JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 15, No. 3, May-June 1992

Maximum Likelihood State and Parameter Estimation Via Derivatives of the V-Lambda Filter

Yaakov Oshman*
Technion—Israel Institute of Technology, Haifa, Israel

Applying the method of maximum likelihood to the problem of parameter and state estimation in linear dynamical systems requires the implementation of a recursive algorithm that consists of a Kalman filter and its derivatives with respect to each of the unknown parameters (the sensitivity equations). Since the conventional Kalman filtering algorithm has been shown to be numerically unstable, a different approach is taken in this paper, which is based on using the V-Lambda square root filtering technique. Equations are developed for the recursive computation of the log-likelihood function gradient (score) and the Fisher information matrix (FIM) in terms of the V-Lambda filter variables, which are the eigenfactors (eigenvalues and eigenvectors) of the estimation error covariance matrix. Based on the singular value decomposition, the recently introduced V-Lambda filters have been shown to be numerically stable and accurate. Therefore, their usage renders the resulting maximum likelihood scheme numerically robust. Moreover, making the covariance eigenfactors available to the user at all estimation stages, which is an inherent and unique property of the V-Lambda class, adds invaluable insight into the heart of the estimation process.

I. Introduction

THE maximum likelihood (ML) method has been widely applied to problems of state and parameter estimation in dynamical systems.[1-3] In principle, the method requires the implementation of a Kalman filter for the estimation of the states (based on the current best estimate of the parameters), as well as the solution of an uncoupled set of sensitivity equations, which consist of derivatives of the Kalman filter equations with respect to the unknown parameters.[4,5] The sensitivity equations, which are driven by the Kalman state estimator as well as by the measurements, are used to compute the log-likelihood function gradient (score) and Fisher information matrix (FIM), which approximates the Hessian in some computational schemes.[6]

The numerical instability of the conventional Kalman filtering algorithm is now well established[7] (see Ref. 8 for a definition and discussion of the numerical stability of the computational algorithm). It has been shown in numerous examples that implementation of this algorithm (either in its continuous-time or in its discrete-time version) on finite word-length computers may lead to the appearance of negative definite covariance eigenvalues, substantial loss of accuracy, and consequent filter divergence.[9,10] In Ref. 11, Bierman and Thornton demonstrated the numerical instability of the conventional Kalman algorithm as well as Joseph's so-called "stabilized" algorithm, by way of an orbit determination case study that was based on a portion of the 1977 Mariner Jupiter-Saturn deep space mission. Contrary to the common belief that the Kalman filter is robust and numerical stability is not an issue if only the problem is well posed in an engineering sense, Bierman and Thornton's work proved that numerical failures and consequent performance degradation may occur in an uncontrived, realistic situation. As is now widely recognized, the best solution to these problems (except for some ad hoc solutions that may or may not work in different situations) is the use of square root filters.[11-15] These algorithms, while being algebraically equivalent to the Kalman algorithm, are numerically stable and highly accurate. Since the estimation error covariance matrix is replaced in these algorithms by its square root factors, which are then propagated and updated, it implicitly retains symmetry and positive definiteness.

As noted previously, the maximum likelihood method for estimation of states and parameters in dynamical systems utilizes the Kalman filter and its derivatives with respect to the parameters. The numerical instability of the conventional Kalman algorithm inevitably renders the resulting ML algorithm numerically unstable. Intuitively, this sensitivity even increases when partial derivatives are taken. It is conceivable, therefore, that replacing the filtering algorithm by a numerically stable equivalent should greatly enhance the numerical characteristics of the ML scheme. The utilization of a square root filter in the context of maximum likelihood estimation was first introduced by Bierman et al.[16] They used the square root information filter (SRIF)[7,10] to derive sensitivity equations that are phrased in terms of the SRIF factors. In this formulation, the square root factors are upper triangular matrices, which are propagated and updated using orthogonal Householder transformations. Procedures were developed for the numerically robust computation of the log-likelihood function gradient and the Fisher information matrix in terms of derivatives of the square root factors and state estimates.

Inspired by the work presented in Ref. 16, this paper introduces a recursive ML algorithm that is based on the V-Lambda square root filter and its derivatives. The recently introduced V-Lambda class of filters[14,15] is based on the spectral decomposition of the error covariance matrix P, i.e., $P \equiv V\Lambda V^T$, where V is the matrix whose columns are the eigenvectors of P and Λ is the diagonal eigenvalue matrix. The discrete time versions of the V-Lambda filters are based on the singular value decomposition (SVD) as a main computational tool, which renders them extremely numerically robust and accurate. The introduction, in this paper, of the derivatives of the V-Lambda filter with respect to the parameters, i.e., the V-Lambda sensitivity equations, facilitates the use of this square root filter in the context of maximum likelihood parameter estimation. Compared with other square root alternatives, which use upper triangular factors of the covariance, the V-Lambda filters are attractive because they continuously provide their user with the eigenfactors of the estimation error covariance matrix. These may be continuously monitored to reveal singularities as they occur and to identify those state subsets that are nearly dependent (see Ref. 7, p. 100; Ref. 17,

---

Received Aug. 21, 1990; revision received April 25, 1991; accepted for publication May 17, 1991.

*Senior Lecturer, Faculty of Aerospace Engineering. Member AIAA.

p. 72; and Ref. 18). Relying on the singular value decomposition, these filters are more computationally expensive than other alternatives, when using conventional, serial implementation. However, the recent advent in the area of parallel computation of the SVD using multiprocessor arrays should alleviate the computational burden and make these filters attractive from this perspective as well.

In Sec. II the ML parameter estimation problem is defined. The V-Lambda filter that is used to obtain the ML estimate is stated in Sec. III. The log-likelihood function, its gradient (the score), and the FIM are expressed in Sec. IV in terms of the V-Lambda variables and derivatives. Algorithms for the recursive computation of these derivatives are presented in Sec. V. These algorithms constitute the square root equivalents of the conventional sensitivity equations. The complete procedure is summarized in Sec. VI. Concluding remarks are given in Sec. VII.

II. Maximum Likelihood State and Parameter Estimation

Consider the following linear, discrete-time, stochastic dynamical system and observation:

$$x(k+1) = F(k,\Theta)x(k) + G(k,\Theta)w(k) \tag{1}$$

$$y(k) = H(k,\Theta)x(k) + v(k) \tag{2}$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $y(k) \in \mathbb{R}^m$ is the measurement vector, and $\{w(k)\} \in \mathbb{R}^q$ and $\{v(k)\} \in \mathbb{R}^m$ are the process and measurement zero mean Gaussian white sequences, respectively, with positive definite covariance matrices $E[w(k)w^T(j)] = Q(k,\Theta)\delta_{kj}$ and $E[v(k)v^T(j)] = R(k,\Theta)\delta_{kj}$. The initial state $x(0)$ is a Gaussian random vector with mean μ_0 and covariance $E\{[x(0)-\mu_0][x(0)-\mu_0]^T\} = P(0)$. It is further assumed that the process and observation noise sequences are not correlated with each other or with the initial state random vector. The matrices $F(k,\Theta)$, $G(k,\Theta)$, $H(k,\Theta)$, $Q(k,\Theta)$, and $R(k,\Theta)$ are assumed to depend on an unknown vector of parameters $\theta \in \mathbb{R}^p$, and the problem at hand is to obtain the maximum likelihood estimate of this vector based on a sequence of measurements, $\mathcal{Y}(N) := \{y(k)\}_{k=1}^N$. For simplicity of notation, the explicit dependence on the parameter vector Θ will be suppressed in the sequel.

