N85-31213

A RESIDUALS APPROACH TO FILTERING, SMOOTHING AND IDENTIFICATION FOR STATIC DISTRIBUTED SYSTEMS

G. Rodriguez
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

ABSTRACT

This paper advances an approach for state estimation and identification of spatially distributed parameters embedded in static distributed (elliptic) system models.

The method of maximum likelihood is used to find parameter values that maximize a likelihood functional for the system model, or equivalently, that minimize the negative logarithm of this functional. To find the minimum, a Newton-Raphson search is conducted that from an initial estimate generates a convergent sequence of parameter estimates. Central to the numerical search are a gradient functional and a Hessian operator, which are respectively the first and second function-space derivatives of the negative-log likelihood functional with respect to the parameter distributions being identified. For simplicity, a Gauss-Markov approach is used to approximate the Hessian in terms of products of first derivatives. The gradient and approximate Hessian are computed by first arranging the negative-log likelihood functional into a form based on the square-root factorization of the predicted covariance of the measurement process. The resulting data-processing approach, referred to here by the new term of predicted-data-covariance square-root filtering, makes the gradient and approximate Hessian calculations very simple. Since the parameter estimates are only approximations to the actual parameter values, there is a parameter estimation error inherent in the estimation process. Cramer-Rao bounds are obtained for the covariance of the estimation error in terms of the information operator associated with the likelihood functional. These error covariance bounds are then used to outline methods for optimal input design.

A closely related set of state estimates is also produced by the maximum likelihood method: smoothed estimates that are optimal in a conditional mean sense and filtered estimates that emerge from the predicted-data-covariance square-root filter. The terms "smoothed" and "filtered" are used because the formulas which generate these estimates, when expressed in operator notation, are symbolically very similar to those used in filtering and smoothing for linear dynamical systems. A key similarity is the presence of a predictor-corrector structure containing estimator gains that, as in a Kalman filter, can be expressed in terms of the state estimation error covariances. In addition, a residual process can be defined, in the usual way, as the difference between the actual data and the predicted data obtained from the filtered state estimate. The residuals have properties nearly identical to those of an innovations process: the residuals are white with a unit covariance; and the residuals and measurements can be obtained from each other by means of reciprocal linear transformations. Because these transformations are not Volterra (causal), the residuals are not a bona fide innovations process. Even though they are not a true innovations process, the residuals are very useful, because they lead to state and parameter estimation schemes for elliptic systems that retain conceptually the simplicity of those obtained by the innovations approach to filtering, smoothing and identification for linear dynamical systems.

1. INTRODUCTION AND SUMMARY

The elliptic models considered in this paper can be cast as

$$A(\Theta)u(\Theta) = B(\Theta)\omega + C(\Theta)f, \tag{1.1}$$

$$y = H(\Theta)u(\Theta) + n, \tag{1.2}$$

where A is a formally self-adjoint elliptic differential operator defined over the spatial domain Ω; B and C are appropriately dimensioned operators that model the influence of the process error ω and the input f on the state u; H is an operator that characterizes the state-to-observations map; ω and n are white-noise model errors forming the model error vector $\varepsilon=[\omega,n]$; and f is a deterministic input. Examples of the application of such models to the problem of static shape determination of large space structures are contained in Ref. [1].

The central aim here is to develop a maximum-likelihood approach to the estimation of the parameters Θ (these parameters could in general be spatially distributed) by using the data y and the system model itself. It is assumed that the true value Θ_0 of the parameter Θ is a deterministic but poorly known quantity. The input f can be selected to optimize the data generated for estimation. A related but somewhat secondary aim is to develop a methodology for computation of the corresponding state estimates.

A Formula for the Negative-Log Likelihood Ratio

It will be shown in Sec. 3 that the negative-log likelihood functional is specified by

$$J(\Theta;y) = \tfrac{1}{2}\,\mathrm{Tr}\log[I+R(\Theta)] + \tfrac{1}{2}[y-m(\Theta)]^*[I+R(\Theta)]^{-1}[y-m(\Theta)] - \tfrac{1}{2}\,y^*y, \tag{1.3}$$

where

$$m(\Theta) = H(\Theta)\Phi(\Theta)C(\Theta)f \quad \text{and} \quad R(\Theta) = H(\Theta)\Phi(\Theta)B(\Theta)B^*(\Theta)\Phi^*(\Theta)H^*(\Theta).$$

The integral operator $\Phi(\Theta)$ is related to $A(\Theta)$ by $A(\Theta)\Phi(\Theta)=I$, with I the identity. The symbol Φ^* denotes the integral operator adjoint to Φ so that $\Phi^*(\Theta)A^*(\Theta)=I$. It will also be shown in Sec. 3 that $m(\Theta)$ and $R(\Theta)$ are respectively the "suspected" mean and covariance of the data y, under the assumption that the model error vector $\varepsilon=[\omega,n]$ is a spatially distributed white-noise process [1] with a covariance operator $E(\varepsilon\varepsilon^*)=I$ equal to the identity. To simplify Eq. (1.3), the following notation has been used:

$$y^*y = \langle y,y\rangle_s \quad \text{and} \quad [y-m(\Theta)]^*[I+R(\Theta)]^{-1}[y-m(\Theta)] = \langle [y-m(\Theta)],\,[I+R(\Theta)]^{-1}[y-m(\Theta)]\rangle_s,$$

where $\langle\cdot,\cdot\rangle_s$ indicates an inner product in the function space to which the data belongs.

Predicted-Data-Covariance Square-Root Form of the Likelihood Ratio

A number of alternative formulas for the negative-log likelihood functional are developed in Sec. 4. To solve the above minimization problem, the most convenient formula is:

$$J(\Theta;y) = \mathrm{Tr}\log[I+K(\Theta)] + \tfrac{1}{2}\, z^*(\Theta)z(\Theta) - z^*(\Theta)y, \tag{1.4}$$

where

$$z(\Theta) = L(\Theta)y + [I-L(\Theta)]m(\Theta), \tag{1.5}$$

$$L(\Theta) = I-[I+R(\Theta)]^{-\frac{1}{2}}, \qquad K(\Theta) = [I+R(\Theta)]^{\frac{1}{2}}-I. \tag{1.6}$$

Equation (1.5) can be viewed as specifying a filter, characterized by the operator $L(\Theta)$, that processes the data y and the suspected mean $m(\Theta)$ to provide a "filtered" state estimate $z(\Theta)$. This filter $L(\Theta)$ will hereafter be referred to as the predicted data-covariance square-root filter because the key calculation required to specify $L(\Theta)$, as in (1.6), is the evaluation of the square-root of the predicted-data-covariance operator $[I+R(\Theta)]$. The equivalence between (1.3) and (1.4) can be established by substitution of (1.5) and (1.6) into (1.4).

Note for later reference that the definitions in (1.6) imply that $K(\Theta)$ and $L(\Theta)$ are related by

$$[I+K(\Theta)]^{-1} = I-L(\Theta). \tag{1.7}$$

Furthermore, (1.7) implies that $K(\Theta) = L(\Theta)+K(\Theta)L(\Theta) = L(\Theta)+L(\Theta)K(\Theta)$.

Gradient of the Likelihood Functional

The gradient functional $\partial J/\partial\Theta$, to be defined more completely in Secs. 5 and 6, is specified by

$$\partial J/\partial\Theta = \mathrm{Tr}[(\partial L/\partial\Theta)(I+K)] + (z-y)^*(\partial z/\partial\Theta), \tag{1.8}$$

where

$$\partial z/\partial\Theta = (\partial L/\partial\Theta)\bar{y} + (I-L)(\partial m/\partial\Theta), \tag{1.9}$$

with $\bar{y} = y-m$, and $\partial L/\partial\Theta$, $\partial m/\partial\Theta$ being the function-space Frechet derivatives of L and m. These equations can be obtained from (1.4) by function-space differentiation with respect to Θ.

The gradient functional $\partial J(\Theta;y)/\partial\Theta$ in (1.8) is the Frechet derivative [2] of the functional J with respect to the parameter Θ. The derivative is a linear transformation (assumed to be bounded) that maps an admissible parameter perturbation $\delta\Theta$ into the corresponding perturbation $\delta J(\Theta,\delta\Theta;y)$ of the likelihood functional by means of the equation $\delta J(\Theta,\delta\Theta;y) = [\partial J(\Theta;y)/\partial\Theta]\,\delta\Theta$. Detailed computation of the function-space derivatives above is conducted in Sec. 6 using a perturbation analysis of the eigensystem of the covariance operator $R = H\Phi BB^*\Phi^*H^*$ obtained in Sec. 5.
Note that in Sec. 7 it will be established that

$$E[\partial J(\Theta;y)/\partial\Theta]_{\Theta=\Theta_o} = 0, \tag{1.10}$$

so that the expected value of the gradient vanishes at the optimal parameter value Θ_o.

Hessian of the Likelihood Functional

Similarly, differentiation of (1.8) leads to

$$\partial^2 J/\partial\Theta^2 = \mathrm{Tr}\,[(\partial^2 L/\partial\Theta^2)(I+K) + (\partial L/\partial\Theta)(\partial K/\partial\Theta)] + (z-v)^*(\partial^2 z/\partial\Theta^2) + (\partial z/\partial\Theta)^*(\partial z/\partial\Theta), \quad (1.11)$$

and to its expected value at $\Theta=\Theta_o$ of $M(\Theta_o) = E[\partial^2 J/\partial\Theta^2]|_{\Theta=\Theta_o}$, i.e.,

$$E[\partial^2 J/\partial\Theta^2]\,|_{\Theta=\Theta_o} = \mathrm{Tr}\,[(\partial L/\partial\Theta)(I+R)(\partial L^*/\partial\Theta)] + E[(\partial z/\partial\Theta)^*(\partial z/\partial\Theta)]. \quad (1.12)$$

Furthermore, substitution of (1.9) in the last term of (1.12) leads to

$$E[\partial^2 J/\partial\Theta^2]\,|_{\Theta=\Theta_o} = 2\mathrm{Tr}\,[(\partial L/\partial\Theta)(I+R)(\partial L^*/\partial\Theta)] + [(I-L)(\partial m/\partial\Theta)]^* [(I-L)(\partial m/\partial\Theta)]. \quad (1.13)$$

Note that the expected value of the Hessian operator $\partial^2 J/\partial\Theta^2$ evaluated at $\Theta=\Theta_c$ is a sum of two terms each of which is positive definite. Consequently, in a probabilistic sense made precise by (1.13), the likelihood functional is strictly convex in the vicinity of the optimal value $\Theta=\Theta_o$. Note that by definition $M(\Theta_o)$ in (1.13) is also the information operator associated with the likelihood functional.

Newton-Raphson Search for the Optimal Parameter Estimates

Since the problem of minimization of $J(\Theta;y)$ in (1.4) has no closed-form solution, it is necessary to consider iterative numerical search techniques for optimization. The following constitutes a function-space Newton-Raphson iteration:

$$\Theta^{n+1} = \Theta^n - M_n^{-1} g_n, \quad (1.14)$$

where $g_n = \partial J(\Theta^n;y)/\partial\Theta$ is the gradient functional (1.8) evaluated at $\Theta=\Theta^n$; and where

$$M_n = \mathrm{Tr}\,[(\partial L/\partial\Theta)(I+R)(\partial L^*/\partial\Theta)] + (\partial z^*/\partial\Theta)(\partial z/\partial\Theta)|_{\Theta=\Theta_n} \quad (1.15)$$

is an approximation to the Hessian operator $\partial^2 J/\partial\Theta^2$ in (1.11). This approximation is obtained from (1.12) by replacing the second term $E[(\partial z^*/\partial\Theta)(\partial z/\partial\Theta)]$ with the actual value $[(\partial z^*/\partial\Theta)(\partial z/\partial\Theta)]$ obtained in a single realization. Under certain conditions, to be

examined in more detail in future work, the sequence Θ^n specified by (1.14) converges to a local minimum of $J(\Theta;y)$, if the initial estimate used to start the search is sufficiently close to the optimal value.

Cramer-Rao Bounds and Optimal Input Design

The above numerical search results in an estimate Θ of the actual parameter value Θ_o. In Sec. 7, a C-R bound for the covariance $E(\Theta_p\Theta_p^*)$ of the estimation error $\Theta_p = \Theta-\Theta_o$ is obtained from the inequality.

$$E(\Theta_p\Theta_p^*) \geq M^{-1}(\Theta_o), \tag{1.16}$$

where the information operator $M(\Theta_o)$ is specified in (1.13). The related mean-square estimation error is bounded by $E(\Theta_p^*\Theta_p) \geq Tr[M^{-1}(\Theta_o)]$.

The information operator $M(\Theta_o)$ can also be used to specify criteria for optimal input design. Perhaps the simplest optimal selection method to implement is that which seeks to maximize $Tr\,[M(\Theta_o)]$ with respect to f, subject to the constraint that f satisfy the normalization condition of $f^*f=1$. This method results in an optimal f which is the eigenvector corresponding to the largest eigenvalue of a positive-definite matrix described in detail in Sec. 7. Other criteria for optimal selection based on calculation of $M^{-1}(\Theta_o)$ may be more difficult to implement but usually lead to superior performance.

The Corresponding State Estimates

Related to the parameter estimation approach are the following two distinct state estimates (denoted by u_o and z_o):

$$u_o = E(u/y) = \Phi Cf + G(y-H\Phi Cf), \quad z_o = \Phi Cf + g(y-H\Phi Cf), \tag{1.17}$$

where G and g are Kalman-like gains (see Sec. 8) specified by

$$G = \sum \sin^2\alpha_k x_k \phi_k^*, \qquad g = \sum (1-\cos\alpha_k)\, x_k \phi_k^*. \tag{1.18}$$

In these equations, ϕ_k are the eigenvectors of the operator $R = H\Phi BB^*\Phi H^*$, so that $R\phi_k = \lambda_k^2\phi_k$, with λ_k^2 being the related eigenvalues. Also, α_k and x_k are defined by $\tan\alpha_k = \lambda_k$ and $x_k = \lambda_k^{-2}\Phi BB^*\Phi^*H^*\phi_k$.

The state estimate $u_o = E(u/y)$ is defined as the conditional expectation of the state given the data y. Since u_o is an optimal estimate of u based on the entire data set (as

opposed to a subset), u_o can be viewed as a best smoothed estimate. The other estimate, z_o in (1.17), will be referred to as a filtered state estimate. The filtered estimate has no known probabilistic interpretation similar to $u_o = E(u/y)$ above. However, in spite of the apparent lack of probabilistic meaning, this estimate is useful in simplifying the gradient and Hessian calculations in (1.8) and (1.11). It will be shown in Sec. 8 that z_o in (1.17) and z, the estimate emerging from the predicted-data-covariance square-root filter, are related by $z = Hz_o$. Hence, z_o is a bona fide estimate of the entire state, whereas $z = Hz_o$ is a partial estimate defined only at the observation locations.

Kalman-like Gains and Error Covariances

The gains G and g in (1.17) can alternatively be specified in terms of the covariance of the state estimation error inherent in u_o and z_o, i.e.,

$$G = \bar{P}H^*, \qquad g = pH^*, \tag{1.19}$$

where

$$\bar{P} = E[(u-u_o)(u-u_o)^*], \quad p + p^* = E[(u-z_o)(u-z_o)^*]. \tag{1.20}$$

The corresponding mean-square state estimation error is

$$E[(u-u_o)^*(u-u_o)] = Tr[\bar{P}], \quad E[(u-z_o)^*(u-z_o)] = Tr[p+p^*]. \tag{1.21}$$

Furthermore, $\bar{P}$ and p are related by

$$\bar{P} = p + p - pH^*Hp. \tag{1.22}$$

Since the term pH*Hp is non-negative, the mean-square estimation error associated with the smoothed estimate u_o is never larger than that of the filtered estimate z_o.

Filtering and Smoothing

While u_o and z_o have been defined somewhat independently in (1.17), they are related by:

$$u_o = z_o + pH^*e, \tag{1.23}$$

where

$$e = y - Hz_o = (I - HpH^*)\bar{y} = (I - L)\bar{y} \tag{1.24}$$

is the residual process defined as the difference between the data y and the observed-state estimate Hz_o. The symbol $\bar{y}$ in (1.24) denotes the mean-centered data

process $\bar{y} = y - H\Phi Cf$. It will be shown in Sec. 8 that (1.22) and (1.23) constitute a generalization to elliptic systems of the forward/backward sweep method for solution of smoothing problems in linear dynamical systems.

The Residuals as a Pseudo-Innovations Process

The residuals in Eq. (1.24) have two properties that are similar (but not identical) to those of an innovations process:

$$E(ee^*) = I, \tag{1.25}$$

$$e = (I - L)\bar{y}, \qquad \bar{y} = (I + K)e. \tag{1.26}$$

Eq. (1.25) reflects whiteness of the residuals. Eq. (1.26) states that the residual and mean-centered data processes e and $\bar{y}$ can be obtained from each other by means of reciprocal transformations, i.e., $(I + K)^{-1} = (I - L)$ as in (1.7). Whiteness of the innovations and reciprocal relationships between innovations and measurements are the two central features of the innovations approach to least-squares estimation for linear dynamical systems. Eqs. (1.25) and (1.26) are similar to these conditions. However, there is a key difference: the transformations (I + K) and (I - L) in (1.26) are Fredholm operators whose domain is the entire measurement space. This is in contrast to the Volterra (causal) operators in the innovations approach for linear dynamical systems. The notion of causality is not even used in this paper, although such a notion can be defined for certain classes of elliptic systems [1]. Because of this difference the residual process is not a bona fide innovations process. However, the residual process is still useful in obtaining the relatively simple formulas in (1.8) - (1.26) for filtering, smoothing and identification.

Paper Outline

This section has at a summary level addressed many of the fundamental issues involved in the maximum likelihood approach to estimation. The subsequent sections of the paper contain a more complete description of the above results.

Section 2. Development of the mathematical framework -- including integral operator models, a priori covariance analysis with white-noise model errors, Fredholm resolvents, and eigenfunction expansions -- required to arrive at formula (1.3) for the likelihood functional and to evaluate the corresponding function-space gradient in (1.8) and the approximate Hessian in (1.15).

Section 3. Derivation of the negative-log likelihood functional in (1.3). This functional is the negative logarithm of the likelihood ratio, associated with the detection of a Gaussian signal in additive Gaussian noise, traditionally encountered in the theory for communication and signal detection.

Section 4. Development of alternative formulas for the likelihood ratio, some of which are more convenient to use than (1.3) in implementing the numerical search for optimization -- in particular, development of the predicted-data-covariance square-root filter form (1.4) upon which the Newton-Raphson search is based. Additional forms of the likelihood ratio which are of interest in their own right

(although not subsequently used in the paper) are: a smoothing form expressed in terms of the best mean-square state estimate; an eigensystem expansion form based on the eignenvalues and eigenvectors of the operator $R=H\Phi BB^*\Phi^*H^*$ in (1.3); a trigonometric operator form with which most of the manipulations involved in the maximum likelihood approach can be visualized using their similarities to simple trigonometric formulas for scalars.

Section 5. Development of a first-order perturbation analysis to evaluate the infinitesimal changes in the eigensystem of the operator $R=H\Phi BB^*\Phi^*H^*$ in (1.3) due to similarly small changes $\delta\Theta$ in the parameter distributions being identified. This is the central calculation required to compute the function-space gradients $\partial J/\partial\Theta$, $\partial z/\partial\Theta$, $\partial L/\partial\Theta$ and $\partial m/\partial\Theta$ in (1.8) and (1.9).

