Les Cahiers du GERAD ISSN: 0711–2440

Parallel coordinate order for high-dimensional data

S. Tilouche, V. Partovi Nia,
S. Bassetto

G–2017–38

May 2017

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2017
– Bibliothèque et Archives Canada, 2017

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2017
– Library and Archives Canada, 2017

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Parallel coordinate order for high-dimensional data

Shaima Tilouche [a]
Vahid Partovi Nia [b]
Samuel Bassetto [a]

[a] *Department of Mathematics and Industrial Engineering, Polytechnique Montréal (Québec) Canada, H3C 3A7*

[b] *GERAD & Department of Mathematics and Industrial Engineering, Polytechnique Montréal (Québec) Canada, H3C 3A7*

shaima-2.tilouche@polymtl.ca
vahid.partovinia@polymtl.ca
samuel-jean.bassetto@polymtl.ca

May 2017
Les Cahiers du GERAD
G–2017–38

Abstract: Visualization of high-dimensional data is counter-intuitive using conventional graphs. Parallel coordinates is proposed, as an alternative, to explore multivariate data more effectively. However, when the data are high-dimensional with thousands of lines overlapping, it is difficult to extract relevant information through the parallel coordinates. The order of the axes determines the perception of information on parallel coordinates. Thus, if coordinates are improperly ordered, the information between attributes remain hidden. Here we propose a general framework to reorder the coordinates. This framework depend on the objective of data visualization. It is also flexible to contain many conventional ordering measures. We also present the coordinate ordering binary optimization problem and enhance towards a computationally efficient greedy approach that suites high-dimensional data. Our approach is applied on wine data and on genetic data.

Keywords: Dimension ordering, high-dimensional data, Kullback-Liebler divergence, parallel coordinates

1 Introduction

When data are high-dimensional, representing each attribute marginally may lead to an incomplete or unclear visualization. Multidimensional graphs such as scatter plot matrices, glyphs, and parallel coordinates are proposed to facilitate multivariate data exploration. Here we focus on parallel coordinates which D'Ocagne [9] invented, primarily as a two-dimensional diagram to approximate the graphical computation of a mathematical function using *nomogram*. Parallel coordinates are further studied in [17] to allow the visualization of multidimensional data on a transformed two-dimensional space [3].

Suppose the data matrix contains n observations in rows and p attributes in columns. A common data visualization representation is scatter plot of data in orthogonal coordinates, where each axis is an attribute and each observation is a point. This representation is limited to maximum $p = 3$ attributes. In parallel coordinates representation, axes are parallel lines and each observation is a line, passing through each coordinate [2]. This technique extends data visualization for $p > 3$.

Several parallel coordinates software have been developed so far. Some of them like *XDAT* and *XMDVTool* are interactive and some others like *Statistica* and *ggparallel* R package are not. Software visualization tools mostly provide options such as applying filters, data clustering, and switching coordinates for a better visualization. Theoretically, there is no limit on the number of observations or the number of attributes. However, when the number of observations is large, many lines overwhelm the display, and the parallel coordinate graph becomes dense to analyze visually. On the other hand, high-dimensional data contains large number of attributes, leading to a wide and an unclear representation.

Several techniques are proposed to improve the visual exploration of data in parallel coordinates. These techniques aim to reorder attributes, so that data exploration becomes more straightforward. These techniques aim to highlight relations between attributes and to reduce data clutter. Our framework attempts to introduce a criterion that adapts to the purpose of parallel coordinate visualization. Here, we mainly focus on two purposes, exploring the dependence between attributes and the data separation. However, our technique is flexible and can be adapted for other purposes like outlier detection. If the purpose of visualization is exploring the dependence between attributes, then the criterion mimics a sort of correlation. If clustering is of interest, then the criterion measures the separation.

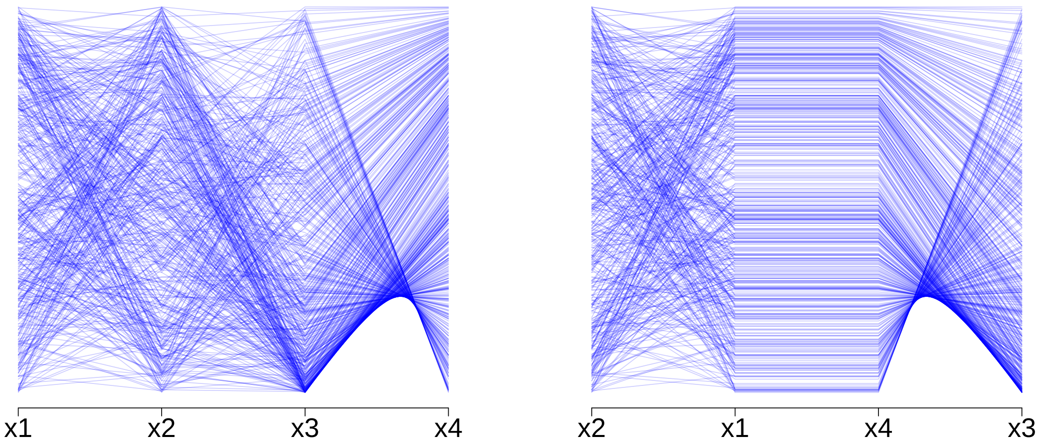
Figure 1 shows the impact of data reordering on highlighting the dependence between attributes even in the case of a small number of attributes, both for dependence and clustering purposes. The left panel shows only the relation between x_3 and x_4. With a coordinate reordering, two relationships appear, one between x_1 and x_4 and another between x_3 and x_4. Figure 2 shows that dimension reordering enhances cluster detection. In the left panel, data are separable only by x_3 and x_4. However, with a proper reordering, the same data are separable by x_1 and x_4 as well.

