

# Generalized Multiple-Model Adaptive Estimation Using an Autocorrelation Approach

Badr N. Alsuwaidan

*National Satellite Technology Program, KACST, Riyadh, Saudi Arabia*

John L. Crassidis

*University at Buffalo, State University of New York, Amherst, NY 14260-4400*

Yang Cheng

*Mississippi State University, Mississippi State, MS 39762*

## Abstract

In this paper a generalized multiple-model adaptive estimator is presented that can be used to estimate unknown model and/or filter parameters, such as the noise statistics in filter designs. The main goal of this work is to provide an increased convergence rate for the estimated model parameters over the traditional multiple-model adaptive estimator. Parameter elements generated from a quasi-random sequence are used to drive multiple parallel filters for state estimation. The current approach focuses on estimating the process noise covariance by sequentially updating weights associated with the quasi-random elements through the calculation of the likelihood function of the measurement-minus-estimate residuals, which also incorporates correlations between various measurement times. For linear Gaussian measurement processes the likelihood function is easily determined. A proof is provided that shows the convergence properties of the generalized approach versus the standard multiple-model adaptive estimator. Simulation results, involving a two-dimensional target tracking problem and a single-axis attitude problem, indicate that the new approach provides better convergence properties over a traditional multiple-model approach.

## Index Terms

Generalized Multiple-Model Adaptive Estimation, Correlation, Residual.

## I. INTRODUCTION

Multiple-model adaptive estimation (MMAE) uses a parallel bank of filters to provide multiple estimates, where each filter corresponds with a dependence on some unknowns. The state estimate is provided through a sum of each filter's estimate weighted by the likelihood of the unknown elements conditioned on the measurement sequence. The likelihood function gives the associated hypothesis that each filter is the correct one. Many applications of MMAE approaches exist [1]. Three generations of MMAE have been characterized. The first is the classical approach presented in Ref. 2. The second is the interacting multiple-model algorithm [3] and the third is variable structure multiple-model estimation [4]. Judiciously determining the filter bank size has also been an active area of research [5].

Adaptive filtering can be divided into four general categories: Bayesian, maximum likelihood, covariance matching, and correlation approaches [6]. Bayesian and maximum likelihood methods may be well suited to multiple-model approaches, but sometimes require large computational loads. Covariance matching

Badr Alsuwaidan is a Research Assistant Professor at the Space Research Institute, KACST. Email: bswaidan@kacst.edu.sa

John Crassidis is a Professor in the Department of Mechanical & Aerospace Engineering. Email: johnc@buffalo.edu

Yang Cheng is an Assistant Professor in the Department of Aerospace Engineering. Email: cheng@ae.msstate.edu

is the computation of the covariances from the residuals of the state estimation problem, but has been shown to give biased estimates of the true covariances. A widely used correlation-based approach for a linear Kalman filter with stationary/Gaussian process and measurement noise is based on “residual whitening” [7]. In particular, the autocorrelation matrix, which can be computed from the measurement-minus-estimate residuals, is used with the system state matrices to provide a least-squares estimate of the Kalman filter error covariance times the measurement output matrix. If the number of unknowns in the process noise covariance is equal to or less than the number of states times the number of outputs, then the error-covariance/output-matrix estimate can be used to find an estimate of the process noise covariance by solving for a set of linear equations. These equations are not linearly independent and one has to choose a linearly independent subset of these equations [7].

Standard MMAE approaches use only the current time measurement-minus-estimate residual in the algorithm. The approach shown in this paper is a generalization of the approach shown in Ref. [8], which uses the time correlation of the filter residuals to assign the likelihood for each of the modeled hypotheses. In particular, the spectral content of the residuals is used and only scalar measurements are assumed in Ref. [8]. The authors also state that if multiple measurements are available, then a diagonal matrix can be used with elements given by the spectral content of each measurement residual, but this assumes that the cross-correlation terms are negligible. Also, the focus of their paper is on the detection of actuator failures with known control-input frequency content.

The main goal of this work is to provide an increased convergence rate for the estimated parameters over the traditional multiple-model adaptive estimator. The new approach shown here, called generalized multiple-model adaptive estimation (GMMAE), is based on calculating the time-domain autocorrelation function [7], which is used to form the covariance of a generalized residual involving any number of backward time steps. This covariance matrix also includes the time correlated terms, thus providing a more rigorous approach than the approach shown in Ref. [8]. The unknown elements in the design are the parameters of the process noise covariance, but the approach shown here is general to any MMAE application and can be used to estimate any model parameters. Process noise covariance elements can be drawn from any sample distribution as long as the resulting covariance matrix remains positive semi-definite. A theoretical proof of the convergence properties of the GMMAE is also shown, which offers insight as to its advantages over the standard MMAE approach.

The organization of the remainder of this paper proceeds as follows. First, the standard Kalman filter equations are summarized, since this filter will be used in the simulations. Then, a review of the standard MMAE algorithm is given. Next, the new adaptive approach is shown, including the assumptions used for linear time-varying systems. The theoretical background for the GMMAE approach and the proof of convergence are then provided. Finally, simulation results involving a two-dimensional target tracking problem and a single-axis attitude estimation problem are shown.

## II. KALMAN FILTER AND MULTIPLE-MODEL ADAPTIVE ESTIMATION

A summary of the discrete Kalman filter is given in Table I, where  $\mathbf{x}_k$  is the state vector,  $\mathbf{u}_k$  is the known control input,  $\Phi_k$  is the discrete-time state transition matrix,  $\Upsilon_k$  is the process noise distribution matrix,  $\mathbf{w}_k$  is the process noise vector which is assumed to be a zero-mean Gaussian noise process with covariance  $Q_k$ ,  $\tilde{\mathbf{y}}_k$  is the discrete-time measurement,  $\mathbf{v}_k$  is the measurement noise vector which is assumed to be a zero-mean Gaussian noise process with covariance  $R_k$ ,  $\hat{\mathbf{x}}_k^-$  and  $\hat{\mathbf{x}}_k^+$  are the propagated and updated state estimates, respectively, and  $P_k^-$  and  $P_k^+$  are the propagated and updated covariances, respectively.

Multiple-model adaptive estimation is a recursive estimator that uses a bank of  $M$  filters that depend on some unknown parameters, denoted by the vector  $\mathbf{p}$ , which is assumed to be constant (at least throughout the interval of adaptation). The MMAE process is shown in Figure 1. Details of the MMAE approach can

TABLE I  
DISCRETE-TIME LINEAR KALMAN FILTER

<b>Model</b>	$\mathbf{x}_{k+1} = \Phi_k \mathbf{x}_k + \Gamma_k \mathbf{u}_k + \Upsilon_k \mathbf{w}_k, \quad \mathbf{w}_k \sim N(\mathbf{0}, Q_k)$ $\tilde{\mathbf{y}}_k = H_k \mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim N(\mathbf{0}, R_k)$
<b>Initialize</b>	$\hat{\mathbf{x}}(t_0) = \tilde{\mathbf{x}}_0$ $P_0 = E \{ \tilde{\mathbf{x}}(t_0) \tilde{\mathbf{x}}^T(t_0) \}$
<b>Gain</b>	$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1}$
<b>Update</b>	$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + K_k [\tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k^-]$ $P_k^+ = [I - K_k H_k] P_k^-$
<b>Propagation</b>	$\hat{\mathbf{x}}_{k+1}^- = \Phi_k \hat{\mathbf{x}}_k^+ + \Gamma_k \mathbf{u}_k$ $P_{k+1}^- = \Phi_k P_k^+ \Phi_k^T + \Upsilon_k Q_k \Upsilon_k^T$

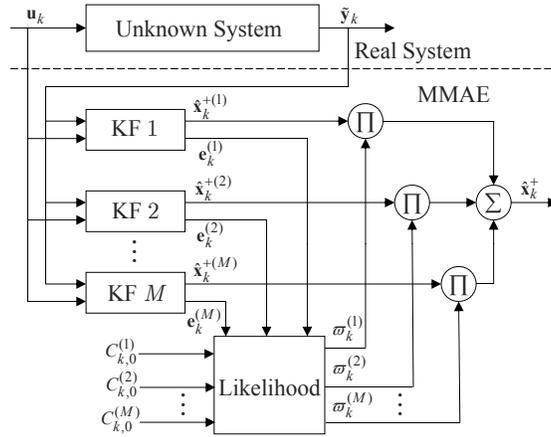


Fig. 1. MMAE Process

be found in Refs. [9] and [10]. Note that we do not need to make the stationary assumption for the state and/or output processes though, i.e. time-varying state and output matrices can be used. It is assumed that the unknown parameters belong to a discrete set of elements generated for each of the  $M$  filters from some known probability density function (pdf) of  $\mathbf{p}$ , denoted by  $p(\mathbf{p})$ , to give  $\{\mathbf{p}^{(\ell)}; \ell = 1, \dots, M\}$ . The goal of the estimation process is to determine the conditional pdf of the  $\ell^{\text{th}}$  element  $\mathbf{p}^{(\ell)}$  given all

the measurements. Application of Bayes' rule yields

$$\begin{aligned} p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_k) &= \frac{p(\tilde{\mathbf{Y}}_k, \mathbf{p}^{(\ell)})}{p(\tilde{\mathbf{Y}}_k)} \\ &= \frac{p(\tilde{\mathbf{Y}}_k|\mathbf{p}^{(\ell)})p(\mathbf{p}^{(\ell)})}{\sum_{j=1}^M p(\tilde{\mathbf{Y}}_k|\mathbf{p}^{(j)})p(\mathbf{p}^{(j)})} \end{aligned} \quad (1)$$

where  $\tilde{\mathbf{Y}}_k$  denotes the sequence  $\{\tilde{\mathbf{y}}_0, \tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_k\}$ . The *a posteriori* probabilities can be computed through [11]

$$\begin{aligned} p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_k) &= \frac{p(\tilde{\mathbf{y}}_k, \mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_{k-1})}{p(\tilde{\mathbf{y}}_k|\tilde{\mathbf{Y}}_{k-1})} \\ &= \frac{p(\tilde{\mathbf{y}}_k|\hat{\mathbf{x}}_k^{-(\ell)})p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_{k-1})}{\sum_{j=1}^M \left[ p(\tilde{\mathbf{Y}}_k|\hat{\mathbf{x}}_k^{-(j)})p(\mathbf{p}^{(j)}|\tilde{\mathbf{Y}}_{k-1}) \right]} \end{aligned} \quad (2)$$

with  $p(\tilde{\mathbf{y}}_k|\tilde{\mathbf{Y}}_{k-1}, \mathbf{p}^{(\ell)})$  given by  $p(\tilde{\mathbf{y}}_k|\hat{\mathbf{x}}_k^{-(\ell)})$  in the Kalman recursion. It is computed as a pdf of the Kalman filter measurement-minus-estimate residual  $\mathbf{e}_k = \tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k^-$ . Suppose that  $\mathbf{e}_k^{(\ell)}$  is a zero-mean Gaussian random vector with covariance  $C_{k,0}^{(\ell)}$ . Then

$$p(\tilde{\mathbf{y}}_k|\tilde{\mathbf{Y}}_{k-1}, \mathbf{p}^{(\ell)}) = 1/\sqrt{\det[2\pi C_{k,0}^{(\ell)}]} \exp \left[ -\frac{1}{2} \mathbf{e}_k^T \left( C_{k,0}^{(\ell)} \right)^{-1} \mathbf{e}_k \right] \quad (3)$$

