Discussion

John F. Elder IV

Visualization is increasingly being recognized as an effective and efficient way not only to communicate patterns in scientific data, but to discover them as well. In the low dimensions of everyday experience, the human ability to find meaningful order in noisy data may never be matched by automatons. So the introduction of a useful visualization procedure, as provided here by Furnas and Buja, is indeed a welcome development. They show that low-dimensional patterns extracted by a combined projection and section operation (a prosection) can imply the existence of similar higher-dimensional structure. In exploratory or inductive data analyses, then, prosections could be used to generate hypotheses about relationships between sampled variables. However, the familiar curse of dimensionality may confine their practical application to point clouds of only moderate dimension.

1. VISUALIZATION ALLOWS MULTIPLE END-POINTS

The great strength of interactive data visualization is that an analyst need not define meaningful structure to be able to recognize it. No great training is required to identify any of a large variety of patterns (i.e., nonrandom distributions) when they are encountered. [In personal communications, John Aitchison (former chair of the University of Virginia Statistics Department) has opined, "If we could see in ten dimensions we wouldn't need statisticians, or mathematicians either." Dave Scott (Rice Statistics) counters that no, there'd be even more statistics to do (though he might concur on that bit about mathematicians!).] In contrast, testing a statistical hypothesis requires specifying goals and assumptions. This step can often be enhanced by visually exploring the data space. [Tweaking those with opposing views in How to Mess Up With Data Analysis, Ned Glick notes that, "Dataholics Anonymous advises that even a peek at data ('just a quick one') or some innocent 'social' discussion of data can lead to heavy thinking," (Glick 1991).]

But straightforward visualization techniques, such as 3-D rotating plots, can reveal only very low-dimensional structure. A fourth potential dimension—movement—is usually harnessed to maintain the illusion of three spatial dimensions on a flat screen, or to transition between views. (If stereo plotting is instead employed with 3-D glasses, movement can be regained, but information from color is lost, though fancy hardware now exists to recover this slight edge.) Further extensions seem only slightly effective. [Color seems best when used to label classes for discrimination tasks. Most changes in

*Research Scientist, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005
Journal of Computational and Graphical Statistics, Volume 3, Number 4, Pages 355–362

355
the properties of the points (e.g., symbol type, point size, or angle of a small ray) can clutter the screen unless the structure is very simple.

Because manual/assisted viewing is limited, peeking into higher dimensions requires that the computer be able to judge views. For example, in "A Projection Pursuit Algorithm for Exploratory Data Analysis" (Friedman and Tukey 1974), one must quantify a projection measure of "interestingness" to be maximized in a search. Even if it were possible to capture a specific version of such a property in an equation, views corresponding to other definitions of interesting would be passed over. Worse, this formulaic reduction is quite imperfect, as a simple example can illustrate.

Many indexes score a projection according to its distance from Normality; which is sensible, as random univariate projections of high-dimensional data appear Normal, almost regardless of the true number of modes (Diaconis and Freedman 1984). The 2-D projection of Figure 1 is the maximum of one such index for 3-D unit sphere data taken from the infamous RANDU pseudo-random number generator. [Note: The projection pursuit capability of XGobi (Swayne et al. 1990, 91, 92) was employed to repeatedly locally optimize the "Natural Hermite" index; that is, the $L_2$ distance between three-term Hermite polynomial approximations of the sampled and Normal distributions.] However, the weakness of that sequential congruential generator is better revealed by the unexpected structure of Figure 2, which shows the points aligned on planes. Yet this view is judged
by the projection index as 97% less interesting! What the index seems to be focusing on in Figure 1, the lack of central mass, is not the salient feature.

[An index in XGobi designed to discover holes in distributions scores the view only slightly better—83% below the maximum. Also, if data inside the unit cube are instead considered, projections that zero out a variable are the local maxima, because the square shape is so un-Normal. In that case, the index focuses on an artifact of the boundary and again ignores the more meaningful striped structure, which reveals an unintended constraint in the generator.]

