Higher Order Parallel Coordinates

Holger Theisel

University of Rostock, Computer Science Department
PostBox 999, 18051 Rostock, Germany
theisel@informatik.uni-rostock.de

Abstract

We extend the concept of parallel coordinates by drawing free-form curves instead of line segments to connect points on adjacent coordinate axes. This way we can use the space between two adjacent axes to encode more information. Using examples we show that we are able to detect correlations between more than two dimensions.

1 Introduction and main idea

Parallel coordinates ([2], [4]) have become a standard tool in both data visualization and information visualization. The data we consider here consists of a number of objects in N-dimensional information space. This way one particular object is described by an N-tuple $(x_1, ..., x_N)$ of numbers. In the usual parallel coordinate approach N parallel coordinate axes $X_1, ..., X_N$ are used. For a particular object $(x_1, ..., x_N)$ these values are mapped onto the points $\mathbf{p}_1, ..., \mathbf{p}_N$ on the corresponding axes $X_1, ..., X_N$. Then the object $(x_1, ..., x_N)$ is represented by the sequence of line segments $(\mathbf{p}_1, \mathbf{p}_2), (\mathbf{p}_2, \mathbf{p}_3), ..., (\mathbf{p}_{N-1}, \mathbf{p}_N)$. Figure 1 gives an illustration for $N = 4$.

Parallel coordinates are a powerful method to visualize high amounts of multidimensional data. Usually, correlations between adjacent dimensions (i.e. dimensions where its coordinate axes in parallel coordinates are adjacent to each other) can be detected. Unfortunately, correlations between non-adjacent axes can hardly be seen. It is a rather laborious approach to change the order of the axes interactively until a complete visual analysis of the data is achieved.

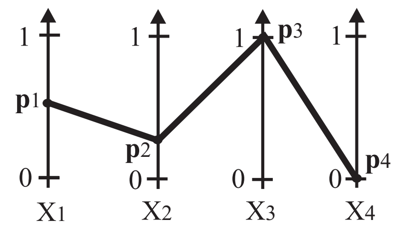

Figure 1: The object $(0.5, 0.25, 1.0, 0.0)$ is described by the line sequence $(\mathbf{p}_1, \mathbf{p}_2), (\mathbf{p}_2, \mathbf{p}_3), (\mathbf{p}_3, \mathbf{p}_4)$.

As an approach to overcome this problem we use the fact that the space between adjacent coordinate axes is sometimes redundant. Figure 2 gives an example.

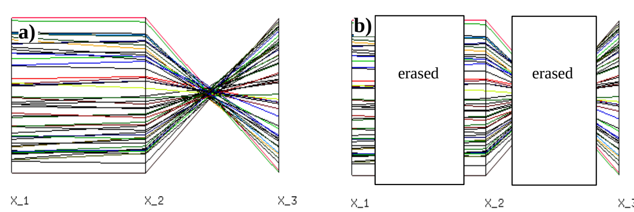

Figure 2: **a)**: Parallel coordinates for two dimensions. In **b)**, areas between the coordinate axes are erased. We can still see the correlations between the coordinate axes. Thus we may place more information into the erased space.

We use the redundant space between two adjacent axes to insert more information. To do so, we replace the line segments between adjacent axes with free-form curves. Curves have the advantage that the human eye reacts rather sensitively to small changes in their shapes. Thus they are promising candidates to encode more information in a small area. The line segments used in the normal parallel coordinate approach can be interpreted as curves of order one. This gives reason to name the approach of replacing the line segments with curves "higher order parallel coordinates".

To define the curves, we place a number of additional coordinate axes from $\{X_1, ..., X_N\}$ between the adjacent axes X_k, X_{k+1}. Then we use the points $\mathbf{p}_k, \mathbf{p}_{k+1}$ and the corresponding points on the newly inserted additional axes as control points of the curve. The shape of the curve is controlled by a weight w which determines the influence of the inner control points. Figure 3 shows an example.

