EFFICIENT ESTIMATION OF INITIAL-CONDITION PARAMETERS FOR PARTIALLY OBSERVABLE INITIAL CONDITIONS

M.D. Shuster, senior member IEEE
D.W. Porter, member IEEE

BUSINESS AND TECHNOLOGICAL SYSTEMS, INC.
Aerospace Building, Suite 440
10210 Greenbelt Road
Seabrook, Maryland 20706

Twenty-Third Conference on Decision and Control
Las Vegas, Nevada - December 12-14, 1984

EFFICIENT ESTIMATION OF INITIAL-CONDITION PARAMETERS FOR PARTIALLY OBSERVABLE INITIAL CONDITIONS

M.D. Shuster, senior member IEEE
D.W. Porter, member IEEE

BUSINESS AND TECHNOLOGICAL SYSTEMS, INC.
Aerospace Building, Suite 440
10210 Greenbelt Road
Seabrook, Maryland 20706

ABSTRACT

Efficient and numerically well-conditioned scoring algorithms are presented for the maximum-likelihood estimation of initial means and covariances from an ensemble of tests when the initial condition is not observable per test. These algorithms take account also of the possibility that the estimated initial covariance may be singular. A sufficient statistic is used to reduce the computational burden and singular-value-decomposition and square root techniques are used to increase the numerical accuracy of the algorithm.

INTRODUCTION

In previous reports [1,2] efficient maximum-likelihood estimators were developed for the simultaneous estimation of system dynamical parameters (time constants and power spectral densities of process noise) and unknown initial means and covariances. The maximum-likelihood estimate in that work was calculated by means of a recursive approximation to scoring for which the Fisher information matrix and the likelihood gradient must be computed. The use of a sufficient statistic [3,4] led to considerable savings in these computations and the two-tier filter of Friedland [5] and Bierman [6] provided an efficient means for computing the sufficient statistic.

That work assumed, however, that the initial condition was observable in every test. In practical cases, unfortunately, this assumption may not hold and the initial condition may not be observable in any single test although the initial mean and covariance (the initial-condition parameters) are observable cumulatively over the entire ensemble of tests. It is this more general case which we treat in the present report.

When the initial condition is not observable per-test the Fisher information matrix for the sufficient statistic, $P_{ML}^{-1}(j)$ becomes singular and expressions derived earlier [1,2] in terms of $P_{ML}(j)$ are no longer meaningful. In addition, in many practical cases some components of the initial mean are not well identified and this is usually accompanied by the estimated initial covariance Σ having negative eigenvalues or, if constrained, being singular.

To obtain an expression for the Fisher information matrix and likelihood gradient in these circumstances we must start with a likelihood function which is valid then and obtain expressions for these quantities which demand the inversion neither of $P_{ML}^{-1}(j)$ nor of Σ. These expressions, though mathematically correct, will be numerically poorly-conditioned due to the loss of significance which accompanies operations with singular matrices. The solution to this problem is to work in a space of smaller dimension where all matrices have full rank. Such a reduced space is provided by the singular-value decomposition (SVD). The SVD requires some care, however, since the expressions for the Fisher information matrix for the initial-condition parameters contain derivatives and the SVD is not always differentiable. The direct differentiation of the SVD can be avoided, fortunately, and the SVD representation provides other computational advantages which can be exploited to bypass the need to compute $P_{ML}^{-1}(j)$ altogether.

Section 2 of this report presents the model for the initial condition and derives expressions for the likelihood gradient for the situation where the initial condition may be unobservable per-test and the estimated initial covariance matrix may be singular. Section 3 uses these results to obtain numerically well-behaved expressions with the aid of the SVD. Section 4 presents a more accurate means for computing the SVD quantities which bypasses the computation of the singular $P_{ML}^{-1}(j)$. The conclusions are stated in Section 5.