The paper is structured as follows. Section 2 summarizes some coordinate reordering techniques and demonstrates the duality between orthogonal coordinates and parallel coordinates. Section 3 introduces our coordinate reordering criterion. Section 4 proposes an optimization algorithm for reordering attributes. The first application is carried out on the wine quality dataset in Subsection 5.1 to explore the dependence with relatively small number of attributes. Another application of our method is shown in Subsection 5.2 on high-dimensional genetic data to explore data separation.

2 Coordinate order

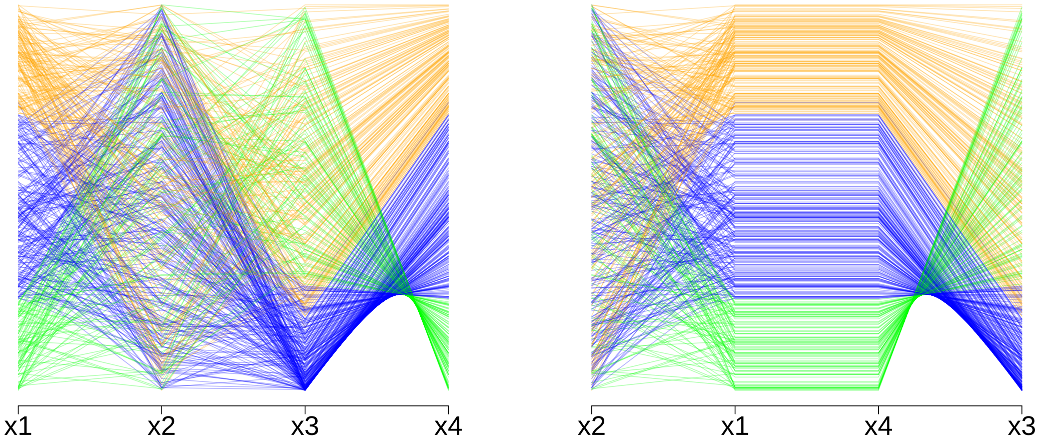
The order of coordinates has a visible impact on the perception of data structure including the visualization of attribute dependence and the detection of clusters. [24] points out the parallel coordinate display visualizes the inter-coordinate dependence between neighboring dimensions, but does not reveal the dependence between non-adjacent coordinates—clearly emphasizing on the importance of coordinate order.

Coordinate reordering helps highlighting data dependencies, promotes visual data mining, and facilitates data exploration. Figure 1 shows an example of four-dimensional data in its original order and after being reordered properly. Figure 1b shows that there is a linear relation between x_1 and x_3 which is not visible



(a) **(b)**

Figure 1: Parallel coordinate graph when the axes are improperly ordered (left panel) versus properly ordered to explore attribute dependence (right panel).



(a) **(b)**

Figure 2: Parallel coordinate graph when the axes are improperly ordered (left panel) versus properly ordered to explore data separation (right panel).

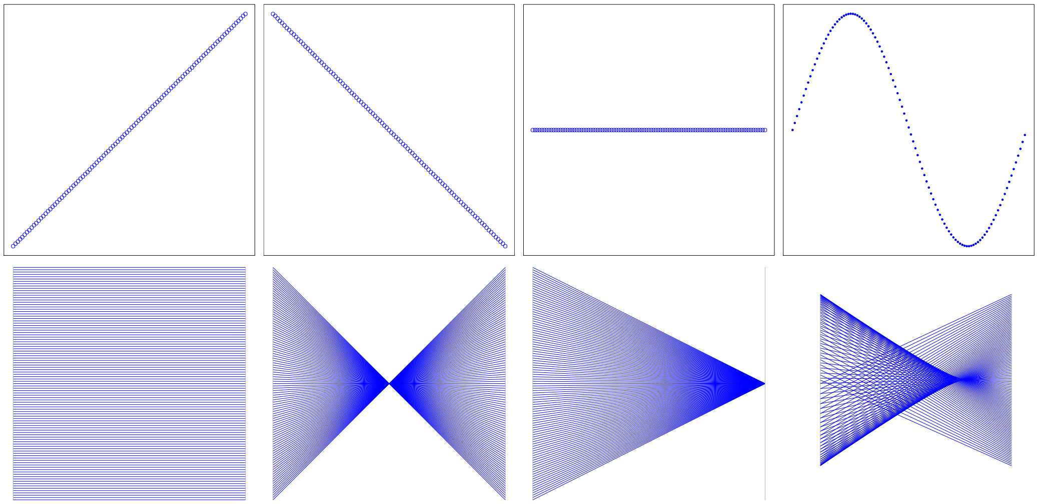
in Figure 1a. This relation is detected through many parallel lines between the two coordinates. Further examples on other dependence are presented in Figure 3. Interactive software enable manual attribute reordering. Users can change the order of attributes by switching axes. Handling the order manually is time consuming, and some relations still may remain undetected. Developing an automatic technique seems essential for a good visualization, specially for large number of attributes.

Some authors proposed automatic techniques to find the best order for data visualization. The proposed techniques focus on highlighting the dependence among attributes. They aim to put a attribute in the neighborhood of the most dependent attribute. For instance, [4] proposes a technique to minimize the dissimilarities or partial dissimilarities between two adjacent attributes. The dissimilarity is often measured through the Euclidean distance over a pair of attributes. It is not difficult to see that minimizing the Euclidean distance coincides with maximizing the squared correlation. Unfortunately, correlation (or Euclidean distance) is unreliable to uncover all types of dependencies. Correlation is a deficient measure to uncover non linear dependencies [6]. For instance if x follows a symmetric distribution such as Gaussian, the correlation between x and x^2 is zero.