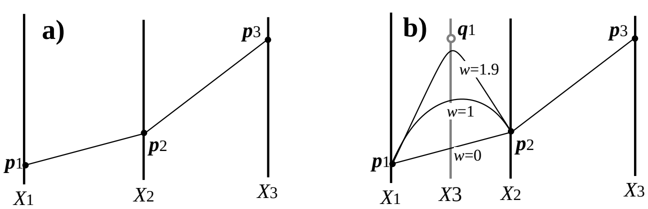

Figure 3: **a)** usual parallel coordinates: one three-dimensional object is visualized by the line segments $(\mathbf{p}_1, \mathbf{p}_2), (\mathbf{p}_2, \mathbf{p}_3)$. **b)** higher order parallel coordinates: the axis X_3 is additionally inserted between X_1 and X_2; $\mathbf{q}_1$ has the same height as $\mathbf{p}_3$. Then the points $\mathbf{p}_1, \mathbf{q}_1, \mathbf{p}_2$ define the curve which replaces the line segment $\mathbf{p}_1, \mathbf{p}_2$ of the usual approach. A weight w controls the influence of $\mathbf{q}_1$.

2 The curve scheme

Considering figure 3 again, we want to specify the impact of the weight parameter w which controls the influence of the additional control point $\mathbf{q}_1$. As a first condition, for $w = 0$ we want the curve to be degenerate into the line segment $(\mathbf{p}_1, \mathbf{p}_2)$. This makes sure that the usual parallel coordinates are a special case of higher order parallel coordinates. As a second condition we demand that if w is increasing to a maximal amount, the curve is going to converge to the line segments $(\mathbf{p}_1, \mathbf{q}_1), (\mathbf{q}_1, \mathbf{p}_2)$. In this case the influence of $\mathbf{q}_1$ to the curve shape is maximal.

At first glance the usage of rational Bezier- or B-spline curves seems to be a promising candidate for higher order parallel coordinates. Unfortunately, rational curve concepts fail the second condition (w becomes maximal) if we have more than two additional axes between two adjacent axes in the usual parallel coordinate approach. So the curve scheme we use here has to be more involved than a simple rational curve approach.

In the following we call the axes of the usual parallel coordinates approach *main axes*; the axes placed between the main axes are called *additional axes*. Given are two adjacent main axes X_k, X_{k+1} with the data points $\mathbf{p}_k, \mathbf{p}_{k+1}$ for a particular object. Between X_k and X_{k+1} we place n additional axes $Y_1, ..., Y_n$ with $\{Y_1, ..., Y_n\} \subseteq \{X_1, ..., X_N\}$. They give the data points $\mathbf{q}_1, ..., \mathbf{q}_n$ for a particular object. Then we define a curve out of $\mathbf{p}_k, \mathbf{q}_1, ..., \mathbf{q}_n, \mathbf{p}_{k+1}$ in the following way: we use a piecewise cubic B-spline approach with the de Boor points $\mathbf{d}_0, ..., \mathbf{d}_{2n+3}$ over the knot sequence $t_0, ..., t_{2n+7}$ (see [1]):

$$
\begin{aligned}
t_0 &= \;... &&= t_3 = 0 \\
t_{2i+2} &= \; i - \frac{1}{4} && \text{for } \; i = 1, ..., n \\
t_{2i+3} &= \; i + \frac{1}{4} && \text{for } \; i = 1, ..., n \\
t_{2n+4} &= \;... &&= t_{2n+7} = n + 1.
\end{aligned}
$$

The de Boor point $\mathbf{d}_i$ is computed as a linear combination of certain auxiliary points $\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i$ for $i = 0, ..., 2n+3$ which are defined as:

$$
\mathbf{a}_i = \frac{2n+3-i}{2n+3}\,\mathbf{p}_k + \frac{i}{2n+3}\,\mathbf{p}_{k+1}
$$

$$
\text{for } \; i = 0, ..., 2n+3
$$

$$\mathbf{b}_0 = \mathbf{p}_k \ , \ \mathbf{b}_1 = \frac{3}{4}\mathbf{p}_k + \frac{1}{4}\mathbf{q}_1$$

$$\mathbf{b}_2 = \frac{1}{4}\mathbf{p}_k + \frac{3}{4}\mathbf{q}_1$$

$$\mathbf{b}_{2i+1} = \frac{3}{4}\mathbf{q}_i + \frac{1}{4}\mathbf{q}_{i+1} \text{ for } i = 1, ..., n-1$$

$$\mathbf{b}_{2i+2} = \frac{1}{4}\mathbf{q}_i + \frac{3}{4}\mathbf{q}_{i+1} \text{ for } i = 1, ..., n-1$$

$$\mathbf{b}_{2n+1} = \frac{3}{4}\mathbf{q}_n + \frac{1}{4}\mathbf{p}_{k+1}$$

$$\mathbf{b}_{2n+2} = \frac{1}{4}\mathbf{q}_n + \frac{3}{4}\mathbf{p}_{k+1}$$

$$\mathbf{b}_{2n+3} = \mathbf{p}_{k+1}$$

$$\mathbf{c}_0 = \mathbf{p}_k \ , \ \mathbf{c}_1 = \mathbf{p}_k$$
$$\mathbf{c}_{2i} = \mathbf{q}_i \ , \ \mathbf{c}_{2i+1} = \mathbf{q}_i \text{ for } i = 1, ..., n$$
$$\mathbf{c}_{2n+2} = \mathbf{p}_{k+1} \ , \ \mathbf{c}_{2n+3} = \mathbf{p}_{k+1}.$$

