

Convergence Properties of Autocorrelation-Based Generalized Multiple-Model Adaptive Estimation

Badr N. Alsuwaidan*

Space Research Institute, KACST, Riyadh, Saudi Arabia

John L. Crassidis[†] and Yang Cheng[‡]

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

In this paper a generalized multiple-model adaptive estimator is presented that can be used to estimate the unknown noise statistics in filter designs. The assumed unknowns in the adaptive estimator are the process noise covariance elements. Parameter elements generated from a quasi-random sequence are used to drive multiple-model 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. For nonlinear Gaussian measurement processes, it is assumed that the linearized output sufficiently captures the statistics of the likelihood function by making the small noise assumption. 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 using an extended Kalman filter, indicate that the new approach provides better convergence properties over a traditional multiple-model approach.

I. Introduction

Modern-day control systems rely heavily on filters (estimators) to provide full-state feedback and to also filter sensor noise. With the advent of micro-electro-mechanical systems (MEMS)-type sensors, state estimation becomes even more crucial in control systems, because these sensors are generally not as accurate as standard sensors. However, they offer several advantages over larger sensors, including the ability to be mass-produced at low prices. MEMS-type sensors and actuators currently provide a multi-billion dollar market in a wide range of areas, such as automotive, industrial, defense, medical, mass data storage and optical switching.^a

Health monitoring and fault detection techniques have increasingly become commonplace with MEMS devices. Most modern-day applications use a model-based approach combined with filter residuals, i.e. measurement minus filter-estimate. The deviation of the residuals from zero is the combined result of noise and faults,¹ Since noise is always present in a sensor, statistical analysis is often used to detect faults. Many approaches exist that can be used for model-based fault detection. In this paper we concentrate on a common approach that incorporates multiple models in an adaptive structure using filtering algorithms.²

Filtering algorithms, such as the extended Kalman filter (EKF),³ the Unscented filter (UF)⁴ and Particle filters (PFs),^{5,6} are commonly used to both estimate unmeasurable states and filter noisy measurements. The EKF and UF assume that the process noise and measurement noise are represented by zero-mean Gaussian white-noise processes. Even if this is true, both filters only provide approximate solutions when the state

*Research Assistant Professor, Space Research Institute, KACST, Riyadh, Saudi Arabia. Email: bsuwaidan@kacst.edu.sa, member AIAA.

[†]Professor, Department of Mechanical & Aerospace Engineering. Email: johnc@buffalo.edu, Associate Fellow AIAA.

[‡]Research Assistant Professor, Department of Mechanical & Aerospace Engineering. Email: cheng3@eng.buffalo.edu. Senior Member AIAA.

^ahttp://www.smalltimes.com/display_article/281842/109/ARTCL/none/none/MEMS-the-word-in-consumer-electronics/

and/or measurement models are nonlinear, since the *a posteriori* density function is most often non-Gaussian. The EKF typically works well only in the region where the first-order Taylor-series linearization adequately approximates the non-Gaussian probability density function (pdf). The Unscented filter works on the premise that with a fixed number of parameters it should be easier to approximate a Gaussian distribution than to approximate an arbitrary nonlinear function. This in essence can provide higher-order moments for the computation of the *a posteriori* function without the need to calculate Jacobian matrices as required in the EKF. Still, the standard form of the EKF has remained the most popular method for nonlinear estimation to this day, and other designs are investigated only when the performance of this standard form is not sufficient.

Like other approximate approaches to optimal filtering, the ultimate objective of a PF is to construct the *a posteriori* pdf of the state vector, or the pdf of the state vector conditioned on all the available measurements. However, the approximation of a PF is vastly different from that of conventional nonlinear filters. The central idea of the PF approximation is to represent a continuous distribution of interest by a finite (but large) number of weighted random samples of the state vector, or particles. Particle filters do not assume the *a posteriori* distribution of the state vector to be a Gaussian distribution or any other distribution of known form. In principle, they can estimate probability distributions of arbitrary form and solve any nonlinear and/or non-Gaussian system.

Even if the process noise and/or measurement noise are Gaussian, all standard forms of the EKF, UF and PFs require knowledge of their characteristics, such as the mean and covariance for a Gaussian process. The covariance and mean of the measurement noise can be inferred from statistical inferences and calibration procedures of the hardware sensing devices. The calibration procedures can also be used to determine the nature of the measurement process distribution. The kurtosis characterizes the relative compactness of the distribution around the mean, relative to a Gaussian distribution. A common kurtosis, called the “Pearson kurtosis,” divides the fourth moment by the second moment.⁷ A greater-than-three Pearson kurtosis indicates a relatively peaked distribution, while a less-than-three Pearson kurtosis indicates a relatively flat distribution. However, the process noise is extremely difficult to characterize because it is usually used to represent modeling errors. Its covariance is usually determined by *ad hoc* or heuristic approaches, which leads to the classical “tuning of the filter” problem. Fortunately, there are tools available to aid the filter designer. For example, several tests can be applied to check the *consistency* of the filter from the desired characteristics of the measurement residuals. These include the normalized error square test, the autocorrelation test and the normalized mean error test.² These tests can, at the very least, provide mechanisms to show that a filter is not performing in an optimal or desired fashion.

In practice the tuning of a filter can be arduous and time consuming. A classic approach to overcome this difficulty is to use adaptive filters. Adaptive filtering can be divided into four general categories: Bayesian, maximum likelihood, covariance matching, and correlation approaches.⁸ Bayesian and maximum likelihood methods may be well suited to a multi-model approaches, but sometimes require large computational loads. Covariance matching is the computation of the covariances from the residuals of the state estimation problem, but have 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.”⁹ 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.⁹

Adaptive filtering for nonlinear systems has recently gained attention. Parlos et al. shows a neural net to constructively approximate the state equations.¹⁰ The proposed algorithms in their work make minimal assumptions regarding the underlying nonlinear dynamics and their noise statistics. Nonadaptive and adaptive state filtering algorithms are presented with both off-line and on-line learning stages. Good performance is shown for a number of test cases. Lho and Painter show an adaptive filter using fuzzy membership functions, where the fuzzy processing is driven by an inaccurate online estimate of signal-to-noise ratio for the signal being tracked.¹¹ Good results are shown for a simple tracking problem. Lee and Alfriend show an adaptive scheme that can be used to estimate the process noise covariance for both the UF and the first-order divided difference filter.¹² The standard update approach requires proper selection of a window size to control the level of the variance update. The innovation of their work is a procedure that

automatically calculates the window size using a derivative-free numerical optimization technique. Good results are shown for satellite orbit determination applications.

A new approach is derived in Ref. 13 for adaptive filtering based on generalizing the standard multiple-model adaptive estimation (MMAE) algorithm.¹⁴ A MMAE algorithm uses a parallel bank of filters to provide multiple estimates, where each filter corresponds with a dependence on some unknowns, which can be the process noise covariance elements if desired. 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. Standard MMAE algorithms use only the current time measurement-minus-estimate residual to test the hypothesis. The approach in Ref. 13 is a generalization of Ref. 15, 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. 15. 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 new approach, called generalized multiple-model adaptive estimation (GMMAE), is based on calculating the time-domain autocorrelation function, 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. The unknown elements in our design are the parameters of the process noise covariance. Process noise covariance elements can be drawn from any sample distribution as long as the resulting covariance matrix remains positive semi-definite. A Hammersley quasi-random sequence¹⁶ is chosen due to its well distributed pattern. The covariance elements are estimated using a weighted sum of the quasi-random elements, similar to the approach used for state estimation in PFs. An expression for the error-covariance of the estimates is also provided, which gives a bound on the process noise parameter estimates. In this paper a theoretical proof of the convergence properties of the GMMAE is shown, which offers insight as to its advantages of the standard MMAE approach.

The organization of the remainder of this paper proceeds as follows. First, the standard EKF 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 a method that can incorporate nonlinear measurement models in the adaptive approach. The theoretical background for the GMMAE approach and proof of convergence is then shown. Finally, simulation results involving a two-dimensional target tracking problem are shown.

