Generalized Multiple-Model Adaptive Estimation Using an Autocorrelation Approach

John L. Crassidis
Dept. of Mech. & Aero. Eng.
University at Buffalo
State University of New York
Amherst, NY 14260-4400 U.S.A.
johnc@eng.buffalo.edu

Yang Cheng
Dept. of Mech. & Aero. Eng.
University at Buffalo
State University of New York
Amherst, NY 14260-4400 U.S.A.
cheng3@eng.buffalo.edu

Abstract - ***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. Simulation results, involving a two-dimensional target tracking problem using an extended Kalman filter, indicate that the new approach is able to correctly estimate the noise statistics.***

Keywords: Multiple-model adaptive estimation, filtering, extended Kalman filter, target tracking.

1 Introduction

Filtering algorithms, such as the extended Kalman filter (EKF) [1], the Unscented filter (UF) [2] and Particle filters (PFs) [3, 4], 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 and/or measurement models are nonlinear, since the posterior 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 posterior 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 posterior 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 posterior 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 [5]. Positive kurtosis indicates a relatively peaked distribution, while negative kurtosis indicates a relatively flat distribution. However, the process noise is extremely difficult to characterize because it is usually used to represent modeling errors. The 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 [6]. 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 [7]. 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" [8]. 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 [8].

Adaptive filtering for nonlinear systems has recently gained attention. Parlos et al. shows a neural net to constructively approximate the state equations [9]. The proposed algorithms in their paper 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 [10]. 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 [11]. The standard update approach requires proper selection of a window size to control the level of the variance update. The innovation of their paper 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.

In this paper a new approach is derived for adaptive filtering based on generalizing the standard multiple-model adaptive estimation (MMAE) algorithm [12]. 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. Our approach is a generalization of Ref. [13], 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. [13]. 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.

Our approach 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 sequence [14] 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.

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. Finally, simulation results involving a two-dimensional target tracking problem are shown.

2 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 mea-

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	$\hat{\mathbf{x}}(t_0) = \hat{\mathbf{x}}_0$ $P_0 = E\left\{\tilde{\mathbf{x}}(t_0)\,\tilde{\mathbf{x}}^T(t_0)\right\}$
Gain	$K_k = P_k^- H_k^T(\hat{\mathbf{x}}_k^-)[H_k(\hat{\mathbf{x}}_k^-)P_k^- H_k^T(\hat{\mathbf{x}}_k^-) + R_k]^{-1}$ $H_k(\hat{\mathbf{x}}_k^-) \equiv \left.\dfrac{\partial \mathbf{h}}{\partial \mathbf{x}}\right\vert_{\hat{\mathbf{x}}_k^-}$
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	$\dot{\hat{\mathbf{x}}}(t) = \mathbf{f}(\hat{\mathbf{x}}(t),\, \mathbf{u}(t),\, t)$ $\dot{P}(t) = F(\hat{\mathbf{x}}(t),\, t)\,P(t) + P(t)\,F^T(\hat{\mathbf{x}}(t),\, t) + G(t)\,Q(t)\,G^T(t)$ $F(\hat{\mathbf{x}}(t),\, t) \equiv \left.\dfrac{\partial \mathbf{f}}{\partial \mathbf{x}}\right\vert_{\hat{\mathbf{x}}(t)}$

surement, $\mathbf{v}_k$ 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:

$$P_{k+1}^- = \Phi_k P_k^+ \Phi_k^T + Q_k \tag{1}$$

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 [15]. First, the following $2n \times 2n$ matrix is formed:

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

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

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

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

$$\Phi_k = \mathcal{B}_{22}^T \tag{4}$$

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

$$Q_k = \Phi_k\,\mathcal{B}_{12} \tag{5}$$

If the sampling interval is "small" enough, then a good approximation for the solution of Q_k is given by

$$Q_k \approx \Delta t\, G(t)\,Q(t)\,G^T(t) \tag{6}$$

Note that 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.

