Signal Processing 135 (2017) 218–238

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Review

Gaussian filters for parameter and state estimation: A general review of theory and recent trends

H.H. Afshari[a], S.A. Gadsden[b,*], S. Habibi[a]

[a] Department of Mechanical Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4L7
[b] School of Engineering, University of Guelph, 50 Stone Road East, Guelph, Ontario, Canada N1G 2W1

ARTICLE INFO

Keywords:
State and parameter estimation
Stochastic dynamic systems
Gaussian filters, Kalman filter
Smooth variable structure filter

ABSTRACT

Real-time control systems rely on reliable estimates of states and parameters in order to provide accurate and safe control of electro-mechanical systems. The task of extracting state and parametric values from system's partial measurements is referred to as state and parameter estimation. The main goal is minimizing the estimation error as well as maintaining robustness against the noise and modeling uncertainties. The development of estimation techniques spans over five centuries, and involves a large number of contributors from a variety of fields. This paper presents a tutorial on the main Gaussian filters that are used for state estimation of stochastic dynamic systems. The main concept of state estimation is firstly described based on the Bayesian paradigm and Gaussian assumption of the noise. The filters are then categorized into several groups based on their applications for state estimation. These groups involve linear optimal filtering, nonlinear filtering, adaptive filtering, and robust filtering. New advances and trends relevant to each technique are addressed and discussed in detail.

1. Introduction

Estimation is the process of extracting the value of a hidden quantity from indirect, inaccurate and uncertain measurements. The hidden quantity may be a parameter or a state variable. In general, the term *parameter* refers to a time-invariant physical quantity that may be a scalar, a vector, or a matrix. Although the term *time-varying parameter* may be appear in some texts, its variations must be slow in comparison to changes in the state variable. The term state usually refers to a vector that evolves over time by the use of an equation which describes the dynamics of a system [4]. There exist two different classes of estimators which include the parameter estimator and the state estimator. The main goal of the estimation task is to minimize the state or parameter estimation error while being robust to uncertainties and perturbations. Noise and perturbations are inherently present in the measurement process, and are caused by instruments and environmental factors. System uncertainties are usually caused by inaccuracies in modeling the process, approximations, nonlinearities, and variations in physical parameters of the system.

Major contributions to the estimation field began in the fifteen century, and included a large number of contributors from a variety of backgrounds. Thomas Bayes (1701–1761), as the first major contributor to this field, introduced the famous Bayesian rule for statistical inference that provides the basic formula for Bayesian estimation methods [1]. The pioneering study that provides an optimal estimate from noisy data was performed by Carl Friedrich Gauss (1777–1855). He invented the famous least square estimation method in 1795 and used it to solve nonlinear estimation problems in mathematical astronomy [2]. Later on, Andrei Markov (1856–1922) introduced the Markov process and Markov chain theories based on probability and statistical methods [3]. The Markov theories formulate transitions in random processes from one state to another, between a finite or countable number of possible states. He proved that the probability distribution of states may be calculated using its current distribution that contains the effects of all the past events of the system [4]. Andrei Kolmogorov (1903–1987) published his well-known book, *Foundations of the Theory of Probability*, in 1933 laying the modern axiomatic foundations of probability theory. Sydney Chapman (1888–1970) continued the research on the Markov processes. Chapman and Kolmogorov independently presented the Chapman-Kolmogorov equations used for solving basic equations in the estimation field [3].

Ronald Aylmer Fisher (1890–1962) became famous for his major contribution, the so-called Fisher information matrix. It represents a measure of the amount of information extracted from a sample of values with a given probability distribution [3]. Norbert Wiener (1894–1964) introduced the so-called Wiener filter formulation in 1949 for

* Corresponding author.
E-mail addresses: afsharh@mcmaster.ca (H.H. Afshari), gadsden@uoguelph.ca (S.A. Gadsden), habibi@mcmaster.ca (S. Habibi).

http://dx.doi.org/10.1016/j.sigpro.2017.01.001
Received 7 March 2016; Received in revised form 30 December 2016; Accepted 2 January 2017
Available online 03 January 2017
0165-1684/

signal processing applications. This filter reduces the amount of noise present in a signal in comparison with an estimation of the desired noiseless signal [5]. Kolmogorov (1903–1987), along with Wiener, made the foundation of estimation theories that were used later to develop the theory of prediction, filtering, and smoothing. His research ultimately led to the derivation of an optimal estimator, which was formulated for continuous-time systems [6]. Meanwhile, Kolmogorov independently derived an optimal linear predictor for discrete-time systems [7]. Their research would later become famous, known as the Wiener-Kolmogorov filter (WF), a predecessor to the Kalman filter [1].