Then the de Boor points are defined as

$$\mathbf{d}_i = \begin{cases} (1-w)\mathbf{a}_i + w\mathbf{b}_i & \text{for } 0 \leq w < 1 \\ (2-w)\mathbf{b}_i + (w-1)\mathbf{c}_i & \text{for } 1 \leq w \leq 2 \end{cases}$$

Figure 4 illustrates for $n = 2$.

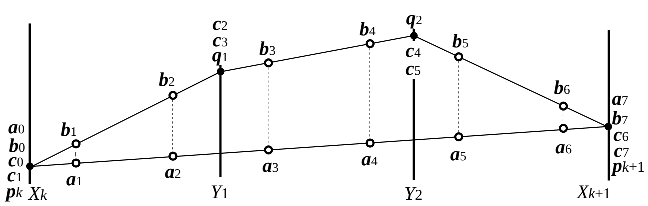

Figure 4: Two additional axes Y_1, Y_2 are placed between the main axes X_k and X_{k+1}. It is shown how the auxiliary points $\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i (i = 0, ..., 7)$ are derived from $\mathbf{p}_k, \mathbf{q}_1, \mathbf{q}_2, \mathbf{p}_{k+1}$. The resulting de Boor points $\mathbf{d}_i$ are linear combinations of $\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i$ using the weight parameter w.

The weight parameter w which ranges between 0 and 2 was introduced to control the influence of the additional axes $Y_1, ..., Y_n$ (and the corresponding additional points $\mathbf{q}_1, ..., \mathbf{q}_n$) on the curve. The parameter w can be moved interactively by the user. For $w = 0$ the curve degenerates into the line segment $(\mathbf{p}_k, \mathbf{p}_{k+1})$. This means that the concept of normal parallel coordinates is still a special case of higher order parallel coordinates. For $w = 2$, the curve turns out to be the sequence of line segments $(\mathbf{p}_k, \mathbf{q}_1), (\mathbf{q}_1, \mathbf{q}_2), ..., (\mathbf{q}_{n-1}, \mathbf{q}_n), (\mathbf{q}_n, \mathbf{p}_{k+1})$. Figure 5 illustrates the influence of w for a configuration where two additional axes are inserted between two adjacent main axes.

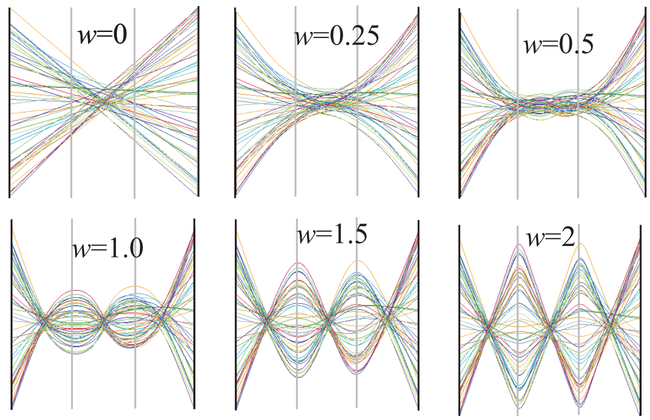

Figure 5: Two additional coordinate axes between two adjacent main axes. Higher order parallel coordinates for different choices of w.

3 Detecting correlations between more than two dimensions

Given a set of objects in N-dimensional information space, the usual parallel coordinate approach is useful for detecting correlations between two dimensions if the corresponding coordinate axes are located adjacent to each other. In real data, correlations between more than two dimensions are possible. Suppose there are correlations between the pairs of dimensions $(d_1, d_2), (d_2, d_3), (d_3, d_1)$. Then a relevant correlation between all three dimensions (d_1, d_2, d_3) may exist or not. (See [3] for a discussion of correlations between more than two dimensions.)