3 Multiple-Model Adaptive Estimation

In this section a review of MMAE is shown. More details can be found in Refs. [16, 17]. Multiple-model adaptive estimation is a recursive estimator that uses a bank of filters that depend on some unknown parameters. In our case these parameters are the process noise covariance, 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 $\{\mathbf{p}^{(\ell)};\, \ell = 1,\, \ldots,\, M\}$. The goal of the estimation process is to determine the conditional pdf of the ℓ^{th} element $\mathbf{p}^{(\ell)}$ given the current-time measurement $\tilde{\mathbf{y}}_k$. Application of Bayes' rule yields

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

where $\tilde{\mathbf{Y}}_k$ denotes the sequence $\{\tilde{\mathbf{y}}_0, \tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_k\}$. The *a posteriori* probabilities can be computed through [18]

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

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

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

where $\varpi_k^{(\ell)} \equiv p\left(\mathbf{p}^{(\ell)}|\tilde{\mathbf{y}}_k\right)$. The weights at time t_0 are initialized to $\varpi_0^{(\ell)} = 1/M$ for $\ell = 1, 2, \ldots, M$. The convergence properties of MMAE are shown in Ref. [18], 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 [19].

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

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

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

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

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

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

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

Equation (12b) can be used to define 3σ bounds on the estimate $\hat{\mathbf{p}}_k$.

4 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.

4.1 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. [8]. Here we assume that the model is linear with

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

$$\tilde{\mathbf{y}}_k = H_k \mathbf{x}_k + \mathbf{v}_k \tag{13b}$$

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

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

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

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

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

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

Carrying Eq. (16) i steps back leads to

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

where

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

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

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

where

$$C_{k-i,\,0} \equiv H_{k-i}\, P^{-}_{k-i}\, H^{T}_{k-i} + R_{k-i} \tag{20}$$

Note that storage of the state model and covariance matrices to the $k-i$ point is required to compute $C_{k,\,i}$ in general.

4.2 Likelihood Function

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

$$\boldsymbol{\epsilon}_i \equiv \begin{bmatrix} \mathbf{e}_k \\ \mathbf{e}_{k-1} \\ \vdots \\ \mathbf{e}_{k-i} \end{bmatrix} \tag{21}$$

The likelihood function associated with $\boldsymbol{\epsilon}_i$ is given by

$$L_i = \frac{1}{[\det(2\pi\,\mathcal{C}_i)]^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_i^T \mathcal{C}_i^{-1} \boldsymbol{\epsilon}_i\right) \tag{22}$$

where $\mathcal{C}_i \equiv E\{\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T\}$ is given by

$$\mathcal{C}_i = \begin{bmatrix} C_{k,\,0} & C_{k,\,1} & C_{k,\,2} & \cdots & C_{k,\,i} \\ C^T_{k,\,1} & C_{k-1,\,0} & C_{k-1,\,2} & \cdots & C_{k-1,\,i-1} \\ C^T_{k,\,2} & C^T_{k-1,\,2} & C_{k-2,\,0} & \cdots & C_{k-2,\,i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C^T_{k,\,i} & C^T_{k-1,\,i-1} & C^T_{k-2,\,i-2} & \cdots & C_{k-i,\,0} \end{bmatrix} \tag{23}$$

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

$$\begin{aligned} L_0 = &\frac{1}{\left\{\det[2\pi\,(H_k P_k^- H_k^T + R_k)]\right\}^{1/2}} \\ &\times \exp\left[-\frac{1}{2}\mathbf{e}_k^T (H_k P_k^- H_k^T + R_k)^{-1} \mathbf{e}_k\right] \end{aligned} \tag{24}$$

This likelihood is widely used in MMAE algorithms [12, 16], 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.