In 1960, Rudolf Kalman, building on the work of others, introduced a new approach to linear filtering and prediction problems; later referred to as the Kalman filter [4]. The Kalman filter was successfully applied by NASA for the Apollo's guidance and navigation system and quickly became popular as the most practical method for state estimation [1,4,8]. The Kalman filter (KF) uses a linear dynamic model and sequential measurements of the system to provide an optimal state estimate in the presence of Gaussian noise. A continuous version of the KF was later developed by Kalman and Bucy which later became known as the Kalman-Bucy filter [9]. Some extensions to the KF formulation, such as linearization and approximation, led to the extended Kalman filter (EKF) and the unscented Kalman filter (UKF), respectively. These extensions allowed the KF to be implemented on nonlinear systems for the purpose of state and parameter estimation. Other advanced variants of the Kalman filter include the quadrature Kalman filter (QKF) [10,11], mixture Kalman filter (MKF) [12], and the cubature Kalman filter (CKF) [13]. Fig. 1 presents the lifeline of a number of main contributions to the estimation theory from the eighteenth century to present.

In the Bayesian state estimation approach, the *a posteriori* probability density function (PDF) of the states is recursively calculated based on the known *a priori* PDF and newer measurements. The calculation includes two main steps: prediction and update. In the prediction stage, the system model is used to predict state values. The predicted values of states are then refined and updated based on measurements from the system. There are three concepts that include smoothing, filtering, and prediction. Smoothing uses the measurements beyond the desired time of interest, $t_{obs} > t_{est}$ to refine the estimates further. Filtering uses measurements up to and including the time of interest, $t_{obs} \leq t_{est}$. Prediction only uses measurements prior to the time of interest and thus predicts the future the system's state, $t_{obs} < t_{est}$ [14]. A state estimation process is generally constructed based on the available knowledge of the system summarized in four items [4]: the state transition model; the measurement model; the input or its probabilistic characterization; and the prior knowledge of the system. State estimation methods are extensively used in modern engineering applications. These include target tracking, control systems, communications, biomedical engineering, and economic systems. During recent years, this field has attracted a significant amount of attention in both theory and applications [15–20]. Fig. 2 shows a block-diagram scheme of state estimation.

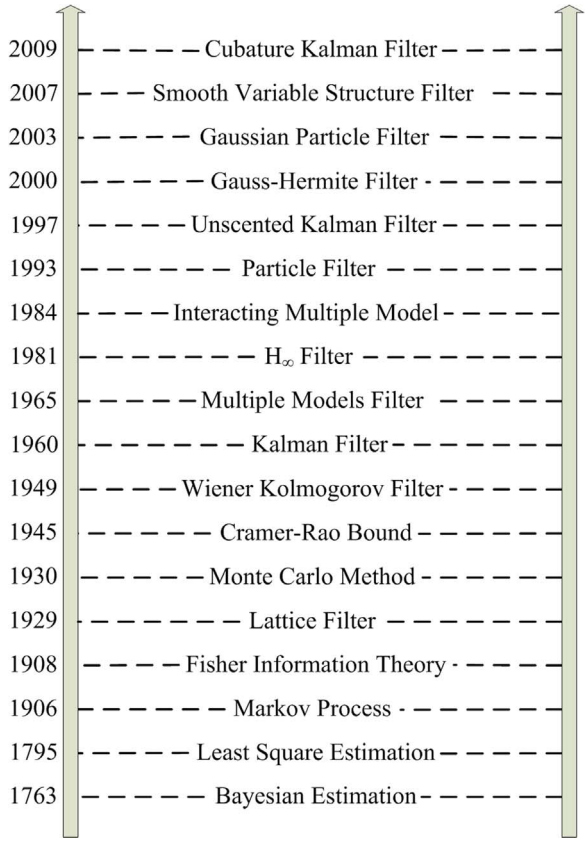


Fig. 1. The 200 year history of main contributions to the estimation theory.

