REFERENCE

WHITTAKER, E. T. and ROBINSON, G. (1924). *The Calculus of Observations.* Blackie, London.

DEPARTMENT OF STATISTICS
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544

GRACE WAHBA

University of Wisconsin-Madison

I would like to thank Professor Huber for a most thought-provoking paper.

I will restrict myself to discussion of projection pursuit regression, in particular to describing an approach to PP regression different than and possibly complementary to that of linear combinations of ridge functions as introduced in Section 9. This approach may be called partial spline modelling. One models a function of, say, $k + d$ variables, parametrically in k variables and as a (thin plate) spline function in the remaining d variables. The role of projection pursuit is to determine which d of the $k + d$ variables must be splined. Partial spline modelling can also be extended to the context of GLIM models, whereby again the dependency on some variables is via the usual GLIM approach while the dependency on other variables is only "smooth." It will turn out that partial spline estimates are linear combinations of (uniquely determined) polynomials and shifted versions of certain spherically symmetric functions (in the d splined variables). These splines are known to nicely model in a nonparametric way the interaction effects among a small number of variables (provided there is enough data), so they in some sense have properties that are complementary to ridge function approximation, and thus may be expected to do well where ridge functions do not. The structure of these estimates hopefully allows the benefits of the lesser data requirements of parametric modelling where that is warranted, while doing smooth nonparametric regression where it is not. By analogy with Huber's definition of "interesting" as "nonnormal" in multivariate density estimation, one could define "interesting" in this context as having a dependence more complicated than that modellable by a low-degree polynomial. With that definition, projection pursuit with the models being proposed here would, if successful, identify the "interesting" directions, particularly if the choice of d is part of the "pursuit."

Several authors have proposed partial spline models with one splined variable, with notable success (Engle, Granger, Rice and Weiss, 1983; Green, Jennison and Seheult, 1983; Anderson and Senthilselvan, 1982; Shiller, 1984). In all of these interesting applications the choice of the (single) splined variable was dictated by the context, so that projection pursuit is not necessary. Gaver and Jacobs (1983), however, consider the problem of predicting low level stratus conditions at Moffet field using surface meteorological measurements of east and north wind velocity, temperature, dew point and existence or nonexistence of low level stratus on preceding days. They use subset selection combined with logistic regression to make this forecast, but suggest that further work with this data set is warranted. We will below describe projection pursuit partial spline models, both in the context of normally distributed data, and in the context of alternatives to (or rather as bona fide generalizations of) logistic regression, and then propose that this approach merits further investigation as a procedure for modelling the Gaver-Jacobs data and similar multivariate data sets.

For ease of exposition, we will describe only a rather simple partial spline model, splined in $d = 2$ variables and linear in k other variables. Several (much) more general forms may be found in Wahba (1984a,b). For normally distributed data, this model is

$$(1) \qquad y_i = f(x_1(i), x_2(i)) + \textstyle\sum_{j=1}^{k} \theta_j z_j(i) + \varepsilon_i, \quad i = 1, 2, \cdots, n$$

where f is a "smooth" function and the $\{e_i\}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$. f and $\theta = (\theta_1, \cdots, \theta_k)$ are obtained as the minimizers, in an appropriate space, of

$$(2) \qquad (1/n) \textstyle\sum_{i=1}^{n} (y_i - f(x_1(i), x_2(i)) - \sum_{j=1}^{k} \theta_j z_j(i))^2 + \lambda J(f)$$

where $J(f)$, the smoothness penalty, is given by

$$(3) \qquad J(f) = \int_{-\infty}^{\infty} \int (f_{x_1x_1}^2 + 2f_{x_1x_2}^2 + f_{x_2x_2}^2)\, dx_1\, dx_2.$$

