DEPARTMENT OF STATISTICS

University of Wisconsin
1210 W. Dayton Street
Madison, WI 53706

TECHNICAL REPORT NO. 732

November 1983

DESIGN CRITERIA AND EIGENSEQUENCE PLOTS FOR SATELLITE COMPUTED TOMOGRAPHY

Grace Wahba

Statistics Department
University of Wisconsin-Madison

This research supported by the National Aeronautics and Space Administration under contract NAG-128.

DESIGN CRITERIA AND EIGENSEQUENCE PLOTS FOR SATELLITE COMPUTED TOMOGRAPHY

Grace Wahba
Statistics Department
University of Wisconsin-Madison

ABSTRACT

The use of the "degrees of freedom for signal" is proposed as a design criteria for comparing different designs for satellite and other measuring systems. It is also proposed that certain eigensequence plots be examined at the design stage along with appropriate estimates of the parameter λ playing the role of noise to signal ratio. The degrees of freedom for signal and the eigensequence plots may be determined using prior information in the spectral domain which is presently available along with a description of the system, and simulated data for estimating λ. This work extends the 1972 work of Weinreb and Crosby.

1. INTRODUCTION

Recently Fleming (1983) has suggested that improved temperature retrievals from satellite soundings may be obtained by use of data from a sensor which scans forward and back along the satellite track, and thus "looks" at a particular point in space from several directions as well as directly down. This idea was suggested by analogy with well known results from computed tomography techniques in use in medicine. Fleming constructed a model temperature field and simulated noisy data from three different ray configurations, one looking straight down only, one having in addition one forward and one rearward angle, and the third having two forward and two rearward angles. See Fig. 1. He then recovered the model temperatures on a two dimensional grid with one axis vertical and one axis along the satellite track, by a numerically efficient iterative procedure for solving large linear systems. He performed the necessary regularization in this ill posed problem by stopping the iteration. See also Fleming (1977), Wahba (1980). Similar methods are common in medical applications. Fleming's results in the example tried were: two additional angles are better than straight down only, and four are better than two, from the point of view of mean square error.

We are interested in the problem of choice of angles, spacing of observations, selection of channels and other questions concerning the design of measuring systems. Weinreb and Crosby (1972) discussed design criteria which can be used to make an evaluation of alternative satellite designs and they applied these criteria to the selection of radiometer channels. In this paper, we begin with what is essentially the design criteria proposed by Weinreb and Crosby. However, we propose using prior information concerning meteorological fields in the frequency or spectral domain, rather than the spatial domain, leading to details which can be different. This approach uses information which is available at the present time (but not in 1972!) and is particularly appropriate for the evaluation and comparison of potential satellite systems that simultaneously use three dimensional information, as well as the evaluation of systems which use combined satellite and radiosonde data. Implicit in the procedures described here is an algorithm for combining satellite and radiosonde data. Our approach also makes clear the role of possibly variable bandwidth parameter(s) in system design, a point which has traditionally been ignored. In Section 2 we derive the design criteria in our form (as opposed to the form used by Weinreb and Crosby) and also note how data from different systems can be combined. In Section 3 we describe the idea of the "effective rank" of a system, which is roughly equivalent to the "degrees of freedom for signal" associated with a design. The "degrees of freedom for signal" is related to but not exactly the same as one of the criteria used by Weinreb and Crosby, and is analogous to the usual degrees of freedom for signal in analysis of variance. We suggest the use of eigensequence plots along with the GCV (generalized cross validation) estimate of the bandwidth (or signal to noise ratio) parameter on these plots, to evaluate and compare different systems, from the point of view of degrees of freedom for signal.

