Dynamic visualization of high-dimensional functions via low-dimension projections and sectioning across 2D and 3D display devices

A thesis submitted for the degree of

Doctor of Philosophy

by

Nicholas S Spyrison

B.Sc. Statistics, Iowa State University

Department of Econometrics and Business Statistics

Monash University

Australia

December 2018

Contents

Chapter 1

Literature Review

1.1 Touring

1.1.1 Overview

In univariate datasets histograms, or smoothed density curves are employed to visualize data. In bivariate data scatterplots and contour plots (2-d density) can be employed. In three dimensions the two most common techniques are: 2-d scatter plot with the 3rd variable as an aesthetic (such as, color, size, height, *etc*.) or rendering the data in a 3-d volume using some perceptive cues giving information describing the seeming depth of the image [1]. When there are 4 variables: 3 variables as spatial-dimensions and a 4th as aesthetic, or a scatterplot matrix consisting of 4 histograms, and 6 unique combinations of bivariate scatterplots.

Let p be the number of numeric variables; how do we visualize data for even modest values of p (say 6 or 12)? It's far too common that visualizing in data-space is dropped altogether in favor of modeling parameter-space, model-space, or worse, long tables of statistics without visuals(Wickham, Cook, and Hofmann, 2015). Yet, we all know of the risks and possible mis-leadingness of relying too heavily on parameters alone(Anscombe,

[1] Graphs of data depicting 3 dimension are typically printed on paper, or rendered on a 2-d monitor, they are intrinsically 2-d images. They are sometimes referred to as 2.5-d, or more frequently erroneously referred to as 3-d, more on this later.

1973; Matejka and Fitzmaurice, 2017). So why do we get aware from visualizing in data-space? Scalability, in a word, we are not familiar with methods that allow us to concisely depict and digest $p \geq 5$ or so dimensions. This is where touring comes in; using the wide range of touring techniques we are able to preserve the visualization of data-space, and the intrinsic understanding of structure and the data, beyond looking at statistic values alone.

Touring is a linear dimensonality reduction technique that orthagonally projects p-space down to d-space. Many of these projections are interpolated while varying the rotation of p-space and viewed in order to the effect of watching an animation of the lower dimensional embedding changing as p-space is manipulated. Shadow puppets offer a useful analogy to aid in conceptualizing touring. Imagine a fixed light source facing a wall. When a hand or puppet is introduced the 3-dimensional object projects a 2-dimensional shadow onto the wall. This is a physical representation of a simple projection, that from $p = 3$ down to $d = 2$. If the object rotates then the shadow correspondingly changes. Observers watching only the shadow are functionally watching a 2-dimensional tour as the 3-dimensional object is manipulated.

1.1.2 History

In 1974 Friedman and Tukey purposed Projection Pursuit(Friedman and Tukey, 1974) (sometimes referred to as PP) while working at Bell Labs. Projection Pursuit involves identifying "interesting" projection, remove a single component of the data, and then iterate in this newly embedded subspace. Within each subspace the projection seeks for a local extrema via gradient descent (historically referred to as hill climbing algorithms), hence the nomenclature pursuit.

Touring was first introduced by Asimov in 1985 with his purposed Grand Tour(Asimov, 1985) at Stanford University. In which, Asimov suggested three types of Grand Tours: torus, at-random, and random-walk. The specifics of which will be discussed below in the Typology section.

...

below is a list of software implementations in ascending year:

```
software & reference & operating system & note \\
\hline\hline
Tourr & [@wickham_tourr_2011] & Linux, Unix, Windows & na \\
Lisp-Stat & [@tierney_lisp-stat:_2009] & na & na \\
CyrstalVision & [@wegman_visual_2003] & Windows & na \\
GGobi & [@swayne_ggobi:_2003] & Linux and Windows & na \\
XGobi & [@swayne_xgobi:_1991] & Unix, Linux, (emulated from) Windows & na \\
ExploreN & [@carr_explorn:_1996] & SGI Unix & na \\
Lisp-Stat & [@tierney_lisp-stat:_2009] & na & na \\
```

software	reference	operating system	note
Tourr	-	Linux, Unix, Windows	na
Lisp-Stat	-	na	na
CyrstalVision	-	Windows	na
GGobi	-	Linux and Windows	na
XGobi	-	Unix, Linux, (emulated from) Windows	na
ExploreN	-	SGI Unix	na
Lisp-Stat	-	na	na

