FRIEDMAN, J. H., STUETZLE, W. and SCHRÖEDER, A. (1984). Projection pursuit density estimation. *J. Amer. Statist. Assoc.* **79** 599–608.

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

HASTIE, T. (1983). Nonparametric logistic regression. Dept. of Statist., Stanford University, Report ORION 016.

HENRY, D. H. (1983). Multiplicative models in projection pursuit. Dept. of Statist., Stanford University, Report ORION 025.

MCDONALD, J. A. (1982). Projection pursuit regression on the ORION-I workstation. Film, Bin 88 Productions, Stanford Linear Accelerator Center, Stanford University.

MITTAL, Y. (1983). Two-dimensional projection pursuit tests for goodness-of-fit and equality of distributions. Dept. of Statist., Stanford University.

TIBSHIRANI, R. (1982). Censored data regression with projection pursuit. Dept. of Statist., Stanford University, Report ORION 013.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

FRED L. BOOKSTEIN

The University of Michigan

Professor Huber neatly reminds us how both principal-components analysis and multiple linear regression may be viewed as special cases of PP. In this note I would like to suggest a ramification of PP that incorporates the considerably wider class of *soft models* (iterated regression protocols), variants of canonical correlations analysis, lately developed by Herman Wold.

1. Section 7 of Huber's article considers "questions of k-dimensional projections." Rephrasing this as "one k-dimensional projection," I shall reverse the multiplicities to consider k one-dimensional projections instead. As Huber notes, for a computation of multiple dimensions to be most easily interpretable, each should be characterized uniquely. The simplest identification represents the measurement space $\mathbb{R}^n$ as the direct product $\prod_{i=1}^{k} \mathbb{R}^{n_i}$ of subspaces. In the language of causal modeling, these are *measurement blocks*.

Partition Huber's random vector $\mathbf{X}$ as $(X_1 : X_2 : \cdots : X_k)$, conformally with the projection vector $\mathbf{a} = (a_1 : a_2 : \cdots : a_k)$. Each X_i or a_i is a vector of length n_i, with $\sum_{i=1}^{k} n_i = d$, the original dimension of $\mathbf{X}$. Our "multiple projection" is then the k-vector $(Z_1, \cdots, Z_k) = (a_1^T X_1, \cdots, a_k^T X_k)$. An appropriate normalization sets each vector a_i to length 1, so that the vector $\mathbf{a}$ has Euclidean norm k rather than 1. The space of multiple projections, then, is no longer the unit d-sphere $\mathbb{S}^{d-1}$, but rather the direct product $\prod_{i=1}^{k} \mathbb{S}^{n_i - 1}$ of unit n_i-spheres.

2. PP finds interesting projections by the numerical examination of an objective function Q that measures "interestingness" in some fashion. For the objective functions $Q(Z)$ of a single projection, interestingness seems to have a useful general interpretation: nonnormality. For the multiple PP I am suggesting here, there is a more fundamental aspect of interest: *dependence*. For instance, for $k = 2$ we might use $Q(Z_1, Z_2) = \{\text{correlation}\}$. The projection vector $\mathbf{a} = (a_1, a_2)$ that

maximizes this Q gives us, of course, the usual first pair of *canonical variates* relating the X_1-block to the X_2-block.

Canonical variates may be computed by a scheme of *alternating regressions* (Lyttkens, 1972). Beginning with an arbitrary projection a_1 of the X_1-block, we regress $a_1^T X_1$ on the indicators of the X_2-block, regress the resulting predictor $a_2^T X_2$ on the indicators of the X_1-block, and so on, back and forth. Upon suitable normalization, the X_1- and X_2-predictors almost always converge to the first pair of canonical variates. For this objective function $Q = \{\text{correlation}\}$, therefore, we achieve the ends of PP by an alternating iteration of projections that considerably reduces the dimensionality of the task. This will prove an interesting paradigm.