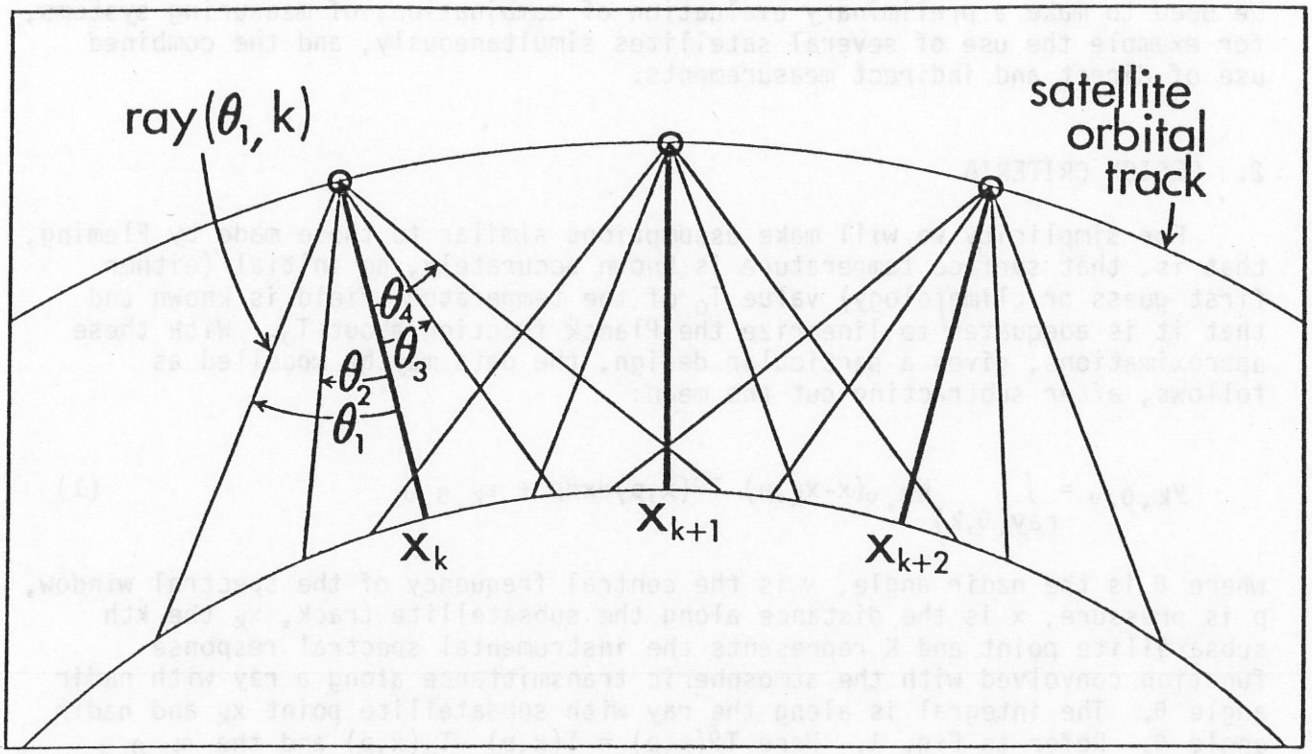

Figure 1. Satellite Viewing Geometry

The details of the approach descibed here are perfectly general and can be used to make a preliminary evaluation of combinations of measuring systems, for example the use of several satellites simultaneously, and the combined use of direct and indirect measurements.

2. DESIGN CRITERIA

For simplicity we will make assumptions similar to those made by Fleming, that is, that surface temperature is known accurately, an initial (either first guess or climatology) value T_0 of the temperature field is known and that it is adequate[1] to linearize the Planck function about T_0. With these approximations, given a particular design, the data may be modelled as follows, after subtracting out the mean:

$$y_{k,\theta,\nu} = \int_{ray(\theta,k)} K_{\theta,\nu}(x-x_k,p)\, T^{\delta}(x,p)\,dx\,dp + \varepsilon_{k,\theta,\nu} \tag{1}$$

where θ is the nadir angle, ν is the central frequency of the spectral window, p is pressure, x is the distance along the subsatellite track, x_k the kth subsatellite point and K represents the instrumental spectral response function convolved with the atmospheric transmittance along a ray with nadir angle θ. The integral is along the ray with subsatellite point x_k and nadir angle θ. Refer to Fig. 1. Here $T^{\delta}(x,p) = T(x,p) - T_0(x,p)$ and the $\varepsilon_{k,\theta,\nu}$ represent measurement, quadrature, and modelling errors. See, e.g. Wark and Fleming (1966), Fritz et al. (1972). We shall assume that the observations have been normalized so that $E\varepsilon^2_{k,\theta,\nu}$ is roughly constant and the $\varepsilon_{k,\theta,\nu}$ are roughly independent.

Next, we shall assume that T^{δ} possess a (generalized) Fourier series expansion in some appropriate basis functions in x and p, for example:

$$T^{\delta}(x,p) = \sum_{\alpha,\gamma} T_{\alpha,\gamma}\, \psi_{\alpha}(x)\, \phi_{\gamma}(p). \tag{2}$$

If the temperature is going to be retrieved around a circle, it may be appropriate to let the ψ_{α} be sines and cosines, the ϕ_{γ} are appropriate (continuous) orthogonal functions in the vertical. If one was carrying out this study on the globe, spherical harmonics might be appropriate. In general, the $\{\psi_{\alpha}(x)\,\phi_{\gamma}(p)\}$ are most conveniently taken to be orthonormal over an appropriate region. In other contexts Hough functions might be used. See Wahba (1982a).

---

[1] For a more careful approach to the nonlinearity, the linearization in O'Sullivan (1983) p. 78 may be used.

The observations are now modelled as:

$$y_{k,\theta,\nu} = \sum_{\alpha,\gamma} T_{\alpha,\gamma} \int_{ray(\theta,k)} K_{\theta,\nu}(x-x_k,p)\, T^{\delta}(x,p) + \epsilon_{k,\theta,\nu} \quad (3)$$

which we can rewrite as

$$y = X\beta + \epsilon \quad (4)$$

where y is the (rearranged) vector of the observations $y_{k,\theta,\nu}$, β is the (rearranged) vector of the $T_{\alpha,\gamma}$'s and ϵ is the rearranged vector of the $\epsilon_{k,\theta,\nu}$. Letting i stand for k,θ,ν and j stand for α,γ we have that the i,jth entry of X is

$$x_{ij} = \int_{ray(\theta,k)} K_{\theta,\nu}(x-x_k,p)\,\psi_\alpha(x)\,\phi_\gamma(p)\,dx\,dp. \quad (5)$$

