Using Penalized Regression with Parallel Coordinates for Visualization of Significance in High Dimensional Data

Shengwen Wang

Cloud Computing and System Integration Division
National Center for High-Performance Computing
Taichung, Taiwan

Yi Yang

School of Statistics
University of Minnesota
Minneapolis, USA

Jih-Sheng Chang

Cloud Computing and System Integration Division
National Center for High-Performance Computing
Taichung, Taiwan

Fang-Pang Lin

Cloud Computing and System Integration Division
National Center for High-Performance Computing
Hsinchu, Taiwan

Abstract—In recent years, there has been an exponential increase in the amount of data being produced and disseminated by diverse applications, intensifying the need for the development of effective methods for the interactive visual and analytical exploration of large, high-dimensional datasets. In this paper, we describe the development of a novel tool for multivariate data visualization and exploration based on the integrated use of regression analysis and advanced parallel coordinates visualization. Conventional parallel-coordinates visualization is a classical method for presenting raw multivariate data on a 2D screen. However, current tools suffer from a variety of problems when applied to massively high-dimensional datasets. Our system tackles these issues through the combined use of regression analysis and a variety of enhancements to traditional parallel-coordinates display capabilities, including new techniques to handle visual clutter, and intuitive solutions for selecting, ordering, and grouping dimensions. We demonstrate the effectiveness of our system through two case-studies.

Keywords—Parallel Coordinates; High-Dimensional Data; Multivariate Visualization; LASSO Regression; Penalized Regression; Dimension Reordering

I. INTRODUCTION

Parallel coordinates is a popular method for exploring multidimensional data. In this method, high-dimensional data are displayed as points on a series of parallel coordinate axes, and relations between pairs of neighboring dimensions are revealed by the pattern of connecting lines between them. Much effort has been dedicated to investigating the properties of high-dimensional data, and to defining appropriate clusters in these data. Previous methods have also explored possible solutions to some of the common problems of parallel coordinates, including visual clutter, dimension space navigation, context and detail enhancement, etc. Attempts to overcome these problems have been made both at the stages of data processing and visualization; however, the current existing solutions may not always be sufficient to solve all of the related issues completely.

In this paper, we describe the development of a tool that is capable of assisting users to understand high dimensional data. Our system includes several novel features in both the stages of data processing and visualization, intended to overcome the classical problems inherent in effectively understanding large, multivariate datasets. In section III, we begin by using a statistical method to analyze the data and extract important information and dimensions in a data processing unit. Then, in the data visualization unit, we explore the use of intuitive methods to properly present the most important high-dimensional features in a two-dimensional form on the screen. In particular, we develop innovative level of detail methods capable of supporting very high dimensional data. We handle the problem of occluding lines through the use of traditional graphics techniques, combined with an analysis of the correlation between neighboring variables, and we enable important patterns to be highlighted using generalized brush tools. Although many different tools with different capabilities have been previously proposed for the visualization of high dimensional data, we believe that our methods provide a novel tradeoff solution that is capable of representing the most important information in many different aspects and different ways.

We apply our tool to two different datasets, featuring information related to housing and to automobiles. The Boston housing dataset by Harrison and Rubinfeld [1, 36] was acquired from the Statistics library that is maintained by Carnegie Mellon University. This real world data concerns housing values in the suburbs of Boston. It includes 506 records, each with 16 attributes, one binary-valued and fifteen continuous, including latitude and longitude. The automobile dataset was obtained from the Machine Learning Repository [1]. It includes 205 instances with 26 attributes each, including three types of entities: (a) the risk factor symbol assigned as an insurance risk rating; (b) normalized losses among different cars; and (c) the specifications of the automobile. Through these applications we attempt to demonstrate that our tool are

capable of assisting users to highlight significant insights and, thus, gain deeper understanding of the data.

II. RELATED WORK

Information in the real world has been growing at an exponential rate, and this explosion of data inspires an increasing need for ways to understand these data. Much attention in the research of multi-variate visualization has been focused on the ongoing effort to develop a variety of tools for visualizing high-dimensional data and solving real related problems. Methods of investigating these extremely high-dimensional variables and representing them in a better format in a two-dimensional window according to the datasets are thus critical to gain insights into characteristics or patterns of the data.

Different kinds of common approaches to multivariate visualization can be categorized and described in several major techniques: glyphs, hierarchical techniques, scatterplots, parallel coordinates, and dimensional reduction techniques. The techniques in the category of glyphs map data values to various primitives, symbols, or curves by using attributes or function [4]. Urness' work uses not only glyphs but also textures in 2D flow fields [6, 7, 8]. Glyphs have also been shown effective in 3D flow fields [5]. This method describes the characteristics of data directly and intuitively using graphical primitives. However, the number of dimensions that can be conveyed effectively is still limited although multiple attributes can be encoded in the glyphs to maximize the number of displayed variables. Hierarchical techniques are developed in order to represent characteristics of variables at different levels in a hierarchy. Examples include the methods of embedded dimensions [12], Dimensional Stacking [10], and Worlds within Worlds [14]. These techniques have advantages in dense data but perform poorly in sparse data. The major disadvantage is that spatial relationships across dimensions may be lost due to the restructuring of the data presentation.