Fits to other simple models involving correlations can be computed by this same strategy. Consider, for instance, an arbitrary projection a^T of one single block X of measurements. Regress $a^T X$ on each of the variates of the X-block separately, then sum all the predicted values. The new linear combination that results is proportional to $\Sigma_{XX} a^T$. Iteration of this cycle of summed simple regressions results in projections $\Sigma_{XX}^2 a^T, \cdots, \Sigma_{XX}^n a^T, \cdots$, that converge in direction, almost always, to the first principal component of the covariance matrix Σ_{XX}.

There is a two-block extension of this, the so-called MIMIC model. Beginning with an arbitrary projection $a_1^T X_1$, we regress it separately upon a list of n_2 predictors $X_{21} \cdots X_{2,n_2}$. The sum of the n_2 predicted values from these separate regressions may be regressed in turn upon the variables of the X_1-block as a whole; the predicted value from this multiple regression is the next candidate for a_1. Iterated, this procedure converges to the projection of the X_1-block that has highest summed squared correlations with the variables of the X_2-block severally. Wold (1975) surveys these and other one- and two-block iterated regression protocols.

3. To pass from these three examples (canonical correlations, principal components, MIMIC) to a prescription for general analysis of $k > 2$ blocks is not difficult. The regression protocol for the general soft model (Wold, 1980, 1982) consists of a series of these same elements—regressions of each projection Z_i upon the variables X_j of other blocks—in a sequence formally embodying a recursive causal scheme. Choices between simple and multiple regressions are involved also, and rules for combination of diverse predictors of the same Z_i from all the blocks to which it is linked in the recursion.

When all its regressions are multiple, any soft model is equivalent to a constrained canonical correlations analysis of the ultimately dependent block upon all the others (Bookstein, 1980, 1982). The constraints express the separation of the predictor blocks and break their symmetry. They drop the canonical r^2 below that of the first pair of ordinary canonical variates but, in exchange, enable us to interpret each projection $Z_i = a_i^T X_i$ as the value of a *latent variable* summarizing the import of all the variables of its block for other variables elsewhere in the scheme. According to an unpublished theorem of mine, for $k > 2$ these simple iterative protocols for such estimation all converge quickly to the true optimal constrained r^2.

This view of soft modeling as constrained canonical correlation is somewhat at odds with its usual presentation. Along with Jöreskog's LISREL, soft modeling is generally couched as a technique for causal modeling in the social sciences (Jöreskog and Wold, 1982). I find it much more realistically conceived as the species of multiblock PP I have just sketched: a search for the suite of projections, one per subspace, that bear to one another the optimum of some extended canonical relation.

4. This said, we may return, at last, to Professor Huber's theme. The simplest multiblock soft model, canonical correlations analysis, may be viewed as an extension of PPR to the case of a vector-valued dependent variable. The Y of Huber's Section 9 becomes a block of some dimension greater than 1. For gentle nonlinearities, at least, PPR could be generalized to this context by exploiting the "soft" strategy of back-and-forth regressions, each now a PPR. An algorithm would begin with a first guess at a summary predictand from one of the blocks (perhaps its ordinary first canonical variate with respect to the other block). There would follow the PPR of this single variable upon all the dimensions of the second block. The resulting predicted value is then regressed (by PPR) upon all the dimensions of the first block, and so on. By careful tuning of PPR, it should be possible to induce this algorithm to converge to the pair of ridge functions, one upon each block, with the highest correlation.

Such an extension of PP to two blocks, by invoking the iterative tactics of soft modeling, would often be very handy. For instance, within the tensor biometrics in which I specialize (Bookstein, 1984), the shape of a triangle is represented as one complex variable—a block of dimension 2. Any shape variable is a projection of this shape 2-space along a direction not known in advance of the analysis. Then any PPR involving shape data must "softly" construct its dependent variable Y at the same time that it produces a prediction function.

Two-block PP might be ideal for relating the *control space* of a complex process (settings of dials, etc.) to a multivariate vector of *state* or *output*. The modeling of a semistable phenomena by catastrophe theory, for instance, is such a search for a highly nonlinear, although stereotyped, structure in the relation of one block to a second.