Section 6. Calculation of the gradient functional and approximate Hessian of the likelihood functional based on the perturbation analysis of Sec. 5. These are the two calculations which are central to implementation of the Newton-Raphson search and which have been used as a basis for computer programs to implement the maximum likelihood approach.

Section 7. Parameter estimation error covariance analysis and Cramer-Rao bounds based on explicit formulas for the Hessian (information) operator in (1.13). Outline of an optimal input design approach based on using the Cramer-Rao bound as an optimality criterion.

Section 8. Analysis of the filtered and smoothed state estimates embedded in the parameter estimation approach. Analysis of the predicted-data-covariance square-root filter resulting in Kalman-like formulas for the filter gain, evaluation of the state estimation error covariance, and relationships between filtered and smoothed estimates.

Section 9. Summary and explanation of calculations required for implementation of the numerical search for the optimal estimates.

Section 10. Conclusions and directions for future work in the areas of development of asymptotic properties of the estimates and of optimal input design.

2. PRELIMINARIES: Notation, Integral Operator Model, Covariance Analysis, Fredholm Resolvents, and Eigenfunction Expansions

The aim of this section is to develop a set of miscellaneous results that will be useful in subsequent sections in conducting detailed derivation of: the negative-log likelihood functional in (1.3) to be minimized, the corresponding function-space gradient in (1.8), and the approximate Hessian operator in (1.15). The main results of the section can be summarized as follows:

- conversion of the partial differential operator model in (1.1) to an equivalent integral operator formulation. This integral operator formulation is introduced because it simplifies the statement and solution of the estimation problems in (1.1) - (1.3).

- evaluation of the observed state covariance operator $R=H\Phi BB^*\Phi^*H^*$ in (1.3), under the assumption that $\epsilon = [\omega,n]$ is a spatially distributed white-noise process with a unit covariance operator. Related to evaluation of this covariance operator R is the similar evaluation of the suspected mean $m = H\Phi Cf$ in (1.3).

- evaluation of the dual observed-state covariance operator $Q=B^*\Phi^*H^*H\Phi B$ - which can be viewed as the covariance of the output of a system model dual to (1.1), under the assumption that this dual system is driven by a white-noise process.

- definition of two sets of eigenvectors ϕ_k and ψ_k of R and Q above, with λ_k^2 being a set of common eigenvalues. These two sets of vectors can be used to expand functions in the input space H_1 and the output space H_3.

- definition of the vectors x_k and p_k in the state space H_2 and its dual H_2^*, related to ψ_k and ϕ_k above by $x_k = \lambda_k^{-1}\Phi B\psi_k$ and $p_k = \lambda_k^{-1}\Phi^*H^*\phi_k$. These two sets of vectors x_k and p_k satisfy a boundary-value problem similar to those traditionally encountered as necessary and sufficient conditions for optimality in quadratic optimal control and estimation problems subject to linear constraints.

- analysis of the basic relationship between R and Q above and their corresponding Fredholm resolvents P and S defined as $P = I-(I+R)^{-1}$ and $S = I-(I+Q)^{-1}$. Expansion of the operators R, Q, P and S in terms of the eigenfunctions ϕ_k and ψ_k defined above.

- development of trigonometric operator forms for R and P. These trigonometric forms allow development of interesting trigonometric alternatives to (1.3) in evaluating the likelihood functional.

While the section concentrates on the development of a mathematical framework to be used in subsequent sections, many of the above results (such as the trigonometric operator formulas for the covariance operators) are of interest in their own right, somewhat independently of their subsequent application.

Hilbert Space Notation

There are three Hilbert spaces of primary interest: the input space H_1 to which the process error ω and the deterministic input f belong; the state space H_2 containing the state u; and the measurement space H_3 where the data y and the observation error n belong. The inner product between two arbitrary elements u and v in the space H_i is denoted by $\langle u,v\rangle_i$ or by the simpler notation $u^*v = \langle u,v\rangle_i$. Similarly, uv^* denotes a Hilbert space outer product.

Conversion to Integral Operator Model

It is convenient for subsequent developments to convert (1.1) to an equivalent integral operator formulation. To this end, define the Green's function $\phi(x/\xi)$ of A as the solution of

$$A_x \phi(x/\xi) = \delta(x-\xi), \tag{2.1}$$

where δ is the impulsive delta function, and where the subscript x in A_x denotes that the spatial differentiations embedded in A are performed with respect to x (as opposed to being performed with respect to ξ). Define then the integral operator Φ whose kernel is the Green's function, i.e.,

$$\Phi v = \int_\Omega \phi(x/\xi) v(\xi) d\xi, \tag{2.2}$$

for all admissible functions v. Note that Φ is the integral operator such that $A\Phi = I$, where I is the appropriately dimensioned identity.

With these definitions at hand, it is possible to recast (1.1) and (1.2) as

$$y = m(\Theta) + H(\Theta)\Phi(\Theta)B(\Theta)\omega + n, \tag{2.3}$$

where $m(\Theta)$ is the "suspected mean"

$$m(\Theta) = H(\Theta)\Phi(\Theta)C(\Theta)f. \tag{2.4}$$

Equation (2.3) can be cast into the following even more compact notation

$$y = m(\Theta) + h(\Theta)\varepsilon, \tag{2.5}$$

where $\varepsilon = [\omega,n]$ is the model error vector [1], and $h(\Theta)$ is the operator $h(\Theta) = [H(\Theta)\Phi(\Theta)B(\Theta) \mid I]$.

Predicted-Data Mean and Covariance

The evaluation of the predicted mean and covariance of y, needed as a preliminary step to arrive at (1.3), is based on the key assumption that the model error vector $\varepsilon = [\omega,n]$ is a zero-mean spatially distributed white-noise process whose covariance operator $E(\varepsilon\varepsilon^*)$ is the identity, i.e.,

$$E(\varepsilon\varepsilon^*) = I, \tag{2.6}$$

where I is the appropriately dimensioned identity. Note that this assumption is not at all restrictive, because the more general case where the model errors $\varepsilon = [\omega,n]$ are correlated (with a nonidentity covariance operator) can be handled within the same formulation by selection of the operator B in (1.1). It is assumed here that BB^* is bounded and trace-class, with kernel $b(x/\xi)$ satisfying $\int_\Omega b(x/x)dx < \infty$.

Remark 2.1 The process y is a random field with mean and covariance specified as

$$E(y) = m(\Theta), \qquad E\{[y-m(\Theta)]\,[(y-m(\Theta)]^*\} = I+R(\Theta), \tag{2.7}$$

where $R(\Theta) = H(\Theta)\Phi(\Theta)B(\Theta)B^*(\Theta)\Phi^*(\Theta)H^*(\Theta)$. That $E(y) = m$ follows from (2.5) and the fact that ϵ is zero-mean. The second of Eqs. (2.7) follows from the following sequence of operations: $E[(y-m)(y-m)^*] = E[h\epsilon\epsilon^*h^*] = hE(\epsilon\epsilon)h^* = hh^* = I+R$. A more detailed development of the above result is contained in Ref. [1].

Remark 2.2 The process u in (1.1), representing the state of the system, is a random field with mean and covariance specified by

$$E(u) = \Phi Cf, \qquad E[(u-\Phi Cf)(u-\Phi Cf)^*] = \bar{R}(\Theta), \tag{2.8}$$

where

$$\bar{R}(\Theta) = \Phi(\Theta)B(\Theta)B^*(\Theta)\Phi^*(\Theta). \tag{2.9}$$

Note that the state covariance $\bar{R}$ and the "observed-state" covariance R in (2.7) and (2.8) are related by

$$R(\Theta) = H\bar{R}(\Theta)H^*. \tag{2.10}$$

Remark 2.3 The state covariance $\bar{R}$ satisfies the partial differential equation

$$A\bar{R}A^* = BB^*, \tag{2.11}$$

a result which can be established by pre-multiplication of $\bar{R}$ in (2.9) by A and subsequent post-multiplication by A^*.

Remark 2.4 The state covariance operator $\bar{R}$ can be represented as the following integral operator

$$\bar{R}v = \int_\Omega r(x/\xi)\, v(\xi)d\xi, \tag{2.12}$$

where the kernel $r(x/\xi)$ satisfies

$$A_x r(x/\xi)A_\xi = b(x/\xi), \tag{2.13}$$

and where $b(x/\xi)$ is the kernel of BB^*. This result can be established by means of the following sequence of operations. Consider an admissible function v (admissible in the sense that it can be operated on by the operators $A\bar{R}A^*$ and BB^* in (2.11) so that $A\bar{R}A^*v = BB^*v$ makes sense). In terms of the corresponding kernels r and b, this last equation becomes

$$A_x\{\int_\Omega r(x/\xi)\,[A_\xi^* v(\xi)]d\xi\} = \int_\Omega [A_x r(x/\xi)A_\xi]\, v(\xi)d\xi = \int_\Omega b(x/\xi)v(\xi)d\xi, \tag{2.14}$$

where the first equality is valid because by definition A_ξ^* is the formal adjoint of A_ξ. Since Eq. (2.14) must be valid for all admissible v, then (2.14) implies (2.13).

Remark 2.5 The state covariance kernel r in (2.12) can be expressed as

$$r(x/\xi) = \int_\Omega \int_\Omega \phi(x/\eta)\, b(\eta/\beta)\, \phi(\beta/\xi)\, d\eta d\beta, \tag{2.15}$$

where ϕ is the Green's function of A, and b is the kernel of BB^*. This result can be established by expressing (2.9) in terms of the operator kernels ϕ and b of Φ and BB^* and by subsequent reversal of the order of integration.

Remark 2.6 In the special case, of interest in many applications, where the process error ω is discretely located at the M locations $[\eta_1,\ldots,\eta_M]$, and the sensors are placed at the N locations $[\xi_1,\ldots,\xi_N]$, then $R = H\Phi BB^*\Phi^*H^*$ is a matrix whose general element R_{ij} is specified by

$$R_{ij} = \sum_{\ell=1}^{M} \phi(x_i/\eta_\ell)\, \phi(\eta_\ell/\xi_j), \tag{2.16}$$

where the summation is taken over the disturbance locations.

The Dual-Model Covariance Operators

Closely related to R and $\bar{R}$ above are the "dual" operators defined as

$$Q(\Theta) = B^*(\Theta)\Phi^*(\Theta)H^*(\Theta)H(\Theta)\Phi(\Theta)B(\Theta), \quad \bar{Q}(\Theta) = \Phi^*(\Theta)H^*(\Theta)H(\Theta)\Phi(\Theta). \tag{2.17}$$

Note that Q and $\bar{Q}$ can be obtained, from R and $\bar{R}$ respectively, by making the substitutions $\Phi \to \Phi^*$ and $B \to H^*$. This observation can be used as the basis for defining the dual system model, illustrated in Fig. 2.1, whose state and output have covariance operators specified by Q and $\bar{Q}$ above.

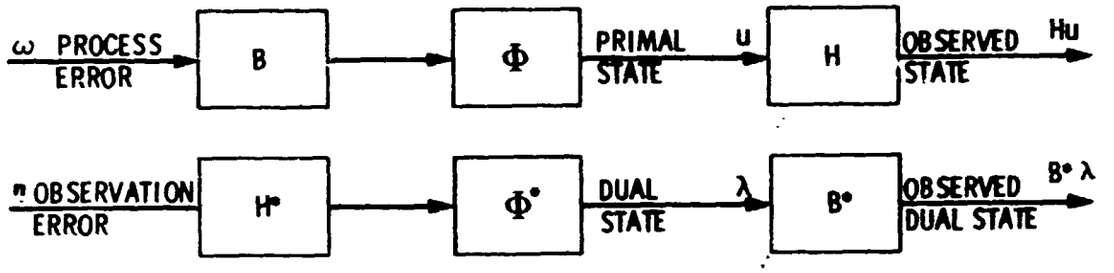

Figure 2.1. Illustration of Primal and Dual Models

The primal system model is based on (2.3), with ω, u and Hu denoting the process error, the system state, and the observed state respectively. For this model, $\bar{R} = E(um^*)$ is the covariance of the state, while $R = H\bar{R}H^*$ is the corresponding covariance of the observed state. It is assumed for the sake of this discussion that the deterministic input f in (1.1) has been set to zero, so that the suspected mean m in (2.3) is zero also. With this assumption, it is not necessary to show m in the block diagram in Fig. 2.1, and the relationship between the primal and dual models is illustrated more easily. The dual system model is characterized by the dual operators H^*, Φ^* and B^*, by the dual or adjoint state λ, and by the observed dual state $B^*\lambda$. It is assumed that the dual model is driven by a unit-covariance white-noise process so that $E(nn^*) = I$. This input process n driving the dual model can be thought of as being the observation error process in (2.3). For this dual model, $\bar{Q} = E(\lambda\lambda^*)$ and $Q = E(B^*\lambda\lambda^*B)$ are respectively the covariances of the state λ and the observed state $B^*\lambda$. Upon multiplication of λ in Fig. 2.1 by A^*, the following partial differential equations result to describe the dual model

$$A^*\lambda = H^*n. \tag{2.18}$$

Note that the dual state covariance $\bar{Q} = \Phi^*H^*H\Phi$ and the dual observed-state covariance $Q = B^*\Phi^*H^*H\Phi B$ are related by

$$Q = B^*\bar{Q}B. \tag{2.19}$$

In the same spirit used to arrive at (2.11)-(2.16), it is now possible to develop the following properties of the operator $\bar{Q}$ and its corresponding kernel q.

Remark 2.7 The dual-state covariance operator $\bar{Q}$ satisfies

$$A^*\bar{Q}A = H^*H, \tag{2.20}$$

a result which can be obtained from $\bar{Q} = \Phi^*H^*H\Phi$ in (2.17) upon multiplication by $A^*(\cdot)A$.

Remark 2.8 In terms of its kernel q, the operator $\bar{Q}$ can be expressed as

$$\bar{Q}v = \int_\Omega q(x/\xi)v(\xi)d\xi, \tag{2.21}$$

where q satisfies the differential equation

$$A_\xi q(\xi/x)A_x = h(\xi/x), \tag{2.22}$$

and where $h(\xi/x)$ is the kernel of H^*H. This result can be established by an approach quite similar to that used in arriving at (2.13). The symbol v denotes again an admissible function defined to be admissible if (2.21) makes sense.

Remark 2.9 The kernel $q(\xi/x)$ of $\bar{Q}$ can be expressed as

$$q(\xi/x) = \int_\Omega \int_\Omega \phi(\eta/\xi)\, h(\eta/\beta)\, \phi(x/\beta)\, d\eta\, d\beta, \tag{2.23}$$

where h is the kernel of H*H. This result can be established in a manner analogous to that used in arriving at (2.15).

Remark 2.10 In the special case most typical in applications, where the range of H and B is finite-dimensional, the $Q = B^*\Phi^*H^*H\Phi B$ is a matrix whose general element Q_{ij} can be expressed as

$$Q_{ij} = \sum_{k=1}^{N} q(\eta_k/\xi_i)\, q(x_j/\eta_k), \tag{2.24}$$

where the summation is taken over the set of sensor locations.

Spectral Representations

Recall that $R = H\Phi BB^*\Phi^*H^*$ is the observed-state covariance operator in (2.7). The eigenvalues of R are defined as the nontrivial solutions of

$$R\phi_k = \lambda_k^2\phi_k, \tag{2.25}$$

with ϕ_k being the corresponding eigenvectors. Note that, in cases where H has a finite-dimensional range, the operator R is an N-by-N matrix with a finite number of eigenvectors. In the more general case where the range of H is infinite-dimensional, then R is usually compact and $\lambda_k^2 \to 0$ as $k \to \infty$. In both of these cases, the following Mercer expansions hold for R and its kernel r

$$R = \sum \lambda_k^2 \phi_k \phi_k^* \quad \text{and} \quad r(x/\xi) = \sum \lambda_k^2 \phi_k(x)\phi_k^T(\xi) \tag{2.26}$$

Furthermore, the normalized eigenvectors ϕ_k form an orthonormal basis for the observation space H_3. This implies that $\Sigma\, \phi_k \phi_k^* = I$, where I is the identity in H_3.

Closely related to the basis ϕ_k above are the dual vectors ψ_k defined as

$$\psi_k = \lambda_k^{-1} B^*\Phi^*H^*\phi_k, \tag{2.27}$$

which can be viewed as the result of applying an input ϕ_k to the dual system model (2.18) and then "balancing" the output by dividing by the eigenvalue λ_k.

Remark 2.11 The vectors ψ_k defined by (2.27) are the eigenvectors of the dual observed-state covariance $Q = B^*\Phi^*H^*H\Phi B$, i.e.,

$$Q\psi_k = \lambda_k^2\psi_k. \tag{2.28}$$

This result can be established by premultiplication of (2.27) by $H\Phi B$ and use of the condition $R\phi_k = \lambda_k^2\phi_k$. Note that, if the dimension of the input space H_1 is greater than that of the output space H_3, then the ψ_k do not span the input space. They do, however, span the range subspace of the operator $B^*\Phi^*H^*$. Consequently, they cannot be used to expand vectors in the null space of $H\Phi B$.

Remark 2.12 The vectors ψ_k are also related to ϕ_k by the equation

$$\phi_k = \lambda_k^{-1} H\Phi B\psi_k, \tag{2.29}$$

a result that can be obtained from (2.27) upon premultiplication by the operator $H\Phi B$ and use of the condition $R\phi_k = \lambda_k^2\phi_k$.

Remark 2.13 The dual-state covariance operator Q and its corresponding kernel q can be expressed as

$$Q = \sum \lambda_k^2 \psi_k \psi_k^* \qquad q = \sum \lambda_k^2 \psi_k(x)\psi_k^T(\xi) \tag{2.30}$$

a set of equations which are analogous to (2.26). This result can be obtained from the observation that $Q = B^*\Phi^*H^*H\Phi B = B^*\Phi^*H^*\,(\Sigma\, \phi_k\phi_k^*)\, H\Phi B = \Sigma\, \lambda_k^2\, \psi_k\psi_k^*$. Use has been made of the condition $\Sigma\, \phi_k\phi_k^* = I$.

The vectors ϕ_k span the observation space H_3. While the vectors ψ_k do not span the input space H_1, they do span the range of $B^*\Phi^*H^*$. So far, no attempt has been made to obtain vectors that can be used to expand functions in the state space H_2 or in its dual space H_2^*. To this end, define

$$x_k = \lambda_k^{-1}\, \Phi B\psi_k, \quad p_k = \lambda_k^{-1}\, \Phi^*H^*\phi_k. \tag{2.31}$$

The vector x_k is in the state space whereas the adjoint variables p_k are in the dual space. In general, neither one of these two vectors however spans the state space H_2.

Remark 2.14 The vectors x_k and p_k are orthonormal with respect to H^*H and BB^* respectively, i.e.,

$$x_k^*H^*Hx_m = 0, \quad p_k^*BB^*p_m = 0, \qquad k \neq m, \tag{2.32}$$

$$x_k^*H^*Hx_k = 1, \quad p_k^*BB^*p_k = 1. \tag{2.33}$$

These results can be established by the following sequence of operations: $x_k^* H^* H x_m = (Hx_k)^* Hx_m = \phi_k^* \phi_m$ and $p_k^* BB^* p_m = (B^* p_k)^* B^* p_m = \psi_k^* \psi_m$. Since ϕ_k and ψ_k are orthonormal, then (2.32) and (2.33) follow.

