# Estimation of Kalman Filter Model Parameters from an Ensemble of Tests

by

B. P. Gibbs, D. R. Haley, W. Levine, D. W. Porter,
and C. J. Vahlberg
Business and Technological Systems, Inc.

## ABSTRACT

This paper presents a methodology for estimating initial mean and covariance parameters in a Kalman filter model from an ensemble of non-identical tests. In addition, the problem of estimating time constants and process noise levels is addressed. The work is motivated by practical problems such as developing and validating inertial instrument error models from laboratory test data or developing error models of individual phases of a test.

## 1.0 INTRODUCTION

This paper presents a methodology for estimating initial mean and covariance parameters in a Kalman filter model from an ensemble of non-identical tests. In addition, the problem of estimating time constants and process noise levels is addressed. The work is motivated by practical problems such as developing and validating inertial instrument error models from laboratory test data or developing error models of individual phases of a test.

Previous results in the literature [2,3] employ a Kalman smoother to obtain a sufficient statistic for the estimation of initial mean and covariance. Then the Expectation-Maximization (EM) algorithm [1] is applied to iteratively obtain maximum likelihood estimates of the entire initial mean vector and covariance matrix. The previous results are extended in this paper to account for parameter constraints such as constraining variables that are physically unrelated to each other to be uncorrelated. Further, the results are extended to consider time constant and process noise level parameters. Previous techniques capable of estimating initial mean and covariance parameters and dynamic parameters require re-reunning a Kalman filter for each value of the parameter vector considered. The new approach presented here is more efficient in that the filter need be re-run only for dynamic parameters.

System testing (for example, of inertial instruments in the laboratory and error mechanisms from flight data) is often done in multiple phases that are physically different but linked dynamically in a given test. In order to obtain models for different phases, the previous results could be applied where the phase dynamics are stacked one on top of the other. New results are presented that provide a simpler and computationally improved approach that deals with each phase individually. The new results are also useful when only one multiple phase test is conducted, it is only desired to estimate the state in each phase, and the state is unobservable in a given phase but observable over all phases.

In practical testing situations, suboptimal filters are often used. Results are presented that account for filter suboptimality.

Theoretical convergence results for the present application of the iterative EM algorithm are presented. Both the case of observable and unobservable per test dynamics are addressed. Also included are some references regarding rate of convergence and the effect of constraints on elements of the estimated covariance matrix.

## 2.0 THE DL/EM APPROACH

The Data Likelihood (DL) algorithm was derived in [2, 3] during analyses directed toward the estimation of mean and covariance parameters of the initial distribution of certain discrete-time multi-dimensional Gaussian random processes. The essential idea is to use Kalman smoother (Bayesian) estimates of realizations of the initial conditions to estimate the mean and variance of these initial condition distributions and to test certain statistical hypotheses thereabout. The iterative DL scheme arose in the process of attempting to divorce the a priori model to be validated from the estimated model, and was observed to be of the form of the Generalized Expectation-Maximization method described by Dempster, et al [1].

The context of the problem is as follows. For each test, $j$, the realization of the r.v. $x$ is assumed to be described by

$$x_{k,j} = \Phi_{k,j} x_{k-1,j} + w_{k,j} \quad , \; k=1,2,\cdots,n_j \; , \; j=1,\cdots,N \; . \tag{1}$$

and to be observed by

$$z_{k,j} = M_{k,j} x_{k,j} + v_{k,j} \; . \tag{2}$$

Here $w$ and $v$ are assumed zero mean, white, uncorrelated, Gaussian, and independent of $x_{o,j}$ for all $j$. It is assumed that $x_{o,j}$ are realizations from a Gaussian distribution with mean $\mu$ and covariance $\Sigma$, $\mu$ and $\Sigma$ unknown. In [2, 3] it is shown that the log likelihood function for the process can be written as

$$\log \; L(\mu,\Sigma) = \log \; L_{ML}(\mu,\Sigma) + R \; ,$$

where

$$\log L_{ML}(\mu,\Sigma) = -\frac{1}{2}\sum_{j=1}^{N} \log \left|\Sigma+P(j)\right| \qquad (3)$$

$$-\frac{1}{2}\sum_{j=1}^{N} (\hat{x}_{o,j}-\mu)^t [\Sigma+P(j)]^{-1} (\hat{x}_{o,j}-\mu)$$

and R is independent of $\mu$ , $\Sigma$ . Here $\hat{x}_{o,j}$ are maximum likelihood (non-Bayesian batch least squares) estimates of $x_{o,j}$ , and $P(j)$ are the associated estimation error covariances. Differentiating Log $L_{ML}$ with respect to $\mu$ and $\Sigma$ and setting the derivative to zero yields the equations

$$\sum_{j=1}^{N} [\Sigma+P(j)]^{-1} [\hat{x}_{o,j}-\mu] = 0 \qquad (4)$$

$$\sum_{j=1}^{N} [\Sigma+P(j)]^{-1} (\hat{x}_{o,j}-\mu) (\hat{x}_{o,j}-\mu)^t [\Sigma+P(j)]^{-1} \qquad (5)$$

$$-\sum_{j=1}^{N} [\Sigma+P(j)]^{-1} = 0 .$$

It is further noted that the Kalman smoother (Bayesian) estimates $\hat{x}_{o,j} \equiv \hat{x}_{o,j}(\mu,\Sigma)$ and associated estimation error covariance $\hat{P}_j \equiv \hat{P}_j(\Sigma)$ are related to $\hat{x}_{o,j}$ and $P(j)$ by