Note that the denominator of (2) is just a normalizing factor to ensure that  $p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_k)$  is a pdf. The *a posteriori* probability update in (2) recursion formula can now be cast into a set of defined weights  $\varpi_k^{(\ell)}$ , so that

$$\varpi_k^{(\ell)} = \frac{\varpi_{k-1}^{(\ell)} p(\tilde{\mathbf{y}}_k|\hat{\mathbf{x}}_k^{-(\ell)})}{\sum_{j=1}^M \varpi_{k-1}^{(j)} p(\tilde{\mathbf{y}}_k|\hat{\mathbf{x}}_k^{-(j)})} \quad (4)$$

where  $\varpi_k^{(\ell)} \equiv p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_k)$ . Note that only the current time measurement  $\tilde{\mathbf{y}}_k$  is needed to update the weights. The weights at time  $t_0$  are initialized to  $\varpi_0^{(\ell)} = 1/M$  for  $\ell = 1, 2, \dots, M$ . The convergence properties of MMAE are shown in Ref. [11], which assumes ergodicity in the proof. The ergodicity assumptions can be relaxed to asymptotic stationarity and other assumptions are even possible for non-stationary situations [12].

The conditional mean estimate is the weighted sum of the parallel filter estimates  $\hat{\mathbf{x}}_k^{-(j)}$ :

$$\hat{\mathbf{x}}_k^- = \sum_{j=1}^M \varpi_k^{(j)} \hat{\mathbf{x}}_k^{-(j)} \quad (5)$$

Also, the error covariance of the state estimate can be computed using

$$P_k^- = \sum_{j=1}^M \varpi_k^{(j)} \left[ \left( \hat{\mathbf{x}}_k^{-(j)} - \hat{\mathbf{x}}_k^- \right) \left( \hat{\mathbf{x}}_k^{-(j)} - \hat{\mathbf{x}}_k^- \right)^T + P_k^{-(j)} \right] \quad (6)$$

The specific estimate for  $\mathbf{p}$  at time  $t_k$ , denoted by  $\hat{\mathbf{p}}_k$ , and error covariance, denoted by  $\mathcal{P}_k$ , are given by

$$\hat{\mathbf{p}}_k = \sum_{j=1}^M \varpi_k^{(j)} \mathbf{p}^{(j)} \quad (7a)$$

$$\mathcal{P}_k = \sum_{j=1}^M \varpi_k^{(j)} \left( \mathbf{p}^{(j)} - \hat{\mathbf{p}}_k \right) \left( \mathbf{p}^{(j)} - \hat{\mathbf{p}}_k \right)^T \quad (7b)$$

Equation (7b) can be used to define  $3\sigma$  bounds on the estimate  $\hat{\mathbf{p}}_k$ . The discrete distribution of the uncertainty of the parameter vector  $\mathbf{p}$  is represented by  $M$  elements of  $\mathbf{p}^{(j)}$ . If  $M$  is large and the significant regions of the parameter space of  $\mathbf{p}$  are well represented by  $\mathbf{p}^{(j)}$ , then (7a) is a good approximation of the conditional mean of  $\mathbf{p}$ .

### III. ADAPTIVE LAW BASED ON AUTOCORRELATION

In this section the adaptive law, based on an autocorrelation approach, for the process noise covariance matrix is shown. First, the autocorrelation for time-varying systems is derived, followed by the associated likelihood functions for the defined measurement residuals.

#### A. Autocorrelation for Time-Varying Systems

In this section the autocorrelation matrix for linear time-varying systems is derived, which is an extension to the approach shown in Ref. [7]. Here we assume that the model is linear with

$$\mathbf{x}_{k+1} = \Phi_k \mathbf{x}_k + \Gamma_k \mathbf{u}_k + \Upsilon_k \mathbf{w}_k \quad (8a)$$

$$\tilde{\mathbf{y}}_k = H_k \mathbf{x}_k + \mathbf{v}_k \quad (8b)$$

Note that the to-be-estimated parameter vector  $\mathbf{p}$  can be function of any unknown model quantity in (8). For example it can be related to parameters in the state model matrices or parameters related to the covariance of the process and/or measurement noise. Here we focus on elements of the process noise covariance. The process noise is assumed to be a zero-mean Gaussian white-noise process so that  $Q_k$  is a constant, denoted by  $Q$ . Consider the following discrete-time residual equation:

$$\begin{aligned} \mathbf{e}_k &\equiv \tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k^- \\ &= -H_k \tilde{\mathbf{x}}_k^- + \mathbf{v}_k \end{aligned} \quad (9)$$

where  $\tilde{\mathbf{x}}_k^- \equiv \hat{\mathbf{x}}_k^- - \mathbf{x}_k$  and  $\hat{\mathbf{x}}_k^- = \Phi_{k-1}(I - K_{k-1}H_{k-1})\hat{\mathbf{x}}_{k-1}^- + \Phi_{k-1}K_{k-1}\tilde{\mathbf{y}}_{k-1}$  with  $K_{k-1}$  being an arbitrary gain. Note that  $K_{k-1}$  is not necessarily the optimal Kalman gain, which is computed using the actual noise statistics. The following autocorrelation matrix can be computed:

$$C_{k,i} = \begin{cases} H_k P_k^- H_k^T + R_k & i = 0 \\ H_k E \{ \tilde{\mathbf{x}}_k^- \tilde{\mathbf{x}}_{k-i}^{-T} \} H_{k-i}^T - H_k E \{ \tilde{\mathbf{x}}_k^- \mathbf{v}_{k-i}^T \} & i > 0 \end{cases} \quad (10)$$

where  $C_{k,i} \equiv E \{ \mathbf{e}_k \mathbf{e}_{k-i}^T \}$  and  $E\{\cdot\}$  denotes expectation. The propagation of  $\tilde{\mathbf{x}}_k^-$  is given by

$$\tilde{\mathbf{x}}_k^- = \Phi_{k-1}(I - K_{k-1}H_{k-1})\tilde{\mathbf{x}}_{k-1}^- + \Phi_{k-1}K_{k-1}\mathbf{v}_{k-1} - \Upsilon_{k-1}\mathbf{w}_{k-1} \quad (11)$$

Carrying (11)  $i$  steps back leads to

$$\begin{aligned} \tilde{\mathbf{x}}_k^- = & \left[ \prod_{j=1}^i \Phi_{k-j} (I - K_{k-j} H_{k-j}) \right] \tilde{\mathbf{x}}_{k-i}^- + \sum_{j=2}^i \left[ \prod_{\ell=1}^{j-1} \Phi_{k-\ell} (I - K_{k-\ell} H_{k-\ell}) \right] \Phi_{k-j} K_{k-j} \mathbf{v}_{k-j} \\ & - \sum_{j=2}^i \left[ \prod_{\ell=1}^{j-1} \Phi_{k-\ell} (I - K_{k-\ell} H_{k-\ell}) \right] \Upsilon_{k-j} \mathbf{w}_{k-j} + \Phi_{k-1} K_{k-1} \mathbf{v}_{k-1} - \Upsilon_{k-1} \mathbf{w}_{k-1} \end{aligned} \quad (12)$$

where

$$\prod_{j=1}^i Z_{k-j} \equiv Z_{k-1} Z_{k-2} \cdots Z_{k-i} \quad (13)$$

Performing the expectations in the definition of  $C_{k,i}$  leads to [13]

$$C_{k,i} = \begin{cases} H_k P_k^- H_k^T + R_k & i = 0 \\ H_k \Phi_{k-1} (P_{k-1}^- H_{k-1}^T - K_{k-1} C_{k-1,0}) & i = 1 \\ H_k \left[ \prod_{j=1}^{i-1} \Phi_{k-j} (I - K_{k-j} H_{k-j}) \right] \Phi_{k-i} (P_{k-i}^- H_{k-i}^T - K_{k-i} C_{k-i,0}) & i > 1 \end{cases} \quad (14)$$

where

$$C_{k-i,0} \equiv H_{k-i} P_{k-i}^- H_{k-i}^T + R_{k-i} \quad (15)$$

In the above equations,  $P_k^-$  and  $P_{k-1}^-$  are computed using the true covariance matrix  $Q$  of the process noise  $\mathbf{w}_k$ . If the  $K_{k-j}$  matrices are computed using the true  $Q$ , too, then  $C_{k,i} = 0$  for  $i \geq 1$ . Note that storage of the state model and covariance matrices to the  $k-i$  point is required to compute  $C_{k,i}$  in general.

