This algorithm is a fast and parsimonious way for representing interaction. For example, if, in their spline bases, $f$ and $g$ have $p$ degrees of freedom, then the minimizing product $fg$ has about $p$ degrees of freedom in it. One adds more multiplicative terms until there is no significant decrease in RSS. Furthermore, the multiplicative terms are easy to interpret.

Unfortunately, numerical results indicate that in the nonindependence case, there are a number of local minima in addition to the global minimum. The algorithm always converges, but it may not converge to the global minimum. This makes the selection of a good starting point important. Our experimental results have been that if we use the starting point given by assuming independence, then the iterates have always converged toward the global minimum.

I am currently working on straightening up the details of this representation of bivariate interaction and hope to go public soon.

## REFERENCES

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* **80** 580–619.

BREIMAN, L. and PETERS, S. (1988). Comparing automatic bivariate smoothers. Technical Report, Dept. Statistics, Univ. California, Berkeley.

FRIEDMAN, J. H. and STUETZLE, W. (1982). Smoothing of scatterplots. Technical Report, Orion 3, Dept. Statistics, Stanford Univ.

HASTIE, T. and TIBSHIRANI, R. (1988). Comment on "Monotone regression splines in action" by J. O. Ramsay. *Statist. Sci.* **3** 450–456.

SMITH, P. (1982). Curve fitting and modelling with splines using statistical variable selection techniques. NASA Contractor Report 66034.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

ZEHUA CHEN, CHONG GU AND GRACE WAHBA[1]

*University of Wisconsin-Madison*

We must begin by thanking the authors for a thought-provoking work. As is well known [Kimeldorf and Wahba (1971) and Wahba (1978)], quadratic penalized likelihood estimates (with nonnegative definite penalty functionals) are Bayes estimates. Let $\mathbf{y} = \mathbf{g} + \boldsymbol{\varepsilon}$ with $\mathbf{g} \sim N(0, \Sigma)$ and $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$, then

$$\hat{\mathbf{g}} = \Sigma(\Sigma + \sigma^2 I)^{-1}\mathbf{y} = A\mathbf{y}, \quad \text{say},$$

which also minimizes $(1/\sigma^2)(\mathbf{y} - \hat{\mathbf{g}})'(\mathbf{y} - \hat{\mathbf{g}}) + \mathbf{g}'\Sigma^{+}\mathbf{g}$, the resulting smoother matrices are all symmetric nonnegative definite with their eigenvalues in $[0, 1)$. This generalizes to the case where $\Sigma$ is improper, which gives eigenvalues $+1$.

[1]Research supported by AFOSR Grant AFOSR 87-0171.

One could make the case that a good linear smoother should be Bayes (including improper Bayes) for some Gaussian prior, equivalently, satisfy some appropriate quadratic variational problem.

We hope you will forgive one of us (G.W.) for claiming the "popular definition" of degrees of freedom for signal in the spline literature [Wahba (1983)]. This can stand as a challenge to Steve Stigler to show that this definition is really due to Laplace!

We have been looking at a (large) class of abstract multivariate smoothing models using reproducing kernel spaces, which include the additive as well as the interaction spline models as special cases, and trying to determine the feasibility of estimating multiple smoothing parameters by GCV. Our comments cover two areas: First, we remark that the backfitting algorithm can be used to fit this class of abstract models, if one specifies the smoothing parameters a priori. Second, we note a new algorithmic development based on matrix decompositions, which allows the estimation of multiple smoothing parameters with large $n$ (our examples involve $n$ from 300–800). The complexity of the algorithm depends on the number of data points, but not the dimensionality of the independent variable. We describe this abstract development and how it relates to the present work, and possibly answer some questions raised by the authors.

Let $\mathbf{x} = (x_1, \dots, x_d) \subset \Omega \in E^d$, $H$ a reproducing kernel (rk) Hilbert space of real-valued functions of $\mathbf{x}$, with an orthogonal decomposition

$$H = H_0 + \sum_{k=1}^{q} H^k,$$

where $H_0$ is spanned by $\phi_1, \dots, \phi_M$, and $H^k$ has the rk $Q^k(\mathbf{x}; \mathbf{x}')$. Thus, letting $H_1$ be $\Sigma_{k=1}^q \oplus H^k$, $H_1$ has the rk $Q(\mathbf{x}; \mathbf{x}') = \Sigma_{k=1}^q Q^k(\mathbf{x}; \mathbf{x}')$. We seek $f = \Sigma_{k=0}^q f_k$ with $f_0 \in H_0$, $f_k \in H^k$, to minimize

