A DOUBLY RECURSIVE ALGORITHM FOR SYSTEM IDENTIFICATION FROM NONSTATIONARY CROSS-SECTIONAL DATA

David W. Porter, member, IEEE
Malcolm D. Shuster, senior member, IEEE
Business and Technological Systems, Inc.
Aerospace Building, Suite 440
10210 Greenbelt Road
Seabrook, Maryland 20706

William S. Levine, senior member, IEEE
Department of Electrical Engineering
University of Maryland
College Park, Maryland 20742

1983 AMERICAN CONTROL CONFERENCE
San Francisco, California June 22-24

A DOUBLY RECURSIVE ALGORITHM FOR SYSTEM IDENTIFICATION FROM NONSTATIONARY CROSS-SECTIONAL DATA

David W. Porter, member, IEEE
Malcolm D. Shuster, senior member, IEEE
Business and Technological Systems, Inc.
Aerospace Building, Suite 440
10210 Greenbelt Road
Seabrook, Maryland 20706

William S. Levine, senior member, IEEE
Department of Electrical Engineering
University of Maryland
College Park, Maryland 20742

SUMMARY

Many practical applications require the simultaneous estimation of unknown dynamical parameters and unknown initial means and covariances from an ensemble of tests. A recursive algorithm which asymptotically obtains the maximum-likelihood estimate of both sets of unknown parameters is presented. The computational requirements of the algorithm are greatly reduced by partitioning the parameter vector into initial and dynamical parameters and making use of a sufficient statistic as an intermediate variable for the estimation of initial condition parameters. The results are illustrated by a numerical example.

1. INTRODUCTION

The majority of work on system identification has been concerned with system parameter identification from single sample longitudinal data. Goodrich and Caines[1] point out that a large range of identification problems require system parameters to be identified from cross-sectional nonstationary data. One such example, which motivates the work reported here, arises when one has a time-varying linear system with unknown constant parameters in the dynamics (so-called Markov parameters) and Gaussian distributed initial conditions with unknown mean and covariance. The presence of initial conditions usually guarantees non-stationary data even in the case of a time-invariant system and the estimation of the initial mean and covariance is impossible if one does not have multiple tests (i.e. cross-sectional data).

An obvious way to compute the parameter estimates from cross-sectional non-stationary data is via maximum likelihood or, more generally, some other prediction-error scheme[2] where these are batch approaches. It is desirable to replace the batch algorithms by a recursive algorithm to reduce data storage, make it possible to respond before all tests are completed, and reduce the amount of computation. Of course, the batch scheme will generally be more accurate for finite amounts of data. It will be shown using ideas from Ljung[3] that, in the limit as the number of tests tends to infinity, the parameter estimate computed via our doubly recursive algorithm equals the batch ML estimate.

For many problems the computational burden associated with estimating an initial covariance matrix of realistic dimension can be impractical. This occurs because a Kalman filter must be differentiated for each element of the covariance matrix. An approach is presented here that avoids differentiation of Kalman filters with respect to initial mean and covariance parameters through the use of a sufficient statistic motivated by the approach in [4,5].

The paper is organized as follows. The system model is described in Section 2. Section 3 describes an algorithm which is recursive from test to test but which requires a batch computation for each test. In Section 4, this batch computation is replaced by a partitioned recursion within each test that greatly reduces the necessary computations. Section 5 contains a numerical example. Finally, Section 6 gives some conclusions and suggestions for further research.