It is known (see e.g. Wahba and Wendelberger (1980) and references cited therein) that if $\lambda > 0$ and the $(k + 3) \times n$ design matrix T for (inhomogenous) linear least squares regression in x_1, x_2 and $z_1, \cdots, z_k$ is of rank $k + 3$, then (2) will have a unique minimizer (in an appropriate space) in (f, θ) with a representation of the form

$$(4) \qquad f_\lambda(x) = b_0 + b_1 x_1 + b_2 x_2 + \textstyle\sum_{i=1}^{n} c_i E(|x - x(i)|)$$

where $x = (x_1, x_2)$, $|x - x(i)|$ is the Euclidean distance between x and $x(i)$, $E(|\tau|) = |\tau|^2 \log|\tau|$, and $T'c = 0$. Due to the radial nature of the E's as a function of x, this form of approximation is quite different than approximation by ridge functions. The smoothness parameter λ may be chosen as the generalized cross-validation (GCV) estimate $\hat{\lambda}$ of λ (see Craven and Wahba, 1979). Given $\hat{\lambda}$, the coefficient vectors b, c and θ satisfy a linear system and may be obtained, for fairly large data sets, by the numerical methods in Bates and Wahba (1983). In particular when n is very large, the sum in (4) can frequently be well approximated by a smaller sum. When f is of the form (4), $J(f)$ is a (known) quadratic form in c. As $\lambda \to \infty$, $c \to 0$, and the result reduces to linear least squares regression on x_1, x_2, and $z_1, \cdots, z_k$. (More general partial spline models result in this limiting least squares linear regression being replaced by low degree polynomial regression in $d + k$ variables with d general.)

Now, if we are doing exploratory semiparametric regression given data sets $(y_i, \tilde{x}_1(i), \tilde{x}_2(i), \cdots, \tilde{x}_{k+2}(i))$ we may ask which variables should play the role of x_1 and x_2. If x_1 and x_2 are to be a two-member subset of $\tilde{x}_1, \cdots, \tilde{x}_{k+2}$, then we only have to choose from among the $\binom{k+2}{2}$ possibilities. If, however, we let x_1 and x_2 be linear combinations of these variables, then we are seeking an "optimal" two-dimensional subspace of Euclidean d space in which the splining takes place, and this leads to challenging search problems. We do not so far claim to know what good projection indices (Q in Huber's notation) are. Two possibilities which suggest themselves are $\hat{\lambda}$ and $J(f_{\hat{\lambda}})$ since both are in some sense measures of the deviation of the model from linear regression. Letting $A(\lambda)$ be the influence matrix (which relates the data vector y to the estimated data vector $\hat{y}$), one might consider as candidates for projection indices the residual sum of squares, RSS

$$\mathrm{RSS}(\hat{\lambda}) = (1/n)\|(I - A(\hat{\lambda}))y\|^2$$

the estimate $\hat{\sigma}^2$ of σ^2 given by

$$\hat{\sigma}^2 = \|(I - A(\hat{\lambda}))y\|^2/\mathrm{tr}(I - A(\hat{\lambda}))$$

or the GCV function V given by

$$V(\hat{\lambda}) = \|(I - A(\hat{\lambda})y\|^2/[\mathrm{tr}(I - A(\hat{\lambda}))]^2.$$

These are candidates for the projection indices since all of them are related to model fit in some way (we remark that $J(f_{\hat{\lambda}}) = \lambda^{-1}y'(I - A(\hat{\lambda}))A(\hat{\lambda})y)$, although they only really make sense if a good value of λ is used. Criteria based on the spline regression diagnostics of Eubank (1984) may possibly be of use, along with some clever procedure which looks for patterns in the residuals. In attempting to develop asymptotic properties of these criteria, one will probably find that the results depend on whether or not there is some partial spline model which describes the "truth." I will stick my neck out and conjecture that if some partial spline model within the class tried is the truth, then the GCV function will asymptotically find it (or, maybe, at least have a local minimum). What happens if no partial spline model within the context tried is true is anybody's guess. Fully splined models have nice convergence properties with respect to all members of certain Sobolev spaces (see e.g. Cox, 1984). Thus, one would very much like to have some criteria for determining whether splining in two of the given variables is adequate or whether one should go on and try to spline in three or more of the available variables.