$$\frac{1}{n} \sum_{i=1}^{n} \left( y_i - \sum_{k=0}^{q} f_k(\mathbf{x}(i)) \right)^2 + \lambda \sum_{k=1}^{q} \theta_k^{-1} \| f_k \|_{H_k}^2, \tag{1}$$

and it is required that the $\mathbf{x}(i)$, $i = 1, \dots, n$, be such that least-squares regression on $H_0$ be unique. Using the theory of reproducing kernels [see Kimeldorf and Wahba (1971)], we have that there exists a unique minimizer and the $\hat{f}_k$ must be of the form

$$\hat{f}_0(\mathbf{x}) = \sum_{\nu=1}^{M} d_\nu \phi_\nu(\mathbf{x}),$$

$$\hat{f}_k(\mathbf{x}) = \sum_{i=1}^{n} c_{ik} Q^k(\mathbf{x}; \mathbf{x}(i)),$$

where the $\mathbf{d} = (d_1, \dots, d_M)'$ and $\mathbf{c}_k = (c_{1k}, \dots, c_{nk})'$ are found as the minimizers of

$$\frac{1}{n} \left\| \mathbf{y} - T\mathbf{d} - \sum_{k=1}^{q} Q^k \mathbf{c}_k \right\|^2 + \lambda \sum_{k=1}^{q} \theta_k^{-1} \mathbf{c}_k' Q^k \mathbf{c}_k, \tag{2}$$

where $T$ is the $n \times M$ matrix with $i\nu$th entry $\phi_\nu(\mathbf{x}(i))$ and $Q^k$ is the $n \times n$ matrix with $ij$th entry $Q^k(\mathbf{x}(i); \mathbf{x}(j))$. $T$ is by assumption of rank $M$. If $Q^k$ is not of full rank, then the minimizing $\hat{\mathbf{c}}_k$ of (2) may not be uniquely determined, but $\hat{f}_k(\mathbf{x})$, and in particular $\hat{\mathbf{f}}_k$, the vector of values of the smoothed component in $H^k$, given by $\hat{\mathbf{f}}_k = Q^k \mathbf{c}_k$, is uniquely determined. To see this, one supposes that $\mathbf{u} = (u_1, \dots, u_n)'$ satisfies $Q^k\mathbf{u} = \mathbf{0}$, then the nonnegative definiteness of $Q^k(\cdot\,;\cdot)$ ensures that $\sum_{i=1}^n u_i Q^k(\mathbf{x}; \mathbf{x}(i)') = 0$, all $\mathbf{x}$. One of us (Z.C.) has observed that the backfitting algorithm can be used here. Let $S_0 = T(T'T)^{-1}T'$, $S_k = Q^k(Q^k + \lambda\theta_k^{-1}I)^{-1}$. One observes in the usual way, by first fixing all of the $\mathbf{c}_k$'s and minimizing (2) with respect to $\mathbf{d}$, and then fixing all but $\mathbf{c}_k$ in turn for each $k$, that the minimizers $\hat{\mathbf{d}} = (\hat{d}_1, \dots, \hat{d}_M)'$ and $\hat{\mathbf{c}}_k$, $k = 1, \dots, q$, satisfy

$$\begin{pmatrix} I & S_0 & \cdots & S_0 \\ S_1 & I & \cdots & S_1 \\ \cdots & \cdots & \cdots & \cdots \\ S_q & S_q & \cdots & I \end{pmatrix} \begin{pmatrix} \hat{\mathbf{f}}_0 \\ \hat{\mathbf{f}}_1 \\ \cdots \\ \hat{\mathbf{f}}_q \end{pmatrix} = \begin{pmatrix} S_0\mathbf{y} \\ S_1\mathbf{y} \\ \cdots \\ S_q\mathbf{y} \end{pmatrix},$$

where $\hat{\mathbf{f}}_0 = T\hat{\mathbf{d}}$, $\hat{\mathbf{f}}_k = Q^k\hat{\mathbf{c}}_k$, $k = 1, \dots, q$.