Though one can design an index that would be maximized by a projection like that of Figure 2, clearly many other forms of structure could be missed that would be obvious to an observer. In some cases iteratively removing structure can help, as is done (for fitted, not perceived structure) in inductive algorithms such as projection pursuit regression (Friedman and Stuetzle 1981), decision trees (Breiman, Friedman, Olshen, and Stone 1984), and polynomial networks (Elder and Brown in press). It has been proposed that combining this with interactive graphics and with automated structure search using a collection of different indexes might take advantage of the complementary properties of the methods (Elder 1992). But inductive modeling methods, which can address high-dimensional relationships among variables, are often outperformed in low dimensions by straightforward visualization techniques (Elder 1993). The close involvement of an
analyst as pattern interpreter appears to be critical to the process of structure discovery. Thus the introduction of a method for extending the reach of the analyst deserves close attention.

2. PROSECUTIONS CAN HELP

As Furnas and Buja illustrate with an empty sphere example, projections alone are insufficient to reveal many types of structure. Other useful methods for reducing dimensionality include summarizing data subspaces with density estimates, and conditional subsampling, or sectioning (e.g., Scott 1992, where a 5-D empty sphere is addressed). The authors note that projections preserve the dimension of objects that can fit in (have lower dimensionality than) the projected space, and that sections preserve the codimension (number of constraints) of objects when these constraints are fewer than the dimensionality of the remaining space. They also point out that any combination of the two operations can be reduced to one of each, in either order, and call this a prospection. In combination, these properties allow one to infer the dimension of an object when its "prospected" dimensionality is observed, and a simple algorithm for tracking the bookkeeping is provided (Subsection 6.1).

Objects with smooth global structure, such as the m-flats in the illustrating game (Section 4), could probably be discovered more easily another way; for example, by inductive modeling. For example, nearly all 2-D prospections of a 5-D investment data set I examined appeared asymmetrically quadratic, suggesting that the utility of combining certain leading indicators wanes if they are over- as well as under-weighted. This basic structural observation was more rapidly provided by fitting a model with fully significant (and noncollinear) quadratic terms.

But the 6-D locus example of Section 2 shows how prospections can reveal local structure that other algorithms would find difficult to capture. It is difficult to formulate the locus problem in a way that a general-purpose estimation or classification algorithm could successfully address. Yet, once the piecewise planar form of the object is revealed by prospections, a specialized computer search could locate joints that summarize the structure better than the sample data (though not as well as the generating equations!). However, the example also hints at the method’s vulnerability of relying critically on sections, which tend to require vast amounts of data.

3. CURSE OF DIMENSIONALITY

Things are so different in high dimensions that intuition gained from life in low-D seems not only useless, but misleading. The basic problem is that, in high-D space, "if neighborhoods are 'local' then they are almost surely 'empty,' whereas if a neighborhood is not 'empty,' then it is not 'local.'" (Scott 1992). A couple of observations suffice. In high dimensions:

1. Nearly every sample views itself as an outlier with respect to the rest of the data.
   For instance: project a spherical Normal distribution onto a line defined by one of its samples and the center of the distribution. The expected separation of the
two points is 3.1 standard deviations in 10 dimensions, and 4.4 in 20 (Friedman 1994).

2. All sample points are close to an edge of the sample space. For example, in the 10-D (uniformly distributed) unit cube, $U^{10}[0, 1]$, the maximum distance to any edge is .5. But, even when using the liberal $L_\infty$ distance metric, the expected distance between a sample and its nearest neighbor approaches .5 when there are 1,000 samples and reduces only slightly to .4 when there are 10,000 (Friedman 1994). In other words, nearest points are, on average, barely closer than the center of the cube is to any edge.

It appears that even moderately dimensioned space is essentially very sparsely populated, and the sections required to contain a significant number of sample points may need to be quite a bit thicker than an idealized plane. For example, in $U^{10}[0, 1]$, to hold 1/10% of a sample would require a cube with edge length just over .5, which is not a very thin slice! As the authors remark (sec. 8.1), “Only with a sufficiently high data density is it possible to find a balance where thick sections capture enough data points to see structure and yet are thin enough to avoid blurring of the global structure as seen through the sections.” What is surprising perhaps, is how high the density has to be as dimension increases.

[Allowing thick sections, one should be able to, for instance, represent a decision tree having $K$ leaves and using multilinear splits as a parallel set of $K$ projections. Probably though, sections with significant thickness hurt the theoretical ability to map a sequence of projection and section operations to a simpler pair.]

Furnas and Buja warn about three other deviations from the ideal often encountered in practice: the presence of noise, the finite extent of point clouds, and nonlinear local patterns. Handling noise seems by far the greatest difficulty for projections; the latter two issues are not as serious when an analyst, rather than an equation, is interpreting the quality of the view.