ESTIMATION OF INITIAL-CONDITION PARAMETERS

As in Refs. 1 and 2 we assume the system to be described by the standard Gaussian linear system model with unknown dynamical parameters subject to the initial condition

$$\underline{x}_0(j) = \underline{a}(j) + C_j\, \underline{b}(j) \quad , \tag{2-1}$$

where $\underline{x}_0(j)$ is the initial state vector for test j , $\underline{a}(j)$ has vanishing mean and covariance a known function of possibly unknown dynamical parameters α , and $\underline{b}(j)$ is sampled from a Gaussian distribution of unknown mean $\underline{\mu}_b$ and covariance Σ_b . The matrix C_j is assumed known. The two contributions $\underline{a}(j)$ and $\underline{b}(j)$ are assumed to be uncorrelated. The initial condition parameters to be estimated are $\underline{\mu}_b$ and Σ_b.

In general, $\underline{b}(j)$ will not contribute to every component of $\underline{x}_0(j)$ so that C_j will have some rows which vanish. Define

$$\underline{x}_0'(j) \equiv M_j\, \underline{b}(j) \quad , \tag{2-2}$$

where M_j is obtained by deleting the vanishing rows from C_j . If $\underline{Y}(j)$ denotes the stacked vector of measurements for test j then we may write

$$\underline{Y}(j) = H(j)\, \underline{x}_0(j) + \underline{\gamma}(j) \quad , \tag{2-3}$$

where $\underline{\gamma}(j)$ has vanishing mean and covariance $\Gamma(j)$. From eqs. (2-1) through (2-3) this may be rewritten as

$$\underline{Y}(j) = H'(j)\, \underline{x}_0'(j) + H(j)\, \underline{a}(j) + \underline{\gamma}(j)$$

$$= H'(j)\, \underline{x}_0'(j) + \underline{\gamma}'(j) \quad , \tag{2-4}$$

where $H'(j)$ is formed from $H(j)$ by deleting appropriate columns and $\underline{\gamma}'(j)$ has vanishing mean and covariance $\Gamma'(j)$.

Following a program similar to that of Levy et al. [3,4] we obtain an expression for the cumulative negative-log-likelihood function when Σ_b is singular and $\underline{x}_0'(j)$ is not completely observable, namely,

$$J(\underline{Y}(1), \underline{Y}(2), \ldots | \underline{\mu}_b, \Sigma_b, \underline{\alpha})$$
$$= \frac{1}{2} \sum_j \{ (\hat{\underline{x}}'_{0ML}(j) - M_j \underline{\mu}_b)^T$$
$$\cdot [I + (P'_{ML}(j))^{-1} M_j \Sigma_b M_j^T]^{-1}$$
$$\cdot (P'_{ML}(j))^{-1} (\hat{\underline{x}}'_{0ML}(j) - M_j \underline{\mu}_b) \tag{2-5}$$
$$+ \log \det [I + (P'_{ML}(j))^{-1} M_j \Sigma_b M_j^T] \}$$
$$+ \text{terms independent of } \underline{\mu}_b \text{ and } \Sigma_b \; ,$$

where

$$[P'_{ML}(j)]^{-1} = H'(j)^T (\Gamma'(j))^{-1} H'(j) \tag{2-6}$$

and $\hat{\underline{x}}'_{0ML}(j)$ a solution of

$$[P'_{ML}(j)]^{-1} \hat{\underline{x}}'_{0ML}(j)$$
$$= H'(j)^T (\Gamma'(j))^{-1} \underline{Y}(j) \quad . \tag{2-7}$$

From eq. (2-5) it is clear that $\hat{\underline{x}}'_{0ML}(j)$ is a sufficient statistic for $\underline{\mu}_b$ and Σ_b and from eqs. (2-4), (2-6) and (2-7) $\hat{\underline{x}}'_{0ML}(j)$ is the maximum-likelihood estimate of $\underline{x}_0(j)$. Note that if $\underline{x}_0'(j)$ is not observable then $[P'_{ML}(j)]^{-1}$ will be singular and $\hat{\underline{x}}'_{0ML}(j)$ will not be unique. Equation (2-5), however, holds for any solution for $\hat{\underline{x}}'_{0ML}(j)$. Typically, one chooses the minimum-length solution for definiteness.