We recall that the Bayes model corresponding to (1) is

$$y_i = \sum_{k=0}^{q} f_k(\mathbf{x}(i)) + \varepsilon_i, \qquad i = 1, \dots, n,$$

where the $f_k(\cdot)$ are independent, zero mean Gaussian processes with $f_0(\mathbf{x}) = \sum_{\nu=0}^M \delta_\nu \phi_\nu(\mathbf{x})$, with $\delta \sim N(0, \xi I)$ with $\xi \to \infty$, and $Ef_k(\mathbf{x})f_k(\mathbf{x}') = b\theta_k Q^k(\mathbf{x}; \mathbf{x}')$, $\varepsilon \sim N(0, \sigma^2 I)$, $n\lambda = \sigma^2/b$.

In recent work on interaction splines [Gu, Bates, Chen and Wahba (1988), Gu (1988) and Wahba (1988)] we have instead pursued the following approach which uses matrix decompositions for $n + M$ unknowns (not $np$!) and allows the explicit computation of the influence matrix and the estimation of multiple smoothing parameters by GCV. We first observe that if we change the squared norm in $H_1$ from $\sum_{k=1}^q \|f_k\|_{H^k}^2$ to $\sum_{k=1}^q \theta_k^{-1}\|f_k\|_{H^k}^2$ we change the rk from $\sum_{k=1}^q Q^k(\mathbf{x}; \mathbf{x}')$ to $\sum_{k=1}^q \theta_k Q^k(\mathbf{x}; \mathbf{x}') = Q_\theta(\mathbf{x}; \mathbf{x}')$, say. Letting $f \in H$ and $P_\theta$ be the projection in $H$ onto $H_1$ with this new norm, (1) is equivalent to: Find $f \in H$ to minimize

$$\frac{1}{n}\sum_{i=1}^n (y_i - f(\mathbf{x}(i)))^2 + \lambda\|P_\theta f\|_{H_1^\theta}^2,$$

where

$$\|P_\theta f\|_{H_1^\theta}^2 = \sum_{k=1}^q \theta_k \|f_k\|_{H_k}^2.$$

Using standard calculations [see, e.g., Wahba (1985)], the minimizer $f_\lambda$ satisfies

$$f_\lambda(\mathbf{x}) = \sum_{\nu=1}^M d_\nu \phi_\nu(\mathbf{x}) + \sum_{i=1}^n c_i Q_\theta(\mathbf{x}; \mathbf{x}(i)),$$

where $\mathbf{c} = (c_1, \dots, c_n)'$ and $\mathbf{d} = (d_1, \dots, d_M)'$ satisfy

$$(Q_\theta + n\lambda I)\mathbf{c} + T\mathbf{d} = \mathbf{y},$$
$$T'\mathbf{c} = 0,$$

and $Q_\theta$ is the $n \times n$ matrix with $ij$th entry $Q_\theta(\mathbf{x}(i); \mathbf{x}(j))$.

Letting the $QR$ decomposition of $T$ be

$$T = (F_1 F_2)\binom{R}{0},$$

we have that

$$\hat{\mathbf{c}} = F_2(F_2'Q_\theta F_2 + n\lambda I)^{-1}F_2'y,$$
$$R\hat{\mathbf{d}} = F_1'(\mathbf{y} - F_1 Q_\theta \mathbf{c}),$$

and the components of the solution are

$$\hat{f}_0(\mathbf{x}) = \sum_{\nu=1}^{M} \hat{d}_\nu \phi_\nu(\mathbf{x}),$$
$$\hat{f}_k(\mathbf{x}) = \sum_{i=1}^{n} \hat{c}_i \theta_k Q^k(\mathbf{x}; \mathbf{x}(i)),$$

which gives

$$\hat{\mathbf{f}}_0 = T\hat{\mathbf{d}}, \tag{3}$$
$$\hat{\mathbf{f}}_k = \theta_k Q^k \hat{\mathbf{c}}, \qquad k = 1, \dots, q. \tag{4}$$

We remark that this smoothing procedure reproduces elements in $H_0$, and we also note that Proposition 3 is a special case of (4) when $M = 0$, $F_2 = I$.