[24] proposes another technique for coordinate reordering, which aims to reorder by minimizing outliers between two neighboring coordinates. An observation is considered as an outlier if it involves no neighboring data. The neighbor is defined by the Euclidean distance after applying a certain threshold. This technique is sensitive to the chosen threshold. [19] suggests a reordering technique using variety of metrics, e.g. maximizing correlation, reducing the number of clutters, etc. This approach gives an effective visualization and exploration of structures within a large multivariate data set, and meanwhile provides enhancement of diverse structures by supplying a range of automatic variable orderings. [11] suggests subspace clustering and coordinate ranking. [7] proposes several reordering metrics such as number of crossing lines, angle of crossing, and mutual information. Another algorithm is independently developed using a genetic algorithm, to highlight important features and allow the detection of irregularities using Pearson correlation [5]. [20] combines singular value decomposition to select the attributes that have the highest contribution and then applies a nonlinear correlation coefficient to order the axes.

The perception of patterns and clusters depends on the choice of the coordinate system. Therefore, it is important to know how to read the coordinate system. Despite the spread of parallel coordinates between practitioners, it is still unknown to many researchers in academica, especially when it comes to the interpretation of the shapes observed in parallel coordinates. Some authors were interested in the graphic transformation from orthogonal coordinates to parallel coordinates already. [17] states that the representation in parallel coordinates is a projective transformation of orthogonal coordinates. [15] studies the transformation of a linear function to parallel coordinates in more details.

Some other dualities are studied in [17] and [27]. In point-line duality, some other mappings can be expressed using the envelope of lines in parallel coordinates [15]. Here we do not review the mathematical details, but rather focus on visual aspects. In Figure 3, some common functions are drawn in orthogonal coordinates and in parallel coordinates.



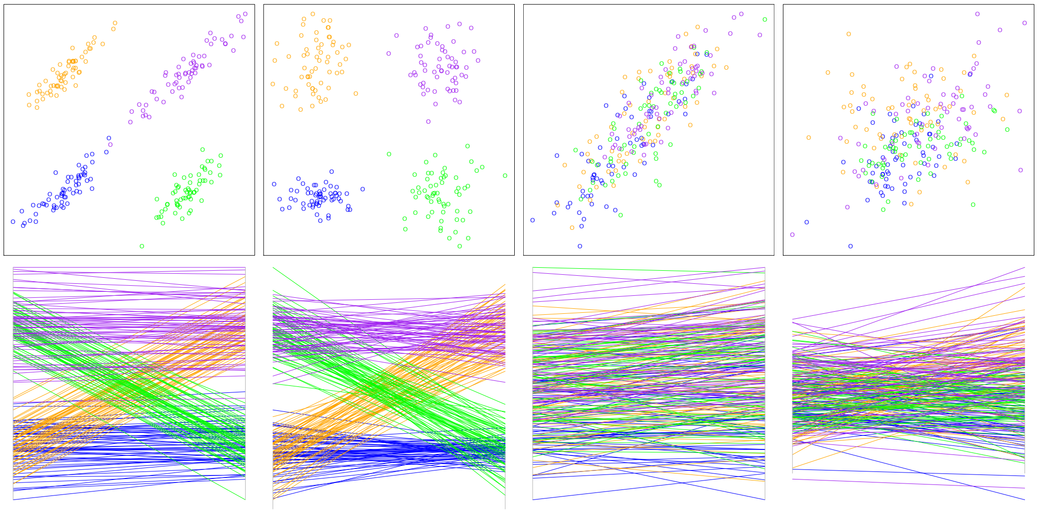
(a) $y = x$ **(b)** $y = -x$ **(c)** $y = 0$ **(d)** $y = \sin(x)$

Figure 3: The duality between the Orthogonal coordinates (top) and the parallel coordinates (bottom) for 4 common functions.

A set of points located on a line is represented in parallel coordinates by a set of lines that intersect at a definite point. The horizontal position of this point depends on the slope of the linear function. If the slope is negative, the intersection point is located between the parallel axes (Figure 3b). Different patterns are observed in a linear function with a slope superior to 1, or inferior to 1. However, as most software normalize the data before treatment, only parallel lines appears for a positive slope. This is illustrated in Figure 3a and Figure 3b.

Figure 3c shows a constant function. This function is illustrated by a set of lines that converge to a single point. A periodic function is translated by 2 sets of lines intersecting in 2 different points as in Figure 3d. Detecting the functions using the parallel coordinate shapes is still confusing, because some shapes resemble. The main difference is in the intersecting points. Despite this uncertainty in the interpretation of some shapes, it is clear that when two attributes are dependent, the parallel coordinate graph shows a certain pattern.

Cluster visualization is different in orthogonal coordinates and in parallel coordinates. Figure 4 illustrates the separation and correlation in both coordinate systems. Figure 4a shows separable and correlated data. The clusters are visible and some patterns appears in parallel coordinates. These patterns translate set of linear functions with different coefficients to set of lines. Figure 4b shows separable and uncorrelated data. The patterns are not much different than Figure 4a, but, the clusters are more distinguishable. Figure 4b translates correlated but non-separable data, and Figure 4d illustrates non-separable and uncorrelated data.



(a) Separable and correlated data **(b) Separable non correlated data** **(c) Non separable correlated data** **(d) Non separable non correlated data**

Figure 4: Separation and correlation in orthogonal coordinates (top panel) and in parallel coordinates (bottom panel).