Likewise, an extension of PPR is, in my view, the only promising means of searching for nonlinearity in the three-block soft models. Potential applications include blocks of variables that depend upon their measured values at several earlier times, and blocks of loosely measured outcomes (e.g., children's achievement) that depend upon precursors of two essentially different sorts (e.g., some neurological, some socioeconomic). In genetics, the relation between a vector of observations on offspring and the same vector of observations for each of two parents might submit to a three-block PPR for which the ridge expressions were constrained to have the same formula in each block. Such dependencies are sure to involve thresholds, interactions, and all the other delicious complexities with which PP seems uniquely suited to cope. In these ways PP should aid us to formulate a dependent variable at the same time that it helps us to untangle its dependencies.

The three-block soft models (Bookstein, 1980, 1982) unfold into PP by interpreting every multiple regression as a PPR. Any estimation would begin with a candidate dependent variable that is the first canonical variate of the dependent block with respect to the pool of all the other blocks. One would then cycle around a loop of three steps: (1a, 1b) separately regress the dependent block score (by PPR) upon each of the two predictor blocks; (2) regress the same dependent block score (by PPR) upon a small hybrid block of dimension 2 consisting of the pair of partial predictors from step 1; (3) regress (by PPR) this bivariate two-block predictor upon the variates of the dependent block. The prediction function becomes a revised dependent variable for step 1, and so forth until convergence, one hopes.

The extension of PP to two-block and multiple-block designs involves two themes: the search for k projections rather than one, and the iterative refinement of projections by alternating regression. Such an incorporation into PP of the two main themes of soft modeling should considerably enhance its power for the point clouds of complicated dimensional structure that arise in biometrics, interdisciplinary developmental studies, and all the other arenas for which "theoretical knowledge," in Wold's phrase, "is scarce." I thank Professor Huber and the editor of these *Annals* for the opportunity to see and explain this connection.

REFERENCES

BOOKSTEIN, F. L. (1980). Data analysis by partial least squares. In *Evaluation of Econometric Models.* (Kmenta and Ramsey, eds.) 75–88, Academic, New York.

BOOKSTEIN, F. L. (1982). The geometric meaning of soft modeling, with some generalizations. In *Systems under Indirect Observation: Causality, Structure, Prediction* (Jöreskog and Wold, eds.) II:55–74. North-Holland, Amsterdam.

BOOKSTEIN, F. L. (1984). Tensor biometrics for changes in cranial shape. *Ann. Hum. Biol.* **11** 413–437.

JÖRESKOG, K. and WOLD, H., eds. (1982). *Systems under Indirect Observation: Causality, Structure, Prediction.* 2 vols. Contributions to Economic Analysis, No. 139. North-Holland, Amsterdam.

KMENTA, J. and RAMSEY, J. B., eds. (1980). *Evaluation of Econometric Models.* Academic, New York.

LYTTKENS, E. (1972). Regression aspects of canonical correlations. *J. Multivariate Anal.* **2** 418–439.

WOLD, H. (1975). Path models latent variables: the NIPALS approach. In *Quantitative Sociology: International Perspective on Mathematical and Statistical Modeling.* (H. M. Blalock et al., eds.) 307–357, Academic, New York.

WOLD, H. (1980). Model construction when theoretical knowledge is scarce. In *Evaluation of Econometric Models* (Kmenta and Ramsey, eds.) 47–74, Academic, New York.

WOLD, H. (1982). Soft modeling: the basic design and some extensions. In *Systems under Indirect Observation: Causality, Structure, Prediction.* (Jöreskog and Wold, eds.) II:1–54, North-Holland, Amsterdam.

CENTER FOR HUMAN GROWTH AND DEVELOPMENT
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109

ANDREAS BUJA AND WERNER STUETZLE

University of Washington

Peter Huber's paper is interesting and important. In our opinion its main contributions are:

- The formulation of abstract versions of PPDE and PPR operating on distributions instead of samples. This complements the more intuitive un-