### B. Likelihood Function

In this section the likelihood function for the measurement residual is shown. First, the following residual is defined:

$$\boldsymbol{\epsilon}_{k,i} \equiv \begin{bmatrix} \mathbf{e}_k \\ \mathbf{e}_{k-1} \\ \vdots \\ \mathbf{e}_{k-i} \end{bmatrix} \quad (16)$$

The likelihood function associated with  $\boldsymbol{\epsilon}_{k,i}$  is given by

$$L_{k,i} = \frac{1}{[\det(2\pi \mathcal{C}_{k,i})]^{1/2}} \exp \left( -\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^T \mathcal{C}_{k,i}^{-1} \boldsymbol{\epsilon}_{k,i} \right) \quad (17)$$

where  $\mathcal{C}_{k,i} \equiv E\{\boldsymbol{\epsilon}_{k,i} \boldsymbol{\epsilon}_{k,i}^T\}$  is given by

$$\mathcal{C}_{k,i} = \begin{bmatrix} C_{k,0} & C_{k,1} & C_{k,2} & \cdots & C_{k,i} \\ C_{k,1}^T & C_{k-1,0} & C_{k-1,1} & \cdots & C_{k-1,i-1} \\ C_{k,2}^T & C_{k-1,1}^T & C_{k-2,0} & \cdots & C_{k-2,i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{k,i}^T & C_{k-1,i-1}^T & C_{k-2,i-2}^T & \cdots & C_{k-i,0} \end{bmatrix} \quad (18)$$

When  $i = 0$  the likelihood function reduces down to

$$L_{k,0} = \frac{1}{\{\det[2\pi (H_k P_k^- H_k^T + R_k)]\}^{1/2}} \exp \left[ -\frac{1}{2} \mathbf{e}_k^T (H_k P_k^- H_k^T + R_k)^{-1} \mathbf{e}_k \right] \quad (19)$$

This likelihood is widely used in MMAE algorithms [9].

### C. MMAE Adaptive Law

The MMAE adaptive law is given by

$$\begin{aligned} \varpi_k^{(\ell)} &= \varpi_{k-1}^{(\ell)} L_{k,0}^{(\ell)} \\ \varpi_k^{(\ell)} &\leftarrow \frac{\varpi_k^{(\ell)}}{\sum_{j=1}^M \varpi_k^{(j)}} \end{aligned} \quad (20)$$

since  $p(\tilde{\mathbf{y}}_k | \hat{\mathbf{x}}_k^{-(\ell)}) = L_{k,0}^{(\ell)}$ , which is defined by

$$L_{k,0}^{(\ell)} = \frac{1}{\{\det[2\pi (H_k P_k^{-(\ell)} H_k^T + R_k)]\}^{1/2}} \exp \left[ -\frac{1}{2} \mathbf{e}_k^{(\ell)T} (H_k P_k^{-(\ell)} H_k^T + R_k)^{-1} \mathbf{e}_k^{(\ell)} \right] \quad (21)$$

where  $\mathbf{e}_k^{(\ell)} \equiv \tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k^{-(\ell)}$ .

The actual statistics of  $\mathbf{e}_k^{(\ell)}$  are determined by  $Q$  and  $K_{k-j}^{(\ell)}$ . The Kalman gains  $K_{k-j}^{(\ell)}$  are computed under the  $\ell^{\text{th}}$  hypothesis that the true covariance of the process noise is  $Q^{(\ell)}$ . In the likelihood function above, the hypothesized residual covariance  $(H_k P_k^{-(\ell)} H_k^T + R_k)$  of  $\mathbf{e}_k^{(\ell)}$  is computed using  $Q^{(\ell)}$  and  $K_{k-j}^{(\ell)}$ . The actual statistics of the residuals match the hypothesized only when  $Q^{(\ell)} = Q$ .

### D. GMMAE Adaptive Law

The GMMAE adaptive law is a modification of the MMAE adaptive law, given by

$$\begin{aligned} \varpi_k^{(\ell)} &= \varpi_{k-1}^{(\ell)} L_{k,i}^{(\ell)} \\ \varpi_k^{(\ell)} &\leftarrow \frac{\varpi_k^{(\ell)}}{\sum_{j=1}^M \varpi_k^{(j)}} \end{aligned} \quad (22)$$

with

$$L_{k,i}^{(\ell)} = \frac{1}{\left[ \det \left( 2\pi \mathcal{C}_{k,i}^{(\ell)} \right) \right]^{1/2}} \exp \left[ -\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \boldsymbol{\epsilon}_{k,i}^{(\ell)} \right] \quad (23)$$

where  $\boldsymbol{\epsilon}_i^{(\ell)}$  is defined as  $\boldsymbol{\epsilon}_i^{(\ell)} \equiv [\mathbf{e}_k^{(\ell)T} \ \mathbf{e}_{k-1}^{(\ell)T} \ \dots \ \mathbf{e}_{k-i}^{(\ell)T}]^T$ , of which the actual statistics are determined by  $Q$  and  $K_{k-j}^{(\ell)}$ , which are computed using  $Q^{(\ell)}$ . The matrix  $\mathcal{C}_{k,i}^{(\ell)}$  is given by (14) and (18), which is evaluated at  $Q^{(\ell)}$  and the *optimal* Kalman gain, which is a function of the actual covariance  $Q$ . Because  $Q$  is unknown, it is approximated by the current best estimate  $\hat{Q}$  of it, computed using  $\hat{\mathbf{p}}_k$ .

Like the MMAE adaptive law, the GMMAE adaptive law is set up such the parameter sets used to compute  $Q^{(\ell)}$  that result in good agreement between the actual and expected residual statistics receive

large weights. On the right-hand side of (22), the more recent weight  $\varpi_{k-1}$  instead of  $\varpi_{k-i}$  is used, which in essence raises the power of the weight and the likelihood function. As a result, the GMMAE estimate given by (7) of the unknown  $\mathbf{p}$  is closer to the maximum of the likelihood than it is towards the conditional likelihood mean or the minimum mean-square error estimate as provided by MMAE. A technique of raising the power of the weights in the MMAE approach has also been proposed in Ref. [14]. But the GMMAE solution presented here provides a clearer understanding of how this is done and it also incorporates correlated terms to increase performance characteristics.

The estimates for  $C_{k,i}^{(\ell)}$ , which are sub-blocks of  $C_{k,i}^{(\ell)}$ , are given by

$$\hat{C}_{k,i}^{(\ell)} = \begin{cases} H_k P_k^{-\ell} H_k^T + R_k & i = 0 \\ H_k \Phi_{k-1} \left( P_{k-1}^{-\ell} H_{k-1}^T - \hat{K}_{k-1} C_{k-1,0}^{(\ell)} \right) & i = 1 \\ H_k \left[ \prod_{j=1}^{i-1} \Phi_{k-j} \left( I - \hat{K}_{k-j} H_{k-j} \right) \right] \Phi_{k-i} \\ \times \left( P_{k-i}^{-\ell} H_{k-i}^T - \hat{K}_{k-i} C_{k-i,0}^{(\ell)} \right) & i > 1 \end{cases} \quad (24)$$

where

$$C_{k-i,0}^{(\ell)} \equiv H_{k-i} P_{k-i}^{-\ell} H_{k-i}^T + R_{k-i} \quad (25)$$

The covariance matrix  $P_k^{-\ell}$  is computed using

$$P_{k+1}^{-\ell} = \Phi_k P_k^{+\ell} \Phi_k^T + Q^{(\ell)} \quad (26a)$$

$$P_k^{+\ell} = \left( I - K_k^{(\ell)} H_k \right) P_k^{-\ell} \quad (26b)$$

$$K_k^{(\ell)} = P_k^{-\ell} H_k^T \left( H_k P_k^{-\ell} H_k^T + R_k \right)^{-1} \quad (26c)$$

where  $Q^{(\ell)}$  is computed using  $\mathbf{p}^{(\ell)}$ . The estimate of the optimal gain is computed using

$$\hat{K}_k = \hat{P}_k^- H_k^T \left( H_k \hat{P}_k^- H_k^T + R_k \right)^{-1} \quad (27)$$

with

$$\hat{P}_{k+1}^- = \Phi_k \hat{P}_k^+ \Phi_k^T + \hat{Q}_k \quad (28a)$$

$$\hat{P}_k^+ = \left( I - \hat{K}_k H_k \right) \hat{P}_k^- \quad (28b)$$

Using the current measurement,  $\tilde{\mathbf{y}}_k$ , along with the  $\ell^{\text{th}}$  element,  $\mathbf{p}^{(\ell)}$ ,  $1 \leq \ell \leq M$ , a bank of filters is executed. For each filter the state estimates,  $\hat{\mathbf{x}}_k^{-\ell}$ , and measurements are used to form the residual,  $\epsilon_{k,i}^{(\ell)}$ , going back  $i$  steps. The filter error covariance,  $P_k^{-\ell}$ , and state matrices,  $\Phi_k^{-\ell}$  and  $H_k^{-\ell}$ , are used to update the estimate of the autocorrelation, denoted by  $\hat{C}_{k,i}^{(\ell)}$ . Note that at each new measurement time, all elements of  $\hat{C}_{k,i}^{(\ell)}$  need to be recalculated since a new estimate  $\hat{\mathbf{p}}_k$  is provided, which is used to compute an estimate of the optimal gain. Unfortunately, this can increase the computational costs. The diagonal elements do not need to be recomputed though, since they are not a function of the optimal gain. The residuals and autocorrelations are then used to evaluate the likelihood functions  $L_{k,i}^{(\ell)}$ . These functions are used to update the weights, which gives the estimate  $\hat{\mathbf{p}}_k$  using (7a).

#### IV. GMMAE CONVERGENCE<sup>1</sup>

A proof of convergence of the GMMAE approach for linear systems is shown, as well as its relationship to the standard MMAE approach.

##### A. Autocorrelation of the Residuals

One way to estimate the correlation of the residuals (the product of two residual values) is by using the ergodic average (time average). However, the GMMAE approach is a realtime adaptive approach and the estimate of the correlation of the residuals has to be determined and provided to the GMMAE likelihood function to test the new measurement-residual optimality for the given parameter  $\mathbf{p}^{(\ell)}$ . The GMMAE estimate  $\hat{C}_{k,i}$  provides an estimate for the correlation of the residuals.

