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.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

of the independent variable X) treat the case of symmetric smoothers S_j with eigenvalues in $[0,1]$. It was not mentioned that the corresponding additive model admits a Bayesian interpretation. I have found this interpretation to be fruitful.

We will utilize the notation of BHT with some extensions given below. Given smoothers S_j which are symmetric and have eigenvalues in $[0,1]$ the prior distribution for $\mathbf{f}_1, \dots, \mathbf{f}_p$ is independent jointly normal with marginals

$$\mathbf{f}_j \sim N_p(0, K_j), \tag{1}$$

where the covariance matrix K_j is given by

$$K_j = \sigma^2 (S_j^- - I)^{-1}, \tag{2}$$

where the distribution is concentrated on the orthogonal complement of the subspace $\mathscr{M}_0(S_j)$. Assuming that the errors are i.i.d. $N(0, \sigma^2)$, it is clear from (19) in BHT that the posterior means for the $\mathbf{f}_j$'s are given by the backfitting algorithm, with the following caveats:

1. It is necessary to restrict the distribution to the orthogonal complement of $\mathscr{M}_0(S_j)$ since if $\mathbf{u} \in \mathscr{M}_0(S_j)$, then $\operatorname{var}[\mathbf{u}^t \mathbf{f}_j] = 0$. Put otherwise, these components are completely eliminated by the smoother S_j as if we have prior knowledge that they are not present.
2. If $\mathscr{M}_1(S_j) \neq 0$ and $\mathbf{u} \in \mathscr{M}_1(S_j)$ with $u \neq 0$, then the prior assigns $\operatorname{var}[\mathbf{u}^t \mathbf{f}_j] = \infty$, that is, we are using the improper Lebesgue prior for such components. Thus, the total prior is "partially improper" as in Wahba (1978) or Ansley and Kohn (1985). Components in $\mathscr{M}_1(S_j)$ are estimated unbiasedly (or equivariantly).
3. It is not really necessary to know the error variance σ^2.

We have only given the prior for the vectors of sampled functions. Typically, one will obtain the covariance matrix for such a random vector by sampling a covariance function for a continuous-time stochastic process. For example, the prior for cubic smoothing splines is given by a scalar multiple of integrated Brownian motion. This continuous-time prior then provides a method for interpolating or extrapolating the fitted values to estimate the function at all values of the independent variable. We will follow BHT in only considering the sampled values.

We now describe an alternative algorithm motivated by the Bayesian model, under the restriction that the eigenvalues of all smoothers are in $(0,1)$, that is, that the priors are proper and nonsingular. Let $K_+ = \sum_{j=1}^p K_j$ be the prior covariance matrix for

$$\mathbf{f}_+ = \sum_{j=1}^p \mathbf{f}_j.$$

Then simple calculations show that the posterior means are given by

$$\hat{\mathbf{f}}_j = K_j (\sigma^2 I + K_+)^{-1} \mathbf{y}.$$

Note that this implies one $n \times n$ solve and then p matrix multiplications, which may be computationally simpler than the backfitting algorithm in this framework. Of course, this is assuming that one has the covariance matrices K_j and it is not necessary to compute them with a matrix inversion as implied by (2) above. If the smoother were given by a kernel estimator, then matrix inversions would be required to compute the covariance K_j, but in the usual Bayesian setting one would begin with the K_j and the smoother would require a solve. It should be possible to generalize this to eigenvalues in $[0,1]$, as in the case of smoothing splines.

A semi-Bayesian approach can provide some shelter from at least one of the difficult inferential problems discussed by BHT, namely checking the adequacy of the additive model. Assume as before proper nonsingular priors for the additive components $\mathbf{f}_j$. We enlarge the model as follows: Let the conditional expectation function be given by

$$f(\mathbf{x}) = \sum_{j=1}^{p} f_j(x_j) + bg(\mathbf{x}),$$

where $g(\mathbf{x})$ represents the interaction term. We assume that $\mathbf{g} \doteq (g(\mathbf{x}_1), \dots, g(\mathbf{x}_n))^t$ is independent of the $\mathbf{f}_j$, Gaussian and mean 0. The scalar multiplier b will be discussed shortly. To avoid confounding various components, we need some orthogonality properties. This is most easily obtained perhaps by direct construction. Let Λ be an "initial" covariance matrix for $\mathbf{g}$ and let