Letting $\beta_j = T_{\alpha,\gamma}$, if T_o has been obtained from climatology and the ψ's and ϕ's have been chosen appropriately, a fair amount of information may be constructed or assumed concerning the prior distribution of the β_j's. See, for example, Baer (1980), Stanford (1979), Kasahara and Puri (1981), Smith and Woolf (1976). An illustration of the explicit use of Stanford's results in this context can be found in Wahba (1982b). We shall suppose that the β_j's have a prior mean of zero, and a prior covariance matrix given by

$$E\beta_j\,\beta_k = b\sigma_{jk}, \quad \Sigma = \{\sigma_{jk}\} \quad .$$

In the sequel we will be assuming that σ_{jk} is known, but the scale factor b may not be. We suppose that the errors can be modelled (approximately) as independent Gaussian random variables with a common (possibly unknown) variance σ^2. Then a regularized estimate of β is β_λ given by the minimizer of

$$\frac{1}{n}||y - X\beta||^2 + \lambda\beta'\,\Sigma^{-1}\beta. \quad (6)$$

The minimizer, β_λ is given by

$$\beta_\lambda = \Sigma X'(X\Sigma X' + n\lambda I)^{-1}y \quad (7)$$

and the temperature estimate is $T_0(x,p) + T_\lambda{}^\delta(x,p)$ where

$$T_\lambda{}^\delta(x,p) = \sum_{\alpha,\gamma} T_{\lambda,\alpha,\gamma}\, \psi_\alpha(x)\, \phi_\gamma(p)$$

and the $T_{\lambda,\alpha,\gamma}$ are the components of β_λ. This can also be shown to be the Bayes estimate (Gandin estimate) of β with the choice $\lambda = \sigma^2/nb$. That is, β_λ is the conditional expectation of β given the data. This result is found in a more general setting in Kimeldorf and Wahba (1971), see also Wahba (1978a). In practice the estimate can be extremely sensitive to the choice of λ and not "robust" to misspecification of σ^2/nb or other modelling assumptions[2] and so λ should be chosen either from experience ("by eyeball") or by a good data based method such as generalized cross validation (GCV) (see e.g. Craven and Wahba (1979), Golub, Heath and Wahba (1979) Halem and Kalnay (1983), Wahba and Wendelberger (1980). We will, for the moment, however, leave λ as a parameter. Now, suppose our criteria for preferring one design over another is to minimize the expected integrated mean square error, (IMSE) where

$$\text{IMSE} = \int_{\substack{\text{area of}\\ \text{interest}}} (T_\lambda{}^\delta(x,p) - T^\delta(x,p))^2 dxdp. \tag{8}$$

Expanding (8) in the $\{\psi_\alpha \phi_\gamma\}$ gives

$$\text{IMSE} = \sum (\beta_j - \beta_{\lambda,j}) q_{jk} (\beta_k - \beta_{\lambda,k}) \tag{9}$$

where, if $j = (\alpha,\gamma)$ and $k = (\alpha',\gamma')$, then

$$q_{jk} = \int_{\substack{\text{area of}\\ \text{interest}}} \psi_\alpha(x)\, \phi_\gamma(p)\, \psi_{\alpha'}(x)\, \phi_{\gamma'}(p) dxdp, \quad j = (\alpha,\gamma),\ k = (\alpha',\gamma')$$

We now take the expected value of (9), over both the distribution of the β_j and the ε_j. Substitution of

$$\beta_\lambda = \Sigma X'(X\Sigma X' + n\lambda I)^{-1}(X\beta+\varepsilon)$$

into (9) gives

---

[2] See Appendix B

$$E\ \mathrm{IMSE}(\lambda) = E\{\beta' M_1{}' Q M_1 \beta + 2\beta' M_1{}' Q M_2 \varepsilon + \varepsilon' M_2{}' Q M_2 \varepsilon\} \quad (10)$$

where Q is the matrix with jkth entry q_{jk} and

$$M_1 = I - \Sigma X'(X\Sigma X' + n\lambda I)^{-1} X$$

$$M_2 = \Sigma X'(X\Sigma X' + n\lambda I)^{-1}.$$

Carrying out the expectation operation in (10), after assuming that $E\varepsilon_i \beta_j = 0$ gives

$$E\ \mathrm{IMSE}(X,\lambda) = \mathrm{Trace}\ \{bM_1\ QM_1\Sigma + \sigma^2 M_2\ QM_2\} \quad (11)$$

Letting $Q^{1/2}$ and $\Sigma^{1/2}$ be the symmetric square roots of Q and Σ, it is shown in Appendix A that rearranging (11) results in

$$\frac{1}{b}\ \mathrm{IMSE}(X,\lambda) = \mathrm{Trace}\ Q^{1/2}\{\Sigma - \Sigma X'(X\Sigma X' + n\lambda I)^{-1} X\Sigma\}Q^{1/2}$$

$$+ \left(\frac{\delta^2}{b} - n\lambda\right)\mathrm{Trace}\ Q^{1/2}\{\Sigma X'(X\Sigma X' + n\lambda I)^{-2} X\Sigma\}Q^{1/2} \quad (12)$$