Remark 2.15 The vectors x_k, ψ_k and p_k, x_k are related by

$$\phi_k = Hx_k \text{ and } \psi_k = B^* p_k. \tag{2.34}$$

This result follows readily from the definitions in (2.27), (2.29) and (2.31).

Remark 2.16 The vectors x_k and p_k satisfy the boundary-value problem:

$$\left[\begin{array}{cc} A & 0 \\ \hline 0 & A^* \end{array}\right] \left[\begin{array}{c} x_k \\ \hline p_k \end{array}\right] = \frac{1}{\lambda_k} \left[\begin{array}{cc} 0 & BB^* \\ \hline H^*H & 0 \end{array}\right] \left[\begin{array}{c} x_k \\ \hline p_k \end{array}\right]. \tag{2.35}$$

This result can be established by operating on x_k in (2.31) by A and on p_k by A^* to obtain

$$Ax_k = \lambda_k^{-1} B\psi_k \text{ and } A^* p_k = \lambda_k^{-1} H^* \phi_k. \tag{2.36}$$

Then, substitution of (2.34) in (2.36) implies (2.35)

Note the similarity between this problem and those traditionally encountered as necessary (and at times sufficient) conditions for optimality in quadratic optimal control and estimation problems for linear systems.

The Fredholm Resolvents of the Covariance Operators Q and R

The Fredholm resolvent of R is defined as that integral operator such that $(I+R)^{-1} = I-P$, a relationship which immediately implies that

$$R = P + RP \text{ and } R = P + PR \tag{2.37}$$

In terms of the corresponding kernels r and p, these equations become

$$r(x/\xi) = p(x/\xi) + \int_\Omega r(x/\eta)p(\eta/\xi)d\eta \tag{2.38}$$

for the case with continuously distributed data. In cases with discrete data, R and P are matrices whose general elements $R_{k,m}$ and $P_{k,m}$ are related by

$$R_{k,m} = P_{k,m} + \sum_{n=1}^{N} R_{k,n} P_{n,m}. \tag{2.39}$$

In both of these equations (2.38) and (2.39), the unknown is the Fredholm resolvent P, whereas the observed-state covariance kernel R is known.

Remark 2.17 The integral operator R and its Fredholm resolvent P commute. This result, which can be stated as

$$RP = PR, \tag{2.40}$$

is a direct consequence of (2.37).

Remark 2.18 Equations (2.37) also imply that

$$P = (I+R)^{-1}R = R(I+R)^{-1}, \qquad R = (I-P)^{-1}P = P(I-P)^{-1}. \tag{2.41}$$

Remark 2.19 In a manner analogous to (2.37) - (2.41), it is possible to define the resolvent S of the dual-state covariance operator Q by the relationship $(I+Q)^{-1} = I-S$ which implies

$$Q = S + QS, \; Q = S + SQ, \; SQ = QS, \tag{2.42}$$

and

$$S = (I+Q)^{-1}Q = Q(I+Q)^{-1}, \qquad Q = (I-S)^{-1}S = S(I-S)^{-1}. \tag{2.43}$$

Remark 2.20 The Fredholm resolvents P and S can be expressed as

$$P = \sum [\lambda_k^2/(1+\lambda_k^2)]\, \phi_k\phi_k^*, \qquad S = \sum [\lambda_k^2/(1+\lambda_k^2)]\psi_k\psi_k^*, \tag{2.44}$$

$$p(x/\xi) = \sum [\lambda_k^2/(1+\lambda_k^2)]\, \phi_k(x)\phi_k^T(\xi), \quad s(x/\xi) = \sum [\lambda_k^2/(1+\lambda_k^2)]\, \psi_k(x)\psi_k^T(\xi). \tag{2.45}$$

These expansions can be established by substituting (2.26) and (2.30) into (2.37) and (2.41).

Trigonometric Operator Forms

Remark 2.21 The predicted-data-covariance operator (I+R) can be expressed as

$$I + R = I + TAN^2\alpha = SEC^2\alpha, \tag{2.46}$$

where $TAN^2\alpha$ and $SEC^2\alpha$ are the operators

$$TAN^2\alpha = \sum \tan^2\alpha_k\phi_k\phi_k^* = R, \qquad SEC^2\alpha = \sum \sec^2\alpha_k\phi_k\phi_k^*, \tag{2.47}$$

and $\tan \alpha_k$ is defined by $\tan \alpha_k = \lambda_k$. Note also for later reference that

$$(I+R)^{1/2} = SEC\alpha = \sum \sec\alpha_k \phi_k \phi_k^*, \qquad R^{1/2} = TAN\alpha = \sum \tan\alpha_k \phi_k \phi_k^*. \tag{2.48}$$

Proof: Recall that $R = \Sigma \lambda_k^2 \phi_k \phi_k^* = \Sigma \tan^2\alpha_k \phi_k \phi_k^*$ thereby establishing (2.47). Use of the formal expression $I = \Sigma \phi_k \phi_k^*$ for the identity I implies that $(I+R) = \Sigma (1+\tan^2\alpha_k) \phi_k \phi_k^*$, which leads to (2.46). Equations (2.48) are obtained from (2.46) and (2.47) by performance of the square-root operation.

Remark 2.22 The Fredholm resolvent defined as $P = I - (I+R)^{-1}$ of the covariance operator R can be expressed as

$$P = SIN^2\alpha, \tag{2.49}$$

where $SIN^2\alpha$ is the operator defined by the expansion

$$SIN^2\alpha = \sum \sin^2\alpha_k \phi_k \phi_k^* = \sum [\lambda_k^2/(1+\lambda_k^2)] \phi_k \phi_k^*. \tag{2.50}$$

Proof: This result can be established by substitution of $\tan^2\alpha_k = \lambda_k^2$ in (2.44).

Remark 2.23 Equations (2.47) and (2.48) together imply that

$$P = R(I+R)^{-1} = TAN^2\alpha[I+TAN^2\alpha]^{-1} = TAN^2\alpha\,[SEC^2\alpha]^{-1} = SIN^2\alpha, \tag{2.51}$$

a trigonometric operator identity that can be viewed as a generalization of a similar identity involving scalars.

3. DERIVATION OF THE LIKELIHOOD FUNCTIONAL

Based on the results of the previous section, it is now possible to derive the likelihood functional in (1.3) to be minimized. Since the development required to achieve this is fairly lengthy, it is convenient to summarize in advance the pivotal steps involved in the derivation:

- the integral operator model $y = m + H\Theta B\omega + n$ in (2.3) is first coverted into an equivalent "spectral" form $y_k = m_k + \lambda_k \omega_k + n_k$, where $y_k = \phi_k^* y$, $\omega_k = \psi_k^* \omega$, $n_k = \phi_k^* n$ are the corresponding spectral coefficients.

- the spectral coefficients y_k of the data y are a sequence of independent Gaussian random variables with mean $E(y_k) = m_k$, covariance $\sigma_k^2 = 1 + \lambda_k^2$ and probability density $\rho_k(y_k;\Theta) = \pi^{-1/2}\sigma_k^{-1} \exp[-(y_k - m_k)^2/2\sigma_k^2]$

- a "finite-dimensional" likelihood ratio is then defined as the product of a finite number N of terms involving the probability densities $\rho_k(y_k;\theta)$ above.

- an "infinite-dimensional" likelihood ratio is obtained by letting the number N of spectral coefficients approach infinity. The related negative-log likelihood functional in (1.3) is obtained by taking the negative logarithm of the functional that results from the limiting process. Of course, in cases where the data is finite-dimensional (obtained by means of a finite number of discretely located measurements), the limiting process involved in this last step is not necessary. In this case, the "finite-dimensional" likelihood function obtained in the previous step is the function to be minimized to obtain the parameter estimates.

The remainde of this section contains a more detailed derivation of the foregoing results.

Recall that

$$y = m + H\Phi B\omega + n, \tag{3.1}$$

where $m = H\Phi Cf$ and f is the input. As outlined above, the first step toward evaluating the likelihood function is to convert (3.1) into an equivalent "spectral" form by using the eigenvectors ϕ_k and ψ_k, i.e.,

$$y = \sum y_k\phi_k, \quad \omega = \sum \omega_k\psi_k, \quad n = \sum n_k\phi_k, \quad m \sum m_k\phi_k. \tag{3.2}$$

Substitution of (3.2) and (3.1) and premultiplication of (3.1) by ϕ_k^* leads to

$$y_k = m_k + \lambda_k\omega_k + n_k. \tag{3.3}$$

Result 3.1 y_k, ω_k and n_k are independent Gaussian random variables with mean and covariance given by

MEAN	COVARIANCE
$E(\omega_k) = E(n_k) = 0$	$E(\bar{y}_k^2) = 1 + \lambda_k^2$
$E(y_k) = m_k$	$E(\omega_k^2) = E(n_k^2) = 1$
	$E(\bar{y}_k\bar{y}_m) = 0 \quad m \neq k$

where $\bar{y}_k = y_k - m_k$. Hence, y_k is a sequence of independent Gaussian random variables with mean m_k and covariance $1 + \lambda_k^2$.

Let $y^N = [y_1,\ldots,y_N]$ be an N-dimensional vector consisting of the first N spectral coefficients y_k of the data y. Because y_k are independent Gaussian random variables with mean m_k and covariance $\sigma_k^2 = 1 + \lambda_k^2$, their corresponding probability densities $\rho_k(y_k;\Theta) = \pi^{-1/2}\sigma_k^{-1}\exp(-\bar{y}_k^2/2\sigma_k^2)$ can be multiplied to obtain the probability density $\rho(y^N;\Theta)$ of the composite N-dimensional vector y^N, i.e.,

$$\rho(y^N;\Theta) = \prod_{k=1}^{N} \rho_k(y_k;\Theta) = \prod_{k=1}^{N} \sigma_k^{-1}\pi^{-1/2}\exp(-\bar{y}_k^2/2\sigma_k^2). \tag{3.4}$$

In order to obtain a likelihood functional for the identification problem with the function-space process y as the data, it would be desirable to let $N\to\infty$ and obtain what would be in the limit a probability density functional (PDF) for the process y. Unfortunately, this limit may not exist because the right side of (3.4) may not converge as $N\to\infty$, and consequently a PDF for the process y cannot be defined in this manner. However, this can be circumvented by dividing by

$$\rho_o(y^N;\Theta) = \prod_{k=1}^{N} \pi^{-1/2}\exp(-y_k^2/2). \tag{3.5}$$

This results in

$$\Lambda(y^N;\Theta) = \prod_{k=1}^{N} \frac{\exp[-(y_k-m_k)^2/2(1+\lambda_k^2)]}{(1+\lambda_k^2)^{1/2}\exp(-y_k^2/2)}, \tag{3.6}$$

which can be viewed as a likelihood ratio consisting of the PDF of the process y^N with the "signals" ω_k and m_k nonzero, divided by the similar PDF of y^N with the signals ω_k and m_k set to zero. The term likelihood ratio used to describe (3.6) is consistent with terminology common in the theory for detection of Gaussian signals in additive Gaussian noise [3].

Although the limits of $\rho(y^N;\Theta)$ and $\rho_o(y^N;\Theta)$ appearing respectively in (3.4) and (3.5) may not exist when taken independently, the limit of their ratio in (3.6) is a well-defined quantity specified by

$$\Lambda(y;\Theta) = \lim_{N\to\infty} \Lambda(y^N;\Theta) = \frac{\exp[(-\tfrac{1}{2})\,(y-m)^*(I+R)^{-1}(y-m)]}{[\det(I+R)]^{1/2}\exp[(-\tfrac{1}{2})y^*y]}. \tag{3.7}$$

This is the desired expression for the likelihood ratio that the maximum-likelihood method seeks to <u>maximize</u>. It can be interpreted as the likelihood ratio for the detection of the "signal" $m + H\Phi B\omega$ in (3.1), in the presence of the noisy Gaussian signal n. Instead of maximizing $\Lambda(y;\Theta)$ directly, it is more convenient to <u>minimize</u> the negative-log likelihood functional defined as $J(\Theta;y) = -\log[\Lambda(y;\Theta)]$, or, more explicitly,

$$J(\Theta;y) = \tfrac{1}{2}\log\det[I+R(\Theta)] + \tfrac{1}{2}[y-m(\Theta)]^*[I+R(\Theta)]^{-1}[y-m(\Theta)] - \tfrac{1}{2}y^*y. \tag{3.8}$$

Note that for the special case with no deterministic input, $m=m=0$ in (3.1), and the negative-log likelihood in (3.8) reduces to

$$J(\Theta;y) = \tfrac{1}{2}\log\det[I+R(\Theta)] - \tfrac{1}{2}y^*P(\Theta)y, \tag{3.9}$$

where $P(\Theta) = I - [I+R(\Theta)]^{-1}$ is the previously defined (in Sec. 2) Fredholm resolvent of the predicted-data-covariance operator R.

The first term in both of these last two equations can be cast into an equivalent and somewhat more convenient form by use of the identity [4]

$$\log\det[I+R(\Theta)] = \mathrm{Tr}\log[I+R(\Theta)]. \tag{3.10}$$

Substitution of (3.10) in (3.9) leads to

$$J(\Theta;y) = \tfrac{1}{2}\mathrm{Tr}\log[I+R(\Theta)] + \tfrac{1}{2}[y-m(\Theta)]^*[I+R(\Theta)]^{-1}[y-m(\Theta)] - \tfrac{1}{2}y^*y, \tag{3.11}$$

which has been recorded previously as (1.3) and constitutes the central aim of this section.

<u>Reorientation</u>

The method of maximum likelihood, as defined here, results in estimates that minimize $J(\Theta;y)$ in (3.11). This minimization problem can be viewed as a function-space nonlinear programming problem subject to the system model constraints that $R(\Theta) = H(\Theta)\Phi(\Theta)^*(\Theta)B^*(\Theta)\Phi^*(\Theta)H^*(\Theta)$ and $m(\Theta) = H(\Theta)\Phi(\Theta)C(\Theta)f$. Since no closed-form solution to this problem exists, it is necessary to use numerical methods for optimization. However, there exist alternative formulas for the likelihood ratio that are more convenient to use in the implementation of the numerical methods. Such formulas are developed in the following section.

4. ALTERNATIVE FORMULAS FOR THE LIKELIHOOD FUNCTIONAL

$\frac{1}{2}\,\mathrm{Tr}\log[I+R(\Theta)] + \frac{1}{2}[y-m(\Theta)]^*[I+R(\Theta)]^{-1}[y-m(\Theta)] - \frac{1}{2}y^*y$	BASIC
$\frac{1}{2}\,\mathrm{Tr}\log[I+R(\Theta)] + \frac{1}{2}[y-m(\Theta)]^*[y-Hu_0(\Theta)] - \frac{1}{2}y^*y$	SMOOTHING
$\frac{1}{2}\sum[\log(1+\lambda_k^2) + (1+\lambda_k^2)^{-1}(y_k-m_k)^2 - y_k^2]$	SPECTRAL
$\mathrm{Tr}\log[I+K(\Theta)] + \frac{1}{2}z^*(\Theta)z(\Theta) - z^*(\Theta)y$	SQUARE-ROOT FILTER
$\mathrm{Tr}\log[SEC\alpha(\Theta)] + \frac{1}{2}z^*(\Theta)z(\Theta) - z^*(\Theta)y$	TRIGONOMETRIC OPERATOR

In the above table, the basic formula is expressed in terms of the suspected mean m and covariance $I+R = I + H\Phi\Im B^*\Phi^*H^*$ of the data y. The smoothing form is specified in terms of the optimal smoothed estimate $u_0 = E(u/y)$, representing the conditional mean of the state u given the data y. The spectral formula is obtained by substitution in (1.3) of the eigensystem expansions $R = \Sigma\lambda_k^2\,\phi_k\phi_k^*$, $y = \Sigma y_k\phi_k$, and $m = \Sigma m_k\phi_k$, where λ_k^2 and ϕ_k are the eigenvalues and eigenvectors of the observed-state covariance operator R. The square-root filter formula, previously recorded in (1.4), is based on the factorization of the predicted-data-covariance operator as $(I+R) = (I+R)^{1/2}(I+R)^{1/2}$ and on the definitions $z = Ly + (I-L)m$ and $(I+K) = (I-L)^{-1} = (I+R)^{1/2}$. Finally, the trigonometric operator formula is obtained from the square-root filter expression by use of the identities $I+R = SEC^2\alpha$ and $L = I-COS\alpha$ developed in Sec. 2.

Although the derivation of the above expressions leads to significant insight about the structure of the likelihood functional, it is not within the scope of the paper to investigate all of these alternatives to the same level of detail. The formula involving the predicted-data-covariance square-root filter appears to be the most convenient to implement the numerical search for the optimal estimates. This section, however, aims to first develop the results summarized above.

Formulas Based on the Optimal Smoothed State Estimate

Result 4.1 The negative-log likelihood functional can be expressed as

$$J(\Theta;y) = \tfrac{1}{2}\,\mathrm{Tr}\log[I+R(\Theta)] + \tfrac{1}{2}[y-m(\Theta)]^*[y-Hu_0(\Theta)] - \tfrac{1}{2}y^*y, \tag{4.1}$$

$$u_c(\Theta) = G(\Theta)y + [I-G(\Theta)H]\Phi(\Theta)Cf, \tag{4.2}$$

$$G(\Theta) = \bar{R}(\Theta)H^*[I+H\bar{R}(\Theta)H^*]^{-1}, \tag{4.3}$$

where $u_0 = E(u/y)$ is the conditional expectation of the state u given the data y, and G is the estimator gain.

Proof: It will be shown in Sec. 8 that u_0 in (4.2) is the conditional mean and that G in (4.3) is the corresponding estimator gain. Therefore, for the sake of the discussion here, assume that (4.2) and (4.3) are valid. Multiply u_0 in (4.2) by H and use (4.3) in the resulting equation to obtain

$$Hu_0 = HGy + (I-HG)m, \tag{4.4}$$

and

$$y - Hu_0 = (I-HG)(y-m), \tag{4.5}$$

where $m = H\Phi Cf$ is as before the suspected mean of the data y. However, recall the identity $HG = H\bar{R}H^*(I+H\bar{R}H^*)^{-1} = I - (I+H\bar{R}H^*)^{-1}$ so that $I-HG = (I+H\bar{R}H^*)^{-1} = (I+R)^{-1}$. Hence, substitution of this last identity in (4.5) leads to

$$y - Hu_0 = (I+R)^{-1}(y-m). \tag{4.6}$$

This is the central result required to establish the equivalence between (4.1) and (3.11). To this end, substitute (4.6) into the second term on the right side of (4.1), and observe the equivalence with (3.11) by inspection.

Result 4.2 The negative-log likelihood functional can be expressed as

$$J(\Theta;y) = \tfrac{1}{2}\,\mathrm{Tr}\log[I+R(\Theta)] + \tfrac{1}{2}[E(u)]^*H^*H[E(u/y)] - \tfrac{1}{2}[E(u)+E(u/y)]^*H^*y \tag{4.7}$$

where $E(u) = \Phi Cf$ and E(u/y) are respectively the unconditional and conditional expected values of the state u.