II. Extended Kalman Filter

A summary of the continuous-discrete EKF is given in Table 1, where $\mathbf{x}(t)$ is the $n \times 1$ state vector, $\mathbf{u}(t)$ is the known control input, $G(t)$ is the process noise distribution matrix, $\mathbf{w}(t)$ is the process noise vector which is assumed to be a zero-mean Gaussian noise process with spectral density $Q(t)$, $\tilde{\mathbf{y}}_k$ 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. Oftentimes, if the sampling interval is below Nyquist's limit, a discrete-time propagation of the covariance is used:

$$P_{k+1}^- = \Phi_k P_k^+ \Phi_k^T + Q_k \quad (1)$$

where Φ_k is the discrete-time state transition matrix of $F(\hat{\mathbf{x}}(t), t)$ and Q_k is the discrete-time process noise covariance matrix. These matrices can be numerically computed for a constant sampling interval using an algorithm given by van Loan.¹⁷ First, the following $2n \times 2n$ matrix is formed:

$$\mathcal{A} = \begin{bmatrix} -F(\hat{\mathbf{x}}(t), t) & G(t) Q(t) G^T(t) \\ 0 & F^T(\hat{\mathbf{x}}(t), t) \end{bmatrix} \Delta t \quad (2)$$

Table 1. Continuous-Discrete Extended Kalman Filter

Model	$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) + G(t) \mathbf{w}(t), \mathbf{w}(t) \sim N(\mathbf{0}, Q(t))$ $\tilde{\mathbf{y}}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k, \mathbf{v}_k \sim N(\mathbf{0}, R_k)$
Initialize	$\hat{\mathbf{x}}(t_0) = \hat{\mathbf{x}}_0$ $P_0 = E \{ \tilde{\mathbf{x}}(t_0) \tilde{\mathbf{x}}^T(t_0) \}$
Gain	$K_k = P_k^- H_k^T(\hat{\mathbf{x}}_k^-) [H_k(\hat{\mathbf{x}}_k^-) P_k^- H_k^T(\hat{\mathbf{x}}_k^-) + R_k]^{-1}$ $H_k(\hat{\mathbf{x}}_k^-) \equiv \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right _{\hat{\mathbf{x}}_k^-}$
Update	$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + K_k [\tilde{\mathbf{y}}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-)]$ $P_k^+ = [I - K_k H_k(\hat{\mathbf{x}}_k^-)] P_k^-$
Propagation	$\dot{\mathbf{x}}(t) = \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), t)$ $\dot{P}(t) = F(\hat{\mathbf{x}}(t), t) P(t) + P(t) F^T(\hat{\mathbf{x}}(t), t) + G(t) Q(t) G^T(t)$ $F(\hat{\mathbf{x}}(t), t) \equiv \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right _{\hat{\mathbf{x}}(t)}$

where Δt is the constant sampling interval. Then, the matrix exponential of Eq. (2) is computed:

$$\mathcal{B} = e^{\mathcal{A}} \equiv \begin{bmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} \\ 0 & \mathcal{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathcal{B}_{11} & \Phi_k^{-1} Q_k \\ 0 & \Phi_k^T \end{bmatrix} \quad (3)$$

where \mathcal{B}_{11} is not needed in the computation of Φ_k and Q_k . The state transition matrix is then given by

$$\Phi_k = \mathcal{B}_{22}^T \quad (4)$$

Also, the discrete-time process noise covariance is given by

$$Q_k = \Phi_k \mathcal{B}_{12} \quad (5)$$

Note that Eqs. (4) and (5) is only valid for time-invariant systems and covariance matrices. However, if the sampling interval is “small” enough, then computing them using the aforementioned approach gives a good approximation to the actual matrices. Note that the first-order approximation of Q_k is given by

$$Q_k \approx \Delta t G(t) Q(t) G^T(t) \quad (6)$$

In most cases $Q(t)$ is a diagonal matrix, while Q_k contains cross-correlation terms due to sampling. These cross-correlations can be ignored if Eq. (6) is a good approximation for Q_k .

III. Multiple-Model Adaptive Estimation

In this section a review of MMAE is shown. More details can be found in Refs. 18 and 19. Multiple-model adaptive estimation is a recursive estimator that uses a bank of 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). Note that we do not necessarily 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. A set of distributed elements is generated from some known 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 ℓ^{th} element $\mathbf{p}^{(\ell)}$ given all the measurements. Application of Bayes' rule yields

$$p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_k) = \frac{p(\tilde{\mathbf{Y}}_k|\mathbf{p}^{(\ell)})p(\mathbf{p}^{(\ell)})}{\sum_{j=1}^M p(\tilde{\mathbf{Y}}_k|\mathbf{p}^{(j)})p(\mathbf{p}^{(j)})} \quad (7)$$

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 through²⁰

$$\begin{aligned} p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_k) &= \frac{p(\tilde{\mathbf{y}}_k, \mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_{k-1})}{p(\tilde{\mathbf{y}}_k|\tilde{\mathbf{Y}}_{k-1})} \\ &= \frac{p(\tilde{\mathbf{y}}_k|\hat{\mathbf{x}}_k^{-(\ell)})p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_{k-1})}{\sum_{j=1}^M \left[p(\tilde{\mathbf{Y}}_k|\hat{\mathbf{x}}_k^{-(j)})p(\mathbf{p}^{(j)}|\tilde{\mathbf{Y}}_{k-1}) \right]} \end{aligned} \quad (8)$$

since $p(\tilde{\mathbf{y}}_k|\tilde{\mathbf{Y}}_{k-1}, \mathbf{p}^{(\ell)})$ is given by $p(\tilde{\mathbf{y}}_k|\hat{\mathbf{x}}_k^{-(\ell)})$ in the Kalman recursion. Note that the denominator of Eq. (8) is just a normalizing factor to ensure that $p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_k)$ is a pdf. The recursion formula can now be cast into a set of defined weights $\varpi_k^{(\ell)}$, so that

$$\begin{aligned} \varpi_k^{(\ell)} &= \varpi_{k-1}^{(\ell)} p(\tilde{\mathbf{y}}_k|\hat{\mathbf{x}}_k^{-(\ell)}) \\ \varpi_k^{(\ell)} &\leftarrow \frac{\varpi_k^{(\ell)}}{\sum_{j=1}^M \varpi_k^{(j)}} \end{aligned} \quad (9)$$

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. 20, 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.²¹

The conditional mean estimate is the weighted sum of the parallel filter estimates:

$$\hat{\mathbf{x}}_k^- = \sum_{j=1}^M \varpi_k^{(j)} \hat{\mathbf{x}}_k^{-(j)} \quad (10)$$

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

$$P_k^- = \sum_{j=1}^M \varpi_k^{(j)} \left[\left(\hat{\mathbf{x}}_k^{-(j)} - \hat{\mathbf{x}}_k^- \right) \left(\hat{\mathbf{x}}_k^{-(j)} - \hat{\mathbf{x}}_k^- \right)^T + P_k^{-(j)} \right] \quad (11)$$

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

$$\hat{\mathbf{p}}_k = \sum_{j=1}^M \varpi_k^{(j)} \mathbf{p}^{(j)} \quad (12a)$$

$$\mathcal{P}_k = \sum_{j=1}^M \varpi_k^{(j)} \left(\mathbf{p}^{(j)} - \hat{\mathbf{p}}_k \right) \left(\mathbf{p}^{(j)} - \hat{\mathbf{p}}_k \right)^T \quad (12b)$$

Equation (12b) can be used to define 3σ bounds on the estimate $\hat{\mathbf{p}}_k$. The entire MMAE process is shown in Figure 1.

IV. 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.

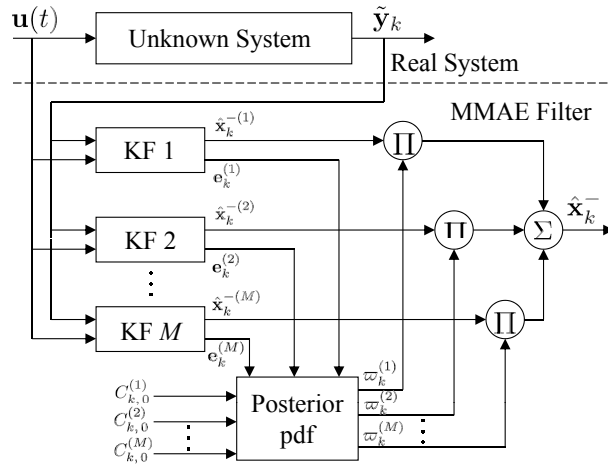


Figure 1. MMAE Process

A. Autocorrelation for Time-Varying Systems

In this section the autocorrelation matrix for time-varying systems is derived, which is an extension to the approach shown in Ref. 9. Here we assume that the model is linear with

$$\mathbf{x}_{k+1} = \Phi_k \mathbf{x}_k + \Gamma_k \mathbf{u}_k + \Upsilon_k \mathbf{w}_k \quad (13a)$$

$$\tilde{\mathbf{y}}_k = H_k \mathbf{x}_k + \mathbf{v}_k \quad (13b)$$

where Υ_k is the discrete-time process noise distribution matrix. Consider the following discrete-time residual equation:

$$\begin{aligned} \mathbf{e}_k &\equiv \tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k^- \\ &= -H_k \tilde{\mathbf{x}}_k^- + \mathbf{v}_k \end{aligned} \quad (14)$$

where $\tilde{\mathbf{x}}_k^- \equiv \hat{\mathbf{x}}_k^- - \mathbf{x}_k$. The following autocorrelation function matrix can be computed:

$$C_{k,i} = \begin{cases} H_k P_k^- H_k^T + R_k & i = 0 \\ H_k E \{ \tilde{\mathbf{x}}_k^- \tilde{\mathbf{x}}_{k-i}^{-T} \} H_{k-i}^T - H_k E \{ \tilde{\mathbf{x}}_k^- \mathbf{v}_{k-i}^T \} & i > 0 \end{cases} \quad (15)$$