Scatterplots are widely used methods for visualizing multi-dimensional datasets and especially useful for investigating any combination of variables. A variety of variations of scatterplots have been developed to enhance the information representation for correlation of paired variables. These methods are very simple, and widely accessible. A scatterplot matrix can provide large amounts of detailed information about all of the variables in a dataset via a collection of 2D plots. However, this method has to compress all of the details into a tiny figure in order to include all combinations of relationships in different small plots. Thus, the major limitation is that the information in the entire figure would be too much to be effective when there are many variables to be visualized. It is also difficult to find and interpret patterns across multiple variables.

Parallel coordinates is a conventional technique for visualizing high-dimensional data and for representing correlations between neighboring coordinates [3, 16, 25]. It has had various applications in broad areas in its long history. It has received more attention since Inselberg [17] discussed this method again with the property of duality and conversion of high-dimensional data points between Cartesian coordinates and parallel coordinates. These properties can connect the patterns observed across coordinates. Tools developed upon parallel coordinates [19, 26] provide ways to gain insights into characteristics of data. Brushing techniques [11, 15] are very useful to highlight interesting portions under user-specified criteria. The generalized parallel coordinate plot (GPCP) [34] was proposed to plot transformed data based on different interpolations, and various other curves [9, 13] were also developed. However, parallel coordinates suffers from the problem of occlusion among crowded lines, so much effort has focused on solving visual clutter. Tile-based parallel coordinates [31] prevents this problem by allowing users merely to show information in each tile. Moustafa's QGPCP [32] reduces visual cluttering by integrating a frequency model into GPCP. Hidden clusters existing in large data can also be uncovered by parallel coordinates [2, 9, 21, 22]. Hierarchical clustering [27] provides the capability of interactively unveiling the patterns of huge data and producing a display at different levels of detail by the combination of techniques such as hierarchical clustering, dimension zooming, extent scaling, dynamic masking, etc. In addition, axis manipulation [15, 28] has been developed to provide an extension of parallel coordinates by variations in the axes' appearance, ordering, spacing, etc. for improved representation of data and the reduction of clutter. Muntzner et al. [29] proposed automatically ranking and selecting axes based on the importance of paired relationships among variables. Ordering of data values on the axes [30] has also been used to enhance the data representation.

Dimensional reduction techniques reduce higher dimensions into low dimensions. Typical methods include multidimensional scaling [18, 20], principal component analysis [24], and self-organizing maps [23]. However, these methods may not scale well, and may introduce additional occluded or overlapped elements into the display. Moreover, with the use of these methods, the relationships among the dimensions is lost, so they may not be the best choice for visualizing high-dimensional data.

Regression problems are relating to analyzing a usual type of data set $(x^i, y_i)_{i=1}^n$ where the y_i's are n independent observations of the response Y given its predictor $x^i = (x_{i1}, \ldots, x_{ip})^T$. Based on Generalized Linear Models (GLM), which is a parametric approach to estimating covariate effects [39], Akaike [37, 38] proposed to select a good model that minimizes the Kullback-Leibler (KL) divergence of the fitted model from the true model. This is the well-known AIC approach. Schwartz [40] proposed a similar idea from a Bayesian perspective that led to BIC. The work of AIC and BIC provides a unified approach to model selection: choose a parameter vector θ that maximizes the penalized likelihood

$$\ln(\theta) - \lambda \parallel \theta \parallel 0 \qquad (1)$$

where the L_0-norm of θ counts the number of non-zero components in θ and $\lambda \geq 0$ is a regularization parameter. However, this approach cannot handle high dimensional cases. This leads to a natural generalization of penalized L_0-regression called Penalized L_1-regression i.e. LASSO regression by Tibshirani [41] in the ordinary regression setting.

Later a linear combination of L_1 and L_2 penalties encourages some grouping effects. This is the elastic net proposed by Zou and Hastie [34].

III. OUR APPROACH

Parallel coordinates is widely used for applications of multivariate visualization, and a variety of its variations have been developed. However, little work has been done using regression problems as well to handle and represent high-dimensional data. In this section, we describe our algorithm of visualizing regression data in terms of two components, data processing and data visualization units, based on LASSO and advanced parallel coordinates.