2. SYSTEM MODEL

Consider a system model of the form

$$\underline{x}^i(t+1) = \underline{A}^i_\theta(t)\,\underline{x}^i(t) + \underline{w}^i(t) \;;\; t=0,1,2,\ldots,n_i \qquad (1)$$

$$i=1,2,\ldots,M$$

$$\underline{y}^i(t) = \underline{C}^i_\theta(t)\,\underline{x}^i(t) + \underline{v}^i(t) \;;\; t=1,2,\ldots,n_i \qquad (2)$$

where

$$\underline{w}^i(t) \sim N(\underline{0},\underline{W}^i_\theta(t)) \quad , \quad \underline{v}^i(t) \sim N(\underline{0},\underline{V}^i_\theta(t))$$

$$\underline{x}^i(0) \sim N(\underline{\mu}_\theta,\underline{\Sigma}_\theta)$$

Here, $\underline{w}^i(t)$, $\underline{v}^i(t)$ are "white" and are independent of each other and of $\underline{x}^i(0)$. The subscript θ indicates dependence on an unknown vector of parameters $\underline{\theta} \in \Theta$. The superscript i denotes the i^{th} test.

There are several points that should be emphasized in connection with the model. First, the time dependence of $\underline{A}^i_\theta(t)$ can be the result of a feedback control system, possibly including a Kalman filter, being applied to the basic dynamical system. Since one might change the feedback control gains as a result of earlier test, $\underline{A}^i_\theta(t)$ can vary with i (as denoted by the superscript i).

3. TEST-TO-TEST RECURSIVE ALGORITHM

The basic problem is to find $\hat{\underline{\theta}}_{ML}$, the maximum-likelihood estimate of $\underline{\theta}$, given the data $\underline{y}^i(t)$ for $t=1,\ldots,n_i$ and $i=1,2,\ldots,M$. By definition,

$$\hat{\underline{\vartheta}}_{ML} = \arg\min_{\underline{\vartheta}} L(\underline{\vartheta},\underline{Y}) \qquad (3)$$

where $L(\underline{\vartheta},\underline{Y})$ denotes the negative log likelihood and $\underline{Y}$ denotes the collection of $\underline{y}^i(t)$ for $t=1,2,\ldots,n_i$ and $i=1,2,\ldots,M$. Of course, under the assumptions we have made $\underline{Y}$ is Gaussian distributed with some mean and variance which depend on $\underline{\vartheta}$.

The first step is to concatenate the data vectors $\underline{y}^i(t)$ for the i^{th} test to form a giant vector $\underline{y}(i) = [\underline{y}^i(1)^T, \underline{y}^i(2)^T, \ldots, \underline{y}^i(n_i)^T]^T$.

Then

$$\underline{y}(i) = \underline{H}^i_{\vartheta}\,\underline{x}^i(0) + \underline{v}(i) \qquad i=1,2,\ldots,M \qquad (4)$$

where $\underline{H}^i_{\vartheta}$ is a giant matrix concatenated from the $\underline{A}^i_{\vartheta}(t)$ and $\underline{C}^i_{\vartheta}(t)$, $\underline{v}(i) \sim N(\underline{0},\underline{R}^i_{\vartheta})$, and $\underline{R}^i_{\vartheta}$ is a giant matrix concatenated from $\underline{A}^i_{\vartheta}(t)$, $\underline{C}^i_{\vartheta}(t)$, $\underline{V}^i_{\vartheta}(t)$ and $\underline{W}^i_{\vartheta}(t)$.

For independent tests we can write Eq. (3) in the form

$$\hat{\underline{\vartheta}}_{ML} = \arg\min_{\underline{\vartheta}} \frac{1}{M}\sum_{i=1}^{M} \ell(\underline{\vartheta},\underline{y}(i)) \qquad (5)$$

where

$$\ell(\underline{\vartheta},\underline{y}(i)) = \frac{1}{2}\,\mathrm{tr}\,[\underline{S}^{i\,-1}_{\vartheta}(\underline{y}(i) - \hat{\underline{y}}(i))(\underline{y}(i) - \hat{\underline{y}}(i))^T] + \frac{1}{2}\log\det \underline{S}^i_{\vartheta} \qquad (6)$$

and

$$\underline{S}^i_{\vartheta} = \underline{H}^i_{\vartheta}\,\underline{\Sigma}_{\vartheta}\,\underline{H}^{i\,T}_{\vartheta} + \underline{R}^i_{\vartheta} = \mathrm{cov}(\underline{y}(i)) \qquad (7)$$

$$\hat{\underline{y}}(i) = E(\underline{y}(i)) = \underline{H}^i_{\vartheta}\,\underline{\mu}_{\vartheta} \qquad (8)$$