The influence matrix $A(\lambda, \theta)$ (called $\mathbf{R}$ by the authors), which satisfies

$$\mathbf{f}_+ = A(\lambda, \theta)\mathbf{y},$$

where $\mathbf{f}_+ = \sum_{k=0}^{q} \mathbf{f}_k$, is given by

$$A(\lambda, \theta) = I - n\lambda F_2(F_2'Q_\theta F_2 + n\lambda I)^{-1}F_2' \tag{5}$$

and has $n - M$ eigenvalues in $[0, 1)$, and $M$ eigenvalues $+1$. Letting $\Sigma_\theta = F_2'Q_\theta F_2$ and $\mathbf{z} = F_2'\mathbf{y}$, the GCV function $V(\lambda, \theta) = \|(I - A(\lambda, \theta))\mathbf{y}\|^2/[\mathrm{tr}(I - A(\lambda, \theta))]^2$ can be written

$$V(\lambda, \theta) = \frac{\mathbf{z}'(\Sigma_\theta + n\lambda I)^{-2}\mathbf{z}}{\left(\mathrm{tr}(\Sigma_\theta + n\lambda I)^{-1}\right)^2}.$$

To minimize $V(\lambda, \theta)$, first put a constraint on $\theta$ so that there are (at most) $q - 1$ independent components, then, for each $\theta$, tridiagonalize $\Sigma_\theta$ by $U'\Sigma_\theta U = \Delta$, where $U$ is orthogonal and $\Delta$ is tridiagonal. A strategy for speeding this step up by a truncated Householder transform is given in Gu, Bates, Chen and Wahba

(1988). For each $\lambda$ do a Cholesky decomposition of $(\Sigma_\theta + n\lambda I)$ as $L'L$, where

$$L = \begin{bmatrix} a_1 & b_1 & & & & \\ & a_2 & b_2 & & & \\ & & \cdot & \vdots & & \\ & & & & \cdot & \\ & & & & a_{n-M-1} & b_{n-M-1} \\ & & & & & a_{n-M} \end{bmatrix}$$

is upper bidiagonal. The calculation of the denominator of the GCV function is based on a trick derived by Eldén (1984); see Gu, Bates, Chen and Wahba (1988). Letting the $j$th row of $L^{-1}$ be $\mathbf{l}_j'$, we have that $\operatorname{tr}(L^{-1}L^{-1'}) = \sum_{j=1}^{n-M}\|\mathbf{l}_j\|^2$, and the following recursion relation can be shown to hold:

$$\|\mathbf{l}_{n-M}\|^2 = a_{n-M}^{-2},$$
$$\|\mathbf{l}_j\|^2 = \left(1 + b_j^2\|\mathbf{l}_{j+1}\|^2\right)a_j^{-2}, \qquad j = n - M - 1,\dots,1,$$

which can be calculated in $O(n - M)$ flops.

The reproducing kernels $Q^k(\mathbf{x};\mathbf{x}')$ for additive and interaction splines are found as follows: Let $W_2^m$ be the Sobolev space

$$W_2^m = \left\{f\colon f, f',\dots, f^{(m-1)} \text{ abs. cont., } f^{(m)} \in L_2[0,1]\right\}$$

with the squared norm

$$\|f\|_{W_2^m}^2 = \sum_{\nu=0}^{m-1}(R_\nu f)^2 + \int_0^1\left(f^{(m)}(x)\right)^2 dx,$$

where

$$R_\nu f = \int_0^1 f^{(\nu)}(x)\,dx, \qquad \nu = 0,1,\dots,m-1.$$

Let $k_l(x) = B_l(x)/l!$, where $B_l$ is the $l$th Bernoulli polynomial, we have $R_\nu B_l = \delta_{\nu-l}$, where $\delta_i = 1$, $i = 0$ and 0 otherwise. With this norm, $W_2^m$ can be decomposed as the direct sum of $m$ orthogonal one-dimensional subspaces $\{k_l\}$, $l = 0,1,\dots,m-1$, where $\{k_l\}$ is the one-dimensional subspace spanned by $k_l$, and $H_*$ which is the subspace (orthogonal to $\Sigma \oplus \{k_l\}$) satisfying $R_\nu f = 0$, $\nu = 0,1,\dots,m-1$, that is,

$$W_2^m = \{k_0\} \oplus \{k_1\} \oplus \cdots \oplus \{k_{m-1}\} \oplus H_*.$$

This construction can be found in, for example, Craven and Wahba (1979). Letting $\otimes^d W_2^m$ be the tensor product of $W_2^m$ with itself $d$ times, we have

$$\bigotimes^d W_2^m = \bigotimes^d \left[\{k_0\} \oplus \cdots \oplus \{k_{m-1}\} \oplus H_*\right]$$

and $\otimes^d W_2^m$ may be decomposed into the direct sum of $(m+1)^d$ fundamental subspaces, each of the form

$$[\ ] \otimes [\ ] \otimes \cdots \otimes [\ ] \quad (d \text{ boxes}) \tag{6}$$

with either $\{k_l\}$ or $H_*$ in each box. The rk for $\{k_l\}$ for the $j$th variable is