[Recall how, in the cubic RANDU example (see note, p. 356), the projection pursuit index was fooled by finite extent, though an analyst would not be. Also, though Furnas and Buja centered their analysis on the simple problem of finding global flats, other examples discussed here show projections finding some local or nonlinear structure.]

4. THE LOCUS EXAMPLE WITH NOISE

The 6-D ultrametric locus example of Section 2, which is said to have motivated the development of the method, is revealed by projections as consisting of five jointed planes (Furnas and Buja, fig. 4, sec. 7.1). Actually, Furnas (1988) mentioned that the full structure is fifteen such planes; so, the example serves also as a reminder of how sections can “throw out (some of) the baby with the bath water.” Note that one’s ability to recognize structure in this problem is enhanced by its regular grid of samples. This imposes an order within which small deviations reveal information, such as depth and skewness. (It is likely, for instance, that if the hollow sphere example were similarly grid-sampled, it would have been easier to recognize without sections.) We now briefly
employ samples that are instead irregular, noisy, and/or sparse, to find the boundaries beyond which the prosection procedure might break down on this example.

On a 9-value 6-D regular grid, a couple of thousand grid points (approximately one in 3,000) satisfy the ultrametric constraints of Section 2. I obtained 1,989 locus points using a step size of .1 in the range [.1,.9], whereas the authors report 23% fewer points on what may be a larger range (see their note, Sec. 2). Despite this discrepancy, the plots are very similar.] Experimentation showed that up to about 60% of those data points could be randomly removed and a diligent search over sections (with 3-D projections) could still reveal (the now somewhat singular) examples of the five-leg structure. [The result is necessarily qualitative, as the procedure is, by definition, not automated, and it is hard to correct for the fact that the analyst knows what is being sought! Note that the new "section search" capability in XGobi performs a grand tour (e.g., Asimov 1985; Buja and Asimov 1986) or local optimization (projection pursuit) over the possible projections, highlighting a section of the data (of variable width) centered at the mid-depth of the viewing plane. That simple constraint though made the new tool not useful for this problem, as the center of the locus is very busy and its neighborhood contains somewhat uninformative views.]

Compensating for the loss of data by thickening the sections tended to smooth out (i.e., hide) the structure.
To study the effect of weakening the regular grid, 2,000 points from a slightly finer grid (one with a step size of .03 rather than .10) were sampled. (Grids are natural for this problem, as the locus has zero content, and if distances were not quantized, no experiments would satisfy the ultrametric.) Figure 3 shows a 2-projection of two 4-sections (i.e., sections with two variables fixed within a small range) represented by different symbols. Even though all sample points are on the ultrametric locus, comparison with Figure 4 (p. 336) in Furnas and Buja shows the reduced regularity of the sampling grid has reduced the visibility of the structure (though less on the circles projection, for which this view was tailored, than the other, for which no good view was found).

Returning to the original sample grid, Figure 4 reveals the effect of adding a small amount of Normal noise to the locus; that is, of accepting points that do not quite satisfy the ultrametric condition. (The standard deviation along each axis, $\sigma_k$, equals .025; i.e., 2.5% of the axes range or 25% of the intergridpoint distance.) With noise, details in the patterns dissipate quickly, and little beyond the main axis of variation remains, suggesting a basically linear or quadratic fit.

5. CONCLUSIONS

Data visualization is a powerful analytic method. Though its practitioners occasionally suffer the raptures of the deep and see nonexistent structure in a data cloud, model
overfit is probably less likely in visual than mechanical data analysis. Furnas and Buja are to be commended for introducing a method in projections that can extend a bit one’s ability to discern structure. Its chief vulnerability stems from critically employing sections that are particularly susceptible to the exponential consequences of increased dimensionality; especially, the outsized effects of noise. Usually though, known characteristics of a problem, external to its data, can be used to summarize or smooth away some distractions. It is likely then that projections, in tandem with more traditional statistical tools and with algorithmic searches for data structure, will contribute to the art of separating wheat from chaff in heaps of data.

ACKNOWLEDGMENTS

Thanks to Dave Scott for useful discussions regarding this material, Andreas Buja and Di Cook for help with XGobi, and the Rice Department of Computational and Applied Mathematics, the Center for Research on Parallel Computing, and the NSF for support.

REFERENCES