All of the expectations and covariances in Eqs. (7) and (8) are evaluated as functions of $\underline{\vartheta}$, as if $\underline{\vartheta}$ were known. Thus, the actual calculation of $\hat{\underline{\vartheta}}_{ML}$ via Eq. (3) or (5) is still fundamentally a batch calculation. However, Eqs. (5) through (8) express $\hat{\underline{\vartheta}}_{ML}$ in terms of the so-called prediction error and Ljung[3] has given a technique for converting a batch prediction error estimator to a recursive estimator whose estimate converges to the batch estimate as the amount of data goes to infinity.

Ljung's procedure amounts to writing

$$\hat{\underline{\vartheta}}(i) = \hat{\underline{\vartheta}}(i-1) + \gamma(i)\,\underline{R}^{-1}(i)\left[-\frac{d\ell}{d\underline{\vartheta}}\right|_{\hat{\underline{\vartheta}}(i-1),\underline{\varepsilon}(i)} \qquad (9)$$

$$\underline{R}(i) = \underline{R}(i-1) + \gamma(i)\left\{\left[\frac{d^2\ell}{d\underline{\vartheta}^2}\right|_{\hat{\underline{\vartheta}}(i-1),\underline{\varepsilon}(i)} + \delta\underline{I} - \underline{R}(i-1)\right\} \qquad (10)$$

where

$\hat{\underline{\vartheta}}(i)$ is the estimate of $\underline{\vartheta}$ based on data $\underline{Y}_i$

$\gamma(i) = 1/i$ although more general forms are possible

$$\underline{\varepsilon}(i) = \underline{y}(i) - E\{\underline{y}(i)\,|\,\hat{\underline{\vartheta}}(i-1)\} \qquad (11)$$

δ is a small positive number that is large enough to ensure $\underline{R}(i) > 0$ for all i

and

$$\ell(\underline{\vartheta},\underline{\varepsilon}(i)) = \frac{1}{2}\,\mathrm{tr}\,[\underline{S}^{i\,-1}_{\vartheta}\,\underline{\varepsilon}(i)\,\underline{\varepsilon}^T(i)] + \frac{1}{2}\log\det \underline{S}^i_{\vartheta} \qquad (12)$$

Ljung[3] proves, under some assumptions we will discuss below; that $\hat{\underline{\vartheta}}(i)$ converges w.p.1 either to the set

$$D_C = \left\{\underline{\vartheta}\,\middle|\,\frac{d}{d\underline{\vartheta}}\,\bar{V}(\underline{\vartheta}) = 0\right\} \qquad (13)$$

where, it can be shown that

$$\bar{V}(\underline{\vartheta}) = \lim_{M\to\infty} L\,(\underline{\vartheta},\underline{Y}_M) \qquad (14)$$

or to the boundary of the model set as $i\to\infty$. Actually convergence is proven for a class of positive semi-definite approximations to

$$\frac{d^2\ell}{d\underline{\vartheta}^2}.$$

Furthermore, among isolated points of D_C, only local minima of $\bar{V}(\underline{\vartheta})$ are possible convergence points. Note that $\hat{\underline{\vartheta}}(i)$ really converges to a local minimum of the negative log likelihood or to a value on the boundary of the admissible parameter set. However, this is all any batch algorithm achieves.

