In this diagram we have a considerable amount of approximate parallelism (relatively few crossings) suggesting positive correlation. This is a graphic representation of the fact that big cars tend to have big engines, a fact most are prepared to believe. Quite striking however is the negative slope going from low weight to moderate cubic inch displacement. This is clearly an outlier which is unusual in neither variable but in their joint relationship. The same observation is highlighted in Figure 7.3.

The relationship between miles per gallon and price is also perhaps worthy of comment. The left-hand side shows an approximate hyperbolic boundary while the right-hand side clearly illustrates the cross-over effect. This suggests for inexpensive cars or poor mileage cars there is relatively little correlation. However, costly cars almost always get relatively poor mileage while good gas mileage cars are almost always relatively inexpensive.

Turning to Figure 7.3, the relationship between gear ratio and miles per gallon is instructive. This diagram is suggestive of two classes. Notice that there are a number of observations represented by line segments tilted slightly to the right of vertical (high positive slope) and a somewhat larger number with a negative slope of about  $-1$ . Within each of these two classes we have approximate parallelism. This suggests that the relationship between gear ratios and miles per gallon is approximately linear, a believable conjecture since low gears = big engines = poor mileage while high gears = small engines = good mileage. What is intriguing, however, is

that there seems to be really two distinct classes of automobiles each exhibiting a linear relationship, but with different linear relationships within each class.

Indeed in Figure 7.4, the third permutation, we are able to highlight this separation into two classes in a truly 5-dimensional sense. The shaded region in Figure 7.4 describes a class of vehicles with relatively poor gas mileage, relatively heavy, relatively inexpensive, relatively large engines and relatively low gear ratios. Figure 7.5 is a repeat of this graphic but with different shading highlighting a class of vehicles with relatively good gas mileage, relatively light weight, relatively inexpensive, relatively small engines and relatively high gear ratios. In 1980, these two characterizations describe respectively domestic automobiles and imported automobiles.

**8. Graphical Extensions of Parallel Coordinate Plots.** The basic parallel coordinate idea suggests some additional plotting devices. We call these respectively the Parallel Coordinate Density Plots, Relative Slope Plots and Color Histograms. These are extensions of the basic idea of parallel coordinates, but structured to exploit additional features or to convey certain information more easily.

**8.1 Parallel Coordinate Density Plots.** While the basic parallel coordinate plot is a useful device itself, like the conventional scatter diagram, it suffers from heavy overplotting with large data sets. In order to get around this problem, we use a parallel coordinate density plot which is computed as follows. Our algorithm is based on Scott (1986) notion of average shifted histogram (ASH) but adapted to the parallel coordinate context. As with an ordinary two dimensional histogram, we decide on appropriate rectangular bins. A potential difficulty arises because a line segment representing a point may appear in two or more bins in the same horizontal slice. Obviously if we have  $k$   $n$ -dimensional observations, we would like to form a histogram based on  $k$  entries. However, since the line segment could appear in two or more bins in a horizontal slice, the count for any given horizontal slice is at least  $k$  and may be bigger. Moreover, every horizontal slice may not have the same count. To get around this, we convert line segments to points by intersecting each line segment with a horizontal line passing through the middle of the bin. This gives us an exact count of  $k$  for each horizontal slice. We construct an ASH for each horizontal slice (typically averaging 5 histograms to form our ASH). We have used contours to represent the two-dimensional density although gray scale shading could be used in a display with sufficient bit-plane memory. An example of an parallel coordinate density plot is given in Figure 8.1. Parallel coordinate density plots have the advantage of being graphical representations of data sets which are simultaneously high dimensional and very large.