To test the ability of higher order parallel coordinates to deal with correlations between more than two dimensions we explore a number of 4-dimensional test data sets. They are visualized by placing two additional axes Y_1, Y_2 between the main axes X_k, X_{k+1}.

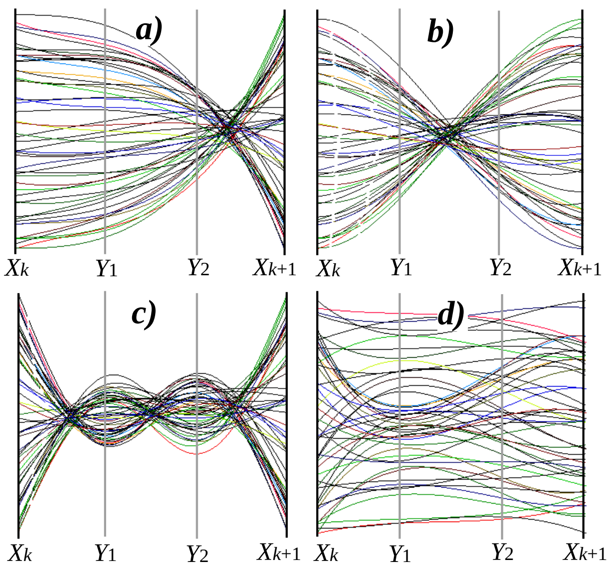

Figure 6: Two additional axes Y_1, Y_2 are placed between the main axes X_k, X_{k+1}. **a)**-**c)**: correlation between all 4 dimensions. **d)**: no correlation between all 4 dimensions. Y_2 was chosen independently of the other dimensions.

The test data set shown in figure 6a consists of equally distributed random values on the main axis X_k. The values for the dimensions Y_1, Y_2 are approximately the same as the values for X_1 (except for slight perturbations). The value of X_{k+1} is approximately the inverse value of X_k (except for a slight perturbations). Thus here is a relevant correlation between the four dimensions. The visualization shows similar and regular patterns.

Figure 6b shows another quadruple of dimensions with relevant correlations between each other. Here X_k and Y_1 have essentially the same values while Y_2 and X_{k+1} have essentially inverse values. Again, similar and regular patterns can be recognized.

Figure 6c shows the correlations between the four dimensions in such a way that X_k and Y_2 have essentially the same values while Y_1 and X_{k+1} have essentially inverse values. Again the visualization looks regular.

In figure 6d Y_1 was chosen independently of the other dimensions. Although there are still correlation between the three dimensions X_k, Y_2, X_{k+1}, the visualization looks "wild". No similar behavior of the curves can be detected.

Figure 6 gives reason for the following statement: higher order parallel coordinates seem to provide a way of detecting and visualizing correlations between the dimensions $X_1, ..., X_m$. We choose two of these axes as main axes and place the remaining axes as additional axes between them. The set of all curves between the main axes gives an impression of the correlations between $X_1, ..., X_m$. If the curves show a similar behavior, correlations between $X_1, ..., X_m$ can be inferred. If the curves behave independently of each other, no symmetric pattern can be recognized in the curve plot. Then no relevant correlation is found.

Another example of constructed test data sets is shown in figures 7 and 8. Both test data set 1 (shown in figure 7) and test data set 2 (shown in figure 8) consist of 70 objects in an 8-dimensional information space. No differences are visible in the visualization using usual parallel coordinates (figures 7a and 8a). We can see correlations between adjacent dimensions: objects with a high value in the axis X_i have to a high probability a high value in X_{i+1} as well. We cannot see whether of not there are correlations between more than two dimensions.

Figures 7b and 8b use higher order parallel coordinates for the test data sets 1 and 2. Between each two main axes two additional axes are inserted. In figure 7b the curves between the main axes do not show a similar behavior. Thus no correlations between quadruples of dimensions are found in test data set 1. In figure 8b the curves between the main axes show a similar behavior; we found quadruples of dimensions with correlations to each other. Here it makes sense to ask for correlations between all 8 dimensions.