It can be shown that the right hand side of (12) is minimized over λ for $n\lambda = \sigma^2/b$. Making this choice for λ gives

$$\frac{1}{b}\mathrm{IMSE}(X) = \mathrm{Trace}\ Q^{1/2}\ \{\Sigma - \Sigma X'(X\Sigma X' + \frac{\sigma^2}{b} I)^{-1} X\Sigma\}\ Q^{1/2} \quad (13)$$

Typically it will be possible to choose the $\{\psi_\alpha \phi_\gamma\}$ so that Σ is diagonal (Usually, information about cross covariances is not readily available anyway.) If the area of interest and the area over which the $\{\psi_\alpha \phi_\gamma\}$ are orthonormal coincide, then Q will be diagonal, thus making (13) more transparent. In any case, we want to choose X so that the right hand side of (13) is as small as possible. We have the following

Theorem: Let X_1 and X_2 be two design matrices of the same dimension and suppose that $\beta' X_1{}' X_1 \beta \geq \beta' X_2{}' X_2 \beta$ for all β. (That is, $X_1{}' X_1 - X_2{}' X_2$ is non negative definite).

Then

$$\mathrm{IMSE}(X_1) \leq \mathrm{IMSE}(X_2) \tag{14}$$

for any non negative definite Q and $\frac{\sigma^2}{b} > 0$.

Proof: See Appendix A

Unfortunately this provides only a partial ordering. We would like to find a more graphic way of evaluating a design, or comparing two designs, independent of Q. We will do this in the next Section.

We remark that if radiosonde information is to be combined with satellite information, then one just increases the dimension of the data vector y in (4). If y_k is a direct measurement of temperature at a point (x_k, p_k) then this just adds a row to the X matrix with entries $x_{kj} = \psi_\alpha(x_k)\phi_\gamma(p_k)$. If different measuring systems are being combined it is appropriate to scale the observations in units chosen so that the ε_i are about the same size.

3. EIGENSEQUENCE PLOTS, EFFECTIVE RANK, AND DEGREES OF FREEDOM FOR SIGNAL.

Letting the dimension of X be nxp, we have not discussed the relative size of n and p. In meteorological work it is frequently reasonable that $p > n$, since meteorological fields contain information at all scales. Certainly in the design phase one should allow p to be as large as computationally feasible consistent with the availability of (measured, theoretical, or conjectured) prior variances. One does not expect to get very good estimates of individual β_j with $p > n$, however, it is $T^\delta(x,p)$ that is actually desired and good estimates of $T^\delta(x,p)$ may be obtainable even though some of the individual coefficient estimates appear poor. Inspection of (7) shows that the number of linearly independent pieces of information in y available for estimating β (and hence T^δ) is limited by the number of eigenvalues of $X\Sigma X'$ which are at least not negligible compared to $n\lambda$[3]. The "signal" along an eigenvector with eigenvalue much less than $n\lambda$ will be down in the "noise". Proceeding under the assumption that $n \leq p$, it is typical nevertheless, in ill posed problems, that the "effective rank" of matrices playing the role of $X\Sigma X'$ is much less than n, when n is large. The "effective rank" of $X\Sigma X'$ can be roughly defined as the number of eigenvalues of $X\Sigma X'$ not small compared to the noise (relative to b) in the system. (See Wahba (1980)). This "noise" in practice includes not only the measurement error, but the errors in modelling the atmospheric transmittance functions, in linearizing Planck's function, and in computing the integrals in (5), using quadrature formulae. The effective rank of $X\Sigma X'$ can easily be studied by plotting the eigenvalues of $X\Sigma X'$ on a log-log plot.

---

[3] Where $n\lambda$ is appropriately chosen, see appendix B.

Figure 2 gives an eigensequence plot of the eigenvalues reprinted from Nychka, Wahba, Goldfarb and Pugh (1983) (NWGP). The problem in NWPG is a mildly ill posed problem concerned with the recovery of three dimensional tumor size distributions from tumor radii observed from two dimensional slices. This is a tomographic problem of a somewhat different form than the one under study. Nevertheless, there are some common problems. There were n = 80 observations, 68 of the 80 eigenvalues appear on this plot. The precipitous drop off of the last few eigenvalues has been attributed to artifacts of the quadrature procedure. Data from an active experiment using the design behind this plot was actually analyzed and $n\hat{\lambda}$ estimated by GCV appears on the figure. In practice $n\hat{\lambda}$ would appear instead of $n\lambda$ in (7), where, in the design phase, $\hat{\lambda}$ would be obtained by simulating realistic examples. One can see that there are only 6 eigenvalues at least as large as $n\hat{\lambda}$. Strictly speaking, comparing the eigensequence plots for $X_1 \Sigma X_1'$ and $X_2 \Sigma X_2'$ does not necessarily provide enough information for choosing between X_1 and X_2 on the basis of criteria (13), nevertheless, these plots can be quite informative.