The residuals of the optimal Kalman filter, in which the Kalman gain is computed using the actual noise covariances, are uncorrelated. However, the residuals of the suboptimal Kalman filter are correlated. There are different ways to estimate the autocorrelation. One way uses the ergodicity property of a stationary random process. The ergodic estimate of the autocorrelation is given by

$$\bar{C}_{k-1,i} = \frac{1}{N} \sum_{j=i}^N \mathbf{e}_j \mathbf{e}_{j-i}^T \quad (29)$$

where  $N$  is a large number of samples. The ergodic estimate of the autocorrelation  $\bar{C}_{k-1,i}$  can be used to estimate the true state error  $\{\hat{\mathbf{x}}_{k-1}^- \hat{\mathbf{x}}_{k-1}^{-T}\}$  which is the approach introduced in Ref. [7]. The estimate of  $\{\tilde{\mathbf{x}}_{k-1}^- \tilde{\mathbf{x}}_{k-1}^{-T}\}$  cannot typically be determined, but rather  $\{\tilde{\mathbf{x}}_{k-1}^- \tilde{\mathbf{x}}_{k-1}^{-T}\} H_{k-i}^T$  is determined instead. The term  $\{\tilde{\mathbf{x}}_{k-1}^- \tilde{\mathbf{x}}_{k-1}^{-T}\} H_{k-i}^T$  is denoted by  $\widehat{P_{k-i}^{-(\ell)}} H_{k-i}^T$  and can be estimated from (14) using the autocorrelation ergodic estimate  $\bar{C}_{k-1,i}$ :

$$[\bar{C}_{k,i}^{(\ell)T} \dots \bar{C}_{k-i+1,1}^{(\ell)T}]^T = A_{k,i} [\widehat{P_{k-i}^{-(\ell)}} H_{k-i}^T - K_{k-i}^{(\ell)} C_{k-i,0}^{(\ell)}] \quad (30)$$

where  $A_{k,i}$  is defined as

$$A_{k,i} = \begin{bmatrix} H_k \left[ \prod_{j=1}^{i-1} \Phi_{k-j} \left( I - \hat{K}_{k-j} H_{k-j} \right) \right] \Phi_{k-i} \\ \vdots \\ H_{k-i+1} \Phi_{k-i} \end{bmatrix} \quad (31)$$

The estimated optimal gain  $\hat{K}_{k-1}$  would be the optimal gain to minimize the state error in a least-squares sense. Noticing  $\hat{K}_{k-i} = \widehat{P_{k-i}^{-(\ell)}} H_{k-i}^T C_{k-i,0}^{-1}$ , the optimal gain is estimated as

$$[\bar{C}_{k,i}^{(\ell)T} \dots \bar{C}_{k-i+1,1}^{(\ell)T}]^T \cong A_{k,i} [\hat{K}_{k-i} - K_{k-i}^{(\ell)}] C_{k-i,0}^{(\ell)} \quad (32)$$

$$\hat{K}_{k-i} \cong K_{k-i}^{(\ell)} + A_{k,i}^\dagger [\bar{C}_{k,i}^{(\ell)T} \dots \bar{C}_{k-i+1,1}^{(\ell)T}]^T C_{k-i,0}^{(\ell)-1} \quad (33)$$

where  $A^\dagger$  is the pseudo-inverse of  $A$ . This provides another estimate to the autocorrelation of the residuals of a suboptimal filter.

The GMMAE approach is a recursive adaptive approach. The result of the previous time evaluated to determine the current estimate will be used to predict future estimate. The GMMAE estimate of the optimal gain  $\hat{K}_{k-1}$  in (24) is determined by the weight  $\varpi_{k-1}$ , which is a measure of the likelihood of

<sup>1</sup>This section is the work of the first two authors.

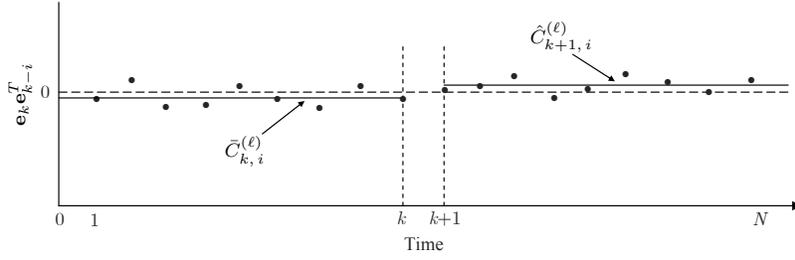


Fig. 2. Correlation of the Residuals of an Optimal Kalman Filter

the residual  $\mathbf{e}_{k-1} = [\tilde{\mathbf{y}}_{k-1} - H_k \hat{\mathbf{x}}_k^-]$  and the previous residuals. The estimate  $\hat{\mathbf{x}}_{k-1}^-$  is the Kalman filter propagation of  $\hat{\mathbf{x}}_{k-2}^+$  updated by  $\hat{K}_{k-2}$ . Therefore, the weight  $\varpi_{k-1}$  is a measure of the performance of  $\hat{K}_{k-2}$  since  $\hat{K}_{k-1}$  has not contributed to the residual  $\mathbf{e}_{k-1}$  yet. The filter that has the highest weight  $\varpi_{k-1}$  has the minimum residual  $\mathbf{e}_{k-1}$  and the Kalman gain estimate  $\hat{K}_{k-2} \rightarrow K_{k-2}^{(\ell)}$  as  $\hat{P}_{k-2} \rightarrow P_{k-2}^{-(\ell)}$ . The measurement  $\tilde{\mathbf{y}}_{k-1}$  complies more with the estimate  $\hat{\mathbf{x}}_k^-$ , where  $\hat{K}_{k-2}$  is most likely to be the optimal gain. The estimate of the optimal gain  $\hat{K}_{k-1}$  given by the GMMAE algorithm in (24) is actually a propagation of  $\hat{K}_{k-2}$  determined by  $\varpi_{k-1}$  based on the likelihood evolution of the residuals performance of the gain estimate  $\hat{K}_{k-2}$  as shown in (27). Any estimate based on the weight  $\varpi_{k-1}$  refers to  $\hat{K}_{k-2}$  in reality. For the suboptimal filter, the ergodic estimate of the correlation of the residuals,  $\bar{C}_{k-1,1}^{(\ell)}$ , is given using (32) by

$$\bar{C}_{k-1,1}^{(\ell)} = A_{k-1,1} [\hat{K}_{k-2} - K_{k-2}^{(\ell)}] C_{k-2,0}^{(\ell)} \quad (34)$$

The GMMAE estimate of the autocorrelation  $\hat{C}_{k,i}^{(\ell)}$  in (24), that is used in the GMMAE likelihood, is based on the gain estimate of  $\hat{K}_{k-1}$ :

$$[\hat{C}_{k,i}^{(\ell)T} \dots \hat{C}_{k-i+1,1}^{(\ell)T}]^T = A_{k,i} [P_{k-i}^{-(\ell)} H_{k-i}^T - K_{k-i}^{(\ell)}] C_{k-i,0}^{(\ell)} \quad (35)$$

In the GMMAE hypothesis, the likelihood of the residual is evaluated with respect to its optimal expected values. The assumption of optimality of the parameter  $\mathbf{p}^{(\ell)}$  is essential to compute the likelihood. Even though the zero residual correlation would be the obvious choice for optimality, the estimate of the correlation,  $\hat{C}_{k,i}^{(\ell)}$ , is based on the available information of the previous correlation  $\bar{C}_{k-1,i}^{(\ell)}$  carried by the GMMAE estimate of the optimal gain  $\hat{K}_{k-i}^{(\ell)}$ . The actual correlation of the residuals of the optimal filter is a random variable and has zero mean. Basically, the optimal filter correlation of the residuals,  $\hat{C}_{k,i}^{(\ell)}$ , is expected to compensate for the previous correlation as shown in Figure 2. If the parameter  $\mathbf{p}^{(\ell)}$  is an optimal parameter, the correlation of the residuals has zero mean while the actual correlation fluctuates around this mean. Considering a linear time-invariant system, and substituting the Kalman gain estimate  $\hat{K}_{k-2}$  from (34) into (35), where  $\hat{K}_{k-1} = \hat{K}_{k-2}$  at steady state, leads to

$$\begin{aligned} \hat{C}_{k,1}^{(\ell)} &= A_{k,1} [P_{k-1}^{-(\ell)} H_{k-1}^T - (K_{k-2}^{(\ell)} + A_{k-1,1}^\dagger \bar{C}_{k-1,1}^{(\ell)})] C_{k-2,0}^{(\ell)-1} C_{k-1,0}^{(\ell)} \\ \hat{C}_{k,1}^{(\ell)} &= -A_{k,1} A_{k-1,1}^\dagger [\bar{C}_{k-1,1}^{(\ell)} C_{k-2,0}^{(\ell)-1}] C_{k-1,0}^{(\ell)} \\ \hat{C}_{k,1}^{(\ell)} &= -\bar{C}_{k-1,1}^{(\ell)} \end{aligned} \quad (36)$$

The predicted estimate  $\hat{C}_{k,1}^{(\ell)}$  has the opposite sign of the ergodic estimate [15]. To ensure the optimality of the Kalman filter, the expected future correlation has to compensate any divergence of the previous

correlation of the zero mean. The autocorrelation  $\hat{C}_{k,1}^{(\ell)}$  is incorporated in the GMMAE likelihood to increase the chance that the parameter  $\mathbf{p}^{(\ell)}$  is chosen as the optimal one. For the case that the residuals are produced by a suboptimal filter, the correlation offset from zero mean is not only a result of the randomness of the residual but also a result of the mis-modeling of the filter parameter. The correlation for the suboptimal filter will converge at the limit to the ergodic correlation estimate  $\bar{C}_{k,1}^{(\ell)}$  and the relation given in (36) will serve as a tool in the GMMAE adaptive law to distinguish between the optimal filter and suboptimal filter residuals by testing for the whiteness of the filter residuals.