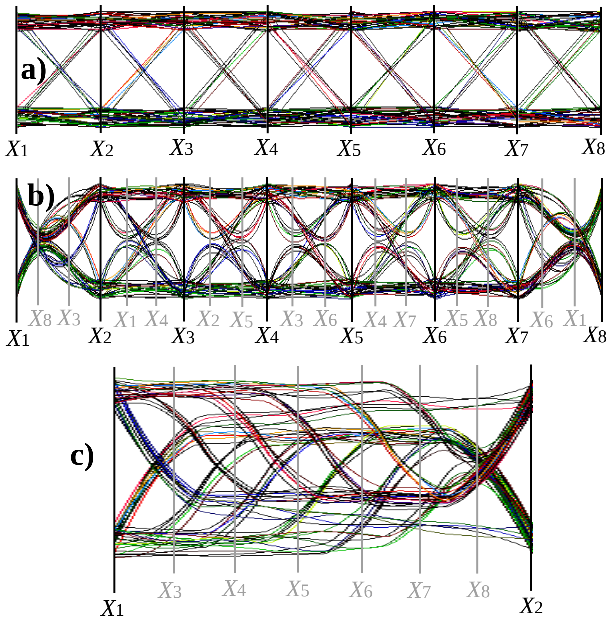

Figure 7: Test data set 1 ; **a)** normal parallel coordinates; **b)** higher order parallel coordinates, between each two main axes, two additional axes are inserted ; **c)** higher order parallel coordinates; two main axes X_1, X_2, 6 additional axes between them.

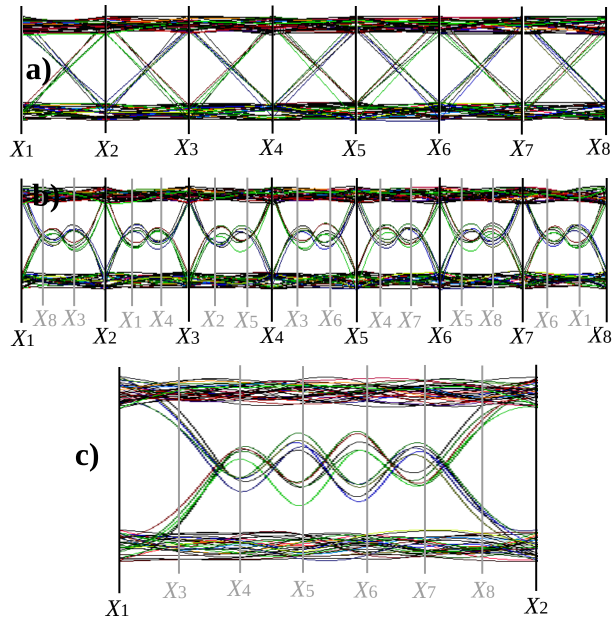

Figure 8: Test data set 2 ; **a)** normal parallel coordinates; **b)** higher order parallel coordinates, between each two main axes, two additional axes are inserted ; **c)** higher order parallel coordinates; two main axes X_1, X_2, 6 additional axes between them.

Figures 7c and 8c show another application of higher order parallel coordinates for the test data sets 1 and 2. Here we have two main axes X_1, X_2, and 6 additional axes between them. Figure 7c shows no correlations between them in data set 1. This is no new information after analyzing figure 7b; it is only for comparison with figure 8c. In figure 8c the curves show a similar behavior. Thus there are correlations between all 8 dimensions. All visualizations for higher order parallel coordinates in figures 7 and 8 were computed using the weight $w = 1$.

4 Application scenario

Given an N-dimensional data set we start out with the usual parallel coordinate approach. Here the order of the axes is subject to interactive change. If correlations between adjacent axes X_k, X_{k+1} are detected, additional axes can be inserted between X_k and X_{k+1} in order to explore whether or not there are correlations between more than two axes (including X_k and X_{k+1}). To do so, the weight parameter is interactively moved between 0 and 2. Figure 9 shows an example for a data set which contains information about 38 automobiles including miles per gallon, weight, drive ratio, horsepower, displacement and number of cylinders. [1]

Figure 9a shows the visualization using usual parallel coordinates. We can detect correlations between the following pairs of dimensions: (MPG, Weight), (Horsepower, Displacement), (Displacement, Cylinders). No relevant correlations are between (Weight, Driveratio), (Driveratio, Horsepower). Figure 9b shows the visualization using higher order parallel coordinates with one additional axis between each pair of adjacent main axes. Correlations between triples of dimensions can be detected for (Horsepower, Weight, Displacement), (Displacement, Horsepower, Cylinders). No relevant correlation are in the

[1] available at http://lib.stat.cmu.edu/DASL /Stories/ClusteringCars.html

triplets (MPG, Driveratio, Weight), (Weight, Horsepower, Driveratio), (Driveratio, Cylinders, Horsepower). The curves for these triples do not show similar patterns.

