A robust fault detection and identification strategy for aerospace systems

Andrew Lee, Waleed Hilal, D. Ciampini, S. Andrew Gadsden, Mohammad Al-Shabi

Andrew S. Lee, Waleed Hilal, D. Ciampini, S. Andrew Gadsden, Mohammad Al-Shabi, "A robust fault detection and identification strategy for aerospace systems," Proc. SPIE 12547, Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII, 1254707 (14 June 2023); doi: 10.1117/12.2663917

Event: SPIE Defense + Commercial Sensing, 2023, Orlando, Florida, United States

A robust fault detection and identification strategy for aerospace systems

Andrew S. Lee[a], Waleed Hilal[a], D. Ciampini[a], S. Andrew Gadsden[a,*], and Mohammad Al-Shabi[b]
[a]McMaster University, 1280 Main St W, Hamilton, ON, Canada L8S 4L8;
[b]University of Sharjah, University City Rd, Sharjah, UAE 27272

ABSTRACT

Fault detection and identification strategies utilize knowledge of the systems and measurements to accurately and quickly predict faults. These strategies are important to mitigate full system failures, and are particularly important for the safe and reliable operation of aerospace systems. In this paper, a relatively new estimation method called the sliding innovation filter (SIF) is combined with the interacting multiple model (IMM) method. The corresponding method, referred to as the SIF-IMM, is applied on a magnetorheological actuator which was built for experimentation. These types of actuators are similar to hydraulic-based ones, which are commonly found in aerospace systems. The method is shown to accurately identify faults in the system. The results are compared and discussed with other popular nonlinear estimation strategies including the extended and unscented Kalman filters.

Keywords: Aerospace systems, fault detection, identification, robustness, Kalman filter, sliding innovation filter

1. BRIEF INTRODUCTION

Electromechanical systems are frequently subject to diverse operational modes resulting from various factors, such as design, environmental conditions, or faults. Modeling each operational mode can facilitate the application of adaptive estimation strategies to enhance estimation accuracy and enable effective fault detection. In the case of the magnetorheological damper examined in this study, it is highly susceptible to changes in temperature and power supply faults that can significantly alter its behaviour. Such sudden and unpredictable variations in the system give rise to substantial degrees of uncertainty. Designing filtering strategies to estimate the output force using adaptive approaches thus is critically reliant on incorporating multiple models to accommodate for all operating modes and to minimize estimation error.

Multiple model (MM) algorithms, which operate on a Bayesian framework for adaptive estimation, comprise various forms including static and dynamic MM, generalized pseudo-Bayesian, and the interacting MM (IMM) [1]–[8]. The fundamental Bayesian principle underlying MM methods is the updating of the probability of a system existing in a particular mode based on new incoming measurements. These algorithms incorporate a finite number of modes and employ state estimates to compute the probability of each mode. The IMM-KF is a widely used MM method that employs a number of Kalman filters (KFs) equivalent to the number of system models running in parallel. The KF is a popular choice due to its optimality and simple corrective gain calculation. Nevertheless, the KF only yields optimal state estimates for linear systems with white noise (or zero-mean with normal distribution properties) [9]. The Kalman gain is computed by minimizing the trace of the *a priori* (predicted) state error covariance, which is a measure of the estimation error [9]–[11]. The KF has been applied in several fields, such as signal processing, fault detection, and target tracking [9]. However, disturbances, nonlinearities, and modeling uncertainties may result in unstable estimates.

Nonlinear behavior is ubiquitous in natural systems. To approximate nonlinear processes, the extended KF (EKF) performs local linearization around the *a priori* state estimate [9]. Specifically, the nonlinear system model and measurement process are approximated using first-order Taylor series to produce Jacobian matrices. The Jacobians can be utilized to compute the corrective Kalman gain for the states and their covariance. However, if the system is highly nonlinear, the EKF estimates may diverge from the true state trajectory [12]. An alternative approach to capturing nonlinear behavior is through sampling. The unscented KF (UKF) generates samples from a probability distribution of states propagated through the system model, known as sigma points [13]. The unscented transform is a deterministic sampling metho that selects a minimal number of sample points around a mean (in this case, the previous state estimate) [9]. Monte Carlo sampling can be used to approximate the mean and covariance of the projected points. Unlike the EKF, the UKF can

Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII, edited by Ivan Kadar,
Erik P. Blasch, Lynne L. Grewe, Proc. of SPIE Vol. 12547, 1254707
 · doi: 10.1117/12.2663917

approximate the updated statistical state mean and state error covariance up to the third order for nonlinear processes [12]. Additionally, the UKF does not necessitate taking partial derivatives of the system model or measurement process. Nonetheless, the unscented transform is generally more computationally expensive than the EKF [9].

Variable structure control (VSC) and sliding mode control (SMC) frameworks have been applied in the design of sliding mode observers (SMOs) [13]. The SMOs employ an innovation-based approach to determine the observer gain that ideally drives the error surface towards the origin [13]. A sliding surface or hyperplane is defined in SMOs to apply a discontinuous switching force that keeps the state estimates bounded within a region of the sliding surface [14]. The smooth variable structure filter (SVSF), based on the concepts of SMOs, was introduced in 2007 [15].. The SVSF gain is computed using the measurement error and a switching term, which bounds the state estimates to the trajectory of the true state values, thereby improving the estimation process stability. While classical model-based filters calculate the corrective gain using the state error covariance, the original formulation of SVSF did not include this covariance [15].. Subsequently, the corrective gain was modified through a process of minimizing the state error covariance to obtain a time-varying smoothing boundary layer [15].. The widths of these layers are adjustable based on the level of uncertainty in the estimation process. Furthermore, the SVSF has been enhanced through the use of a chattering function for higher-order solutions and fault detection [13], [16], [17].

The sliding innovation filter (SIF) was first presented in 2020 based on SMOs as an improvement over the SVSF [18]. The SIF retains robustness to uncertainties but uses a more concise gain structure and higher estimation accuracy. This paper proposes a novel IMM strategy that uses an extension of the SIF for nonlinear systems known as the extending sliding innovation filter (ESIF). Like the EKF, the ESIF uses the Jacobian matrix for linear approximation of the system to calculate the *a priori* state error covariance [19], [20]. The IMM algorithm is combined with the ESIF to form the IMM-ESIF [21]. The efficacy of this proposed estimation strategy is compared with the IMM-EKF and IMM-UKF. These filters are applied using a highly nonlinear polynomial model of an experimental MR damper to estimate the force exerted by the damper.

The paper is organized as follows. The estimation methods used in this paper are provided in Section II, followed by the IMM algorithm in Section III. Details about the experimental setup are described in Section IV. The mathematical model of the MR damper is provided in Section IV, followed by experimental results in Section VI.

2. REVIEW OF ESTIMATION STRATEGIES

In this section, we provide a review of the main estimation strategies used in this paper. In particular, the Kalman filter and sliding innovation filter are summarized.

2.1 Kalman Filter

Although it is the most popular method, the KF will be summarized here for completeness. The KF provides the optimal solution to the linear estimation problem which is described by (2.1.1) and (2.1.2). The goal of any estimator is to obtain the true state value x_{k+1} using noisy measurements z_{k+1}.