A measure of comparison between X_1 and X_2 which depends only on the respective eigenvalues and λ is the "degrees of freedom for signal." We may define d.f. signal (X,λ^*) as

$$\text{d.f. signal } (X,\lambda^*) = \text{trace}(X\Sigma X')(X\Sigma X' + n\lambda^* I)^{-1}\text{[4]}$$

$$= \sum_{\nu=1}^{n} \frac{\lambda_\nu}{\lambda_\nu + n\lambda^*}$$

where λ_ν, $\nu=1,2,\ldots n$ are the eigenvalues of $X\Sigma X'$, and λ^* is a good choice of λ. To understand this definition, which is analogous to similar definitions in analysis of variance, observe that y can be decomposed into signal and noise as follows

$$y = y_{\lambda^*} + \varepsilon_{\lambda^*}$$

---

[4] Weinreb and Crosby's trace M, of their eqn. (10) would correspond to trace $X\Sigma^2 X'(X\Sigma X' + n\lambda^* I)^{-1}$. The present criteria is likely to be less sensitive to misspecification of Σ.

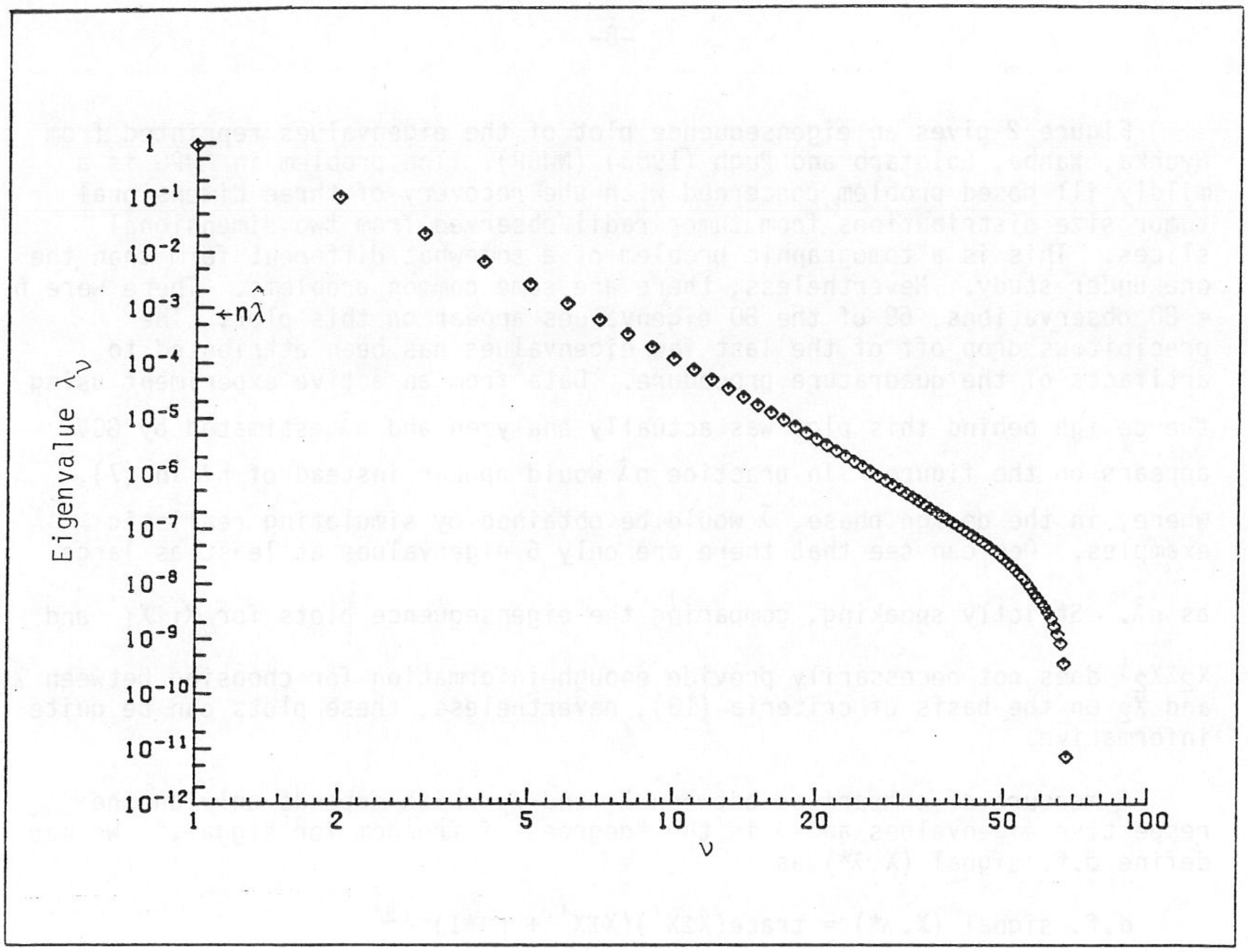

Figure 2. Eigensequence plot from NWGP.