3 General information criterion

Various methods are used to order coordinates, from Euclidean distance to correlation. As only two coordinates are visualized at a time, it looks promising to order coordinates through some measures defined over the bivariate data distribution. Take two arbitrary attributes, say x_1, x_2. Define two hypothetical bivariate probability measures over the product of their sample space, and over the same sigma algebra $\mathcal{F}$. In other words, define two probability spaces $\big(\Omega, \mathcal{F}, F\big)$, and $\big(\Omega, \mathcal{F}, H\big)$ for (x_1, x_2). For the simplicity of notation we denote the probability measures F and H by their imposed distribution functions $F(x_1, x_2)$ and $H(x_1, x_2)$. Let $F(x_1, x_2)$ and $H(x_1, x_2)$ impose different probability measures, i.e.

$$\exists (x_1, x_2) \in \mathbb{R}^2 \text{ such that } F(x_1, x_2) \neq H(x_1, x_2).$$

Define the *general information* as

$$\mathrm{GI}(x_1, x_2) = \frac{1}{G''(1)} \int\int G\left\{\frac{dF(x_1, x_2)}{dH(x_1, x_2)}\right\} dH(x_1, x_2), \tag{1}$$

where $dF(x_1, x_2)/dH(x_1, x_2)$ is the Radon-Nikodym derivative, $G(.)$ is a univariate smooth function and $G''(1)$ is the second derivative of $G(.)$ at 1. The second derivative $G''(1)$ in (1) adjusts for scaling. Criterion (1) is closely related to the Kullback-Leibler divergence, the cross entropy, and the joint entropy.

The choice of F relative to H defines the measuring concept and the choice of $G(.)$ defines the measuring statistic. A common choice of F and H is the data joint distribution and the product of marginal data distributions, respectively. In this case, the measuring concept reduces to dependence. The Pearson correlation as a measure of dependence arises if $F(x_1, x_2)$ is bivariate Gaussian.

A common choice of $G(.)$ is $G(u) = u \log(u)$ which brings the Kullback-Leibler divergence of F relative to H. Our suggestion for $G(u)$ is a univariate function that

i) vanishes at 1, i.e. $G(1) = 0$,

ii) its first derivative is smooth at 1 , i.e. $|G''(u)|$ is bounded in an infinitesimal neighborhood $u \in (1 - \epsilon, 1 + \epsilon)$.

The first condition ensures that GI is well-defined. In other words, GI $= 0$ if and only if the reference probability measures F and H coincide. The second condition ensures the asymptotic statistical behavior of GI as the number of observations n increases.

One may choose the statistic of interest by varying $G(u)$. It is easier to understand the role of $G(.)$ in the context of discrete random variables. If (x_1, x_2) is a pair discrete random variables, $H(x_1, x_2) = F(x_1)F(x_2)$, then various famous statistics of contingency tables are derived by varying $G(u)$

- $G(u) = 2u \log u$ gives the log likelihood ratio statistic,
- $G(u) = (u - 1)^2$ gives the Pearson chi-square statistic,
- $G(u) = u(1 - 1/\sqrt{u})$ gives the Freeman-Tukey statistic,
- $G(u) = (1 - u)^2/u$ gives the Neyman statistic,
- $G(u) = u(\sqrt[3]{u^2} - 1)$ gives the Cressie-Read statistic,

and more importantly $G(u) = u \log u$ is the mutual information

$$\mathrm{GI}(x_1, x_2) = \sum_{x_1} \sum_{x_2} p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_1)p(x_2)},$$

where $p(x_1, x_2)$ is the joint probability mass, $p(x_1)$ and $p(x_2)$ are the marginal masses.

If visualization towards data clutter is preferred, it is more meaningful to measure the separation instead of dependence. Therefore, one may define $F(x_1, x_2)$ to be a k component distribution

$$dF(x_1, x_2) = \sum_{c=1}^{k} p_c g_{\boldsymbol{\mu}_c}(x_1, x_2) dx_1 dx_2 \tag{2}$$

and $H(x_1, x_2)$ to be a single component distribution

$$dH(x_1, x_2) = g_{\boldsymbol{\mu}}(x_1, x_2) dx_1 dx_2, \tag{3}$$

where $g_{\boldsymbol{\mu}}(., .)$ is a density family indexed by the location parameter $\boldsymbol{\mu}$. Such a measure mimics the silhouettes [25] if g is Gaussian bivariate density.

Ordering with respect to outliers is feasible through assigning a heavy-tailed, such as the Student's t-distribution, for F and a bivariate Gaussian for H. Many other concepts such as dispersion, skewness, prediction power, multi-collinearity, etc, can be quantified through the general information criterion (1), and then be used to order the coordinates for further visual inspection.

4 Order optimization

Suppose data contain p attributes. The total number of coordinate permutation is $p!$ impossible to check visually for large p. It is natural to put the most informative coordinates early in the graph. This is specially helpful while data are high-dimensional.