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{2.1.1}$$

$$z_{k+1} = Cx_{k+1} + v_{k+1} \tag{2.1.2}$$

The system and measurement noises are represented by w_k and v_k, respectively. A, B, and C represent the system (or process) matrix, input gain matrix, and measurement matrix, respectively. In (2.1.1) and (2.1.2), it is assumed that these matrices are fixed and do not change with time. The input to the system is defined as u_k. For the KF and most estimation methods, it is assumed that the system and measurement noises are statistically zero mean with Gaussian distribution [1]. The system and measurement noise are generated using the covariance matrices Q and R, respectively.

The KF is formulated as a predictor-corrector estimator and is an iterative process. The prediction stage involves calculating the state estimates based on the previous state values and knowledge of the system, as per (2.1.3). The corresponding state error covariance matrix is calculated in (2.1.4) and is used in the update stage to calculate the KF gain in (2.1.5) and update the state error covariance as per (2.1.7).

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k \tag{2.1.3}$$

$$P_{k+1|k} = AP_{k|k}A^T + Q_k \tag{2.1.4}$$

The update stage is summarized by (2.1.5) through (2.1.7). The gain calculated in (2.1.5) is used to update the state estimates in (2.1.6) based on the measurement error (or innovation). The gain is also used along with the predicted state error covariance to update the state error covariance in (2.1.7).

$$K_{k+1} = P_{k+1|k}C^T\left(CP_{k+1|k}C^T + R_{k+1}\right)^{-1} \tag{2.1.5}$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\left(z_{k+1} - C\hat{x}_{k+1|k}\right) \tag{2.1.6}$$

$$\begin{aligned} P_{k+1|k+1} = (I - K_{k+1}C)P_{k+1|k}(I - K_{k+1}C)^T \ldots \\ \ldots + K_{k+1}R_{k+1}K_{k+1}^T \end{aligned} \tag{2.1.7}$$

Note that k refers to the time step, $k|k$ refers to the updated values at the previous iteration, and $k+1|k$ refers to the predicted values at time $k+1$ based on information at time k. Equations (2.1.3) through (2.1.7) represent the KF estimation process for linear systems and measurements defined by (2.1.1) and (2.1.2), respectively. The process is iterative and repeats every time step k. Note that (2.1.7) is known as the Joseph covariance form, and is considered to be numerically stable. The basic nonlinear form of the KF, known as the extended Kalman filter (EKF), is based on linearizing the nonlinear system and/or measurement equations by first-order Taylor series expansions. This is described later at the end of Section 2.3.

2.2 Unscented Kalman Filter

Another method of dealing with nonlinearities involves using statistical linear regression of sample points projected using the nonlinear system model [22]. The unscented Kalman filter (UKF) is a popular formulation of the sigma-point Kalman filter (SPKF). The UKF generates sigma points based on the previous state estimate and covariances. The sigma points are then projected using the nonlinear system model to form the *a priori* state estimate and state error covariance in a process known as the unscented transform [23], [24]. Additionally, the points are also projected using the nonlinear measurement function as well. This method removes the need for linearization and generally produces a more accurate estimate than the Jacobian approximation of the nonlinear system [23], [25]–[27].

The UKF algorithm is detailed in the following equations [28]. Given a state space with dimension n, the state x_k can be represented with $2n+1$ sigma points denoted by X. The sigma points have a mean of $\hat{x}_{k|k}$ and a covariance of $P_{k|k}$. The initial sigma point $X_{0,k|k}$ and corresponding weight W_0 are given as follows:

$$X_{0,k|k} = \hat{x}_{k|k} \tag{2.2.1}$$

$$W_0 = \kappa/(n+\kappa) \tag{2.2.2}$$

where κ is a design parameter. The next $2n$ number of sigma points are calculated as follows:

$$X_{i,k|k} = \hat{x}_{k|k} + \left(\sqrt{(n+\kappa)P_{k|k}}\right)_i \tag{2.2.3}$$

$$W_i = 1/[2(n+\kappa)] \tag{2.2.4}$$

where the value $X_{i,k|k}$ is the i^{th} sigma point and W_i is the weight that is associated with the i^{th} sigma point [29]. The sigma points are projected ($\hat{X}_{i,k+1|k}$) through the nonlinear system function f and added together with their corresponding weights to produce the *a priori* state estimate $\hat{x}_{k+1|k}$ as follows [9]:

$$\hat{X}_{i,k+1|k} = f\left(X_{i,k|k}, u_k\right) \tag{2.2.5}$$

$$\hat{x}_{k+1|k} = \sum_{i=0}^{2n} W_i \hat{X}_{i,k+1|k} \tag{2.2.6}$$

The previous calculations are used to calculate the *a priori* state error covariance as follows [9]:

$$P_{k+1|k} = \sum_{i=0}^{2n} W_i\left(\hat{X}_{i,k+1|k} - \hat{x}_{k+1|k}\right)\left(\hat{X}_{i,k+1|k} - \hat{x}_{k+1|k}\right)^T \tag{2.2.7}$$

The sigma points are also propagated through the nonlinear measurement function. Unlike the KF and EKF, the UKF calculates a predicted measurement $\hat{z}_{k+1|k}$ which is used to produce the innovation covariance $P_{zz.k+1|k}$.

$$\hat{Z}_{i,k+1|k} = h\left(\hat{X}_{i,k+1|k}, u_k\right) \tag{2.2.8}$$

$$\hat{z}_{k+1|k} = \sum_{i=0}^{2n} W_i \hat{Z}_{i,k+1|k} \tag{2.2.9}$$

$$P_{zz.k+1|k} = \sum_{i=0}^{2n} W_i \left(\hat{Z}_{i,k+1|k} - \hat{z}_{k+1|k}\right) \dots \\ * \left(\hat{Z}_{i,k+1|k} - \hat{z}_{k+1|k}\right)^T \tag{2.2.10}$$

The cross-covariance (with respect to the state and measurement) is calculated as follows [9]:

$$P_{xz.k+1|k} = \sum_{i=0}^{2n} W_i \left(\hat{X}_{i,k+1|k} - \hat{x}_{k+1|k}\right) \dots \\ * \left(\hat{Z}_{i,k+1|k} - \hat{z}_{k+1|k}\right)^T \tag{2.2.11}$$

The cross-covariance $P_{xz.k+1|k}$ and innovation covariance $P_{zz.k+1|k}$ are combined to produce the corrective gain K_{k+1} as follows:

$$K_{k+1} = P_{xz.k+1|k} P_{zz.k+1|k}^{-1} \tag{2.2.12}$$

To conclude the update state of the UKF, the *a posteriori* state estimate and *a posteriori* state error covariance are given as follows [9]:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\left(z_{k+1} - \hat{z}_{k+1|k}\right) \tag{2.2.13}$$

$$P_{k+1|k+1} = P_{k+1|k} - K_{k+1} P_{zz,k+1|k} K_{k+1}^T \tag{2.2.14}$$

In the case of the UKF, there is a trade-off between computational cost and accuracy. While the EKF only propagates a single state estimate through a nonlinear process, the UKF uses $2n + 1$ sigma points to achieve a more accurate state estimate and state error covariance The UKF is comparable to the EKF for mildly nonlinear systems but has better performance when the nonlinear process cannot be approximated by a first order Taylor series [11].