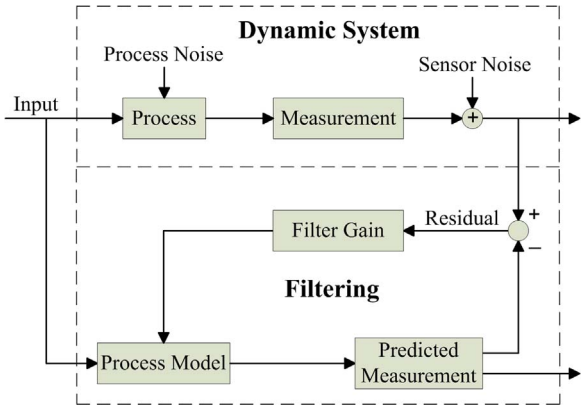


Fig. 2. Block-diagram scheme of a state estimation process.

In this paper, the main Gaussian estimators used for state estimation are reviewed, and new advances and trends are addressed. Section 2 presents the main foundation of the state estimation problem which is based on the Bayesian paradigm and Gaussian assumption. Section 3 reviews the main linear Gaussian filters, and Section 4 surveys popular nonlinear Gaussian filters. Section 5 and Section 6 respectively present adaptive filtering and robust filtering for state estimation.

2. State estimation of stochastic dynamic systems

The task of extracting state values from inaccurate, uncertain, and noisy measurements is referred to as state estimation. The main objective is to minimize the estimation error when projected to the output space. This error is referred to as the residual or innovation vector. It is important to note that due to the presence of noise and uncertainties (caused by the measurement process, instrumentation, and environment), the measurements cannot reflect exact values of the state variables. In order to construct a framework for the state estimation of stochastic dynamic systems, one may assume a first-order Markov process that is modeled as follows:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k), \tag{1}$$

$$\mathbf{z}_{k+1} = \mathbf{h}(\mathbf{x}_k, \mathbf{v}_k), \tag{2}$$

where $\mathbf{x}_k$,$\mathbf{u}_k$, and $\mathbf{z}_k$ are the state, input, and measurement vectors, respectively. Moreover, $\mathbf{w}_k$ and $\mathbf{v}_k$ are the process noise and measurement noise at time step k, respectively. It is assumed that $\mathbf{f}$, $\mathbf{h}$, and $\mathbf{u}_k$ are known, when $\mathbf{w}_k$ and $\mathbf{v}_k$ are mutually independent white stochastic processes. The filtering problem is formulated by recursively calculating an estimate of the state vector $\mathbf{x}_k$. This is achieved by constructing a Bayesian paradigm based on the sequence of measurements $\mathbf{Z}_k$ up to

time k. There are two main concepts in statistics that help to computationally simplify the process of state estimation. They are the Bayesian paradigm and the Gaussian distribution of states, which will be explained in the subsequent subsections.

2.1. Bayesian paradigm for state estimation

The main purpose of using a Bayesian paradigm in state estimation is to calculate the conditional *a posteriori* state PDF $p(\mathbf{x}_{k+1}|\mathbf{Z}_{k+1})$, where $\mathbf{Z}_{k+1} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k\}$ is the vector of noisy measurements. In order to formulate the state's *a posterior* PDF, a two stage recursive algorithm can be used, when the state *a priori* PDF $p(\mathbf{x}_k|\mathbf{z}_k)$is available. It is assumed that the initial PDF of the state is $p(\mathbf{x}_0) = p(\mathbf{x}_0|\mathbf{z}_0)$. The filtering process contains two stages that include prediction and update. The Chapman-Kolmogorov equation can be used to perform the prediction stage using the system model of (1), as follows [21]:

$$p(\mathbf{x}_{k+1}|\mathbf{Z}_k) = \int p(\mathbf{x}_{k+1}|\mathbf{x}_k)\ p(\mathbf{x}_k|\mathbf{Z}_k)\ d\mathbf{x}_k \tag{3}$$

where the state transition probability $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ is obtained from the state Eq. (1). Then after, the Bayesian rule is used to provide the basis for the update stage given by [21]:

$$p(\mathbf{x}_{k+1}|\mathbf{Z}_{k+1}) = \frac{p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})p(\mathbf{x}_{k+1}|\mathbf{Z}_k)}{p(\mathbf{z}_{k+1}|\mathbf{Z}_k)}, \tag{4}$$

where $p(\mathbf{z}_{k+1}|\mathbf{Z}_k)$ is the normalizing constant, and is obtained by [21]:

$$p(\mathbf{z}_{k+1}|\mathbf{Z}_k) = \int p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})\ p(\mathbf{x}_{k+1}|\mathbf{Z}_k)\ d\mathbf{x}_k, \tag{5}$$

This value depends on the likelihood function $p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})$ that is obtained from the measurement Eq. (2). From the *a posteriori* PDF, a theoretically optimal state estimate may be computed using an approach such as the minimum mean square error (MMSE), as follows [21]:

$$\hat{\mathbf{x}}^{MMSE}_{k+1|k+1} \triangleq \int \mathbf{x}_{k+1} p(\mathbf{x}_{k+1}|\mathbf{Z}_{k+1})\ d\mathbf{x}_{k+1}, \tag{6}$$

Alternatively, the maximum *a posteriori* (MAP) method may be used, as follows [21]:

$$\hat{\mathbf{x}}^{MAP}_{k|k} \triangleq \arg\ \max_{\mathbf{x}_k} p(\mathbf{x}_{k+1}|\mathbf{Z}_{k+1}), \tag{7}$$

The above calculations are based on two assumptions: 1- the state transitions follow a first order Markov process, i.e., $p(\mathbf{x}_{k+1}|\mathbf{X}_k, \mathbf{Z}_{k+1}) = p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ where $\mathbf{X}_k = \{\mathbf{x}_0, \dots, \mathbf{x}_k\}$; 2- the measurements are conditionally independent given the states, i.e., $p(\mathbf{z}_{k+1}|\mathbf{X}_{k+1}, \mathbf{Z}_k) = p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})$ [22].

The main purpose of filtering is to construct an accurate posterior PDF of the state based on all available information. Eqs. (3)–(5) provide the basis for recursive estimation schemes; with emphasis that they present only a conceptual solution, which in some scenarios cannot be calculated analytically. It is possible to solve the recursive equation of the estimated posteriori PDF analytically for the estimation problem with a linear state transition and measurement model, subjected to additive noise and uncertainties with Gaussian PDF. As a statistical point of view, in linear systems with Gaussian uncertainties, $p(\mathbf{x}_k|\mathbf{Z}_k)$ contains all statistical information about $\mathbf{x}_k$. In this way, it is expected to convert the estimation problem to the point estimation in which the mode, mean, or median are estimated. In such cases, the *a posteriori* PDF can be expressed with simply the mean and covariance terms; the *a posteriori* mean and covariance can be predicted and updated recursively. However, this approach is not applicable to nonlinear systems or systems with non-Gaussian uncertainties. Fig. 3 compares the main concept of point estimation for systems with Gaussian and non-Gaussian uncertainties respectively. For systems with Gaussian distributions, the mode, mean and median are the same. The most popular method used to solve the linear estimation problem when subjected to the white Gaussian noise is the Kalman filter (KF) [8, 21].