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

$$\tilde{\mathbf{x}}_k^- = \Phi_{k-1} (I - K_{k-1} H_{k-1}) \tilde{\mathbf{x}}_{k-1}^- + \Phi_{k-1} K_{k-1} \mathbf{v}_{k-1} - \Upsilon_{k-1} \mathbf{w}_{k-1} \quad (16)$$

Carrying Eq. (16) i steps back leads to

$$\begin{aligned} \tilde{\mathbf{x}}_k^- &= \left[\prod_{j=1}^i \Phi_{k-j} (I - K_{k-j} H_{k-j}) \right] \tilde{\mathbf{x}}_{k-i}^- + \sum_{j=2}^i \left[\prod_{\ell=1}^{j-1} \Phi_{k-\ell} (I - K_{k-\ell} H_{k-\ell}) \right] \Phi_{k-j} K_{k-j} \mathbf{v}_{k-j} \\ &\quad - \sum_{j=2}^i \left[\prod_{\ell=1}^{j-1} \Phi_{k-\ell} (I - K_{k-\ell} H_{k-\ell}) \right] \Upsilon_{k-j} \mathbf{w}_{k-j} + \Phi_{k-1} K_{k-1} \mathbf{v}_{k-1} - \Upsilon_{k-1} \mathbf{w}_{k-1} \end{aligned} \quad (17)$$

where

$$\prod_{j=1}^i Z_{k-j} \equiv Z_{k-1} Z_{k-2} \cdots Z_{k-i} \quad (18)$$

Performing the expectations in the definition of $C_{k,i}$ leads to¹³

$$C_{k,i} = \begin{cases} H_k P_k^- H_k^T + R_k & i = 0 \\ H_k \Phi_{k-1} (P_{k-1}^- H_{k-1}^T - K_{k-1} C_{k-1,0}) & i = 1 \\ H_k \left[\prod_{j=1}^{i-1} \Phi_{k-j} (I - K_{k-j} H_{k-j}) \right] \Phi_{k-i} (P_{k-i}^- H_{k-i}^T - K_{k-i} C_{k-i,0}) & i > 1 \end{cases} \quad (19)$$

where

$$C_{k-i,0} \equiv H_{k-i} P_{k-i}^- H_{k-i}^T + R_{k-i} \quad (20)$$

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:

$$\boldsymbol{\epsilon}_{k,i} \equiv \begin{bmatrix} \mathbf{e}_k \\ \mathbf{e}_{k-1} \\ \vdots \\ \mathbf{e}_{k-i} \end{bmatrix} \quad (21)$$

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

$$L_{k,i} = \frac{1}{[\det(2\pi \mathcal{C}_{k,i})]^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^T \mathcal{C}_{k,i}^{-1} \boldsymbol{\epsilon}_{k,i}\right) \quad (22)$$

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

$$\mathcal{C}_{k,i} = \begin{bmatrix} C_{k,0} & C_{k,1} & C_{k,2} & \cdots & C_{k,i} \\ C_{k,1}^T & C_{k-1,0} & C_{k-1,1} & \cdots & C_{k-1,i-1} \\ C_{k,2}^T & C_{k-1,1}^T & C_{k-2,0} & \cdots & C_{k-2,i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{k,i}^T & C_{k-1,i-1}^T & C_{k-2,i-2}^T & \cdots & C_{k-i,0} \end{bmatrix} \quad (23)$$

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

$$L_{k,0} = \frac{1}{\{\det[2\pi (H_k P_k^- H_k^T + R_k)]\}^{1/2}} \exp\left[-\frac{1}{2} \mathbf{e}_k^T (H_k P_k^- H_k^T + R_k)^{-1} \mathbf{e}_k\right] \quad (24)$$

This likelihood is widely used in MMAE algorithms,^{14,18} but ignores correlations between different measurement times. However, it is simpler to evaluate than the general likelihood function in Eq. (22) since no storage of data or system matrices is required.

C. GMMAE Adaptive Law

In this section the new adaptive law based on the autocorrelation is shown. In the traditional MMAE approach only the current-time measurement information is used in the update law given by Eq. (9). Therefore, the update law is given by

$$\begin{aligned} \varpi_k^{(\ell)} &= \varpi_{k-1}^{(\ell)} L_{k,0}^{(\ell)} \\ \varpi_k^{(\ell)} &\leftarrow \frac{\varpi_k^{(\ell)}}{\sum_{j=1}^M \varpi_k^{(j)}} \end{aligned} \quad (25)$$

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

$$L_{k,0}^{(\ell)} = \frac{1}{\left\{ \det[2\pi (H_k P_k^{-\ell}) H_k^T + R_k] \right\}^{1/2}} \exp \left[-\frac{1}{2} \mathbf{e}_k^{(\ell)T} (H_k P_k^{-\ell}) H_k^T + R_k)^{-1} \mathbf{e}_k^{(\ell)} \right] \quad (26)$$

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

The GMMAE adaptive law is based on carrying Eq. (8) i steps back to give the new update law:

$$\begin{aligned} \varpi_k^{(\ell)} &= \varpi_{k-1}^{(\ell)} L_{k,i}^{(\ell)} \\ \varpi_k^{(\ell)} &\leftarrow \frac{\varpi_k^{(\ell)}}{\sum_{j=1}^M \varpi_k^{(j)}} \end{aligned} \quad (27)$$

with

$$L_{k,i}^{(\ell)} = \frac{1}{\left[\det \left(2\pi C_{k,i}^{(\ell)} \right) \right]^{1/2}} \exp \left[-\frac{1}{2} \boldsymbol{\epsilon}_i^{(\ell)T} \left(C_{k,i}^{(\ell)} \right)^{-1} \boldsymbol{\epsilon}_i^{(\ell)} \right] \quad (28)$$

where $\boldsymbol{\epsilon}_i^{(\ell)}$ is defined as $\boldsymbol{\epsilon}_i^{(\ell)} \equiv [\mathbf{e}_k^{(\ell)T} \mathbf{e}_{k-1}^{(\ell)T} \cdots \mathbf{e}_{k-i}^{(\ell)T}]^T$. The matrix $C_{k,i}^{(\ell)}$ is given by Eqs. (19) and (23) evaluated at the ℓ^{th} covariance and the *optimal* Kalman gain. Unfortunately, the optimal gain is a function of the actual covariance Q_k , which is not known. Specifically, if K_k from Table 1 is substituted into Eq. (19), then for $i \geq 1$ the correlated terms $C_{k,i}$ will always be zero. One way to overcome this problem is to estimate the $C_{k,i}^{(\ell)}$ terms using the residuals, which is the approach taken in Ref. 9. But, this requires a stationary process and a sufficiently large set of measurements over time, which would not work properly for time-varying system matrices and/or a sequential updating scheme. A different approach is taken here, which is also expanded for nonlinear systems. This assumes that the measurement noise is small compared to the signal so that the Gaussian nature of the measurement residuals is maintained. Estimates for $C_{k,i}^{(\ell)}$ are given by

$$\hat{C}_{k,i}^{(\ell)} = \begin{cases} H_k(\hat{\mathbf{x}}_k^{-(\ell)}) P_k^{-\ell} H_k^T(\hat{\mathbf{x}}_k^{-(\ell)}) + R_k & i = 0 \\ H_k(\hat{\mathbf{x}}_k^{-(\ell)}) \Phi_{k-1}(\hat{\mathbf{x}}_{k-1}^{-(\ell)}) \left[P_{k-1}^{-\ell} H_{k-1}^T(\hat{\mathbf{x}}_{k-1}^{-(\ell)}) - \hat{K}_{k-1} C_{k-1,0}^{(\ell)} \right] & i = 1 \\ H_k(\hat{\mathbf{x}}_k^{-(\ell)}) \left\{ \prod_{j=1}^{i-1} \Phi_{k-j}(\hat{\mathbf{x}}_{k-j}^{-(\ell)}) \left[I - \hat{K}_{k-j} H_{k-j}(\hat{\mathbf{x}}_{k-j}^{-(\ell)}) \right] \right\} \Phi_{k-i}(\hat{\mathbf{x}}_{k-i}^{-(\ell)}) \\ \times \left[P_{k-i}^{-\ell} H_{k-i}^T(\hat{\mathbf{x}}_{k-1}^{-(\ell)}) - \hat{K}_{k-i} C_{k-i,0}^{(\ell)} \right] & i > 1 \end{cases} \quad (29)$$

where

$$C_{k-i,0}^{(\ell)} \equiv H_{k-i}(\hat{\mathbf{x}}_{k-i}^{-(\ell)}) P_{k-i}^{-\ell} H_{k-i}^T(\hat{\mathbf{x}}_{k-i}^{-(\ell)}) + R_{k-i} \quad (30)$$