### B. GMMAE Convergence

We expect that for the GMMAE approach to converge, one of the elements  $\mathbf{p}^{(\ell)}$  is nearly equal to the true parameter  $\mathbf{p}^{\text{true}}$  so that  $p(\mathbf{p}^{\text{true}}|\tilde{\mathbf{Y}}_{k,i}) \rightarrow 1$  as  $k \rightarrow \infty$  and  $p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_{k,i}) \rightarrow 0$  as  $k \rightarrow \infty$  for  $\mathbf{p}^{(\ell)} \neq \mathbf{p}^{\text{true}}$ . Also, the  $\mathbf{p}^{(\ell)}$  elements that have values far from the true parameter have faster rates of convergence to zero,  $p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_{k,i}) \rightarrow 0$  as  $k \rightarrow \infty$ , than the ones closer to the true parameter. This result will be proved with the assumption of ergodicity of residuals and the use of the following lemma of standard matrix theory [12]. Let  $A$  and  $B$  be two  $n \times n$  positive definite matrices, then

$$n + \ln \left( \frac{|A|}{|B|} \right) - \text{tr}[B^{-1}A] \leq 0 \quad (37)$$

where  $\text{tr}$  denotes the matrix trace. The convergence of a linear time-invariant model is first considered here. Recall the adaptive law for the GMMAE approach:

$$\varpi_k^{(\ell)} = \frac{L_{k,i}^{(\ell)} \varpi_{k-1}^{(\ell)}}{\sum_{j=1}^M L_{k,i}^{(j)} \varpi_{k-1}^{(j)}} \quad (38)$$

For simplicity the normalization in the denominator can be replaced by the dominating weight since all others will converge to zero eventually:

$$\varpi_k^{(\ell)} = \frac{L_{k,i}^{(\ell)} \varpi_{k-1}^{(\ell)}}{L_{k,i}^{\text{true}}} \quad (39)$$

where  $L_{k,i}^{\text{true}}$  is the likelihood evaluated using  $\mathbf{p}^{\text{true}}$ . Substituting for  $L_{k,i}$  from (23) into (39) gives

$$\varpi_k^{(\ell)} = \frac{\left| \mathcal{C}_{k,i}^{(\ell)} \right|^{-1/2} \exp \left[ -\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \boldsymbol{\epsilon}_{k,i}^{(\ell)} \right]}{\left| \mathcal{C}_{k,i}^{(O)} \right|^{-1/2} \exp \left[ -\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(O)T} \left( \mathcal{C}_{k,i}^{(O)} \right)^{-1} \boldsymbol{\epsilon}_{k,i}^{(O)} \right]} \varpi_{k-1}^{(\ell)} \quad (40)$$

where  $\mathcal{C}_{k,i}^{(O)}$  is the autocorrelation matrix associated with the optimal filter using  $\mathbf{p}^{\text{true}}$ . Then, taking the natural logarithm of both sides gives

$$\ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right) = \frac{1}{2} \ln \left( \frac{\left| \mathcal{C}_{k,i}^{(O)} \right|}{\left| \mathcal{C}_{k,i}^{(\ell)} \right|} \right) - \frac{1}{2} \text{tr} \left[ \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] + \frac{1}{2} \text{tr} \left[ \boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} \left( \mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \quad (41)$$

Next, extending this equation by going backwards in time though the recursive law in (39) yields

$$2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left[ \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right] - \text{tr} \left[ \frac{1}{k} \sum_{j=1}^k \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] + \text{tr} \left[ \frac{1}{k} \sum_{j=1}^k \boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} \left( \mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \quad (42)$$

Rearranging (42) yields the following form:

$$2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right) = \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) - \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr}(I) - \text{tr} \left[ \frac{1}{k} \sum_{j=1}^k (\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)}) \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] \quad (43)$$

where  $I$  is the identity matrix of the size of  $\mathcal{C}_{k,i}^{(\ell)}$ . The first three terms in (39) are non-positive values based on (37). The ergodicity principle gives  $\frac{1}{k} \sum_{j=1}^k \boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} \rightarrow \mathcal{C}_{k,i}^{(O)}$  as  $k \rightarrow \infty$ . Let us rewrite (43) in shorthanded form as

$$2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = -\alpha^{(\ell)} - \beta^{(\ell)} \quad (44)$$

where

$$\begin{aligned} \alpha^{(\ell)} &= -\ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) + \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] - \text{tr}(I) \\ \beta^{(\ell)} &= \text{tr} \left[ \frac{1}{k} \sum_{j=1}^k (\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)}) \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] \end{aligned} \quad (45)$$

The terms  $\alpha^{(\ell)}$  are non-negative if  $\mathbf{p}^{(\ell)} \neq \mathbf{p}^{\text{true}}$ . This leads to

$$\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} = e^{-\frac{1}{2}(\alpha^{(\ell)} + \beta^{(\ell)})k} \quad (46)$$

Note,  $\alpha^{(O)}$  and  $\beta^{(O)}$  are strictly zero for the optimal case as  $\frac{1}{k} \sum_{j=1}^k \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \rightarrow \mathcal{C}_{k,i}^{(\ell)} \rightarrow \mathcal{C}_{k,i}^{(O)}$ , which leads to  $\frac{\varpi_k^{(\ell)}}{\varpi_0^{(\ell)}} \rightarrow 1$  as  $k \rightarrow \infty$ .

It is worth investigating a special case of the GMAE approach where the correlation between residuals is not incorporated into the likelihood. In the uncorrelated case the autocorrelation matrix is a block

diagonal matrix and  $C_{k,i}^{(\ell)} = 0$  for  $i \neq 0$ . For this special case,  $\alpha$  and  $\beta$  can be reduced to

$$\begin{aligned}\alpha_U^{(\ell)} &= -\ln \left( \frac{|C_{k,i}^{(O)}|}{|C_{k,i}^{(\ell)}|_U} \right) + \text{tr} [C_{k,i}^{(O)} C_{k,i}^{(\ell)-1}] - \text{tr}(I) \\ &= \sum_{s=k-i}^k \left[ -\ln \left( \frac{|C_{s,0}^{(O)}|}{|C_{s,0}^{(\ell)}|} \right) + \text{tr} [C_{s,0}^{(O)} C_{s,0}^{(\ell)-1}] - \text{tr}(I) \right] \\ &\equiv (i+1) \left[ -\ln \left( \frac{|C_{k,0}^{(O)}|}{|C_{k,0}^{(\ell)}|} \right) + \text{tr} [C_{k,0}^{(O)} C_{k,0}^{(\ell)-1}] - \text{tr}(I) \right] = (i+1)\alpha_0^{(\ell)}\end{aligned}\quad (47)$$

$$\begin{aligned}\beta_U^{(\ell)} &= \text{tr} \left[ \frac{1}{k} \sum_{j=1}^k \sum_{s=j-i}^j (\mathbf{e}_s^{(\ell)} \mathbf{e}_s^{(\ell)T} - C_{s,0}^{(O)}) (C_{k,0}^{(\ell)})^{-1} \right] \\ &\equiv (i+1) \text{tr} \left[ \frac{1}{k} \sum_{j=1}^k (\mathbf{e}_j^{(\ell)} \mathbf{e}_j^{(\ell)T} - C_{j,0}^{(O)}) (C_{j,0}^{(\ell)})^{-1} \right] = (i+1)\beta_0^{(\ell)}\end{aligned}\quad (48)$$

and the convergence rate will be

$$\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} = e^{-\frac{1}{2}(\alpha_0^{(\ell)} + \beta_0^{(\ell)})(i+1)k} \quad (49)$$

Setting  $i = 0$  compares this result with the convergence rate of the traditional MMAE approach:

$$\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} = e^{-\frac{1}{2}(\alpha_0^{(\ell)} + \beta_0^{(\ell)})k} \quad (50)$$

This shows that even without considering the autocorrelation in GMMAE approach, it converges faster than the traditional MMAE.

Incorporating the autocorrelation in the GMMAE makes the convergence to the optimal parameter faster. This shown by comparing the convergence of GMMAE to the uncorrelated GMMAE as follows:

$$\left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_G / \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_U = \frac{\left( |C_{k,i}|^{-1/2} \exp \left[ -\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} (C_{k,i})^{-1} \boldsymbol{\epsilon}_{k,i}^{(\ell)} \right] \right)_G}{\left( |C_{k,i}|^{-1/2} \exp \left[ -\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} (C_{k,i})^{-1} \boldsymbol{\epsilon}_{k,i}^{(\ell)} \right] \right)_U} \quad (51)$$

where the subscripts  $U$  and  $G$  refer to the uncorrelated GMMAE and the standard form of the GMMAE, respectively. Taking the natural logarithm of both sides in (51) gives

$$\begin{aligned}2 \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_G - 2 \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_U &= -\ln \frac{|C_{k,i}^{(\ell)}|_G}{|C_{k,i}^{(\ell)}|_U} - \text{tr} \left[ \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} (C_{k,i})^{-1} \right]_G \\ &\quad + \text{tr} \left[ \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} (C_{k,i})^{-1} \right]_U\end{aligned}\quad (52)$$

If  $\mathbf{p}^{(\ell)}$  is the true parameter  $\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \rightarrow (\mathcal{C}_{k,i})_G$ , where the off-diagonal terms are the predicted estimates  $\bar{C}_{k,i}$ , and thus

$$2 \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_G - 2 \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_U = - \ln \frac{|\mathcal{C}_{k,i}|_G}{|\mathcal{C}_{k,i}|_U} - \text{tr} I + \text{tr} \left[ (\mathcal{C}_{k,i})_G (\mathcal{C}_{k,i})_U^{-1} \right] \quad (53)$$

The right side of (53) yields a positive quantity which means the weight  $\varpi_k^{(\ell)}$  is higher in the GMMAE if  $\mathbf{p}^{(\ell)}$  is true parameter. This shows that the GMMAE approach converges faster toward the optimal estimate than the uncorrelated GMMAE. The autocorrelation helps the generalized adaptive law to makeup for the decline in the convergence rate in the previous steps due to the residuals randomness. However, if  $\mathbf{p}^{(\ell)}$  is not the true parameter then the weight convergence is given from (43):

$$2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] + \text{tr}(I) - \text{tr} \left[ \frac{1}{k} \sum_{j=1}^k (\boldsymbol{\epsilon}_{k,i}^{(j)} \boldsymbol{\epsilon}_{k,i}^{(j)T} - \mathcal{C}_{k,i}^{(O)}) (\mathcal{C}_{k,i}^{(\ell)G^{-1}}) \right] \quad (54)$$

Let  $[\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]^{(\ell)} = \frac{1}{k} \sum_{j=1}^k \boldsymbol{\epsilon}_{k,i}^{(j)} \boldsymbol{\epsilon}_{k,i}^{(j)T}$  and rewrite (54) as

$$2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] + \text{tr}(I) - \text{tr} \left[ ([\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] \quad (55)$$