$$\hat{x}_{o,j} = [\Sigma+P(j)]\,\Sigma^{-1}\,[\hat{\hat{x}}_{o,j}-\mu] , \qquad (6)$$

$$\hat{P}_j = (P(j)^{-1}+\Sigma^{-1})^{-1} . \qquad (7)$$

Then the iterative DL loop can be defined:

$$\hat{\mu}_s = \frac{1}{N} \sum_{j=1}^{N} \hat{\hat{x}}_{o,j} (\hat{\mu}_{s-1}, \hat{\Sigma}_{s-1}) , \tag{8}$$

$$\hat{\Sigma}_s = \frac{1}{N} \sum_{j=1}^{N} \{\hat{P}_j(\hat{\Sigma}_{s-1}) + [\hat{\hat{x}}_{o,j}(\hat{\mu}_{s-1}, \hat{\Sigma}_{s-1}) - \hat{\mu}_s] \times [\hat{\hat{x}}_{o,j}(\hat{\mu}_{s-1}, \hat{\Sigma}_{s-1}) - \hat{\mu}_s]^t\} , \tag{9}$$

$$\hat{P}_j(\hat{\Sigma}_s) = [\hat{\Sigma}_s^{-1} + P_j^{-1}]^{-1} , \tag{10}$$

$$\hat{\hat{x}}_{o,j}(\hat{\mu}_s, \hat{\Sigma}_s) = \hat{\mu}_s + \hat{P}_j(\hat{\Sigma}_s) \, P_j^{-1} \, [\hat{x}_{o,j} - \hat{\mu}_s] , \tag{11}$$

with iteration on $s$ .

## 2.1 Theoretical Convergence

Theoretical convergence of the DL algorithm is addressed in detail in references [4] and [5]. The first note begins by proving that the DL algorithm is a Generalized Expectation Maximization algorithm. It then follows that the DL algorithm produces a monotone increasing sequence of likelihoods. It also follows, under the additional assumption that there exists a pair $\underline{\Sigma}^*$ , $\underline{\mu}^*$ that maximizes the likelihood, that $\underline{\Sigma}^*$ , $\underline{\mu}^*$ is a fixed point of the DL algorithm. Finally, it is shown that the sequence $\{\underline{\mu}_s, \underline{\Sigma}_s\}_{s=0}^{\infty}$ converges to some $\underline{\mu}^o$, $\underline{\Sigma}^o$ assuming that $\underline{\Sigma}_s \leq c\underline{I}$ for all $s$ and some $c$ .

There are two defects in this result. First, it is not guaranteed that $\underline{\mu}^o$ , $\underline{\Sigma}^o$ maximize the log likelihood. There is probably nothing to be done about this. Dempster et al [1] remark, as is probably true, that

$\underline{\mu}^0$ , $\underline{\Sigma}^0$ will, in almost all applications, occur at a local, if not global, maximum of the likelihood. Second, it would be desirable to relax the assumption that $\Sigma_s \leq c\ \underline{I}$ . In [5] it is shown that $\Sigma_s \leq c\underline{I}$ is automatically true provided all the tests are identically set-up. It seems reasonable, but has not yet been proven, that a similar result holds in general.

## 2.2 Constrained DL Estimates

In many potential applications of the DL method, the random variable of interest (whose mean and covariance we desire to estimate) is of rather large dimension. It is also often true that several of the parameters to be estimated are simultaneously poorly estimable and of relatively little interest. The judicious constraint of some parameters thus presents itself as a reasonable possibility. For example, if there is reason to believe that some components of the random variable are physically uncorrelated, and any correlation is believed to be largely irrelevant, little is likely to be lost if the estimated covariance is constrained to exhibit zero correlation.

It is desired to obtain those values $\mu$ and $\Sigma$ of the mean and covariance of the r.v. under consideration which maximize the log likelihood

$$L = \log L(\mu,\Sigma)_{ML} = -\frac{1}{2}\sum_{1}^{N} \log\left|\Sigma+P(j)\right| - \frac{1}{2}\sum_{1}^{N} (\hat{x}_{o,j}-\mu)^t\ [\Sigma+P(j)]^{-1}\ (\hat{x}_{o,j}-\mu)\ , \tag{12}$$

subject to appropriate constraints. Three forms of constraints have been explicitly considered. Within the DL iterative context, it seems that a rather wide variety of constraints may be handled quite easily.

The general approach taken here is to use Lagrange multipliers to reduce the constrained maximization problem to a modified but unconstrained problem.