The maximum likelihood estimate $\hat{\Theta}_{ML}$ is defined as that value of Θ that maximizes the joint probability density function (pdf) of the measurements. Equivalently, it is that value of Θ that maximizes the following log-likelihood function[5]:

$$L\left[\Theta \mid \mathcal{Y}(N)\right] = -\frac{1}{2}\sum_{k=1}^{N}\left[\tilde{y}^T(k|k-1)\mathfrak{N}^{-1}(k|k-1)\tilde{y}(k|(k-1) + \log\det\mathfrak{N}(k|k-1)\right] \tag{3}$$

which is derived from the joint pdf of the innovations process $\{\tilde{y}(k|k-1)\}_{k=1}^N$. The innovations process covariance is $\mathfrak{N}(k|k-1)$. The innovations process and the covariance of that process are conventionally computed via a Kalman filter as

$$\tilde{y}(k|k-1) = y(k) - H(k)\hat{x}(k|k-1) \tag{4}$$

$$\mathfrak{N}(k|k-1) = H(k)P(k|k-1)H^T(k) + R(k) \tag{5}$$

where $\hat{x}(k|k-1)$, the a priori state estimate, and $P(k|k-1)$, the a priori estimation error covariance matrix at time k, are based on the measurement history $\mathcal{Y}(N-1)$. Since the parameter vector Θ enters the log-likelihood function in a nonlinear manner, the computation of the ML estimate is carried out via an iterative, nonlinear mathematical programming algorithm of the form[6]

$$\hat{\Theta}_{i+1} = \hat{\Theta}_i + \rho_i\mathcal{K}_i^{-1}(\hat{\Theta}_i)g_i, \qquad i = 0,1,\ldots \tag{6}$$

where g_i is the gradient of the log-likelihood function

$$g_i = \nabla_\Theta L\left[\Theta \mid \mathcal{Y}(N)\right]\big|_{\Theta=\hat{\Theta}_i} \tag{7}$$

and

$$\nabla_\Theta(\cdot) \equiv \left[\frac{\partial(\cdot)}{\partial\Theta(1)}, \frac{\partial(\cdot)}{\partial\Theta(2)}, \ldots, \frac{\partial(\cdot)}{\partial\Theta(p)}\right]^T$$

$\mathcal{K}_i$ is a gain matrix that is typically computed via some approximation to the Hessian of $L[\Theta \mid \mathcal{Y}(N)]$

$$\mathcal{K}_i = \frac{\partial^2 L\left[\Theta \mid \mathcal{Y}(N)\right]}{\partial\Theta^2}\bigg|_{\Theta=\hat{\Theta}_i} \tag{8}$$

and ρ_i is a scalar step size control parameter. In the Gauss-Newton method, the Hessian is approximated by the FIM, which is its expected value (up to a minus sign). This computation involves only the first derivatives of the log-likelihood function, which are needed anyway in the evaluation of the score. To solve problems of singularity or near-singularity of the computed information matrix, the Levenberg-Marquardt algorithm uses the following gain:

$$\mathcal{K}_i = \mathfrak{M}_i + \mathfrak{D}_i \tag{9}$$

where $\mathfrak{M}_i$ is the Fisher information matrix (based on the current estimate $\hat{\Theta}_i$), and $\mathfrak{D}_i$ is a diagonal matrix chosen to insure positive definiteness of $\mathcal{K}_i$.

This paper presents a method for the computation of the gradient of the log-likelihood function and the Fisher information matrix, using the square root V-Lambda filtering algorithm implementation. Avoiding the numerical instability problem inherent in the conventional Kalman filter, the new algorithm is numerically superior to the conventional mechanization. The V-Lambda filtering algorithm is stated in the next section.

III. V-Lambda State Filtering

In this section the V-Lambda filtering algorithm, which will be used in the sequel, is presented. The algorithm is a hybrid one, consisting of a time update that is performed in covariance mode and a measurement update that is performed in information mode. Although the V-Lambda class contains complete algorithms in both covariance and information modes, this special arrangement of the algorithm is necessary for the ensuing development of the ML parameter estimator in the next sections. The algorithm is presented without proof; for a detailed derivation, the reader is referred to Refs. 14 and 15.

As stated in Sec. I, the V-Lambda filter is based on the spectral decomposition of the covariance (or information) matrix, $P \equiv V\Lambda V^T$. The spectral factors are propagated in time and updated across measurement according to the algorithms that follow.

V-Lambda Time Update (Covariance Mode)

Given the measurement-updated factors $V(k-1|k-1)$ and $\Lambda^{1/2}(k-1|k-1)$ of the a posteriori covariance $P(k-1|k-1)$ at time $k-1$ based on the measurement history $\mathcal{Y}(N-1)$, the following algorithm computes the time-propagated factors $V(k|k-1)$ and $\Lambda^{1/2}(k|k-1)$ of the a priori covariance $P(k|k-1)$:

1) Define the augmented matrix $\mathcal{A}(k-1) \in \mathcal{R}^{n,n+q}$ as

$$\mathcal{A}(k-1) := \left[F(k-1)V(k-1|k-1)\Lambda^{1/2}(k-1|k-1) \quad G(k-1)Q^{1/2}(k-1)\right] \tag{10}$$

where $Q^{1/2}(k-1)$ is the matrix square root of $Q(k-1)$, which can be computed via a Cholesky decomposition.[19]

2) Use the singular value decomposition of $\mathcal{A}(k-1)$ to write

$$\mathcal{A}(k-1) = \Xi(k-1)\left[\Sigma(k-1) \quad 0\right]Z^T(k-1) \tag{11}$$

where $\Sigma(k-1)$ is the diagonal singular value matrix, and $\Xi(k-1)$ and $Z(k-1)$ are, respectively, the left and right singular vector matrices.

3) From the SVD [Eq. (11)] read the time-updated spectral factors of $P(k|k-1)$:

$$V(k|k-1) = \Xi(k-1) \tag{12}$$

$$\Lambda^{1/2}(k|k-1) = \Sigma(k-1) \tag{13}$$

4) Compute the state estimate as

$$\hat{x}(k|k-1) = F(k-1)\hat{x}(k-1|k-1) \tag{14}$$

V-Lambda Measurement Update (Information Mode)

Given the time-propagated eigenfactors $V(k|k-1)$ and $\Lambda^{1/2}(k|k-1)$ of the a priori information matrix $P^{-1}(k|k-1)$, and the a priori normalized state estimate $\hat{q}(k|k-1)$, where

$$\hat{q}(k|k-1) := \Lambda^{1/2}(k|k-1)V^T(k|k-1)\hat{x}(k|k-1) \tag{15}$$

the following algorithm computes the a posteriori eigenfactors $V(k|k)$ and $\Lambda^{-1/2}(k|k)$, and the updated normalized estimate $\hat{q}(k|k)$, defined as

$$\hat{q}(k|k) := \Lambda^{-1/2}(k|k)V^T(k|k)\hat{x}(k|k) \tag{16}$$

1) Define the augmented matrix $\mathcal{B}(k) \in \mathcal{R}^{n+m,n}$ as

$$\mathcal{B}(k) := \begin{bmatrix} \Lambda^{-1/2}(k|k-1)V^T(k|k-1) \\ R^{-1/2}(k)H(k) \end{bmatrix} \tag{17}$$

2) Perform a singular value decomposition of $\mathcal{B}(k)$ to obtain

$$\mathcal{B}(k) = Y(k)\begin{bmatrix} \Psi(k) \\ 0 \end{bmatrix}\Phi^T(k) \tag{18}$$

where $\Psi(k)$ is the diagonal singular value matrix and $Y(k)$ and $\Phi(k)$ are, respectively, the left and right singular vector matrices.

3) The measurement updated spectral factors are related to the SVD factors of $\mathcal{B}(k)$ as follows:

$$V(k|k) = \Phi(k) \tag{19}$$

$$\Lambda^{-1/2}(k|k) = \Psi(k) \tag{20}$$

4) Define the vector $b(k) \in \mathcal{R}^{n+m}$ as

$$b(k) := \begin{bmatrix} \hat{q}(k|k-1) \\ R^{-1/2}(k)y(k) \end{bmatrix} \tag{21}$$

5) The updated normalized estimate is found by partitioning $Y^T(k)b(k)$ corresponding to the partition of $b(k)$ as

$$Y^T(k)b(k) = \begin{bmatrix} \hat{q}(k|k) \\ e(k) \end{bmatrix} \tag{22}$$

where the m vector $e(k)$ is the normalized estimation residual (i.e., the innovation, normalized by its covariance square root) as will be shown in the sequel.

Returning to Eq. (14) and using the definitions Eqs. (15) and (16) and the expressions for the propagated and updated spectral factors yields the following time-update equation for the *normalized state estimate*:

$$\hat{q}(k|k-1) = \Sigma^{-1}(k-1)\Xi^T(k-1)F(k-1)\Phi(k-1) \times \Psi^{-1}(k-1)\hat{q}(k-1|k-1) \tag{23}$$

which is equivalent to Eq. (14).

In the next section, the log-likelihood function and its gradient are expressed in terms of the V-Lambda estimator variables and their derivatives with respect to the parameters.