A. Data Processing Unit

In many high dimensional problems we often want to find a smaller subset of input covariates $X_1, \ldots, X_p$ that contribute most to a specified output Y. Consider the high dimensional regression model

$$Y = \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon = x^T \beta + \epsilon \qquad (2)$$

where ϵ is a random noise and $\beta = (\beta_1, \ldots, \beta_p)^T$ is a p-dimensional vector. The linear relation between Y and $X_1, \ldots, X_p$ captured in β is estimated using n pairs of training data $\{x_i, y_i\}$ for $i = 1, \ldots, N$ where $x_i \in \mathbb{R}^p$. Some traditional regularization methods such as ridge regression [35, 36] are adopted for solving the problem when the number of explanatory variables is much larger than the number of observations, i.e. $p \gg N$. Ridge regression is defined for minimizing the residual sum of squares, subject to a constraint on the L_2-norm of the regression coefficients, $\|\beta\|_2^2 \leq t$, which is equivalent to an optimization problem with an L_2 penalty on the regression coefficients:

$$\hat{\beta}^{Ridge} = \min_\beta \sum_{i=1}^N (y_i - x_i^T \beta)^2 + \lambda \|\beta\|_2^2 \qquad (3)$$

Ridge regression can nicely handle correlated predictors. If two predictors are highly correlated, ridge regression will equally scale each predictor. However, ridge regression cannot do variable selection, for it only proportionally shrinks the coefficients, but does not set any to exact zero. LASSO proposed by Tibshirani [41] successfully combines the shrinkage property and subset selection; LASSO is defined for minimizing the residual sum of squares, subject to a constraint on the L_1-norm of the regression coefficients, equivalently:

$$\hat{\beta}^{LASSO} = \min_\beta \sum_{i=1}^N (y_i - x_i^T \beta)^2 + \lambda \|\beta\|_1 \qquad (4)$$

The L_1 penalty in LASSO results in a variable selection property due to the singularity of L_1 function at zero. LASSO can exclude unimportant variables from the model by shrink their coefficients to be exact zero. However if two predictors are highly correlated, LASSO will select only one and completely dropped the other.

More recently Zou and Hastie [34] proposed the elastic net which combines the strengths of the previous two approaches, it is a mixture of the L_1 (lasso) and L_2 (ridge regression): to maximize the likelihood subject to the constraint $x_i = (1 - \alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1 \leq t$, equivalently,

$$\hat{\beta}^{Net} = \min_\beta \sum_{i=1}^N (y_i - x_i^T \beta)^2 + \lambda(1 - \alpha)\|\beta\|_2^2 + \lambda\alpha\|\beta\|_1 \quad (5)$$

when $\alpha = 1$ the problem reduces to the LASSO penalty. The advantages of elastic net are threefold. First of all, by the property of its constraint on parameter α, making λ enough large will set some of the coefficients to be exactly zero, hence left us a smaller subset of input variables with nonzero associated estimates. Secondly, this model selection process is continuous hence very stable: when λ is sufficiently large all estimate are zero. While gradually decreasing λ will make variables that minimize the penalized residual sum of squares become nonzero. Thus predictors with strongest effect will enter the model first. Thirdly, the elastic net has a nice grouping effect, it can select groups of correlated variables, i.e. strongly correlated predictors tend to be in or out of the model simultaneously.

B. Data Visualization Unit

Parallel coordinates is the main visualization unit of our tool used for visualizing high dimensional data in this paper. We augment the traditional approach with a combination of techniques including axis manipulation, axis grouping, line and axis coloring, and region highlighting for enhancing multivariate visualization and investigating patterns in the data.

1) Axis Manipulation: Using regression analysis with high dimensional data for model selection could provide important information about the significance of variables, and it is very handy to develop techniques by using axis manipulation to visualize the information effectively. The techniques we use include axis reordering, axis navigation, and axis flip.

Different methods [15, 28] have been proposed for ordering axes; however, the order we adopt from the model selection should be the most appropriate for regression problems. We apply elastic net regression to the high dimensional data to obtain a sequence of ordered variables and render the Y variable followed by a subset of subsequent ordered variables on a 2D screen. The number of variables that are displayed in a screen can be dynamically designated by the user according to the screen resolution, the dataset characteristics, and the user's preference. Since the number of variables that can be displayed at one time in one screen is limited, our tool addresses a way for observing large numbers of different variables through the capability of navigating subsets of variables in distinct screens. These variables still appear in an ordered sequence but on different pages, in which the first coordinate remains the Y axis. We can not only focus on the most important data and their relationships in one screen by automatically ordering significant variables but we can also investigate all other variables by navigating among them. In addition, we enable the feature of axis flip, which provides the assist of better representation of line connection between neighboring coordinates. Parallel coordinates features a strong visualization of adjacent coordinates, whose correlation ranges from negative one to positive one. When the correlation between two variables is near positive one, lines connecting two variables of each record tend to be parallel. Otherwise,

complex crossing of lines would appear. Flipping the axes properly according to the correlation of each pair of adjacent coordinates ensures the positive correlation between neighboring coordinates. The ability of switching between flip and non-flip statuses improves the observation of characteristics between two variables since human perception distinguishes the degree of parallel lines well. We design the axes to be of a certain width with a solid green triangle on the bottom edge or a solid red triangle on the top edge. The bottom and top of all of the axes are originally minimum and maximum values of corresponding variables respectively. Theses axes have solid green triangles on the bottom edges, representing their non-flipped state featuring increasing values from the bottom. Once an axis is flipped, its triangle is moved to the top and given the red color. All flipping can be performed automatically, and the icons provide effective hints for the status of the axes. Users can also decide to switch between statuses on their own.

