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 $f\colon X \to \mathbb{R}$ 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 $y_i^0 = \{x\colon x_i = 0\}$, and $y_i^1 = \{x\colon X_i = 1\}$ form a projection base. The Radon transform amounts to asking how many students answered the $i$th 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 useful alternatives to log-linear models, multidimensional scaling or correspondence analysis. It sometimes points to structures that other analyses have missed.

It seems likely that a form of projection pursuit regression can be developed for "taking out" structure as it is detected. For simple projection bases, such as the affine hyperplanes in the space of binary $k$-tuples, Fourier analysis shows that any function can be exactly written as a finite linear combination of nonlinear functions as in my work with Mehrdad Shahshahani. There is much interesting work to be done in studying properties of a few linear combinations, or in studying more general projection bases.

John Tukey has suggested an alternative: mapping discrete data into a Euclidean space in several ways (say using a weighted linear combination $Ax$ of the binary vectors in the example) and then using projection pursuit. I wonder if Professor Huber sees other approaches to extending the usefulness of projection pursuit to discrete data.

The project sketched above is closely connected to my joint work with David Freedman. We showed that for data sets in "general position" in $\mathbb{R}^p$, most projections would be close to Gaussian. Of course, the Gaussian distribution has maximum entropy for fixed scale, while the uniform distribution has maximum entropy for discrete data. In both cases, theory points to the need for projection pursuit in the following form: projections are "interesting" if they have *minimum* entropy among projections being considered. David Aldous has pointed to work by V. N. Sudakov that has some overlap with these considerations.

Professor Huber mentions some calculations in connection with "grand tours": how many projections must be viewed to be sure of coming within a "squint angle" $\varepsilon$ of every view. I assume that the bounds follow from the following kind of reasoning: consider projecting onto "lines." Find the volume $V(\varepsilon)$ of a cap of radius $\varepsilon$ on the unit sphere $S_{p-1}$ in $p$-dimensions (the whole sphere has volume 1). Then at least $N(\varepsilon) = 1/V(\varepsilon)$ caps of radius $V(\varepsilon)$ are needed. Thus, at least $N(\varepsilon)$ projections are needed. Of course, these lower bounds are quite sloppy. To see the reason for my concern, Peter Matthews has used computations in Asimov (1985) to get upper and lower bounds on the number of planes needed to cover all two-dimensional views in $p$-dimensions. For example, in four dimensions, about 3258 planes must be used (see Table 1). If one plane is viewed per second, this would take about an hour. The upper bound for four dimensions is about 15 hours. It would help to know where the numbers in the present paper come from.

TABLE 1
*Upper and lower bounds for the number of two-dimensional projections needed*

| | Lower bound | Upper bound |
|---|---|---|
| $P = 3$ | 66 | 263 |
| $P = 4$ | 3258 | 51684 |
| $P = 5$ | 143529 | $0.9 \times 10^7$ |
| $P = 6$ | $0.6 \times 10^7$ | $0.2 \times 10^{10}$ |
| $P = 7$ | $0.2 \times 10^9$ | $0.2 \times 10^{12}$ |

Matthews (1985) has carried out many further computations in his Stanford Ph.D. thesis. He works in the context of a random walk on a group and derives the distribution until the walk hits (or is suitably close to) every point. This relates to projection pursuit via Asimov's scheme for the "grand tour." Asimov considers projections that "wiggle around" by small random rotations. His results agree with those reported by Huber in the following sense, it takes a long time to get close to most projections in high dimensions. Therefore, some form of projection pursuit is needed. On the other hand, once an interesting projection has been located, it seems useful to have some kind of grand tour to "wiggle around" in a neighborhood, to try to explore the features of the interesting projection.

## REFERENCES


ASIMOV, D. (1985). The Grand Tour. *SIAM J. Sci. Statist. Comput.* **6** 128–143.

DIACONIS, P. (1983). Projection pursuit for discrete data. Stanford University Technical Report #198.

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

DIACONIS, P. and GRAHAM, R. L. (1983). The Radon transform on $Z_2^k$. Stanford University Technical Report #206. To appear in *Pacific J. Math.*

DIACONIS, P. and SHAHSHAHANI, M. (1984). Some theory for projection pursuit regression. *SIAM J. Sci. Statist. Comput.* **5** 175–191.

MATTHEWS, P. (1985). Covering problems for random walks on spheres and finite groups. Doctoral dissertation, Dept. of Statist., Stanford University.

SUDAKOV, V. N. (1980). Typical distributions of linear functionals chosen at random. Colloques Internationaux Du Centre National de la Recherche Scientifique No. 307, pp. 501–511.



DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305



DAVID DONOHO, IAIN JOHNSTONE, PETER ROUSSEEUW, AND WERNER STAHEL

*University of California at Berkeley, Stanford University, Technische Hogeschool Delft, and ETH Zürich*


In our discussion of this very stimulating paper, we will mostly confine our remarks to some of the general issues Huber raises in the introductory paragraphs.

**1. The curse of dimensionality.** In paragraph four of the introduction, Huber writes "··· the most exciting feature of PP is that it is one of the very few multivariate methods able to bypass the 'curse of dimensionality' ··· "

Actually, Huber gives no precise definition of the "curse." Perhaps this is best, because there are several curses of dimensionality. Adverse effects of increasing dimension can include: less robustness, greater computational costs, worse mean squared error, and slower convergence to limiting distributions.

In this instance, Huber is concerned with the effects of increasing dimension on the mean-squared-error of smoothers. He points out that kernel and related