IV. Log-Likelihood Function, Its Gradient, and Fisher Information Matrix: V-Lambda Formulation

Log-Likelihood Function

In Sec. II the log-likelihood function was defined in Eq. (3), which is repeated here for convenience:

$$L\left[\Theta \mid \mathcal{Y}(N)\right] = -\frac{1}{2}\sum_{k=1}^{N}\left[\tilde{y}^T(k|k-1)\mathfrak{N}^{-1}(k|k-1)\tilde{y}(k|k-1) + \log\det\mathfrak{N}(k|k-1)\right] \tag{24}$$

The ultimate goal of this work is to express the score and the FIM in terms of the V-Lambda filter variables, which will enable the mechanization of the ML identification scheme via the V-Lambda filter. To this end, first we need to express Eq. (24) in terms of the V-Lambda filter variables. This is done next in Theorem 1.

Theorem 1: In terms of the V-Lambda filter variables (Sec. II), the log-likelihood function Eq. (24) can be expressed as

$$L\left[\Theta \mid \mathcal{Y}(N)\right] = -\frac{1}{2}\sum_{k=1}^{N}\left[\|e(k)\|^2 + \log\frac{\det R(k)\cdot\det\Lambda(k|k-1)}{\det\Lambda(k|k)}\right] \tag{25}$$

which can also be written as

$$L\left[\Theta \mid \mathcal{Y}(N)\right] = -\frac{1}{2}\sum_{k=1}^{N}\left[\|e(k)\|^2 + \log\det R(k) + \sum_{i=1}^{n}\log\frac{\lambda_i(k|k-1)}{\lambda_i(k|k)}\right] \tag{26}$$

where $\{\lambda_i(k|k-1)\}_{i=1}^{n}$ are the eigenvalues of the a priori estimation error covariance.

Proof

The proof consists of two parts, in which it is shown, respectively, that

$$\tilde{y}^T(k|k-1)\mathfrak{N}^{-1}(k|k-1)\tilde{y}(k|k-1) = \|e(k)\|^2 \tag{27}$$

and

$$\log\det\mathfrak{N}(k|k-1) = \log\frac{\det R(k)\cdot\det\Lambda(k|k-1)}{\det\Lambda(k|k)} \tag{28}$$

To prove Eq. (27), we start by rewriting Eq. (22) as

$$\begin{bmatrix} Y_{11}^T(k) & Y_{21}^T(k) \\ Y_{12}^T(k) & Y_{22}^T(k) \end{bmatrix} \begin{bmatrix} \Lambda^{-1/2}(k|k-1)V^T(k|k-1)\hat{x}(k|k-1) \\ R^{-1/2}(k)y(k) \end{bmatrix} = \begin{bmatrix} \Lambda^{-1/2}(k|k)V^T(k|k)\hat{x}(k|k) \\ e(k) \end{bmatrix} \tag{29}$$

Here, matrix $Y^T(k)$ was written in partitioned form and use was made of Eqs. (15), (16), and (21). From Eq. (29) we readily obtain

$$e(k) = Y_{12}^T(k)\Lambda^{-1/2}(k|k-1)V^T(k|k-1)\hat{x}(k|k-1) + Y_{22}^T(k)R^{-1/2}(k)y(k) \tag{30}$$

Next, using Eqs. (17), (19), and (20) in Eq. (18) and again writing $Y^T(k)$ in partitioned form, we have

$$\begin{bmatrix} \Lambda^{-1/2}(k|k)V^T(k|k) \\ 0 \end{bmatrix} = \begin{bmatrix} Y_{11}^T(k) & Y_{21}^T(k) \\ Y_{12}^T(k) & Y_{22}^T(k) \end{bmatrix} \begin{bmatrix} \Lambda^{-1/2}(k|k-1)V^T(k|k-1) \\ R^{-1/2}(k)H(k) \end{bmatrix} \tag{31}$$

from which

$$Y_{12}^T(k)\Lambda^{-1/2}(k|k-1)V^T(k|k-1) = -Y_{22}^T(k)R^{-1/2}(k)H(k) \tag{32}$$

To prove Eq. (28), note that there exists an orthogonal transformation τ such that[20]

$$\begin{bmatrix} R^{1/2}(k) & H(k)P^{1/2}(k|k-1) \\ 0 & P^{1/2}(k|k-1) \end{bmatrix}\tau = \begin{bmatrix} \mathfrak{N}^{1/2}(k|k-1) & 0 \\ P(k|k-1)H^T(k)\mathfrak{N}^{-T/2}(k|k-1) & P^{1/2}(k|k) \end{bmatrix}$$

Hence

$$\det R^{1/2}(k)\det P^{1/2}(k|k-1) = \det \mathfrak{N}^{1/2}(k|k-1)\det P^{1/2}(k|k) \tag{38}$$

Substituting the following relations

$$\det P^{1/2}(k|k-1) = \det \Lambda^{1/2}(k|k-1)$$

$$\det P^{1/2}(k|k) = \det \Lambda^{1/2}(k|k)$$

into Eq. (38) yields Eq. (28), which completes the proof. □

Log-Likelihood Gradient

Having developed the expressions Eqs. (25) and (26) for the log-likelihood function in terms of the V-Lambda filter variables, we now turn to the computation of the score. The log-likelihood gradient is the vector of partial derivatives of the log-likelihood function with respect to (w.r.t.) each of the components of the parameter vector Θ, i.e.,

$$\nabla_\Theta L[\Theta|\mathcal{Y}(N)] \equiv \left\{\frac{\partial}{\partial\Theta(1)}L[\Theta|\mathcal{Y}(N)], \frac{\partial}{\partial\Theta(2)}L[\Theta|\mathcal{Y}(N)], \ldots, \frac{\partial}{\partial\Theta(p)}L[\Theta|\mathcal{Y}(N)]\right\}^T$$

Using Eq. (32) in Eq. (30) yields

$$e(k) = Y_{22}^T(k)R^{-1/2}(k)\tilde{y}(k|k-1) \tag{33}$$

where the innovation $\tilde{y}(k|k-1)$ was defined in Eq. (4). From Eq. (33), the norm of $e(k)$ is computed as

$$\|e(k)\|^2 = \tilde{y}^T(k|k-1)R^{-T/2}(k)Y_{22}(k)Y_{22}^T(k)R^{-1/2}(k)\tilde{y}(k|k-1) \tag{34}$$

Employing the orthogonality of the singular vector matrix $Y(k)$, we have

$$Y_{22}(k)Y_{22}^T(k) = I - Y_{21}(k)Y_{21}^T(k) \tag{35}$$

From Eq. (31), $Y_{21}(k)$ is expressed as

$$Y_{21}(k) = R^{-1/2}(k)H(k)V(k|k)\Lambda^{1/2}(k|k) \tag{36}$$

Substituting Eqs. (35) and (36) into Eq. (34) results, after some algebra, in

$$\|e(k)\|^2 = \tilde{y}^T(k|k-1)\left[R^{-1}(k) - R^{-1}(k)H(k)P(k|k)H^T(k)R^{-1}(k)\right]\tilde{y}(k|k-1) \tag{37}$$

Finally, recalling the information mode Kalman measurement update,

$$P^{-1}(k|k) = P^{-1}(k|k-1) + H^T(k)R^{-1}(k)H(k)$$

using the matrix inversion lemma (Ref. 5, p. 30) and the innovation covariance definition in Eq. (5), the term inside the brackets in Eq. (37) is recognized to be $\mathfrak{N}^{-1}(k|k-1)$, from which Eq. (27) follows. This completes the first part of the proof.

where $\Theta(j)$ is the jth component of the parameter vector Θ. For simplicity, we will refer in the sequel to the partial derivative w.r.t. only one of the components. Correspondingly, the symbol $[\partial(\cdot)/\partial\Theta]$ will be used hereinafter to denote the partial derivative w.r.t. $\Theta(j)$, where $1 \leq j \leq p$.

Computing the gradient of the model dependent part Eq. (28) of the log-likelihood function and using the identity

$$\frac{\partial \log \det A}{\partial\Theta} = \operatorname{tr}\left(A^{-1}\frac{\partial A^T}{\partial\Theta}\right)$$

yields

$$\frac{\partial}{\partial\Theta}\log\frac{\det R(k)\cdot\det\Lambda(k|k-1)}{\det\Lambda(k|k)} = \operatorname{tr}\left[R^{-1}(k)\frac{\partial R(k)}{\partial\Theta}\right] + \operatorname{tr}\left[\Lambda^{-1}(k|k-1)\frac{\partial\Lambda(k|k-1)}{\partial\Theta}\right] - \operatorname{tr}\left[\Lambda^{-1}(k|k)\frac{\partial\Lambda(k|k)}{\partial\Theta}\right] \tag{39}$$

Also, since the eigenvalue matrices apearing in Eq. (39) are diagonal, this can be equivalently rewritten as

$$\frac{\partial}{\partial\Theta}\log\frac{\det R(k)\cdot\det\Lambda(k|k-1)}{\det\Lambda(k|k)} = \operatorname{tr}\left[R^{-1}(k)\frac{\partial R(k)}{\partial\Theta}\right] + \sum_{i=1}^{n}\frac{1}{\lambda_i(k|k-1)}\frac{\partial\lambda_i(k|k-1)}{\partial\Theta} - \sum_{i=1}^{n}\frac{1}{\lambda_i(k|k)}\frac{\partial\lambda_i(k|k)}{\partial\Theta} \tag{40}$$

Algorithms for computing the derivatives of the eigenvalues will be developed in the following section. Notice that only the eigenvalues (and not the eigenvectors) are involved in this part of the gradient. The indirect participation of the eigenvectors will be revealed in the sequel.