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

$$P_{k+1}^{-\ell} = \Phi_k(\hat{\mathbf{x}}_k^{-(\ell)}) P_k^{+(\ell)} \Phi_k^T(\hat{\mathbf{x}}_k^{-(\ell)}) + Q^{(\ell)} \quad (31a)$$

$$P_k^{+(\ell)} = \left[I - K_k^{(\ell)} H_k(\hat{\mathbf{x}}_k^{-(\ell)}) \right] P_k^{-\ell} \quad (31b)$$

$$K_k^{(\ell)} = P_k^{-\ell} H_k^T \left[H_k(\hat{\mathbf{x}}_k^{-(\ell)}) P_k^{-\ell} H_k^T(\hat{\mathbf{x}}_k^{-(\ell)}) + R_k \right]^{-1} \quad (31c)$$

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

$$\hat{K}_k = \hat{P}_k^- H_k^T(\hat{\mathbf{x}}_k^-) \left[H_k(\hat{\mathbf{x}}_k^-) \hat{P}_k^- H_k^T(\hat{\mathbf{x}}_k^-) + R_k \right]^{-1} \quad (32)$$

with

$$\hat{P}_{k+1}^- = \Phi_k(\hat{\mathbf{x}}_k^-) \hat{P}_k^+ \Phi_k^T(\hat{\mathbf{x}}_k^-) + \hat{Q}_k \quad (33a)$$

$$\hat{P}_k^+ = \left[I - \hat{K}_k H_k(\hat{\mathbf{x}}_k^-) \right] \hat{P}_k^- \quad (33b)$$

where \hat{Q}_k is computed using $\hat{\mathbf{p}}_k$.

Using the current measurement, $\tilde{\mathbf{y}}_k$, along with the ℓ^{th} element, $\mathbf{p}^{(\ell)}$, $1 \leq \ell \leq M$, a bank of filters are executed. For each filter the state estimates, $\hat{\mathbf{x}}_k^{-(\ell)}$, and measurements are used to form the residual, $\boldsymbol{\epsilon}_{k,i}^{(\ell)}$, going back i steps. The filter error covariance, $P_k^{-(\ell)}$, and state matrices, $\Phi_k^{-(\ell)}$ and $H_k^{-(\ell)}$, evaluated at the current estimates 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 significantly 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 Eq. (12a).

V. GMMAE Theoretical Concept

The theoretical aspect of Kalman filter for linear systems is very sound, derived from a rigorous analysis. In practice ‘‘tuning’’ of Kalman filter can be arduous and very time consuming. The approach presented here provides a recursive method to estimate the parameter of system. Convergence of adaptive approaches is an important issue. For fault detection methods using MMAE approaches, it is imperative that the fault be detected as quickly as possible. The MMAE approach is applicable to the stationary noise process, but can be extended to non-stationary cases as long as the convergence of the adaptive algorithm is faster than the change in the covariance matrix to be identified. In this section a proof of convergence of the GMMAE approach is shown, as well as its relationship to the standard MMAE approach.

The goal of the GMMAE approach is to determine the value of the unknown parameter \mathbf{p}_k from a Bayesian prospective, where the best point estimate of $\hat{\mathbf{p}}_k$ is the determined quantity of the *a posteriori* distribution of \mathbf{p}_k . Thus, the *a posteriori* distribution of \mathbf{p}_k plays a central role in the estimation process. Both the traditional MMAE and GMMAE approaches involve discretization of the parameter space of \mathbf{p} into set of a fixed points, $\mathbf{p}^{(\ell)}$, $\ell = 1, \dots, M$. Their difference is mainly in the way the *a posteriori* distribution of \mathbf{p} and the weights $\varpi^{(\ell)}$ associated with \mathbf{p} are updated:

$$\varpi_k^{(\ell)} \propto p(\mathbf{p}^{(\ell)} | \tilde{\mathbf{Y}}_{k,i}) \propto p(\tilde{\mathbf{Y}}_{k,i} | \mathbf{p}^{(\ell)}) p(\mathbf{p}^{(\ell)}) \quad (34)$$

The joint likelihood $p(\tilde{\mathbf{Y}}_{k,i} | \mathbf{p}^{(\ell)})$ is central to GMMAE estimation process. The idea comes from the fact that $\mathbf{p}^{(\ell)}$ is just a parameter used to determine the Kalman gain $K^{(\ell)}$ and thus the residual $\boldsymbol{\epsilon}_{k,i}^{(\ell)}$, while the true parameter \mathbf{p}^{true} (i.e. the process noise Q) is present in the actual system and governs the true dynamics. First, the following pdf is defined:

$$p(\tilde{\mathbf{Y}}_{k,i} | \mathbf{p}^{(\ell)}) \equiv p(\boldsymbol{\epsilon}_{k,i}^{(\ell)} | \mathbf{p}^{\text{true}}) \quad (35)$$

Since the true parameter \mathbf{p}^{true} is unknown, it is approximated by $\hat{\mathbf{p}}$, the current estimate of \mathbf{p}^{true} . Such a approximation is common in estimation techniques, such as the EKF, when the needed value is unavailable. Because $\hat{\mathbf{p}}$ may not be identical to any $\mathbf{p}^{(\ell)}$, the residuals $\boldsymbol{\epsilon}_k^{(\ell)}$, which are still zero mean but may be correlated in time. The autocorrelation can be estimated using Eq. (29). The GMMAE approach takes into account the autocorrelation in the joint likelihood function. The joint likelihood function is a Gaussian distribution as in the MMAE case. However, the GMMAE approach uses a set of the last i^{th} residuals to compute the likelihood in a moving average manner. The concept is that for large k , a higher likelihood $p(\boldsymbol{\epsilon}_k^{(\ell)} | \mathbf{p}^{\text{true}})$ yields a higher weight $\varpi^{(\ell)}$, which corresponds to $\mathbf{p}^{(\ell)}$ providing the best estimate of \mathbf{p}^{true} . Since $\hat{\mathbf{p}}$ may not be initially an accurate estimate of \mathbf{p}^{true} , accordingly the autocorrelation has to be recalculated every time as a new estimate $\hat{\mathbf{p}}$ is provided. Eventually, the autocorrelation will be computed using $\hat{\mathbf{p}}$ that is very close to \mathbf{p}^{true} as the GMMAE algorithm converges.

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 that maintains the necessary optimal null hypothesis of the GMMAE likelihood. The assumptions that have been taken into consideration in the evaluation of the correlation of the residuals are as follows: 1) the filter that produces the residuals that is assumed to be an optimal filter uses an optimal Kalman gain, in the strict sense, and the GMMAE estimate of the optimal gain is equivalent to the optimal gain that reflects the real plant measurement; 2) the ergodic average of the correlation of the residuals is zero at the limit when k goes to infinity; 3) the residuals are random variables even for the optimal case as they are a result of random noise inputs (specifically measurement and process noise); 4) the actual correlation of the residuals is also random and fluctuates around zero mean and any estimate of the correlation of the residuals at a given time may encounter an instantaneous offset from the zero mean; and, 5) the optimal filter is predicted to compensate for this offset in future time steps.