1.1.3 Typology

Movement

A fundamental aspect of touring is the path of rotation. There are four primary methods of defining such paths(Buja et al., 2005):

- Random choice such as Asimov's grand tour(Asimov, 1985).
- Precomputed choice, *e.g.* the little tour(McDonald, 1982).
- Data driven - a guided tour performing (stochastic) gradient descent on some objective function(Hurley and Buja, 1990).
- Manual control, a constrained rotation on selected manipulation variable and magnitude(Cook and Buja, 1997).

… - torus: where a p-dimensional torus, T^p is created from a Cartesian product of p unit circles with $T^p \in \mathbb{R}^p$. Unfortunately uniformity of the parameters do not correlate to uniform points on the surface of the torus. If step distance between frames is fixed, disproportionate time is spent between subspaces. If step distance is change to account for uniform points on the torus then the continuity of the tour is lost.
- at-random: where each 2-frame is chosen at random without replacement. This affords an assured uniform distribution of subspaces, but is far to discontinuous for observation. It also leaves no parameters to control. - random-walk: combines the continuity of the torus method and the uniformity of the at-random method while leaving room for a control parameter.

Geoms

Scatterplots offer a simple, general case for viewing lower-dimension embeddings of higher-dimensions. Such visualization offer p-dim down to d-dim, typically two in the case of a standard monitor. Yet, no intrinsic value stops touring being used in other graphics or geoms (geometrics). For instance using parallel coordinate plots (PCP)[@…], Andrews plots [@…], Chernoff faces [@…], all offer perfectly valid graphs in p-dimensions. these can also be toured.

This works well when the number of dimensions being toured is small (in the neighborhood of 5-10), yet the number of view, or 2-frames and we can produce from p-space suffers from the so called blessing/curse of dimensionality. In which the plethora of degrees of freedom either offer many (non-unique) solutions to a problem or something that becomes ever increasing unlikely, *ie*. a bivariate square needs .

1.1.4 Linear vs non-linear dimensonality reduction

1.2 Virtual reality

Bibliography

Anscombe, FJ (1973). Graphs in Statistical Analysis. *The American Statistician* **27**(1), 17–21. (Visited on 12/19/2018).

Asimov, D (1985). The grand tour: a tool for viewing multidimensional data. *SIAM journal on scientific and statistical computing* **6**(1), 128–143.

Buja, A, D Cook, D Asimov, and C Hurley (2005). "Computational Methods for High-Dimensional Rotations in Data Visualization". en. In: *Handbook of Statistics*. Vol. 24. Elsevier, pp.391–413. http://linkinghub.elsevier.com/retrieve/pii/S0169716104240147 (visited on 04/15/2018).

Cook, D and A Buja (1997). Manual Controls for High-Dimensional Data Projections. *Journal of Computational and Graphical Statistics* **6**(4), 464–480. (Visited on 04/15/2018).

Friedman, J and J Tukey (1974). A Projection Pursuit Algorithm for Exploratory Data Analysis. en. *IEEE Transactions on Computers* **C-23**(9), 881–890. (Visited on 06/22/2018).

Hurley, C and A Buja (1990). Analyzing High-Dimensional Data with Motion Graphics. *SIAM Journal on Scientific and Statistical Computing* **11**(6), 1193–1211. (Visited on 11/27/2018).

Matejka, J and G Fitzmaurice (2017). Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. en. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. Denver, Colorado, USA: ACM Press, pp.1290–1294. http://dl.acm.org/citation.cfm?doid=3025453.3025912 (visited on 12/19/2018).

McDonald, JA (1982). INTERACTIVE GRAPHICS FOR DATA ANALYSIS.

Wickham, H, D Cook, and H Hofmann (2015). Visualizing statistical models: Removing the blindfold: Visualizing Statistical Models. en. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **8**(4), 203–225. (Visited on 03/16/2018).