Comment

D. M. Titterington

1. INTRODUCTION

It is a pleasure to comment on the paper by Dr. O'Sullivan. The paper represents admirable blends of review and new ideas, together with theory and application.

It seems that the world is saturated with inverse problems. At least, I am continually being surprised to discover further manifestations of the general structure and, sometimes, substantively innovative developments. I was particularly grateful to discover the work of Backus and Gilbert and to learn about the notion and use of representers.

The paper, of course, discusses a particular class of inverse problems, those which are ill-posed. Perhaps the most surprising feature of the literature on this topic is the comparatively late stage at which statisticians have made an impact. After all, a major reason for the inherent difficulties is the existence of the random noise terms, ε_t, in the model, and we note that the ubiquitous prescription for estimation is of the ridge-regression type, so it is certainly appropriate territory for statisticians. I should like to base the bulk of my remarks on the theme of what particular contributions statisticians can make to the development of the area.

Before I launch into this, I should admit that, as the paper points out, other mathematical specialties are also essential to a full treatment of the problem. Particular areas are those of functional analysis, matrix theory (singular-value decomposition), and numerical methods for optimization. So far as the last topic is concerned, the paper has concentrated on *linear* or *linearized* problems, so that the optimality criterion is a quadratic function for which the minimizer can be written down explicitly. In other cases we are left with a nonquadratic criterion, which leads to the requirement of numerical methods; see, for instance, the use of simulated annealing in finding a regularized image restoration by Geman and Geman (1984).

2. THE IMPACT OF STATISTICIANS

In this section I shall follow the pattern of the paper in concentrating on linear problems. As a result, and with some apologies to the author for apparently trivializing his achievements, the problems are "solved" by ridge-regression estimators of which (3.1) is an example. Crucial features of the prescription are a matrix Ω_2 and a scalar, λ.

Depending on one's statistical leanings, (λ, Ω_2) have different interpretations. For a Bayesian, they are hyperparameters in a notional prior density and the ridge-regression estimator is itself interpretable as a posterior mode. This Bayesian basis has the advantage, in principle, of permitting the construction of confidence regions for the true quantities of interest. Of course, the validity of such regions is dependent on whether the notional prior is a meaningful one. So far as repeated sampling confidence statements are concerned, more work requires to be done on the lines of Wahba (1983b) to see to what extent Bayesian statements carry similar confidence values from a frequentist point of view.

Non-Bayesians interpret λ and Ω_2 somewhat differently. They regard Ω_2 as the kernel of a roughness penalty function, usually chosen to reflect some (admittedly "prior") ideas about the local smoothness of the underlying functions and/or to lead to tractable prescriptions for the regularized estimators, in the form of splines, for instance. If one can extrapolate from the literature about kernel-based density estimation, the choice of Ω_2 (cf., the choice of kernel function) should not be crucial to the performance of the resulting estimator, computational difficulties apart. Certainly, from the non-Bayesian point of view, no one Ω_2 seems sacrosanct. This last statement appears to conflict with the views of the adherents of maximum entropy regularization, who contend that, in a wide range of problems, a roughness penalty based on Shannon entropy is fundamentally special, an opinion I do not share (Titterington, 1984).

The other parameter, λ, called variously the smoothing, ridge, or regularization parameter, is the one to which the estimators should be more sensitive. Furthermore it is here that the statistical impact is most obvious. In principle there is no problem to the Bayesian, in that λ is a parameter of the prior which is, of course, known! To other statisticians, it is natural to base the choice of λ on some criterion of how close the estimator is to the true, on average. As a result, we obtain the mean squared error criteria of Section 5 and the associated databased versions such as cross-validatory choice, now familiar in several types of smoothing problems. It has required

---

D. M. Titterington is Titular Professor and Head of the Statistics Department, University of Glasgow, Glasgow G12 8QW, Scotland.