The residuals of the optimal filter 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

$$\bar{C}_{k-1,i} = \frac{1}{N} \sum_{j=i}^N \mathbf{e}_j \mathbf{e}_{j-i}^T \quad (36)$$

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 $\{\tilde{\mathbf{x}}_{k-1}^- \tilde{\mathbf{x}}_{k-1}^{-T}\}$ which is the approach introduced in Ref. 9. 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 Eq. (19) using the autocorrelation ergodic estimate $\bar{C}_{k-1,i}$:

$$[\bar{C}_{k,i}^{(\ell)T} \dots \bar{C}_{k-i+1,1}^{(\ell)T}]^T = A_{k,i} [\widehat{P_{k-i}^{-(\ell)} H_{k-i}^T} - K_{k-i}^{(\ell)} C_{k-i,0}^{(\ell)}] \quad (37)$$

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

$$A_{k,i} = \begin{bmatrix} H_k(\tilde{\mathbf{x}}_k^{-(\ell)}) \left\{ \prod_{j=1}^{i-1} \Phi_{k-j}(\tilde{\mathbf{x}}_k^{-(\ell)}) \left[I - \hat{K}_{k-j} H_{k-j}(\tilde{\mathbf{x}}_{k-i}^{-(\ell)}) \right] \right\} \Phi_{k-i}(\tilde{\mathbf{x}}_{k-i}^{-(\ell)}) \\ \vdots \\ H_{k-i+1}(\tilde{\mathbf{x}}_{k-i+1}^{-(\ell)}) \Phi_{k-i}(\tilde{\mathbf{x}}_{k-i}^{-(\ell)}) \end{bmatrix} \quad (38)$$

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

$$[\bar{C}_{k,i}^{(\ell)T} \dots \bar{C}_{k-i+1,1}^{(\ell)T}]^T \cong A_{k,i} [\hat{K}_{k-i} - K_{k-i}^{(\ell)}] C_{k-i,0}^{(\ell)} \quad (39)$$

$$\hat{K}_{k-i} \cong K_{k-i}^{(\ell)} + A_{k,i}^\dagger [\bar{C}_{k,i}^{(\ell)T} \dots \bar{C}_{k-i+1,1}^{(\ell)T}]^T C_{k-i,0}^{(\ell)-1} \quad (40)$$

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 Eq. (29) is determined by the weight ϖ_{k-1} , which is a measure of the likelihood of the residual $\mathbf{e}_{k-1} = [\tilde{\mathbf{y}}_{k-1} - H_k \tilde{\mathbf{x}}_k^-]$ and the previous residuals. The estimate $\tilde{\mathbf{x}}_{k-1}^-$ is the Kalman filter propagation of $\tilde{\mathbf{x}}_{k-2}^+$ updated by \hat{K}_{k-2} . Therefore, the weight ϖ_{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 ϖ_{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 $\tilde{\mathbf{y}}_{k-1}$

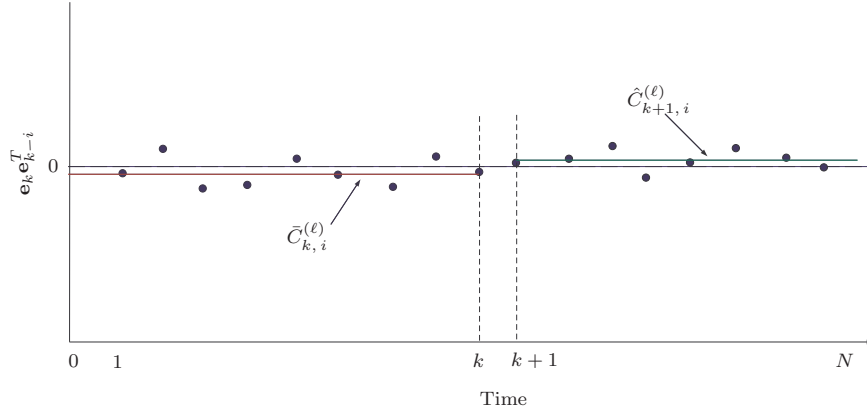


Figure 2. Correlation of the Residuals of an Optimal Kalman Filter

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 Eq. (29) is actually a propagation of \hat{K}_{k-2} determined by ϖ_{k-1} based on the likelihood evolution of the residuals performance of the gain estimate \hat{K}_{k-2} as shown in Eq. (32). Any estimate based on the weight ϖ_{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 Eq. (39) by

$$\bar{C}_{k-1,1}^{(\ell)} = A_{k-1,1} [\hat{K}_{k-2} - K_{k-2}^{(\ell)}] C_{k-2,0}^{(\ell)} \quad (41)$$

The GMMAE estimate of the autocorrelation $\hat{C}_{k,i}^{(\ell)}$ in Eq. (29), that is used in the GMMAE likelihood, is based on the gain estimate of \hat{K}_{k-1} :

$$[\hat{C}_{k,i}^{(\ell)T} \dots \hat{C}_{k-i+1,1}^{(\ell)T}]^T = A_{k,i} [P_{k-i}^{-(\ell)} H_{k-i}^T - K_{k-i}^{(\ell)} C_{k-i,0}^{(\ell)}] \quad (42)$$

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 Eq. (41) into Eq. (42), where $\hat{K}_{k-1} = \hat{K}_{k-2}$ at steady state, leads to

$$\begin{aligned} \hat{C}_{k,1}^{(\ell)} &= A_{k,1} [P_{k-1}^{-(\ell)} H_{k-1}^T - (K_{k-2}^{(\ell)} + A_{k-1,1}^\dagger \bar{C}_{k-1,1}^{(\ell)}) C_{k-2,0}^{(\ell)-1} C_{k-1,0}^{(\ell)}] \\ \hat{C}_{k,1}^{(\ell)} &= -A_{k,1} A_{k-1,1}^\dagger [\bar{C}_{k-1,1}^{(\ell)} C_{k-2,0}^{(\ell)-1}] C_{k-1,0}^{(\ell)} \\ \hat{C}_{k,1}^{(\ell)} &= -\bar{C}_{k-1,1}^{(\ell)} \end{aligned} \quad (43)$$

The predicted estimate $\hat{C}_{k,1}^{(\ell)}$ has the opposite sign of the ergodic estimate.²² 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 Eq. (43) 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, $p(\mathbf{p}^{\text{true}}|\tilde{\mathbf{Y}}_{k,i}) \rightarrow 1$ as $k \rightarrow \infty$ and $p(\mathbf{p}^{(\ell)}|\tilde{\mathbf{Y}}_{k,i}) \rightarrow 0$ as $k \rightarrow \infty$ for $\mathbf{p}^{(\ell)} \neq \mathbf{p}^{\text{true}}$. This result will be proved with assumption of ergodicity of residuals and the use of the following lemma of standard matrix theory.²¹ Let A and B be two $n \times n$ positive definite matrices, then

$$n + \ln \left(\frac{|A|}{|B|} \right) - \text{tr}[B^{-1}A] \leq 0 \quad (44)$$

where 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:

$$\varpi_k^{(\ell)} \leftarrow \frac{L_{k,i}^{(\ell)} \varpi_{k-1}^{(\ell)}}{\sum_{j=1}^M L_{k,i}^{(j)} \varpi_{k-1}^{(j)}} \quad (45)$$

or more specifically

$$\varpi_k^{(\ell)} = \frac{L_{k,i}^{(\ell)} \varpi_{k-1}^{(\ell)}}{L_{k,i}^{\text{true}}} \quad (46)$$

where $L_{k,i}^{\text{true}}$ is the likelihood evaluated using \mathbf{p}^{true} . Substituting for $L_{k,i}$ from Eq. (28) into Eq. (46) gives

$$\varpi_k^{(\ell)} = \frac{|\mathcal{C}_{k,i}^{(\ell)}|^{-1/2} \exp \left[-\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left(\mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \boldsymbol{\epsilon}_{k,i}^{(\ell)} \right]}{|\mathcal{C}_{k,i}^{(O)}|^{-1/2} \exp \left[-\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(O)T} \left(\mathcal{C}_{k,i}^{(O)} \right)^{-1} \boldsymbol{\epsilon}_{k,i}^{(O)} \right]} \varpi_{k-1}^{(\ell)} \quad (47)$$

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

$$\ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right) = \frac{1}{2} \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) - \frac{1}{2} \text{tr} \left[\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left(\mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] + \frac{1}{2} \text{tr} \left[\boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} \left(\mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \quad (48)$$

Next, extending this equation by going backwards in time though the recursive law in Eq. (46) yields

$$2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left[\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right] - \text{tr} \left[\frac{1}{k} \sum_{j=1}^k \boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left(\mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] + \text{tr} \left[\frac{1}{k} \sum_{j=1}^k \boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} \left(\mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \quad (49)$$

Rearranging Eq. (49) yields the following form:

$$2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right) = \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) - \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr}(I) - \text{tr} \left[\frac{1}{k} \sum_{j=1}^k (\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)}) \left(\mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] \quad (50)$$

where I is the identity matrix of the size of $\mathcal{C}_{k,i}^{(\ell)}$. The first three terms in Eq. (46) are non-positive values based on Eq. (44). 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 Eq. (50) in shorthanded form as

$$2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = -\alpha^{(\ell)} - \beta^{(\ell)} \quad (51)$$

where

$$\begin{aligned}\alpha^{(\ell)} &= -\ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) + \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] - \text{tr}(I) \\ \beta^{(\ell)} &= \text{tr} \left[\frac{1}{k} \sum_{j=1}^k (\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)}) \left(\mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right]\end{aligned}\quad (52)$$