**8.2 Relative Slope Plots.** We have already seen that parallel line segments in a parallel coordinate plot correspond to high positive correlation (linear relationship). As in our automobile example, it is possible for two or more sets of linear relationships to exist simultaneously. In an ordinary parallel coordinate plot, we see these as sets of parallel lines with distinct slopes. The work of Cleveland and McGill (1985) suggests that comparison of slopes (angles) is a relatively inaccurate judgement task and that it is much easier to compare magnitudes on the same scale. The relative slope plot is motivated by this. In an  $n$ -dimensional relative slope plot there are  $n - 1$  parallel axes, each corresponding to a pair of axes, say  $x_i$  and  $x_j$ , with  $x_j$  regarded as the lower of the two coordinate axes. For each observation, the slope of the line segment between the pair of axes is plotted as a magnitude between  $-1$  and  $+1$ . The maximum positive slope is coded as  $+1$ , the minimum negative slope as  $-1$  and a slope of  $\infty$  as  $0$ . The magnitude is calculated as  $\cos \eta$  where  $\eta$  is the angle between the  $x_j$  axis and the line segment corresponding to the observation. Each individual observation in the relative slope plot corresponds to a vertical section through the axis system. An example of a relative slope plot is given in Figure 8.2. Notice that since slopes are coded as heights, simply laying a straightedge will allow us to discover set of linear relationships within the pair of variables  $x_i$  and  $x_j$ .

**8.3 Color Histograms.** The basic set-up for the color histogram is similar to the relative slope plots. For an  $n$ -dimensional data set, there are  $n$  parallel axes. A vertical section through the diagram corresponds to an observation. The idea is to code the magnitude of an observation along a given axis by a color bin, the colors being chosen to form a color gradient. We typically choose 8 to 15 colors. The diagram is drawn by choosing an axis, say  $x_k$ , and sorting the observations in ascending order. Along this axis, we see blocks of color arranged according to the color gradient with the width of the block being proportional to the number of observations falling into the color bin. The observations on the other axes are arranged in the order corresponding to the  $x_k$  axis and color coded according to their magnitude. Of course, if the same color gradient shows up say on the  $x_m$  axis as on the  $x_k$ , then we know  $x_k$  is positively "correlated" with  $x_m$ . If the color gradient is reversed, we know

the "correlation" is negative. We used the phrase "correlation" advisedly since in fact if the color gradient is the same but the color block sizes are different, the relationship is nonlinear. Of course if the  $x_m$  axis show color speckle, there is no "correlation" and  $x_k$  is unrelated to  $x_m$ . An example of a color histogram is given in Figure 8.3 (for purposes of reproduction here it is really a gray-scale histogram).

**9. Implementations.** Our parallel coordinates data analysis software has been implemented in two forms, one a PASCAL program operating on the IBM RT under the AIX operating system. This code allows for up to four simultaneous windows and offers simultaneous display of parallel coordinates and scatter diagram displays. It offers highlighting, zooming and other similar features and also allows the possibility of nonlinear rescaling of each axis. It incorporates the axes permutations described in Section 6 and also includes Parallel Coordinate Density Plots, Relative Slope Plots and Color Histograms.

Our second implementation is under development in PASCAL for MS-DOS machines and include similar features. In addition, it has a mouse-driven painting capability and can do real-time rotation of 3-dimensional scatter plots. Both programs use EGA graphics standards, with the second also using VGA or Hercules monochrome standards.

We regard the parallel coordinate representation as a device complementary to scatterplots. A major advantage of the parallel coordinate representation over the scatterplot matrix is the linkage provided by connecting points on the axes. This linkage is difficult to duplicate in the scatterplot matrix. Because of the projective line-point duality, the structures seen in a scatterplot can also be seen in a parallel coordinate plot. Moreover, the work of Cleveland and McGill (1984b) suggests that it is easier and more accurate to compare observations on a common scale. The parallel coordinate plot and the derivatives of it de facto have a common scale and so for example a sense of variability and central tendency among the variables are easier to grasp visually in parallel coordinates when compared with the scatterplot matrix. On the other hand, one might interpret all the ink generated by the lines as a significant disadvantage of the parallel coordinate plot. Our experience on this is mixed. Certainly for large data sets on hard copy this is a problem. On the other hand with traditional scatterplots viewed on an interactive graphics screen particularly a high resolution screen, we have often found that individual points can get lost because they are simply not bright enough. That does not happen in a parallel coordinate plot. However, if many points are plotted using parallel coordinates on monochrome screen, it is hard to distinguish them. We have gotten around this problem by plotting distinct points in different colors. In an EGA implementation, this means 16 colors. This is surprisingly effective in separating points. In one experiment, we plotted 5000 5-dimensional random vectors using 16 colors, and in spite of total overplotting, we were still able to see structure. In data sets of somewhat smaller scale, we have implemented a scintillation technique. With this technique, when there is overplotting we cause the screen view to scintillate between the colors representing the overplotted points. The speed of scintillation is proportional to the number of points overplotted and by carefully tracing colors, one can follow an individual point through the entire diagram.