2.3 Sliding Innovation Filter

The sliding innovation filter (SIF) is a predictor-corrector estimator based on sliding mode concepts [2]. The difference between the KF and SIF is the structure of the corrective gain matrix. The SIF gain is calculated using the measurement matrix, innovation, and sliding boundary layer term. An initial estimate is pushed towards the sliding boundary layer which is based on the upper limit of uncertainties in the estimation process [2]. If the estimate is within the sliding boundary layer, the estimates are forced to switch about the true state trajectory by the SIF gain. Figure 1 illustrates the SIF estimation concept. This section describes the linear SIF estimation process. The prediction stage is given by the following equations:

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k \tag{2.2.1}$$

$$P_{k+1|k} = AP_{k|k}A^T + Q_{k+1} \tag{2.2.2}$$

$$\tilde{z}_{k+1|k} = z_{k+1} - C\hat{x}_{k+1|k} \tag{2.2.3}$$

where x refers to the state, $\hat{x}$ refers to the estimated state, u refers to the system input, z refers to the measurement, $\tilde{z}$ refers to the innovation (or measurement error), and k refers to the time step. In addition, A, B, C, P, Q, and R, are respectively defined as the system matrix, input gain matrix, measurement matrix, state error covariance matrix, system noise covariance, and measurement noise covariance. Note also that $k + 1|k$ and $k + 1|k + 1$ refer to predicted and updated values, respectively.

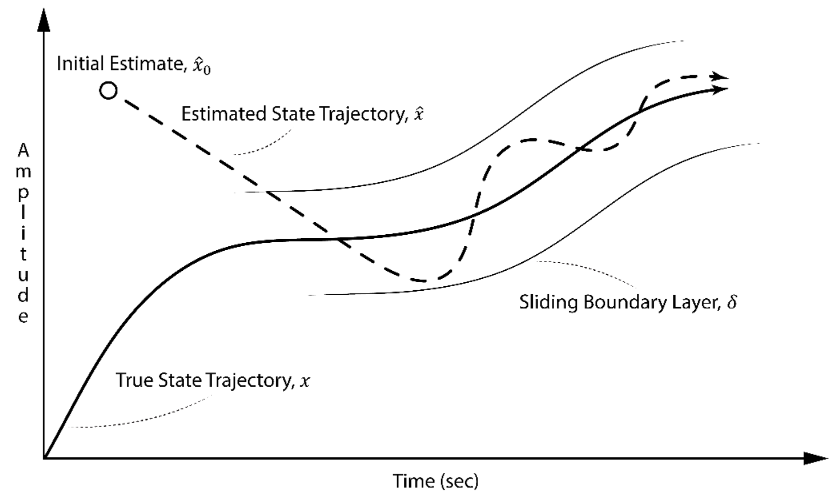

Figure 1. The sliding innovation filter (SIF) concept illustrating the effects of the switching gain and sliding boundary layer [2].

The states are predicted in (2.2.1) before being updated in (2.2.5) using the innovation defined in (2.2.3) which is also used in the gain formulation in (2.2.4). The state error covariance matrix is predicted in (2.2.2) before being updated in (2.2.6). Note that the gain (2.2.4) is also used to update the state error covariance (2.2.6). The <u>update state</u> is summarized by the following equations:

$$K_{k+1} = C^{+}\overline{sat}\left(\left|\tilde{z}_{k+1|k}\right|/\delta\right) \tag{2.2.4}$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\tilde{z}_{k+1|k} \tag{2.2.5}$$

$$\begin{aligned} P_{k+1|k+1} = (I - K_{k+1}C^{+})P_{k+1|k}(I - K_{k+1}C^{+})^{T} \ldots \\ \ldots + K_{k+1}R_{k+1}K_{k+1}^{T} \end{aligned} \tag{2.2.6}$$

where C^{+} refers to the pseudoinverse of the measurement matrix, $\left|\tilde{z}_{k+1|k}\right|$ refers to the absolute innovation value, T refers to transpose of a vector or matrix, δ refers to the fixed sliding boundary layer width, and $\overline{sat}$ refers to the diagonal matrix of the saturated vector values. The sliding boundary layer term may be tuned based on designer knowledge of the system (e.g., level of noise) in an effort to minimize the state estimation error.

Equations (2.2.1) through (2.2.6) represent the SIF estimation process for linear systems and measurements. The SIF proof of stability was discussed in detail in [2]. A Lyapunov function was defined based on the updated innovation, and was used to prove stability. Note that the nonlinear version of the SIF, the extended SIF (ESIF), is similar to the SIF with the main difference being the formulation of the gain [2]. Similar to the EKF, the ESIF uses Jacobian matrices to linearize the nonlinear system $f(\hat{x}_{k|k}, u_k)$ and nonlinear measurement $h(\hat{x}_{k+1|k})$ functions, respectively as follows:

$$F_k = \left.\frac{\partial f}{\partial x}\right|_{(\hat{x}_{k|k}, u_k)} \tag{2.2.7}$$

$$H_{k+1} = \left.\frac{\partial h}{\partial x}\right|_{(\hat{x}_{k+1|k})} \tag{2.2.8}$$

In its current formulation, the state error covariance matrix P defined in the SIF estimation process is not used to update the state estimates. However, as will be shown in Section III, it is used to derive a time-varying sliding boundary layer.

3. FORMULATION OF THE SIF-BASED IMM STRATEGY

The interacting multiple model (IMM) method incorporates a finite number of models and filtering strategies that run in parallel. Each filter associated to a particular model produces its own state estimate, sate error covariance, and likelihood that the model is correct. The likelihood is a function of the innovation (measurement error) and its covariance. This in turn is used to calculate the mode probabilities which represent the probability of the system existing in a particular mode based on the current information.

The IMM method's access to additional modeling information presents a clear advantage over single model strategies. Combining the IMM with the ESIF adds stability and robustness while increasing adaptability and accuracy with access to multiple models. In this paper, the efficacy of this strategy is tested against previous IMM strategies such as the IMM-EKF and IMM-UKF when applied to a highly nonlinear MR damper system.

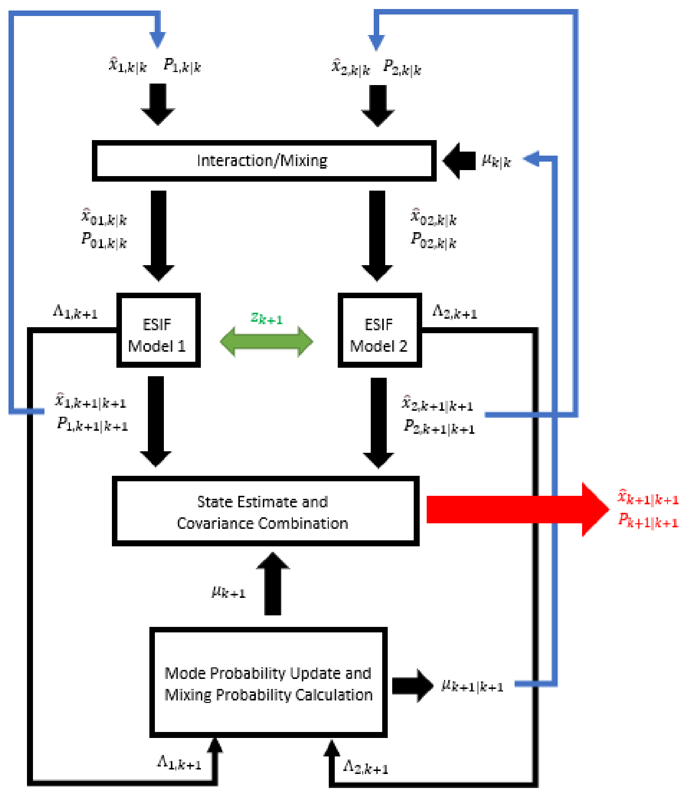

Figure 2. Overview of the proposed IMM-ESIF algorithm.