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

$$\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} = e^{-\frac{1}{2}(\alpha^{(\ell)} + \beta^{(\ell)})k} \quad (53)$$

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

$$\frac{\varpi_k^{(\ell)}}{\varpi_0^{(\ell)}} \rightarrow 1 \text{ as } k \rightarrow \infty.$$

Now, it is worth to look at a special case of the GMMAE 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 $\mathcal{C}_{k,i}^{(\ell)} = 0$ for $i \neq 0$. For this special case, α and β can be reduced to

$$\begin{aligned}\alpha_U^{(\ell)} &= -\ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) + \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] - \text{tr}(I) \\ &= \sum_{s=k-i}^k \left[-\ln \left(\frac{|\mathcal{C}_{s,0}^{(O)}|}{|\mathcal{C}_{s,0}^{(\ell)}|} \right) + \text{tr} \left[\mathcal{C}_{s,0}^{(O)} \mathcal{C}_{s,0}^{(\ell)-1} \right] - \text{tr}(I) \right]\end{aligned}\quad (54)$$

$$\equiv (i+1) \left[-\ln \left(\frac{|\mathcal{C}_{k,0}^{(O)}|}{|\mathcal{C}_{k,0}^{(\ell)}|} \right) + \text{tr} \left[\mathcal{C}_{k,0}^{(O)} \mathcal{C}_{k,0}^{(\ell)-1} \right] - \text{tr}(I) \right] = (i+1)\alpha_0^{(\ell)}$$

$$\beta_U^{(\ell)} = \text{tr} \left[\frac{1}{k} \sum_{j=1}^k \sum_{s=j-i}^j (\mathbf{e}_s^{(\ell)} \mathbf{e}_s^{(\ell)T} - \mathcal{C}_{s,0}^{(O)}) \left(\mathcal{C}_{k,0}^{(\ell)} \right)^{-1} \right] \quad (55)$$

$$\equiv (i+1) \text{tr} \left[\frac{1}{k} \sum_{j=1}^k (\mathbf{e}_j^{(\ell)} \mathbf{e}_j^{(\ell)T} - \mathcal{C}_{j,0}^{(O)}) \left(\mathcal{C}_{j,0}^{(\ell)} \right)^{-1} \right] = (i+1)\beta_0^{(\ell)}$$

and the convergence rate will be

$$\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} = e^{-\frac{1}{2}(\alpha_0^{(\ell)} + \beta_0^{(\ell)})(i+1)k} \quad (56)$$

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

$$\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} = e^{-\frac{1}{2}(\alpha_0^{(\ell)} + \beta_0^{(\ell)})k} \quad (57)$$

Clearly, the GMMAE converges faster than MMAE even without autocorrelation.

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:

$$\left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_G \bigg/ \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_U = \frac{\left(|\mathcal{C}_{k,i}|^{-1/2} \exp \left[-\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} (\mathcal{C}_{k,i})^{-1} \boldsymbol{\epsilon}_{k,i}^{(\ell)} \right] \right)_G}{\left(|\mathcal{C}_{k,i}|^{-1/2} \exp \left[-\frac{1}{2} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} (\mathcal{C}_{k,i})^{-1} \boldsymbol{\epsilon}_{k,i}^{(\ell)} \right] \right)_U} \quad (58)$$

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 Eq. (58) gives

$$2 \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_G - 2 \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_U = - \ln \frac{|\mathcal{C}_{k,i}^{(\ell)}|_G}{|\mathcal{C}_{k,i}^{(\ell)}|_U} - \text{tr} \left[\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} (\mathcal{C}_{k,i})^{-1} \right]_G + \text{tr} \left[\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} (\mathcal{C}_{k,i})^{-1} \right]_U \quad (59)$$

If $\mathbf{p}^{(\ell)}$ is the optimal 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 estimate $\bar{\mathcal{C}}_{k,i}$, then

$$2 \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_G - 2 \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right)_U = - \ln \frac{|\mathcal{C}_{k,i}|_G}{|\mathcal{C}_{k,i}|_U} - \text{tr} I + \text{tr} \left[(\mathcal{C}_{k,i})_G (\mathcal{C}_{k,i})_U^{-1} \right] \quad (60)$$

The right side of Eq. (60) yields a positive quantity which means the weight $\varpi_k^{(\ell)}$ is higher in the GMMAE if $\mathbf{p}^{(\ell)}$ is optimal. This shows that the GMMAE approach converges faster to the optimal estimate than the uncorrelated GMMAE. The autocorrelation helps the generalized adaptive law to make up for the decline in the convergence rate in the previous steps due to the residuals randomness. However, if $\mathbf{p}^{(\ell)}$ is not the optimal parameter then the weight convergence is given from Eq. (50):

$$2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr}(I) - \text{tr} \left[\frac{1}{k} \sum_{j=1}^k (\boldsymbol{\epsilon}_{k,i}^{(j)} \boldsymbol{\epsilon}_{k,i}^{(j)T} - \mathcal{C}_{k,i}^{(O)}) (\mathcal{C}_{k,i}^{(\ell)})^{-1} \right] \quad (61)$$

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 Eq. (61) as

$$2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr}(I) - \text{tr} \left[([\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i}^{(\ell)-1} \right] \quad (62)$$

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

$$2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr}(I) - \text{tr} \left[([\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_d^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i}^{(\ell)-1} \right] - \text{tr} \left[([\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_f^{(\ell)}) \mathcal{C}_{k,i}^{(\ell)-1} \right] \quad (63)$$

As stated earlier the estimate of the off-diagonal elements $E \{ \mathbf{e}_k \mathbf{e}_{k-i}^T \} = \bar{\mathcal{C}}_{k,i}$ is given in Eq. (36), assuming k is large enough.⁹ Then, since the estimate $\bar{\mathcal{C}}_{k,i} = -\hat{\mathcal{C}}_{k,i}$ as given in Eq. (43), we have

$$[\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_f^{(\ell)} = -\mathcal{C}_{k,i}^{(\ell)} + \mathcal{C}_{k,i}^{(\ell)} \quad (64)$$

Substituting Eq. (64) into Eq. (63) gives

$$2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr}(I) - \text{tr} \left[([\bar{\boldsymbol{\epsilon}}_i \bar{\boldsymbol{\epsilon}}_i^T]_d^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr} \left[(\hat{\mathcal{C}}_{k,i}^{(\ell)} - \hat{\mathcal{C}}_{k,i}^{(\ell)}) \mathcal{C}_{k,i}^{(\ell)-1} \right] \quad (65)$$

Now, Eq. (65), after algebraic rearrangement, becomes

$$\begin{aligned}
2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) &= \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) - \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i_G}^{(\ell)-1} \right] + \text{tr}(I) \\
&\quad - \text{tr} \left[([\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i_G}^{(\ell)-1} \right] \\
&\quad - \ln \left(\frac{|\mathcal{C}_{k,i}^{(\ell)}|_G}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) + \text{tr}(I) - \text{tr} \left[\mathcal{C}_{k,i_U}^{(\ell)} \mathcal{C}_{k,i_G}^{(\ell)-1} \right]
\end{aligned} \tag{66}$$

which reduces down to

$$\begin{aligned}
2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) &= \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) + \text{tr}(I) - \text{tr} \left[([\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)}) \mathcal{C}_{k,i_G}^{(\ell)-1} \right] \\
&\quad - \ln \left(\frac{|\mathcal{C}_{k,i}^{(\ell)}|_G}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) + \text{tr}(I) - \text{tr} \left[\mathcal{C}_{k,i_U}^{(\ell)} \mathcal{C}_{k,i_G}^{(\ell)-1} \right]
\end{aligned} \tag{67}$$

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 Eq. (67) and rearranging, we obtain

$$\begin{aligned}
2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) &= \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|_U} \right) - \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i_U}^{(\ell)-1} \right] + \text{tr}(I) \\
&\quad - \text{tr} \left[([\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} - \mathcal{C}_{k,i}^{(O)}) \mathcal{C}_{k,i_U}^{(\ell)-1} \right] \\
&\quad - \text{tr} \left[[\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} (\mathcal{C}_{k,i_G}^{(\ell)-1} - \mathcal{C}_{k,i_U}^{(\ell)-1}) \right] \\
&\quad + \ln \left(\frac{|\mathcal{C}_{k,i}^{(\ell)}|_U}{|\mathcal{C}_{k,i}^{(\ell)}|_G} \right) - \text{tr} \left[\mathcal{C}_{k,i_U}^{(\ell)} \mathcal{C}_{k,i_G}^{(\ell)-1} \right] + \text{tr}(I)
\end{aligned} \tag{68}$$