4.3 New Adaptive Law

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

$$\begin{aligned} \varpi_k^{(\ell)} &= \varpi_{k-1}^{(\ell)} L_0^{(\ell)} \\ \varpi_k^{(\ell)} &\leftarrow \frac{\varpi_k^{(\ell)}}{\sum_{j=1}^{M} \varpi_k^{(j)}} \end{aligned} \tag{25}$$

since $p\,(\tilde{\mathbf{y}}_k | \hat{\mathbf{x}}_k^{-(\ell)}) = L_0^{(\ell)}$, which is defined by

$$\begin{aligned} L_0^{(\ell)} = &\frac{1}{\left\{\det[2\pi\,(H_k P_k^{-(\ell)} H_k^T + R_k)]\right\}^{1/2}} \\ &\times \exp\left[-\frac{1}{2}\mathbf{e}_k^{(\ell)T} (H_k P_k^{-(\ell)} H_k^T + R_k)^{-1} \mathbf{e}_k^{(\ell)}\right] \end{aligned} \tag{26}$$

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

The new adaptive law, which we call the generalized multiple-model adaptive estimation (GMMAE) algorithm, is based on carrying Eq. (8) i steps back to give the new update law:

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

with

$$L_i^{(\ell)} = \frac{1}{\left[\det\left(2\pi\,\mathcal{C}_i^{(\ell)}\right)\right]^{1/2}} \exp\left[-\frac{1}{2}\boldsymbol{\epsilon}_i^{(\ell)T} \left(\mathcal{C}_i^{(\ell)}\right)^{-1} \boldsymbol{\epsilon}_i^{(\ell)}\right] \tag{28}$$

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 $\mathcal{C}_i^{(\ell)}$ 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. [8]. 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 residual is maintained. Estimates for $C_{k,\,i}^{(\ell)}$ are given by

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

where

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

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

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

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

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

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

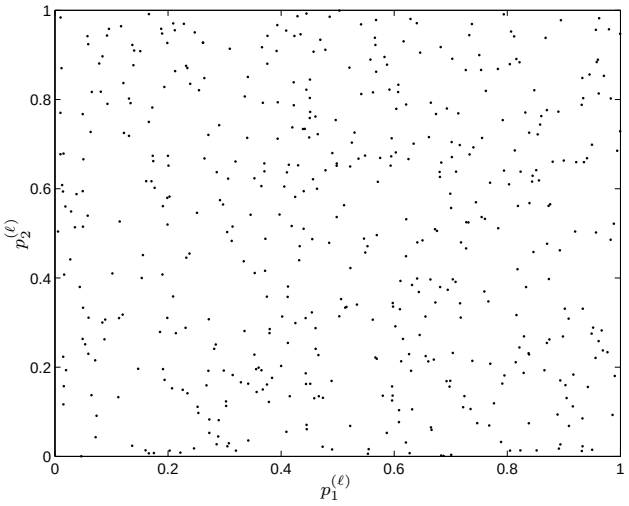
$$\hat{P}_{k+1}^- = \Phi_k(\hat{\mathbf{x}}_k^-) \hat{P}_k^+ \Phi_k^T(\hat{\mathbf{x}}_k^-) + \hat{Q}_k \quad (33a)$$

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

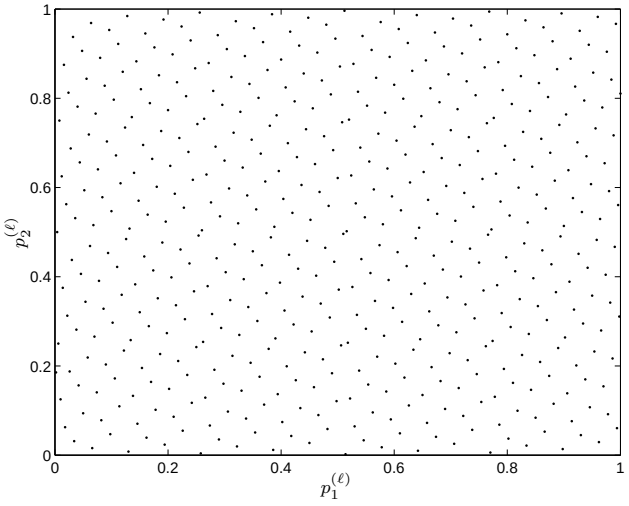
where $\hat{Q}_k$ is computed using $\hat{\mathbf{p}}_k$.