The first forms of constraint on $\mu = [\mu_i]$ and $\Sigma = [\sigma_{jk}]$ under consideration here are

$$\phi_i = \delta_i(\mu_i - \mu_i^*) = 0 \ , \tag{13}$$

$$\psi_{jk} = \delta_{jk}(\sigma_{jk} - \sigma_{jk}^*) = 0 \ . \tag{14}$$

Here $[\delta_i]$ and $[\delta_{jk}]$ are "selectors",

$$\delta_i = \begin{cases} 0 \text{ if } \mu_i & \text{unconstrained} \\ 1 \text{ if } \mu_i & \text{constrained} \end{cases} \tag{15}$$

$$\delta_{jk} = \begin{cases} 0 \text{ if } \sigma_{jk} & \text{unconstrained} \\ 1 \text{ if } \sigma_{jk} & \text{constrained} \end{cases} \tag{16}$$

Clearly, reasonableness dictates that $[\delta_{jk}]$ and $[\sigma_{jk}^*] = \Sigma^*$ be symmetric, and that $\Sigma \geq 0$ . The problem now becomes that of obtaining an unconstrained maximum to

$$L^* = L^*(\mu, \Sigma) = L + \sum_i \lambda_i \ \delta_i(\mu_i - \mu_i^*) + \sum_{j,k} \lambda_{jk} \ \delta_{jk}(\sigma_{jk} - \sigma_{jk}^*) \ . \tag{17}$$

(Again, it is clear that $[\lambda_{jk}]$ must be symmetric.) To extremize $L^*$ , we set

$$\frac{\partial L^*}{\partial \mu} = \frac{\partial L}{\partial \mu} + \text{diag}(\delta\ )[\lambda\ ]$$

$$= \frac{\partial L}{\partial \mu} + \Delta\lambda = 0\ , \tag{18}$$

and

$$\frac{\partial L^*}{\partial \Sigma} = \frac{\partial L}{\partial \Sigma} + [\lambda_{jk}\delta_{jk}]$$

$$= \frac{\partial L}{\partial \Sigma} + \Lambda = 0\ . \tag{19}$$

The explicit computations and solutions become rather tedious and are not reproduced here. The complete details and several examples are found in [6]. It should be noted that the use of Lagrange multipliers has an important advantage that is not mentioned in [6]. It is shown in [7] that $\lambda_{jk} = \frac{\partial L}{\partial \sigma_{jk}}\Big|_{\sigma_{jk}=\sigma^*_{jk}}$ and $\lambda_i = \frac{\partial L}{\partial M_i}\Big|_{\mu_i=\mu^*_i}$. In words, solving Lagrange multipliers gives the sensitivity of the log likelihood to the constraint.

Consider the case where the mean $\mu$ is partitioned as $\mu = \begin{bmatrix}\mu_u \\ \mu_c\end{bmatrix}$ into its constrained and unconstrained parts. It is rather straightforward to show (see [6]) that the solution for $\hat{\mu}$ at each DL step is given by

$$\hat{\mu} = \begin{bmatrix}\hat{\mu}_u \\ \mu_c\end{bmatrix} = \begin{bmatrix}\frac{1}{N}\left[\sum_1^N \hat{x}_i^u + \Sigma_{12}\Sigma_{22}^{-1}\left\{N\mu_c^* - \sum_1^N \hat{x}_i^c\right\}\right] \\ \mu_c^*\end{bmatrix}\ , \tag{20}$$

from which the effect of the constraints on $\hat{\mu}$ at each iteration is clearly visible. (Here the state estimates $\hat{x}_i$ are also assumed partioned as $\mu$ ,

$$\hat{\hat{x}}_i = \begin{bmatrix} \hat{\hat{x}}_i^u \\ \\ \hat{\hat{x}}_i^c \end{bmatrix} .)$$

Such a formal solution is not generally available for the covariance equations. However, several specific cases yield results of some interest. One such case arises when $\Sigma$ is partioned as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} ,$$

and we wish to constrain $\Sigma_{12} = \Sigma_{12}^*$ , $\Sigma_{21} = \Sigma_{12}^t = \Sigma_{12}^{*t}$ , and $\Sigma_{22} = \Sigma_{22}^*$ . Denoting

$$\Sigma^* - M = \tilde{M} = \begin{bmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{bmatrix} , \tag{21}$$

where

$$M = \frac{1}{N} \{ \sum_1^N (\hat{\hat{x}}_i - \mu)(\hat{\hat{x}}_i - \mu)^t + \hat{P}_i \} , \tag{22}$$

It is shown in [6] that

$$\Sigma_{11} = M_{11} + \Sigma_{12} \Sigma_{22}^{-1} \tilde{M}_{12}^t + \tilde{M}_{12} \Sigma_{22}^{-1} \Sigma_{21} - \Sigma_{12} \Sigma_{22}^{-1} \tilde{M}_{22} \Sigma_{22}^{-1} \Sigma_{21} , \tag{23}$$

$$\Sigma_{12} = \Sigma_{12}^* , \tag{24}$$

$$\Sigma_{22} = \Sigma_{22}^* . \tag{25}$$

One slightly disconcerting question which does arise though regards the positivity of $\Sigma_{11}$ in (23). It is presently not clear what conditions on $\Sigma_{12}^*, \Sigma_{22}^*$ yield $\Sigma_{11}^* > 0$. The conditions on $\Sigma_{12}^*$ are more open to question since $\Sigma_{22}^*$ is more naturally restricted. Perhaps a relevant question is "how does one place a reasonable a priori constraint on a cross-covariance matrix?" One reasonable choice for the problem at hand might well be $\Sigma_{12}^* = 0$. In this case $\Sigma_{11} = M_{11}$, and the difficulty regarding definiteness disappears.

A slightly different sort of result arises from the more specific desire to constrain $\Sigma$ to be of the form

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} ,$$

with $\Sigma_{12} = \Sigma_{21}^t = 0$, and $\Sigma_{22} = \text{diag}\,(\sigma_{22i})$.