Suppose the general information matrix, call the symmetric weight matrix, is computed for all pairs of attributes $\mathbf{W}_{p\times p} = [\mathrm{GI}(x_i, x_j)]$. The problem of finding optimal neighboring coordinates is reduced to estimation of a binary symmetric adjacency matrix $\mathbf{A} = [a_{ij}]$ that maximizes the total information

$$\hat{\mathbf{A}} = \operatorname{argmax} \|\mathbf{A} \odot \mathbf{W}\| \tag{4}$$

$$\text{s.t.}$$

$$a_{ij} = 0 \text{ or } a_{ij} = 1 \tag{5}$$

$$\mathbf{a}_i^\top \mathbf{1} = \mathbf{1}^\top \mathbf{a}_j = 2 \tag{6}$$

$$a_{ij} = a_{ji}, \tag{7}$$

$$\|\mathbf{A}\| \leq 2q \tag{8}$$

where $\mathbf{a}_i^\top$ is the ith row of $\mathbf{A}$, $\mathbf{a}_j$ is the jth column of $\mathbf{A}$, $\odot$ is the Hadamart product, and $\|\mathbf{A}\| = \sum_i \sum_j |a_{ij}|$ is the L_1 Frobenius norm.

The objective function $\sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} w_{ij}$ in (4) computes the utility of incorporating some adjacent coordinates. The constraint (5) ensures whether or not a coordinate is neighbor to another. The constraint (6) ensures a coordinate is neighbor to only two other coordinates. The constraint (7) imposes symmetry on the adjacency matrix. The constraint (8), for a $q < p$, selects only q out of p coordinates for visualization.

Standard solvers such as CPLEX can be used to solve this integer-linear optimization program after fixing q. If $q \geq p$, the integer program only finds the adjacent coordinates and relaxes the selection. For high-dimensional data, this optimization is cumbersome to solve even with powerful computers. We propose a faster algorithm by optimizing the objective function (4) hierarchically as follows.

The first pair of coordinates are the one that maximize the objective function at the first iteration

$$\begin{array}{rcl} (\hat{x}_1, \hat{x}_2) & = & \operatorname{argmax} \mathrm{GI}(x_i, x_j) \\ 1 \leq i \leq p-1 & & i+1 \leq j \leq p. \end{array} \tag{9}$$

The jth, $j = 3, \ldots, q$ coordinates is

$$\begin{array}{c} \hat{x}_j = \operatorname{argmax} \mathrm{GI}(\hat{x}_{j-1}, x_i), \\ i \in \{1, \ldots, p\} \backslash \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{j-1}\}. \end{array} \tag{10}$$

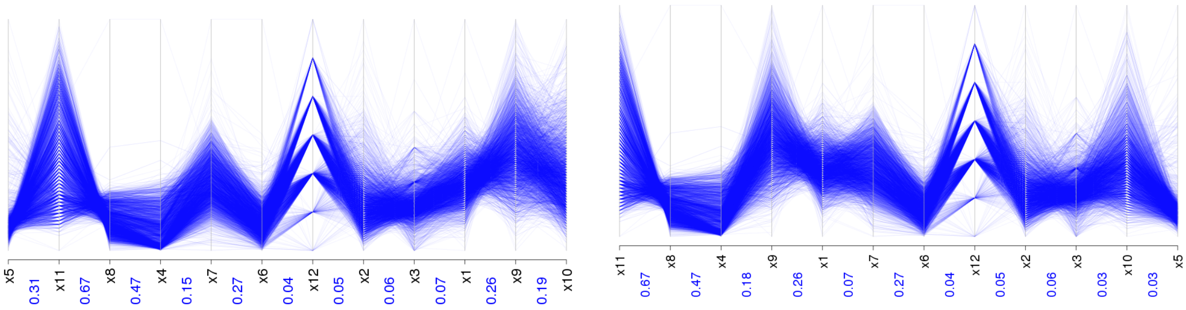
The computation of this greedy algorithm is of time complexity $O(p^2)$ and dominated by the first step of the algorithm (9). A faster algorithm of order $O(qp)$ can be achieved by fixing the first coordinate manually and order the remaining coordinates using (10). This technique is scalable with the number of coordinates p, specially for high-dimensional data while $q \ll p$.

5 Application

To test the proposed algorithm, we used two well-known datasets. The first is the white wine quality data [1]. This dataset includes 12 attributes. The second dataset is Golub genetic data [13]. It is a high-dimensional data and only 50 attributes out of 2030 are selected for visualization.

5.1 Wine dataset

These data are the result of a chemical analysis of white wines taken from [1]. The data include 4898 observation measurements over 12 attributes: fixed acidity (x_1), volatile acidity (x_2), citric acid (x_3), residual sugar (x_4), chlorides (x_5), free sulfur dioxide (x_6), total sulfur dioxide (x_7), density (x_8), pH (x_9), sulphates (x_{10}), alcohol (x_{11}) and quality (x_{12}, a score between 0 and 10). This dataset is analyzed for several reasons such as outlier detection, classification, and regression. [7] used this database to evaluate the dimension reordering techniques in parallel coordinates using crossing angles and mutual information. First, the optimal order for mutual information of problem was compared to the solution found by greedy algorithm. The optimization problem was solved using IBM ILOG CPLEX Optimization Studio 12.7.1 a 2.20 GHz Intel core i7-2702MQ processor and 16.00 Go RAM taking around 17 seconds compared to around 1 second for our greedy algorithm. The optimal solution given by CPLEX is a circle-like neighborhood matrix. To transform this neighborhood matrix it into a list, the circle is cut at the pair with the minimum mutual information. Figure 5 presents a comparison between the order given by CPLEX and the order given with our greedy algorithm. Many pairs of adjacent attributes appear in both panels (x_8, x_{11}), (x_8, x_4), (x_7, x_6), (x_6, x_{12}), (x_{12}, x_2), (x_2, x_3), and (x_1, x_9).



(a) Order with CPLEX, $\sum_i \sum_j \mathrm{GI}(x_i, x_j) = 2.53$.