Writing Eq. (68) in shorthand form gives

$$2k^{-1} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} \right) = -[\alpha_U + \beta_U + \delta_{GU} + \alpha_{GU}] \tag{69}$$

with

$$\alpha_{GU} = -\ln \frac{|\mathcal{C}_{k,i}|_U}{|\mathcal{C}_{k,i}|_G} - \text{tr} I + \text{tr} \left[(\mathcal{C}_{k,i})_U (\mathcal{C}_{k,i})_G^{-1} \right] \tag{70a}$$

$$\delta_{GU} = \text{tr} \left[[\bar{\epsilon}_i \bar{\epsilon}_i^T]_d^{(\ell)} (\mathcal{C}_{k,i_G}^{(\ell)-1} - \mathcal{C}_{k,i_U}^{(\ell)-1}) \right] \tag{70b}$$

where α_{GU} has non-negative value and δ_{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. The autocorrelation matrix $\mathcal{C}_{k,i_G}^{(\ell)}$ is positive definite and nearly a diagonal symmetric matrix where $(C^{-1})_{i,i} \geq (C_{i,i})^{-1}$.²³ This shows δ_{GU} is a positive quantity. Then, the convergence rate is given as

$$\begin{aligned}
\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} &= e^{-\frac{1}{2}[\alpha_U^{(\ell)} + \beta_U^{(\ell)} + \delta_{GU} + \alpha_{GU}]k} \\
&= e^{-\frac{1}{2}[(\alpha_0^{(\ell)} + \beta_0^{(\ell)})(i+1) + \delta_{GU} + \alpha_{GU}]k}
\end{aligned} \tag{71}$$

Clearly, the GMMAE approach converges faster 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 linearized or linear time-varying model can be obtained, by recalling Eq. (48), as

$$\begin{aligned} \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right) &= \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|^{1/2}}{|\mathcal{C}_{k,i}^{(\ell)}|^{1/2}} \right) - \frac{1}{2} \text{tr} \left[\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} \left(\mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] \\ &\quad + \frac{1}{2} \text{tr} \left[\boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} \left(\mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \end{aligned} \quad (72)$$

Rearranging Eq. (72) yields

$$\begin{aligned} 2 \ln \left(\frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} \right) &= \ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) - \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] + \text{tr}(I) \\ &\quad - \text{tr} \left[\left(\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)} \right) \left(\mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] \\ &\quad + \text{tr} \left[\left(\boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} - \mathcal{C}_{k,i}^{(O)} \right) \left(\mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \end{aligned} \quad (73)$$

which can be written in short hand as

$$2 \ln \frac{\varpi_k^{(\ell)}}{\varpi_{k-1}^{(\ell)}} = -\alpha_k^{(\ell)} - \beta_k^{(\ell)} \quad (74)$$

where

$$\begin{aligned} \alpha_k^{(\ell)} &= -\ln \left(\frac{|\mathcal{C}_{k,i}^{(O)}|}{|\mathcal{C}_{k,i}^{(\ell)}|} \right) + \text{tr} \left[\mathcal{C}_{k,i}^{(O)} \mathcal{C}_{k,i}^{(\ell)-1} \right] - \text{tr}(I) \\ \beta_k^{(\ell)} &= -\text{tr} \left[\left(\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)} \right) \left(\mathcal{C}_{k,i}^{(\ell)} \right)^{-1} \right] + \text{tr} \left[\left(\boldsymbol{\epsilon}_{k,i}^{(O)} \boldsymbol{\epsilon}_{k,i}^{(O)T} - \mathcal{C}_{k,i}^{(O)} \right) \left(\mathcal{C}_{k,i}^{(O)} \right)^{-1} \right] \end{aligned} \quad (75)$$

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

$$\alpha^{(\ell)} = \frac{1}{k} \sum_{j=1}^k \alpha_j^{(\ell)} \quad , \quad \beta^{(\ell)} = \frac{1}{k} \sum_{j=1}^k \beta_j^{(\ell)} \quad (76)$$

Then, the overall convergence rate is given as

$$\frac{\varpi_k^{(\ell)}}{\varpi_1^{(\ell)}} = e^{-\frac{1}{2}(\alpha^{(\ell)} + \beta^{(\ell)})k} \quad (77)$$

The α term is always non-negative from Eq. (44). The non-negativeness of β still holds for linear time-variant systems even if the β_k terms are from different times. This is because the residual error $(\boldsymbol{\epsilon}_{k,i}^{(\ell)} \boldsymbol{\epsilon}_{k,i}^{(\ell)T} - \mathcal{C}_{k,i}^{(O)})$ 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 $\Phi_k(\hat{\mathbf{x}}_k)$ and $H_k(\hat{\mathbf{x}}_k)$ are bounded and provide close approximation to the real system where $\Phi_k(\hat{\mathbf{x}})\hat{\mathbf{x}} \approx \mathbf{f}(\mathbf{x}_k)$ and $H_k(\hat{\mathbf{x}}_k)\hat{\mathbf{x}}_k \approx \mathbf{h}(\mathbf{x})$. The 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 β 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 $\mathbf{p}^{(true)} = 10$. This shows that the GMMAE approach provides superior performance and that the correlated terms do play a role in the convergence.

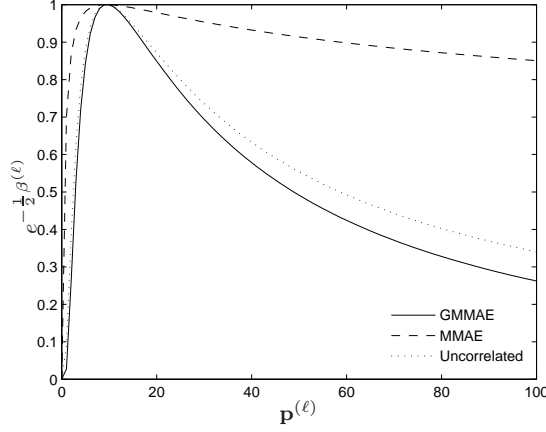


Figure 3. Convergence of Linearly Distrusted $\mathbf{p}^{(\ell)}$

VI. Numerical Simulations

In this section a numerical simulation result is given. A two-dimensional target tracking problem is used to compare the performance of the MMAE and the GMAE approaches. 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:

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

with

$$\Phi = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (79)$$

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

$$Q_k = \begin{bmatrix} q_x \begin{bmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t \end{bmatrix} & 0_{2 \times 2} \\ 0_{2 \times 2} & q_y \begin{bmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t \end{bmatrix} \end{bmatrix} \quad (80)$$

The true values of q_x and q_y are chosen as $q_x = q_y = 10$. In the MMAE and GMAE 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 the MMAE or the GMAE approach is 250. The measurement model is given by

$$\tilde{\mathbf{y}}_k = \begin{bmatrix} \sqrt{x^2 + y^2} \\ \arctan(y/x) \end{bmatrix} + \mathbf{v}_k \quad (81)$$

It is assumed that the noise in range and azimuth measurements is 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 GMAE 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

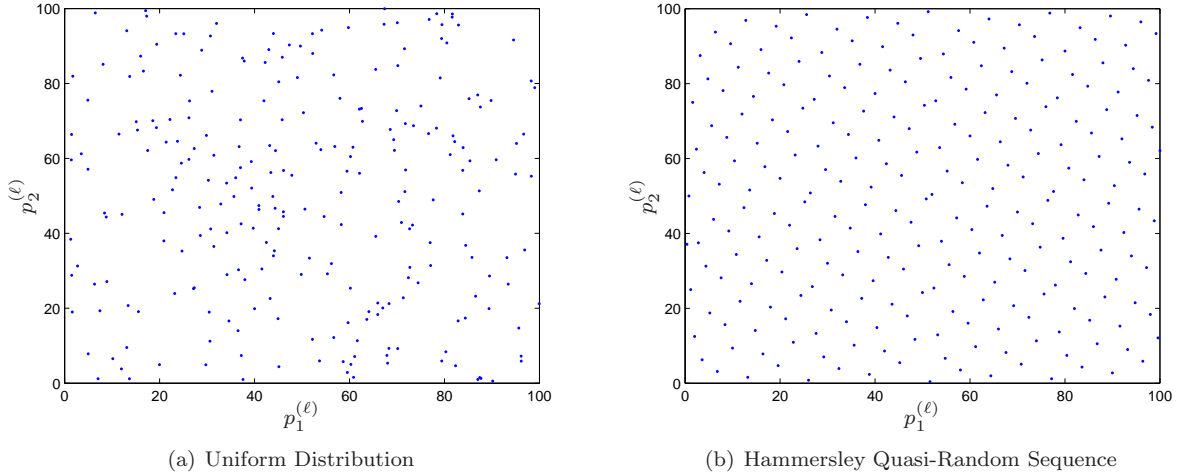


Figure 4. Uniform Distribution and Hammersley Quasi-Random Sequence Comparison

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. We instead choose a Hammersley quasi-random sequence¹⁶ 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. Clearly, the Hammersley quasi-random sequence provides a better “spread” of elements than the uniform distribution. In low dimensions, the multidimensional Hammersley sequence quickly “fills up” the space in a well-distributed pattern. However, for very high dimensions, the initial elements of the Hammersley sequence can be very poorly distributed. Only when the number of sequence elements is large enough relative to the spatial dimension is the sequence properly behaved. This isn’t much of a concern for the process noise covariance adaption problem since the dimension of the elements will be much larger than the dimension of the unknown process noise parameters. Remedies for this problem are given in Ref. 24 if needed.