Then it is shown in [6] that the solution is

$$\Sigma_{11} = M_{11} , \tag{26}$$

$$\Sigma_{12} = \Sigma_{12}^* = 0 , \tag{27}$$

$$\Sigma = \begin{cases} M_{22ii} & i=j, \\ \sigma_{22ij}^* = 0 , & i=j . \end{cases} \tag{28}$$

It is clear that the constrained lack of correlation forces $\Sigma_{11}$ and $\Sigma_{22}$ to exactly follow the data observed for each. It is important to note that the above result can be applied to cases where elements of $\Sigma_{12}$ or off-diagonal elements of $\Sigma_{22}$ are constrained to known non-zero values. Suppose the $x_{o,j}$ can be expressed as

$$x_{o,j} = \mu + x_{1j} + x_{2j} \tag{29}$$

where $x_{2j}$ is uncorrelated with $x_{1j}$, the covariance of $x_{2j}$ is known and carries the known non-zero values mentioned above, and the covariance of $x_{1j}$ is to be estimated subject to the constraints $\Sigma_{12} = \Sigma_{21}^T = 0$, $\Sigma_{22} = \text{diag}\,(\sigma_{22i})$. Then the results of equations (26), (27), (28) can be applied where $x_{o,j}$ is re-defined to be

$$x_{o,j} = \mu + x_{1j} \tag{30}$$

A second form of constraint on $\mu$ and $\Sigma$ is

$$\mu = \alpha\mu^*, \quad \text{and}$$

$$\Sigma = \beta\Sigma^* .$$

Here $\mu^*, \Sigma^*$ are assumed to be given, and $\alpha, \beta$ are undetermined constants. For purposes of analysis, these constraints may be more conveniently stated as

$$\phi_i = \mu_1^*\mu_i - \mu_i^*\mu_1 = 0 , \qquad p \geq i \geq 2, \tag{31}$$

$$\psi_{jk} = \sigma_{11}^*\sigma_{jk} - \sigma_{jk}^*\sigma_{11} = 0 \qquad 1 \leq j,k \leq p , \quad (j,k) \neq (1,1) \tag{32}$$

The Lagrange multipliers $\lambda_i, \lambda_{jk}$, then enter in the unconstrained maximization of

$$L^* = L + \sum \lambda_i\phi_i + \sum \lambda_{jk}\psi_{jk} . \tag{33}$$

The explicit extremizing solution for $\alpha$ is given by

$$\alpha = \frac{1}{N} \frac{\mu^{*t} \sum^{-1} \{\sum_{\hat{}}^{N} x_{0,i}\}}{\mu^{*t} \sum^{-1} \mu^{*}} \quad . \tag{34}$$

A similar but quite complicated analytic solution for $\beta$ is obtained. As was the case for identity constraints, this value of $\beta$ and derivational details are contained in [6].

A third form of constraint of potential interest, applying only to the covariance estimation problem, is that of a specified correlation matrix, $R = [\rho_{ij}]$, so that the covariance takes the form

$$\sum = \text{diag}\,(\sqrt{\sigma_{ii}})\ R\ \text{diag}\,(\sqrt{\sigma_{jj}}) \tag{35}$$

with $\sigma_{ii}$ being the individual variance components. Then the constraint functions are

$$\phi_{ij} = \sigma_{ij} - \sqrt{\sigma_{ii}}\ \rho_{ij}\ \sqrt{\sigma_{jj}} = 0 \ . \qquad (i \neq j) \tag{36}$$

Clearly the usual properties of a correlation matrix are required.

The modified likelihood function is then

$$\begin{aligned} L^{*} &= L + \sum_{i \neq j} \lambda_{ij}\ \phi_{ij} \ . \\ &= L+S \ . \end{aligned} \tag{37}$$

We wish to find $\sum$ such that $\frac{\partial L^{*}}{\partial \sum} = 0$. Again, the details of the solution are found in [6].

## 2.3 Observability

The effect of unobservable tests on the DL algorithm is discussed in detail in reference [8]. The results can be summarized as follows:

(1) The state space can be divided into an observable subspace and an unobservable subspace.

(2) On the observable subspace the DL algorithm performs as if the observable subspace is the whole space. Data and estimates of $\underline{\Sigma}$ and $\underline{\mu}$ on the unobservable space have no effect on the DL algorithm as applied to the observable subspace.

(3) The data and estimates on the observable subspace do effect the algorithms results on the unobservable subspace and the correlations between unobservable and observable subspaces.

(4) If the DL algorithm is initialized with zero correlation between observable and unobservable subspaces then the correlation will remain zero and the DL algorithm will not change the mean and covariance on the unobservable subspace.

## 2.4 Estimation of Markov Parameters

In addition to estimating parameters of initial distributions, it is often of interest to use data from multiple tests to estimate dynamic parameters of the system, particularly parameters of Markov processes. Several possibilities exist for such estimation, and three are discussed very briefly here. It should be noted that such estimation likely is most useful for consistency checking because of the innately poor identifiability of such parameters during system tests of short time duration. Further, all methods for their estimation are likely to be computationally costly.