Equation (2-5) may be transformed to

$$J = \frac{1}{2} \sum_j \{ (\hat{\underline{b}}_{ML}(j) - \underline{\mu}_b)^T [I + P^{-1}_{bML}(j) \Sigma_b]^{-1}$$
$$\cdot P^{-1}_{bML}(j) (\hat{\underline{b}}_{ML}(j) - \underline{\mu}_b)$$
$$+ \log \det [I + P^{-1}_{bML}(j) \Sigma_b] \} \tag{2-8}$$
$$+ \text{terms independent of } \underline{\mu}_b \text{ and } \Sigma_b \; ,$$

where

$$P^{-1}_{bML}(j) = M_j^T (P'_{ML}(j))^{-1} M_j \tag{2-9}$$

and $\hat{\underline{b}}_{ML}(j)$ is a solution of

$$P^{-1}_{bML}(j)\, \hat{\underline{b}}_{ML}(j) = M_j^T H'(j)^T (\Gamma'(j))^{-1} \underline{Y}(j) \; . \tag{2-10}$$

Clearly $\hat{\underline{b}}_{ML}(j)$ also is a sufficient statistic for $\underline{\mu}_b$ and Σ_b.

The estimation of $\underline{\mu}_b$ and Σ_b is straightforward. As before [1, 2], we assume that $H'(j)$ does not depend on unknown dynamical parameters. For many systems of interest (e.g., inertial navigation, orbit determination) this is often the case. With this assumption, the Fisher information matrix has the simple form

$$F = \left[\begin{array}{c|c|c} F_{\mu_b \mu_b} & 0 & 0 \\ \hline 0 & F_{\Sigma_b \Sigma_b} & F_{\Sigma_b \alpha} \\ \hline 0 & F_{\alpha \Sigma_b} & F_{\alpha\alpha} \end{array} \right] \tag{2-11}$$

with

$$F_{\mu_{bm} \mu_{bn}} = \sum_j D_{mn}(j) \quad , \tag{2-12}$$

$$F_{\Sigma_{bmn} \Sigma_{bpq}} = C_{mn} C_{pq} \sum_j [D_{mp}(j) D_{nq}(j) \tag{2-13}$$
$$+ D_{mq}(j) D_{np}(j)] \; ,$$

$$F_{\Sigma_{bmn} \alpha_\ell} = - C_{mn} \sum_j \frac{\partial}{\partial \alpha_\ell} D_{mn}(j) \quad , \tag{2-14}$$

where

$$C_{mn} = \begin{cases} 1/2 & m = n \\ 1 & m \neq n \end{cases} \tag{2-15}$$

and from eq. (2-8)

$$D(j) = [I + P^{-1}_{bML}(j) \Sigma_b]^{-1} P^{-1}_{bML}(j) \quad . \tag{2-16}$$

The submatrix $F_{\alpha\alpha}$ must be calculated separately from the other submatrices using the methods of [1,2]. For the special case when $P^{-1}_{bML}(j)$ is invertible eq. (2-16) reduces to

$$D(j) = [\Sigma_b + P_{bML}(j)]^{-1} \tag{2-17}$$

which appeared in the previous work. It might seem that eq. (2-16) can be obtained by a simple matrix operation from eq. (2-17). This, however, is not the case since eq. (2-17) does not hold under the present circumstances.

The likelihood gradient with respect to initial-condition parameters has the form

$$\frac{\partial J}{\partial \underline{\mu}_b} = - \sum_j D(j) \, (\hat{\underline{b}}_{ML}(j) - \underline{\mu}_b) \quad , \tag{2-18}$$

$$\frac{\partial J}{\partial \Sigma_{bmn}} = - C_{mn} \sum_j [D(j) \, (\hat{\underline{b}}_{ML}(j) - \underline{\mu}_b) \cdot (\hat{\underline{b}}_{ML}(j) - \underline{\mu}_b)^T D(j) - D(j)]_{mn} \, . \tag{2-19}$$

Note that the likelihood gradients above have a unique value even though $\underline{b}_{ML}(j)$ may not be unique.

The expression for $D(j)$ given by eq. (2-16) while simple in appearance may be poorly conditioned. While $D(j)$ is necessarily symmetric the first matrix factor in eq. (2-16) is not and may be the inverse of a nearly singular matrix. Therefore, we must seek a numerically superior method of calculating the Fisher information matrix and likelihood gradients.