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 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. Resampling and Markov-Chain Monte Carlo or regularization techniques used in particle filtering can be used in order to prune elements with very small weights and dynamically generate new elements with high weights in the neighborhood of existing elements, while maintaining a constant number of elements during the estimation process. However, since a long data sequence is usually required (and sometimes maneuvers are also required) to estimate the process noise covariance and to discriminate between “good” elements and “bad” elements, straightforward application of resampling and Markov-Chain Monte Carlo or regularization based on just a few consecutive data points does not yield a satisfactory result. It is found that “good” elements could be pruned in the early stage and “promising” elements could end up with small weights. Further investigation of the problem is under way.

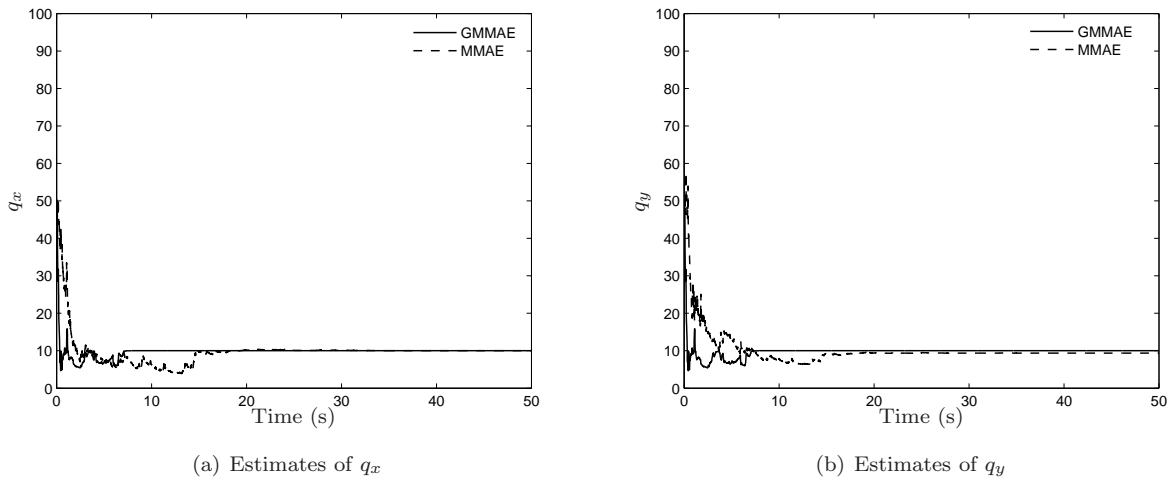


Figure 5. Simulation of MMAE and GMMAE Estimates

VII. Conclusion

In this paper a generalized approach for multiple-model adaptive estimation was shown, which can be used for time-varying and nonlinear systems. This approach is based on using the autocorrelation of the measurement-minus-estimate residual. A formulation was developed for the extended 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 the generalized approach's likelihood raises the sensitivity toward optimality. Simulation results indicated that the new multiple-model adaptive estimation approach can provide better convergence to the best estimate over the standard approach.

References

- ¹Gertler, J. J., "Survey of Model-Based Failure Detection and Isolation in Complex Plants," *IEEE Control Systems Magazine*, Vol. 8, No. 9, Dec. 1988, pp. 3–11.
- ²Bar-Shalom, Y., Li, X. R., and Kirubarajan, T., *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, New York, NY, 2001, pp. 232–245.
- ³Kalman, R. E. and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, March 1961, pp. 95–108.
- ⁴Wan, E. and van der Merwe, R., "The Unscented Kalman Filter," *Kalman Filtering and Neural Networks*, edited by S. Haykin, chap. 7, John Wiley & Sons, New York, NY, 2001.
- ⁵Gordon, N. J., Salmond, D. J., and Smith, A. F. M., "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEE Proceedings, Part F - Communications, Radar and Signal Processing*, Vol. 140, No. 2, Seattle, WA, April 1993, pp. 107–113.
- ⁶Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T., "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, Feb. 2002, pp. 174–185.
- ⁷Pearson, K., "Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder," *Biometrika*, Vol. 4, 1905, pp. 169–212.
- ⁸Odelson, B. J., Rajamani, M. R., and Rawlings, J. B., "A New Autocovariance Least-Squares Method for Estimating Noise Covariances," Tech. Rep. 2003-04, Texas-Wisconsin Modeling and Control Consortium, Sept. 2003.
- ⁹Mehra, R. K., "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Transactions on Automatic Control*, Vol. AC-15, No. 2, April 1970, pp. 175–184.
- ¹⁰Parlos, A. G., Menon, S. K., and Atiya, A. F., "An Algorithmic Approach to Adaptive State Filtering Using Recurrent Neural Networks," *IEEE Transactions on Neural Networks*, Vol. 12, No. 6, Nov. 2001, pp. 1411–1432.
- ¹¹Lho, Y. H. and Painter, J. H., "A Fuzzy-Tuned Adaptive Kalman Filter," *Proceedings of the Third International IEEE Conference on Industrial Fuzzy Control and Intelligent Systems*, Houston, TX, Dec. 1993, pp. 144–148.
- ¹²Lee, D.-J. and Alfriend, K. T., "Adaptive Sigma Point Filtering for State and Parameter Estimation," *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, Providence, RI, Aug. 2004, AIAA-2004-5101.

- ¹³Crassidis, J. L. and Cheng, Y., "Generalized Multiple-Model Adaptive Estimation Using an Autocorrelation Approach," *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006, paper 223.
- ¹⁴Magill, D. T., "Optimal Adaptive Estimation of Sampled Stochastic Processes," *IEEE Transactions on Automatic Control*, Vol. AC-10, No. 4, Oct. 1965, pp. 434–439.
- ¹⁵Hanlon, P. D. and Maybeck, P. S., "Multiple-Model Adaptive Estimation Using a Residual Correlation Kalman Filter Bank," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-36, No. 2, April 2000, pp. 393–406.
- ¹⁶Hammersley, J. M., "Monte Carlo Methods for Solving Multivariable Problems," *Annals of the New York Academy of Sciences*, Vol. 86, 1960, pp. 844–874.
- ¹⁷van Loan, C. F., "Computing Integrals Involving the Matrix Exponential," *IEEE Transactions on Automatic Control*, Vol. AC-23, No. 3, June 1978, pp. 396–404.
- ¹⁸Brown, R. G. and Hwang, P. Y. C., *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, New York, NY, 3rd ed., 1997, pp. 353–361.
- ¹⁹Stengel, R. F., *Optimal Control and Estimation*, Dover Publications, New York, NY, 1994, pp. 402–407.
- ²⁰Anderson, B. D. O. and Moore, J. B., *Optimal Filtering*, chap. 10.1, Dover Publications, Mineola, NY, 2005.
- ²¹Anderson, B. D. O., Moore, J. B., and Hawkes, R. M., "Model Approximations via Prediction Error Identification," *Automatica*, Vol. 14, No. 6, Nov. 1978, pp. 615–622.
- ²²Alsuwaidan, B., *Generalized Multiple Model Adaptive Estimation*, Ph.D. dissertation, University at Buffalo, The State University of New York, Amherst, NY, 2008.
- ²³Sprent, P., "Inversion of Nearly Diagonal Matrices," *The Mathematical Gazette*, May 1965, pp. 184–189.
- ²⁴Kocis, L. and Whiten, W. J., "Computational Investigations of Low-Discrepancy Sequences," *ACM Transactions on Mathematical Software*, Vol. 23, No. 2, June 1997, pp. 266–294.