2.2. Gaussian assumption for the Bayesian estimation paradigm

In order to simplify complex equations of the Bayesian filtering paradigm, Gaussian distributions for the noise and uncertainties are assumed. This assumption provides a Gaussian distribution for the state *a priori* PDF $p(\mathbf{x}_{k+1}|\mathbf{Z}_k)$ and the filter likelihood density $p(\mathbf{z}_{k+1}|\mathbf{Z}_k)$ which alternatively results in a Gaussian distribution for the state *a posteriori* PDF $p(\mathbf{x}_{k+1}|\mathbf{Z}_{k+1})$. In this context, a class of Bayesian filters is formulated under the Gaussian assumption and is referred to as the Gaussian filters. Following this formulation, recursive computations of the former Bayesian filter convert to recursive algebraic computations of the first moment (mean) and the second moment (covariance) of existing conditional PDFs. This procedure is followed for both time and measurement updates, which follow [13]:

A. Time Update [13]:

In this step, the state's *a priori* mean $\hat{\mathbf{x}}_{k+1|k}$ and the state estimation error's *a priori* covariance $\mathbf{P}_{k+1|k}$of the Gaussian distribution are calculated using the expectation operator as follows [13]:

$$\hat{\mathbf{x}}_{k+1|k} = E\{f(\mathbf{x}_k, \mathbf{u}_k)|\mathbf{Z}_k\} = \int_{\mathbb{R}^{n_x}} f(\mathbf{x}_k, \mathbf{u}_k) \times N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) d\mathbf{x}_k, \tag{8}$$

$$\begin{aligned}\mathbf{P}_{k+1|k} &= E\{(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T|\mathbf{Z}_k\} \\ &= \int_{\mathbb{R}^{n_x}} f(\mathbf{x}_k, \mathbf{u}_k) \times f^T(\mathbf{x}_k, \mathbf{u}_k) \times N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) \\ &\quad d\mathbf{x}_k - \hat{\mathbf{x}}_{k+1|k}\hat{\mathbf{x}}^T{}_{k+1|k} + \mathbf{Q}_k.\end{aligned} \tag{9}$$

where $N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$ denotes the Gaussian density function.

B. **Measurement Update** [13]:

Since the error in the *a priori* measurement is a zero-mean white process [23], it is possible to approximate the error to be Gaussian and restate the filter likelihood density as follows [13]:

$$p(\mathbf{z}_{k+1}|\mathbf{Z}_k) = N(\mathbf{z}_{k+1}; \hat{\mathbf{z}}_{k+1|k}, \mathbf{P}_{zz,k+1|k}), \tag{10}$$

where the *a priori* measurement is given by [13]:

$$\hat{\mathbf{z}}_{k+1|k} = \int_{\mathbb{R}^{n_x}} h(\mathbf{x}_{k+1}, \mathbf{u}_{k+1}) \times N(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k}) dx_k. \tag{11}$$

and the *a priori* covariance and cross-covariance are respectively calculated as follows [13]:

$$\begin{aligned}\mathbf{P}_{zz,k+1|k} &= \int_{\mathbb{R}^{n_x}} h(\mathbf{x}_{k+1}, \mathbf{u}_{k+1}) h^T(\mathbf{x}_{k+1}, \mathbf{u}_{k+1}) \times N(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k}) \\ &\quad d\mathbf{x}_k - \hat{\mathbf{z}}_{k+1|k}\hat{\mathbf{z}}^T{}_{k+1|k} + \mathbf{R}_{k+1}.\end{aligned} \tag{12}$$

$$\begin{aligned}\mathbf{P}_{xz,k+1|k} &= \int_{\mathbb{R}^{n_x}} \mathbf{x}_{k+1} h^T(\mathbf{x}_{k+1}, \mathbf{u}_{k+1}) \times N(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k}) \\ &\quad d\mathbf{x}_k - \hat{\mathbf{x}}_{k+1|k}\hat{\mathbf{z}}^T{}_{k+1|k}.\end{aligned} \tag{13}$$

The Gaussian filter calculates the state *a posteriori* PDF using the new measurement $\mathbf{z}_{k+1}$, as follows [13]:

$$p(\mathbf{x}_{k+1}|\mathbf{Z}_{k+1}) = N(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|k+1}, \mathbf{P}_{k+1|k+1}), \tag{14}$$

and hence, the *a posteriori* state and error covariance may be calculated by [13]:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}(\mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1|k}), \mathbf{P}_{k+1|k+1} \\ &= \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1}\mathbf{P}_{zz,k+1|k}\mathbf{K}^T_{k+1}, \mathbf{K}_{k+1} = \mathbf{P}_{xz,k+1|k}\mathbf{P}_{zz,k+1|k}.\end{aligned} \tag{15}$$