2) Axis Grouping: Our tool not only arranges axes in an ordered sequence with respect to variables' significance but also provides information of significance in different levels. A series of grouped highly correlated variables are produced in data processing stages and converted into advanced parallel coordinates in a two-dimensional figure. Users can view the figure embedded with an implicit hierarchical structure of grouped variables and further observe relationships across multiple variables with strong correlations.

We apply techniques of different spacing, curves, and axes to group of variables. Intuitively, variables in the same group have similar properties so they tend to be placed more closely. Thus, spacing of axes in a same group is given narrower. We further make the group tight in human perception by replacing axes from rectangle boxes to thick lines so that the groups can be perceived directly. An issue is raised while this design is developed. When we compress more information into groups with less spacing, lines between axes tend to become more cluttered and unclear. Flipping axes may alleviate this problem but it is not sufficiently effective in all situations. Therefore, we use curves to reduce the distance of line deviances in the group. Many possible splines could be used; we found that B-spline curves met our needs well. The curves are located within the convex hall of the control points, which are the values of variables in the record, and they show the trend of the lines. Degrading dramatically crossing lines into smooth curves alleviates unwanted crowding effects in narrower regions.

3) Line and Axis Color: Colors can be very effective for conveying information. We can encode a variety of information in the appearance of lines or axes by using different coloring schemes. Originally, we had only mono-colored lines in between coordinates; however, various saturated colors can encode the degree of correlation of two neighboring variables. We use the most saturated blue and red colors to represent correlations of positive and negative one respectively. The saturation of the color decreases as the absolute correlation decreases. Dark or black colors imply that the two adjacent variables are not correlated. When we

apply the axis flip, all correlations of two adjacent variables will be positive and only blue and black colors of lines will be seen. We can further encode values of Y into the colors of the lines by using a rainbow color scale. The rainbow color scale includes multiple hues, and we map these colors (excluding the white portion in the scale) to the Y axis bar, which is the first axis in each page. Each record is given the color according to the color of the Y axis where it starts. The color for each record is the same throughout all line segments connecting variables' values. In this way, we can easily observe the corresponding Y value based on the line color of each record.

Our thick axis bars are capable of encoding line frequency, deviances, or other values in patterns and colors on the bars. There are many records connecting two neighboring coordinates, many of which may coincide, leading to the loss of information about some records in the display. To avoid this, we develop frequency indicators by dividing our bar into twenty segments and filling in a color for each segment based on the line frequency ending at that segment. The more lines a segment ends with, the darker the color of that segment will be. If the color of a segment is white, then there are no lines connected to that segment. Our axis coloration scheme can also encode values like deviances by drawing a red color in a portion of a bar. If we consider a full red axis bar to be 100 percent and an empty bar to be 0 percent, all other deviances can be represented with a red partial axis bar according to its values. This bar can certainly encode different values, depending on which values users want to visualize.

Visual clutter is an intrinsic problem in parallel coordinates visualization when the number of records is high. By using the transparency capabilities of OpenGL on top of our line coloring methods, we can provide an adequate solution and appealing results. Where multiple lines occlude or coincide, the colors of the lines will tend to appear more saturated. Otherwise, they are less salient. Thus this method emphasizes the lines that appear with higher frequency, and those saturated lines tend to receive more attention automatically.

4) Region Highlight: Highlighting regions containing groups of lines focuses attention on those lines and allows the unveiling of insights into data in those regions. Showing the relationships of variables among groups is important. We paint the background with a light blue color within each region of grouped coordinates. Coloring the regions of each group brings users' attention to these areas and emphasizes the relations of the line connections.