$k_l(x_j)k_l(x_j')$ and the rk for $H_*$ is $Q_*(x_j, x_j')$ given by

$$Q_*(x_j, x_j') = k_m(x_j)k_m(x_j') + (-1)^{m-1}k_{2m}([x_j - x_j']),$$

where $[u]$ is the fractional part of $u$ [see Craven and Wahba (1979)]. For additive splines $H$ is the direct sum of all of the fundamental subspaces with at most one entry in the boxes of (6) which is not $\{k_0\}$, $H_0$ is the direct sum of the fundamental subspaces with no $H_*$ entry and each $H^k$ is a fundamental subspace with all entries $\{k_0\}$ except $H_*$ in the $k$th position. The rk's for direct products and sums of orthogonal rk spaces are the corresponding products and sums of the component spaces; see Aronszajn (1950). Recalling that $k_0(\cdot) = 1$, we have the rk $Q_\theta(\mathbf{x}; \mathbf{x}')$ for additive splines is

$$Q_\theta(\mathbf{x}; \mathbf{x}') = \sum_{j=1}^{q} \theta_j Q_*(x_j, x_j')$$

and an element in $H_0$ is of the form $f_0(\mathbf{x}) = d_0 + \sum_{j=1}^{d}\sum_{\nu=1}^{m-1} d_{j,\nu}\phi_\nu(x_j)$. This construction results in the same penalty function for the additive splines as given in the paper (with $\lambda\theta_j^{-1} = \lambda_j$) and hence this approach can be used to fit the additive spline model and simultaneously estimate multiple smoothing parameters. The two-factor interaction subspaces are direct sums of fundamental subspaces with two boxes filled with other than $\{k_0\}$ and so forth; see Gu, Bates, Chen and Wahba (1988) for examples and rk's.

Now as far as comparing the direct approach using matrix decompositions versus the backfitting algorithm (assuming that smoothing parameters are not estimated), the backfitting algorithm could be expected to be faster when the relevant matrices are sparse (as can be arranged for the main effects smoothing spline case). It is possible that Girard's method [Girard (1987)] for evaluating the denominator of the GCV function may prove useful in this case, if $n$ is sufficiently large. This method estimates $\operatorname{tr} A$ by evaluating $\delta' A \delta$, where $\delta$ is a pseudorandom $N(0, I)$ vector. When no special structure is available, then the matrix decomposition approach is probably more appropriate, to the extent that it is feasible. We have done examples with three $\theta$'s and with $n$ as large as $n = 800$ on the Cray and $n = 300$ on a Sun workstation. Transportable code is available [Gu (1988)] by writing gu@stat.wisc.edu. [Added in proof: Further algorithmic and numerical results appear in Gu and Wahba (1988).]

Note that the confidence intervals in Nychka (1988) (which would apply here) use $A = \mathbf{R}$ and not $\mathbf{R}^2$. As the authors note, convergence results are available under various assumptions on $f$ for univariate (as well as other) splines. Best possible convergence rates can be identified with the rate of decay of the eigenvalues of the reproducing kernel(s) [see, for instance, Micchelli and Wahba (1981)], and rates for nicely distributed $\mathbf{x}(i)$ are also known to be related to the eigenvalues of the rk.

Now that we have this rather huge family of smoothers to choose from, by selecting $H_0$ possessing unique regression and $q$ rk's $Q^k$, which only have to be nonnegative definite, what is so special about splines? Stein, in a series of papers [Stein (1987, 1988)] has explored the effects of misspecifying covariance kernels (i.e., rk's), the end result being that asymptotically, one only needs to ensure that one is in the correct equivalence class of covariance kernels, conversely, as far as estimating covariances from data, one can only asymptotically get in the right equivalence class anyway. [See also Section 2 of Wahba (1981).] These results apply to $f$ a stochastic process, if $f$ is considered to be a fixed function in a certain rk space, estimation in certain bigger spaces can still give optimum rates but topologically equivalent norms are going to give the same rates. The theory of equivalence and perpendicularity for zero mean Gaussian processes (i.e., rk's) is fairly complete; see, for example, Rosenblatt (1963) and Hâjek (1962). Anyway, the wonderful thing about spline spaces is that their rk's are the most "parsimonious" members of a big group of equivalence classes.

