REFERENCES

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., and STONE, C. J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, CA.

DIACONIS, P. (1985). Theories of data anlysis from magical thinking through mathematical statistics. In *Exploring Tables, Trends and Shapes* (D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds.). Wiley, New York.

DIACONIS, P. and FRIEDMAN, J. H. (1983). *M* and *N* plots. *Recent Advances in Statistics* (M. H. Rizvi, J. Rustagi, and D. Siegmund, eds.). 425–447, Academic, New York.

FRIEDMAN, J. H. (1977). A variable metric decision rule for nonparametric classification. *IEEE Trans. Comput.* **C-26** 404–408.

FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717.

FRIEDMAN, J. H. and RAFSKY, L. C. (1981). Graphics for the multivariate two-sample problem. *J. Amer. Statist. Assoc.* **76** 277–287.

GORDON, L. and OLSHEN, R. A. (1978). Asymptotically efficient solutions to the classification problem. *Ann. Statist.* **6** 515–533.

KOLATA, G. (1982). Computer graphics comes to statistics. *Science* **217** 919–920.

REAVEN, G. M. and MILLER, R. G. (1979). An attempt to define the nature of chemical diabetes using a multidimensional analysis. *Diabetologia* **16** 17–24.

SYMONS, M. J. (1981). Clustering criteria and multivariate normal mixtures. *Biometrics* **37** 35–43.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

L. A. SHEPP

AT&T Bell Laboratories

I would like to discuss the many interesting similarities and differences between projection pursuit (PP) and computed tomography (CT) as emphasized by Professor Huber.

First, just in case PP becomes a successful tool in exploratory data analysis (EDA), I hasten to point out that CT has indeed had some influence on PP. Indeed, the basic (to PP) concept of ridge function is an idea which was introduced to CT already in Logan's paper in the *Duke Math. J.*, 1975. The idea of superposing filtered projections, which is the basis of all algorithms used in commercial CT, is analogous to superposing filtered projections in PP even though "filtered" in PP is used in the sense that all but the "interesting few of the projections" are discarded while in CT "filtered" means convolution of each projection with a "filter function" (this is also due to Logan). Actually PP seems even closer to emission CT than to transmission CT—see Vardi, Shepp and Kaufman (1985), *J. Amer. Statist. Assoc.* **80** 8–37.

Emphasizing next the differences between PP and CT for accuracy of discussion, we note:

(a) Parallel linear projections in PP are not fundamental—one could easily imagine nonparallel projections, fan-beam, or even curvilinear projections rather than straight line projections. Thus in PP the density is sampled,

projections are estimated, and then the density is approximated from the estimated projections. There is no intrinsic need for the intermediate step of forming projections in PP, while in CT the raw measurement data are linear projections because of the nature of the measurements and so the projections are intrinsically needed in CT.

(b) The estimates of interesting densities in CT are quantitatively accurate (for head sections) to .5% and achieve nominal spatial resolution of better than 2 mm, whereas PP is aiming at mere qualitative agreement.

(c) So far the algorithms of PP involve only "trivial" mathematics, there are no transforms or other subtle mathematical constructs.

(d) Finally, CT is widely used.

Are these criticisms of PP fair? Not really. For one thing, CT is aimed at a very specific problem—getting a cross-sectional image of an anatomical object. PP doesn't address this problem or even any precise problem, but like EDA instead attempts to take an almost arbitrary unspecified and unknown data set and find some structure or departure from white noise in it. One can only admire the ambition in such an undertaking. It is as if a new program was attempting to automatically translate a passage from an arbitrary and unknown language into English in the face of the failure of programs designed to translate from say Russian alone. However the success of EDA is well acclaimed. Thus not only has mission impossible been launched but it apparently was successful! Hats off to Peter Huber for taking the trappings of a known technique and expanding them to solve a central problem of EDA. I worry however that there may be dangers to using such general methodology. Returning to CT for helpful analogy, there are certainly densities in CT which simply cannot be reconstructed from a specified (under) sampling of their projections; thus the (signed) density q in Figure 1 is indistinguishable from the zero density if only projections along lines with slopes 0, ± 1, ∞ are taken and similar examples abound for any finite sampling. Of course this means that adding a small multiple of this signed density to a strictly positive one will leave similar ambiguities due to the undersampling. This is not a serious problem for CT only because human cross-sections don't involve q. I think it will become a gainful exercise for PP'ers to take a similar negative view to see if they can find similar examples of interest to a user of PP
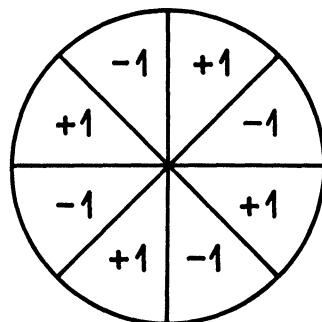

FIG. 1

where PP actually misinforms. Some examples are given in the last chapter of the paper. Are there any really bad ones? CT wasn't really accepted by radiologists until the existing counterexamples and artifacts were well understood and this was only achieved (I think) because the set of things one might want to place in the aperture of a CT scanner is severely limited. Not so for PP because of its great generality. Somehow the universe of possible data set candidates for PP should be defined and limited by a mathematical model.

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

Paul Switzer

Stanford University

Huber has given us an organized and well-classified account of diverse problems in statistics which may be approached from the point of view of projection pursuit. He also has brought to light certain connections which are not immediately obvious and indicated a number of important and challenging areas for further research. Very appropriately Huber warns us of the need for benchmarks and stopping criteria especially where iterative or stepwise searches are used to get at ever finer data structures. What follows are elaborations of several topics raised in the Huber paper.

Location, scale and structure. Huber has clearly pointed out the connection between the invariance structure of the projection index and the kind of problem which gets solved. For many interesting multivariate problems, global location and global scale questions are incidental and one might deal with orthonormalized data to begin with. At least the location and scale should be handled separately from the search for other structure. It seems that the same should also be true for density estimation although it is not clear whether or not Huber would agree.

Search methods. It seems worthwhile to distinguish between reconnaissance and pinpoint searches. Reconnaissance means that the p-dimensional orthonormalized data space is scanned unguided by jumping quickly through all orthants. A reasonable procedure might select interesting data projections from the class of $3^p/2$ projections of the form

$$\sum c_i x_i, \quad \text{where} \quad c_i = -c, 0, +c \quad \text{for} \quad i = 1, \cdots, p,$$

and c is a suitable constant such as 1.0 or 2.0. Some of the interesting projections may then be refined by a guided localized interactive pinpoint search. Reconnaissance of this kind is important if there may be multiple but well-separated projections of interest. Thus reconnaissance is limited to $p < 10$ or 15 and for larger p one must be resigned to having large unexplored possibilities.

Sometimes one may be interested in finding any interesting projection as opposed to all interesting projections or even a globally optimized projection.