Highlighting lines or curves within some specified regions enhances the capability of our tool for visualizing interesting data. Suppose that S is a set of n-dimensional data defined as $S = \{x_k | x_k \text{ is a record in the } n\text{-dimensional data}\}$. x_k is an n-dimensional record defined as $x_k = (x_{k_0}, x_{k_1}, \ldots, x_{k_{n-1}})$. We can define an interval $I_{i_j} = [a_{i_j}, b_{i_j}]$ of the axis i and an axis restriction $A_{i_j} = \{x_k | x_k \in I_{i_j}\}$ based on this interval I_{i_j}. A highlighted region R_i between coordinates i and $i+1$ is thus defined as $R_{i_j} = A_{i_j} \cap A_{(i+1)_j}$. Similarly, we have brush

regions corresponding to Y coordinates and defined as $R_{Y_j} = A_{Y_j} \cap A_{0_j}$, where $A_{Y_j} = \left\{ x_k | f(x_k) \in I_{Y_j} \right\}$, $f(x_k)$ is the response of x_k, and I_{Y_j} is an interval of Y. With these basic blocks, generalized brush regions are given as: $R = \left(\cap_i \left(\cup_j R_{i_j} \right) \right) \cap \left(\cup_j R_{Y_j} \right)$. Any complex brush regions consist of the combination of these basic restrictions and can be supported by our models. Only records belonging to these brush regions will be rendered with original colors, and all others will be rendered with less salient colors like gray.
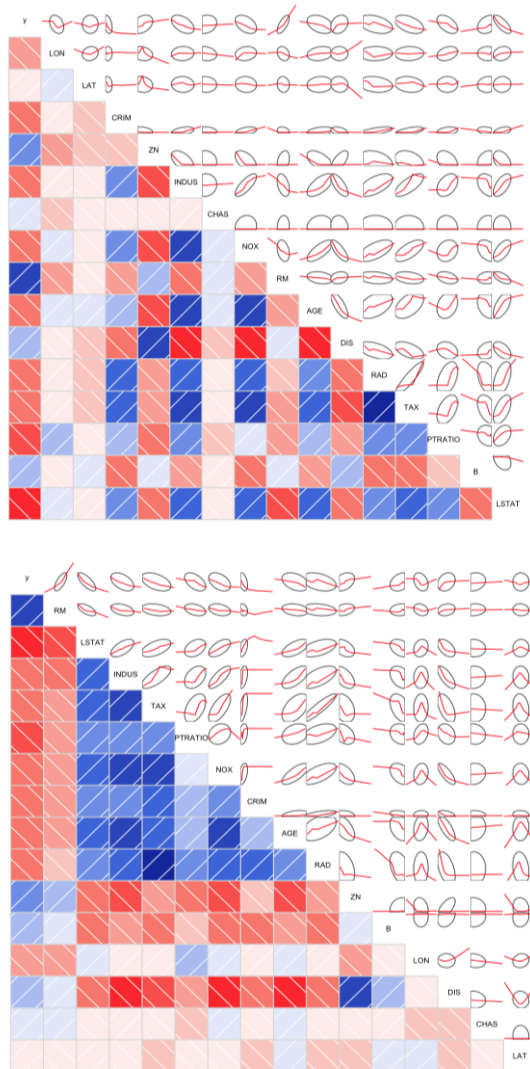
Fig. 1. (Top) The regular scatterplot is drawn for correlations of response and predictors. (Bottom) The scatterplot is drawn with ranking of predictors.

IV. CASE STUDY APPLICATIONS

By applying our tool to two datasets, a housing dataset and an automobile dataset, we demonstrate the features of our method and attempt to show how our approach can assist the effective investigation of data.

These case study results are presented in two sections below.

A. Housing Data

We visualize the Boston Housing Dataset from 1978 [1, 36] with our method. The data frame has 506 instances and 16 predictors concerning housing price. After ranking and grouping the variables, the most significant variables along with house price MEDV are displayed in order in an advanced parallel coordinates plot. As expected, we find strong patterns within the groups and highlight them in our plots.
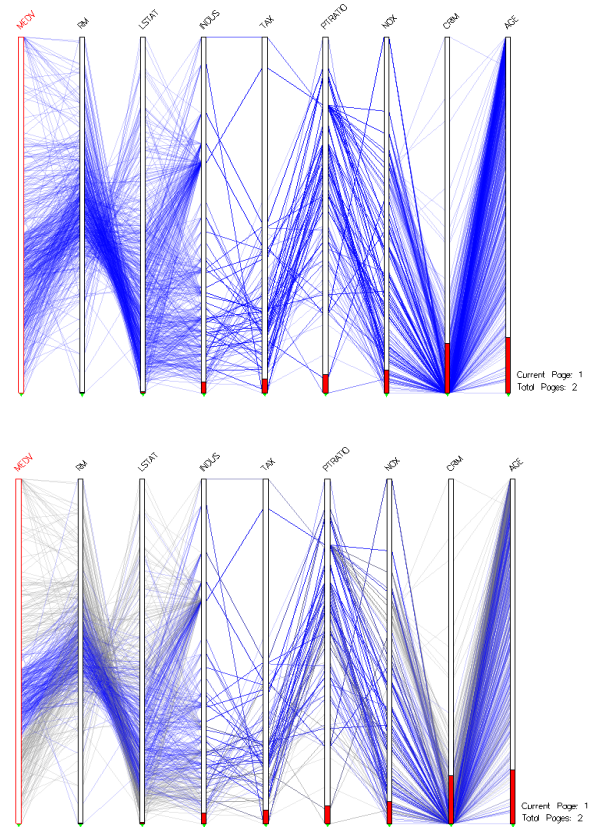

Fig. 2. (Top) The most important eight variables plus the response variable are shown in this first page. Lines are drawn with monotonic colors and emphasis of higher frequency. Axes are coded with deviances. (Bottom) A brushed region is applied to the second quarter of house prices. Lines in the brushed region remain blue, and other lines are given a gray color.