Proof: This result can be established as a corollary to the Result 4.1 by combining the last two terms on the right side of (4.1) and use of the equation $m = H\Phi Cf$.

Both of these results express the likelihood functional in terms of a quantity u_0 in (4.2) which is the conditional expectation E(u/y) of the state given the data y. This quantity is also known to be the best linear mean-square estimate as well as the optimal least-squares estimate. The coincidence of the best mean-square estimate and the optimal least-squares estimate, both of which can be computed by the conditional expectation formula (4.2), is explored at length in Ref. [1].

Result 4.3 The negative-log likelihood functional can be expressed as

$$J(\Theta;y) = \tfrac{1}{2} \sum [\log(1+\lambda_k^2) + (1+\lambda_k^2)^{-1} (y_k-m_k)^2 - y_k^2], \tag{4.8}$$

where $y_k = \phi_k^* y$ and $m_k = \phi_k^* m$ are the spectral coefficients of the data and the suspected mean m, and λ_k^2 are the eigenvalues of R. By substitution of $\lambda_k = \tan\alpha_k$ in (4.8), this equation can be cast as

$$J(\Theta;y) = \tfrac{1}{2} \sum [\log (\sec^2\alpha_k) + \cos^2\alpha_k (y_k - m_k)^2 - y_k^2]. \tag{4.9}$$

Proof: Equation (4.8) can be established by taking the negative log of $\Lambda(y^N;\Theta)$ in (3.6) and letting $N\to\infty$. Use of the identity $\lambda_k = \tan\alpha_k$ in (4.8) leads to (4.9).

Result 4.4 The negative-log likelihood ratio can be expressed as

$$J(\Theta;y) = \sum [\tfrac{1}{2} \log (1+\lambda_k^2) + \tfrac{1}{2} z_k^2 - z_k y_k], \tag{4.10}$$

where

$$z_k = L_k y_k + (1-L_k)m_k \quad \text{and} \quad L_k = 1-\cos\alpha_k. \tag{4.11}$$

Proof: Define the "residual" process

$$e_k = y_k - z_k, \tag{4.12}$$

as the difference between the data y_k and the "filtered" estimate z_k. Observe that $e_k = \cos\alpha_k(y_k-m_k)$ by substitution of (4.11) in (4.12). Substitute this last equation into the second term on the right side of (4.9) to obtain (4.10).

The formula for the likelihood functional in (4.10) can be viewed as the "spectral" version of the predicted-data-covariance square-root formula described below.

Predicted-Data-Covariance Square-Root Formula for the Likelihood Functional

Result 4.5 The negative-log likelihood functional can be expressed as

$$J(\Theta;y) = \mathrm{Tr} \log [I+K(\Theta)] + \tfrac{1}{2} z^*(\Theta)z(\Theta) - z^*(\Theta)y, \tag{4.13}$$

where

$$z(\Theta) = L(\Theta)y + [I-L(\Theta)]m(\Theta), \tag{4.14}$$

with $L(\Theta)$ and $K(\Theta)$ defined as

$$L(\Theta) = I - [I+R(\Theta)]^{-1/2}, \qquad K(\Theta) = [I+R(\Theta)]^{1/2} - I. \tag{4.15}$$

Proof: Conversion of the first term (1/2) Tr log $[I+R(\Theta)]$ in (3.11) into Tr log $[I+K(\Theta)]$ follows because (1/2) Tr $\log(I+R)$ = (1/2) Tr log $(I+K)^2$ = Tr log $(I+K)$. Conversion of the last two terms on the right side of (3.11) into the desired form in (4.13) follows from the identity

$$(y-m)^*(I+R)^{-1}(y-m) = [(I-L)(y-m)]^*[(I-L)(y-m)] = (y-z)^*(y-z), \tag{4.16}$$

where use has been made of the fact that $(I+R)^{-1} = (I-L^*)(I-L)$.

Result 4.6 The operators L and K can be represented in terms of the following eigensystem expansions:

$$L = \sum (1-\cos\alpha_k)\,\phi_k\phi_k^* \qquad K = \sum (\sec\alpha_k - 1)\,\phi_k\phi_k^*, \tag{4.17}$$

where $\alpha_k = \tan^{-1}\lambda_k$, and ϕ_k are the eigenvectors of R.

Proof: Let

$$L = \sum L_k\phi_k\phi_k^* \text{ with } L_k = \phi_k^* L\phi_k, \tag{4.18}$$

and then evaluate the as yet undetermined coefficients L_k from $L = I - (I+R)^{-1/2}$ in (4.15). To this end, premultiply L in (4.15) by ϕ_k^* and postmultiply by ϕ_k to obtain $L_k = 1 - (1+\lambda_k^2)^{-1/2} = 1-\cos\alpha_k$, which is the desired result.

Similarly, to obtain the desired expansion for K, seek to determine the coefficients K_k in

$$K = \sum K_k\,\phi_k\phi_k^* \text{ with } K_k = \phi_k^* K\,\phi_k. \tag{4.19}$$

Multiplication of K in (4.15) by ϕ_k^* and ϕ_k leads to $K_k = \phi_k^* K\phi_k = (1+\lambda_k^2)^{1/2} - 1 = \sec\alpha_k - 1$.

Trigonometric Operator Formulas for the Likelihood Functional

Result 4.7 The log-likelihood functional can be expressed as

$$J(\Theta;y) = \text{Tr log } [SEC_Z(\Theta)] + \tfrac{1}{2}\, z^*(\Theta)z(\Theta) - z^*(\Theta)y, \tag{4.20}$$

where

$$z(\Theta) = [I - COS\alpha(\Theta)]y + COS\alpha(\Theta)m(\Theta), \tag{4.21}$$

$$\text{with } COS\alpha(\Theta) = [I+R(\Theta)]^{-1/2} = \sum \cos\alpha_k \phi_k \phi_k^*. \tag{4.22}$$

Proof: Recognize that (4.17) implies that

$$L(\Theta) = I - COS\alpha(\Theta) \quad \text{and} \quad K(\Theta) = SEC\alpha(\Theta) - I, \tag{4.23}$$

and use these identities in (4.13) and (4.14) to obtain (4.20) and (4.21) respectively.

Result 4.8 The negative-log likelihood functional can be expressed as

$$J(\Theta;y) = \sum [\log \sec\alpha_k(\Theta) + \tfrac{1}{2} z_k^2(\Theta) - z_k(\Theta)y_k(\Theta)], \tag{4.24}$$

where z_k and y_k are the "spectral" coefficients

$$z_k(\Theta) = \phi_k^*(\Theta)z(\Theta), \qquad y_k(\Theta) = \phi_k^*(\Theta)y, \tag{4.25}$$

and as before $\sigma_k = \tan^{-1}\lambda_k$, with λ_k^2 being the eigenvalues of R.

Proof: This result, which is closely related to Result 4.4 above, can be established by observing that $SEC\alpha(\Theta)$, $z(\Theta)$ and y in (4.20) can be expanded as

$$SEC\alpha(\Theta) = \sum \sec\alpha_k \phi_k \phi_k^*, \quad z = \sum z_k \phi_k \quad \text{and} \quad y = \sum y_k \phi_k. \tag{4.26}$$

Selection of Preferred Formula for Numerical Search Implementation

In principle, all of the above formulas for the likelihood functional $J(\Theta;y)$ can be used as a point of departure to compute the gradient $\partial J/\partial\Theta$ and the corresponding Hessian $\partial^2 J/\partial\Theta^2$ - and to thereby obtain the necessary ingredients to implement the Newton-Raphson search for optimization. The calculations involved in the numerical search can vary significantly, however, depending on which of the forms is used as a starting point. It is therefore of interest to conduct a detailed investigation of the relative advantages and disadvantages of the various methods to implement the search that arise from the various forms of the likelihood functional. Such an investigation is currently in progress and will be reported on in future work. In this paper however, the formula selected to compute the gradient and Hessian is that based on the predicted-data-covariance square-root filter in (4.13).

5. COVARIANCE EIGENSYSTEM SENSITIVITY TO SMALL PARAMETER CHANGES

As a preliminary to the evaluation of $\partial J/\partial\Theta$ and $\partial^2 J/\partial\Theta^2$ involved in the numerical search for optimization, it is necessary to conduct an analysis of the perturbations $\delta\lambda_k$ and $\delta\phi_k$ of the eigenvalues and eigenvectors of $R = H\Phi BB^*\Phi^*H^*$, with respect to

variations $\delta\Theta$ of the parameter distribution Θ. Such an analysis will provide the mathematical tools that will be used in subsequent sections to evaluate $\partial J/\partial\Theta$ and $\partial^2 J/\partial\Theta^2$.

By definition, λ_k^2 and ϕ_k are the nontrivial solutions of

$$R(\Theta)\phi_k(\Theta) = \lambda_k^2(\Theta)\phi_k(\Theta), \tag{5.1}$$

where the dependence on Θ of R, ϕ_k and λ_k has been explicit. The ultimate objective of this section is to develop analytical formulas for calculating the first-order perturbations $\delta\lambda_k$ and $\delta\phi_k$ of λ_k and ϕ_k with respect to small changes $\delta\Theta$ in the parameter distributions Θ.

Definition of $\delta\lambda_k$, $\partial\lambda_k/\partial\Theta$, $\delta\phi_k$ and $\partial\phi_k/\partial\Theta$

It is assumed here that the Frechet differential [2] of λ_k at Θ exists and that it can be computed by

$$\delta\lambda_k(\Theta;\delta\Theta) = [d\lambda_k(\Theta+\gamma\delta\Theta)/d\gamma]_{\gamma=0}, \tag{5.2}$$

where γ is a scalar and $\delta\Theta$ is an admissible perturbation of Θ. Equation (5.2) is actually the formula typically used for computation of the Gateaux differential. However, it is assumed here that both of these derivatives exist and coincide and that therefore (5.2) can be used to calculate the Frechet derivative.

Since λ_k is Frechet differentiable (admittedly by assumption, as an investigation of the technical conditions required for differentiability is not within the scope of this paper), its differential $\delta\lambda_k(\Theta;\delta\Theta)$ can be expressed as

$$\delta\lambda_k(\Theta;\delta\Theta) = [\partial\lambda_k(\Theta)/\partial\Theta]\delta\Theta, \tag{5.3}$$

where $\partial\lambda_k(\Theta)/\partial\Theta$ is a bounded linear functional referred to as the Frechet derivative of λ_k at Θ. The transformation $\partial\lambda_k/\partial\Theta$ can also be viewed as a function space gradient of λ_k at Θ. Similarly, the eigenvector differential $\delta\phi_k(\Theta;\delta\Theta)$ is defined as

$$\delta\phi_k(\Theta;\delta\Theta) = [\partial\phi_k(\Theta;\delta\Theta)/\partial\Theta]\,\delta\Theta. \tag{5.4}$$

where $[\partial\phi_k(\Theta)/\partial\Theta]$ is the Frechet derivative, assumed to be linear and bounded.

Calculation of $\delta\lambda_k$ and $\partial\lambda_k/\partial\Theta$

Recall that the ϕ_k in (5.1) are orthonormal so that

$$\phi_k^*\phi_k = 1 \quad \text{and} \quad \phi_k^*\phi_m = 0, \qquad m \neq k. \tag{5.5}$$

Multiplication of (5.1) by ϕ_k^* and use of $\phi_k^*\phi_k = 1$ leads to

$$\lambda_k^2 = \phi_k^* R \phi_k, \tag{5.6}$$

which can be taken as the point of departure for calculation of $\delta\lambda_k$ and $\partial\lambda_k/\partial\Theta$.

Result 5.1 The Frechet differential $\delta\lambda_k(\Theta;\delta\Theta)$ can be expressed as

$$\delta\lambda_k(\Theta;\delta\Theta) = -\lambda_k^2(\Theta)\,[p_k^*(\Theta)\delta A(\Theta;\delta\Theta)x_k(\Theta)], \tag{5.7}$$

where $\delta A(\Theta;\delta\Theta)$ is the differential of A defined as

$$\delta A(\Theta;\delta\Theta) = [dA(\Theta+\gamma\delta\Theta)/d\gamma]_{\gamma=0} \tag{5.8}$$

and p_k and x_k are the vectors defined as $p_k = \lambda_k^{-1}\Phi^*H^*\phi_k$ and $x_k = \lambda_k^{-1}\Phi B\psi_k$ in Sec. 2.

Proof: Performance of a first-order perturbation on (5.6), and use of the condition $\phi_k^*\delta\phi_k = 0$ leads to

$$\delta\lambda_k = (2\lambda_k)^{-1}\,\phi_k^*\delta R\phi_k, \tag{5.9}$$

where $\delta R(\Theta;\delta\Theta) = [dR(\Theta+\gamma\delta\Theta)/d\gamma]$ evaluated at $\gamma = 0$. However, since $\delta R = \delta(H\Phi BB^*\delta\Phi^*)H^*$, then

$$\delta R = H(\delta\Phi)BB^*\Phi^*H^* + H\Phi BB^*(\delta\Phi^*)H^*. \tag{5.10}$$

It can be observed from (5.10) that evaluation of $\delta\Phi$ is the central calculation required to determine δR. In order to simplify notation, without loss of generality, it has been assumed in arriving at (5.10) that B and H do not depend on Θ. In most practical cases, this assumption is satisfied because the poorly known parameters occur in the operator A.

To compute $\delta\Phi$, as required by (5.10), recall that $A(\Theta)\Phi(\Theta) = I$, so that $(\delta A)\Phi + A(\delta\Phi) = 0$, and

$$\delta\Phi = -\Phi(\delta A)\Phi. \tag{5.11}$$

Substitution of (5.11) in (5.10) leads to

$$\delta R = -H\Phi(\delta A)\Phi BB^*\Phi^*H^* - H\Phi BB^*\Phi^*(\delta A)^*\Phi^*H^*. \tag{5.12}$$

Multiplication by $\phi_k^*(\cdot)\phi_k$ results in

$$\phi_k^* \delta R \phi_k = -(\phi_k^* H \Phi)\delta A(\Phi B B^* \Phi^* H^* \phi_k) - (\phi_k^* H \Phi B B^* \Phi^*)\delta A^*(\Phi^* H^* \phi_k). \qquad (5.13)$$

Finally, use of the definitions $p_k = \lambda_k^{-1} \Phi^* H^* \phi_k$ and $x_k = \lambda_k^{-1} \Phi B \psi_k$ in (5.13), and substitution in (5.9), implies (5.7). In performing this last step, it has been assumed that $A = A^*$ is formally self-adjoint, a condition that is valid on most problems of practical interest.

Discussion and Additional Assumptions on A

The above result, although a step in the right direction, is still somewhat intermediate because the differential $\delta\lambda_k$ in (5.7) is expressed in terms of the yet to be determined differential δA. To proceed further, it is convenient to make two additional assumptions (typically satisfied in practice):

- $A(\Theta)$ is linear in Θ so that $A(\Theta_1 + \Theta_2) = A(\Theta_1) + A(\Theta)_2$ for two admissible distributions Θ_1 and Θ_2.
- $A(\Theta)$ can be factored as $A(\Theta) = D^*(\Theta)D$, where D and its corresponding formal adjoint D^* may in general be matrix differential operators.

Based on these assumptions, it is now possible to derive the following more explicit formulas for $\delta\lambda_k$ and $\partial\lambda_k/\partial\Theta$.

Result 5.2 The Frechet derivative $\partial\lambda_k(\Theta)/\partial\Theta$ of λ_k is

$$\partial\lambda_k(\Theta)/\partial\Theta = \lambda_k^2 \, Dp_k(\Theta) \cdot Dx_k(\Theta). \qquad (5.14)$$

Proof: Since A has been assumed to be linear and factorizable

$$\delta\lambda_k = -\lambda_k^2 p_k^* D^*(\delta\Theta) D x_k = \lambda_k^2 \langle (Dp_k)\, \delta\Theta\, (Dx_k) \rangle_2 \qquad (5.15)$$

where the last equality is a consequence of a process analogous to integration by parts.

Result 5.3 Since $\partial\lambda_k/\partial\Theta$ has been assumed to be a bounded linear functional in X, it must be expressible as

$$[\partial\lambda_k(\Theta)/\partial\Theta]\delta\Theta = \langle \partial\lambda_k(\Theta;\cdot)/\partial\Theta, \delta\Theta \rangle_X, \qquad (5.16)$$

where $\partial\lambda_k(\Theta;\cdot)/\partial\Theta$ is an element of $X^*(\Omega)$. Futhermore, $[\partial\lambda_k(\Theta;\cdot)/\partial\Theta]$ can be evaluated from

$$\partial\lambda_k(\Theta;x)/\partial\Theta = \lambda_k^2 \, Dp_k(\Theta;x) \cdot Dx_k(\Theta;x). \qquad (5.17)$$

Proof: The rigorous derivation of this result is not as yet available. The result is accepted somewhat formally on the basis that a bounded linear functional can be represented by an element in the dual to the space in which the functional is defined.

Calculation of $\delta\phi_k$ and $\partial\phi_k/\partial\theta$

Result 5.4 The Frechet differential $\delta\phi_k(\theta;\delta\theta)$ of ϕ_k can be expressed as

$$\delta\phi_k(\theta;\delta\theta) = \sum_{m\neq k} [(\phi_m^* \delta R \phi_k)/(\lambda_k^2 - \lambda_m^2)]\,\phi_m, \tag{5.18}$$

where δR is the differential of the observed-state covariance operator R.

Proof: Since $R\phi_k = \lambda_k^2\phi_k$,

$$(\delta R)\phi_k + R\delta\phi_k = 2\lambda_k(\delta\lambda_k)\phi_k + \lambda_k^2\delta\phi_k. \tag{5.19}$$

Now, seek an expansion for $\delta\phi_k$ in terms of the orthonormal basis ϕ_m, i.e.,

$$\delta\phi_k = \sum_{m\neq k} c_{km}\phi_m, \qquad c_{km} = \phi_m^*\delta\phi_k, \tag{5.20}$$

where c_{km} are scalar coefficients to be determined. Note that the orthonormality of ϕ_k implies that $c_{kk} = 0$, so that $\delta\phi_k$ does not have a component in the direction of ϕ_k. To evaluate c_{km}, premultiply (5.20) by ϕ_m^* to obtain

$$\phi_m^*\delta R\phi_m + \phi_m^* R\delta\phi_k = \lambda_k^2\phi_m^*\delta\phi_k. \tag{5.21}$$

Use of the conditions $\phi_m^* R = \lambda_m^2\phi_m^*$ and $c_{km} = \phi_m^*\delta\phi_k$ and rearrangement of terms leads to

$$c_{km} = (\phi_m^*\delta R\phi_k)/(\lambda_k^2 - \lambda_m^2). \tag{5.22}$$

Substitution of (5.22) in (5.20) leads to (5.18), thereby establishing the result.

Equation (5.18) is similar in nature to (5.9) in that it expresses the desired differential in terms of the yet to be determined quantity δR.

Result 5.5 The Frechet differential $\delta\phi_k(\theta;\delta\theta)$ of ϕ_k can be expressed as

$$\delta\phi_k(\theta;\delta\theta) = \sum_{m\neq k} [\lambda_k\lambda_m/(\lambda_k^2-\lambda_m^2)]\,[\lambda_k p_m^*(\delta A)x_k + \lambda_m x_m^*(\delta A)^* p_k]\phi_m, \tag{5.23}$$

where $p_k = \lambda_k^{-1}\Phi^*H^*\phi_k$ and $x_k = \lambda_k^{-1}\Phi B\psi_k$.