$$L = \left(I - K_+\left(\sigma^2 I + K_+\right)^{-1}\right)\Lambda\left(I - \left(\sigma^2 I + K_+\right)^{-1} K_+\right)$$

be the "final" covariance. This is the covariance matrix of $\gamma - K_+(\sigma^2 I + K_+)^{-1}\gamma$, where γ has a $N_p(0, \Lambda)$ distribution and $K_+(\sigma^2 I + K_+)^{-1}\gamma$ is the vector of fitted values from fitting the Bayesian additive model to the data vector γ. One can then show [Koh (1989)] that the locally most powerful invariant test for $\mathbf{H}_0$: $b = 0$ vs. $\mathbf{H}_1$: $b > 0$ rejects if the test statistic

$$\frac{1}{\sigma^2}\mathbf{y}^t\left(\sigma^2 I + K_+\right)^{-1} L \left(\sigma^2 I + K_+\right)^{-1}\mathbf{y}$$

is too large. See Chen (1986) and Cox, Koh, Wahba and Yandell (1988) for related LMPI tests. Perhaps a more meaningful approach in the context of estimation is to construct an estimator for the interaction (Bayesian or otherwise) and then compute its sum of squares and determine if it is sufficiently small for the application at hand.

REFERENCES

ANSLEY, C. and KOHN, R. (1985). Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *Ann. Statist.* **13** 1286–1316.

CHEN, Z. (1986). A testing procedure for selecting interactions in the purely periodic interaction spline model. Technical Report, Dept. Statistics, Univ. Wisconsin, Madison.

COX, D., KOH, E., WAHBA, G. and YANDELL, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16** 113–119.

KOH, E. (1989). A smoothing spline based test of model adequacy in nonparametric regression. Ph.D. dissertation, Univ. Wisconsin, Madison.

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.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
CHAMPAIGN, ILLINOIS 61820

R. L. EUBANK AND P. SPECKMAN

Texas A & M University and University of Missouri, Columbia

The authors are to be congratulated on this interesting and thought-provoking paper. They have raised a number of important questions and issues concerning additive model methodology. We will discuss some of these below. Throughout, our comments will be restricted to the case of symmetric smoothers having eigenvalues in $[0, 1]$.

1. Exact and approximate concurvity. This paper contains a thorough treatment of the fundamental issues of existence and uniqueness of solutions for the normal equations arising from additive model estimation. The authors show that these equations will have multiple solutions in certain cases. This raises questions as to how analyses should proceed in the presence of exact concurvity. Results from linear models would suggest that if $\mathbf{f}$ represents any solution to the normal equations, then one should only examine functionals $\mathbf{l}^t\mathbf{f}$ of the solution that are "*estimable*" in the sense that $\mathbf{l}^t\mathbf{g} = 0$ whenever $\hat{\mathbf{P}}\mathbf{g} = 0$. Such functionals are invariant under all choices of solutions to the normal equations and will have unique expectations. According to Theorem 5 of the paper, "estimable" functionals are provided by np-vectors in the orthogonal complement of the linear span of vectors $\mathbf{g}^t = (\mathbf{g}_1^t, \dots, \mathbf{g}_p^t)$ with $\mathbf{g}_j \in M_1(S_j)$ and $\mathbf{g}_+ = \mathbf{0}$. In particular we see that $\mathbf{f}_+$ is derived using "estimable" functionals.

Another approach to solving the normal equations for linear models of less than full rank is to reparameterize to obtain a full rank model. This is essentially what the authors have accomplished in Section 4.4 by extracting the projection parts from the smoothers, if linear dependencies are also eliminated from $M_1(S_1) + \cdots + M_1(S_p)$. The $\tilde{\mathbf{f}}_j$ are therefore obtained using "estimable" functionals and perhaps they are what should be studied when there is exact concurvity.

However, it seems to us that instances where an analysis should actually proceed in the presence of exact concurvity without some type of remedial action are rare. For example, in the case of smoothing splines, $M_1(S_j)$ is the linear span of the constant vector and $\mathbf{x}_j$. By Theorem 5 the concurvity space consists only of the constant vector unless the $\mathbf{x}_j$ are linearly dependent. In this latter case at least one of the variables should be dropped from the analysis to obtain meaningful estimates.

The real issue here seems to be approximate concurvity. As before we will draw an analogy with the linear regression case. In that setting approximate