The scatterplot on the top in figure 1 represents the correlation among response and predictors. Saturated blue and red show strong positive and negative correlations respectively. After ranking and grouping by elastic net, we find the squares with dark shades mainly distributed in the upper left corner of the bottom of the plot. This shows that variables with strong correlation have been placed and grouped together.

Figure 2 shows the first page of a regression analysis of the Boston housing data, which includes the response (the coordinate of the red boundary), and the most significant eight predictors (the coordinates of the black boundaries). Axes are coded with deviances by red bars and lines are coded with emphasized blue colors. The image in the bottom of figure 2 shows the application of a brushed region to the second quarter of housing prices. All lines in this region remain in their original colors while the rest are shown in gray.
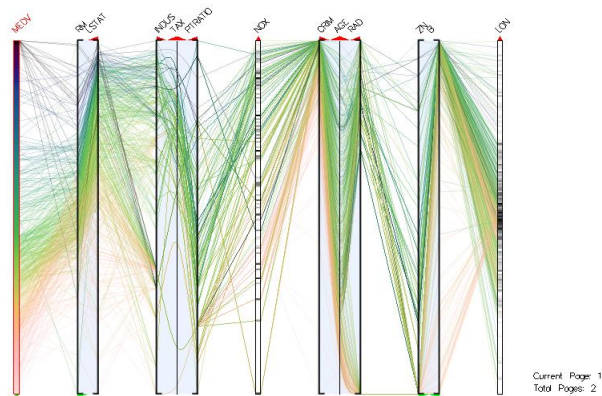
Fig. 3. A Rainbow color scale is applied so that line colors are drawn with respect to the house price. Information of axis groups with flipping is included in this figure and single axes are drawn with line frequency.
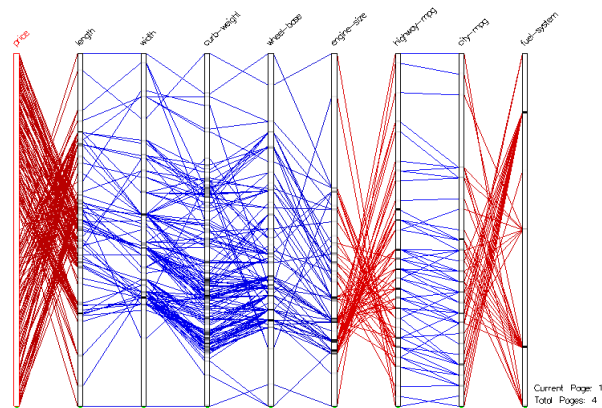

Fig. 4. Positive and negative correlations between neighboring coordinates are displayed with blue and red colors respectively.

Figure 3 demonstrates the use of advanced parallel coordinates with grouped coordinates. A rainbow color scale is applied to the lines and curves with the first axis displaying the color variation corresponding to the response variable. Line frequency is encoded in the single axis and our curves within groups depict the trend of in-between variables. The axis-flip technique is used to enhance the perception of relationships and reduce clutter. After flipping the negative correlations into positive, the strong with-group effect is even more obvious. After price we have three important groups, and within each group there are three members. Positive correlation shown in form of parallel patterns can be easily found within three groups especially in groups one and two.

B. Automobile Data

Figure 4 shows an application of our parallel coordinates to the automobile data, mainly showing the relationships and correlations between coordinates. Correlations are depicted by the saturation of blue and red colors. The fully saturated blue and red colors represent correlations of positive one and negative one respectively. All of the axes are not flipped, so the values increase from minimum to maximum from the bottom of the axis. Figure 5 shows the strong relationships in the predictors, which are depicted in three large groups. Each

group has three variables and demonstrates the tendency with similar trends in the group. Rainbow colors help to provide information about how the variables' values correspond to the price of automobiles.

V. CONCLUSION

With the intrinsic properties of elastic net penalized regression, the combination of advanced parallel coordinates and elastic net penalized regression enables these superior advantages:

- Variables are ranked by significance, and so are the corresponding coordinates. Important variables that enter the model earlier come first in the parallel coordinates plot. Our tool supports an ordering of axes in which the more significant variables are visualized earlier than the rest.