where

$$y_\lambda{}^* = X\beta_\lambda{}^* = X\Sigma X'(X\Sigma X' + n\lambda^* I)^{-1} y \qquad \text{(estimated signal)}$$

$$\varepsilon_\lambda{}^* = n\lambda(X\Sigma X' + n\lambda^* I)^{-1} y, \qquad \text{(estimated noise)}$$

and

$$n = \text{trace } I = \text{trace } X\Sigma X'(X\Sigma X' + n\lambda^* I)^{-1} \qquad \text{(d.f. for signal)}$$

$$+ \text{ trace } n\lambda^*(X\Sigma X' + n\lambda^* I)^{-1} \qquad \text{(d.f. for noise).}$$

It is necessary, of course, that the λ^* used provides a good partition of y into signal and noise for this definition to be valid. It is clear that one wants as many eigenvalues as possible to be large compared to $n\lambda^*$. One can make a loose association of the d.f. for signal with the "effective rank." Thus, X_1 is to be preferred to X_2 if

$$\text{d.f. signal } (X_1, \lambda_1{}^*) > \text{d.f. signal } (X_2, \lambda_2{}^*). \qquad (15)$$

We have deliberately allowed $\lambda_1{}^*$ and $\lambda_2{}^*$ to be different, and not necessarily equal to σ^2/b, since in practice, as well as in Monte Carlo experiments with a small number of examples, the optimum λ may depend on X as well as T^δ and the noise in the system.

The GCV estimate $\hat{\lambda}$ of λ is the minimizer of

$$V(\lambda) = \frac{\frac{1}{n}\|y - X\Sigma X'(X\Sigma X' + n\lambda I)^{-1} y\|^2}{\left[\frac{1}{n}\,\text{Trace}(I - X\Sigma X'(X\Sigma X' + n\lambda I)^{-1})\right]^2}$$

and can be obtained as part of a realistic Monte Carlo study. Of course, it is quite possible that the eigensequence plots will show that the choice between X_1 and X_2 on the basis of d.f. signal is insensitive to the choice of λ.

Eigenvalues of symmetric nonnegative definite matrices of dimension up to several hundred can be computed using double precision EISPACK (Smith et al. (1976)). If Σ is diagonal, it may be cheaper and more accurate to compute the singular values of $\Sigma^{1/2}X'$ using the singular value decomposition in LINPACK (Dongarra et al. (1979)). The eigenvalues of $X\Sigma X'$ are the squares of the singular values of $\Sigma^{1/2}X'$. Approximate information concerning very much larger matrices may be obtained using the truncated singular value decomposition in Bates and Wahba (1982). It is conjectured that eigensequence plots comparing different satellite scanning designs will show that, e.g. combining side looking scans from successive passes of a satellite along with data in the plane of the orbit (as suggested by Suomi (1983)) would have highly desirable properties.

We close with a few remarks. Quadrature error, in e.g. evaluating the x_{ij} in (5) can be surprisingly important in ill posed problems and should not be treated cavalierly, either at the design stage or at the data analysis stage. This point is discussed in some detail in NWGP, where the use of matched quadrature for ill posed problems is discussed. Eigensequence plots obtained via inaccurate quadrature may present a different appearance than those from a highly accurate quadrature, and a poor quadrature procedure or unrealistic value of λ may mask differences between systems.

ACKNOWLEDGEMENT

The author would like to acknowledge numerous helpful discussions with D. R. Johnson, and the assistance of Nira Dyn in the proof of the theorem. I thank Bill Raynor for bringing Weinreb and Crosby's work to my attention.

REFERENCES

Baer, F. (1980): Three dimensional scaling and structure of atmospheric energetics. J. Atmos. Sci., 38, 52-68.

Bates, D. M. and Wahba, G. (1982): Computational methods for generalized cross-validation with large data sets. pp. 283-296. In Treatment of Integral Equations by Numrical Methods, C. T. H. Baker and G. F. Miller (ed.), London: Academic Press.

Craven, P. and Wahba, G. (1979): Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. Numer. Math., 31, 377-403.

Crosby, D. (1983): Applications of a linear model to the retrieval of atmospheric temperatures from satellite measurements. Talk given at the American Statistical Association annual meeting, Toronto, Aug. 17, also, manuscript.

Dongarra, J., Bunch, J., Moler, C., and Stewart, G. (1979): Linpack User's Guide. Philadelphia, PA: SIAM.

Fleming, H. (1977): Comparison of linear inversion methods by examination of the duality between iterative and inverse matrix methods. pp. 325-355. In Inversion Methods in Atmospheric Remote Sounding, A. Deepak (ed.), Academic Press.

Fleming, H. (1983): Satellite remote sensing by the technique of computed tomography. J. Applied Meteorology, 21, 1539-1549.

Fritz, S., Wark, D., Fleming, J., Smith, W., Jacobowitz, H., Hilleary, D., and Alishouse, J. (1972): Temperature sounding from satellites. NOAA Tech. Rept. NESS 59, National Oceanic and Atmospheric Administration, Washington, D. C.

Golub, G. H., Heath, M., and Wahba, G. (1979): Generalized cross validation as a method for choosing a good ridge parameter. Technometrics, 21, 215-224.