(b) Order with greedy algorithm, $\sum_i \sum_j \mathrm{GI}(x_i, x_j) = 2.13$

Figure 5: Comparison between the order found with CPLEX (top panel) and the order with our Greedy algorithm (bottom panel). The blue values between a pair of coordinates are $\mathrm{GI}(x_i, x_j)$**.**

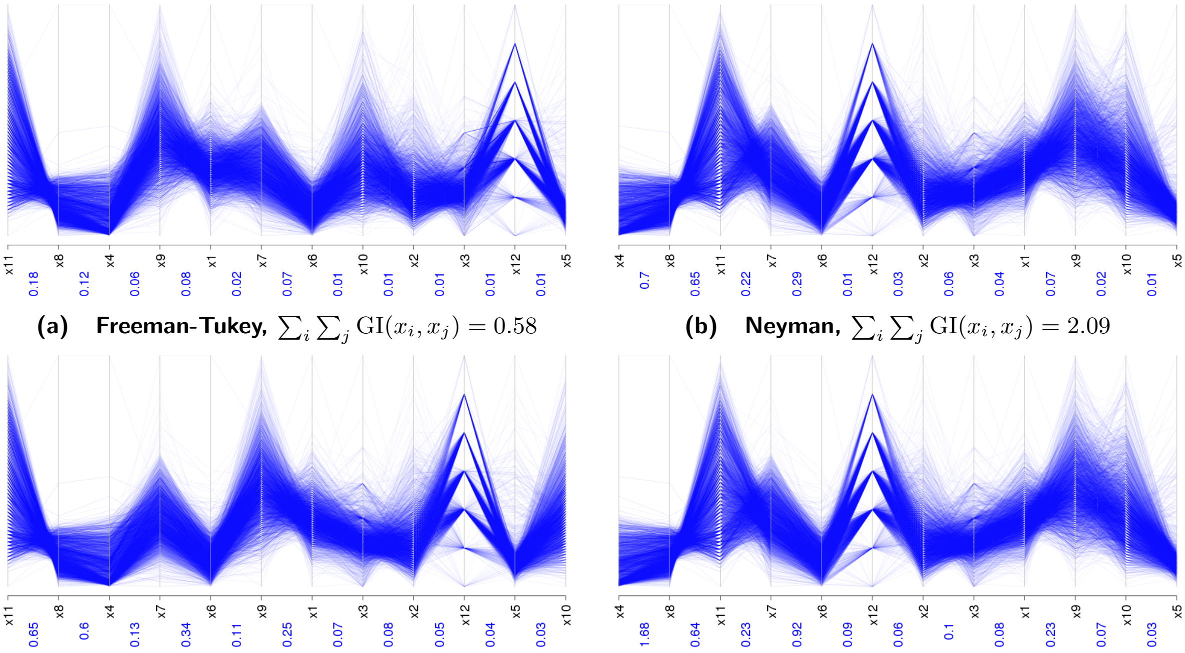
These data are used also to evaluate the impact of changing the statistic by varying $G(.)$ when the concept measure (F relative to H) is set to dependence. Therefore, F is set to the joint probability, $F(x_1, x_2)$, and H is set to the product of probability masses, $F(x_1)F(x_2)$.

The results are illustrated in Figure 6. The values between each 2 adjacent attributes are the numerical values of the criterion. All the algorithms started with the highest information and tend to decrease.

The order changes as the statistic varies, compare for instance Figure 6a with Figure 6d. The first 3 coordinates are ordered similarly by Mutual information, Cressie and Freeman-Tukey. Again, Mutual information and Tukey-Freeman selected the same 7 first attributes and give a different order for the last 5 attributes. In statistics literature it is known that the behaviour of Neyman and Pearson statistics are alike. Here, Neyman and Pearson statistics give exactly the same order. Tukey statistic starts with a different attribute. However, Cressie-Read represents the dependence on attributes along with other statistics, for instance (x_2, x_12) and $(x7, x6)$ are also represented by Pearson statistic.

Comparing the total information of each statistic, $\sum_i \sum_j \mathrm{GI}(x_i, x_j)$, shows that Pearson gives the highest value of 4.13, followed by Cressie, mutual information, Neyman, with total information around 2, and Freeman statistic with a total information of 0.58. As the Pearson statistic provides the highest total information between adjacent attributes, we suggest to use Pearson statistic to reorder attributes for this data set.

The order proposed by all criteria places the more dependent attributes first ending with nearly independent attributes. Through this data example, we notice that changing the criterion change the order globally, however many coordinates are placed in the neighborhood of one another overall.