- An additional group selection capability is provided. Variables with high correlation are included into the model together once one variable among them is selected. This maintains a high within-group correlation and low between-group correlations. Our parallel coordinates plot identifies and emphasizes the within-group relationships effectively.

- The multiple capabilities provided by our tool can significantly enhance parallel coordinates plots. We have integrated additional techniques such as axis navigation, axis flip, axis spacing, curves, line and axis colors, frequency indicators, and brush regions to assist users in gaining insights into features of high-dimensional data.

We develop a new tool using advanced parallel coordinates that incorporates regression analysis. Our work demonstrates the potential to unveil insights into high-dimensional data and to achieve effective multivariate visualization.
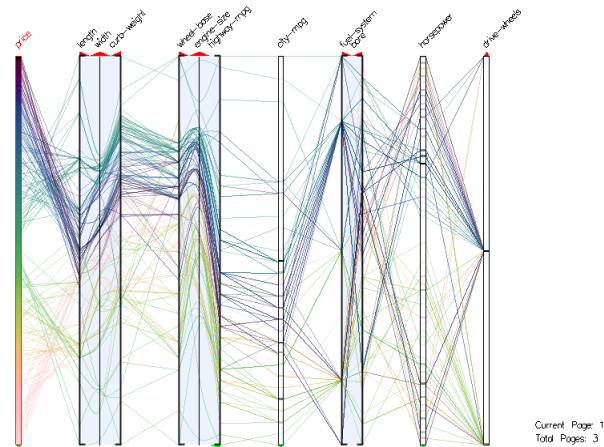
Fig. 5. Grouped coordinates with rainbow color-coded lines and frequency-coded axes is shown in this figure. Axes are flipped according to the correlations with adjacent axes.

REFERENCES

[1] C. L. Blake and C. J. Merz, UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California, Irvine, 1998.

[2] J. Zhou, S. Konecni, K. Marx, and G. Grinstein, "A Visual Approach to Improve Clustering Based on Cluster Ensembles," Proceedings of SPIE 7530, 2010.

[3] R. E. Moustafa and E. J. Wegman, "On Some Generalization to Parallel Coordinate Plots," Seeing a Million, A Data Visualization Workshop, 2002.

[4] R. J. Littlefield, "Using the GLYPH concept to Create User-Definable Display Formats," Proceedings NCGA '83., pp. 697-706, June 1983.

[5] S. Wang, V. Interrante, and E. Longmire, "Multi-Variate Visualization of 3D Turbulent Flow Data," Proceedings of SPIE The International Society for Optical Engineering, 7530, Jan. 2010.

[6] T. Urness and V. Interrante, "Streamline Visualization of Multiple 2D Vector Fields," Proceedings of SPIE The International Society for Optical Engineering, 6809, Jan. 2008.

[7] T. Urness, V. Interrante, I. Marusic, E. Longmire, and B. Ganapathisubramani, "Effectively Visualizing Multi-Valued Flow Data Using Color and Texture," Proceedings of IEEE Visualization 2003, pp. 16, 2003.

[8] T. Urness, V. Interrante, E. Longmire, I. Marusic, S. O'Neill, and T. W. Jones, "Strategies for the Visualization of Multiple 2D Vector Fields," IEEE Computer Graphics and Applications, vol. 26, no. 4, pp. 74-82, 2006.

[9] K. T. McDonnell and K. Mueller, "Illustrative Parallel Coordinates," Computer Graphics Forum, vol. 27, no. 3, pp. 1031-1038, 2008.

[10] J. LeBlanc, M. O. Ward, and N. Wittels, "Exploring N-dimensional Databases," Proceedings of Visualization '90, pp. 230-237, 1990.

[11] A. R. Martin and M. O. Ward, "High Dimensional Brushing for Interactive Exploration of Multivariate Data," Proceedings of IEEE Visualization '95, pp. 126, 1995.

[12] T. Mihalisin, E. Gawlinski, J. Timlin, and J. Schwegler, "Visualizing Multivariate Functions, Data, and Distributions," IEEE Computer Graphics and Applications, vol. 11, pp. 28-37, 1991.

[13] M. Graham and J. Kennedy, "Using Curves to Enhance Parallel Coordinate Visualizations," Proceedings Of Intl. Conf. on Information Visualization, pp. 10-16, 2006.

[14] C. Beshers and S. Feiner, "AutoVisual: Rule-Based Design of Interactive Multivariate Visualizations," IEEE Computer Graphics and Applications, vol. 13, no. 4, pp. 41-49, 1993.

[15] H. Hauser, F. Ledermann, and H. Doleisch, "Angular Brushing of Extended Parallel Coordinates," Proceedings of the IEEE Symposium on Information Visualization (InfoViz'02), pp. 127, 2002.

