REJOINDER

DEBORAH J. DONNELL, ANDREAS BUJA AND WERNER STUETZLE

Bellcore

We thank Professor Flury for his insightful remarks. In particular, we are glad that he pointed to a Gnanadesikan–Wilks reference that fell by the wayside as we shortened the paper. Since points of disagreement often expose critical ideas, we first take up Flury's assertion that comparisons between smallest additive principal components and *principal curves* are unavoidable. Here is a comparison of the major points of difference between the two methodologies:

Applicability.

- Principal curves and surfaces are successful if applied to data that are close to a smooth manifold of dimension 1 or 2.
- Smallest APCs are successful if applied to data that are close to a smooth manifold of low codimension, typically 1 or 2, sometimes 3, but rarely higher. In addition, it must be possible to describe the manifold by additive equations.

Output.

- The output of principal curves and surfaces is a parametrized representation of a curve or two-dimensional surface, with no constraint on the parameterization other than smoothness.
- The output of smallest APCs is a small number of implicit equations that are of an additive, and hence highly interpretable, form.

Uses.

- The main use of principal curves and surfaces is for visualization. They therefore provide in J. W. Tukey's terminology a "sharpening" tool that exposes geometric features of multivariate pointscatters.
- Smallest APCs have two uses: first as an analytical tool for fitting additive implicit equations to data; second as an adjunct of additive regression for diagnosing concurvity—the analog of collinearity in additive regression.

There is a narrow window where both principal curves/surfaces and APCs could be used: namely when the dimension is 1 or 2, the codimension is small and the manifold underlying the data is additive. Under these conditions both a principal curve/surface and a set of additive equations can be fitted. This may apply in three and sometimes even four dimensions, but any problem with a larger number of variables does not permit meaningful application of both methods.

An even crisper divergence occurs in the use of the methods, as parametrizations and implicit equations serve very different purposes. For example, the parametrization of principal curves is quite useless if the problem is to diagnose additive degeneracies among the predictor variables of an additive regression. On the other hand, implicit equations from APCs are not helpful if the problem is to capture nonlinearities in basically unidimensional data. As a general rule, if a problem asks for the application of one of these methods, it probably excludes the application of the other.

Having clarified the relation between APCs and principal curves (hopefully), we are still left with the nagging suspicion expressed by Flury: that, between these two methods, APCs may have a harder time catching on because of the horseshoe effect. Flury's reaction to artifactual parabolic and other nonlinear transforms is identical to reactions we observed after we pointed out the horseshoe effect for the ACE method [Buja and Kass (1985)]. To some, this seems to invalidate the approach as a whole, and the method is summarily dismissed. To us, this is a prime example of how we can be led astray by the study of theory alone. Unhappily, we do not seem to have put this theory in proper perspective. Below, we explain why we think that horseshoes are an entirely manageable problem in practice and give concrete rules for handling this phenomenon.

First, we relate some interesting new work in the theory of principal curves. Recent results by Duchamp and Stuetzle (1993) show that the population theory of principal curves is even more intricate than that of APCs. Similar to APCs, simple bivariate distributions such as the uniform on a rectangle can have multiple principal curves, some of which are nonlinear and as artifactual as horseshoes. Thus the conceptual advantage that Flury presumed for principal curves is nonexistent.

We return to the discussion of the horseshoe problem in APCs. The most important point is that horseshoes can be diagnosed in practice. Here is a multistep procedure that will pinpoint and eliminate the vast majority of horseshoes in real data:

1. Restrict attention to strongest APCs and ignore those whose eigenvalues are too large, that is, too close to 1. Weak APCs are more likely to have horseshoes.
2. Compare the transforms for a given variable across smallest APCs. If the variable has a very strong monotone transform in one APC, it is more likely to have horseshoe transforms in other APCs.
3. Use strong but uninteresting APCs for variable elimination: they point to relations among variables that destabilize the interesting APCs. Then recompute APCs of the remaining variables.

Each of these steps can be illustrated by the analysis of the full set of variables of the ozone data in subsection 3.2. Comparing Figures 1 through 3, one sees immediately that the number of parabolic transforms increases as the order of the APC goes up. Some of these parabolas are likely to be horseshoes. Comparing the transforms of *Sandburg temperature* across Figures 1 through 3, we find two strong monotone transforms in the first two APCs, hence our suspicion falls on the parabola for this variable in Figure 3. Similarly, the extremely strong showing of *inversion base temperature* in the first APC would indicate a horseshoe for the parabolas in the remaining figures. Elimination of some of these variables (in particular *inversion base temperature* which is highly related with *Sandburg temperature*) gets rid of most of these horseshoes as can be seen from Figures 5 and 6.

The third step of variable elimination is of course analogous to variable elimination in regression. When APCs are used for analysis of data dependencies, rather than for diagnostics in additive regression, the goal is again fitting stable equations so variable elimination may be desirable.

As a final point, note that the nonmonotonic transforms of the *day of year* variable are real on subject matter grounds: they reflect yearly periodicities in climate. Arguments of this kind are needed to establish the reality of nonmonotonic transforms.

Although the rules listed above serve well in practice, any data fitting method will always pose questions of adequacy of particular data descriptions. For example, a degenerate uniform distribution on the unit circle is well described by the equation $X^2 + Y^2 = 1$ (which is found by smallest APCs with an eigenvalue 0); on the other hand, a bivariate standard normal is not well described by any equation (as reflected by APC eigenvalues that are all identically equal to 1). Consider now a family of distributions that interpolates these two cases: $p(x,y) \propto (1 - x^2 - y^2)^{a-1}$ for $x^2 + y^2 < 1$ and 0 otherwise, where $0 < a < \infty$. (this is the circular family of Pearson Type II distributions. It was used to illustrate the horseshoe effect for the alternating least squares and ACE methods [Buja (1990)].) For $a \to \infty$ the family approaches the degenerate uniform on the unit circle, and for $a \to \infty$ it approaches (after suitable rescaling) the bivariate standard normal. The smallest APC eigenvalue is $\lambda = a/(a + 1/2)$ and the corresponding additive implicit equation is $X^2 + Y^2 = 1$ for all a [Buja (1990), page 1054 ff, and Section 4.6 above]. The question is: at what point do we consider this equation an adequate description of the data? Clearly, for values of a close to 0, that is, distributions close to the uniform on a circle, this is a good summary. For values $a > 1$, we may agree that it is worthless since the corresponding densities are bell-shaped. But for $a = 1$ (corresponding to the uniform on the unit disk), $\lambda = 2/3$ is still not a convincing smallest eigenvalue. How far do we need to go?

There is of course no single answer to this question. For comparison, in regression problems, there is no single recommended value of R^2 that would assure a meaningful regression either, yet this vagueness does not invalidate regression as a data fitting method. We have become used to testing, diagnosing and otherwise assessing regression equations, and similar efforts are necessary to validate APC equations.

REFERENCES

DUCHAMP and STUETZLE (1993).

BELLCORE
BOX 1910
445 SOUTH STREET
MORRISTOWN, NEW JERSEY 07962-1910