As with the usual parallel coordinate approach, using higher order parallel coordinates is an interactive process. The user has to find appropriate sequences of main axes as well as the additional axes between them. To explore the behavior of the curve the user can move the weight parameter interactively.

5 Future research

The research reported in this paper is still in progress. Especially the question of what kind of correlations between more than two dimensions can be visualized by using higher order parallel coordinates is still under research. Also the question of finding an appropriate start sequence of main and additional axes is open. Nevertheless we found the results up to now sufficiently promising to continue research in this area.

Acknowledgements: The author would like to thank Prof. Heidrun Schumann from Rostock University for her constant encouragement and support of the research presented in this paper. Thanks also to Martin Köller for developing a testbed for higher order parallel coordinates.

References

[1] Gerald Farin. Curves and Surfaces for Computer Aided Geometric Design. Fourth edition. Akademic Press, 1997.

[2] A. Inselberg, B. Dimsdale. Parallel Coordinates: A Tool for Visualizing multi-dimensional Geometry. Proceedings IEEE Visualization 90, pp. 361-375, Los Alamitos, 1990.

[3] Holger Theisel. Analyse und Visualisierungshilfe für mehrdimensionale wissenschaftliche Daten. Informatik Forschung und Entwicklung 10(2), pp. 91-98. Springer, 1995 (in German).

[4] R. Wegenkittl, H. Löffelmann, E. Gröller. Visualizing the Behavior of Higher Dimensional Dynamical Systems. Proceedings IEEE Visualization 97, pp. 119-125, Los Alamitos, 1997.

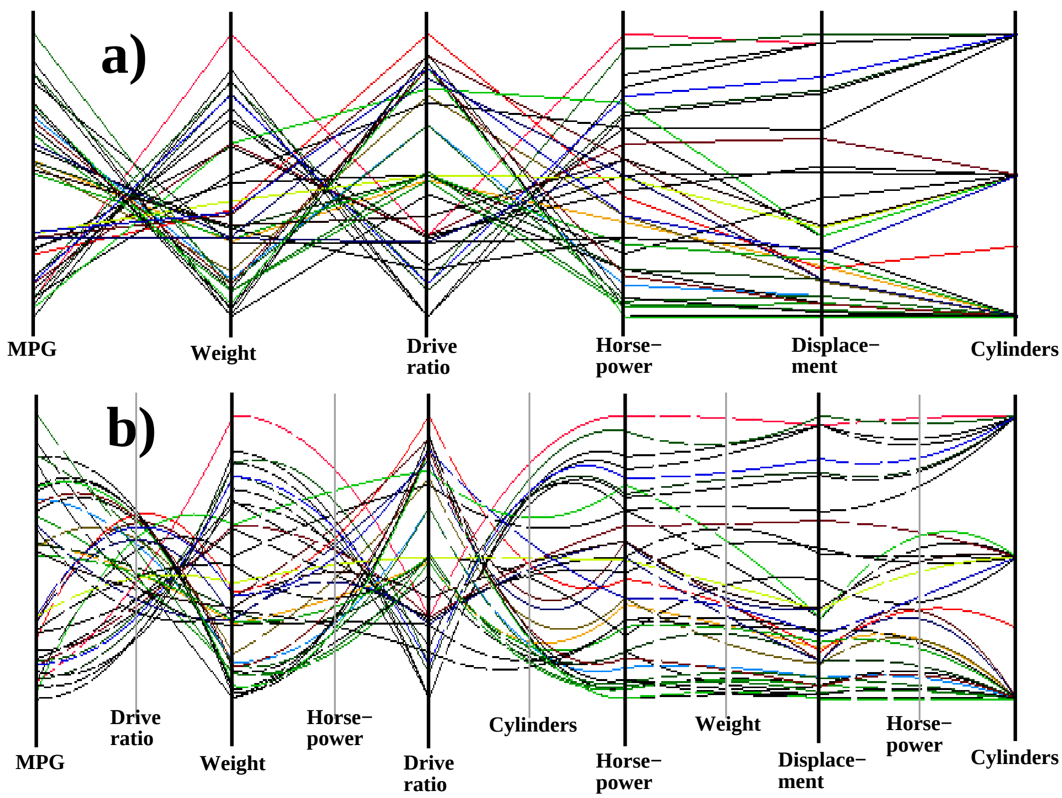

Figure 9: Car data set; **a)** usual parallel coordinates; **b)** higher order parallel coordinates with one additional axis between each pair of adjacent main axes