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, $\boldsymbol{\epsilon}_i^{(\ell)}$, 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 $\hat{\mathcal{C}}_i^{(\ell)}$. Note that at each new measurement time, all elements of $\hat{\mathcal{C}}_i^{(\ell)}$ 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 $L_i^{(\ell)}$. These functions are used to update the weights, which gives the estimate $\hat{\mathbf{p}}_k$ using Eq. (12a).

There are many possibilities for the chosen distribution of the process noise covariance parameters. A simple approach is to a assume a uniform distribution. We instead choose a Hammersley quasi-random sequence [14] 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 1. 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. [20] if needed.



(a) Uniform Distribution



(b) Hammersley Quasi-Random Sequence

Figure 1: Uniform Distribution and Hammersley Quasi-Random Sequence Comparison

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

$$\mathbf{x}_{k+1} = \Phi \mathbf{x}_k + \mathbf{w}_k \quad (34)$$

with

$$\Phi = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (35)$$

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

$$Q_k = \begin{bmatrix} q_x \begin{bmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t \end{bmatrix} & 0_{2\times 2} \\ 0_{2\times 2} & q_y \begin{bmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t \end{bmatrix} \end{bmatrix} \quad (36)$$

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

$$\tilde{\mathbf{y}}_k = \begin{bmatrix} \sqrt{x^2+y^2} \\ \arctan(y/x) \end{bmatrix} + \mathbf{v}_k \tag{37}$$

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 GMMAE 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.

The two adaptive estimators are run 50 seconds to process the same measurement data. As discussed in the previous section, the GMMAE goes back i time steps in order to form the residual. We choose $i = 4$ for the GMMAE. The size of the corresponding residual $\boldsymbol{\epsilon}_i^{(\ell)}$ is 10 by 1. The results are given in Figures 2 and 3. It can be seen that both estimators converge within 50 seconds, but the MMAE takes more than twice as much time as the GMMAE 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 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.