In [9], Goodrich and Caines have presented a methodology for maximum likelihood idenfication of system parameters based on data from repeated independent tests. The likelihood function is based, as in the original derivation of [3], on Kalman filter innovations, and the assumption of independence of realization yields a rather tractable form. Methods for modification of the procedure to allow for correlated tests should be studied further. The computational burden can be high for this approach since multiple Kalman filter passes are needed at each iteration.

Sun [10] has presented an application of the E-M procedure to the simultaneous estimation of system initial state, process and measurement noise levels, and system dynamics based on data from a single test. The paper indicates that an extension to repeated tests may be possible. Again, further study and extension seem necessary.

A third possibility combines several aspects of the DL methods as previously described, the ideas of Goodrich and Caines, and other work in maximum likelihood estimation.

Consider a dynamical system as described in equation (1), where $\Phi$ and $Q$ may depend on some parameter vector $\alpha$ (e.g., time constants and process noise levels). Whether to solve for $\alpha,\mu,\Sigma$ simultaneously or separately seems unclear as yet. For a given value of $\alpha$, one might obtain via DL the maximum likelihood estimates $\hat{\mu}(\alpha),\hat{\Sigma}(\alpha)$. Then, fixing $\mu,\Sigma$, numerical/gradient methods could be used to obtain the value $\alpha$ to maximize the likelihood. Also open to question is the variability of $\alpha$ -- one might assume $\alpha$ to be universally constant, constant over groups of tests or unique from test to test.

An illustration of a possible implementation loop on such a procedure is found in Figure 1. We desire, again, to estimate $\mu=E(x_{o,i})$, $\Sigma=E((x_{o,i}-\mu)(x_{o,i}-\mu)^t)$, and the Markov parameter vector $\alpha$. Dropping the individual test indicator $i$ for the moment, we have

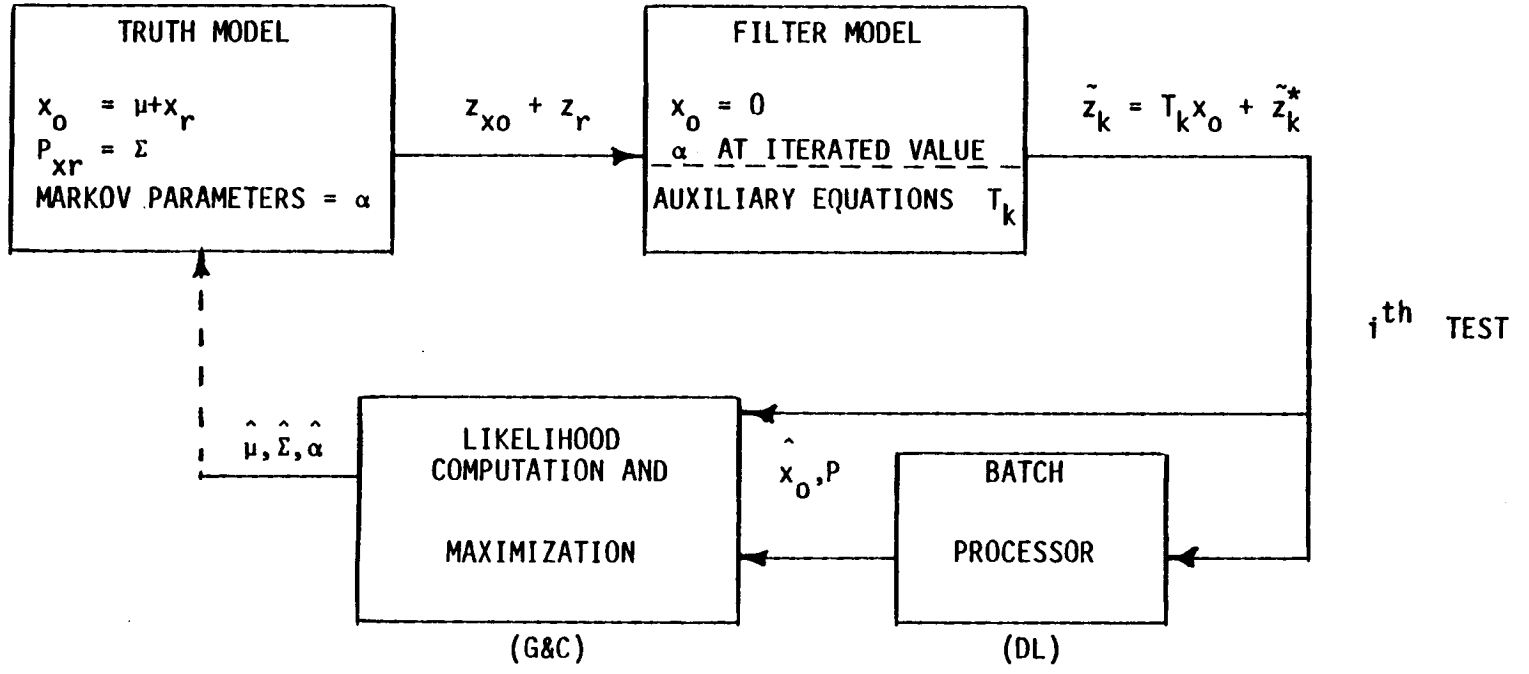


Figure 1 Proposed Markov Parameter Estimation Loop

$$x_o = \mu + x_r \ ,$$

$x_r$ the random part of $x_o$ , and

$$P_{xr} = \Sigma \ .$$

The observation $z$ may be decomposed into

$$z = z_{xo} + z_r \ ,$$

a part due to $x_o$ and a random part. In Reference [11] it is shown that the innovations $\tilde{z}_k$ may be decomposed as