Proof: Substitute (5.12) in (5.18) and use the definitions for p_k and x_k.

Equation (5.23) is valid without making the additional assumption that $A(\Theta)$ is linear in Θ and factorizable as $A(\Theta) = D^*(\Theta)D$. If these two assumptions are now made, the following result can be obtained.

Result 5.6 The Frechet derivative $\partial\phi_k(\Theta)/\partial\Theta$ is specified by

$$\partial\phi_k(\Theta)/\partial\Theta = \sum_{m\neq k} [\lambda_k\lambda_m/(\lambda_k^2-\lambda_m^2)]\,[\lambda_k Dp_m \cdot Dx_k + \lambda_m Dx_m \cdot Dp_k]\phi_m. \tag{5.24}$$

Proof: This result follows by substitution of $\delta A(\Theta) = D^*(\delta\Theta)D$ in (5.23).

Closely related to $\delta\phi_k$ is the differential

$$\delta(\phi_k\phi_k^*) = \phi_k\delta\phi_k^* + (\delta\phi_k)\phi_k^* \tag{5.25}$$

of the outer product $\phi_k\phi_k^*$. The corresponding Frechet derivative $\partial(\phi_k\phi_k^*)/\partial\Theta$ is evaluated in the following result.

Result 5.7 The Frechet derivative $[\partial(\phi_k\phi_k^*)/\partial\Theta]$ is specified by

$$\partial(\phi_k\phi_k^*)/\partial\Theta = \sum_{m\neq k} [\lambda_k\lambda_m/(\lambda_k^2-\lambda_m^2)]\,[\lambda_k Dp_m \cdot Dx_k + \lambda_m Dx_m \cdot Dp_k]\,[\phi_m\phi_k^* + \phi_k\phi_m^*]. \tag{5.26}$$

Proof: Use (5.24) to evaluate the right side of (5.25) and recall that $\delta(\phi_k\phi_k^*) = [\partial(\phi_k\phi_k^*)/\partial\Theta]\delta\Theta$.

Discussion

The results obtained above provide the key tools required to evaluate the function-space gradient $\partial J/\partial\Theta$ and Hessian $\partial^2 J/\partial\Theta^2$ of the likelihood functional. The most useful formulas are (5.17) for the derivative $\partial\lambda_k/\partial\Theta$ of the eigenvalue λ_k, (5.24) for the derivative $\partial\phi_k/\partial\Theta$ of the eigenvector ϕ_k, and (5.26) for the derivative $\partial(\phi_k\phi_k^*)/\partial\Theta$ of the outer product $(\phi_k\phi_k^*)$. These formulas will be used repeatedly in the following section.

6. SPECTRAL REPRESENTATIONS FOR THE GRADIENT, APPROXIMATE HESSIAN, AND NEWTON-RAPHSON SEARCH

Implementation of the modified Newton-Raphson search for the optimal parameter estimates requires calculation of the gradient $\partial J/\partial\Theta$ and of an approximation to the Hessian operator $\partial^2 J/\partial\Theta^2$. These calculations are best achieved using the predicted-data-covariance square-root filter in Result 4.5 that expresses the likelihood functional as

$$J(\Theta;y) = \mathrm{Tr}\log[I+K(\Theta)] + \tfrac{1}{2}\, z^*(\Theta)z(\Theta) - z^*(\Theta)y, \tag{6.1}$$

where $z(\Theta) = L(\Theta)y + [I-L(\Theta)]\, m(\Theta)$. Function space differentiation of (6.1) with respect to Θ leads to the gradient functional

$$g(\Theta;y) = \partial J(\Theta;y)/\partial\Theta = \mathrm{Tr}[(\partial L/\partial\Theta)(I+K)] + (z-y)^*(\partial z/\partial\Theta), \tag{6.2}$$

and to the approximate Hessian operator

$$M(\Theta;y) = \mathrm{Tr}[(\partial L/\partial\Theta)(I+R)(\partial L/\partial\Theta)] + (\partial z/\partial\Theta)^*(\partial z/\partial\Theta), \tag{6.3}$$

upon which the Newton-Raphson numerical search is to be based. An updated estimate $\Theta^{n+1} = \Theta^n - \delta\Theta^n$ is obtained by specification of the parameter change $\delta\Theta^n$ defined as

$$\delta\Theta^n = M^{-1}(\Theta^n;y)g(\Theta^n;y). \tag{6.4}$$

The main objective of this section is to replace the operator equations (6.2) and (6.3) with a set of equivalent matrix equations more convenient for calculations. The fundamental approach to be used consists of representing the function space derivatives $\partial L/\partial\Theta$, $\partial m/\partial\Theta$ and $\partial z/\partial\Theta$ - which have only been derived in terms of operator symbols in (6.2) and (6.3) - in terms of a specific orthonormal basis defined by the eigenvectors ϕ_k of the observed-state covariance operator R.

Spectral Representation for the Gradient

Result 6.1 The Frechet derivative $\partial L/\partial\Theta$ of the predicted-data-covariance square-root filter L can be represented as

$$\partial L/\partial\Theta = \sum_k \sum_m a_{km}\phi_k\phi_m^*, \tag{6.5}$$

where the spectral coefficients $a_{km} = \phi_k^*(\partial L/\partial\Theta)\phi_m$ are specified by

$$a_{kk} = \sin^2\alpha_k Dp_k \cdot Dx_k, \tag{6.6}$$

$$a_{km} = [\lambda_k\lambda_m/(\lambda_m^2-\lambda_k^2)][\cos\alpha_k - \cos\alpha_m][\lambda_k Dp_m \cdot Dx_k + \lambda_m Dx_m \cdot Dp_k] \quad k\neq m. \tag{6.7}$$

Note that a_{km} defines a matrix whose diagonal elements are provided by (6.6) and whose corresponding nondiagonal elements are given by (6.7).

Proof: Observe $L = \sum (1-\cos\alpha_i)\phi_i\phi_i^*$ implies

$$\partial L/\partial\theta = \sum \{\sin\alpha_i (\partial\alpha_i/\partial\theta)\phi_i\phi_i^* - \cos\alpha_i [\partial(\phi_i\phi_i^*)/\partial\theta]\}. \tag{6.8}$$

Substitution of this equation in $a_{km} = \phi_k^*(\partial L/\partial\theta)\phi_m$ and use of orthonormality of ϕ_k lead to

$$a_{kk} = \sin\alpha_k \cos^2\alpha_k (\partial\lambda_k/\partial\theta),\ a_{km} = -\cos\alpha_m\phi_k^*(\partial\phi_m/\partial\theta) - \cos\alpha_k(\partial\phi_k^*/\partial\theta)\phi_m, \tag{6.9}$$

where $\partial\lambda_k/\partial\theta$ and $\partial\phi_k/\partial\theta$ are the function-space derivatives evaluated in (5.18) and (5.25). Substitution of these two equations from Sec. 5 in (6.9) leads to (6.6) and (6.7) thereby establishing the result.

Result 6.2 The Frechet derivative $\partial m/\partial\theta$ of the suspected mean $m(\theta)$ is represented by

$$\partial m/\partial\theta = \sum (\partial m/\partial\theta)_k\phi_k, \tag{6.10}$$

with the spectral coefficients $(\partial m/\partial\theta)_k$ specified by

$$(\partial m/\partial\theta)_k = \lambda_k (Dp_k \cdot D\Phi Cf), \tag{6.11}$$

and ΦCf in (6.11) denoting the suspected value of the state u.

Proof: Since $m = H\Phi Cf$, then $\delta m = H\delta\Phi Cf = -H\Phi A(\delta\theta)\Phi Cf$, where the last equality follows from the condition $\delta\Phi = -\Phi A(\delta\theta)\Phi$. Define now $(\delta m)_k$ as the k^{th} spectral coefficient of δm, i.e.,

$$(\delta m)_k = \phi_k^*\,\delta m = -\phi_k^*\,H\Phi A(\delta\theta)\Phi Cf = -\lambda_k p_k^* A(\delta\theta)\Phi Cf, \tag{6.12}$$

where as before $p_k = \lambda_k^{-1}\Phi^*H^*\phi_k$. Use of the identity $p_k^*A(\delta\theta)\Phi Cf = -Dp_k \cdot D(\Phi Cf)\delta\theta$ in (6.12) results in $(\delta m)_k = (\partial m/\partial\theta)_k\,\delta\epsilon$, with $(\partial m/\partial\theta)_k$ given by (6.11).

Result 6.3 In the special case in which the deterministic input f is assumed to be a vector $f = [f_1,\ldots,f_M]$ of M inputs applied at the discrete locations ξ_i, an alternative to (6.11) in evaluating $(\partial m/\partial\theta)_k$ is

$$(\partial m/\partial\theta)_k = \sum_{m=1}^{M} \lambda_k\ Dp_k(x) \cdot D\phi(x/\xi_m)f_m \qquad \text{for } k = 1,\ldots,N, \tag{6.13}$$

where $\phi(x/\xi)$ is the Green's function of the system model operator A.

Result 6.4 The gradient $\partial z/\partial\theta = (\partial L/\partial\theta)y + (I-L)(\partial m/\partial\theta)$ of the filtered state estimate z can be represented as

$$\partial z/\partial\theta = \sum (\partial Z/\partial\theta)_k \phi_k, \tag{6.14}$$

where the spectral coefficients $(\partial z/\partial\theta)_k = \phi_k^* (\partial z/\partial\theta)$ are given by

$$(\partial z/\partial\theta)_k = \sum_{m=1}^{N} a_{km}(x)\,\bar{y}_m + \sum_{m=1}^{M} b_{km}(x) f_m, \tag{6.15}$$

with a_{km} specified in (6.6) and (6.7) and

$$b_{km}(x) = \sin\alpha_k\, Dp_k(x) \cdot D\phi(x/\xi_m). \tag{6.16}$$

Proof: Substitute $\partial L/\partial\theta$ and $\partial m/\partial\theta$ from (6.5) and (6.10) into $\partial z/\partial\theta = (\partial L/\partial\theta)\bar{y} + (I-L)(\partial m/\partial\theta)$ and then compute the spectral coefficients $(\partial z/\partial\theta)_k$ in (6.14) from $(\partial z/\partial\theta)_k = \phi_k^*(\partial z/\partial\theta)$.

Result 6.5 The gradient $g(\theta;y)$ in (6.2) can be represented as

$$g(\theta;y) = \sum [\sin^2\alpha_k \tan\alpha_k\, (Dp_k \cdot Dx_k) - e_k (\partial z/\partial\theta)_k], \tag{6.17}$$

where $e_k = \phi_k^* e$ are the spectral coefficients of the residual process $e = y - z$, and $(\partial Z/\partial\theta)_k$ are given by (6.15).

Proof: Substitute $\partial L/\partial\theta$ in (6.5), $\partial z/\partial\theta$ in (6.14), $e = \Sigma e_k \phi_k$ and $I + K = \Sigma \sec\alpha_k \phi_k \phi_k^*$ into (6.2) and use orthonormality of ϕ_k.

Equation (6.17) provides the means to evaluate the likelihood functional gradient, one of the key ingredients of the Newton-Raphson iteration. The approximate Hessian operator $M(\theta;y)$, which is the other major element required to implement the search, is evaluated below.

Evaluation and Inversion of the Approximate Hessian

Result 6.6 The approximate Hessian $M(\theta;y)$ in (6.3) is an integral operator whose kernel $M(x/\xi)$ is specified by

$$M(x/\xi) = \sum [\sec^2\alpha_k\, a_{kk}(x)\, a_{kk}(\xi) + z'_k(x) z'_k(\xi)], \tag{6.18}$$

where $z'_k = (\partial z/\partial\Theta)_k = \phi_k^*(\partial z/\partial\Theta)$ is the k^{th} spectral coefficient of $\partial z/\partial\Theta$.

Proof: Substitute (2.26) and (6.14) into (6.3) and use the orthonormality of ϕ_k.

Implementation of an iteration step in the Newton-Raphson search requires calculation of $\delta\Theta^n = M^{-1}(\Theta^n;y)g(\Theta^n;y)$, representing the incremental change in the parameter estimate. Inversion of $M(\Theta^n;y)$ is therefore required at every step of the search. This inversion is achieved by solving an integral equation as outlined in the following result.

Result 6.7 The incremental parameter change $\delta\Theta^n$ can be computed as the solution of the following integral equation

$$\int_\Omega M_n(x/\xi)\,\delta\Theta^n(\xi)d\xi = g_n(x), \tag{6.19}$$

where M_n is the approximate Hessian kernel in (6.18), and $g_n(x)$ is the value of the gradient at the spatial location x. The subscript n in M_n and g_n denotes that the corresponding quantities are evaluated at the n^{th} parameter estimate $\Theta=\Theta^n$.

Proof: Observe that $\delta\Theta^n = M_n^{-1}g_n$ implies $M_n\delta\Theta^n = g_n$, and express this last equation in terms of the kernel M_n to obtain (6.19).

7. PARAMETER ESTIMATION ERROR, CRAMER-RAO BOUNDS AND OPTIMAL INPUT DESIGN

The objectives here are to obtain a C-R bound for the covariance of the parameter estimation error and to begin an investigation of the problem of optimal input design by using the C-R bound as a criterion for optimal input selection.

Recall that the covariance of an unbiased estimate Θ satisfies the inequality

$$E(\Theta_p\Theta_p^*) \geq M^{-1}(\Theta_o), \tag{7.1}$$

where $M(\Theta_o)$ is the information operator defined as

$$M(\Theta_o) = E[\partial^2 J/\partial\Theta^2]_{\Theta=\Theta_o} = E[(\partial J/\partial\Theta)(\partial J/\partial\Theta)^*]_{\Theta=\Theta_o}. \tag{7.2}$$

The corresponding mean-square estimation error $E(\Theta_p^*\Theta_p)$ satisfies the related inequality

$$E(\Theta_p^*\Theta_p) \geq Tr[M^{-1}(\Theta_o)] \tag{7.3}$$

It can be observed that the key calculation required to obtain the C-R bound is the computation of $E[\partial^2 J/\partial\Theta^2]$ as outlined below.

Cramer-Rao Bound for the Estimation Error

Result 7.1 The information operator $M(\Theta_o)$ is specified by

$$M(\Theta_o) = E[\partial^2 J/\partial\Theta^2]_{\Theta=\Theta_o} = 2\,\mathrm{Tr}[(\partial L/\partial\Theta)(I+R)(\partial L^*/\partial\Theta)] + (\partial m^*/\partial\Theta)(I-L^*)(I-L)(\partial m/\partial\Theta), \tag{7.4}$$

where $R = H\Phi EE^*\Phi^*H^*$ is the data-covariance operator, $(\partial L/\partial\Theta)$ is the derivative of $L = I - (I+R)^{-1/2}$, and $(\partial m/\partial\Theta)$ is the derivative of the data mean $m=H\Phi Cf$.

Proof: Differentiate $g(\Theta;y)$ in (6.2) to obtain

$$\partial^2 J/\partial\Theta^2 = \mathrm{Tr}[(\partial^2 L/(\partial\Theta^2)(I+K) + (\partial L/\partial\Theta)(\partial K/\partial\Theta)] + (z-y)^*(\partial z/\partial\Theta) + (\partial z/\partial\Theta)^*(\partial z/\partial\Theta). \tag{7.5}$$

Take the expected value in (7.5) above, evaluate at $\Theta = \Theta_o$, and simplify to obtain

$$E[\partial^2 J/\partial\Theta^2]_{\Theta=\Theta_o} = \mathrm{Tr}\,[(\partial L/\partial\Theta)(I+R)(\partial L/\partial\Theta)^*] + E[(\partial z^*/\partial\Theta)(\partial z/\partial\Theta)]. \tag{7.6}$$

Finally, use (7.4) in (7.6) to arrive at (7.4).

Result 7.2 In spectral form, the information operator $M(\Theta_o)$ is specified by

$$M(\Theta_o) = \sum [2\sec^2\alpha_k\, a_{kk}(x)\, a_{kk}(\xi) + \cos^2\alpha_k m'_k(x) m'_k(\xi)], \tag{7.7}$$

where a_{kk} and $m'_k = (\partial m/\partial\Theta)_k$ are defined in (6.9) and (6.11) respectively.

Proof: Use an approach similar to that used to arrive at (6.18).

Inspection of (7.4) reveals that the information operator $M(\Theta_o)$ consists of the sum of two terms both of which are positive definite. In the first term, the data-covariance operator (I+R) appears as a "weighting" that is multiplied by the sensitivity filter $\partial L/\partial\Theta$. Note parenthetically that in fact L is self-adjoint so that $L = L^*$. The second term, on the other hand, will be shown to be a quadratic function of the input f.

Result 7.3 Assume that $f = [f_1,\ldots,f_M]$ is a vector of M inputs applied at the M discrete locations ξ_m. The information operator $M(\Theta_o)$ is an integral operator whose kernel $M(x/\xi)$ can be expressed as

$$M(x/\xi) = U(x/\xi) + f^T V(x/\xi) f, \tag{7.8}$$

where

$$U(x/\xi) = \sum_k \sin^4\alpha_k \tan^2\alpha_k \, [Dp_k(x) \cdot \Gamma p_k(x)] \, [Dp_k(\xi) \cdot Dx_k(\xi)], \tag{7.9}$$

$$V(x/\xi) = \sum_k \sin^2\alpha_k \, b_k(x) \, b_k^T(\xi), \tag{7.10}$$

and where $b_k^T(\xi)$ is the M-dimensional vector

$$b_k^T(\xi) = [Dp_k(\xi) \cdot D\phi(\xi/\xi_1), \ldots, Dp_k(\xi) \cdot D\phi(\xi/\xi_M)], \tag{7.11}$$

with ϕ being the Green's function of A in (1.1).

Proof: Substitute the eigensystem expansions for R in (2.26), for L in (4.17), for $(\partial L/\partial\Theta)$ in (6.5), and for $\partial m/\partial\Theta$ in (6.10) into (7.4) to obtain (7.9) and (7.10).

The second term in (7.8) is a quadratic form in the input signal f. This property can be used as a basis for optimal input design.

Optimal Input Design

The information operator can be used to state criteria for optimal input design. While several possible criteria exist, the one that is easiest to use is perhaps the maximization of Tr $M(\Theta_o)$:

$$\max \text{Tr } M(\Theta_o) = U + f^T V f, \quad f^T f = 1, \tag{7.12}$$

where

$$U = \int_\Omega U(x/x)dx \text{ and } V = \int_\Omega V(x/x)dx. \tag{7.13}$$

The optimal input f_o, which is the solution to the above optimization problem, is the eigenvector corresponding to the largest eigenvalue of the M-by-M matrix V.

Other criteria for optimal input selection include: minimization of Tr (M^{-1}), which would correspond to minimizing the Cramer-Rao bound; and minimization of $\lambda_{max}(M^{-1})$, where λ_{max} is the maximum eigenvalue of M^{-1}. While these last two criteria could be superior to (7.12), they both have the disadvantage of requiring inversion of the operator $M(\ _o)$. However, the requirement for such an inversion may not be a serious additional drawback because a similar calculation is required to implement the Newton-Raphson search outlined in the previous sections.

