of ill-posed problems, of loss-based methods for choosing smoothing parameters, supplemented by empirical checks that the resulting smoothed estimates are acceptable from a practical point of view. I look forward, in particular, to reading about the future exploits of the present author in this important area!

ADDITIONAL REFERENCES

BESAG, J. (1986). On the statistical analysis of dirty pictures (with discussion). To appear in *J. Roy. Statist. Soc. Ser. B.*

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.* **PAMI-6** 721–741.

HALL, P. and TITTERINGTON, D. M. (1986a). On some smoothing techniques used in image processing. To appear in *J. Roy. Statist. Soc. Ser. B.*

HALL, P. and TITTERINGTON, D. M. (1986b). Common structure of techniques for choosing smoothing parameters in regression problems. Revised manuscript in preparation.

SILVERMAN, B. W. (1984). A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *J. Amer. Statist. Assoc.* **79** 584–589.

TITTERINGTON, D. M. (1984). The maximum entropy method for data analysis (with reply). *Nature* **312** 381–382.

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

Comment

Grace Wahba

Professor O'Sullivan has given us a nice overview of some of the issues in ill-posed inverse problems as well as some new ideas. The most important of these new ideas I believe are the following: a) the extension of the idea of averaging kernel to reproducing kernel spaces, with the resulting formula

$$\operatorname*{supremum}_{\|\theta\|^2 \le \mu^2} |\theta(t) - E\hat{\theta}(t)|^2 = \|e_t - A(t)\|^2 \mu^2$$

and b) a new approach to the history matching problem of reservoir engineering. The formula bears a not coincidental relationship to Scheffé's S method of multiple comparisons (Scheffé, 1959, page 65). In atmospheric sciences and possibly elsewhere, extensive historical data allows the construction of a prior covariance for the unknown θ, from which reasonable norms can often be constructed via the well known duality between prior covariances and optimization problems in reproducing kernel spaces. An example of the use of prior covariances based on historical meteorological data to establish penalty functions can be found in Wahba (1982a). The problems of reservoir engineering are extremely important and would benefit from the attention of statisticians. Letting

$$z_{ij} = u(x_i, t_j, a) + \varepsilon_i,$$

as in Section 4.2, the method of regularization estimate of a is the minimizer of

$$\frac{1}{n} \sum_{ij} (z_{ij} - u(x_i, t_j, a))^2 + \lambda J(a)$$

(see especially Kravaris and Seinfeld, 1985). This problem is particularly difficult since, not only is u a nonlinear function of a, but in general the relationship is only known implicitly as the solution to a partial differential equation. It is a good conjecture that the GCV for nonlinear problems as proposed in O'Sullivan and Wahba (1985) can be used to choose λ in this problem. The details are far from obvious but it looks like the present paper provides an important first step. Of course this history matching setup leads to some juicy experimental design problems—choice of the forcing function q, the location of the wells, and the times of observation.

Concerning robustness of the PMSE criteria (that is, minimizing PMSE also tends to minimize other, possibly more interesting loss functions), further remarks on that can be found in Wahba (1985, page 1381). The GCV extension proposed by the author is an interesting one. Let C be the matrix with ijth entry $c_i' c_j$. If C is the identity then the extension is the same as GCV. If C is a well conditioned matrix, then it appears that one can show that the minimizer of $EV(\lambda)$ is asymptotically near the minimizer of $EL(\lambda)$, the associated (estimable) loss function. You need $(1/T)\text{tr } HC$ to be small near the minimizer of EL. I think a problem may arise if you try to choose C to approximate $L(\lambda)$ of the form

$$L(\lambda) = \frac{1}{T} \sum_{i=1}^{T} |\theta(t_i) - \hat{\theta}_\lambda(t_i)|^2$$

where the problem is very ill-posed. Consider the operator which maps θ to euclidean m space via the formula $\theta - (\eta(x_1, \theta), \ldots, \eta(x_m, \theta))$. In practice the theoretical dimension of the range space of this

Grace Wahba is Professor, Department of Statistics, 1210 West Dayton Street, Madison, Wisconsin 53706-1693.

operator can be m, whereas the "computer dimension" is much less than m. In that case $L(\lambda)$ may be "theoretically" estimable but practically C behaves like the inverse of a matrix with computer rank much less than m. For an example of a convolution operator with a range space with effective dimension very much smaller than m see Wahba (1982b).

There are many interesting open questions remaining in connection with ill-posed inverse problems and I trust Professor O'Sullivan's paper will generate more interest in them among statisticians. Some of the open questions are really at the intersection of statistics and numerical analysis, in particular, those involving extremely large data sets such as occur in x-ray and satellite tomography (the three-dimensional recovery of the atmospheric temperature distribution) and as occur in nonlinear problems and implicit problems like the history matching problem. One needs good approximation theoretic methods to solve extremely large, sometimes nonquadratic optimization problems, and, in the case of the history matching problem, partial differential equations. One would like the approximations to simultaneously be the right sort of "low pass" filters. For example, instead of solving a variational problem exactly in some function space, one solves it in a carefully chosen finite-dimensional subspace. This lowers the complexity of the numerical problem, while at the same time, if the subspace is chosen appropriately, performs further low pass filtering. One would like to choose the approximation theoretic methods so that they simultaneously give a desirable result from a statistical and a numerical analytic point of view.

ADDITIONAL REFERENCES

SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.

WAHBA, G. (1982a). Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data. In *Multivariate Approximation Theory II.* (W. Schempp and K. Zeller, eds.) **2** 407–429. Birkhäuser, Basel.

WAHBA, G. (1982b). Constrained regularization for ill-posed linear operator equations, with applications in meteorology and medicine. In *Statistical Decision Theory and Related Topics III.* (S. S. Gupta and J. O. Berger, eds.) **2** 383–418. Academic, New York.

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.

Comment

John A. Rice

Finbarr O'Sullivan has presented us with a very nice survey and discussion of topics in ill-posed inverse problems. There are many practical problems of this kind in which one is given noisy direct or indirect measurements of an object which one then wishes to reconstruct. The object is often inherently infinite-dimensional whereas there are only a finite number of measurements. In this context one is forced into the healthy exercise of directly confronting problems of bias, which have typically been swept under parametric rugs by professional statisticians. Backus-Gilbert kernels provide a simple and easily interpretable means of qualitatively assessing bias. It is interesting that texts on linear statistical models rarely show figures which give the kernels for linear and quadratic regression. These kernels are, of course, just the rows (or columns) of the matrix $X(X^TX)^{-1}X^T$, if the x_i are listed in increasing order. It is amusing to plot the kernels for higher order polynomials as well.

Examination of the functionals O'Sullivan denotes by η is often very instructive. As in O'Sullivan's first example, the data often consist of the result of a linear operator applied to the object of interest plus noise. By carrying out a singular value decomposition of the operator and plotting the singular values and vectors, one can often see what information is being inherently degraded by the data collection process, that is, which features of the solution can be resolved well and which cannot.

The applications of regularization presented in this paper make use of a prior assumption of smoothness of the solution. Other sorts of prior information can be useful as well. An assumption of positivity, or monotonicity, can be very effective in eliminating highly oscillatory solutions (cf. Wahba, 1982). One of the reasons for the highly advertised effectiveness of maximum entropy solutions is that they are forced to be positive. In various forms of spectroscopy, the solutions are known a priori to be composed of very

John A. Rice is Professor in the Department of Mathematics, University of California, San Diego, La Jolla, California 92093.