We have found painting to be an extraordinarily effective technique in parallel coordinates. We have a painting scheme that not only paints all lines within a given rectangular area, but also all line lying between to slope constraints. This is very effective in separating clusters. We also use invisible paint to eliminate observation points from the data set temporarily. This is a natural way of doing a subset selection.

## References

- Andrews, D. F. (1972), "Plots of high dimensional data," *Biometrics*, 28, 125-136.
- Asimov, Daniel (1985), "The grand tour: a tool for viewing multidimensional data," *SIAM J. Scient. Statist. Comput.*, 6, 128-143.
- Carr, D. B., Nicholson, W. L., Littlefield, R., Hall, D. L. (1986), "Interactive color display methods for multivariate data," in *Statistical Image Processing and Graphics*, (Wegman, E. and DePriest, D., eds.), New York: Marcel Dekker, Inc.
- Chernoff, H. (1973), "Using faces to represent points in k-dimensional space," *J. Am. Statist. Assoc.*, 68, 361-368.
- Cleveland, W. S. and McGill, R. (1984a), "The many faces of the scatterplot," *J. Am. Statist. Assoc.*, 79, 807-822.

- Cleveland, W. S. and McGill, R. (1984b), "Graphical perception: theory, experimentation, and application to the development of graphical methods," *J. Am. Statist. Assoc.*, 79, 531-554.
- Diaconis, P. and Friedman, J. (1983), "M and N plots," in *Recent Advances in Statistics*, 425-447, New York: Academic Press, Inc.
- Dimsdale, B. (1984), "Conic transformations and projectivities," IBM Los Angeles Scientific Center Report #6320-2753.
- Fienberg, S. (1979), "Graphical methods in statistics," *Am. Statistician*, 33, 165-178.
- Fishback, W. T. (1962), *Projective and Euclidean Geometry*, New York: John Wiley and Sons, Inc.
- Friedman, J. and Tukey, J. W. (1973), "PRIM-9" a film produced by Stanford Linear Accelerator Center, Stanford, CA Bin 88 Productions, April, 1973.
- Friedman, J. and Tukey, J. W. (1974), "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, C-23, 881-889.
- Griffen, H. D. (1958), "Graphic computation of tau as a coefficient of disarray," *J. Am. Statist. Assoc.*, 53, 441-447.
- Hartigan, John A. (1975), *Clustering Algorithms*, New York: John Wiley and Sons, Inc.
- Holmes, S. D. (1928), "Appendix B: a graphical method for estimating R for small groups," 391-394 in *Educational Psychology* (Peter Sandiford, auth.), New York: Longmans, Green and Co.
- Inselberg, A. (1985), "The plane with parallel coordinates," *The Visual Computer*, 1, 69-91.
- Kendall, M. G. (1961), *A Course in the Geometry of n-Dimensions*, London: Charles Griffen and Company, Ltd.
- Plastock, R. A. and Kalley, G. (1986), *Computer Graphics*, New York: McGraw-Hill Book Company.
- Scott, D. W. (1985), "Average shifted histograms: effective nonparametric density estimators in several dimensions," *Ann. Statist.*, 13, 1024-1040.