The IMM-ESIF algorithm is shown in Fig. 2. The green arrows indicate measurement input, the blue arrows indicate recursion, and the red arrow indicates the overall IMM-RSIF output. A number of SIFs equivalent to the number of models are run in parallel. While Fig 2. shows two models for conciseness, there is no limit to the number of models that can be incorporated. However, it should be noted that processing time scales linearly with each additional model. The IMM-ESIF estimator consists of five steps: mixing probability calculation, ESIF mode-matched filtering, mode probability update, and combination of the state estimate and covariance.

The mixing probabilities $\mu_{i|j,k|k}$ represent the probability of the system in mode i and switching to mode j at the next time step. The mixing probabilities are calculated as follows [8]:

$$\mu_{i|j,k|k} = \frac{1}{\bar{c}_j} p_{ij} \mu_{i,k} \tag{3.1}$$

$$\bar{c}_j = \sum_{i=1}^{r} p_{ij} \mu_{i,k} \tag{3.2}$$

where p_{ij} is the mode transition probability which is a design parameter, μ_{ik} is the probability of the system existing in mode i, and r is the number of system modes. The previous mode matched state $\hat{x}_{i,k|k}$ and covariance $P_{i,k|k}$ are used to calculate the mixed initial conditions (state $\hat{x}_{0j,k|k}$ and covariance $P_{0j,k|k}$) for the filter matched to mode j as follows [8]:

$$\hat{x}_{0j,k|k} = \sum_{i=1}^{r} \hat{x}_{ij,k|k} \mu_{i|j,k|k} \tag{3.3}$$

$$P_{0j,k|k} = \sum_{i=1}^{r} \mu_{i|j,k|k} \{P_{i,k|k} + \cdots$$
$$\left(\hat{x}_{i,k|k} - \hat{x}_{0,k|k}\right)\left(\hat{x}_{i,k|k} - \hat{x}_{0,k|k}\right)^T\} \tag{3.4}$$

These mixed initial conditions are then fed into the filters matched to mode j. Each ESIF uses the measurement z_{k+1} as well as any system inputs u_k to calculate the updated states and corresponding state error covariance. The initial state estimate $\hat{x}_{0j,k|k}$ and corresponding state error covariance $P_{0j,k|k}$ for each mode j are used to calculate the *a priori* states $\hat{x}_{j,k+1|k}$ error covariance $P_{j,k+1|k}$ as follows:

$$\hat{x}_{j,k+1|k} = f_j\left(\hat{x}_{0j,k|k}, u_k\right) \tag{3.5}$$

$$P_{j,k+1|k} = F_j P_{0j,k|k} {F_j}^T + Q_k \tag{3.6}$$

where f_j is the nonlinear system state equations of mode j and F_j is the Jacobian matrix of said equations.

The mode-matched innovation covariance $S_{j,k+1|k}$ and mode-matched *a priori* measurement error $e_{j,z,k+1|k}$ are calculated as follows [8]:

$$S_{j,k+1|k} = C_j P_{j,k+1|k} {C_j}^T + R_{k+1} \tag{3.7}$$

$$e_{j,z,k+1|k} = z_{k+1} - C_j \hat{x}_{j,k+1|k} \tag{3.8}$$

where the measurement matrix C_j is considered linear and constant for the purposes of this paper.

The update stage is described by the following four equations. The mode-matched ESIF gain $K_{j,k+1}$ is calculated via (3.9) and used to update the state estimate $\hat{x}_{j,k+1|k+1}$ (3.10).

$$K_{j,k+1} = {H_j}^{+} \overline{sat}\left(\left|e_{j,z,k+1|k}\right|/\delta\right) \tag{3.9}$$

$$\hat{x}_{j,k+1|k+1} = \hat{x}_{j,k+1|k} + K_{j,k+1} e_{j,z,k+1|k} \tag{3.10}$$

The updated state error covariance matrix $P_{j,k+1|k+1}$ is generated via (4.11) and is used to produce the a posteriori measurement error $e_{j,z,k+1|k+1}$ (4.12).

$$P_{j,k+1|k+1} = \left(I - K_{j,k+1} C_j\right) P_{j,k+1|k} \left(I - K_{j,k+1} C_j\right)^T \ldots$$
$$+ K_{j,k+1} R_{k+1} {K_{j,k+1}}^T \tag{3.11}$$

$$e_{j,z,k+1|k+k+1} = z_{k+1} - H_j \hat{x}_{j,k+1|k+1} \tag{3.12}$$

Using the mode-mode matched innovation matrix $S_{j,k+1|k}$ and the mode-matched updated measurement error $e_{j,z,k+1|k}$, a corresponding likelihood function $\Lambda_{j,k+1}$ is calculated as follows [8]:

$$\Lambda_{j,k+1} = \mathcal{N}\left(z_{k+1}; e_{j,z,k+1|k}, S_{j,k+1|k}\right) \tag{3.13}$$

The likelihood is calculated by applying measurement z_{k+1} to a Gaussian probability density function with mean $e_{j,z,k+1|k}$ and covariance $S_{j,k+1|k}$. The likelihood can be rewritten as the following equation [8]:

$$\Lambda_{j,k+1} = \frac{1}{\sqrt{\left|2\pi S_{j,k+1|k}\right|}} \exp\left(\frac{-\frac{1}{2} e_{j,z,k+1|k}^T e_{j,z,k+1|k}}{S_{j,k+1|k}}\right) \tag{3.14}$$

The mode-matched likelihood function $\Lambda_{j,k+1}$ is then used to update the mode probability $\mu_{i,k}$ as shown [8]:

$$\mu_{i,k} = \frac{1}{c}\Lambda_{j,k+1}\sum_{i=1}^{r} p_{ij}\mu_{i,k} \tag{3.15}$$

where the normalizing constant c is defined as follows [8]:

$$c = \sum_{j=1}^{r}\Lambda_{j,k+1}\sum_{i=1}^{r} p_{ij}\mu_{i,k} \tag{3.16}$$

Finally, the IMM-ESIF outputs the overall state estimates $\hat{x}_{k+1|k+1}$ and corresponding state error covariance $P_{k+1|k+1}$ which are calculated as follows [8]:

$$\hat{x}_{k+1|k+1} = \sum_{j=1}^{r}\mu_{i,k+1}\,\hat{x}_{j,k+1|k+1} \tag{3.17}$$

$$P_{k+1|k+1} = \sum_{j=1}^{r}\mu_{i,k+1}\left\{P_{i,k+1|k+1} + \cdots \right.$$
$$\left.\left(\hat{x}_{j,k+1|k+1} - \hat{x}_{k+1|k+1}\right)\left(\hat{x}_{j,k+1|k+1} - \hat{x}_{k+1|k+1}\right)^{T}\right\} \tag{3.18}$$

The formulation of the IMM-ESIF can be summarized by (3.1) – (3.18). Note that the estimator's overall output $\hat{x}_{k+1|k+1}$ from (3.17) and $P_{k+1|k+1}$ from (3.18) are not used in the algorithm recursions [8]. The IMM-EKF and IMM-UKF follow similar processes with the main difference being their respective corrective gain calculations.