Halem, M. and Kalnay, E. (1983): Variational global analysis of satellite temperature soundings, manuscript, GLAS.

Hoffman, R. (1982): SASS wind ambiguity removal by direct minimization. Monthly Weather Revies, 110, 434-445.

Kasahara, A. and Puri, K. (1981): Spectral representation of three-dimensional global data by expansion in normal mode functions. Monthly Weather Review, 109, 37-51.

Kimeldorf, G. and Wahba, G. (1971): Some results on Tchebycheffian spline functions. J. Math. Anal. Applic., 33, 82-95.

Marshall, A. and Olkin, I. (1979): Inequalities: Theory of Majorization and Its Applications. New York: Academic Press.

Nychka, D., Wahba, G., Goldfarb, S., and Pugh, T. (1983): Cross-validated spline methods for the estimation of three dimensional tumor size distributions from observations on two dimensional cross sections. University of Wisconsin-Madison Statistics Dept. TR 711.

O'Sullivan, F. (1983): The analysis of some penalized likelihood estimation schemes. Thesis, University of Wisconsin-Madison, Department of Statistics. TR No. 726.

Smith, B., Boyle, J., Dongarra, J., Garbow, B., Ikebe, Y., Klema, V., and Moler, C. (1976): Matrix Eigensystem Routines-EISPACK Guide, New York: Springer-Verlag.

Smith, W. and Woolf, H. (1976): The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations. J. Atmos. Sci., 33, 1127-1140.

Stanford, J. (1979): Latitudinal-wavenumber power spectra of stratospheric temperature fluctuations. J. Atmos. Sci., 36, 921-931.

Suomi, V. (1983): High equatorial orbit satellites, Talk, Weather Forecasting and Analysis Seminar Madison, WI, Sept. 20.

Wahba, G. (1977a): Optimal smoothing of density estimates. pp. 423-458. In Classification and Clustering, J. Van-Ryzin (ed.), Academic Press.

Wahba, G. (1977b): Practical approximate solutions to linear operator equations when the data are noisy. SIAM J. Numerical Analysis, 14, 651-667.

Wahba, G. (1978a): Improper priors, spline smoothing and the problem of guarding against model errors in regression. J. R. Statist. Soc. B, 40, 364-372.

Wahba, G. (1978b): Comments on Curve fitting and optimal design for prediction by A. O'Hagan. J. Roy. Stat. Soc. B, 40, 35-36.

Wahba, G. (1980): Ill posed problems: Numerical and statistical methods for mildly, moderately, and severely ill posed problems with noisy data. University of Wisconsin-Madison, Statistics Department Technical Report No. 595 (to appear in the Proceedings of the International Conference on Ill Posed Problems, M.Z. Nashed, ed.)

Wahba, G. (1982a): Variational methods in simultaneous optimum interpolation and initialization, in The Interaction between Objective Analyses and Initialization, D. Williamson, ed. Publication in Meteorology 127, Atmospheric Analysis and Prediction Division, NCAR, Boulder, CO, 178-185.

Wahba, G. (1982b): Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data. In Multivariate Approximation Theory, Vol. 2, W. Schempp and K. Zeller, eds., 407-429, Birkhauser Verlog.

Wahba, G. and Wendelberger, J. (1980): Some new mathematical methods for variational objective analysis using splines and cross-validation. Monthly Weather Review, 108, 36-57.

Wark, D. and Fleming, H. (1966): Indirect measurements of atmospheric temperature profiles from satellites. Monthly Weather Review, 94, 351-362.

Weinreb, M. and Crosby, D. (1972): Optimization of spectral intervals for remote sensing of atmospheric temperature profiles. Remote sensing of environment, 2, 193-201.

Appendix A. Proofs.

Proof of (12).

Letting $Q^{1/2}$ be the symmetric square root of Q,

$E\beta' M_1' Q M_1 \beta =$

$$= b \text{ Trace } M_1' Q M_1 \Sigma$$

$$= b \text{ trace } Q^{1/2} M_1 \Sigma M_1' Q^{1/2}$$

$$= b \text{ Trace } Q^{1/2}\{\Sigma - 2\Sigma X'(X\Sigma X' + n\lambda I)^{-1} X\Sigma + \Sigma X'(X\Sigma X' + n\lambda I)^{-1} X\Sigma X' (X\Sigma X' + n\lambda I)^{-1} X\Sigma\} Q^{1/2} \quad (A.1)$$

$E\varepsilon' M_2' Q M_2 \varepsilon$

$$= \sigma^2 \text{ Trace } M_2' Q M_2$$

$$= \sigma^2 \text{ Trace } Q^{1/2} M_2 M_2' Q^{1/2}$$

$$= \sigma^2 \text{ Trace } Q^{1/2} \{\Sigma X'(X\Sigma X' + n\lambda I)^{-2} X\Sigma\} \quad (A.2)$$

Using

$$(X\Sigma X' + n\lambda I)^{-1} X\Sigma X'(X\Sigma X' + n\lambda I)^{-1} = (X\Sigma X' + n\lambda I)^{-1} - n\lambda(X\Sigma X' + n\lambda I)^{-2}$$

and adding (A.1) and (A.2) gives,

$$\text{Trace } Q^{1/2}\Sigma^{1/2} \{b[I - \Sigma^{1/2}X(X\Sigma X' + n\lambda I)^{-1}X\Sigma^{1/2}]$$

$$+ (\sigma^2 - n\lambda b)\Sigma^{1/2}X(X\Sigma X' + n\lambda I)^{-2}X\Sigma^{1/2}\}\Sigma^{1/2}Q^{1/2} \quad (A.3)$$

which gives (12).

