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.

flat regions with occasional spikes. It might be useful to develop smoothing functionals J that mirror this.

Regarding the empirical selection of smoothing parameters, Rice (1986) sounds a cautionary note by constructing simple examples in which a choice of smoothing parameter giving a good value of predictive mean square error gives unacceptable errors for estimating θ and vice versa.

ADDITIONAL REFERENCES

RICE, J. (1986). Choice of smoothing parameter in deconvolution problems. *Contemp. Math.* **59** 137–151.

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

Comment

Freeman Gilbert

In a typical geophysical inverse problem one has

$$d_j = D_j(f) + r_j\sigma_j, \qquad j \in \{1, \ldots, J\}, \tag{1}$$

where

- d is a datum,
- D is the functional that maps f into d,
- f is the model,
- r is a unit variance random variable,
- σ is the assigned error, usually taken to be the standard deviation (Gaussian errors).

An error statistic is introduced, usually the χ^2 statistic

$$\chi^2(f) = \sum_j [d_j - D_j(f)]^2/\sigma_j. \tag{2}$$

One defines the set

$$\{F_0(f)\text{: all } f \text{ such that } \chi^2(f) \le X_0^2\},$$

where χ_0^2 is chosen to be the 99% or 95% confidence level, for example.

Except in very unusual circumstances, $F_0(f)$ is either empty or infinite-dimensional. In the former case, one increases χ_0^2 and seeks to fill $F_0(f)$. In the latter case, one desires to know about the members of $F_0(f)$.

One procedure is to use the method of regularization (MOR) to find a particular member of $F_0(f)$ (e.g., the smallest, the smoothest, the one closest to a particular f_0, the maximum entropy solution, $\max\{-f \log f\}$, etc.). Another procedure is to use a resolution method to find what features all f have in common or what are the resolvable averages of f. In any case one may wish to assert a priori conditions on f, such as prejudices about the shape or size of f that can be cast in the form of equation (1).

O'Sullivan has shown that the two procedures are connected and, taken together, can lead to improved methods of estimating bias. By generalizing the concept of averaging kernel, i.e., requiring the averaging kernel to assume certain shapes, one can estimate average bias as well as local bias. For linear problems, the matter appears to be resolved and depends only on the number and quality of the data and the span of their representers. For nonlinear problems one is confined to the neighborhood of the subject. O'Sullivan is to be congratulated for his original contribution to it.

Freeman Gilbert is Professor of Geophysics in the Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093.

Rejoinder

Finbarr O'Sullivan

It is a pleasure to thank the discussants Professors Gilbert, Rice, Titterington, and Wahba for their most interesting and stimulating comments. The ubiquity of inverse problems in areas like geophysics, medical imaging, and meteorology presents statisticians with wonderful opportunities to contribute to the development of science and technology. As Professor Wahba notes there are lots of open research questions many