4. EXPERIMENTAL SYSTEM AND RESULTS

In this section, the experimental MR damper setup and results are discussed.

4.1 Experimental Setup

The primary component of the experimental setup used in this paper is the RD-8041-1 MR damper acquired from LORD. MR dampers have numerous applications in the automotive and aerospace industry such as isolating vibrations to passengers using adaptive suspension systems [30]. A typical MR damper consists of the MR fluid itself, housing, piston, diaphragm, and magnetic coil [31]. An electrical current is supplied to the damper in order to increase the viscosity of the MR fluid which in turn, increases the damping force. The change in viscosity is attributed to the rearrangement of the ferromagnetic particles suspended in the fluid. In the presence of a magnetic field, the particles align to form linear chain structures [31]. As the MR damper is driven, the MR fluid moves between different chambers via small orifices in the piston assembly and converts mechanical energy into friction losses [31].

The experimental setup was developed at McMaster University and the University of Guelph. In order to mathematically model the MR damper, an A1 series linear actuator from Ultramotion was used to drive the damper. A RAS1-500S-S resistive load cell acquired from Loadstar was used to measure the damping force and a Korad programmable power supply was used to supply current to the MR damper. Data acquisition and commands were delivered using RS232 serial communication on a laboratory computer. The components were mounted together using an extruded t-slotted aluminum frame as seen in Fig. 3.

Figure 3. Magnetorheological testing setup used in this study.

The RD-8041-1 is a linear MR damper with continuous variable damping determined by the yield strength of the MR fluid in response to a magnetic field. The damper responds in less than 15 milliseconds to changes in the magnetic field and can operate at 1 A continuously or 2 A intermittently at 12 V DC. The RD-8041-1 is a monotube shock containing high pressure nitrogen gas (300 psi) which fully extends the piston under no load. At ambient temperatures the resistance of the coil is 5 Ω and at 71° C the resistance increases to 7 Ω. Extreme temperature changes can drastically alter the performance of the MR damper [32].

The Ultramotion linear actuator used to drive the MR damper is a standard servo cylinder with an acme screw to prevent back-drive and operates at 180 W. The actuator is capable of 445 N of continuous force and 1001 N at its peak with a maximum speed of 178 mm/s. There are several onboard sensors to measure states such as position, torque, temperature, and humidity. The position of the linear actuator is measured using the phase index absolute position sensor. This sensor is a multi-turn magnetic encoder with a resolution of 1024 counts per revolution used for absolute position feedback and commutation. The measurement noise covariance of the sensor is discussed in subsequent sections. The torque feedback is calculated using closed loop current feedback on each motor phase. This is then translated into actuator output force. Since using current feedback is not an accurate method of calculating output force this results in significant error and noise.

In general, there is a direct relationship between motor torque and actuator output force. However, there are some complicating factors that can significantly impact this relationship. Rotational inertial loads, lubricant viscosity, and seal friction can all contribute to output force variability. Factory test data was used in order to convert motor torque into actuator output force. The data is collected on each actuator during the acceptance test procedure (ATP) before leaving the factory [33]. The current-force curves that are generated are unique to each actuator. However, there is still significant noise in force output. In order to reduce some of the noise in the torque sensor, a first order Butterworth filter was applied with a cutoff frequency between 0 and 0.05 of the Nyquist rate.

The RAS1-500S-S is a resistive S-Beam load cell capable of measuring both compressive and tensile force measurement. The load cell is made from tool steal and has a capacity of 2224 N and a sample rate of 1000 Hz. The calibration measurement equipment is traceable to NIST via Pacific Calibration Services. This sensor was used to test the efficacy of applying adaptive filtering strategies on the current feedback of the linear actuator. While the noise covariance of the loadcell is 26.535 N, the noise covariance of the Ultramotion motor torque sensor is 622.407 N. The comparatively high noise distribution of the onboard Ultramotion motor torque sensor makes it a suitable candidate for applying adaptive filtering strategies.

Force-velocity hysteresis curves have been modeled extensively by [33] and [34]. However, at low velocities over long stroke lengths, the force of the diaphragm and compressed nitrogen gas is not negligible. Thus, a force-position hysteresis curve was modeled by driving the MR damper at a constant velocity over one full stroke. For the MR model used in this paper, the actuator speed was set to 30 mm/s and the damping force was recorded by the loadcell over a stroke length of 57 mm. Approximately 200 strokes (extension and retraction) were used to model the behavior of the behavior at each operational mode (normal, overcurrent, undercurrent). The conditions of the operational modes are discussed below.

There are several different types of faults that can be experienced during MR damper operations. The viscosity of the MR fluid is sensitive to extreme temperatures [31] and the particles in the MR fluid are also subject to degradation over time [35]. However, this paper focuses on faults caused by minor temperature changes or faulty power supplies which alters the current supplied to the MR damper. Undercurrent and overcurrent fault modes were modeled in addition to the normal operating current. The undercurrent, normal, and overcurrent operational modes are denoted by a supply current of 0 mA, 120 mA, and 220 mA, respectively.

A sample of experimental data used to model the MR damper can be seen in Fig 4. The figures show the actuator extending and retracting at a constant speed with MR force being recorded by the loadcell and actuator current sensor. The figures also show the application of a first order Butterworth filter on the actuator current sensor in order to reduce some of the noise before apply adaptive filtering strategies.

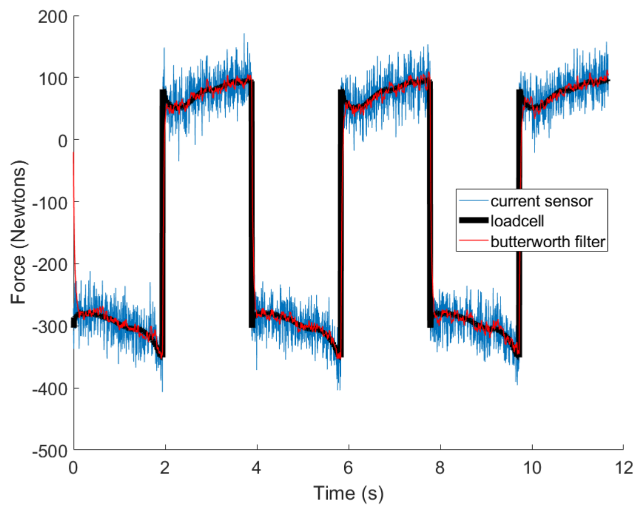

Figure 4. Sample of experimental data used to model the MR damper under normal operating conditions.