Proof of Theorem

Suppose that $\beta X_1'X_1\beta' \geq \beta X_2'X_2\beta'$ for any β. We will show that this implies that

$$\text{Trace } Q^{1/2}\Sigma X_1'(X_1\Sigma X_1' + n\lambda I)^{-1} X_1\Sigma Q^{1/2}$$
$$\geq \text{Trace } Q^{1/2}\Sigma X_2'(X_2\Sigma X_2' + n\lambda I)^{-1}X_2\Sigma Q^{1/2} \quad (A.6)$$

for any Q and λ.

We will assume that Σ is nonsingular. Then our hypotheses imply that $\beta'\Sigma^{1/2}X_1'X_1\Sigma^{1/2}\beta \geq \beta'\Sigma^{1/2}X_2'X_2\Sigma^{1/2}\beta$ for any β, in other words,

$$\Sigma^{1/2}X_1'X_1\Sigma^{1/2} \succcurlyeq \Sigma^{1/2}X_2'X_2\Sigma^{1/2},$$

where $A \succcurlyeq B$ means A-B is nonnegative definite.

Let $A = \Sigma^{1/2}X_1$, $B = \Sigma^{1/2}X_2$, where A and B are p x n.

We have to show that

$$AA' \succsim BB' \Rightarrow A(A'A + n\lambda I)^{-1} A' \succsim B(B'B + n\lambda I)^{-1}B' .$$

We first show that $A(A'A + n\lambda I)^{-1} A' \equiv AA'(AA' + n\lambda I)^{-1}$.

This is equivalent to showing

$$A(A'A + n\lambda I)^{-1} A' \ (AA' + n\lambda I) = AA' .$$

Expanding the left hand side gives

$$A(A'A + n\lambda I)^{-1} (A'A) A' + n\lambda A(A'A + n\lambda I)^{-1} A'$$

$$= A(A'A + n\lambda I)^{-1} (A'A + n\lambda I) A' - n\lambda A(A'A + n\lambda I)^{-1} A' + n\lambda A(A'A + n\lambda I)^{-1}A'$$

$$= AA' \quad .$$

Now, let $(AA' + n\lambda I) = C$ and $(BB' + n\lambda I) = D$.

Therefore

$$AA' \ (AA' + n\lambda I)^{-1} = (C - n\lambda I)C^{-1} = I - n\lambda C^{-1}$$

$$BB' \ (BB' + n\lambda I) = (D - n\lambda I)D^{-1} = I - n\lambda D^{-1} \quad .$$

Now $AA' \succsim BB' \Rightarrow C \succsim D$, and $C \succsim D \Rightarrow C^{-1} \precsim D^{-1}$ (See, e.g. Marshall and Olkin, p. 464), which in turn implies that $I - n\lambda C^{-1} \succsim I - n\lambda D^{-1}$, so the proof is finished.

Appendix B. Remarks on the specification of σ^2/nb.

Suppose that T^{δ} has an infinite series expansion in the $\psi_a \phi_{\gamma}$. Under the assumption that $E\beta_j^2 = b\sigma_{jj}$, $j = 1,2,..\infty$, and (for mathematical convenience only) $E\beta_j\beta_k = 0$, $j \neq k$, then for each $p = 1,2,...,$

$$\frac{1}{p} E\beta'(p)\ \Sigma^{-1}(p)\beta(p) = \frac{1}{p}\sum_{j=1}^{p} \frac{b\sigma_{jj}}{\sigma_{jj}} = b, \tag{B.1}$$

where $\beta(p)$ and $\Sigma(p)$ are the first p and $p \times p$ components of β and Σ respectively. A different modelling assumption is, that T^{δ} has the property that

$$\sum_{j=1}^{\infty} \frac{\beta_j^2}{\sigma_{jj}} < \infty \quad . \tag{B.2}$$

Under this assumption β_λ of (7) is still an appropriate estimate of the first p components of β, for appropriately chosen λ. (See, e.g. Wahba (1977a)), but

$$\frac{1}{p}\ \beta'(p)\ \Sigma^{-1}(p)\ \beta(p) \to 0$$

as $p \to \infty$ so that b is not readily defined independent of p. GCV will return a good estimate of λ under either assumption (B.1) or (B.2) (see Wahba 1977b) and the design criteria resulting from the assumptions of this paper (i.e. assumption B.1) appear eminently plausible even if (B.2) is true. A related but somewhat harder to study design criteria under assumption (B.2) appears in Wahba (1978b).