$$\tilde{z}_k = T_k \, x_o + \tilde{z}_k^* \ . \tag{38}$$

where $\tilde{z}_k^*$ is computed based on assumed filter and truth models of $x_o=0$ . It is then possible to write the log likelihood function, given $\mu,\Sigma,\alpha,$ and indexing repeated tests by $i$,

$$-2 \log p(z|\mu,\Sigma,\alpha) + \text{constant} =$$

$$\sum_i \{\log|\Sigma+P(i)| + (\hat{x}_{o,i}-\mu)^t \, (\Sigma+P(I))^{-1} \, (x_{o,i}-\mu) \tag{39}$$

$$- \log|P(i)| - \hat{x}_{o,i}^t \, P^{-1}(i) \hat{x}_{o,i}\}$$

$$+ \sum_{i,k} \{\log|P_{\tilde{z}_{i,k}}| + \tilde{z}_{i,k}^t \, P_{\tilde{z}_{i,k}}^{-1} \, \tilde{z}_{i,k}\} \ .$$

If $\alpha$ is known, this procedure reduces to the DL algorithm. If $\alpha$ is unknown then gradient procedures may be used in the maximization. This requires differentiating the estimate for $\alpha$ , but not for $\mu,\Sigma$ . Another idea is to solve for $\alpha$ for each test, and only for $\mu,\Sigma$ for repeated tests. The proposed iterative loop is illustrated as the dotted closure in Figure 1.

In summary, the detailed analysis of the effect of Markov parameters is a difficult problem which has really only been recently addressed. It is felt that substantial additional effort may be required to fully develop adequate analysis methodology, but that failure to attempt to address the problem in detail may lead to inadequate analysis capabilities in some areas.

## 3.0 MULTIPLE PHASE THEORY

Many systems are operated and evaluated in a sequence of phases. The analysis of performance in one phase is carried out relatively independent of results from other phases, and then results are combined at the end. Although this procedure is not a constraint for many systems, there is interest in studying it. This section presents the theory necessary to combine the results from several phases on one test, then to analyze the results using a cumulative methodology such as DL. An approximation is presented which allows for reasonably accurate quick cumulative evaluations. An extension of the theory is discussed at the end.

Some framework, nomenclature, and assumptions need to be stated before the theory is presented. It is assumed that the effects of least squares estimates of errors from previous phases have been removed from the data prior to its analysis in a Bayesian per phase filter, or equivalently removed after data analysis is complete. The per phase analysis is actually done with a Bayes filter, but the theory is developed starting with a least squares (infinite prior) filter. The least squares estimate of errors in the $i^{th}$ phase can be represented by (see Reference [12])

$$\hat{x}_{i_{LS}} = \mu_i + x_{r_i} + \tilde{x}_i - \tilde{x}_{T_{i-1}} , \tag{40}$$

where

$\mu_i$ = Systematic error in this phase,

$x_{r_i}$ = Random error introduced in this phase,

$\tilde{x}_i$ = Residual estimation error from this phase due to $Q_i$ , $R_i$ ,

$\tilde{x}_{T_{i-1}}$ = Residual estimation error from previous phase due to $Q_{i-1}$, $R_{i-1}$ .

The covariance of this estimate is

$$E((\hat{x}_{LS}-\mu)(\hat{x}_{LS}-\mu)^T) = \Sigma_i + P_i + P_{T_{i-1}} \tag{41}$$

where

$$\Sigma_i = E(x_{r_i} x_{r_i}^T)$$

$$P_i = E(\tilde{x}_i \tilde{x}_i^T)$$

$$P_{T_{i-1}} = E(\tilde{x}_{T_{i-1}} \tilde{x}_{T_{i-1}}^T) .$$

The error at the transition time can be represented by (see References [11], [13])

$$\tilde{x}_T = C\, \tilde{x}_i + \tilde{x}'_{T_i} \tag{42}$$

where the following statistics are obtained suppressing the $i$ subscript

$$\begin{aligned} E(\tilde{x}_i \tilde{x}_i)^T &= P \\ E(\tilde{x}_T \tilde{x}_T^T) &= CPC^T + P'_T = P_T \\ E(\tilde{x}_i \tilde{x}_T^T) &= PC^T = P_C \\ E(\tilde{x}_T \tilde{x}_o^T) &= 0 \end{aligned} \tag{43}$$

These statistics can be calculated following data analysis using a conventional Bayesian filter. If the initial states are augmented to the state vector to provide a fixed-point estimate of errors, all necessary covariances and correlations are obtained. For the state vector definition

$$x^* = \begin{bmatrix} x_i \\ x_T \end{bmatrix} \tag{44}$$

the covariance of estimation errors obtained from a fixed-point Bayesian smoother will be

$$P_N^* = \begin{bmatrix} P_N & P_{CN}^T \\ P_{CN} & P_{TN} \end{bmatrix} \tag{45}$$

Using the above, the correlation matrix C and the covariance of transition-time errors which are independent of initial estimation errors can be calculated

$$\begin{aligned} C &= P_{CN}^T P_N^{-1} \\ P_T' &= P_{TN} - CP_{CN} \end{aligned} \tag{46}$$