Vanishing Bias of the Gradient

Closely related to the above analysis is an investigation of the bias in the parameter estimate Θ. The central result is as follows.

Result 7.4 The expected value of the gradient functional $g(\Theta;y)$ vanishes at $\Theta = \Theta_o$, i.e.,

$$E[g(\Theta;y)]\,|_{\Theta=\Theta_o} = 0. \tag{7.14}$$

Proof: Observe that $\partial z/\partial\Theta = (\partial L/\partial\Theta)\bar{y} + (I-L)(\partial m/\partial\Theta)$, and recall that $\bar{y} = (I+K)e$. Substitute this in (6.2) and take the expected value. Finally, use the whiteness of the residual process, to be established in (8.46).

8. FILTERING, SMOOTHING AND THE RESIDUAL PROCESS

The central aim of this section is to conduct an analysis of the smoothed estimate u_o and of the filtered state estimate z_o that emerges from the predicted-data-covariance square-root filter. This analysis leads to the following major results:

- The smoothed estimate u_o is optimal in a conditional mean sense.

- The formulas that generate u_o and z_o have a predictor-corrector structure in which the final state estimate is the sum of: a prediction term-based on application of known inputs to the system model; and a correction term based on the difference between the actual and predicted data. The key element in these formulas is an estimator gain that provides the relative weighting between the two terms.

- The covariance of the state estimation error inherent in both estimates can be evaluated by means of equations which, if written in operator notation, resemble those encountered in filtering and smoothing for linear dynamical systems.

- Investigation of a residual process associated with the filtered state estimate z_o that has properties nearly identical to those of an innovations process: the residuals are a white noise process with a unit covariance; the residuals and the measurements can be obtained from each other by means of reciprocal linear transformations. Because these transformations are not causal, the residuals are not a bona fide innovations process. However, they are as useful in deriving filtering, smoothing and identification solutions for elliptic systems as the innovations process is in deriving similar solutions for linear dynamical systems.

- Development of relationships between the filtering and smoothing estimates that can be thought of as extensions to elliptic systems of the forward/backward sweep method for solution of filtering and smoothing problems in linear dynamical systems.

- Development of spectral representations for the predicted-data-covariance square-root filter and the optimal smoother in terms of the eigensystem of the state covariance $R = \Phi BB^*\Phi^*$. This leads to simple ways to implement filtering and smoothing solutions on a computer.

Smoothed and Filtered Estimates

The smoothed and filtered state estimates u_0 and z_0 have been defined in (1.17) as

$$u_0 = \Phi Cf + G(y-H\Phi Cf), \qquad z_0 = \Phi Cf + g(y-H\Phi Cf), \tag{8.1}$$

where G and g are Kalman-like gains specified by

$$G = \sum \sin^2\alpha_k x_k \phi_k^*, \quad g = \sum (1-\cos\alpha_k) x_k \phi_k^*. \tag{8.2}$$

The estimate u_0 is referred to as a smoothed estimate because it is the minimum-variance estimate of the state given the entire data set. This is established by the following result.

Result 8.1 The smoothed estimate u_0 in (8.1) is the conditional mean $u_0 = E(u/y)$ of the state given the data. Furthermore, the estimator gain G in (8.2) can be expressed alternatively as

$$G = \bar{R}H^*(I + H\bar{R}H^*)^{-1}, \tag{8.3}$$

in terms of the state covariance $\bar{R} = \Phi BB^*\Phi^*$.

Proof: Recall the general formula

$$E(u/v) = E(uv^*)\,[E(vv^*)]^{-1}v \tag{8.4}$$

derived in [4] for the conditional expected value of a zero-mean random process u given the related zero-mean random process v. Note that this formula requires calculation of the "cross-covariance" operator $E(uv^*)$ and the auto-covariance operator $E(vv^*)$. Define now the mean-centered state $\bar{u} = u-\Phi Cf = \Phi B\omega$ and the mean-centered data $\bar{y} = H\bar{u}+n$. By this definition, $\bar{u}$ and $\bar{y}$ are zero-mean. Therefore (8.4) can be used directly to compute $\bar{u}_0 = E(\bar{u}/\bar{y})$, i.e.,

$$\bar{u}_0 = E(\bar{u}/\bar{y}) = E(\bar{u}\bar{y}^*)\,[E(\bar{y}\bar{y}^*)]^{-1}\bar{y}, \tag{8.5}$$

which indicates that to evaluate $\bar{u}_o$, it is necessary to first evaluate the covariance operators $E(\bar{u}\bar{y}^*)$ and $E(\bar{y}\bar{y}^*)$. These calculations are: $E(\bar{u}\bar{y}^*) = E(\Phi B\omega\omega^* B^*\Phi^*) = \Phi BB^*\Phi^*$ and $E(\bar{y}\bar{y}^*) = E[(H\bar{u}+n)(H\bar{u}+n)^*] = I + H\bar{R}H^*$. Use of this in (8.5) leads to

$$E(\bar{u}/\bar{y}) = G\bar{y}. \tag{8.6}$$

This together with the definition of $\bar{u}$ and $\bar{y}$ in terms of u and y implies (8.1). The equivalence between the two different expressions for G in (8.2) and (8.3) is established by use of the spectral expansions in Sec. 2. In particular, use expansions (2.46) - (2.47) for I + R and the definition for x_k in (2.31).

As established by this result, the estimate u_o has a very well defined probabilistic interpretation. It is not presently known if the filtered estimate z_o has a similar interpretation. Nonetheless, this estimate plays a very significant role in the filtering, smoothing and identification methodology for elliptic systems under development here. Its role is analogous to that of the filtered estimate emerging from a Kalman filter in the case of dynamical systems. This is further investigated below.

Predictor-Corrector Structure

To examine this structure, consider the equation for u_o in (8.1) and illustrated in Fig. 8.1. Use of the deterministic input $f^{[1]}$ and the system model $\Phi C^{[2]}$ leads to a predicted estimate [3]. The difference process $y-H\Phi Cf$ [4] is then formed and operated on by the estimator gain G [5] to obtain the correction term $G(y-H\Phi Cf)$ [6]. Firally, the correction term is added to the predicted estimate to obtain the optimal estimate u_o. The equation for the filtered estimate z_o in (8.1) also has a predictor-corrector structure. The key difference between the two equations in (8.1) is that the estimator gains are different. A relationship between these two different gains G and g is explored later in this section.

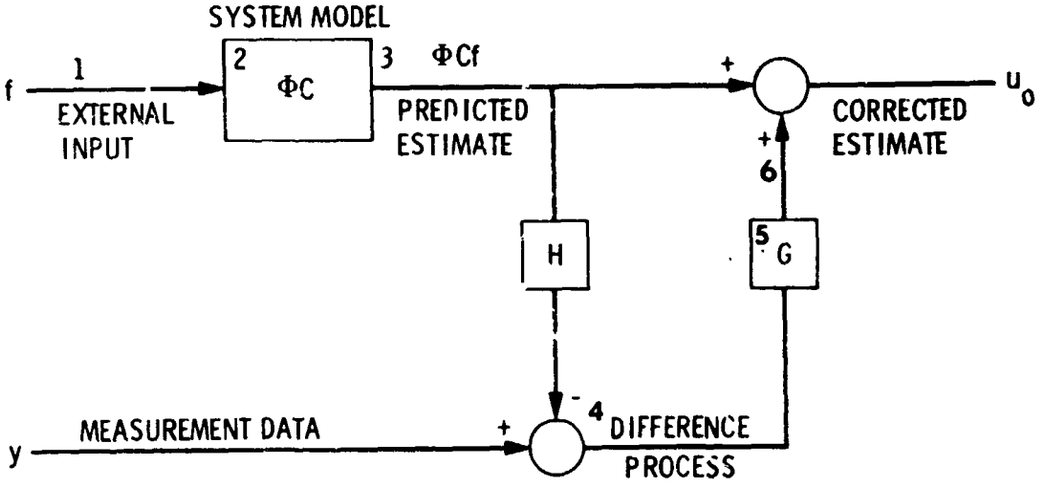

Fig. 8.1 Predictor-Corrector Form of the Smoothed State Estimator

Estimation Error Covariance and Kalman-like Gains: Smoothing

Since u_o and z_o are only estimates of the actual state u, it is of interest to investigate the inherent estimation error $u_p = u-u_o$ and $z_p = u-z_o$. In particular, the aim is to determine the estimation error covariance, under the assumption that the actual model errors ω and n in (1.1) and (1.2) are white-noise processes.

Result 8.2 The covariance $\bar{P} = \bar{P}^* = E(u_p u_p^*)$ of the state estimation error $u_p = u-u_o$ is specified by the following alternative formulas:

$$\bar{P} = (I - GH)\bar{R}(I - GH)^* + GG^*, \tag{8.7}$$

$$\bar{P} = \bar{R} - \bar{R}H^* (I + H\bar{R}H^*)^{-1} H\bar{R}, \tag{8.8}$$

$$\bar{P} = (I - GH)\bar{R} = \bar{R}(I - GH)^*, \tag{8.9}$$

$$\bar{P} = \Phi B (I + B^*\Phi^*H^*H\Phi B)^{-1}B^*\Phi^*. \tag{8.10}$$

Proof: To show (8.7), observe that $u = \Phi Cf + \Phi B\omega$ and $u_o = \Phi Cf + G(y - H\Phi Cf)$ imply that $u_p = u - u_o$ is

$$u_p = (I - GH)\Phi B\omega - Gn. \tag{8.11}$$

Hence $E(u_p u_p^*) = E[(I-GH)\Phi B\omega\omega^*B^*\Phi^* (I-GH)^* + Gnn^*G^*] = (I-GH)\bar{R}(I-GH)^* + GG^*$, where use has been made of the fact that $\epsilon = [\omega,n]$ is a white-noise process with covariance $E(\epsilon\epsilon^*) = I$. To show (8.8), observe that (8.7) implies

$$\bar{P} = \bar{R} - GH\bar{R} - \bar{R}H^*G^* + G(+ H\bar{R}H^*)G^*. \tag{8.12}$$

Substitution of $G = \bar{R}H^* (I + H\bar{R}H^*)^{-1}$ in (8.12) leads to (8.8). To show (8.9), observe that (8.8) can be expressed as $\bar{P} = \bar{R} (I-GH)^* = (I-GH)\bar{R}$ by using $G = \bar{R}H^*(I + H\bar{R}H^*)^{-1}$ in the last two terms of (8.12). To establish (8.10), substitute $\bar{R} = \Phi BB^*\Phi^*$ in (8.8) and use the identities $B^*\Phi^*H^* (I+H\Phi BB^*\Phi^*H^*)^{-1}H\Phi B = (I+B^*\Phi^*H^*H\Phi B)^{-1}B^*\Phi^*H^*H\Phi B = I - (I+B^*\Phi^*H^*H\Phi B)^{-1}$.

Result 8.3 The operator $H\bar{P}H^*$ is the Fredholm resolvent of $H\bar{R}H^*$ so that

$$(I + H\bar{R}H^*)^{-1} = I-H\bar{P}H^*. \tag{8.13}$$

Proof: Compute $H\bar{P}H^*$ from $\bar{P}$ in (7.8) to obtain $H\bar{P}H^* = H\bar{R}H^* [I - (I+H\bar{R}H^*)^{-1}]$. Use the identity $(I+H\bar{R}H^*)^{-1}H\bar{R}H^* = I - (I+H\bar{R}H^*)^{-1}$ twice in this last equation to obtain (8.13).

The aim now is to use (8.13) in (8.2) to obtain an alternative expression for the estimator gain.

Result 8.4 The estimator gain $G = \bar{R}H^*(I+H\bar{R}H^*)^{-1}$ can also be expressed as

$$G = \bar{P}H^*, \tag{8.14}$$

where $\bar{P} = E(u_p u_p^*)$ is the covariance of the smoothed state estimation error u_p.

Proof: Recall that $\bar{P}H^* = (I - GH)\bar{R}H^* = \bar{R}H^*[I - (I + H\bar{R}H^*)^{-1}H\bar{R}H^*]$. Since $(I + H\bar{R}H^*)^{-1}H\bar{R}H = I - (I + H\bar{R}H^*)^{-1}$, then $\bar{P}H^* = \bar{R}H^*(I + H\bar{R}H^*)^{-1} = G$.

Result 8.5 The mean-square smoothed state estimation error is given by

$$E(u_p^* u_p) = \mathrm{Tr}\,[\bar{P}]. \tag{8.15}$$

Proof: This follows from the definition of $\bar{P}$ as $\bar{P} = E(u_p u_p^*)$.

Note that many of the above formulas are very similar in form to the ones traditionally encountered in Kalman filtering for dynamical systems. For instance, Eqs. (8.3) and (8.14) are very similar to those used to compute the gain G for a Kalman filter in which $\bar{R}$ and $\bar{P}$ are the covariances of the estimation error associated with the predicted and corrected state estimates. Note also that (8.8) implies that $\bar{P}$ is always smaller than $\bar{R}$, which implies that the covariance of the estimation error after the observation y has been accounted for is smaller than the error covariance before the estimate correction occurs.

Estimation Error Covariance and Kalman-like Gains: Filtering

The aim here is to obtain results similar to results (8.2) - (8.5) above, but that are applicable to the filtered estimate z_o.

Result 8.6 The covariance $E(z_p z_p^*)$ of the filtered state estimation error $z_p = u - z_o$ is given by

$$E(z_p z_p^*) = (I-gH)\,\bar{R}(I-gH)^* + gg^*, \tag{8.16}$$

where $\bar{R} = \Phi BB^*\Phi^*$ is the state covariance, and g is the filter gain in (8.2).

Proof: Note that $u = \Phi Cf + \Phi B\omega$. This and (8.1) imply that

$$z_p = (I-gH)\,\Phi B\omega - gn, \tag{8.17}$$

where use has been made of $y - H\Phi Cf = H\Phi B\omega + n$ in (8.1). Calculation of $E(z_p z_p^*)$, ..., (8.17) and the conditions $E(\omega\omega^*) = I$ and $E(nn^*) = I$, leads to (8.16).

This result applicable to the filtered estimate is analogous to (8.7) of the smoothed estimates. To obtain results that are analogous to (8.8) - (8.10) requires, however, a few preliminary definitions and results. The need for these preliminaries arises from the ultimate desire to find a spectral decomposition for the state covariance $\bar{R} = \Phi BB^*\Phi^*$. It is straightforward to obtain the spectral representation for the

observed-state covariance $H\bar{R}H^*$. However, finding a similar decomposition of $\bar{R}$ is not as simple. The primary reason for this lack of simplicity is that the vectors $\psi_j = \lambda_j^{-1} B^*\Phi^*H^*\phi_j$ may not necessarily span the entire space H. This is particularly true in cases in which the dimension of the input space H_1 is greater than the dimension of the observation space H_3. In order to consider this case, assume that the operator $H\Phi B$ has finite-dimensional range. This corresponds to the situation where there are only a finite number N of sensors and the observed-state covariance $R = H\Phi BB^*\Phi^*H^*$ is an N-by-N matrix. Assume also that the input space is either infinite-dimensional or finite-dimensional with dimension M greater than N. This second assumption corresponds to cases where the uncertainty is distributed at M discrete locations or throughout the entire spatial domain Ω.

Result 8.7 The identity operator I mapping H_1 into itself can be decomposed as

$$I = I_o + I_\perp, \tag{8.18}$$

where

$$I_o = I - B^*\Phi^*H^*R^{-1}H\Phi B, \qquad I_\perp = B^*\Phi^*H^*R^{-1}H\Phi B, \tag{8.19}$$

and $R = H\Phi BB^*\Phi^*H^*$ is the observed-state covariance. In addition, I_o is in the null-space of the operator

$$R(\cdot) = H\Phi B(\cdot)B^*\Phi^*H^*, \tag{8.20}$$

mapping the space of bounded linear transformations on $H_1 \times H_1$ into the space of N-by-N matrices. Furthermore, I_o and $I_\perp$ are orthogonal complements so that

$$I_o^* I_\perp = \mathrm{Tr}\,[I_o I_\perp] = 0. \tag{8.21}$$

Proof: This result and its corresponding proof are illustrated graphically in Fig. 8.2. Eq (8.18) follows from (8.19). Substitution of I_o in (8.19) into (8.20) shows that $R(I_o) = 0$ so that I_o is in the null space of $R(\cdot)$. That I_o and $I_\perp$ are orthogonal complements follows from substitution of (8.19) in (8.21) by calculation of $\mathrm{Tr}\,[I_o I_\perp]$ using (8.19).

Result 8.8 The identity operator I mapping H_1 into itself can be expressed as

$$I = I_o + \sum_{j=1}^{N} \psi_j \psi_j^* \tag{8.22}$$

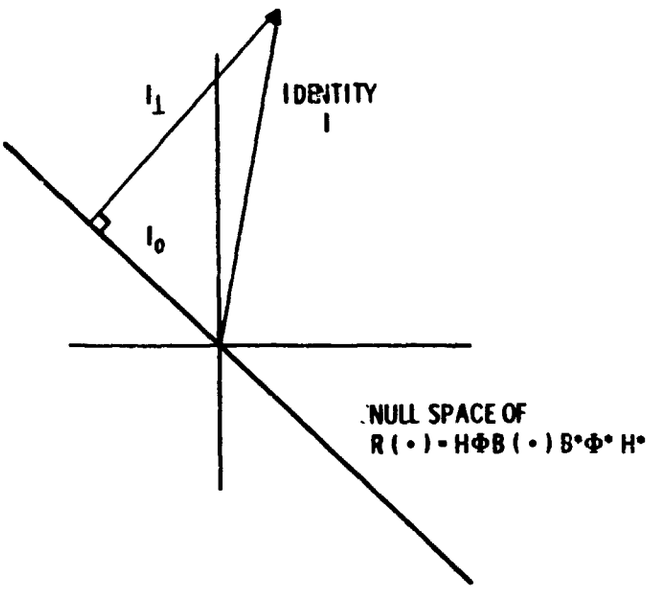

Fig. 8.2 Orthogonal Complement Decomposition of the Identity in $H_1 \times H_1$.

Proof: Substitute $R = \lambda_j^2 \phi_j \phi_j^*$ into $I_\perp$ in (8.19) and use $\psi_j = \lambda_j^{-1} B^* \Phi^* H^* \phi_j$.

The above result simply reflects the fact that the ψ_j do not span H_1, because (by assumption) there are only a finite number of them, and this number is smaller than the dimension of the input space.

Result 8.9 The state covariance $\bar{R} = \Phi B B^* \Phi^*$ can be decomposed as

$$\bar{R} = \bar{R}_0 + \sum_{j=1}^{N} \lambda_j^2 x_j x_j^*, \tag{8.23}$$

where

$$\bar{R}_0 = \Phi B I_0 B^* \Phi^* \tag{8.24}$$

Furthermore,

$$H\bar{R}_0 H^* = 0, \; H\bar{R}_0 = 0, \; \bar{R}_0 H^* = 0. \tag{8.25}$$

Proof: To show (8.23), substitute I from (8.22) into $\bar{R}(I) = \Phi B(I) B^* \Phi^*$, and use $x_j = \lambda_j^{-1} \Phi B \psi_j$. To show (8.25), substitute I_0 from (8.19) into (8.24) and (8.25).