SVD FORMULATION

The problem of lost significance in computing the Fisher information matrix and the likelihood gradients can be greatly reduced through the use of a singular-value decomposition (SVD) [7]. Thus, one may write

$$[P'_{ML}(j)]^{-1} = U_j \left[\begin{array}{c|c} \Delta^{-1}(j) & 0 \\ \hline 0 & 0 \end{array} \right] U_j^T \quad , \tag{3-1}$$

where U_j is orthogonal and $\Delta(j)^{-1}$ is diagonal and positive definite. (We express all quantities initially in terms of $(P'_{ML}(j))^{-1}$ because it is this quantity rather than $P_{bML}^{-1}(j)$ which is supplied by the Kalman Filter [1, 5, 6]. If U_j is partitioned as

$$U_j = [U_{1j} \vdots U_{2j}] \quad , \tag{3-2}$$

where U_{1j} has the same rank as $\Delta^{-1}(j)$, then it is simple to show by a transformation of variables in eq. (2-8) that

$$\hat{\underline{\xi}}_{ML}(j) = U_{1j}^T M_j \hat{\underline{b}}_{ML}(j) \equiv V_j \hat{\underline{b}}_{ML}(j) \tag{3-3}$$

is a sufficient statistic for $\underline{\mu}_b$ and Σ_b. By appropriately transforming the results of Section 2, we arrive at identical expressions for the submatrices of the Fisher information matrix in terms of the matrix $D(j)$ where now

$$D(j) = V_j^T [V_j \Sigma_b V_j^T + \Delta(j)]^{-1} V_j \equiv V_j^T d(j) V_j \tag{3-4}$$

and the likelihood gradient takes the form

$$\frac{\partial J}{\partial \mu_b} = - \sum_j V_j^T d(j) \, (\hat{\underline{\xi}}_{ML}(j) - V_j \underline{\mu}_b) \quad , \tag{3-5}$$

$$\frac{\partial J}{\partial \Sigma_{bmn}} = - C_{mn} \sum_j \{V_j^T [d(j)(\hat{\underline{\xi}}_{ML}(j) - V_j \underline{\mu}_b) (\hat{\underline{\xi}}_{ML}(j) - V_j \underline{\mu}_b)^T d(j) - d(j)] V_j\}_{mn} \quad . \tag{3-6}$$

The vector $\hat{\underline{\xi}}_{ML}(j)$ is unique since V_j annihilates the ambiguous component of $\hat{\underline{b}}_{ML}(j)$.

The computation of $F_{\Sigma_b \alpha}$ requires special care since the factors of the SVD of a matrix are not necessarily separately differentiable. However, as a consequence of the model we note that every null vector of $[P'^{-1}_{ML}(j)]$ is also a null vector of $\frac{\partial}{\partial \alpha_\ell} [P'_{ML}(j)]^{-1}$. Therefore, we can write

$$\frac{\partial}{\partial \alpha_\ell} P_{bML}^{-1}(j) = V_j^T \Delta_\ell^{-1}(j) \, V_j \quad , \tag{3-7}$$

where

$$\Delta_\ell^{-1}(j) \equiv U_{1j}^T \left\{ \frac{\partial}{\partial \alpha_\ell} [P'_{ML}(j)]^{-1} \right\} U_{1j} \quad . \tag{3-8}$$

The matrix $\Delta_\ell^{-1}(j)$ is symmetric but otherwise has no special properties. Note that $\Delta_\ell^{-1}(j)$ is <u>not</u> a component of the singular value decomposition of $\frac{\partial}{\partial \alpha_\ell} [P'_{ML}(j)]^{-1}$ and need not be either diagonal or positive definite. Evaluation of the derivative appearing in eq. (2-14) yields directly

$$\frac{\partial}{\partial \alpha_\ell} D(j) = V_j^T d(j) \, \Delta(j) \, \Delta_\ell^{-1}(j) \, \Delta(j) \, d(j) \, V_j \, . \tag{3-9}$$

Except for V_j and its transpose only symmetric non singular matrices now appear in the computation of the Fisher information matrix and the likelihood gradients.