Note that for the case with linear state and measurement functions

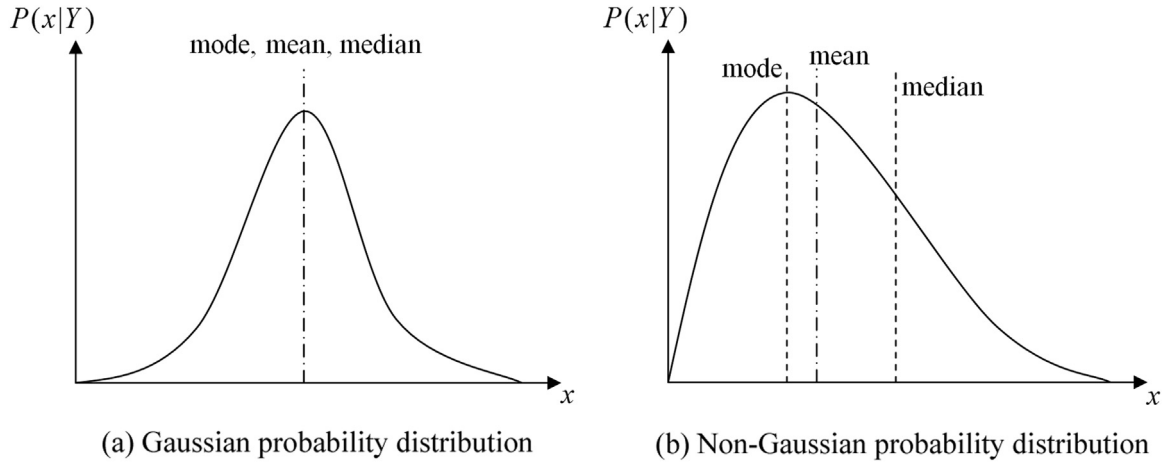


Fig. 3. Effects of the probability distribution on point state estimation [22].

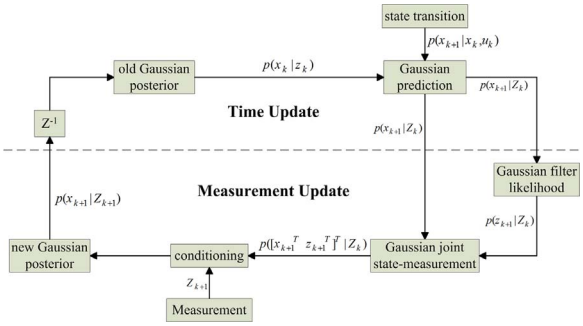


Fig. 4. A block-diagram scheme of a one-cycle Gaussian filtering [13].

subjected to an additive zero-mean white Gaussian noise, the above formulation reduces to the Kalman filter. However, the main basis of the Gaussian filter is concentrated on how to calculate the Gaussian weighted integrals that are all formulated as nonlinear functions with Gaussian densities [13]. Fig. 4 presents a block-diagram concept of a one-cycle Gaussian filtering process.

In the case of nonlinear systems with non-Gaussian noise and disturbances, however, it is impossible to obtain an exact analytical solution. Techniques such as linearization or PDF approximation may be considered to solve the estimation problem. The extended Kalman filter (EKF) technique is the most common Gaussian method for solving recursive nonlinear estimation problems through linearization [2,3,14]. The unscented Kalman filter (UKF) or the Cubature Kalman filter (CKF) are extensions to the Kalman filter. They use a transformation to approximate a posterior Gaussian distribution by capturing its mean and covariance accurately to the second order. The corresponding approximation error will be in the third order or higher [2,3,14]. In the past decade, due to increased computational power, the particle filter (PF) has attracted considerable interest as a powerful tool for solving nonlinear estimation problems with a non-Gaussian noise distribution. The PF technique uses a random set of weighted particles that approximate nonlinear characteristics or distributions in the state *a posteriori* PDF.