[16] E. J. Wegman, "Hyperdimensional Data Analysis Using Parallel Coordinates," J. Am. Stat. Assn., vol. 85, no. 411, pp. 664-675, 1990.

[17] A. Inselberg, "The Plane with Parallel Coordinates," Special Issue on Computational Geometry of the Visual Computer, vol. 1, no. 2, pp. 69-91, 1985.

[18] A. Mead, "Review of the Development of Multidimensional Scaling Methods," The Statistician, vol. 33, pp. 27-35, 1992.

[19] D. Yang, E. A. Rundensteiner, and M. O. Ward, "Analysis Guided Visual Exploration of Multivariate Data" Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology, pp. 38-90, 2007.

[20] S. Weinberg, "An Introduction to Multidimensional Scaling," Measurement and Evaluation in Counseling and Development, vol. 24 pp. 12-36, 1991.

[21] A. O. Artero, M. C. F. de Oliveira, and H. Levkowitz, "Uncovering Clusters in Crowded Parallel Coordinates Visualizations," Proceedings of the IEEE Symposium on Information Visualization (InfoVis'04), pp. 81-88, 2004.

[22] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen, "Visual Clustering in Parallel Coordinates," Computer Graphics Forum, vol. 27, no. 3, pp. 1047-1054, 2008.

[23] T. Kohonen, "The Self-Organizing Map," Proceedings of IEEE, pp. 1464-80, 1978.

[24] J. Jolliffe, "Principal of Component Analysis," Springer Verlag, 1986.

[25] A. Inselberg and B. Dimsdale, "Parallel Coordinates: A Tool for Visualizing Multidimensional Geometry," Proceedings of IEEE Visualization '90, pp. 361-378, 1990.

[26] M. O. Ward, "XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data," Proceedings of IEEE Visualizatin '94, pp. 326-333, 1994.

[27] Y.-H. Fua, M. O. Ward, and A. E. Rundensteiner, "Hierarchical Parallel Coordinates for Exploration of Large Datasets," Proceedings of IEEE Visualizatin '99, pp. 43-50, 1999.

[28] J. Yang, W. Peng, M. O. Ward, and E. A. Rundensteiner, "Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration of High Dimensional Datasets," Proceedings of the IEEE Symposium on Information Visualization, pp. 105-112, 2003.

[29] H.-C. Hege, I. Hotz, and T. Muntzner, "Pairwise Axis Ranking for Parallel Coordinates of Large Multivariate Data," Eurographics-IEEE TCVG Symposium on Visualization, vol. 28, no. 3, 2009.

[30] G. E. Rosario, E. A. Rundensteiner, D. C. Brown, M. O. Ward, and S. Huang, "Mapping Nominal Values to Numbers for Effective Visualization," Proceedings of the IEEE Symposium on Information Visualization (InfoVis'04), vol. 3, no. 2, pp. 80-95, 2004.

[31] J. Alsakran, Y. Zhao, and X. Zhao, "Tile-Based Parallel Coordinates and its Application in Financial Visualization," Proceedings of SPIE 7530, 2010.

[32] R. E. Moustafa, "QGPCP: Quantized Generalized Parallel Coordinate Plots for Large Multivariate Data Visualization," J. Comput. Graph. Stat., vol. 18, no. 1, pp. 32-51, 2009.

[33] D. Harrison and D. L. Rubinfeld, "Hedonic Prices and the Demand for Clean Air," Journal of Environmental Economics and Management, vol. 5, pp. 81-102, 1978.

[34] H. Zou and T. Hastie, "Regularization and Variable Selection via the Elastic Net," J. Roy. Statist. Soc. Ser. B, vol. 67, pp. 301-320, 2005.

[35] N. R. Draper and H. Smith, Applied Regression Analysis. Wiley Series in Probability and Statistics, 3rd ed., John Wiley & Sons, New York, 1998.

[36] R. R. Hocking, Methods and Applications of Linear Models: Regression and the Analysis of Variance. Wiley Series in Probability and Statistics, John Wiley & Sons, New York, 1996.

[37] H. Akaike, "Information Theory and an Extension of the Maximum likelihood Principle," Second International Symposium on Information Theory (Edited by B.N. Petrov and F. Csaki), Akademiai Kiado, Budapest, pp. 267-281, 1973.

[38] H. Akaike, "A New Look at the Statistical Model Identification," IEEE Trans. Auto. Control, vol. 19, pp. 716-723, 1974.

[39] P. McCullagh and J. A. Nelder, "Generalized Linear Models," Chapman and Hall, London, 1989.

[40] G. Schwartz, "Estimating the Dimension of a Model," Ann. Statist., vol. 6, pp. 461-464, 1978.

[41] R. Tibshirani, "Regression Shrinkage and Selection via the LASSO," J. Roy. Statist. Soc. Ser. B, vol. 58, pp. 267-288, 1996.