COMPUTATION OF SVD QUANTITIES

The matrices U_j, $\Delta(j)^{-1}$, and $\Delta_\ell(j)^{-1}$ may be computed more efficiently by avoiding the computation of $[P'^{-1}_{ML}(j)]$ altogether. In actual practice $[P'_{ML}(j)]^{-1}$ is computed from the output of a bias-free Kalman filter [1,5,6]. Assume without loss of generality that the measurements are scalar and define

$$G_{k,j} \equiv B_{Ik,j}^{-1/2} H_{k,j} T'_{k,j} \, , \quad k=1,\ldots,n_j \quad , \tag{4-1}$$

where $H_{k,j}$ is the measurement sensitivity matrix, $H_{k,j} T'_{k,j}$ is the sensitivity of the bias-free residuals to $\underline{x}_0(j)$ and $B_{Ik,j}$ is the bias-free residual variance. The number of (scalar) measurements in test j is n_j.

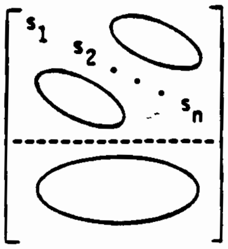
Then

$$[P'_{ML}(j)]^{-1} = \sum_{k=1}^{n_j} G_{k,j}^T G_{k,j} = G_j^T G_j \quad , \tag{4-2}$$

with $G_j \equiv [G_{1,j}^T, G_{2,j}^T, \ldots, G_{n,j}^T]^T$. The SVD of G_j is simply

$$G_j = \Psi_{\ell j}\, S_G(j)\, \Psi_{rj}^T \quad , \qquad (4\text{-}3)$$

where $\Psi_{\ell j}$ and Ψ_{rj} are orthogonal and

$S_G =$  . (4-4)

It follows directly that

$$U_j = \Psi_{rj} \quad , \qquad (4\text{-}5)$$

$$\left[\begin{array}{c|c} \Delta^{-1}(j) & 0 \\ \hline 0 & 0 \end{array}\right] = S_G^T(j)\, S_G(j) \quad , \qquad (4\text{-}6)$$

and the singular values of $[P'_{ML}(j)]^{-1}$ are s_i^2, $i=1,\ldots,n$. Further, if in analogy to eq. (3-8) we define

$$S_G(j,\ell) = \Psi_{\ell j}^T \left[\frac{\partial}{\partial \alpha_\ell} G_j\right] \Psi_{rj} \quad , \qquad (4\text{-}7)$$

then

$$\left[\begin{array}{c|c} \Delta_\ell^{-1}(j) & 0 \\ \hline 0 & 0 \end{array}\right] = S_G^T(j,\ell)\, S_G(j) + S_G^T(j)\, S_G(j,\ell). \qquad (4\text{-}8)$$

Since the singular values of $S_G(j)$ are the square roots of the singular values of $[P'_{ML}(j)]^{-1}$, there is far less loss of significance in dealing with these quantities.

CONCLUSIONS

Algorithms have been developed for the maximum-likelihood estimation of initial means and covariances when the initial condition is not observable per test and when the estimated initial covariance may be singular. These algorithms are both computationally efficient and well-conditioned numerically.

REFERENCES

[1] D.W. Porter, M.D. Shuster, B.P. Gibbs, and W.S. Levine, "A Partitioned Recursive Algorithm for the Estimation of Dynamical and Initial-Condition Parameter from Cross-Sectional Data," Twenty-Second Conference on Decision and Control, San Antonio, Texas, December 1983.

[2] D.W. Porter, M.D. Shuster, and W.S. Levine, "A Doubly Recursive Algorithm for System Identification from Cross-Sectional Data," American Control Conference, San Francisco, California, June 1983.

[3] L.J. Levy, R.H. Shumway, D.E. Olsen, and F.C. Deal, Jr., "Model Validation from an Ensemble of Kalman Filter Tests," Proc. 21st Midwestern Symposium on Circuits and Systems, Ames, Iowa, 1978.

[4] R.H. Shumway, D.E. Olsen, and L.J. Levy, "Estimation and Tests of Hypothesis for the Initial Mean and Covariance in the Kalman Filter Model," Communications in Statistics: Theory and Methods, Vol. A10, No. 16, 1981.

[5] B. Friedland, "Treatment of Bias in Recursive Filtering," IEEE Trans. Automat. Contr., Vol. AC-14, no. 1, pp. 359-367, August 1969.

[6] G. Bierman, "The Treatment of Bias in the Square Root Information Filter Smoother," Conference on Decision and Control, 1973.

[7] C.L. Lawson and R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, 1974.