Result 8.10 The dual state covariance $\bar{Q} = \Phi^* H^* H \Phi$ can be expressed as

$$\bar{Q} = \sum_{j=1}^{N} \lambda_j^2 \, p_j p_j^* \tag{8.26}$$

where $p_j = \lambda_j^{-1} \Phi^* H^* \phi_j$.

Proof: Since the ϕ_j span the observation space $H_3 = R^N$, then

$$I^N = \sum_{j=1}^{N} \phi_j \phi_j^* \tag{8.27}$$

where I^N denotes the identity in $R^N \times R^N$. To obtain (8.26), substitute (8.27) in $Q = \Phi^* H^* I^N H \Phi$ and use definition of p_j.

Define now the quantities

$$r = (1/2)\,\bar{R}_o + \sum_{j=1}^{N} (\sec\alpha_j - 1)\, x_j x_j^*, \quad q = \sum_{j=1}^{N} (\sec\alpha_j - 1)\, p_j p_j^* \tag{8.28}$$

and note the following key identities.

Result 8.11 The state covariance $\bar{R} = \Phi B B^* \Phi^*$ and r defined in (8.28) are related by

$$\bar{R} = r + r^* + r H^* H r. \tag{8.29}$$

Furthermore,

$$I + H \bar{R} H^* = (I + H r H^*)\,(I + H r^* H^*). \tag{8.30}$$

Proof: Substitute r from (8.28) and $\bar{R}$ from (8.23) into (8.29). Use the orthonormality of x_j with respect to H^*H. This establishes (8.29). Equation (8.30) follows from (8.29) by forming $I + H\bar{R}H^*$ from (8.29) and rearranging terms.

Result 8.12 The dual state covariance $\bar{Q}$ in (8.26) and q in (8.28) satisfy the identity

$$\bar{Q} = q + q^* + q B B^* q^*. \tag{8.31}$$

Furthermore,

$$I + B^* \bar{Q} B = (I + B^* q B)\,(I + B^* q B). \tag{8.32}$$

Proof: Substitute $\bar{Q}$ in (3.26) and q in (8.28) into (8.31). Use the orthonormality of p_j with respect to BB*. This establishes (8.31). To establish (8.32), form $I + B^*\bar{Q}B$ using (8.31, and rearrange terms in the resulting equation.

These are the preliminary results needed to evaluate the covariance of the estimation error associated with the filtered state estimate z_o.

Result 8.13 The filter gain g defined in (8.2) can be expresssed alternatively as

$$g = rH^* (I+HrH^*)^{-1}, \tag{8.33}$$

where r is defined in (8.28).

Proof: Substitute r from (8.28) into (8.33) and use $\bar{R}_o H^* = 0$ and $\phi_j^* = x_j^* H^*$. This recovers g in (8.2).

Note the similarity between (8.3) and (8.33). The equation in (8.3) expresses the smoother gain G in terms of the state covariance $\bar{R} = \Phi BB^*\Phi$. Eq. (8.33) is a similar equation for the filter gain in terms of r. The operator $\bar{R}$ in G can be interpreted as the state covariance. No similar probalistic interpretation for r is known. However, its introduction is very useful because it allows development of formulas for the estimation error covariance and for the filter gain that very closely resemble those obtained for the smoothing solutions.

Result 8.14 The covariance $E(z_p z_p^*)$ of the filtered state estimation error $z_p = u - z_o$ is

$$E(z_p z_p^*) = p + p^*, \tag{8.34}$$

where $p = p^*$ is specified by the alternative formulas

$$p = (I-gH)r(I-gH)^* + gg^* \tag{8.35}$$

$$p = r - rH^* (I+HrH^*)^{-1} Hr, \tag{8.36}$$

$$p = (I-gH)r = r(I-gH)^*, \tag{8.37}$$

$$p = (1/2)\bar{R}_o + \sum (1-\cos\alpha_j) x_j x_j^*. \tag{8.38}$$

Proof: To establish (8.34) and (8.35), substitute (8.29) in (8.16) and use the identity

$$(I-gH)\, rH^* = rH^* (I + HrH^*)^{-1} = g. \tag{8.39}$$

To establish (8.36), observe that (8.35) implies that

$$p = r - gHr - rH^*g^* + g(I + HrH^*)g^*. \tag{8.40}$$

Substitute $g = rH^* (I + HrH^*)^{-1}$ in (8.40) to obtain (8.36). To obtain (8.37) observe that the second term of (8.36) can be expressed alternatively as gHr and rH^*g^*. To obtain (8.38), substitute r in (8.28) into (8.36) and use orthonormality of ϕ_j.

Result 8.15 The mean-square estimation error associated with the filtered state estimate is given by

$$E(z_p^* z_p) = \mathrm{Tr}\,[p + p^*] = \mathrm{Tr}\,[\bar{R}_o] + 2 \sum_{j=1}^{N} (1-\cos\alpha_j)\, x_j^* x_j. \tag{8.41}$$

Proof: This result follows from (8.34) and (8.38).

Result 8.16 The filter gain g can be expressed as

$$g = pH^*, \tag{8.42}$$

where p is related to the filtered state estimation error covariance by $E(z_p z_p^*) = p + p^*$.

Proof: Since $g = rH^* (I+HrH^*)^{-1}$, then $g = rH^* (I-Hg)^* = rH^* (I-g^*H^*) = r (I-H^*g^*)H^* = r (I-gH)^*H^* = pH^*$.

This equation is analogous to (8.14) in that it expresses an estimator gain in terms of the covariance of the state estimation error.

Result 8.17 The operators I + HrH* and I - HpH* are reciprocal, i. e.,

$$(I + HrH^*)^{-1} = I - HpH^*. \tag{8.43}$$

Proof: Recall $(I + HrH^*)^{-1} = I-Hg = I-HpH^*$, where the last equality holds because $g = pH^*$.

Note that this result implies that the operator HpH* is the Fredholm resolvent of HrH*. The identity also immediately implies whiteness of the residuals process as investigated in more detail below.

Pseudo-Innovations Properties of the Residuals

Define the residual process in the usual way, as the difference between the actual measurements and the predicted data emerging from the predicted-data-covariance square-root filter, i.e.,

$$e = y - Hz_o \tag{8.44}$$

This process turns out to have two key properties that are nearly identical to those of an innovations process: the residuals are white with a unit covariance; the residuals

and the measurements can be obtained from each other by means of reciprocal relationships. These two properties are established in the following results.

Result 8.18 The residual process defined in (8.44) is white with a unit covariance, i.e.,

$$E(ee^*) = I. \tag{8.45}$$

Proof: Observe from (8.1) that $e = (I-Hg)(y-H\Phi cf)$. Hence, $E(ee^*) = (I-Hg)(I+H\bar{R}H^*)(I-Hg)^* = I$. This last equality follows from $E[(y-H\Phi cf)(y-H\Phi Cf)^*] = I + H\bar{R}H^*$ and from (8.42) and (8.43).

Result 8.19 The residuals $e = y-Hz_o$ and the mean-centered measurement process $\bar{y} = y-H\Phi Cf$ can be obtained from each other by means of reciprocal linear transformations, i.e.,

$$e = (I - HpH^*)\bar{y}, \qquad \bar{y} = (I+HrH^*)e \tag{8.46}$$

where

$$(I + HrH^*)^{-1} = (I-HpH^*). \tag{8.47}$$

Proof: Eq. (8.47) has been established in (8.43) and is restated here only to emphasize its relationship to the properties of the residual process. Eq. (8.1) implies $e = (I-Hg)\bar{y}$.

Relationships Between Filtered and Smoothed Estimates

While the smoothed and filtered estimates have been defined somewhat independently of each other, these estimates are in fact very closely related. It is possible to obtain one in terms of the other, as outlined in the following result.

Result 8.20 The smoothed and filtered estimates u_o and z_o are related by

$$u_o = z_o + ge, \tag{8.48}$$

where

$$e = y-Hz_o \tag{8.49}$$

is the residual process, and g is the predicted-data-covariance square-root filter gain.

Proof: Observe that (8.1) and (8.3) imply $u_o = \Phi Cf + \bar{R}H^*(I + H\bar{R}H^*)^{-1}(y-H\Phi Cf)$. Use of (8.46) leads to $u_o = \Phi Cf + \bar{R}H^*(I+HrH^*)^{-1}e$. Similarly, z_o in (8.1) and r in (8.33) lead to $z_o = \Phi Cf + rH^*e$. Hence, $u_o-z_o = [\bar{R}H^*(I+HrH^*)^{-1} - rH^*]e$. Use of the identity (8.29) in this implies that $u_o-z_o = ge$, which is the desired result. Note that (8.48) can be written in the alternative form

$$u_o = (I-gH)z_o + gy. \tag{8.50}$$

Closely related to the above relationship between filtered and smoothed state estimates is a relationship between the corresponding covariances of the state estimation errors. This is developed below.

Result 8.21 The filtered state estimation error $z_p = u-z_o$ and the residual process $e = y-Hz_o$ are related by

$$e = n + Hz_p, \tag{8.51}$$

where n is the measurement error.

Proof: Note that $e = y-Hz_o = Hu+n-Hz_o = H(u-z_o) + n = Hz_p + n$

Result 8.22 The covariance $\bar{P} = E(u_p u_p^*)$ of the smoothed state estimation error $u_p = u-u_o$ can be expressed as

$$\bar{P} = p + p^* - pH^*Hp, \tag{8.52}$$

where $p + p^* = E(z_p z_p^*)$ is the covariance of the filtered state estimation error $z_p = u-z_o$. Furthermore,

$$I-H\bar{P}H^* = (I-HpH^*)(I-HpH^*). \tag{8.53}$$

Proof: Use (8.52) to obtain

$$E(ez_p^*) = E(nz_p^*) + HE(z_p z_p^*) = E(nz_p^*) + H(p+p^*). \tag{8.54}$$

Now use (8.17) to compute $E(nz_p^*)$, i.e.,

$$E(nz_p^*) = -g^*, \tag{8.55}$$

since $E(n\omega^*) = 0$ by assumption. Substitution of (8.55) in (8.54) and use of $g = pH^*$ leads to

$$E(ez_p^*) = g^*. \tag{8.56}$$

Since $u_p = u-u_o$, then $u_p = z_p - ge$ from (8.48). Hence,

$$E(u_p u_p^*) = E(z_p z_p^*) - g\,E(ez_p^*) - E(z_p e^*)g^* + gE(ee^*)g^*. \tag{8.57}$$

Now use (8.34), (8.43), (8.45) and (8.56) to obtain (8.52). Equation (8.53) follows immediately from (8.52) by forming $I - H\bar{P}H^*$ and rearranging terms in the resulting expression. Note that (8.52) implies that the gains G and g are related by

$$G = g + g - gHg. \tag{8.58}$$

The last three results can be viewed as a generalization to elliptic systems of relationships encountered in filtering and smoothing for dynamical systems. For example, Equation (8.48) is a generalization to elliptic systems of the forward/backward sweep method for solution of two-point boundary-value problems. This method in general terms states that the smoothed states estimates can be obtained as a result of a two-stage process: forward filtering by means of a Kalman filter to obtain a filtered state estimate and a residual process; and backward smoothing to process the residuals and obtain a smoothed state estimate. This two-stage data processing approach has been extensively studied for linear dynamical systems. Eqs. (8.48) and (8.49) have exactly the same structure. This structure is illustrated in Fig. 8.3.

The overall diagram illustrates how the data $y^{[1]}$ and the deterministic input $f^{[2]}$ are processed to arrive at a smoothed estimate $u_o^{[3]}$. The estimation process consists of two stages: a FILTERING stage that results in a filtered estimate $z_o^{[4]}$ and a residual process[5]. This filtering stage is characterized by a predictor-corrector structure where a predicted estimate[6] is first produced and then corrected by a correction term.[7] The results of the filtering stage are then processed by the SMOOTHING stage. Central to both the filtering and smoothing stages is the gain $g^{[8]}$. The foregoing structure is nearly identical to that of the forward/backward sweep method in linear dynamical systems. There are, however, some key differences. One of the differences is that the filtering stage in the case of dynamical systems is based on the Kalman filter, whereas in the elliptic case under consideration here, this filter is replaced by the predicted-data-covariance square-root filter. Another key difference is that the Kalman filter is causal whereas the predicted-data covariance square-root filter is not, i.e., the filter gain g is a Fredholm operator as opposed to being a Volterra operator. In the same vein, the smoothing stage for dynamical systems is backward (in time) or anticausal. In the elliptic system case, however, the smoothing stage is also characterized by Fredholm operators. The notion of causality is not even introduced here although it is possible to do this for certain classes of elliptic systems [1].

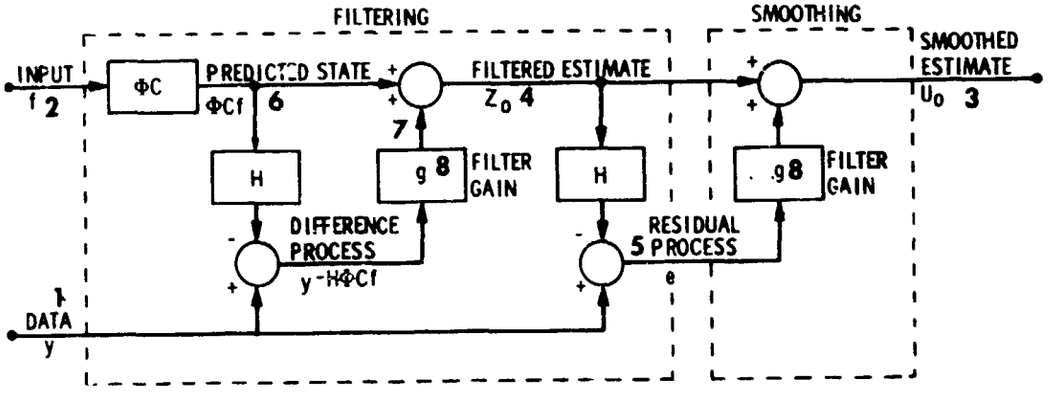

Fig. 8.3 Filtering and Smoothing

Spectral Representations: Smoothing, Filtering, and the Residuals

The aims here are: to obtain spectral representations for the filtered and smoothed estimates u_o and z_o and the corresponding error covariances $\bar{P}$ and p; to explore the predictor-corrector structure of the spectral representations of the filter and smoother; and to investigate the pseudo-innovations properties of the spectral representation of the residual process. The term "spectral representation" means the use of an expansion in terms of the eigensystem ϕ_j of R and of the related functions

$\psi_j = \lambda_j^{-1} B^* \Phi^* H^* \phi_j$, $x_j = \lambda_j^{-1} \Phi B \psi_j$ and $p_j = \lambda_j^{-1} \Phi^* H^* \phi_j$.

Result 8.23 The smoothed state estimate u_o can be represented as

$$u_o = \Phi C f + \sum \sin^2 \alpha_j \, (y_j - m_j) \, x_j, \tag{8.59}$$

where $y_j = \phi_j^* y$ and $m_j = \phi_j^* m$ are the spectral components of the data y and the suspected mean $m = H\Phi C f$. The related observed-state estimate Hu_o is specified by

$$Hu_o = m + HG\,(y-m), \qquad Hu_o = (I-HG)\,m + HGy. \tag{8.60}$$

In spectral form, $Hu_o = \sum u_o^j \phi_j$ where

$$u_o^j = m_j + \sin^2 \alpha_j \, (y_j - m_j), \quad u_o^j = \cos^2 \alpha_j \, m_j + \sin^2 \alpha_j y_j, \tag{8.61}$$

and $u_o^j = \phi_j^* H u_o$. Let $u_p = u - u_o$ denote the estimation error. The error covariances $\bar{P} = E(u_p u_p^*)$ and $H\bar{P}H^* = E(u_p^* H^* H u_p)$ are represented by

$$\bar{P} = \bar{R}_o + \sum \sin^2 \alpha_j \, x_j x_j^* \qquad H\bar{P}H^* = \sum \sin^2 \alpha_j \phi_j \phi_j^*. \tag{8.62}$$

Furthermore, the corresponding mean-square estimation errors $E(u_p^* u_p) = Tr[\bar{P}]$ and $E(u_p^* H^* H u_p) = Tr\,[H\bar{P}H^*]$ are

$$E(u_p^* u_p) = Tr\,[\bar{R}_o] + \sum \sin^2 \alpha_j \, x_j^* x_j, \qquad E(u_p^* H^* H u_p) = \sum \sin^2 \alpha_j \tag{8.63}$$

Proof: To establish (8.59), substitute $y = \Sigma y_j \phi_j$ and $m = \Sigma m_j \phi_j$ in (8.1). To show (8.60) multiply u_o in (8.1) by H and recall that $m = H\Phi C f$. To establish 8.61, multiply (8.60) by ϕ_j^*. The equation for $\bar{P}$ in (8.62) follows by substitution of (8.23) in (8.8) and use of the conditions $H\bar{R}_o = \bar{R}_o H^* = H\bar{R}_o H^* = 0$. The equation for $H\bar{P}H^*$ in (8.62) follows from $\bar{P}$ and use of $\phi_j = Hx_j$. Eq (8.63) follows from (8.62) and the orthonormality of ϕ_j.

Result 8.24 The filtered state estimate z_o can be represented by

$$z_o = \Phi C f + \sum (1-\cos\alpha_j)(y_j - m_j)\, x_j, \tag{8.64}$$

The related observed state estimate $z = Hz_o$ is

$$z = m + Hg(y-m), \qquad z = (I-Hg)\, m + Hgy. \tag{8.65}$$

In spectral form, $z = \sum z_j \phi_j$

$$z_j = m_j + (1-\cos\alpha_j)(y_j - m_j), \qquad z_j = \cos\alpha_j\, m_j + (1-\cos\alpha_j)\, y_j. \tag{8.66}$$

Let $z_p = z - z_o$ denote the estimation error. The estimation error covariances $B(z_p z_p^*) = p + p^*$ and $B(Hz_p z_p^* H^*) = H(p + p^*)H^*$ can be represented as

$$p = (1/2)\, \bar{R}_o + \sum (1-\cos\alpha_j)\, x_j x_j^*, \qquad HpH^* = \sum (1-\cos\alpha_j)\, \phi_j \phi_j^*. \tag{8.67}$$

Furthermore, the corresponding mean-square estimation errors are

$$E(z_p^* z_p) = Tr(p + p^*), \quad E(z_p^* H^* H z_p) = Tr\,[H(p + p^*)H^*], \tag{8.68}$$

where

$$Tr\,[p] = (1/2)\, Tr\,[\bar{R}_o] + \sum (1-\cos\alpha_j)\, x_j^* x_j, \quad Tr\,[HpH^*] = \sum (1-\cos\alpha_j). \tag{8.69}$$

Proof: To show (8.64), substitute $y = \Sigma y_j \phi_j$ and $m = \Sigma m_j \phi_j$ into z_o in (8.1). Eq. (8.65) follows from multiplication of (8.64) by H and use of $m = H\Phi C f$. Eq. (8.66) is obtained from (8.65) upon multiplication by ϕ_j^* and use of the orthonormality of ϕ_j. The equation for p in (8.67) has been established in (8.38) and is repeated here only for convenience. The second of Eq. (8.67) follows from use of the identity $\phi_j = Hx_j$. Eq. (8.68) follows from the definition of $p = p^*$ in (8.38). Eq. (8.69) is established by performing the trace operation on (8.67).

