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.

Such may be the case, for example, if one is looking for a pair of separated clusters. Then the search is often facilitated by the presence of redundant variables, albeit they are globally orthogonal, as well as useless variables. The implication is that interesting projections would exist for which relatively few variables have nonzero coefficients. Then some stepwise or hierarchical version of a reconnaissance search might permit the inspection of high-dimensional data sets.

Clusters. The first goal of a data analysis of multivariate observations is commonly to determine whether they should be regarded as homogeneous or whether they should be split into subsets each of which would permit a simpler description than the data set as a whole. Here reduction of dimensionality is not the explicit goal; it is not so important to find an optimal projection as opposed to one that does the job of dichotomizing the data. Interesting projections will be those which exhibit bimodality or multimodality as opposed to projections based on variance criteria or other criteria not explicitly tied to cluster separation.

Then one searches for further dichotomies using separate projections of each subset of the first dichotomy. Such a procedure was used, for example, in Switzer (1971) and accords with Huber's suggestion that iterative projections should always remove the structure uncovered at preceding steps of the iteration. It is important to note that clustering of data may occur on both large and small scales in the observation space. Once a partition of the data set into clusters has been tentatively achieved then the large clusters might each be separately reanalyzed by some form of principal components in order to find a concise description or reduction in dimensionality for each such cluster. The small clusters might be treated in the fashion of outlier data.

Classification. When the multivariate data are partitioned a priori into classes then one typically searches for a low-dimensional projection of the data in which the class centroids are well separated relative to the overall variation. This is the approach of Fisher's linear discriminant analysis, robustified by Huber in his section on two-sample problems. However, such analyses are usually preliminary to the real task of finding an assignment rule for future unclassified data points. This may then be accomplished by partitioning the low-dimensional space into assignment regions using bisectors between centroid pairs or some other algorithm which tries harder to minimize misclassification error rates.

However, it would seem that the search for a projection ought to be guided *from the start* by the objective of getting good assignment rules for future observations or, equivalently, characterizing how distributions differ in different classes—without forcing something like a location shift perspective on the problem. When there are more than two classes, especially when there are many classes, there are likely to be clusters of classes, that is, several scales of class separation. Then tree-like sequences of univariate projections, corresponding to hierarchical assignment rules are likely to be more useful than any single two- or three-dimensional projection.

Deconvolution of time series. Huber has described a class of projection pursuit procedures wherein the segments of length d from a univariate time series are treated as the basic d-dimensional observations. Projections which give rise to least normal univariate distributions are candidates for the desired filter or inverse filter to be applied to the time series. While considerable success is claimed for such procedures, their rationale seems to depend in part on supposing that the deconvolved series should be an i.i.d. sequence. In many geophysical problems the deconvolved series is expected to look like a step function corresponding to stratigraphy. This suggests that the projection index should pay some attention to the time order of the deconvolved series. For example, one might consider an index based on the scaled total variation of the deconvolved series.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

JOHN TUKEY

Princeton University

As always, it is a pleasure to see the carefully thought through and neatly organized result of Huberizing a field.

I shall confine my detailed comments to Section 19, where the $x_{p,t}$ are essentially the column sums of a Buys-Ballot table (e.g. Whittaker and Robinson, 1924).

This approach was once more conventional than the periodogram (not then yet invented). We can improve its behavior somewhat by replacing equally weighted sums, of X_{t+kp} over p, by windowed sums, where the (data) window tends finitely to zero at the nearest points which would have appeared in the sum if their values had been observable, but which were not observed.

The difficulty with harmonics and subharmonics can be minimized by beginning with the largest Fourier amplitude $|c(p)|^2$, which will also be better calculated with a data window (and, further, if more refined assessment of periods is desired, padded rather extensively with zeroes), and then using Buys-Ballot technique to identify—and then subtract—a general periodic constituent whose period is sufficiently close to the Fourier-selected period. A new set of Fourier amplitudes can then be found (cheaply by an FFT), and the cycle repeated.

Notice that

$$\sum (1/m^2)\cos 2\pi(2^{-m}f_0)t$$

shows that we cannot hope, whatever our approach, to always avoid selecting a harmonic of a frequency also present. So we must be prepared to also have a revision process, in which, once we have a good finite sum of periodic terms, we look for harmonic relations among their periods and corresponding reductions of the number of periodic terms. This is needed for the approach suggested above, as well as for any other approach.