The discretized state space equations can be written as follows:

$$x_{1,k+1} = x_{1,k} + T \cdot x_{2,k} \tag{4.1.1}$$

$$x_{2,k+1} = x_{2,k} \tag{4.1.2}$$

$$x_{3,k+1} = \begin{cases} \sum_{k=0}^{9} a_{uk} x_{1,k} \; ; \; x_{2,k} < 0 \\ \sum_{k=0}^{9} a_{dk} x_{1,k} \; ; \; x_{2\ k} > 0 \\ \sum_{k=0}^{9} \frac{1}{2}(a_{uk} + a_{dk}) x_{1,k} \; ; \; x_{2,k} = 0 \end{cases} \tag{4.1.3}$$

where x_1, x_2, x_3, are the position, velocity, and force of MR damper and T is the sampling rate. The system and measurement noise covariance matrices are defined respectively as follows, based on factory testing:

$$Q = R \cdot 10^{-1} \tag{4.1.4}$$

$$R = \begin{bmatrix} 5.5134 \cdot 10^{-4} & 0 & 0 \\ 0 & 7.797 \cdot 10^{-4} & 0 \\ 0 & 0 & 622.407 \end{bmatrix} \tag{4.1.5}$$

The system noise was not measured directly but was assumed to be one magnitude smaller than the measurement noise.

4.2 Experimental Results and Discussion

The linear actuator drove the MR damper for a total of 11.62 seconds with constant velocity (30 mm/s) during extension and retraction. The position and velocity profile captured by the actuator encoder can be seen in Fig. 9. The initial current of 120 mA was applied to MR damper which represents normal operation. The MR damper was allowed to fully extend and retract before an overcurrent fault (220 mA) was introduced at 3.86 seconds. After another full period of motion, an undercurrent fault (0 mA) was introduced to the MR damper at 7.73 seconds before completing a final extension and retraction.

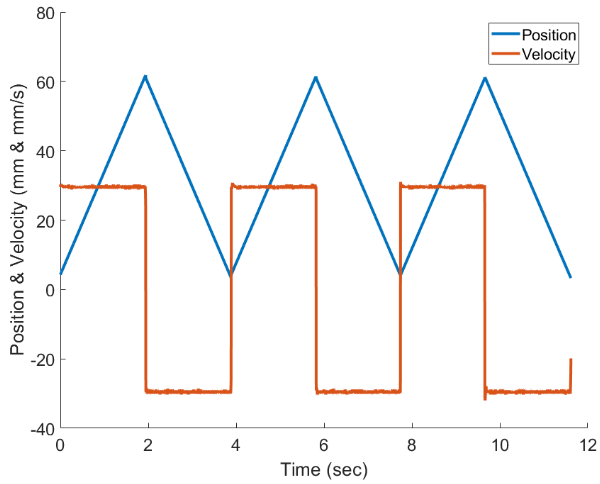

Figure 9. Sample of experimental data used to model the MR damper under normal operating conditions.

The fixed boundary layer applied in the ESIF was tuned based on minimizing the force state estimation error. The smoothing boundary layer widths are given by the following:

$$\delta = \begin{bmatrix} 5.5134 \cdot 10^{-4} & 0 & 0 \\ 0 & 7.797 \cdot 10^{-4} & 0 \\ 0 & 0 & 80 \end{bmatrix} \tag{4.2.1}$$

For all estimation strategies, the initial conditions were set to the following:

$$\hat{x}_0 = [4.2788 \quad 30.2792 \quad -303.0187]^T \tag{4.2.2}$$

$$P_{0|0} = 10 * Q \tag{4.2.3}$$

For the experiments conducted in this paper, it is assumed that the MR damper operates normally 65% of the time and has an equal likelihood of experiencing an undercurrent or overcurrent fault. The initial mode probability $\mu_{i,0}$ is given as follows:

$$\mu_{i,0} = [0.65 \quad 0.175 \quad 0.175]^T \tag{4.2.4}$$

Based on experimental procedures, the mode transition matrix $p_{i,j}$ is defined by a 3 by 3 diagonal matrix with 0.65 on the diagonal and 0.175 on the off-diagonal. This transition matrix signifies that there is a 65% probably that the system will remain in the current mode. For example, if the system is experiencing normal operation, there is a 65% chance the system will continue to undergo normal operation in the next time step.

As described previously, the experiment consisted of a test in which all three modes (normal, overcurrent, undercurrent) were experienced. After 1 period of actuation in a certain mode, the system transitioned to a different one until all modes were introduced. Fig. 10 shows the results of the IMM-EKF, IMM-UKF, and IMM-ESIF for estimating the force exerted by the MR damper during testing.

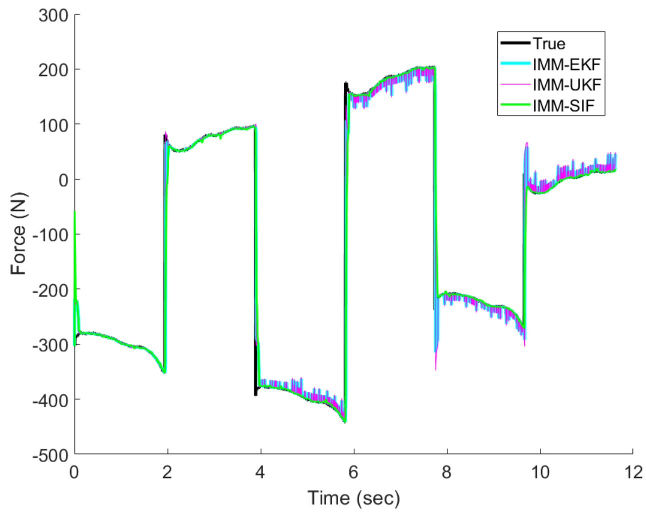

Figure 10. Force estimation of the MR damper undergoing mixed operation with normal, overcurrent, and undercurrent modes.

The RMSE (root mean squared error) for each estimator was calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \tag{4.2.5}$$

where n is the number of steps. The values shown in Tables 1 and 2 are the average RMSE of the 20 separate trials similar to the one shown in Fig. 10. The order in which the modes were experienced were randomized for each trial.

The IMM-EKF, IMM-UKF, and IMM-ESIF perform comparatively well when the MR is normal operation. As shown in Table 1, the IMM-ESIF performs slightly better than the IMM-EKF and IMM-UKF under normal operation. However, the benefit of the increased robustness demonstrated by Table 2 which shows the RMSE for mixed operation. In the presence of faults and modeling uncertainty, the IMM-ESIF shows a clear advantage over its counterparts. There is an 83.7% improvement over the IMM-EKF and 89.4% improvement over the IMM-UKF. It is interesting to note that while the UKF generally performs better than the EKF for highly nonlinear system, the EKF outperformed the UKF during mixed operation.

Table 1. RMSE for normal operation.

Estimation Strategy	RMSE (Newtons)
IMM-EKF	2.37
IMM-UKF	2.36
IMM-ESIF	1.97

Table 2. RMSE for mixed operation.