To compute the gradient of the data dependent part Eq. (27), norms are taken from both sides of Eq. (22), which yields

$$\|e(k)\|^2 = \|b(k)\|^2 - \|\hat{q}(k|k)\|^2 \tag{41}$$

Using Eq. (21) in Eq. (41) results in

$$\|e(k)\|^2 = \|\hat{q}(k|k-1)\|^2 - \|\hat{q}(k|k)\|^2 + \|R^{-1/2}(k)y(k)\|^2$$

from which, upon taking derivatives w.r.t. Θ, we obtain

$$\frac{\partial\|e(k)\|^2}{\partial\Theta} = 2\hat{q}^T(k|k-1)\frac{\partial\hat{q}(k|k-1)}{\partial\Theta} - 2\hat{q}^T(k|k)\frac{\partial\hat{q}(k|k)}{\partial\Theta} + 2y^T(k)R^{-T/2}(k)\frac{\partial R^{-T/2}(k)}{\partial\Theta}y(k) \tag{42}$$

Procedures for computing the derivatives of the normalized state estimates $\hat{q}(k|k)$ and $\hat{q}(k|k-1)$ will be derived in the next section. For the computation of $[\partial R^{-1/2}(k)/\partial\Theta]$ we proceed, following Ref. 16, by differentiating $R^{-1/2}(k)R(k)R^{-T/2}(k) = I$ w.r.t. Θ, which yields

$$R^{-1/2}(k)R_\Theta(k)R^{-T/2}(k) = -R_\Theta^{-1/2}(k)R^{1/2}(k) - \left[R_\Theta^{-1/2}(k)R^{1/2}(k)\right]^T \tag{43}$$

where, to simplify notation, $(\cdot)_\Theta$ denotes partial derivative w.r.t. Θ, $[\partial(\cdot)/\partial\Theta]$. Since $R^{1/2}(k)$ is upper triangular, the product $-R_\Theta^{-1/2}(k)R^{1/2}(k)$ is upper triangular as well. Denoting by $U(k)$ the upper triangular part of the matrix on the left side of Eq. (43) and by $D(k)$ its diagonal, we have

$$-R_\Theta^{-1/2}(k)R^{1/2}(k) = U(k) + \tfrac{1}{2}D(k) \tag{44}$$

Equation (44) is a triangular linear system that is easily solved for $R_\Theta^{-1/2}(k)$.

Summarizing the results so far, the components of the score are given by

$$\begin{aligned}\frac{\partial}{\partial\Theta}L\left[\Theta\,|\,\mathcal{Y}(N)\right] = &-\sum_{k=1}^{N}\Bigg\{\hat{q}^T(k|k-1)\frac{\partial\hat{q}(k|k-1)}{\partial\Theta} - \hat{q}^T(k|k)\frac{\partial\hat{q}(k|k)}{\partial\Theta} + y^T(k)R^{-T/2}(k)\frac{\partial R^{-T/2}(k)}{\partial\Theta}y(k)\\ &+\frac{1}{2}\operatorname{tr}\left[R^{-1}(k)\frac{\partial R(k)}{\partial\Theta}\right] + \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\lambda_i(k|k-1)}\frac{\partial\lambda_i(k|k-1)}{\partial\Theta} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\lambda_i(k|k)}\frac{\partial\lambda_i(k|k)}{\partial\Theta}\Bigg\}\\ = &\sum_{k=1}^{N}\Bigg\{\hat{q}^T(k|k-1)\frac{\partial\hat{q}(k|k-1)}{\partial\Theta} - \hat{q}^T(k|k)\frac{\partial\hat{q}(k|k)}{\partial\Theta} + y^T(k)R^{-T/2}(k)\frac{\partial R^{-T/2}(k)}{\partial\Theta}y(k)\\ &+\frac{1}{2}\operatorname{tr}\left[R^{-1}(k)\frac{\partial R(k)}{\partial\Theta}\right] + \frac{1}{2}\operatorname{tr}\left[\Lambda^{-1}(k|k-1)\frac{\partial\Lambda(k|k-1)}{\partial\Theta}\right] - \frac{1}{2}\operatorname{tr}\left[\Lambda^{-1}(k|k)\frac{\partial\Lambda(k|k)}{\partial\Theta}\right]\end{aligned} \tag{45}$$

where all variables involved are either directly available from the V-Lambda filter or are computed based on it via procedures that will be specified in the sequel.

Fisher Information Matrix

To compute the FIM, recall the definition

$$\mathfrak{M} := E\left\{\nabla_\Theta L\left[\Theta\,|\,\mathcal{Y}(N)\right]\nabla_\Theta^T L\left[\Theta\,|\,\mathcal{Y}(N)\right]\right\} \tag{46}$$

Since the score $\nabla_\Theta L[\Theta\,|\,\mathcal{Y}(N)]$ has zero mean, the FIM is the covariance of the score. The gradient of the log-likelihood function Eq. (3) can be written as the sum

$$\nabla_\Theta L\left[\Theta\,|\,\mathcal{Y}(N)\right] = \sum_{k=1}^{N}s(k|k-1) \tag{47}$$

where $s(k|k-1)$ is the conditional score, i.e., the gradient of the conditional log-likelihood function given the past measurement history $\mathcal{Y}(k-1)$:

$$s(k|k-1) = -\tfrac{1}{2}\nabla_\Theta\left[\tilde{y}^T(k|k-1)\mathfrak{N}^{-1}(k|k-1)\tilde{y}(k|k-1) + \log\det\mathfrak{N}(k|k-1)\right]$$

Using previous development, the components of the conditional score are given by

$$\begin{aligned}&-\frac{1}{2}\frac{\partial}{\partial\Theta}\left[\tilde{y}^T(k|k-1)\mathfrak{N}^{-1}(k|k-1)\tilde{y}(k|k-1) + \log\det\mathfrak{N}(k|k-1)\right] = \hat{q}^T(k|k-1)\frac{\partial\hat{q}(k|k-1)}{\partial\Theta}\\ &-\hat{q}^T(k|k)\frac{\partial\hat{q}(k|k)}{\partial\Theta} + y^T(k)R^{-T/2}(k)\frac{\partial R^{-T/2}(k)}{\partial\Theta}y(k) + \frac{1}{2}\operatorname{tr}\left[R^{-1}(k)\frac{\partial R(k)}{\partial\Theta}\right]\\ &+\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\lambda_i(k|k-1)}\frac{\partial\lambda_i(k|k-1)}{\partial\Theta} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\lambda_i(k|k)}\frac{\partial\lambda_i(k|k)}{\partial\Theta}\end{aligned} \tag{48}$$

Note that the FIM is the covariance of the sum Eq. (47) whose terms are uncorrelated. Following Ref. 21, we propose to estimate the FIM from the sample as follows:

$$\hat{\mathfrak{M}} = \sum_{k=1}^{N}s(k|k-1)s^T(k|k-1) - \frac{1}{N}\nabla_\Theta L\left[\Theta\,|\,\mathcal{Y}(N)\right]\nabla_\Theta^T L\left[\Theta\,|\,\mathcal{Y}(N)\right] \tag{49}$$

where $s(k|k-1)$ is computed via Eq. (48) and the score $\nabla_\Theta L[\Theta\,|\,\mathcal{Y}(N)]$ is computed in Eq. (45). This estimate of the FIM is asymptotically unbiased and consistent.[21]

Having obtained the expressions for the score and FIM in terms of the derivatives of the V-Lambda filter variables, these derivatives are developed next.

V. V-Lambda Sensitivity Equations

In this section we derive algorithms for the recursive computation of the derivatives of the V-Lambda filter variables, which are needed for the score computation, as shown in Sec. IV [Eq. (45)]. These include the derivatives of the eigenfactors and those of the state estimates.