Now, back to the practical problems of applying these interesting methods. Looking at the scatterplots in Figure 6, one certainly could doubt that the noise is white. Changing variances probably do not hurt methods like cross-validation too much, but correlated errors can be a problem. We do not at the moment know how many smoothing parameters is too many (except we want different parameter values to correspond to different equivalence classes) but we are doing some experimenting.

It is not surprising that there is a relationship between inversion base height and inversion base temperature. If the relationship were exactly linear, there would be exact concurvity in the additive spline model, as formally evidenced by the fact that you could not fit a unique least-squares plane through data on the plane which is restricted to a line. We liked the authors' plots which displayed the effects of this problem. To make the most of additive and other semiparametric models in many variables, we are going to need better tools for diagnosing and dealing with practical (near) concurvity problems, which are undoubtedly going to be encountered frequently. The first step is obviously to check the condition of the $T$ matrix, but, for subtle kinds of concurvity we might need more subtle methods. We trust will see more from the authors on this point in the future.

## REFERENCES

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403.

ELDÉN, L. (1984). A note on the computation of the generalized cross-validation function for ill-conditioned least squares problems. *BIT* **24** 467–472.

GIRARD, D. (1987). A fast "Monte-Carlo cross-validation" procedure for large least squares problems with noisy data. RR 687-M, Informatique et Mathematiques Appliquees de Grenoble, Grenoble.

GU, C. (1988). RKPACK—a general purpose minipackage for spline modeling. Technical Report 832, Dept. Statistics, Univ. Wisconsin, Madison.

GU, C. and WAHBA, G. (1988). Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method. Technical Report 847, Dept. Statistics, Univ. Wisconsin, Madison.

GU, C., BATES, D. M., CHEN, Z. and WAHBA, G. (1988). The computation of GCV functions through Householder tridiagonalization with application to the fitting of interaction spline models. Technical Report 823, Dept. Statistics, Univ. Wisconsin, Madison.

HÁJEK, J. (1962). On linear statistical problems in stochastic processes. *Czech. Math. J.* **87** 404–444.

KIMELDORF, G. and WAHBA, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95.

MICCHELLI, C. and WAHBA, G. (1981). Design problems for optimal surface interpolation. In *Approximation Theory and Applications* (Z. Ziegler, ed.) 329–348. Academic, New York.

NYCHKA, D. (1988). Bayesian confidence intervals for smoothing splines. *J. Amer. Statist. Assoc.* **83** 1134–1143.

ROSENBLATT, M., ed. (1963). *Proc. Symposium on Time Series Analysis Brown Univ.* Wiley, New York.

STEIN, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *J. Amer. Statist. Assoc.* **82** 765–772.

STEIN, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

WAHBA, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *Ann. Statist.* **9** 146–156.

WAHBA, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B* **45** 133–150.

WAHBA, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Statist.* **13** 1378–1402.

WAHBA, G. (1988). Multiple smoothing parameters in semiparametric multivariate model building. In *Computer Science and Statistics: Proceedings of the 20th Symposium on the Interface* (E. Wegman, ed.). Amer. Statist. Assoc., Washington.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
1210 W. DAYTON ST.
MADISON, WISCONSIN 53706

## DENNIS D. COX[1]

*University of Illinois, Urbana-Champaign*

Fitting the additive model using the backfitting algorithm with symmetric smoothers having eigenvalues in [0, 1] amounts to a Bayesian procedure. This statistical interpretation is interesting in its own right, but also suggests other algorithms and provides a framework for solving some of the inferential problems left open by Buja, Hastie and Tibshirani.

The paper by Buja, Hastie and Tibshirani (referred to hereafter as BHT) makes several important contributions. On a trivial note, the discussion of "degrees of freedom" hopefully clarifies the ambiguity of the term when applied to smoothers which are not orthogonal projections. The tantalizing remarks on concurvity may well be the first salvo in a whole barrage of results on such notions. However, the main contribution is the development of the backfitting algorithm. There is an aesthetic elegance in computing estimates for the complex additive model by concatenation of estimates for simpler unidimensional models. From the practical perspective, it provides a method whereby users can "wire together" existing pieces of software to solve a seemingly difficult problem. There are clearly opportunities for many spinoffs, such as implementations on distributed processing systems. Most of the theorems for general $p$ (the dimension

[1]Research supported by National Science Foundation Grant DMS-86-03083.