This convergence result is based on three assumptions:

A1: $$\bar{V}(\underline{\vartheta}) = \lim_{M\to\infty} E\{L(\underline{\vartheta},\underline{Y}_M)\} \qquad (15)$$

A2: $\hat{\underline{y}}(i|\underline{Y}_{i-1})$ is computed via equations of the form

$$\underline{x}(i+1) = \underline{F}(\underline{\vartheta})\ \underline{x}(i) + \underline{G}(\underline{\vartheta})\ \underline{y}(i) \tag{16}$$

$$\hat{\underline{y}}(i|\underline{Y}_{i-1}) = \underline{H}(\underline{\vartheta})\ \underline{x}(i) + \underline{H}(\underline{\vartheta})\ \underline{u}_{\vartheta} \tag{17}$$

$\underline{F}(\underline{\vartheta})$ has all its eigenvalues strictly inside the unit circle and $\underline{F}(\underline{\vartheta})$, $\underline{G}(\underline{\vartheta})$ and $\underline{H}(\underline{\vartheta})$ are twice differentiable for all $\underline{\vartheta}$ in the compact set of possible parameter values. Not that the second term on the right hand side of Eq. (17) is not included in Ljung's version. The extra term is needed to handle unknown initial conditions. The proof of convergence is straightforward.

A3: The test procedure that actually generates the data is "exponentially stable". That is, the influence of any test on future tests decreases exponentially.

Assumption (A1) is proven by expanding (15) to

$$\overline{V}(\underline{\vartheta}) = \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} E\{\ell(\underline{\theta},\underline{\varepsilon}(i))\} \tag{18}$$

where expectation is with respect to the true $\underline{\theta}$ and $\underline{\varepsilon}(i)$ is calculated using some other $\underline{\vartheta}$. So,

$$\overline{V}(\underline{\vartheta}) = \lim_{M\to\infty} \frac{1}{2M} \sum_{i=1}^{M} [\log \det \underline{S}^i_{\vartheta} + \mathrm{tr}\,[(\underline{S}^i_{\vartheta})^{-1}\ \underline{P}^i_{\varepsilon}(\underline{\theta})]] \tag{19}$$

where $P^i_{\varepsilon}(\underline{\theta})$ = true covariance of $\underline{\varepsilon}(i)$. In order for the limit on the right-hand side of the above equation to exist we have to impose some condition on the way tests can vary. A sufficient condition is for the change in conditions from test to test to go to zero as $i\to\infty$. If this condition is satisfied then

$$\overline{V}(\underline{\vartheta}) = \frac{1}{2}\log\det \underline{S}_{\vartheta} + \frac{1}{2}\mathrm{tr}\,[\underline{S}^{-1}_{\vartheta}\ \underline{P}_{\varepsilon}(\underline{\theta})] \tag{20}$$

It is trivial to show that the remaining assumptions are satisfied when the tests are independent $\underline{F}(\underline{\vartheta}) = \underline{0}$, $\underline{H}(\underline{\vartheta}) = \underline{H}^i_{\vartheta}$, $\overline{G}(\underline{\vartheta}) = \underline{0}$. When tests are coupled via some dependence of test conditions on previous tests then the assumptions impose conditions on the form of coupling. The key is Eqs. (16) and (17).

4. CALCULATION OF QUANTITIES FOR INDIVIDUAL TESTS

In general, even for independent tests, the measurements within a test will be highly correlated. Hence, the computation of

$$\frac{d\ell(\underline{\theta},\underline{\varepsilon}(i))}{d\underline{\theta}}$$

and of

$$\frac{d^2\ell(\underline{\vartheta},\underline{\varepsilon}(i))}{d\underline{\theta}^2}$$

is potentially very burdensome. The balance of this section describes the reduction of the computational requirements to a reasonable level.