(a) Freeman-Tukey, $\sum_i \sum_j \mathrm{GI}(x_i, x_j) = 0.58$

(b) Neyman, $\sum_i \sum_j \mathrm{GI}(x_i, x_j) = 2.09$

(c) Cressie-Read, $\sum_i \sum_j \mathrm{GI}(x_i, x_j) = 2.39$

(d) Pearson chi-square, $\sum_i \sum_j \mathrm{GI}(x_i, x_j) = 4.13$

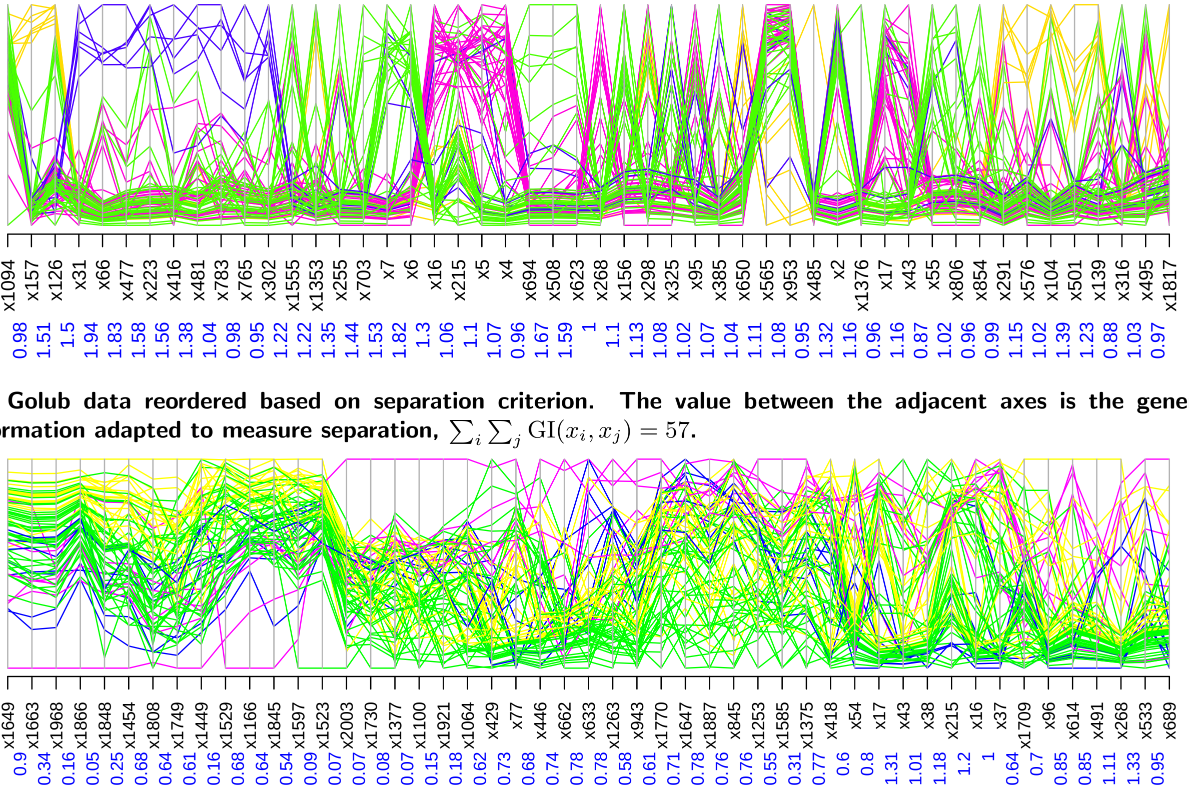
Figure 6: Wine data reordered based on different statistics. $\sum_i \sum_j \mathrm{GI}(x_i, x_j)$ **is the sum of the values written between the pair of coordinates.**

5.2 Genetic dataset

We applied the developed approach to [13]. Golub dataset consists of 47 patients with acute lymphoblastic leukemia (ALL) and 25 patients with acute myeloid leukemia (AML). The observations have been assayed with Affymetrix Hgu6800 chips, resulting in 7129 gene expressions (Affymetrix probes). The data was preprocessed, giving 2030 attributes [21]. This data is high-dimensional so, selecting the most informative attribute subset is crucial. Finding the genes that separates the data are more appealing than dependence in genetic application. Therefore, we apply the separation statistic described earlier, by choosing F to be a bivariate k-component Gaussian (2), and H to be a single component Gaussian (3).

The visualized dimensions are those which maximize the criterion of the list. To find the appropriate order, we tried to run the optimization algorithm (4) using CPLEX, but it did not converge for $q = 50$. Therefore, we only present the result of our greedy method. To improve the computational complexity of the greedy algorithm, the first attribute is selected to be the one with the highest univariate separation criterion. Fixing the first attribute avoids computation of general information criterion (1) for all pairs of attributes. This is a huge gain while data are high dimensional.

When the number of clusters is not known, we suggest to use a large number of clusters for the reordering and then adjust the colors for a better visualization. This data are clustered into 7 clusters. Then, 3 small clusters are re-grouped to visualize only 4 groups. Figure 7b illustrates the results. The top panel shows clustered data reordered based on separation metric and the bottom panel shows clustered data reordered based on Pearson correlation. It is clear that for the purpose of cluster detection, separation criterion highlights the data separation more clearly. The sum of separation criterion is around 57 for the order found based on the separation criterion and 30 for the order based on Pearson correlation. It is natural to expect that the total information for separation is higher when the attributes are reordered for that purpose. Not only the total information, but also the parallel coordinate graph clearifies the effect of choosing the right measure for the visualization purpose. The result confirms that when the purpose of reordering is data separation, or cluster detection as discussed in Section 1, then, F and H needs to defined in the direction of visualization purpose.



(a) Golub data reordered based on separation criterion. The value between the adjacent axes is the general information adapted to measure separation, $\sum_i \sum_j \mathrm{GI}(x_i, x_j) = 57$.

(b) Golub data reordered based on Pearson correlation. The value between the adjacent axes is the general information adapted to measure separation, $\sum_i \sum_j \mathrm{GI}(x_i, x_j) = 30$.

Figure 7: Golub data reordered based on separation criterion and on Pearson correlation.

6 Conclusion

This paper presents a new and a general framework for coordinate ordering. The new framework is general enough to cover many existing ordering methods. This framework is based on a general information criterion defined to cover wide range of ordering measures. A computationally efficient ordering algorithm is developed to cover high-dimensional data visualization. The tests showed that according to the purpose of the reordering, the criterion and the statistic need to be chosen appropriately in order to achieve a useful coordinate order.

References

[1] "Wine data set." [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Wine

[2] H. Albazzaz and X. Z. Wang, "Historical data analysis based on plots of independent and parallel coordinates and statistical control limits," Journal of Process Control, 16(2) 103–114, 2006.