The max-likelihood information matrix is obtained as

$$P^{*-1} = \begin{bmatrix} P_N & P_{CN}^T \\ P_{CN} & P_{TN} \end{bmatrix}^{-1} - \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{47}$$

Assuming that it is invertible, the max-likelihood covariance is obtained and is given by

$$P^* = \begin{bmatrix} P & P_C^T \\ P_C & P_T \end{bmatrix} = \begin{bmatrix} P & CP \\ PC^T & CPC^T + P_T' \end{bmatrix} \tag{48}$$

The discussion thus far has focused upon manipulation of data and covariances from one phase of a multiple phase system. The result is the

max-likelihood covariance and by similar procedures, the estimate for the combined state vector at the initial and transition times. Since the error in the max-likelihood estimate, as represented by Equation (40) is unbiased, the error in the estimate of the $i^{th}$ phase, although correlated with the previous phase, is uncorrelated with all phases previous to that. Hence, the multiphase max-likelihood covariance of the max-likelihood estimate for the stacked vector of $\mu_i + x_{ri}$ vectors is of the banded form (see Reference [12]).

$$P = \begin{bmatrix} P_1 & -P_{c_1}^T & 0 & 0 & 0 \\ -P_{c_1} & P_2+P_{T_1} & -P_{c_2}^T & 0 & 0 \\ 0 & -P_{c_2} & P_3+P_{T_2} & -P_{c_3}^T & 0 \\ 0 & 0 & -P_{c_3} & P_4+P_{T_3} & -P_{c_4}^T \\ 0 & 0 & 0 & -P_{c_4} & P_5+P_{T_4} \end{bmatrix} \qquad (49)$$

This banded form has some interesting properties that lead to a useful result, especially when the following often practical assumptions are made:

(1) The derivations presented have already assumed that all states in each phase are observable -- so that the max-likelihood information matrix is invertible

(2) Interphase correlations can be ignored for preliminary cumulative analysis giving an algorithm suboptimal in the sense that information is thrown away but not in the sense that an approximation is made.

Although the first assumption may not always be true, it may be possible to redefine the state vector so that the unobservable states do

not enter the system until the phase in which they are observable. This can be accomplished automatically in a mathematical sense utilizing the Singular Value Decomposition (SVD) algorithm. Thus, although the first assumption may not be able to be satisfied explicitly, there are ways to accomplish its effect without degrading the fidelity of the model.

## 4.0 USE OF SUBOPTIMAL STATE ESTIMATES IN MULTIPHASE ANALYSIS

It is often the case that data from the several phases under consideration are obtained from Kalman-Schmidt filter/fixed point smoother algorithms. Thus the algorithm for processing a test phase should be capable of handling suboptimal gains. Even if the filter were optimal, the equations for the suboptimal case would be applicable, and, in some situations, might be preferable to equations assuming optimality.

It is also desirable that the processing for each phase be done independently. In some cases, the processing for different phases may be done by different organizations. Thus, the per phase data reduction must use no information from other phases. The combination of phase estimates is done as the final step in the data reduction.

Reference [14] defines the equations required for the phase data reduction. These equations are fairly general and would apply to most suboptimal filters. Also presented there are the additional recursive equations which must be computed in a consider filter so that the phases may be combined. These equations only apply to a Kalman-Schmidt filter (which automatically computes the correct covariance matrix) but could be modified for other suboptimal filters. The following section presents the algorithm for combining the suboptimal (or optimal) estimates from different phases.

### 4.1 Multiphase Reduction Using Suboptimal Estimates

The output of each phase will be a suboptimal, smoothed estimate of the state at initial time and transition time. Also obtained are the various covariance and sensitivity matrices. The true state at the epoch of each phase is assumed to have a mean and random component; i.e., for phase $i$

$$x_0^i = \mu^i + x_r^i \,. \tag{50}$$

The DL method attempts to estimate $\mu^i$ and cov $(x_r^i)$ by combining results of different phases and tests. To do this, the state estimates for each phase are manipulated so that they are in the familiar form

$$z = Hx_0 + v \ . \tag{51}$$

Consider the combination of phase 1 and 2 shown in Figure 2. The value of $x_T^1$ is

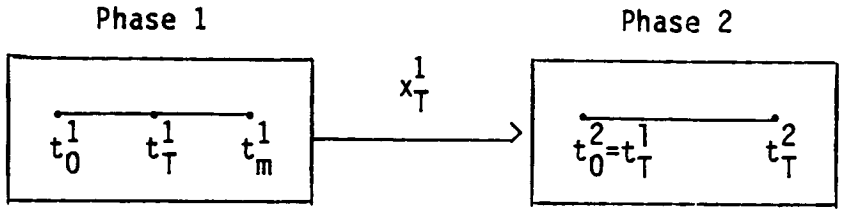


Figure 2 Phase Combination

directly included in the phase 2 initial condition, i.e.,

$$x_0^2 = \mu^2 + x_r^2 + Tx_T^1 \tag{52}$$

where $T$ is a transformation matrix (not to be confused with transition time $t_T$ ).