There are two well known simplifications. First, write ℓ in terms of the innovations process within a single test. That is

$$\ell(\underline{\theta},\underline{\varepsilon}(i)) = \frac{1}{2}\sum_{k=1}^{n_i} \{\underline{\nu}^{iT}_{\vartheta}(k)\ \underline{R}^{i\,-1}_{\vartheta}(k)\ \underline{\nu}^i_{\vartheta}(k) + \log\det \underline{R}^i_{\vartheta}(k)\} \tag{21}$$

where

$$\underline{\nu}^i_{\vartheta}(k) = \underline{y}^i(t{=}k) - E_{\vartheta}\{\underline{y}^i(t{=}k)\,|\,\underline{y}^i(1),\ldots\,\underline{y}^i(t{=}k{-}1)\} \tag{22}$$

and

$$\underline{R}^i_{\vartheta}(k) = E_{\vartheta}\{\underline{\nu}^i_{\vartheta}(k)\ \underline{\nu}^{iT}_{\vartheta}(k)\} \tag{23}$$

The calculation of $\frac{d\ell}{d\underline{\theta}}$ is straightforward and gives (suppressing the test index i and the dependence on $\underline{\theta}$)

$$\frac{\partial\ell(\underline{\theta},\underline{\varepsilon}(i))}{\partial\theta_m} = \sum_{k=1}^{n} \{\underline{\nu}^T(k)\ \underline{R}^{-1}(k)\ \frac{\partial\nu}{\partial\theta_m}(k) - \frac{1}{2}\underline{\nu}^T(k)\ \underline{R}^{-1}(k)\ \frac{\partial R(k)}{\partial\theta_m}\ \underline{R}^{-1}(k)\ \underline{\nu}(k) + \frac{1}{2}\mathrm{tr}\,[\underline{R}^{-1}(k)\ \frac{\partial R(k)}{\partial\theta_m}]\} \tag{24}$$

Second, approximate $\frac{d^2\ell}{d\underline{\theta}^2}$ by

$$\frac{\partial^2\ell}{\partial\theta_m\partial\theta_n} = \sum_{k=1}^{n} \{\frac{1}{2}\mathrm{tr}\,[\underline{R}^{-1}(k)\ \frac{\partial R(k)}{\partial\theta_m}\ \underline{R}^{-1}(k)\ \frac{\partial R(k)}{\partial\theta_n}] + \mathrm{tr}\,[\underline{R}^{-1}(k)\ \frac{\partial\nu(k)}{\partial\theta_m}\ \frac{\partial\nu^T(k)}{\partial\theta_n}]\} \tag{25}$$

Note that the calculations required in Eq. (24) require that one construct and run a Kalman filter to give $\underline{\beta}(k)$ and $\underline{\nu}(k)$. Then, these equations must be differentiated with respect to $\underline{\theta}$ to give difference equations which must be sequentially solved to give

$$\frac{\partial \underline{\beta}(k)}{\partial \theta_m} \quad \text{and} \quad \frac{\partial \underline{\nu}(k)}{\partial \theta_m} \quad \text{for each} \quad m .$$

The computational burden of these operations can be considerable.

A major reduction in computations is achieved by avoiding differentiation of Kalman filters with respect to the elements of $\underline{\mu}_\theta$ and $\underline{\Sigma}_\theta$. For convenience of presentation, it is assumed here that the parameters to be estimated are the elements of $\underline{\mu}_\theta = \underline{\mu}$ and $\underline{\Sigma}_\theta = \underline{\Sigma}$ along with the parameters in $\underline{A}^i_\theta(t)$, $\underline{C}^i_\theta(t)$, $\underline{W}^i_\theta(t)$ and $\underline{R}^i_\theta(t)$. The $\underline{\mu},\underline{\Sigma}$ parameters are contained in the vector $\underline{\beta}$ and the remaining parameters are denoted by $\underline{\alpha}$ so $\underline{\theta} = [\underline{\beta}^T, \underline{\alpha}^T]^T$.

The $\underline{\mu},\underline{\Sigma}$ derivatives are avoided by expressing the likelihood in terms of the per-test maximum-likelihood estimate of $\underline{x}^i(0)$ denoted by $\hat{\underline{b}}^i$. This can be done because $\hat{\underline{b}}^i$, $i = 1,\ldots,M$ is a sufficient statistic for $\underline{\beta}$ as shown in references 4 and 5.

