LANCASTER, H. O. (1969). *The Chi-squared Distribution*. Wiley, New York.
LEBART, L., MORINEAU, A. and WARWICK, K. M. (1984). *Multivariate Descriptive Statistical Analysis*. Wiley, New York (translated by E. M. Berry).
NAOURI, J. C. (1970). Analyse factorielle des correspondances continue. *Publ. Inst. Statist. Univ. Paris* **19** 1–100.
STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.* **22** 118–184.
STYAN, G. P. (1973). Hadamard products and multivariate statistical analysis. *Linear Algebra Appl.* **6** 217–240.
VAN RIJCKEVORSEL, J. (1982). Canonical analysis with B-splines. *COMPSTAT, IASC* 393–398.
YOUNG, F. W., TAKANE, Y. and DE LEEUW, J. (1978). The principal components of mixed measurement level multivariate data. *Psychometrika* **43** 279–281.

DEBORAH J. DONNELL
STATSCI, SUITE 500
1700 WESTLAKE AVENUE NORTH
SEATTLE, WASHINGTON 98109

ANDREAS BUJA
AT&T BELL LABORATORIES
ROOM 2C-261
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974-0636

WERNER STUETZLE
STATISTICS DEPARTMENT, GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

DISCUSSION

BERNARD D. FLURY
Indiana University

The past 15 years have shown considerable progress in generalizations and modifications of classical principal component analysis. This includes distribution theory for sample principal components in elliptical families [Muirhead (1982) and references therein], robust estimation and testing [Campbell (1980), Devlin, Gnanadesikan and Kettenring (1981) and Tyler (1981, 1983)], *common principal components* in several groups [Flury (1988)], principal components of patterned covariance matrices [Neuenschwander (1991) and Flury and Neuenschwander (1993)] and last but not least, generalizations to nonlinear situations, which constitute perhaps the thorniest area. Donnell, Buja and Stuetzle (DBS) give a fundamental building block in this field, the previous most significant building block being the *principal curves* of Hastie and Stuetzle (1989).

DBS stress that their method is rooted in the psychometric literature, but there is at least one (to my knowledge) direct predecessor in the statistical literature as well: Gnanadesikan and Wilk's (1969) *Generalized Principal Component Analysis*, described in detail also in Gnanadesikan (1977), Seber (1984) and Jackson (1991), which imitates the successful use of polynomials in regression. Gnanadesikan and Wilk's approach was to introduce powers and products of the original variables **X**, thus generating an expanded vector **Y**, and to compute the eigenvectors and eigenvalues of Cov(**Y**). Indeed, if all N data points $\mathbf{x}_1, \dots, \mathbf{x}_N$ satisfy a polynomial equation $r(\mathbf{x}_i) = 0$ of degree (say) q, then the coefficients of the polynomial, except for the intercept, can be recovered from the eigenvectors associated with the zero root of Cov(**Y**), provided all the necessary powers and products are contained in **Y**. By a vague continuity principle one would then expect this approach to reveal approximate polynomial relationships in multivariate data, but practical experience suggests that generally the method is not very successful because it is heavily affected by small errors in the data. This is not too surprising because with every additional power or product in the list the dimensionality of the problem is increased, giving small changes in the data an opportunity to influence the summary statistics. What both Gnanadesikan's approach and APCs have in common is their focus on the "small" side by searching for functions of the data which are almost constant. However, APCs go far beyond the mere addition of powers and products: they constitute what I consider a genuine generalization of principal components (see the next paragraph), and most importantly, they are theoretically well founded. Many data-analytic methods, particularly when they are formulated in terms of an algorithm to be applied to a finite number of data points, suffer from a lack of theoretical understanding. I find it always enlightening to study the properties of a method when it is applied to theoretical distributions rather than just to observed or fictitious data. The theory of population APCs developed in Section 4 of DBS is a prime example of what I mean.

Among the many attempts to generalize or modify principal components [see Jolliffe (1986), Chapters 11 and 12], not all are equally well-founded, so let me elaborate a bit on what I consider important criteria. What the more successful generalizations mentioned earlier have in common is that they focus on a defining property of principal components (like the ones at the beginning of Section 2 of DBS), and then find more general situations or models in which the same property still holds. In the case of *common principal components*, the defining property is the existence of an orthogonal matrix which diagonalizes the covariance matrix, and the generalization is to data in several groups. In *principal curves*, the defining property is orthogonal least squares, and the generalization is to nonlinear manifolds. In the case of APCs, the defining property is the minimization of variance, and the generalization is to sums of nonlinear functions. This is quite different from applying a standard method to situations for which it was not meant to work in the first place, like binary data or powers and products. Even more importantly, good generalizations allow a theoretical analysis; that is, there is a population (or model) version on which properties of the method can be studied.

Although, as DBS explain, APCs and principal curves focus on opposite ends of the spectrum, the two methods will inevitably be compared to each other. Having studied both methods theoretically, but having no practical experience with either of them, I believe that APCs will have more difficulty to become a widely applicable methodology. What I am most concerned about, despite the elegant horseshoe theory developed in Section 4.7, is the emergence of nonlinear functions in the simplest of situations, the multivariate normal. Users of multivariate statistical techniques often have a tendency to overinterpret results, a tendency which is likely to be reinforced if APCs are used uncritically. Inevitably, people will attempt to interpret "almost-constant" nonlinear relationships even if there is absolutely nothing in the data that would justify nonlinearity, as is the case with the multivariate normal. Given that nonlinear APC functions are found, how can we decide whether the structure is meaningful (like in the example of Section 3.5) or whether it is an artifact? It seems that the best solution is to recommend using only the smallest APC, but not the subsequent ones.

Otherwise, I can only support DBS's own remarks and precautions. Many crucial problems remain open, most prominently the questions of how to choose the class of nonlinear functions and how to handle multiple eigenvalues. The importance of the latter problem is evident from the ozone example, where the interpretation of the second and third smallest APCs might change drastically under small perturbations of the data.

REFERENCES

CAMPBELL, N. A. (1980). Robust procedures in multivariate analysis. I. Robust covariance estimation. *J. Roy. Statist. Soc. Ser. C* **29** 231–237.

DEVLIN, S. J., GNANADESIKAN, R. and KETTENRING, J. R. (1981). Robust estimation of dispersion matrices and principal components. *J. Amer. Statist. Assoc.* **76** 354–362.

FLURY, B. (1988). *Common Principal Components and Related Multivariate Models.* Wiley, New York.

FLURY, B. and NEUENSCHWANDER, B. (1993). Principal components and proportionality in patterned covariance matrices. In *Statistical Modelling and Latent Variables* (K. Haagen, D. J. Bartholomew and M. Deistler, eds.) 113–125. North-Holland, Amsterdam.

GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* Wiley, New York.

GNANADESIKAN, R. and WILK, M. B. (1969). Data analytic methods in multivariate statistical analysis. In *Multivariate Analysis* (P. R. Krishnaiah, ed.) **2** 593–638. Academic, New York.

JACKSON, J. E. (1991). *A User's Guide to Principal Components.* Wiley, New York.

JOLLIFFE, I. T. (1986). *Principal Component Analysis.* Springer, New York.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

NEUENSCHWANDER, B. (1991). Common principal components for dependent random vectors. Ph.D. thesis, Dept. Statistics, Univ. Bern.

SEBER, G. A. F. (1984). *Multivariate Observations.* Wiley, New York.

TYLER, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736.

TYLER, D. E. (1983). A class of asymptotic tests for principal component vectors. *Ann. Statist.* **11** 1243–1250.

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
RAWLES HALL
BLOOMINGTON, INDIANA 47405