Sensitivity Equation for the Time-Updated Eigenvalue Matrix

To obtain the sensitivity derivatives of the a priori eigenvalues, we use the covariance mode time-updated equation

$$\mathcal{C}(k-1) = V(k|k-1)\left[\Lambda^{1/2}(k|k-1) \quad 0\right]Z^T(k-1) \tag{50}$$

which represents the singular value decomposition of $\mathcal{C}(k-1)$. Assuming that all of the variables in Eq. (50) can be computed via the V-Lambda state estimator (based on the current estimate of Θ), and also that $\mathcal{A}_\Theta(k-1) := [\partial\mathcal{A}(k-1)/\partial\Theta]$ is available from previous computations, the problem is to compute the derivative of the eigenvalue matrix $\Lambda(k|k-1)$ w.r.t. Θ, $[\partial\Lambda(k|k-1)/\partial\Theta]$. Since the right side of Eq. (50) is the singular value decomposition of $\mathcal{A}(k-1)$, it follows that the elements of the diagonal matrix $\Lambda(k|k-1)$ are the eigenvalues of the matrix $\mathcal{A}(k-1)\mathcal{A}^T(k-1)$. The problem translates, therefore, to the computation of the derivatives of the eigenvalues of a symmetric matrix that depends on a parameter. To this end, the result summarized in the following theorem is needed.

Theorem 2: Let $P(s)$ be a symmetric $n \times n$ real matrix function that depends on the real parameter s, and let $P(s)$ possess continuous first derivatives for every $s \in \mathbb{R}$. Let $\{\lambda_i(s)\}_{i=1}^n$ be the eigenvalues of $P(s)$ (which need not be distinct), and let $V(s)$ be the orthogonal matrix whose columns are the corresponding eigenvectors $\{v_i(s)\}_{i=1}^n$ of $P(s)$. Then $\{\lambda_i(s)\}_{i=1}^n$ and $V(s)$ possess continuous first derivatives for every $s \in \mathbb{R}$; furthermore the following is true:

1)
$$\frac{d\lambda_i(s)}{ds} = v_i^T(s)\frac{dP(s)}{ds}v(s), \qquad i = 1,2,\ldots,n \tag{51}$$

2)
$$\frac{dv_i(s)}{ds} = \sum_{\substack{j=1\\ \lambda_j\neq\lambda_i}}^{n} \frac{v_j^T[dP(s)/ds]v_i}{\lambda_i - \lambda_j}v_j, \quad i = 1,2,\ldots,n \tag{52}$$

Proof: For the existence of analytic eigenvalue/eigenvector functions for every $s \in \mathbb{R}$ in the more general case of an analytic, self-adjoint matrix $P(s)$, see Ref. 22. The derivatives of the eigenfactors are developed in Ref. 23. □

Using Theorem 2, we can express the sensitivity derivatives of the eigenvalues as functions of the available factors as follows:

$$\frac{\partial\lambda_i(k|k-1)}{\partial\Theta} = v_i^T(k|k-1)\frac{\partial[\mathcal{A}(k-1)\mathcal{A}^T(k-1)]}{\partial\Theta}v_i(k|k-1) \tag{53}$$

where

$$\frac{\partial[\mathcal{A}(k-1)\mathcal{A}^T(k-1)]}{\partial\Theta} = \mathcal{A}(k-1)\mathcal{A}_\Theta^T(k-1) + [\mathcal{A}(k-1)\mathcal{A}_\Theta^T(k-1)]^T \tag{54}$$

Using Eq. (54) in Eq. (53) yields

$$\frac{\partial\lambda_i(k|k-1)}{\partial\Theta} = 2v_i^T(k|k-1)\mathcal{A}(k-1)\mathcal{A}_\Theta^T(k-1)v_i(k|k-1) \tag{55}$$

Employing Eq. (50), $\mathcal{A}(k-1)$ can be expressed as

$$\mathcal{A}(k-1) = \sum_{j=1}^{n}\lambda_j^{1/2}(k|k-1)v_j(k|k-1)Z_j^T(k-1) \tag{56}$$

where $Z_j(k-1)$ is the jth column of the matrix $Z(k-1)$ [notice that only the first n columns of the $(n+q)\times(n+q)$ matrix $Z(k-1)$ are involved]. Using the last expression for $\mathcal{A}(k-1)$, Eq. (55) becomes

$$\begin{aligned}\frac{\partial\lambda_i(k|k-1)}{\partial\Theta} &= 2v_i^T(k|k-1)\\ &\times\left[\sum_{j=1}^{n}\lambda_j^{1/2}(k|k-1)v_j(k|k-1)Z_j^T(k-1)\right]\\ &\times\mathcal{A}_\Theta^T(k-1)v_i(k|k-1)\\ &= 2\lambda_i^{1/2}(k|k-1)Z_i^T(k-1)\mathcal{A}_\Theta^T(k-1)v_i(k|k-1)\end{aligned} \tag{57}$$

where use was made of the eigenvectors orthonormality property. To simplify the notation in the sequel, define $\gamma_{ii}(k|k-1)$ as

$$\gamma_{ii}(k|k-1) := Z_i^T(k-1)\mathcal{A}_\Theta^T(k-1)v_i(k|k-1) \tag{58}$$

and let $\Gamma(k|k-1)$ be the diagonal matrix

$$\Gamma(k|k-1) := \operatorname{diag}[\gamma_{11}(k|k-1), \gamma_{22}(k|k-1), \ldots, \gamma_{nn}(k|k-1)] \tag{59}$$

Using these definitions, Eq. (57) becomes

$$\frac{\partial\Lambda(k|k-1)}{\partial\Theta} = 2\Lambda^{1/2}(k|k-1)\Gamma(k|k-1) \tag{60}$$

which is the sought for sensitivity equation for the time-updated eigenvalue matrix $\Lambda(k|k-1)$. Using this result in the expression for the log-likelihood gradient, the corresponding term in Eq. (45) is rewritten as

$$\frac{1}{2}\operatorname{tr}\left[\Lambda^{-1}(k|k-1)\frac{\partial\Lambda(k|k-1)}{\partial\Theta}\right] = \operatorname{tr}\left[\Lambda^{-1/2}(k|k-1)\Gamma(k|k-1)\right] \tag{61}$$

Note that all matrices involved here are diagonal, which renders the trace computation trivial. Note also that throughout the derivation $\mathcal{A}_\Theta(k-1)$ has been assumed known from previous calculations. We will return to the details of the explicit computation of this matrix derivative at a later point.

Sensitivity Equation for the Measurement-Updated Eigenvalue Matrix

Recall, from Sec. III, that the V-Lambda information mode measurement update can be written as

$$\mathcal{B}(k) = Y(k)\begin{bmatrix}\Lambda^{-1/2}(k|k)\\ 0\end{bmatrix}V^T(k|k) \tag{62}$$

which is the singular value decomposition of $\mathcal{B}(k)$. Obviously, $\Lambda^{-1}(k|k)$ is the eigenvalue matrix of $\mathcal{B}^T(k)\mathcal{B}(k)$, and the problem is, therefore, to compute the sensitivities of the eigenvalues of a symmetric matrix. Again, we use Theorem 2 to obtain

$$\frac{\partial\lambda_i^{-1}(k|k)}{\partial\Theta} = v_i^T(k|k)[\mathcal{B}^T(k)\mathcal{B}(k)]_\Theta v_i(k|k) \tag{63}$$

where

$$[\mathcal{B}^T(k)\mathcal{B}(k)]_\Theta = \mathcal{B}^T(k)\mathcal{B}_\Theta(k) + [\mathcal{B}^T(k)\mathcal{B}_\Theta(k)]^T \tag{64}$$

Using Eq. (64) in Eq. (63) yields

$$\frac{\partial\lambda_i^{-1}(k|k)}{\partial\Theta} = 2v_i^T(k|k)\mathcal{B}^T(k)\mathcal{B}_\Theta(k)v_i(k|k) \tag{65}$$