[3] H. Albazzaz, X. Z. Wang, and F. Marhoon, "Multidimensional visualisation for process historical data analysis: a comparative study with multivariate statistical process control," Journal of Process Control, 15(3) 285–294, 2005.

[4] M. Ankerst, S. Berchtold, and D. A. Keim, "Similarity clustering of dimensions for an enhanced visualization of multidimensional data," in Information Visualization, 1998. Proceedings. IEEE Symposium on. IEEE, 1998, 52–60.

[5] T. Boogaerts, L. Tranchevent, A. Pavlopoulos, G., J. Aerts, and J. Vandewalle, "Visualizing high dimensional datasets using parallel coordinates: Application to gene prioritization," in Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on. IEEE, 2012, 52–57.

[6] C. J. Cellucci, A. M. Albano, and P. E. Rapp, “Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms,” Physical Review E, 71(6) p. 066208, 2005.

[7] A. Dasgupta and R. Kosara, “Pargnostics: Screen-space metrics for parallel coordinates,” IEEE Transactions on Visualization and Computer Graphics, 16(6) 1017–1026, 2010.

[8] A. Dix and G. Ellis, “By chance enhancing interaction with large data sets through statistical sampling,” in Proceedings of the Working Conference on Advanced Visual Interfaces. ACM, 2002, 167–176.

[9] M. d’Ocagne, Coordonnées parallèles & axiales: méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles. Gauthier-Villars, 1885.

[10] G. Ellis and A. Dix, “Density control through random sampling: an architectural perspective,” in Sixth International Conference on Information Visualisation. IEEE, 2002, 82–90.

[11] B. J. Ferdosi and J. B. Roerdink, “Visualizing high-dimensional structures by dimension ordering and filtering using subspace analysis,” in Computer Graphics Forum, 30(3). Wiley Online Library, 2011, 1121–1130.

[12] Y. H. Fua, M. O. Ward, and E. A. Rundensteiner, “Hierarchical parallel coordinates for exploration of large datasets,” in Proceedings of the conference on Visualization’99: celebrating ten years. IEEE Computer Society Press, 1999, 43–50.

[13] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, C. Bloomfield, and E. Lander, “Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.” Science, 286(5439) 531–537, 1999.

[14] H. Hauser, F. Ledermann, and H. Doleisch, “Angular brushing of extended parallel coordinates,” in IEEE Symposium on Information Visualization, INFOVIS 2002. IEEE, 2002, 127–130.

[15] J. Heinrich and D. Weiskopf, “State of the art of parallel coordinates,” STAR Proceedings of Eurographics, 2013, 95–116, 2013.

[16] J. Hlinka, D. Hartman, M. Vejmelka, D. Novotna, and M. Palus, “Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity,” Climate dynamics, 42(7-8) 1873–1886, 2014.

[17] A. Inselberg, “The plane with parallel coordinates,” The Visual Computer, 1(2) 69–91, 1985.

[18] J. Johansson, P. Ljung, M. Jern, and M. Cooper, “Revealing structure within clustered parallel coordinates displays,” in IEEE Symposium on Information Visualization, INFOVIS 2005. IEEE, 2005, 125–132.

[19] S. Johansson and J. Johansson, “Interactive dimensionality reduction through user-defined combinations of quality metrics,” IEEE Transactions on Visualization and Computer Graphics, 15(6) 993–1000, 2009.

[20] L. Lu, M. Huang, and J. Zhang, “Two axes re-ordering methods in parallel coordinates plots,” Journal of Visual Languages & Computing, 33, 3–12, 2016.

[21] P. D. McNicholas and T. B. Murphy, “Model-based clustering of microarray expression data via latent gaussian mixture models,” Bioinformatics, 26(21) 2705–2712, 2010.

[22] G. M. Oyeyemi, “Principal component chart for multivariate statistical process control,” The Online Journal of Science and Technology, 1(2) 2011.

[23] G. Palmas, M. Bachynskyi, A. Oulasvirta, H. P. Seidel, and T. Weinkauf, “An edge-bundling layout for interactive parallel coordinates,” in Visualization Symposium (PacificVis), 2014 IEEE Pacific. IEEE, 2014, 57–64.

[24] W. Peng, M. O. Ward, and E. A. Rundensteiner, “Clutter reduction in multi-dimensional data visualization using dimension reordering,” in IEEE Symposium on Information Visualization, 2004. INFOVIS 2004. IEEE, 2004, 89–96.

[25] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” Journal of computational and applied mathematics, 20, 53–65, 1987.

[26] J. Shearer, M. Ogawa, K.-L. Ma, and T. Kohlenberg, “Pixelplexing: Gaining display resolution through time,” in IEEE Pacific Visualization Symposium, PacificVIS’08. IEEE, 2008, 159–166.

[27] E. J. Wegman, “Hyperdimensional data analysis using parallel coordinates,” Journal of the American Statistical Association, 85(411) 664–675, 1990.

[28] P. C. Wong and R. D. Bergeron, “Multiresolution multidimensional wavelet brushing,” in Visualization ’96 Proceedings. IEEE, 1996, 141–148.

[29] H. Zhou, W. Cui, H. Qu, Y. Wu, X. Yuan, and W. Zhuo, “Splatting the lines in parallel coordinates,” in Computer Graphics Forum, 28(3). Wiley Online Library, 2009, 759–766.

[30] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen, “Visual clustering in parallel coordinates,” in Computer Graphics Forum, 27, no. 3. Wiley Online Library, 2008, 1047–1054.