DONOHO, D. and JOHNSTONE, I. M. (1985b). On convergence of projection pursuit algorithms. Manuscript in preparation.

FILL, J. A. and JOHNSTONE, I. M. (1984). On projection pursuit measures of multivariate location and dispersion. *Ann. Statist.* **12** 127–141.

HAMPEL, F. R. (1975). Beyond location parameters: robust concepts and methods. *Bull. Internat. Statist. Inst.* **46,** 375–382.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

KRASKER, W. and WELSCH, R. E. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.* **77** 595–604.

ROUSSEEUW, P. (1983). Multivariate estimation with high breakdown point. To appear in *Fourth Pannonian Symp. Math. Statist.*

ROUSSEEUW, P. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

ROUSSEEUW, P. and YOHAI, V. (1985). Robust regression by means of S-estimators. To appear in *Proc. Heidelberg workshop on Robust and Non-linear methods in Time Series Analysis.* Springer, Berlin.

D. DONOHO
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA AT BERKELEY
BERKELEY, CALIFORNIA 94720

I. JOHNSTONE
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

P. ROUSSEEUW
TECHNISCHE HOGESCHOOL DELFT
DEPARTMENT OF MATHEMATICS
AND INFORMATICS
JULIANALAAN 132
2628 BL DELFT
THE NETHERLANDS

W. STAHEL
SEMINAR FÜR STATISTIK
SOL, ETH
8092 ZÜRICH
SWITZERLAND

R. GNANADESIKAN AND J. R. KETTENRING

Bell Communications Research

With the recent flurry of interest in Projection Pursuit (PP), it is both timely and useful to have this paper by Professor Huber which attempts to unify these developments and more classical multivariate dimensionality reduction techniques.

We will limit our comments mostly to PP for finding clusters. The focus will be on the significant gap between the general ideas and existing theory about PP and what is important to and needed by practitioners.

The emphasis in PP on linear combinations and, in this paper, on affine invariance is in conflict with the frequent need to identify subsets of the given variables which contain the cluster structure. PP may help, but it is far from the end of the line as far as reduction of dimensionality and interpretation are concerned. Furthermore, the use of such invariant procedures can be misleading. For instance, the Class III type (squared) distance function, $[a'(x_i - x_j)]^2/a'Sa$, which utilizes an overall covariance matrix S, whether determined robustly or not, may be quite inappropriate for capturing cluster structure because S says nothing about within-cluster variability. Similar comments would seem to apply to Huber's invariant version of the Friedman-Tukey index.

Huber's claim in the Introduction that PP can avoid the "curse of dimensionality" is appropriately softened in the end: the sample size may need to be quite large relative to the dimensionality to avoid the effects of spurious structure. So, while PP may avoid the more blatant drawbacks of distance-based methods in handling extraneous variables, it does not provide a panacea for the challenging problem of handling a large number of variables simultaneously.

The computational problems associated with the use of general purpose optimization algorithms should not be overlooked. Users will need to worry about starting points, tuning constants, global vs. local optima, etc. There is clearly a trade-off between the use of PP indices which require such algorithms and those which can be solved as eigenvalue problems. In this latter class are many of the classical multivariate statistics as treated by Roy (1957) via the union-intersection principle (Roy, 1953). Indeed, this principle encompasses the spirit of PP as discussed in Section 6 of Huber's paper.

There are some quite useful projection methods directed at detecting "interesting" directions, too, that require no optimization algorithms at all; see, e.g., Andrews (1972) and Andrews et al. (1971, 1973). And there are others that piggyback on the results of direct clustering methods by making eigenvector-based projections of the clusters; see Gnanadesikan et al. (1982).