In the Gaver-Jacobs data, we will suppose that y is an indicator function for low level stratus and is binomial $B_i(1, p)$ with $p = p(\tilde{x}_1, \cdots, \tilde{x}_{k+d})$. In the special case $p = p(x_1, x_2)$, O'Sullivan (1983) and O'Sullivan, Yandell and Raynor (1984) have proposed estimating the logit $f(x)$

$$f(x) = \ln[p(x)/(1 - p(x))]$$

via a maximum penalized likelihood method which estimates f as the minimizer of

$$-\{\textstyle\sum_{i=1}^{n} y_i f(x(i)) - \ln(1 + e^{f(x(i))})\} + \lambda J(f). \tag{5}$$

Note that the term in brackets is just the log likelihood

$$\ln \pi_i\{p_i^{y_i}(1 - p_i)^{1-y_i}\} = \textstyle\sum y_i[p_i/(1 - p_i)] + \ln(1 - p_i)$$

The minimizer f_λ is known to have a representation of the form (4). O'Sullivan and O'Sullivan et al. provide an approximate form of GCV for estimating λ and an iterative method for computing f_λ. The penalized GLIM models in O'Sullivan (1983), of which this is one example, immediately extend to partially penalized GLIM models by (for example, here) simply replacing $f(x(i))$ in the term in brackets by $f(x(i)) + \sum_{j=1}^{k} \theta_j z_j(i)$. Since

$$p = \exp(f + \sum \theta_j z_j)/(1 + \exp(f + \sum \theta_j z_j))$$

this model reduces to the (usual) logistic regression model as $\lambda \to \infty$. We hesitate even further here to suggest a projection index. The modified GCV function as proposed by O'Sullivan and O'Sullivan et al. is one possibility. In examples such as the Gaver-Jacobs problem where huge quantities of data are available one might just use the classical cross-validation procedure of reserving a large data set and letting Q depend on predictive ability with respect to the omitted data.

REFERENCES

ANDERSON, J. A. and SENTHILSELVAN, A. (1982). A two-step regression model for hazard functions. *Appl. Statist.* **31** 44–51.

BATES, D. and WAHBA, G. (1983). A truncated singular value decomposition and other methods for generalized cross-validation. University of Wisconsin-Madison Statistics Dept. (TR No. 715).

COX, D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31** 377–403.

ENGLE, R., GRANGER, C., RICE, J. and WEISS, A. (1983). Nonparametric estimates of the relation between weather and electricity demand. Discussion paper 83-17, Dept. of Economics, University of California, San Diego.

EUBANK, R. (1984). The hat matrix for smoothing splines. *Statist. Probab. Lett.* **2** 9–14.

GAVER, D. and JACOBS, P. (1983). Low-level stratus prediction using binary statistical regression. Naval Postgraduate School, Monterey, CA (Report No. NPS55-83-034).

GREEN, P., JENNISON, C. and SEHEULT, A. (1983). Comments to Nearest Neighbour (NN) Analysis of Field Experiments by Wilkinson, et al., *J. Roy. Statist. Soc. Ser. B.* **45** 193–195.

O'SULLIVAN, F. (1983). The analysis of some penalized likelihood estimation schemes. University of Wisconsin-Madison, Statistics Dept. (TR No. 726).

O'SULLIVAN, F., YANDELL, B. and RAYNOR, W. (1984). Automatic smoothing of regression functions in generalized linear models. University of Wisconsin-Madison, Statistics Dept. (TR No. 734).

SHILLER, R. J. (1984). Smoothness priors and nonlinear regression. *J. Amer. Statist. Assoc.* **79** 609–615.

WAHBA, G. (1984a). Partial spline models for the semiparametric estimation of functions of several variables. In *Statistical Analyses of Time Series.* Institute of Statistical Mathematics, Tokyo, 319–329.

WAHBA, G. (1984b). Cross validated spline methods for the estimation of multivariate functions from data on functionals. In *Statistics: An Appraisal, Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory,* (H. T. David and H. A. David, eds.) Iowa State University Press.

WAHBA, G. and WENDELBERGER, J. (1980). Some new mathematical methods for variational analysis using splines and cross-validation. *Monthly Weather Rev.* **108** 36–57.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53707