have

$$
\begin{aligned}
|f(x) - \bar{f}^{(k)}(x)| &\le (2\pi)^{-d/2} \int |f(x) - f(x - \sigma_k y)| e^{-\|y\|^2/2}\,dy \\
&\le (2\pi)^{-d/2} \Big\{ 2 \sup_x f(x) \int_{\|y\|>R} e^{-\|y\|^2/2}\,dy \\
&\qquad + \int_{\|y\|\le R} |f(x) - f(x - \sigma_k y)| e^{-\|y\|^2/2}\,dy \Big\},
\end{aligned} \tag{8}
$$

whose first term can be made arbitrarily small by choosing a large R since $\sup_x f(x) < \infty$ follows from the uniform continuity of f, and whose second term, for fixed R, can be made arbitrarily small (uniformly in x) by choosing a small σ_k (k large) again from the uniform continuity of f. This and (3) imply the uniform convergence of $\bar{g}^{(k)}$ to f. □

The uniform continuity condition on f is much weaker than the condition in Proposition 14.3 that f can be deconvoluted with a normal density.

Our last remark concerns the choice of σ_k in the smoother (1), which depends on the knowledge of τ_k. An *optimal* choice of σ_k can be obtained by equating the convergence rates of $\bar{g}^{(k)} - \bar{f}^{(k)}$ and $\bar{f}^{(k)} - f$. Let us further assume that f satisfies the Lipschitz condition of order λ

$$|f(x_1) - f(x_2)| \le C|x_1 - x_2|^{\lambda},$$

where C is independent of x_1, x_2. Then $|f(x) - \bar{f}^{(k)}(x)|$ in (8) is bounded above by $C'\sigma_k^{\lambda}$. This and the rate $\tau_k^{1/2}\sigma_k^{-d}$ in (3) are of the same order if

$$\sigma_k = c\tau_k^{1/2(d+\lambda)}.$$

REFERENCE

ROYDEN, H. L. (1968). *Real Analysis*. 2nd ed. Macmillan, New York.

INSTITUTE OF SYSTEMS SCIENCE
ACADEMIA SINICA
BEIJING 100080, PEOPLE'S REPUBLIC OF CHINA

1210 W. DAYTON ST.
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

D. R. COX

Imperial College, London

Dr. Huber's scholarly paper invests the impressive techniques of projection pursuit with a halo of mathematical formalism. Key questions clearly concern the choice of properties that it is *scientifically* fruitful to pursue. My judgment, based on totally inadequate experience, is that, except in fairly extreme cases, peculiarities of univariate distributional form are often of fairly fleeting interest and hence may not be good pointers to multidimensional structure. Consideration of bivariate properties and, in particular, the search for two directions with maximum curvature of regression (Cox and Small, 1978, Section 4.2) may be more promising. Certainly that gives quite direct diagnosis of both smooth nonlinearity and groups of points away from a broadly linear form. There is much scope for empirical and theoretical study.

REFERENCE

COX, D. R. and SMALL, N. J. H. (1978). Testing multivariate normality. *Biometrika* **65** 263–272.

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE
LONDON, ENGLAND SW7 2BZ

PERSI DIACONIS

Stanford University

This work makes a great contribution by introducing the unifying notion that projections are interesting if they minimize indices of randomness. Before, there was a sea of isolated, seemingly disjoint, ideas. Now there is some order, and a way of connecting the applied success stories of projection pursuit to more classical statistics. This often suggests new research projects.

One project involves notions of projection suitable for discrete data such as contingency tables and the analysis of preferences. I have introduced one such notion which involves projecting discrete data along "lines" of things like finite geometries. More formally, let X be a finite set (such as all binary k-tuples). Let f: $X \to \mathbb{R}$ be a summary of the data (the proportion of students with a given pattern of correct/incorrect in a k item test). Let Y be a class of subsets of X. The *Radon transform* of f at $y \in Y$ is the sum

$$\bar{f}(y) = \sum_{x \in Y} f(x)$$

The class Y is a *projection base* if it partitions into $y_1, \cdots, y_t$ where each y_i is itself a partition of X.

In the example, the sets $y_i^0 = \{x: x_i = 0\}$, and $y_i^1 = \{x: X_i = 1\}$ form a projection base. The Radon transform amounts to asking how many students answered the ith question correctly.

If the sets y_i are considered as lines in a geometry with points x, a projection base corresponds to the Euclidean axiom: for each point and line, there is a unique line through the point parallel to the given line. If $X = \mathbb{R}^p$, and Y is taken as all affine hyperplanes, the Radon transform gives ordinary projection.

A theory can be built in this generality. Many of the basic results seem to go through: for most data sets, most projections are close to uniform. Thus projections are interesting if they are far from uniform, and projection pursuit is forced on us.

I have analyzed several sets of discrete data using this approach. It leads to