Employing Eq. (62), $\mathcal{B}^T(k)$ can be expressed as

$$\begin{aligned}\mathcal{B}^T(k) &= V(k|k)[\Lambda^{-1/2}(k|k) \quad 0]Y^T(k)\\ &= \sum_{j=1}^{n}\lambda_j^{-1/2}(k|k)v_j(k|k)Y_j^T(k)\end{aligned} \tag{66}$$

where $Y_j(k)$ is the jth column of the matrix $Y(k)$ [again, notice that only the first n columns of the $(n+m)\times(n+m)$

matrix are involved]. Using the last expression for $\mathcal{B}^T(k)$, Eq. (65) becomes

$$\frac{\partial \lambda_i^{-1}(k|k)}{\partial \Theta} = 2v_i^T(k|k)\left[\sum_{j=1}^{n} \lambda_j^{-\frac{1}{2}}(k|k)v_j(k|k)Y_j^T(k)\right] \times \mathcal{B}_\Theta(k)v_i(k|k) = 2\lambda_i^{-\frac{1}{2}}(k|k)Y_i^T(k)\mathcal{B}_\Theta(k)v_i(k|k) \quad (67)$$

where use was made of the eigenvectors orthonormality. Analogously to definitions Eqs. (58) and (59), we define

$$\gamma_{ii}(k|k) := Y_i^T(k)\mathcal{B}_\Theta(k)v_i(k|k) \quad (68)$$

$$\Gamma(k|k) := \operatorname{diag}\left[\gamma_{11}(k|k), \gamma_{22}(k|k), \ldots, \gamma_{nn}(k|k)\right] \quad (69)$$

Using these definitions, Eq. (67) can be written as

$$\frac{\partial \Lambda^{-1}(k|k)}{\partial \Theta} = 2\Lambda^{-\frac{1}{2}}(k|k)\Gamma(k|k) \quad (70)$$

from which we obtain

$$\frac{\partial \Lambda(k|k)}{\partial \Theta} = -2\Lambda^{3/2}(k|k)\Gamma(k|k) \quad (71)$$

Equation (71) is the sensitivity equation for the measurement-updated eigenvalue matrix $\Lambda(k|k)$. Using this result in the expression Eq. (45) for the log-likelihood gradient, the last term of that expression may be written as

$$-\frac{1}{2}\operatorname{tr}\left[\Lambda^{-1}(k|k)\frac{\partial \Lambda(k|k)}{\partial \Theta}\right] = \operatorname{tr}\left[\Lambda^{\frac{1}{2}}(k|k)\Gamma(k|k)\right] \quad (72)$$

Having derived the sensitivity equations for the eigenvalue matrices, we notice that it has been assumed in both the time-updated and the measurement-update sensitivity equations, respectively, that $\mathcal{A}_\Theta(k-1)$ and $\mathcal{B}_\Theta(k)$ are known from previous stages. Recalling Eqs. (10) and (17), we express these derivatives as follows:

$$\frac{\partial \mathcal{A}(k-1)}{\partial \Theta} = \left[\frac{\partial\left[F(k-1)V(k-1|k-1)\Lambda^{\frac{1}{2}}(k-1|k-1)\right]}{\partial \Theta} \times \frac{\partial\left[G(k-1)Q^{\frac{1}{2}}(k-1)\right]}{\partial \Theta}\right] \quad (73)$$

$$\frac{\partial \mathcal{B}(k)}{\partial \Theta} = \begin{bmatrix} \dfrac{\partial\left[\Lambda^{-\frac{1}{2}}(k|k-1)V^T(k|k-1)\right]}{\partial \Theta} \\ \dfrac{\partial\left[R^{-\frac{1}{2}}(k)H(k)\right]}{\partial \Theta} \end{bmatrix} \quad (74)$$

Examining both of these equations, we observe that to compute the derivatives $[\partial \mathcal{A}(k-1)/\partial \Theta]$ and $[\partial \mathcal{B}(k)/\partial \Theta]$ we have to use sensitivities that belong to the following four categories:

1) Derivatives of mathematical model matrices $[\partial F(k)/\partial \Theta]$, $[\partial G(k-1)/\partial \Theta]$, and $[\partial H(k)/\partial \Theta]$; these are specified as part of the parameter estimation problem.

2) Derivatives of the square roots of the noises covariance matrices $[\partial R^{-\frac{1}{2}}(k)/\partial \Theta]$ and $[\partial Q^{\frac{1}{2}}(k-1)/\partial \Theta]$; these are computed from the specified $[\partial R(k)/\partial \Theta]$ and $[\partial Q(k-1)/\partial \Theta]$ via the procedure outlined in Sec. IV.

3) Eigenvalue matrices derivatives $[\partial \Lambda(k|k)/\partial \Theta]$ and $[\partial \Lambda(k|k-1)/\partial \Theta]$; procedures for the computation of these derivatives were derived previously in this section.

4) Eigenvector matrices derivatives $[\partial V(k|k)/\partial \Theta]$ and $[\partial V(k|k-1)/\partial \Theta]$. These derivatives are needed for the computation of $[\partial \mathcal{A}(k-1)/\partial \Theta]$ and $[\partial \mathcal{B}(k)/\partial \Theta]$, though they do not appear explicitly in the log-likelihood gradient, Eq. (45).

Sensitivity equations for the eigenvector derivatives are derived next.

Sensitivity Equation for the Time-Updated Eigenvector Matrix

To compute $[\partial V(k|k-1)/\partial \Theta]$, we observe from the singular value decomposition Eq. (50) that $V(k|k-1)$ is the eigenvector matrix of $\mathcal{A}(k-1)\mathcal{A}^T(k-1)$. Therefore the problem is to obtain the sensitivities of the eigenvectors of the symmetric matrix $\mathcal{A}(k-1)\mathcal{A}^T(k-1)$ that depend on the parameter Θ. The solution to this problem is given in terms of Theorem 2, as follows.

Let $v_i(k|k-1)$ be an eigenvector corresponding to $\lambda_i(k|k-1)$ [the ith column of $V(k|k-1)$], then

$$\frac{\partial v_i(k|k-1)}{\partial \Theta} = \sum_{\substack{j=1 \\ \lambda_j \neq \lambda_i}}^{n} \frac{v_j^T(k|k-1)\left[\mathcal{A}(k-1)\mathcal{A}^T(k-1)\right]_\Theta v_i(k|k-1)}{\lambda_i - \lambda_j} \times v_j(k|k-1) \quad (75)$$

To simplify notation, define the functions $\omega_{ji}(k|k-1)$ as

$$\omega_{ji}(k|k-1) = \begin{cases} \dfrac{v_j^T(k|k-1)\left[\mathcal{A}(k-1)\mathcal{A}^T(k-1)\right]_\Theta v_i(k|k-1)}{\lambda_i - \lambda_j} & \lambda_i \neq \lambda_j \\ 0 & \lambda_i = \lambda_j \end{cases} \quad (76)$$

Using Eqs. (54) and (56) in Eq. (76), we obtain

$$v_j^T(k|k-1)\left[\mathcal{A}(k-1)\mathcal{A}^T(k-1)\right]_\Theta v_i(k|k-1) = \lambda_i^{\frac{1}{2}}(k|k-1)Z_i^T(k-1)\mathcal{A}_\Theta^T(k-1)v_j(k|k-1) + \lambda_j^{\frac{1}{2}}(k|k-1)Z_j^T(k-1)\mathcal{A}_\Theta^T(k-1)v_i(k|k-1) \quad (77)$$

Defining, analogously to Eq. (58) the functions $\gamma_{ij}(k|k-1)$ as

$$\gamma_{ij}(k|k-1) := Z_i^T(k-1)\mathcal{A}_\Theta^T(k-1)v_j(k|k-1) \quad (78)$$

and using this definition and Eq. (77) in Eq. (76), $\omega_{ji}(k|k-1)$ becomes

$$\omega_{ji}(k|k-1) = \begin{cases} \dfrac{\lambda_i^{\frac{1}{2}}(k|k-1)\gamma_{ij}(k|k-1) + \lambda_j^{\frac{1}{2}}(k|k-1)\gamma_{ji}(k|k-1)}{\lambda_i - \lambda_j} & \lambda_i \neq \lambda_j \\ 0 & \lambda_i = \lambda_j \end{cases} \quad (79)$$

Now let $\Omega(k|k-1)$ be the matrix whose (i,j) element is $\omega_{ij}(k|k-1)$. Then, using this notation, Eq. (75) can be rewritten as

$$\frac{\partial v_i(k|k-1)}{\partial \Theta} = \sum_{\substack{j=1 \\ \lambda_j \neq \lambda_i}}^{n} \omega_{ji}(k|k-1) v_j(k|k-1) \tag{80}$$

and the sought for derivative of the eigenvector matrix finally becomes

$$\frac{\partial V(k|k-1)}{\partial \Theta} = V(k|k-1)\Omega(k|k-1) \tag{81}$$

Sensitivity Equation for the Measurement-Updated Eigenvector Matrix

Following along the lines of the preceding derivation, we observe from Eq. (62) that the columns of $V(k|k)$ are the eigenvectors of the symmetric matrix $\mathcal{B}^T(k)\mathcal{B}(k)$. Hence, to compute the sensitivity $[\partial V(k|k)/\partial\Theta]$ we employ Theorem 2 to write

$$\frac{\partial v_i(k|k)}{\partial \Theta} = \sum_{\substack{j=1 \\ \lambda_j \neq \lambda_i}}^{n} \frac{v_j^T(k|k)\left[\mathcal{B}^T(k)\mathcal{B}(k)\right]_\Theta v_i(k|k)}{\lambda_i - \lambda_j} v_j(k|k) \tag{82}$$