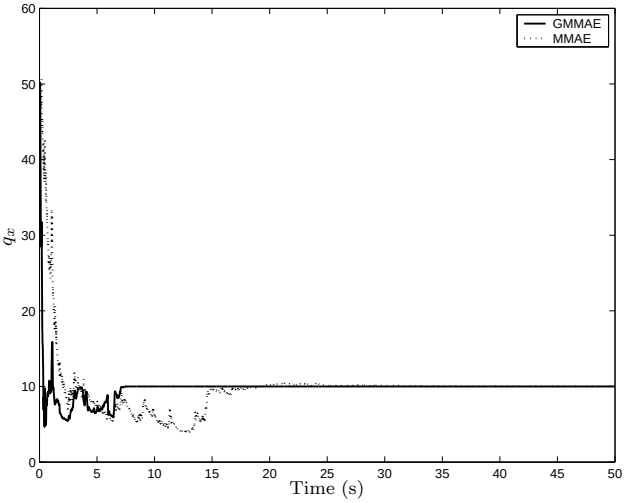


Figure 2: Estimates of q_x

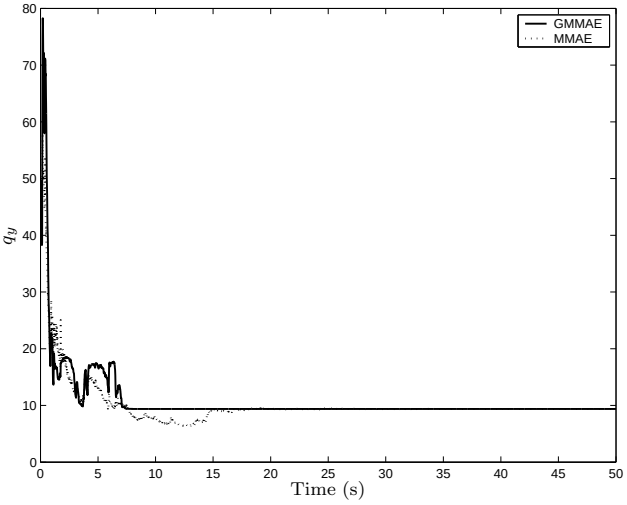


Figure 3: Estimates of q_y

6 Conclusions

In this paper a new approach for multiple-model adaptive estimation was derived, 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. Another approach involves using the estimated values in the state matrices and us-

ing the elements in the Kalman gain computation. Further studies are under way to assess the differences between these approaches. Simulation results indicated that the new multiple-model adaptive estimation approach can provide better convergence properties than the standard approach.

Acknowledgment

This work was supported through an Independent Research and Development (IRAD) award by L-3 Communications Integrated Systems, Greenville, TX, under the supervision of Dr. Gerald L. Fudge. The authors greatly appreciate the support.

References

[1] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, March 1961, pp. 95–108.

[2] E. Wan and R. van der Merwe, "The Unscented Kalman Filter," *Kalman Filtering and Neural Networks*, edited by S. Haykin, chap. 7, John Wiley & Sons, New York, NY, 2001.

[3] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "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.

[4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "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.

[5] K. Pearson, "Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder," *Biometrika*, Vol. 4, 1905, pp. 169–212.

[6] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, New York, NY, 2001, pp. 232–245.

[7] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings, "A New Autocovariance Least-Squares Method for Estimating Noise Covariances," Tech. Rep. 2003-04, Texas-Wisconsin Modeling and Control Consortium, Sept. 2003.

[8] R. K. Mehra, "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Transactions on Automatic Control*, Vol. AC-15, No. 2, April 1970, pp. 175–184.

[9] A. G. Parlos, S. K. Menon, and A. F. Atiya, "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.

[10] Y. H. Lho and J. H. Painter, "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.

[11] D.-J. Lee and K. T. Alfriend, "Adaptive Sigma Point Filtering for State and Parameter Estimation," *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, Providence, RI, Aug. 2004, AIAA-2004-5101.

[12] D. T. Magill, "Optimal Adaptive Estimation of Sampled Stochastic Processes," *IEEE Transactions on Automatic Control*, Vol. AC-10, No. 4, Oct. 1965, pp. 434–439.

[13] 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, April 2000, pp. 393–406.

[14] J. M. Hammersley, "Monte Carlo Methods for Solving Multivariable Problems," *Annals of the New York Academy of Sciences*, Vol. 86, 1960, pp. 844–874.

[15] C. F. van Loan, "Computing Integrals Involving the Matrix Exponential," *IEEE Transactions on Automatic Control*, Vol. AC-23, No. 3, June 1978, pp. 396–404.

[16] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, New York, NY, 3rd ed., 1997, pp. 353–361.

[17] R. F. Stengel, *Optimal Control and Estimation*, Dover Publications, New York, NY, 1994, pp. 402–407.

[18] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, chap. 10.1, Dover Publications, Mineola, NY, 2005.

[19] B. D. O. Anderson, J. B. Moore, and R. M. Hawkes, "Model Approximations via Prediction Error Identification," *Automatica*, Vol. 14, No. 6, Nov. 1978, pp. 615–622.

[20] L. Kocis and W. J. Whiten, "Computational Investigations of Low-Discrepancy Sequences," *ACM Transactions on Mathematical Software*, Vol. 23, No. 2, June 1997, pp. 266–294.