Let  $[\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]^{(\ell)} = [\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_d^{(\ell)} + [\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_f^{(\ell)}$  where  $[\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_d^{(\ell)}$  contains the diagonal elements of matrix  $[\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]^{(\ell)}$  and  $[\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_f^{(\ell)}$  contains the off-diagonal elements, so that

$$2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] + \text{tr}(I) - \text{tr} \left[ ([\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_d^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] - \text{tr} \left[ ([\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_f^{(\ell)}) \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] \quad (56)$$

As stated earlier the estimate of the off-diagonal elements  $E \{ \mathbf{e}_k \mathbf{e}_{k-i}^T \} = \bar{C}_{k,i}$  is given in (29), assuming  $k$  is large enough [7]. Then, since the estimate  $\bar{C}_{k,i} = -\hat{C}_{k,i}$  as given in (36), we have

$$[\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_f^{(\ell)} = -\mathcal{C}_{k,i}^{(\ell)G} + \mathcal{C}_{k,i}^{(\ell)U} \quad (57)$$

Substituting (57) into (56) gives

$$2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] + \text{tr}(I) - \text{tr} \left[ ([\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_d^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] + \text{tr} \left[ (\hat{\mathcal{C}}_{k,i}^{(\ell)} - \hat{\mathcal{C}}_{k,i}^{(\ell)U}) \mathcal{C}_{k,i}^{(\ell)G^{-1}} \right] \quad (58)$$

Now, (58), after algebraic rearrangement, becomes

$$\begin{aligned}
2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) &= \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) - \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i_G}^{(\ell)-1} \right] + \text{tr}(I) \\
&\quad - \text{tr} \left[ ([\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i_G}^{(\ell)-1} \right] \\
&\quad - \ln \left( \frac{|\mathcal{C}_{k,i}^{(\ell)}|_G}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) + \text{tr}(I) - \text{tr} \left[ \mathcal{C}_{k,i_U}^{(\ell)} \mathcal{C}_{k,i_G}^{(\ell)-1} \right]
\end{aligned} \tag{59}$$

which reduces down to

$$\begin{aligned}
2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) &= \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) + \text{tr}(I) - \text{tr} \left[ ([\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)}) \mathcal{C}_{k,i_G}^{(\ell)-1} \right] \\
&\quad - \ln \left( \frac{|\mathcal{C}_{k,i}^{(\ell)}|_G}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) + \text{tr}(I) - \text{tr} \left[ \mathcal{C}_{k,i_U}^{(\ell)} \mathcal{C}_{k,i_G}^{(\ell)-1} \right]
\end{aligned} \tag{60}$$

By adding the following terms  $-\text{tr} \left[ [\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} (\mathcal{C}_{k,i_U}^{(\ell)-1} - \mathcal{C}_{k,i_U}^{(\ell)-1}) \right] = 0$  and  $-\text{tr} \left[ (\mathcal{C}_{k,i}^{(O)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i_U}^{(\ell)-1} \right] = 0$  to (60) and rearranging, we obtain

$$\begin{aligned}
2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) &= \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) - \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i_U}^{(\ell)-1} \right] + \text{tr}(I) \\
&\quad - \text{tr} \left[ ([\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i_U}^{(\ell)-1} \right] \\
&\quad - \text{tr} \left[ [\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} (\mathcal{C}_{k,i_G}^{(\ell)-1} - \mathcal{C}_{k,i_U}^{(\ell)-1}) \right] \\
&\quad + \ln \left( \frac{|\mathcal{C}_{k,i}^{(\ell)}|_U}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[ \mathcal{C}_{k,i_U}^{(\ell)} \mathcal{C}_{k,i_G}^{(\ell)-1} \right] + \text{tr}(I)
\end{aligned} \tag{61}$$

Writing (61) in shorthand form gives

$$2k^{-1} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = -[\alpha_U + \beta_U + \delta_{GU} + \alpha_{GU}] \tag{62}$$

with

$$\alpha_{GU} = -\ln \frac{|\mathcal{C}_{k,i}^{(O)}|_U}{|\mathcal{C}_{k,i}^{(O)}|_G} - \text{tr} I + \text{tr} \left[ (\mathcal{C}_{k,i}^{(O)})_U (\mathcal{C}_{k,i}^{(O)})_G^{-1} \right] \tag{63a}$$

$$\delta_{GU} = \text{tr} \left[ [\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} (\mathcal{C}_{k,i_G}^{(\ell)-1} - \mathcal{C}_{k,i_U}^{(\ell)-1}) \right] \tag{63b}$$

where  $\alpha_{GU}$  has non-negative value and  $\delta_{GU}$  is a trace of a product of matrix differences. Since the term  $[\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)}$  is a diagonal matrix, the trace carries the diagonal elements of the second term only. At convergence the autocorrelation matrix  $\mathcal{C}_{k,i_G}^{(\ell)}$  is positive definite and nearly a diagonal symmetric matrix

where  $(C^{-1})_{i,i} \geq (C_{i,i})^{-1}$  [16]. This shows  $\delta_{GU}$  is a positive quantity. Then, the convergence rate is given as

$$\begin{aligned} \frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} &= e^{-\frac{1}{2}[\alpha_U^{(\ell)} + \beta_U^{(\ell)} + \delta_{GU} + \alpha_{GU}]k} \\ &= e^{-\frac{1}{2}[(\alpha_0^{(\ell)} + \beta_0^{(\ell)})(i+1) + \delta_{GU} + \alpha_{GU}]k} \end{aligned} \quad (64)$$

Incorporating the autocorrelation in the GMMAE approach shows faster convergence than the uncorrelated GMMAE and even faster than the traditional MMAE approach. This proves the convergence properties of the GMMAE approach using the Kalman filter as an elemental filter where a linear model is used with stationary measurement and/or process noise.

Similarly, the convergence rate for the linear time-varying model can be obtained, by recalling (41), as

$$\begin{aligned} \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right) &= \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|^{1/2}}{|\mathcal{C}_{k,i}^{(\ell)}|^{1/2}} \right) - \frac{1}{2} \text{tr} \left[ \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] \\ &\quad + \frac{1}{2} \text{tr} \left[ \boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} \left( \mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \end{aligned} \quad (65)$$

Rearranging (65) yields

$$\begin{aligned} 2 \ln \left( \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right) &= \ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) - \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr}(I) \\ &\quad - \text{tr} \left[ \left( \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)} \right) \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] \\ &\quad + \text{tr} \left[ \left( \boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} - \mathcal{C}_{k,i}^{(O)} \right) \left( \mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \end{aligned} \quad (66)$$

which can be written in short hand as

$$2 \ln \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} = -\alpha_k^{(\ell)} - \beta_k^{(\ell)} \quad (67)$$

where

$$\begin{aligned} \alpha_k^{(\ell)} &= -\ln \left( \frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) + \text{tr} \left[ \mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] - \text{tr}(I) \\ \beta_k^{(\ell)} &= -\text{tr} \left[ \left( \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)} \right) \left( \mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] + \text{tr} \left[ \left( \boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} - \mathcal{C}_{k,i}^{(O)} \right) \left( \mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \end{aligned} \quad (68)$$

Then, to obtain  $\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}}$ , let

$$\alpha^{(\ell)} = \frac{1}{k} \sum_{j=1}^k \alpha_j^{(\ell)} \quad , \quad \beta^{(\ell)} = \frac{1}{k} \sum_{j=1}^k \beta_j^{(\ell)} \quad (69)$$

Then, the overall convergence rate is given as

$$\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} = e^{-\frac{1}{2}(\alpha^{(\ell)} + \beta^{(\ell)})k} \quad (70)$$

The  $\alpha$  term is always non-negative from (37). The non-negativeness of  $\beta$  still holds for linear time-variant systems even if the  $\beta_k$  terms are from different times. This is because the residual error  $(\epsilon_{k,i}^{(\ell)} \epsilon_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)})$  is normalized by  $(\mathcal{C}_{k,i}^{(\ell)})^{-1}$ , which follows the normalized Kalman innovations of the Kalman filter. The same approach is valid for the nonlinear case where the first-order Taylor series approximates the system functions. This is limited when the linearized  $\Phi_k(\hat{\mathbf{x}}_k)$  and  $H_k(\hat{\mathbf{x}}_k)$  are bounded and provide close approximations to the real system. The extended Kalman filter (EKF) is derived with this assumption and the approach is assumed valid as long as the EKF is valid for the problem. Simulations using the EKF indicate that the non-negativeness property of  $\beta$  term is true however. Figure 3 shows the value of  $e^{-\frac{1}{2}\beta^{(\ell)}}$  for different  $\mathbf{p}^{(\ell)}$ . The simulation is performed for the system in next section except  $\mathbf{p}^{(\ell)}$  is linearly distributed equally in the range  $[0, 100]$  where the true parameter  $\mathbf{p}^{\text{true}} = 10$ . This shows that the GMMAE approach provides superior performance and that the correlated terms do play a role in the convergence.

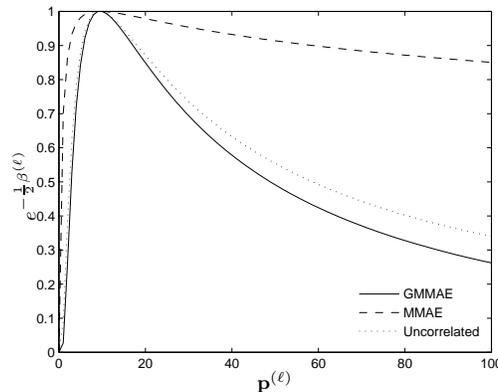


Fig. 3. Convergence of Linearly Distributed  $\mathbf{p}^{(\ell)}$

## V. NUMERICAL SIMULATIONS

Two practical examples are presented in this section to compare the performance of the MMAE and GMMAE approaches for noise covariance adaptation.