Define $\omega_{ji}(k|k)$ to be

$$\omega_{ji}(k|k) = \begin{cases} \dfrac{v_j^T(k|k)\left[\mathcal{B}^T(k)\mathcal{B}(k)\right]_\Theta v_i(k|k)}{\lambda_i - \lambda_j} & \lambda_i \neq \lambda_j \\ 0 & \lambda_i = \lambda_j \end{cases} \tag{83}$$

Using Eqs. (64) and (66) in Eq. (83), we obtain

$$\begin{aligned} &v_j^T(k|k)\left[\mathcal{B}^T(k)\mathcal{B}(k)\right]_\Theta v_i(k|k) \\ &= \lambda_i^{-1/2}(k|k)Y_i^T(k)\mathcal{B}_\Theta(k)v_j(k|k) \\ &+ \lambda_j^{-1/2}(k|k)Y_j^T(k)\mathcal{B}_\Theta(k)v_i(k|k) \end{aligned} \tag{84}$$

Define the functions $\gamma_{ij}(k|k)$ as

$$\gamma_{ij}(k|k) := Y_j^T(k)\mathcal{B}_\Theta(k)v_i(k|k) \tag{85}$$

Then, using Eqs. (84) and (85) in Eq. (83), $\omega_{ji}(k|k)$ becomes

$$\omega_{ji}(k|k) = \begin{cases} \dfrac{\lambda_i^{-1/2}(k|k)\gamma_{ij}(k|k) + \lambda_j^{-1/2}(k|k)\gamma_{ji}(k|k)}{\lambda_i - \lambda_j} & \lambda_i \neq \lambda_j \\ 0 & \lambda_i = \lambda_j \end{cases} \tag{86}$$

Letting $\Omega(k|k)$ be the matrix whose entries are the functions $\omega_{ij}(k|k)$ Eq. (82) can be rewritten as

$$\frac{\partial v_i(k|k)}{\partial \Theta} = \sum_{\substack{j=1 \\ \lambda_j \neq \lambda_i}}^{n} \omega_{ji}(k|k) v_j(k|k) \tag{87}$$

or, in matrix form,

$$\frac{\partial V(k|k)}{\partial \Theta} = V(k|k)\Omega(k|k) \tag{88}$$

which is the required sensitivity equation for the measurement-updated eigenvector matrix.

Having obtained the sensitivity equations for the eigenfactors, we finally derive the corresponding equations for the state estimates.

Sensitivity Equation for the A Priori Normalized State Estimate

Rewriting Eq. (23) as

$$\begin{aligned} \hat{q}(k|k-1) &= \Lambda^{-1/2}(k|k-1)V^T(k|k-1)F(k-1) \\ &\times V(k-1|k-1)\Lambda^{1/2}(k-1|k-1)\hat{q}(k-1|k-1) \end{aligned} \tag{89}$$

differentiating w.r.t. Θ and using the sensitivity equations [Eqs. (60), (71), (81), and (88)] for the a priori and a posteriori eigenfactors yields

$$\begin{aligned} \frac{\partial \hat{q}(k|k-1)}{\partial \Theta} &= \Big[-\Lambda^{-1}(k|k-1)\Gamma(k|k-1)V^T(k|k-1) \\ &\times F(k-1)V(k-1|k-1)\Lambda^{1/2}(k-1|k-1) \\ &- \Lambda^{-1/2}(k|k-1)\Omega(k|k-1)V^T(k|k-1)F(k-1) \\ &\times V(k-1|k-1)\Lambda^{1/2}(k-1|k-1) \\ &+ \Lambda^{-1/2}(k|k-1)V^T(k|k-1)F_\Theta(k-1)V(k-1|k-1) \\ &\times \Lambda^{1/2}(k-1|k-1) + \Lambda^{-1/2}(k|k-1)V^T(k|k-1) \\ &\times F(k-1)V(k-1|k-1)\Omega(k-1|k-1)\Lambda^{1/2}(k-1|k-1) \\ &- \Lambda^{-1/2}(k|k-1)V^T(k|k-1)F(k-1)V(k-1|k-1) \\ &\times \Lambda(k-1|k-1)\Gamma(k-1|k-1)\Big]\hat{q}(k-1|k-1) \\ &+ \Lambda^{-1/2}(k|k-1)V^T(k|k-1)F(k-1)V(k-1|k-1) \\ &\times \Lambda^{1/2}(k-1|k-1)\hat{q}_\Theta(k-1|k-1) \end{aligned} \tag{90}$$

where use was made of the fact that $\Omega(k|k-1)$ is a skew-symmetric matrix [see Eq. (79)]; $F_\Theta(k-1)$ is given in the problem definition, and $\hat{q}_\theta(k-1|k-1)$ is computed via the a posteriori sensitivity equation that is derived next.

Sensitivity Equation for the A Posteriori Normalized State Estimate

To obtain the equation for $[\partial\hat{q}(k|k)/\partial\Theta]$, we rewrite Eq. (62) as

$$\mathcal{B}^T(k)Y(k) = \left[V(k|k)\Lambda^{-1/2}(k|k) \quad 0\right] \tag{91}$$

Multiplying Eq. (91) by Eq. (22) yields

$$\mathcal{B}^T(k)b(k) = V(k|k)\Lambda^{-1/2}(k|k)\hat{q}(k|k) \tag{92}$$

where use was made of the orthogonality of $Y(k)$. Taking derivative w.r.t. Θ and using the expressions (71) and (88) for the eigenfactors sensitivities, we obtain, after some manipulation,

$$\begin{aligned} \frac{\partial \hat{q}(k|k)}{\partial \Theta} &= \Lambda^{1/2}(k|k)V^T(k|k)\mathcal{B}_\Theta(k)b(k) \\ &+ \Lambda^{1/2}(k|k)V^T(k|k)\mathcal{B}^T(k)b_\Theta(k) \\ &- \Lambda^{1/2}(k|k)\left[\Gamma(k|k) + \Omega(k|k)\Lambda^{-1/2}(k|k)\right]\hat{q}(k|k) \end{aligned} \tag{93}$$

where

$$b_\Theta(k) = \begin{bmatrix} \dfrac{\partial \hat{q}(k|k-1)}{\partial \Theta} \\ \dfrac{\partial R^{-\frac{1}{2}}(k)}{\partial \Theta} y(k) \end{bmatrix} \tag{94}$$

Now rewrite Eq. (91) as

$$\Lambda^{\frac{1}{2}}(k|k)V^T(k|k)\mathcal{B}^T(k) = \begin{bmatrix} I & 0 \end{bmatrix} Y^T(k) = \tilde{Y}^T(k) \tag{95}$$

where $\tilde{Y}(k) \in \mathcal{R}^{n+m,n}$ is the matrix containing the first n columns of $Y(k)$. Using the last result in Eq. (93), we obtain the final form of the sensitivity equation:

$$\frac{\partial \hat{q}(k|k)}{\partial \Theta} = \Lambda^{\frac{1}{2}}(k|k)V^T(k|k)\mathcal{B}_\Theta(k)b(k) + \tilde{Y}^T(k)b_\Theta(k)$$
$$- \Lambda^{\frac{1}{2}}(k|k)\left[\Gamma(k|k) + \Omega(k|k)\Lambda^{-\frac{1}{2}}(k|k)\right]\hat{q}(k|k) \tag{96}$$

Having obtained the sensitivity equation for the normalized state estimate, the new, square root, maximum likelihood algorithm is completed. For the reader's convenience, it is summarized in the next section.

VI. Implementation Summary of the Maximum Likelihood Algorithm

Log-Likelihood Gradient Computation

The log-likelihood gradient is given in Eq. (45). The following procedure summarizes the algorithm for computing the V-Lambda variables and sensitivity derivatives appearing in that expression.