Estimation Strategy	RMSE (Newtons)
IMM-EKF	17.52
IMM-UKF	19.20
IMM-ESIF	2.04

The IMM-EKF, IMM-UKF, and IMM-ESIF were all able to properly detect the mode probabilities with varying degrees of confidence. Figures 11-13 show the mode probabilities calculated by each estimation strategy. In order to clear depict the mode probabilities, the overall trends are shown as solid lines while spikes in the mode probability are represented as dots. The mode probabilities show that the IMM-ESIF misclassifies the correct mode when the velocity of the MR damper changes direction. However, the overall classification accuracy of the IMM-ESIF is higher than its counterparts. A value “1” for a mode probability refers to a 100% confidence that the system is experiencing that mode while a “0” refers to a probability of 0%. Tables 3-5 show confusion matrices for each estimator which are commonly used in fault detection and diagnosis. The vertical axis typically represents the predicted mode while the horizontal axis represents the actual mode being experienced by the MR damper. For Tables 3-5 mode 1 represents normal operation, mode 2 represents an overcurrent fault, and mode 3 represents an undercurrent fault.

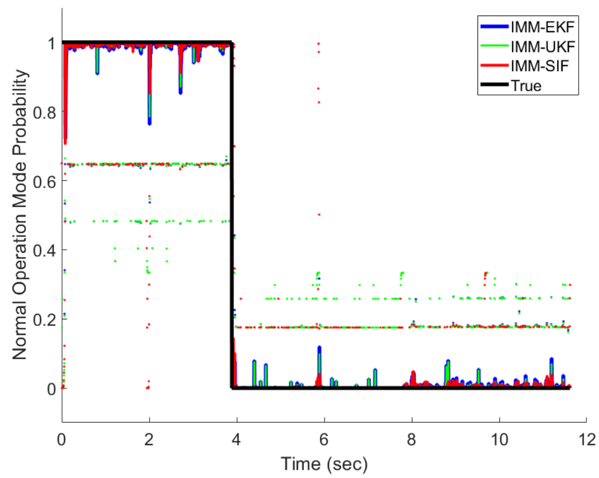

Figure 11. Normal operation mode probability.

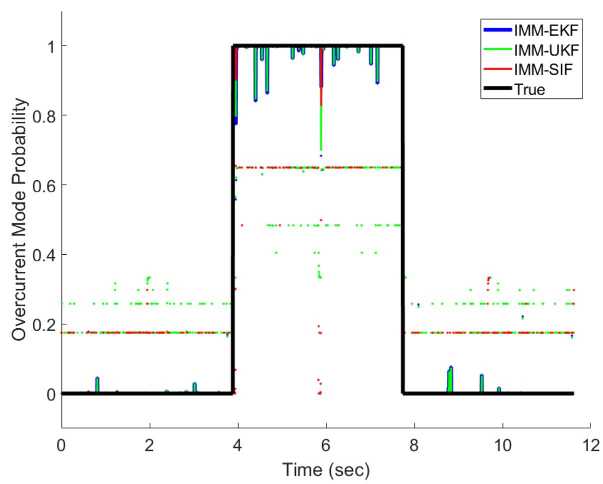

Figure 12. Overcurrent fault mode probability.

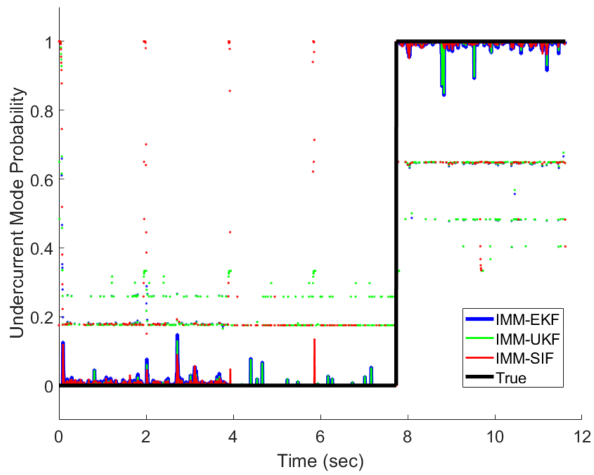

Figure 13. Undercurrent fault mode probability.

Table 3. IMM-EKF confusion matrix.

Predicted \ Actual	1	2	3
1	88.75 %	4.86 %	5.62 %
2	4.49 %	90.58 %	5.08 %
3	6.76 %	4.55 %	89.30 %

Table 4. IMM-UKF confusion matrix.

Predicted \ Actual	1	2	3
1	88.99 %	4.84 %	5.43 %
2	4.48 %	90.61 %	5.07 %
3	6.53 %	4.55 %	89.50 %

Table 5. IMM-ESIF confusion matrix.

Predicted \ Actual	1	2	3
1	93.78 %	2.66 %	2.09 %
2	1.26 %	94.58 %	1.55 %
3	4.97 %	2.76 %	96.36 %

The confusion matrices show that the IMM-EKF, IMM-UKF, and IMM-ESIF were all able to predict the correct mode of operation with relatively high confidence. In general, the classification accuracy of normal operation was the lowest. This is because the damping force of normal operation falls between the overcurrent and undercurrent modes as shown in Fig. 8. Likewise, the classification overcurrent fault had the highest accuracy because it has greater separation from the normal operation than the undercurrent fault. The IMM-UKF had slightly higher classification accuracy than the IMM-EKF. However, the IMM-ESIF shows a 4-5% higher accuracy when classifying the correct mode when compared to the IMM-EKF and IMM-UKF. Overall, the IMM-ESIF showed significant improvement in both estimation accuracy (RMSE) and classification (confusion matrix) when compared to the IMM-EKF and IMM-UKF.

5. CONCLUSIONS

In this paper, a relatively new estimation method called the sliding innovation filter (SIF) is combined with the interacting multiple model (IMM) method. The corresponding method, referred to as the SIF-IMM (or ESIF-IMM for nonlinear systems and measurements), is applied on a magnetorheological actuator which was built for experimentation. These types of actuators are similar to hydraulic-based ones, which are commonly found in aerospace systems. The method is shown to accurately identify faults in the system. The results are compared and discussed with other popular nonlinear estimation strategies including the extended and unscented Kalman filters. Future work will study data-driven approaches such as those based on machine learning. A comprehensive study will follow comparing model-based and data-driven fault detection and identification strategies.

REFERENCES

[1] S. A. Gadsden, Y. Song, and S. R. Habibi, "Novel model-based estimators for the purposes of fault detection and diagnosis," *IEEE/ASME Transactions on Mechatronics*, vol. 18, no. 4, pp. 1237–1249, 2013, doi: 10.1109/TMECH.2013.2253616.

[2] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: Wiley, 2001. doi: 10.1002/0471221279.

[3] G. A. Ackerson and K. S. Fu, "On State Estimation in Switching Environments," *IEEE Trans Automat Contr*, vol. AC-15, no. 1, pp. 10–17, 1970, doi: 10.1109/TAC.1970.1099359.

[4] A. G. Jaffer and S. C. Gupta, "Recursive Bayesian estimation with uncertain observation," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 5, pp. 614–616.

[5] A. G. Jaffer and S. C. Gupta, "Optimal sequential estimation of discrete processes with Markov interrupted observations," *IEEE Trans. Autom. Control*, vol. AC-16, no. 5, pp. 471–475.

[6] C. B. Chang and M. Athans, "State estimation for discrete systems with switching parameters," *IEEE Trans. Aerosp. Electron. Syst*, vol. AES-14, no. 5, pp. 418–425, doi: 10.5711/1082598316323.