### A. Target Tracking

The state vector is chosen as the position and velocity in  $x$  and  $y$  directions, given by  $\mathbf{x} = [x \ \dot{x} \ y \ \dot{y}]^T$ . The target motion model is given by a linear dynamical system:

$$\mathbf{x}_{k+1} = \Phi \mathbf{x}_k + \mathbf{w}_k \quad (71)$$

with

$$\Phi = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (72)$$

The sampling period  $\Delta t = 0.01$ . The  $4 \times 4$  process noise covariance matrix is parameterized by  $q_x$  and  $q_y$ , given by

$$Q_k = \begin{bmatrix} q_x \begin{bmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t \end{bmatrix} & 0_{2 \times 2} \\ 0_{2 \times 2} & q_y \begin{bmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t \end{bmatrix} \end{bmatrix} \quad (73)$$

The true values of  $q_x$  and  $q_y$  are chosen as  $q_x = q_y = 10$ . In the MMAE and GMMAE approaches, the elements of  $q_x$  and  $q_y$  for the individual Kalman filters are generated using a two-dimensional Hammersley sequence under the assumption that  $q_x$  and  $q_y$  are independently uniformly distributed in  $[0, 100]$ . The number of elements in both the MMAE and the GMMAE approach is 250. The measurement model is given by

$$\tilde{\mathbf{y}}_k = \begin{bmatrix} \sqrt{x^2 + y^2} \\ \arctan(y/x) \end{bmatrix} + \mathbf{v}_k \quad (74)$$

It is assumed that the noise terms in range and azimuth measurements are uncorrelated, so the measurement noise covariance matrix is diagonal, given by  $R_k = \text{diag}[r_\rho \ r_A]$ . We choose  $r_\rho = 0.01$  and  $r_A = 0.000001$  in the simulations. Since the measurement model is nonlinear in the state vector, EKFs are used in the MMAE and GMMAE approaches with the sensitivity matrix  $H_k$  evaluated at the current estimate. The EKFs are provided with good initial estimates of the state vector, so they do not suffer any divergence problems. An alternative approach to working with the nonlinear measurement model is to first convert the original range and azimuth measurements to the effective position measurements on  $x$  and  $y$  and then apply the linear Kalman measurement update. In the latter approach, the covariance matrix for the effective measurement does not take as simple a form as the shown  $R_k$  and becomes data-dependent. The resulting residual sequence is not stationary or ergodic in either case. There are many possibilities for the chosen distribution of the process noise covariance parameters. A simple approach is to assume a uniform distribution simulated by pseudo-random numbers. We instead choose a Hammersley quasi-random sequence [17] due to its well distributed pattern. A comparison between the uniform distribution and the Hammersley quasi-random sequence for 500 elements is shown in Figure 4. The Hammersley quasi-random sequence provides a better ‘‘spread’’ of elements than the uniform distribution.

The two adaptive estimators are run 50 seconds to process the same measurement data. As discussed in the previous section, the GMMAE approach goes back  $i$  time steps in order to form the residual. We choose  $i = 4$  for the GMMAE approach. The size of the corresponding residual  $\epsilon_{k,i}^{(\ell)}$  is 10 by 1. The results are given in Figures 5(a) and 5(b). It can be seen that both estimators converge within 50 seconds, but the MMAE approach takes more than twice as much time as the GMMAE approach to converge. Both estimators converge with a relative small set of elements in the above-mentioned example. A closer examination of the automatically-generated element set shows that of the 250 elements there is a element that is close to the true values of  $q_x$  and  $q_y$ . If all of the 250 elements are far away from the true values, then the MMAE and GMMAE approaches with 250 *static* elements may fail to yield satisfactory performance, since the estimates of  $q_x$  and  $q_y$  are given by the average of the elements. Increasing the element size is always a solution, but as the dimensionality of the problem increases, the computational complexity involved may quickly become prohibitive. Meanwhile, it should be noted that

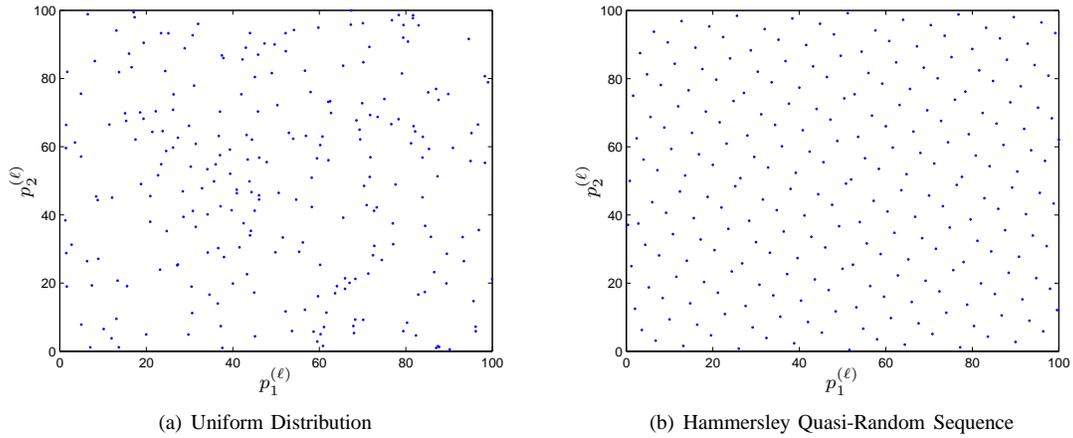


Fig. 4. Uniform Distribution and Hammersley Quasi-Random Sequence Comparison

most of the elements have almost zero weights at the end of the simulation and their contributions to the final estimates of  $q_x$  and  $q_y$  are negligible.

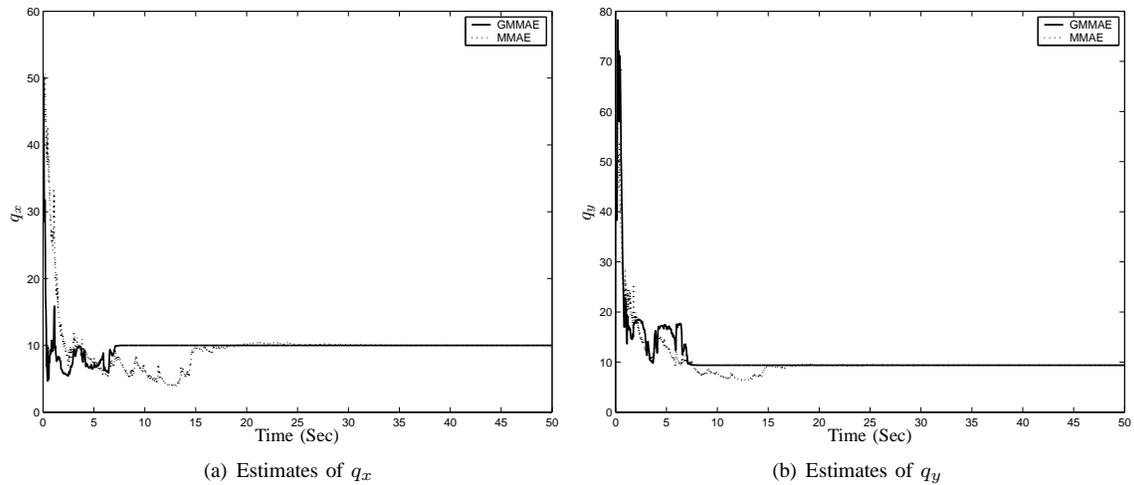


Fig. 5. Simulation of MMAE and GMMAE Estimates

### B. Single-Axis Attitude Estimation

The single axis attitude estimation problem is concerned with using angle-attitude measurements and rate information from gyros to estimate the attitude error and gyro drift. Two error sources are generally present in gyros. The first is a short-term component of instability, referred to as random drift, and the second is a random walk. These noise terms affect the Kalman filter estimates if they are not well represented in the model. The MMAE and GMMAE are used to estimate the statistical properties of these

noise terms. The attitude rate  $\dot{\theta}$  is assumed to be related to the gyro output  $\tilde{\omega}$  by

$$\dot{\theta} = \tilde{\omega} - \beta - \eta_v \quad (75)$$

where  $\beta$  is the gyro drift rate and  $\eta_v$  is a zero-mean Gaussian white-noise process with spectral density given by  $\sigma_v^2$ . The drift rate is modeled by a random walk process, given by

$$\dot{\beta} = \eta_u \quad (76)$$

where  $\eta_u$  is a zero-mean Gaussian white-noise process with spectral density given by  $\sigma_u^2$ . The parameters  $\sigma_v^2$  and  $\sigma_u^2$  are estimated by the GMMAE and MMAE approaches. The linear model used in the Kalman filter follows the attitude angle measurement model"

$$\tilde{y}_k = \theta_k + v_k \quad (77)$$

where  $v_k$  a zero-mean Gaussian white-noise process with variance given by  $R = \sigma_n^2$ . The discrete-time system used in the Kalman filter can now be written as

$$\mathbf{x}_{k+1} = \Phi \mathbf{x}_k + \Gamma \tilde{\omega} + \mathbf{w}_k \quad (78)$$

$$\tilde{y}_k = H \mathbf{x}_k + v_k \quad (79)$$

where  $\mathbf{x}_k = [\theta, \beta]_k^T$ ,  $\Gamma = [\Delta t, 0]$ ,  $H = [1, 0]$ ,

$$\Phi = \begin{bmatrix} 1 & -\Delta t \\ 0 & 1 \end{bmatrix}$$

and

$$Q = E\{\mathbf{w}_k \mathbf{w}_k^T\} = \begin{bmatrix} \sigma_v^2 \Delta t + \frac{1}{3} \sigma_u^2 \Delta t^3 & -\frac{1}{2} \sigma_u^2 \Delta t^2 \\ -\frac{1}{2} \sigma_u^2 \Delta t^2 & \sigma_u^2 \Delta t \end{bmatrix}$$

where  $\Delta t$  is the constant sampling interval. Synthetic measurements are created using a true constant angle-rate given by  $\dot{\theta} = 0.0001$  rad/sec and a sampling period of 1 second. The noise parameters are given by  $\sigma_n = 17 \times 10^{-6}$  rad,  $\sigma_u = 1 \times 10^{-10}$  rad/sec<sup>3/2</sup> and  $\sigma_v = 1 \times 10^{-7}$  rad/sec<sup>1/2</sup>. The initial bias  $\beta_0$  is given as 0.1 deg/hr. A set of only eleven filters run in parallel with the initial covariance matrix set to  $P_0 = \text{diag}[1 \times 10^{-4}, 1 \times 10^{-12}]$  for each. The only difference in the filters is the predetermined values for  $\sigma_u$  and  $\sigma_v$ , which are randomly selected in a neighborhood of the true values. As in the target tracking example, both MMAE and GMMAE are used to estimate the unknown parameters  $\sigma_u$  and  $\sigma_v$ . Figure 6 shows the performance of both approaches in estimating the attitude error and the gyro drift rate. Figures 7 and 8 show the estimates using the MMAE and GMMAE approaches, respectively, for 10 Monte Carlo runs. The results show that the GMMAE approach, using only four steps back, outperforms the MMAE approach by converging in about one-third of the time and with much smaller standard deviation between runs. A study on choosing different steps back in the GMMAE approach has also been done. Figure 9 shows the convergence rate for the MMAE approach and GMMAE approach with  $i = 2, 4, 6$ . Note that even though the theoretical derivations show that the convergence rate increases with the number of steps back taken in the GMMAE likelihood, a practical limit exists for this number. This is due to the limited number of samples chosen in the parameter space and results can vary from application to application.