Now consider the estimates obtained from phase 1:

$$\begin{aligned} \hat{x}_0^1 &= \mu^1 + x^1 + \tilde{x}_0^1 \\ &= D_{10}^1(\mu^1+x_r^1) + D_{20}^1 w^T + v_0^1 \end{aligned} \tag{53}$$

$$\hat{x}_T^1 = D_{1T}^1 \phi_T x_0^1 + D_{2T}^1 w_T + v_T^1 \tag{54}$$

and from phase 2

$$\begin{aligned}\hat{x}_0^2 &= \mu^2 + x_r^2 + Tx_T^1 + \tilde{x}_0^2 \\ &= D_{10}^2(\mu^2 + x_r^2 + Tx_T^1) + v_0^2 \ .\end{aligned} \tag{55}$$

These equations can be combined in matrix notation as:

$$\begin{bmatrix} \hat{x}_0^1 \\ \hat{x}_T^1 \\ \hat{x}_2^2 \end{bmatrix} = \begin{bmatrix} D_{10}^1 & D_{20}^1 & 0 \\ D_{1T}^1\phi_T & D_{2T}^1 & 0 \\ D_{10}^2{}^T\phi_T & D_{10}^2{}^T & D_{10}^2 \end{bmatrix} \begin{bmatrix} y^1 \\ w_T \\ y^2 \end{bmatrix} + \begin{bmatrix} v_0^1 \\ v_T^1 \\ v_0^2 \end{bmatrix} \tag{56}$$

where $y^1 = \mu^1 + x_T^1$ and $y^2 = \mu^2 + x_T^2$. Notice that $v_0^1$ and $v_T^1$ are correlated but are uncorrelated with $v_0^2$. Also notice that there is no <u>a priori</u> information on $y^1$ and $y^2$ but that the <u>a priori</u> variance of $w_T$ is $Q_T$. Thus, equation (56) can be treated as three measurements in a Bayesian least squares estimator for $\hat{y}^1$, $\hat{w}_T$ and $\hat{y}^2$

where

$$E\left\{\begin{bmatrix} V_0^1 \\ V_T^1 \\ V_0^2 \end{bmatrix} \begin{bmatrix} V_0^{1T} & V_T^{1T} & V_0^{2T} \end{bmatrix}\right\}$$

is calculated in [14].

If all states are observable, then $\hat{y}^1$ and $\hat{y}^2$ should be maximum likelihood estimates.

In order to better understand the result of this processing, we assume that the first two measurements in equation (56) were processed first (to estimate $\hat{y}^1$ and $\hat{w}_T$ ) the third measurement was processed separately to estimate

$$y_s^2 = y^2 + Tx_T \ . \tag{57}$$

This can be done because the measurement errors are uncorrelated. Then we want to combine the estimates. Since $y^2$ had infinite a priori variance, all information in the third measurement will be used to estimate $y^2$ if $y^1$ and $w_T$ were observable from the first two measurements, i.e., the estimates of $y^1$ and $w_T$ will not change. Thus,

$$\hat{y}_2 = \hat{y}_s^2 - T(\phi_T\hat{y}^1+\hat{w}_T) = y^2 + \tilde{y}_s^2 - T(\phi_T\tilde{y}^1+\tilde{w}_T) \tag{58}$$

and the covariance of the error in the estimate of $y_2$ is

$$P_{\tilde{y}_s^2} + T[\phi_T ; I] \, P_{\substack{\tilde{y}^1 \\ \tilde{w}T}} \begin{bmatrix} \phi_T^T \\ \hline I \end{bmatrix} T^T \ . \tag{59}$$

The above analysis is similar to that given previously in the sense that data processing and requirements analysis can be done phase by phase.

## 5.0 BIBLIOGRAPHY


[1] Dempster, A. P., N. M. Land and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", J. Roy. Stat. Soc., Series B, Vol. 39, No. 1, 1977.

[2] Levy, L. J., 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, 1978.

[3] Shumway, R. H., D. E. Olsen and L. J. Levy, "Estimation and Tests of Hypothesis for the Initial Mean and Covariance in the Kalman Filter", American Statistical Association, San Diego, August, 1978.

[4] Levine, W. L., "Convergence of the DL Algorithm", BTS, May, 1980.

[5] Levine, W. L., "Boundedness of $\hat{\Sigma}_k$ in the DL Algorithm", BTS, July, 1980.

[6] Haley, D. R., "Modifications and Extensions to the DL Algorithm to Allow for Constrained Max Likelihood Estimation", BTS, May, 1980.

[7] Varaiya, P. O., Notes on Optimization, Von Nostrand Reinhold, New York, 1972.

[8] Levine, W. L., "Behavior of the DL Algorithm for Unobservable Tests", BTS, July, 1980.

[9] Goodrich, R. L., and P. E. Caines, "Linear System Identification from Nonstationary Cross-Sectional Data", IEEE Trans. AC. Vol. AC-24, No. 3, June, 1979.

[10] Sun, F. K., "An Alternative Approach for Maximum Likelihood Identification", 18th Conf. Dec. Cont., Ft. Lauderdale, Florida, December 1979.

[11] Porter, D. W., "Relationships Required to Implement Generalized Likelihood Approach", BTS, October, 1978.

[12] Porter, D. W., To Appear.

[13] Porter, D. W., and T. S. Englar, Jr., "Multi-Object Tracking via a Recursive Generalized Likelihood Approach", 18th Conf. Dec. Cont., Ft. Lauderdale, Florida, December, 1979.

[14] Gibbs, B. P., "Suboptimal Per-Phase Data Reduction", BTS, May, 1980.