In terms of the sufficient statistic the negative log likelihood can be written as

$$- \log p(\underline{y}(1) \,|\, \underline{\mu},\underline{\Sigma})$$

$$= \frac{1}{2}\Big[\log \det (\underline{\Sigma} + \underline{P}(1)) + (\hat{\underline{b}}^i - \underline{\mu})^T (\underline{\Sigma} + \underline{P}(1))^{-1} (\hat{\underline{b}}^i - \underline{\mu})\Big] + T(\underline{\alpha}) \qquad (26)$$

where $T(\underline{\alpha})$ is independent of the initial-condition parameter $\underline{\beta}$ and where $\underline{P}(1)$ is the error covariance for $\hat{\underline{b}}^T$. The per-test estimate $\hat{\underline{b}}^i$ and its covariance can be obtained from the Kalman filter used to provide the innovations in (21). This is accomplished by augmenting the filter to provide a smoothed estimate of the initial condition $\hat{\underline{x}}^i(0)$ and the estimate-error covariance $\underline{P}^i(0)$ and then removing the $\underline{\mu},\underline{\Sigma}$ prior according to

$$\underline{P}(1) = (\underline{P}^i(0)^{-1} - \underline{\Sigma}^{-1})^{-1} \qquad (27)$$

$$\hat{\underline{b}}^i = \underline{P}(1)\,(\underline{P}^i(0)^{-1}\,\hat{\underline{x}}^i(0) - \underline{\Sigma}^{-1}\,\underline{\mu}) \qquad (28)$$

Note that in practice $\underline{\mu},\underline{\Sigma}$ often are to be estimated only on a subset of states of $\underline{x}^i$ that are not driven by process noise where the mean and covariance on the remaining states are either known or functions of $\underline{\alpha}$. The above results easily generalize to this case and the filter need not be augmented because states not driven by process noise are not smoothable.

Now the derivatives of the negative log likelihood with respect to $\underline{\mu},\underline{\Sigma}$ are straightforward to compute from (26) and do not involve differentiation of a Kalman filter. If there are many $\underline{\mu},\underline{\Sigma}$ parameters, particularly off diagonal elements of $\underline{\Sigma}$, this provides an enormous computational savings. The Fisher information relative to $\underline{\mu},\underline{\Sigma}$ can be used to approximate the Hessian of $\ell(\underline{\theta},\underline{\varepsilon}(\overline{1}))$

$$\frac{\partial^2 \ell(\theta,\varepsilon(i))}{\partial \underline{\mu}\,\partial \underline{\mu}^T} = (\underline{\Sigma} + \underline{P}(i))^{-1} \qquad (29)$$

$$\frac{\partial^2 \ell(\theta,\varepsilon(i))}{\partial \mu_j\,\partial \Sigma_{mn}} = 0 \quad \text{for all} \quad j, m, n \qquad (30)$$

$$\frac{\partial^2 \ell(\theta,\varepsilon(i))}{\partial \Sigma_{mn}\,\partial \Sigma_{pq}} = C_{mn}\,C_{pq}\,\big([\underline{\Sigma} + \underline{P}(i))^{-1}]_{mp}\,[(\underline{\Sigma} + \underline{P}(i))^{-1}]_{nq} + [(\underline{\Sigma} + \underline{P}(i)^{-1}]_{mq}\,[(\underline{\Sigma} + \underline{P}(i))^{-1}]_{np}\big) \qquad (31)$$

where C_{mn} is $\frac{1}{2}$ for $m = n$ and 1 for $m > n$.

The approximate Hessian relative to $\underline{\alpha}$ is still provided by equation (25), which requires a standard Kalman filter and its derivatives relative to $\underline{\alpha}$. If the cross terms in the Hessian between $\underline{\mu},\underline{\Sigma}$ and $\underline{\alpha}$ are set to zero, then the convergence of $\hat{\underline{\theta}}(i)$ follows from the type of results given in Section 3. Thus, the algorithm is a complete recursive algorithm guaranteed asymptotically to perform as well as batch maximum likelihood. Further, derivatives of the Kalman filter with respect to $\underline{\mu},\underline{\Sigma}$ have been completely eliminated.