The choice of index is a critical issue as Friedman and Tukey cautioned in their 1974 paper. Whatever the index, it is essential to have some knowledge about its statistical behavior. Huber points out the lack of this knowledge in the regression case, but the same comment seems to apply to other situations in varying degrees as well. Proper scaling of the variables may be crucial to the performance of some indices. For many problems, there may be no answer that is both easy and effective. In the case of the Friedman-Tukey index, inappropriate scaling may affect the weights assigned to noise variables to the point where they have too large an impact on the distances between points along the projections. Unlike indices that are based either on variance maximization (e.g. principal components) or on maximal clustering (e.g. Kruskal, 1969), the Friedman-Tukey index combines both aspects. From an interpretational viewpoint, one wonders about the relative advantages of such a "combination."

Some perspective on the state of PP is revealed by comparing the development of it with the usual methods of cluster analysis. Direct clustering methods have evolved over a long period along a variety of fronts as discipline-oriented researchers found techniques and refinements of them for the real problems they faced. The result has been a wide range of imperfect yet workable and rather well-understood tools, but not much in the way of cohesive statistical theory. The novel aspects and clever ideas of PP, as addressed in this paper, have been put forth by a relatively small circle of statisticians. Some elegant results have been obtained, not only in Huber's paper but also in Diaconis and Freedman (1984). What is lacking is a broad experience base and perspective as to what PP methods will, and will not, do on a variety of real and simulated problems.

Hopefully, one after-effect of Huber's paper will be extensive empirical work to help pin down some of these issues that practitioners need to have resolved before they can use PP with confidence.

REFERENCES

ANDREWS, D. F (1972). Plots of high dimensional data. *Biometrics* **28** 125–36.

ANDREWS, D. F., GNANADESIKAN, R. and WARNER, J. L. (1971). Transformations of multivariate data. *Biometrics* **27** 825–40.

ANDREWS, D. F., GNANADESIKAN, R. and WARNER, J. L. (1973). Methods for assessing multivariate normality. In *Multivariate Analysis III.* (P. R. Krishnaiah, ed.) 95–116. Academic, New York.

DIACONIS, P. and FREEDMAN, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* **12** 793–815.

FRIEDMAN, J. H. and TUKEY, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.* **C-23** 881–889.

GNANADESIKAN, R., KETTENRING, J. R. and LANDWEHR, J. M. (1982). Projection plots for displaying clusters. In *Statistics and Probability: Essays in Honor of C. R. Rao.* (G. Kallianpur et al., eds.) 269–80, North-Holland, Amsterdam.

KRUSKAL, J. B. (1969). Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new 'index of condensation.' In *Statistical Computation.* (R. C. Milton and J. A. Nelder, eds.) Academic, New York.

ROY, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24** 220–38.

ROY, S. N. (1957). *Some Aspects of Multivariate Analysis.* Wiley, New York.

BELL COMMUNICATIONS RESEARCH
MORRISTOWN, NEW JERSEY 07960

TREVOR HASTIE AND ROBERT TIBSHIRANI[1]

Stanford University and Stanford Linear Accelerator Center

We would like to thank Professor Huber for this far-reaching yet penetrating discussion of projection pursuit methods.

Our comments will touch upon three areas: inference, the relation of PPDE to the Iterative Proportional Scaling algorithm, and the extension of PPR models to other settings.

1. Inference. Professor Huber discusses only briefly (Section 21) the problem of inference for PP models. But if PP is to be used for data analysis, we feel that this is an important question. We will concentrate on the PPR model, although qualitatively our findings should apply to PPDE and perhaps to other PP procedures as well. Suppose that we have fit a one-term PPR model of the form $\hat{y} = g(\hat{\mathbf{a}}'\mathbf{x})$ to a set of data with p predictors and n observations. An important question is: Is the direction $\hat{\mathbf{a}}$ really "significant," or just an artifact of our search over all possible directions? We can answer this by comparing the observed decrease in the corrected sum of squares $D = \sum (y_i - \bar{y})^2 - \sum (y_i - \hat{y}_i)^2$

[1] This work was supported by the Department of Energy under contracts DE-AC03-76SF and DE-AT03-81-ER10843, and by the Office of Naval Research under contract ONR N00014-81-K-0340, and by the U.S. Army Research Office under contract DAAG29-82-K-0056.