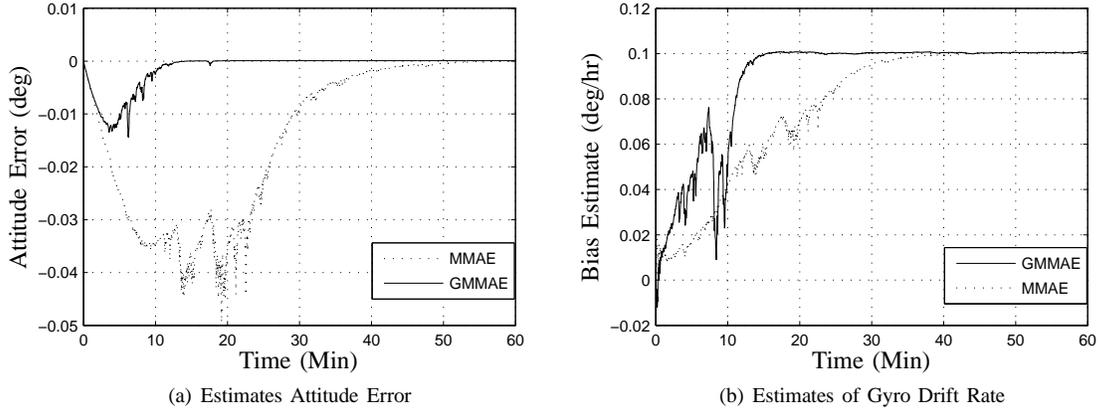


Fig. 6. Simulation of MMAE vs. GMAE Gyro State Estimates

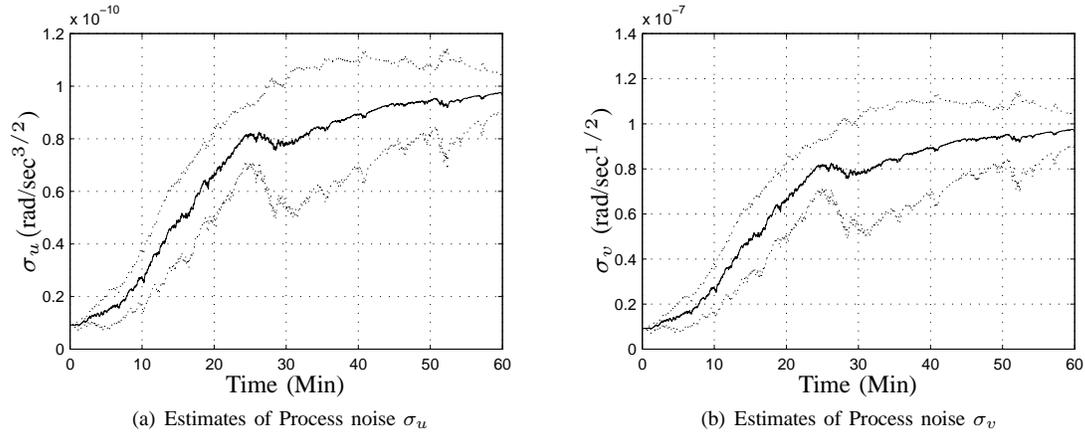


Fig. 7. Simulation of MMAE Gyro Estimates

## VI. CONCLUSION

In this paper a generalized approach for multiple-model adaptive estimation was shown, which can be used for time-varying systems. The main advantage of this approach is that it provides faster convergence properties than the traditional multiple-model adaptive estimation approach. The new approach is based on using the autocorrelation of the measurement-minus-estimate residual. A formulation was developed for the Kalman filter, however computation of the off-diagonal elements of the autocorrelation matrix is intractable in practice since these matrix elements are a function of the true parameters. This difficulty was overcome by using the estimated values in the Kalman gain. It is proven that the generalized approach provides better convergence properties than the standard approach because the autocorrelation incorporated into generalized approach's likelihood raises the sensitivity toward the parameter of an optimal filter. Simulation results indicated that the new multiple-model adaptive estimation approach can provide better convergence to the best estimate over the standard approach. Similar results have been achieved when the approach is applied to the tracking problem using range and angle measurements, which involves the use

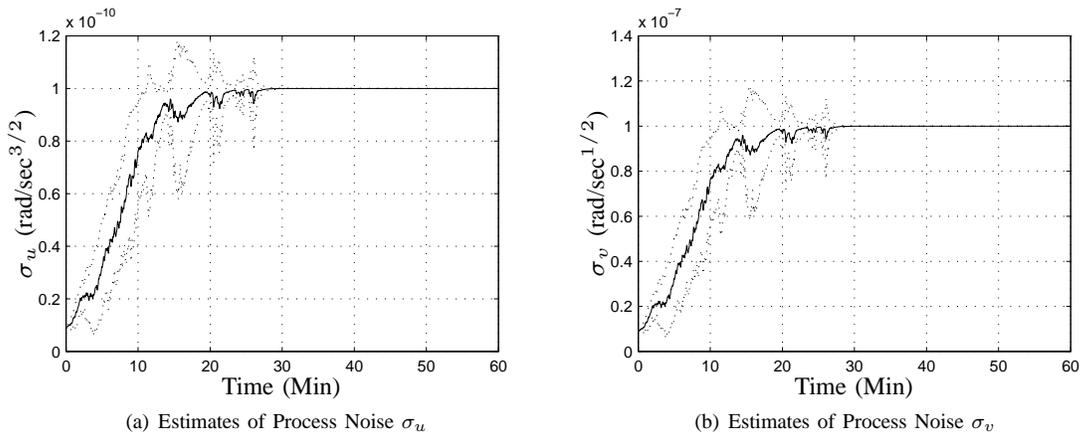


Fig. 8. Simulation of GMMAE Gyro Estimates

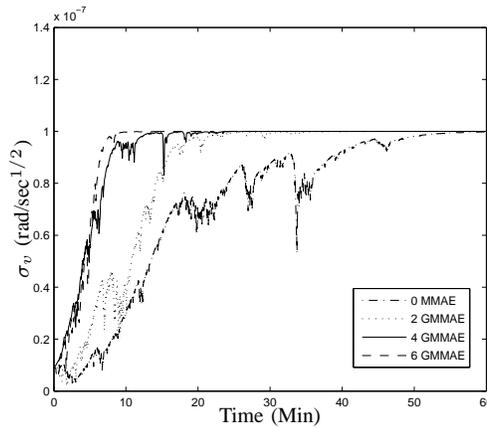


Fig. 9. Simulation of MMAE and Varying Steps Back GMMAE of Process Noise  $\sigma_v$  Estimates

of the extended Kalman filter for state estimation. For nonlinear model and/or measurement processes, it is assumed that the linearized output sufficiently captures the statistics of the likelihood function by making the small noise assumption, and thus the time-varying solution provided in this paper is applicable for this case.

#### REFERENCES

- [1] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. part v. multiple-model methods," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1255–1321, Oct. 2005.
- [2] D. T. Magill, "Optimal adaptive estimation of sampled stochastic processes," *IEEE Transactions on Automatic Control*, vol. 10, no. 4, pp. 434–439, Oct. 1965.
- [3] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 3, no. 8, pp. 780–783, Aug. 1988.
- [4] X. R. Li and Y. Bar-Shalom, "Multiple-model estimation with variable structure," *IEEE Transactions on Automatic Control*, vol. 41, no. 4, pp. 478–493, April 1996.

- [5] J. R. Vasquez and P. S. Maybeck, "Enhanced motion and sizing of bank in moving-bank mmae," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 770–779, July 2004.
- [6] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings, "A new autocovariance least-squares method for estimating noise covariances," Texas-Wisconsin Modeling and Control Consortium, Tech. Rep. 2003-04, Sept. 2003.
- [7] R. K. Mehra, "On the identification of variances and adaptive kalman filtering," *IEEE Transactions on Automatic Control*, vol. 15, no. 2, pp. 175–184, April 1970.
- [8] P. D. Hanlon and P. S. Maybeck, "Multiple-model adaptive estimation using a residual correlation kalman filter bank," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-36, no. 2, pp. 393–406, April 2000.
- [9] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, 3rd ed. New York, NY: John Wiley & Sons, 1997, pp. 353–361.
- [10] R. F. Stengel, *Optimal Control and Estimation*. New York, NY: Dover Publications, 1994, pp. 402–407.
- [11] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Mineola, NY: Dover Publications, 2005, ch. 10.1.
- [12] B. D. O. Anderson, J. B. Moore, and R. M. Hawkes, "Model approximations via prediction error identification," *Automatica*, vol. 14, no. 6, pp. 615–622, Nov. 1978.
- [13] J. L. Crassidis and Y. Cheng, "Generalized multiple-model adaptive estimation using an autocorrelation approach," in *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006, paper 223.
- [14] P. D. Hanlon, "Practical implementation of multiple model adaptive estimation using neyman-pearson based hypothesis testing and spectral estimation tools," Ph.D. dissertation, School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH, September 1996.
- [15] B. Alsuwaidan, "Generalized multiple-model adaptive estimation," Ph.D. dissertation, University at Buffalo, The State University of New York, Amherst, NY, 2008.
- [16] P. Sprent, "Inversion of nearly diagonal matrices," *The Mathematical Gazette*, pp. 184–189, May 1965.
- [17] J. M. Hammersley, "Monte carlo methods for solving multivariable problems," *Annals of the New York Academy of Sciences*, vol. 86, pp. 844–874, 1960.