V-Lambda Variables: $\Lambda(k|k-1)$, $\hat{q}(k|k-1)$, *and* $\hat{q}(k|k)$

These quantities are computed recursively via the V-Lambda state estimator outlined in Sec. III.

Eigenvalue Sensitivity Derivatives: $[\partial\Lambda(k|k-1)/\partial\Theta]$ and $[\partial\Lambda(k|k)/\partial\Theta]$

At each time step k:

1) Using the values of $[\partial V(k-1|k-1)/\partial\Theta]$, $[\partial\Lambda(k-1|k-1)/\partial\Theta]$, and $[\partial Q^{\frac{1}{2}}(k-1)/\partial\Theta]$ that were computed at the previous time step $(k-1)$, compute $\mathcal{C}_\Theta(k-1)$ via Eq. (73).

2) Using $\mathcal{C}_\Theta(k-1)$, compute $[\partial\Lambda(k|k-1)/\partial\Theta]$ via Eq. (60).

3) Using $\mathcal{C}_\Theta(k-1)$, compute $[\partial V(k|k-1)/\partial\Theta]$ via Eq. (81).

4) Compute $[\partial R^{-\frac{1}{2}}(k)/\partial\Theta]$ by solving the triangular system, Eq. (44).

5) Using $[\partial V(k|k-1)/\partial\Theta]$, $[\partial\Lambda(k|k-1)/\partial\Theta]$ and $[\partial R^{-\frac{1}{2}}(k)/\partial\Theta]$, compute $\mathcal{B}_\Theta(k)$ via Eq. (74).

6) Using $\mathcal{B}_\Theta(k)$, compute $[\partial\Lambda(k|k)/\partial\Theta]$ via Eq. (71).

7) Using $\mathcal{B}_\Theta(k)$, compute $[\partial V(k|k)/\partial\Theta]$ via Eq. (88). Also, solving a triangular system analogous to Eq. (44), compute $[\partial Q^{\frac{1}{2}}(k)/\partial\Theta]$. These two sensitivity derivatives are needed for the computation of $\mathcal{C}_\Theta(k)$ at the following time step (see stage 1).

Normalized State Sensitivity Derivatives: $[\partial\hat{q}(k|k-1)/\partial\Theta]$ and $[\partial\hat{q}(k|k)/\partial\Theta]$

1) Using $\hat{q}_\theta(k-1|k-1)$ (which is available from the previous time step), $[\partial\hat{q}(k|k-1)/\partial\Theta]$ is computed via Eq. (90).

2) The $b_\Theta(k)$ is computed via Eq. (94) using $[\partial R^{-\frac{1}{2}}(k)/\partial\Theta]$ and $[\partial\hat{q}(k|k-1)/\partial\Theta]$.

3) Using $b_\Theta(k)$ and $\mathcal{B}_\Theta(k)$, $[\partial\hat{q}(k|k)/\partial\Theta]$ is computed via Eq. (96).

Fisher Information Matrix Computation

The FIM is estimated from the sample via Eq. (49), where the conditional scores are computed in Eq. (48) and the score is given in Eq. (45). Note that all of the variables needed for the FIM have already been computed for the score.

VII. Conclusions

The maximum likelihood (ML) parameter estimation method is implemented in this paper using a hybrid (covariance/information) V-Lambda square root state estimator. Derivatives of this filter with respect to the parameter vector are developed that comprise a complete set of square root sensitivity equations. Based on the V-Lambda filter variables and their derivatives, formulas are derived for the computation of the log-likelihood function, its gradient, and the Fisher information matrix.

Compared to the SRIF-based ML scheme proposed by Bierman et al.,[16] the new algorithm is more computationally expensive, relying mainly on the SVD procedure. However, in view of the excellent numerical characteristics of the SVD-based V-Lambda filter, the new scheme should be numerically superior in stability and accuracy. Moreover, the direct utilization of the covariance spectral factors adds physical insight into the estimation process as an additional merit of the proposed method.

Acknowledgments

This research was supported by the Technion V.P.R. Fund—Seniel Ostrow Research Fund.

References

[1] Astrom, K. J., and Kallstrom, C. G., "Identification of Ship Steering Dynamics," *Automatica,* Vol. 12, 1976, pp. 9–22.

[2] Mehra, R. K., and Tyler, J. S., "Case Studies in Aircraft Parameter Identification," *Proceedings of the 3rd IFAC Symposium on Identification and System Parameter Estimation,* North Holland, Amsterdam, The Netherlands, 1973.

[3] Iliff, K. W., "Parameter Estimation for Flight Vehicles," *Journal of Guidance, Control, and Dynamics,* Vol. 12, No. 5, 1989, pp. 609–622.

[4] Astrom, K. J., "Maximum Likelihood and Prediction Error Methods," *Automatica,* Vol. 16, 1980, pp. 551–574.

[5] Mendel, J. M., *Lessons in Digital Estimation Theory,* Prentice-Hall, Englewood Cliffs, NJ, 1987.

[6] Gupta, N. K., and Mehra, R. K., "Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations," *IEEE Transactions on Automatic Control,* Vol. AC-19, No. 6, 1974, pp. 774–783.

[7] Bierman, G. J., *Factorization Methods for Discrete Sequential Estimation,* Academic, New York, 1977.

[8] Laub, A. J., "Numerical Linear Algebra Aspects of Control Design Computations," *IEEE Transactions on Automatic Control,* Vol. AC-30, No. 2, 1985, pp. 97–108.

[9] Bellantoni, J. F., and Dodge, K. F., "A Square Root Formulation of the Kalman-Schmidt Filter," *AIAA Journal,* Vol. 5, No. 7, 1967, pp. 1309–1314.

[10] Dyer, P., and McReynolds, S., "Extension of Square Root Filtering to Include Process Noise," *Journal of Optimization Theory and Applications,* Vol. 3, No. 6, 1969, pp. 444–459.

[11] Bierman, G. J., and Thornton, C. L., "Numerical Comparison of the Kalman Filter Algorithms: Orbit Determination Case Study," *Automatica,* Vol. 13, 1977, pp. 23–35.

[12] Bierman, G. J., "Measurement Updating Using the U-D Factorization," *Automatica,* Vol. 12, July 1976, pp. 375–382.

[13] Thornton, C. L., and Bierman, G. J., "Gram-Schmidt Algorithms for Covariance Propagation," *Proceedings of the IEEE Conference on Decision and Control,* Inst. of Electrical and Electronics Engineers, New York, 1975, pp. 489–498.

[14] Oshman, Y., and Bar-Itzhack, I. Y., "Square Root Filtering via Covariance and Information Eigenfactors," *Automatica,* Vol. 22, No. 5, 1986, pp. 599–604.

[15]Oshman, Y., "Gain-Free Square Root Information Filtering Using the Spectral Decomposition," *Journal of Guidance, Control, and Dynamics,* Vol. 12, No. 5, 1989, pp. 681–690.

[16]Bierman, G. J., Belzer, M. R., Vandergraft, J. S., and Porter, D. W., "Maximum Likelihood Estimation Using Square Root Information Filters," *Proceedings of the American Control Conference,* Institute of Electrical and Electronics Engineers, Piscataway, NJ, 1989, pp. 2646–2652.

[17]Lawson, C. L., and Hanson, R. J., *Solving Least Squares Problems,* Prentice-Hall, Englewood Cliffs, NJ, 1974.

[18]Ham, F. M., and Brown, R. G., "Observability, Eigenvalues, and Kalman Filtering," *IEEE Transactions on Aerospace and Electrical Systems,* Vol. AES-19, No. 2, 1983, pp. 269–273.

[19]Golub, G. H., and Van Loan, C. F., *Matrix Computations,* Johns Hopkins Univ. Press, Baltimore, MD, 1983.

[20]Jover, J. M., and Kailath, T., "A Parallel Architecture for Kalman Filter Measurement Update nad Parameter Estimation," *Automatica,* Vol. 22, No. 1, 1986, pp. 43–57.

[21]Segal, M., and Weinstein, E., "A New Method for Evaluating the Log-Likelihood Gradient, the Hessian, and the Fisher Information Matrix for Linear Dynamic Systems," *IEEE Transactions on Information Theory,* Vol. 35, May 1989, pp. 682–687.

[22]Gohberg, I., Lancaster, P., and Rodman, L., *Matrix Polynomials,* Academic, New York, 1982.

[23]Oshman, Y., and Bar-Itzhack, I. Y., "Eigenfactor Solution of the Matrix Riccati Equation—A Continuous Square Root Algorithm," *IEEE Transactions on Automatic Control,* Vol. AC-30, No. 10, 1985, pp. 971–978.