[7] H. P. Blom, "An efficient filter for abruptly changing systems," in *An efficient filter for abruptly changing systems*, Las Vegas, NV, USA: 23rd IEEE Conference on Decision and Control, 1984, pp. 656–658. doi: 10.1109/cdc.1984.272089.

[8] H. A. P. Blom and Y. Bar-Shalom, "The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients," *IEEE Trans Automat Contr*, vol. 33, no. 8, pp. 780–783, 1988, doi: 10.1109/9.1299.

[9] H. H. Afshari, S. A. Gadsden, and S. Habibi, "Gaussian filters for parameter and state estimation: A general review of theory and recent trends," *Signal Processing*, vol. 135, no. March 2016, pp. 218–238, 2017, doi: 10.1016/j.sigpro.2017.01.001.

[10] B. Ristic, S. Arulampalam, and N. Gordon, "Beyond the Kalman filter: Particle filters for tracking applications," vol. 685. Artech House, Boston, 2004.

[11] S. Haykin, *Kalman Filtering and Neural Networks*. New York: John Wiley and Sons Inc, 2001. doi: 10.1002/0471221546.

[12] S. A. Gadsden, S. Habibi, and T. Kirubarajan, "Kalman and smooth variable structure filters for robust estimation," *IEEE Trans Aerosp Electron Syst*, vol. 50, no. 2, pp. 1038–1050, 2014, doi: 10.1109/TAES.2014.110768.

[13] S. K. Spurgeon, "Sliding mode observers: A survey," *Int J Syst Sci*, vol. 39, no. 8, pp. 751–764, 2008, doi: 10.1080/00207720701847638.

[14] J. J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ: Prentice-Hall, 2001.
[15] S. A. Gadsden, M. el Sayed, and S. R. Habibi, "Derivation of an optimal boundary layer width for the smooth variable structure filter," in *Proceedings of the American Control Conference*, Piscataway, NJ, USA: 2011 American Control Conference, 2011, pp. 4922–4927. doi: 10.1109/acc.2011.5990970.
[16] M. Al-Shabi, S. A. Gadsden, and S. R. Habibi, "Kalman filtering strategies utilizing the chattering effects of the smooth variable structure filter," *Signal Processing*, vol. 93, no. 2, pp. 420–431, 2013, doi: 10.1016/j.sigpro.2012.07.036.
[17] M. Alshabi and A. Elnady, "Recursive Smooth Variable Structure Filter for Estimation Processes in Direct Power Control Scheme under Balanced and Unbalanced Power Grid," *IEEE Syst J*, vol. 14, no. 1, pp. 971–982, 2020, doi: 10.1109/JSYST.2019.2919792.
[18] S. Andrew Gadsden and M. Al-Shabi, "The Sliding Innovation Filter," *IEEE Access*, vol. 8, pp. 96129–96138, 2020, doi: 10.1109/ACCESS.2020.2995345.
[19] W. Hilal, S. A. Gadsden, S. A. Wilkerson, and M. A. Al-Shabi, "A square-root formulation of the sliding innovation filter for target tracking," 2022. doi: 10.1117/12.2618965.
[20] W. Hilal, S. A. Gadsden, S. A. Wilkerson, and M. A. Al-Shabi, "Combined particle and smooth innovation filtering for nonlinear estimation," 2022. doi: 10.1117/12.2618973.
[21] J. Goodman, W. Hilal, S. A. Gadsden, and C. D. Eggleton, "Adaptive SVSF-KF estimation strategies based on the normalized innovation square metric and IMM strategy," *Results in Engineering*, vol. 16, 2022, doi: 10.1016/j.rineng.2022.100785.
[22] R. van der Merwe and E. A. Wan, "Sigma-point Kalman filters for integrated navigation," in *Proceedings of the Annual Meeting - Institute of Navigation*, Proceedings of the Institute of Navigation Annual Meeting, 2004, pp. 641–654.
[23] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," *In Pract*, vol. 7, no. 1, pp. 1–16, 2006, doi: 10.1.1.117.6808.
[24] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 12, p. 1958, 2004, doi: 10.1109/JPROC.2004.837637.
[25] Grewal M.S. and A. A.P., *Kalman Filtering, Theory and Practice Using MATLAB*. New York: John Wiley and Sons, Inc., 2001.
[26] J. T. Ambadan and Y. Tang, "Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems," *J Atmos Sci*, vol. 66, no. 2, pp. 261–285, 2009, doi: 10.1175/2008JAS2681.1.
[27] X. Tang, X. Zhao, and X. Zhang, "The square-root spherical simplex unscented Kalman filter for state and parameter estimation," *International Conference on Signal Processing Proceedings, ICSP*, pp. 260–263, 2008, doi: 10.1109/ICOSP.2008.4697120.
[28] A. Kolmogorov, V. Tikhomirov, and Ed. Dordrecht, "Interpolation and Extrapolation of Stationary Random Sequences," *Selected Works of A. N. Kolmogorov*. Kluwer Academic Publishers, he Netherlands, pp. 272–280, 1992. doi: 10.1007/978-94-011-2260-3_28.
[29] S. J. Julier and J. K. Uhlmann, "A New Method for Nonlinear Transformation of Means and Covariances in Filters and Estimator," in *Proceedings of the IEEE,* pp. 401–422. doi: 10.1109/TAC.1973.1100244.
[30] J. S. Oh, K. S. Kim, Y. S. Lee, and S. B. Choi, "Dynamic simulation of a full vehicle system featuring magnetorheological dampers with bypass holes," *J Intell Mater Syst Struct*, vol. 31, no. 2, pp. 253–262, 2020, doi: 10.1177/1045389X19876880.

[31] "LORD Magneto-Rheological," *Mid Atlantic Rubber 2019*. http://www.lordmrstore.com/home (accessed Jun. 30, 2021).
[32] M. K. Thakur and C. Sarkar, "Influence of Graphite Flakes on the Strength of Magnetorheological Fluids at High Temperature and its Rheology," *IEEE Trans Magn*, vol. 56, no. 5, 2020, doi: 10.1109/TMAG.2020.2978159.
[33] "Ultramotion Servo Cylinder," 2021. https://www.ultramotion.com/linear-actuators/ (accessed Jun. 30, 2021).
[34] X. Q. Ma, S. Rakheja, and C. Y. Su, "Development and relative assessments of models for characterizing the current dependent hysteresis properties of magnetorheological fluid dampers," *J Intell Mater Syst Struct*, vol. 18, no. 5, pp. 487–502, 2007, doi: 10.1177/1045389X06067118.
[35] S. B. Choi, S. K. Lee, and Y. P. Park, "A hysteresis model for the field-dependent damping force of a magnetorheological damper," *J Sound Vib*, vol. 245, no. 2, pp. 375–383, 2001, doi: 10.1006/jsvi.2000.3539.
[36] J. S. Kumar, P. S. Paul, G. Raghunathan, and D. G. Alex, "A review of challenges and solutions in the preparation and use of magnetorheological fluids," *International Journal of Mechanical and Materials Engineering*, vol. 14, no. 1, 2019, doi: 10.1186/s40712-019-0109-2.