Result 8.25 The residual process $e = y - Hz_o$ can be represented as

$$e = \sum e_j \phi_j, \quad e_j = \phi_j^* e. \tag{8.70}$$

The spectral components e_j are independent random variables with zero-mean and unit covariance, i.e.,

$$E(e_i e_j) = 0 \text{ for } i \neq j, \qquad E(e_i^2) = 1. \tag{8.71}$$

Furthermore, the spectral components e_i and y_i of the residual and difference processes $e = y - Hz_0$ and $y = \bar{y} - m$ are related by the reciprocal relationships

$$e_i = \cos\alpha_i \, \bar{y}_i, \qquad \bar{y}_i = \sec\alpha_i \, e_i. \tag{8.72}$$

Proof: Eq. (8.70) is valid because ϕ_i are orthonormal in H To show (8.71), observe that $E(e_i e_j) = \phi_i^* \, \Xi(ee^*) \, \phi_j$ and then use (8.45) and the orthonormality of ϕ_j. Equations (8.72) are the spectral representations of the reciprocal relationships (8.47). Note that (8.72) can also be established by the simple trigonometric identity $(1/\cos\alpha_i) = \sec\alpha_i$.

9. NUMERICAL SEARCH CALCULATION SUMMARY

Since the development of the estimation approach is rather lengthy, it is convenient to summarize the steps that are required to implement the numerical search.

It is assumed that the process starts with a known input f, a set of data y and an initial parameter estimate Θ^n. To conduct an iteration in the numerical search requires that the following steps be performed:

1. Compute the suspected mean and covariance $m = H\Phi Cf$ and $R = H\Phi BB^*\Phi^*H^*$.

2. Compute the eigenvalues λ_k^2 and eigenvectors ϕ_k of R.

3. Compute the related vectors $p_k = \lambda_k^{-1}\Phi^*H^*\phi_k$, $\psi_k = B^*p_k$ and $x_k = \lambda_k^{-1}\Phi B\psi_k$.

4. Conduct a spectral analysis of the data and of the suspected mean to obtain the spectral coefficients $y_k = \phi_k^* y$ and $m_k = \phi_k^* m$.

5. Use Result 6.5 to evaluate the gradient $\partial J/\partial\Theta$ of the likelihood functional.

6. Use Results 6.6 and 6.7 to compute the Hessian M_n and to determine the incremental change $\delta\Theta^n$ of the parameter estimates.

7. Obtain a new parameter estimate $\Theta^{n+1} = \Theta^n - \delta\Theta^n$, return to step 1 above, and iterate through steps 1 to 6 until convergence is achieved.

If Cramer-Rao bounds and/or an optimal input are desired use (7.6) - (7.13). If the covariance of the state estimation error is desired use Result 8.2 and/or 8.13.

The calculations involved in conducting a single iteration in the maximum-likelihood parameter estimation approach are summarized in block diagram form in Fig. 9.1 A single iteration consists of all of the computational steps required to obtain an updated parameter estimate Θ^{n+1} by processing the available data, the known deterministic input, and the current parameter estimate Θ^n.

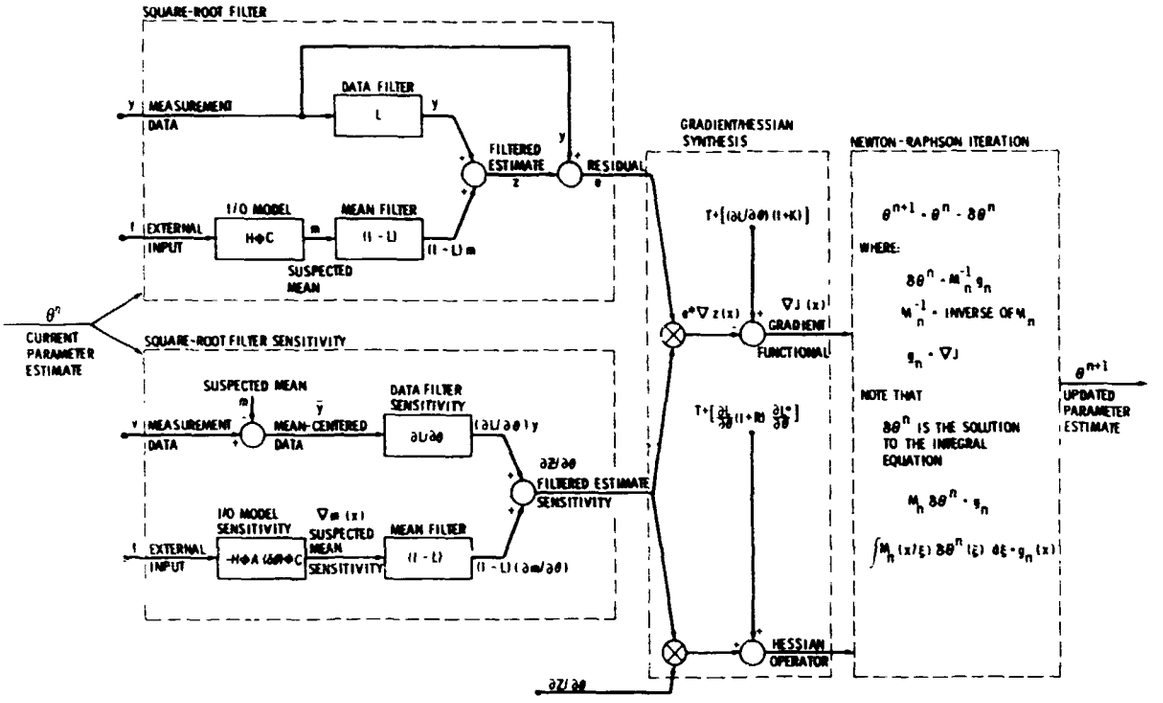

Fig. 9.1 Calculations Required for Single Iteration in Modified Newton-Raphson Search

To simplify the description of these computations, the steps performed in a single iteration have been grouped into the following four major blocks (delineated by the broken lines in the diagram):

- a SQUARE-ROOT FILTER block that processes the measurement data y and the external input f to obtain a filtered estimate z and a corresponding residual process e, defined as the difference between the data and the filtered state estimate. The square-root filter implements the equations $z = Ly + (I-L)m$ and $e = y-z$. The central computation in the square-root filter block is that provided by the operator $L = I - (I+R)^{-1/2}$ defined in terms of the square-root of the predicted-data-covariance $(I + R)$. This operator appears in two distinct places in the diagram: in the data filter, whose primary function is to process the measurements y; and in the mean filter, whose main function is to process the suspected mean m. The suspected mean is in turn obtained from the known external input by means of the input-output model.

- a SQUARE-ROOT FILTER SENSITIVITY block that processes the measurement data y and the deterministic input f to obtain the filtered estimate sensitivity $\partial z/\partial\Theta$. This block implements the equation $\partial z/\partial\Theta = (\partial L/\partial\Theta)(y-m) + (I-L)(\partial m/\partial\Theta)$. The computation of the sensitivity $\partial L/\partial\Theta$ is the main calculation performed in this block.

- a GRADIENT-HESSIAN SYNTHESIS block that forms: the function-space gradient $\partial J/\partial\Theta$ of the likelihood functional by means of the equation $\partial J/\partial\Theta = Tr[(\partial L/\partial\Theta)(I+K) - (\partial z/\partial\Theta)e^*]$; and the function-space approximate Hessian by means of the equation $M = Tr[(\partial L/\partial\Theta)(I+R)(\partial L^*/\partial\Theta)] + (\partial z/\partial\Theta)^*(\partial z/\partial\Theta)$. Note that the quantity that is actually evaluated in this block is the kernel $M(x/\xi)$ of the Hessian operator. This kernel is a function of two spatial variables x and ξ defined over a "square" domain $(x/\xi)\epsilon\Omega \times \Omega$, where Ω is as before the spatial domain of definition of the system model.

- a NEWTON-RAPHSON ITERATION block whose input is the gradient and the approximate Hessian and that generates as an output the updated parameter distribution Θ^{n+1} for the next iteration. The central calculation in this block is the solution of the integral equation $M_n\delta\Theta^n = g_n$ that results in the parameter estimate update $\delta\Theta^n$.

After specification of the parameter estimate Θ^{n+1}, the square-root filter $L(\Theta^n)$ and its sensitivity $\partial L(\Theta^n)/\partial\Theta$ are redesigned by letting $\Theta^n \rightarrow \Theta^{n+1}$, and the steps outlined above are repeated in order to conduct the next step in the iterative process for optimization.

The predicted-data-covariance square-root filter processes the data y and the suspected mean m to produce a filtered state estimate z and a set of residuals e = y-z. This is done by means of the equation $z = Ly + (I-L)m$, where $L = I-(I+R)^{-1/2}$. This equation, while providing a very succinct symbolic description of the square-root filter, does not by itself provide a recipe to conduct computations. In order to provide such a recipe, it is convenient to use the corresponding spectral form $z_k = (1-\cos\alpha_k)y_k + \cos\alpha_k m_k$, which expresses the spectral amplitudes $z_k = \phi_k^* z$ of the filtered state estimate z as a linear combination of the data and suspected mean spectral amplitudes y_k and m_k. Such a spectral form of the predicted-data-covariance square-root filter is illustrated in Fig. 9.2.

The diagram in the figure illustrates the main calculations involved in the square-root filter. On the upper branch of the diagram, a set of data[1] $y = [y_1,\ldots,y_N]$ is assumed to be available at N discrete locations. A spectral analysis[2] is conducted on this data to obtain the data spectral amplitudes[3] $[y^1,\ldots,y^N]$. These spectral amplitudes are then multiplied by the coefficients $(1-\cos\alpha_k)$ in the data filter[4], resulting in the terms $(1-\cos\alpha_k)y^k$. On the lower branch of the diagram, the deterministic inputs f_i[6] are

processed by the input/output system model[7] to obtain the suspected mean $m = [m_1,\ldots,m_N]$[8]. The spectral amplitudes $m^k = \phi_k^* m$[9] of the suspected mean are then computed and subsequently multiplied by the coefficients $\cos\alpha_k$ in the mean filter[10] to produce the terms $\cos\alpha_k m^k$ [11]. This last term is then added to $(1-\cos\alpha_k)y^k$ in[5] resulting in the filtered state spectral amplitudes z_k[12] and the residuals e_k[13]. Note that the physical state estimate z and the residual e can be recovered from z_k and e_k by means of the summations $z = \Sigma z_k\phi_k$ and $e = \Sigma e_k\phi_k$, although for simplicity this last transformation is not shown on the diagram.

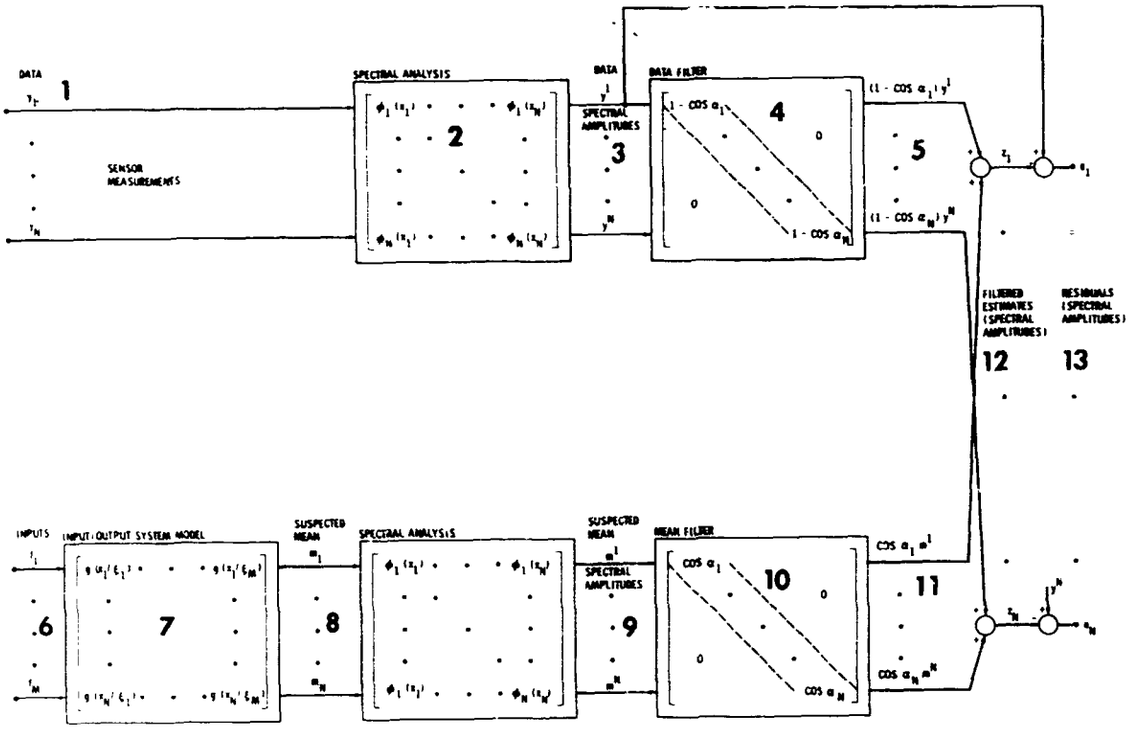

Fig. 9.2 Spectral Form of Predicted-Data-Covariance Square-Root Filter

The foregoing remarks have scrutinized the spectral form of the square-root filter equation $z = Ly + (I-L)m$. The immediate aim now is to conduct a similar detailed analysis of the spectral representation of the square-root filter sensitivity equation $\partial z/\partial\Theta = (\partial L/\partial\Theta)y + (I-L)(\partial m/\partial\Theta)$. The spectral form of this equation is stated in Eq. (6.15) and illustrated in the block diagram in Fig. 9.3. The overall primary

function of the square-root filter sensitivity is to process the N mean-centered data spectral amplitudes[1] and the M deterministic inputs[2] in order to obtain the spectral amplitudes of[3] of the filtered state estimate sensitivity $\partial z/\partial\theta$. An intermediate calculation embedded within this overall process involves processing of the mean-centered data spectral amplitudes y^{k}[1] by means of the N-by-N matrix, with general elements a_{km}, representing the data filter sensitivity $\partial L/\partial\theta$[4]. Other intermediate steps involve: processing of the deterministic inputs f_m [2] by the input/output model sensitivity matrix b_{km} [5] to generate the suspected mean spectral amplitudes $(\partial m/\partial\theta)_k$ [6]; and subsequent processing of these coefficients by the mean filter [7] to obtain the terms $\cos\alpha_k\ (\partial m/\partial\theta)_k$ [8].

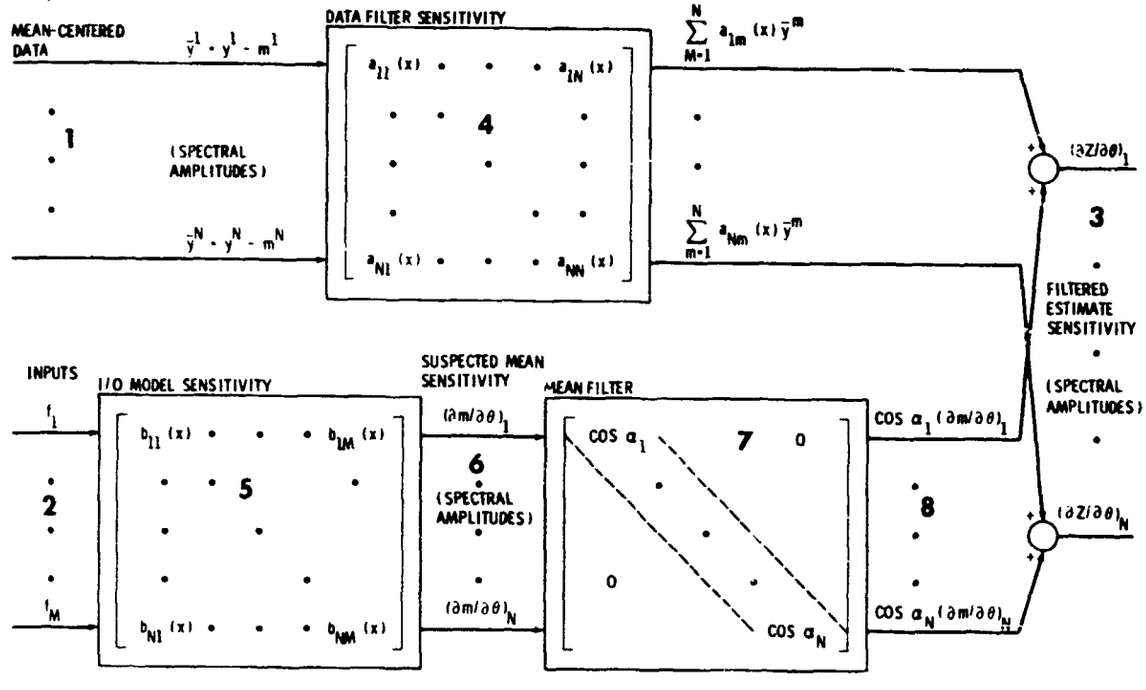

Fig. 9.3 Spectral Form of Square-Root Filter Sensitivity

10. CONCLUDING REMARKS AND FUTURE DIRECTIONS

The area of estimation for elliptic systems is so full of interesting research problems that, in spite of all that this paper has covered, much more remains to be done. These are some of the problems that lie ahead:

- Conduct of an asymptotic statistical property analysis that explores the convergence of the parameter estimates as the number of observations increases.

- Development of approximation approaches that rigorously arrive at finite-dimensional approximations to the infinite-dimensional solutions advanced here.

- More complete investigation of the optimal input design problem. In particular, development of "spectral" domain design approaches which would do for elliptic systems what the frequency domain methods achieve for linear time - invariant dynamical systems.

- Development of more precise mathematical arguments to justify function-space differentiation, eigensystem expansions, covariance calculations, likelihood-ratio derivations, etc.

- Investigation of alternative (to the square-root) factorization of the predicted-data-covariance that could result in easier calculation of the function-space derivatives necessary for the Newton-Raphson search.

- Numerical experimentation with the filtering, smoothing and identification algorithms to gain further insight into the state and parameter estimation approaches and solutions [5].

As a final remark, this paper is a concrete example of the power of the functional analysis approach to estimation advanced in Ref. [4]. Because of the conceptual simplicity of the method, it has been possible to solve in this paper problems that would have defied solution by any other method. It has also made it possible to conceive areas for future research that would otherwise have been left unidentified.

REFERENCES

1. Rodriguez, G., "A Function Space Approach to State and Model Error Estimation for Elliptic Systems," Proceedings of the Workshop on Applications of Distributed System Theory to the Control of Large Space Structures, JPL Publication 83-46, July 1983.

2. Luenberger, D. G., Optimization by Vector Space Methods, Prentice-Hall, 1971.

3. Balakrishnan, A. V., Communication Theory, McGraw-Hill, 1968.

4. Balakrishnan, A. V., Applied Functional Analysis, Springer-Verlag, 1977.

5. Scheid, R. E. and G. Rodriguez, "Numerical Experimentation with Maximum Likelihood Identification in Static Distributed Systems", Proceedings of NASA/JPL/LRC Workshop on Identification and Control of Flexible Space Structures, San Diego, CA., June 1984.

ACKNOWLEDGEMENT

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.