5. NUMERICAL EXAMPLE

As an example we have considered the simple second-order system

$$\dot{x}^i_1(t) = x^i_2(t) \qquad (32)$$

$$\dot{x}^i_2(t) = \lambda^i(t) + n^i(t) \qquad (33)$$

where $\lambda^i(t)$ is a known deterministic input and $n(t)$ is stationary, colored, zero-mean Gaussian noise which is given by

$$\dot{n}^i(t) = -\frac{1}{\tau}\,n^i(t) + w^i(t)$$

with $w^i(t)$ a zero-mean white Gaussian process with power spectral density q. The initial-conditions $x^i_1(0)$, $x^i_2(0)$ are assumed to uncorrelated and

$$x^i_1(0) \sim N(u_1, \Sigma_{11}) \qquad x^i_2(0) \sim N(u_2, \Sigma_{22}) \qquad (34)$$

The system was subjected to uncorrelated measurements of the form

$$z^i(k) = x_1^i(k) + v^i(k) \tag{35}$$

with

$$v^i(k) \sim N(0, .0025) \tag{36}$$

The estimated parameters are

$$\underline{\beta} = [\mu_1, \Sigma_{11}, \mu_2, \Sigma_{22}]^T \quad , \quad \underline{\alpha} = [q \ , \ \tau]^T \tag{37}$$

Table 1 gives the results for the estimation process comparing the recursive algorithm developed here with batch maximum likelihood estimation. Twenty tests were processed each having 200 measurements 1 sec. apart. In order to start the recursive algorithm it was initialized with the batch result for the first five tests.

The results are seen to be in good agreement. As expected two-thirds of the results fall inside the $1-\sigma$ error bracket calculated from the Fisher information matrix.

Parameter	True Value	Starting Value	Batch Estimate	Recursive Estimate
μ_1	1.0	0.	1.27 ± .20	1.35
Σ_{11}	1.0	0.5	.77 ± .26	.84
μ_2	.5	0.	.44 ± .16	.50
Σ_{22}	.25	0.5	.43 ± .15	.49
$1/\tau$	.05	.04	.054 ± .005	.052
q	.025	.04	.0252 ± .001	.0252

Table 1
Parameter Estimates for Twenty Tests

6. CONCLUSIONS

We have exhibited a partitioned recursive algorithm for calculating the maximum likelihood estimate of the unknowns $\underline{\theta}$. We have proven that the recursion converges to the same estimates as the batch maximum likelihood approach when the number of tests tends to ∞. The partitioning of the parameters into initial condition and dynamic parameters provides a crucial improvement in numerical properties.

Several areas of further research are believed to be important. Many of the potential applications of these ideas involve nonlinear systems. Since the results presented here rely primarily on derivatives of the log likelihood there is hope that they could be extended to apply to reasonable classes of nonlinear systems. Also, it would be very useful to have estimates of how the rate of convergence depends on the scheme for finding $\underline{R}(i)$. Another practical problem arises from the fact that tests can be correlated due, for example, to common test instrumentation. Further, it would be desirable to obtain the per-test maximum likelihood estimates $\hat{\underline{b}}^i$ directly rather than fom an augmented state smoother.

REFERENCES

[1] R. L. Goodrich and P. E. Caines, "Linear System Identification from Nonstationary Cross-Sectional Data", IEEE Trans. Automat. Contr., Vol. AC-24, no. 3 pp. 403-411, June 1979.

[2] K. J. Astrom, "Maximum Likelihood and Prediction-Error Methods", Automatica, Vol. 16, no. 5, pp. 551-574, Sept. 1980.

[3] L. Ljung, "Analysis of a General Recursive Prediction Error Identification Algorithm", Automatica, Vol. 17, no. 1, pp. 89-99, January 1981.

[4] 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, IA, 1978.

[5] 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", American Statistical Association Conference, San Diego, CA, August 1978.