The Kalman-type filtering methods are primarily designed based on the assumption that the system model is known and that noise is white. In real applications, there may be considerable uncertainties about the model structure, physical parameters, level and distribution of noise, and initial conditions. In some situations, the system dynamic is too complex to be modeled exactly, or *a priori* knowledge is not available about parameters such as noise levels or distributions. In other situations, the system structure or parameters may change by time unpredictably. In order to overcome such potential difficulties, there are two approaches in state estimation, when the Kalman-type filtering methods diverge or present unacceptable performance. These two approaches are referred to as the robust state estimation and the adaptive state estimation. The main objective of robust estimation is designing a fixed filter that presents an acceptable performance for a wide range of modeling uncertainties [81]. The main robust state estimation methods found in the literature are the robust Kalman filter (or H_2), the H_∞ filter and the variable structure filter (VSF). Otherwise, the adaptive estimation approach is primarily used to estimate both the unknown state and the unknown noise parameters, when in some cases they may considerably change over time. There are two main approaches for adaptive estimation that include the adaptive filter with gain adaptation approach and the multiple models (MM) approach. In the first approach, the filter gain and parameters are adjusted based on statistical characteristics of noise and uncertainties. In the MM approach, several models of the system, each representing a particular operating regime, are stored and used for state estimation. The final state and covariance estimates are then calculated through a weighted summation of each filter output. Fig. 5 shows a general classification of main Gaussian filters used for state estimation.

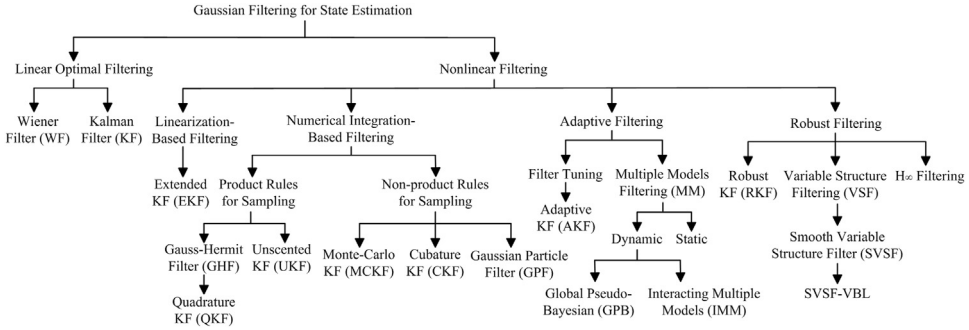


Fig. 5. A general classification of main Gaussian filters for state estimation.

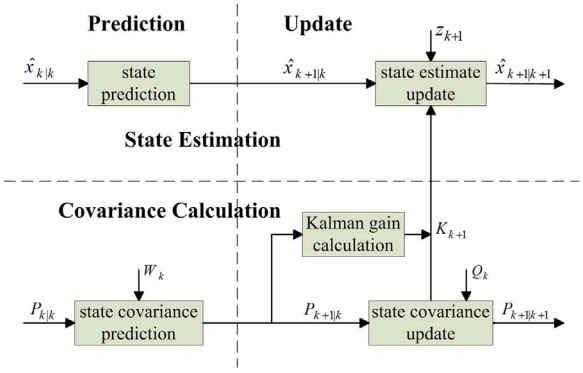


Fig. 6. Block-diagram scheme of a one cycle Kalman filter [3].

3. Gaussian filters for linear systems

Gaussian filters may be used to estimate states of systems with linear or nonlinear state transition models. For linear systems, there are two main approaches including optimal filtering and robust filtering. In the linear optimal filtering, the main purpose is minimizing the estimation error. The optimal filtering for linear Gaussian systems leads to the Wiener-Kolmogorov filter and its extension, the well-known Kalman filter (KF). Optimal state estimation is the task of extracting state values from system measurements by minimizing the MSE. The Wiener-Kolmogorov filter is the first contribution into the optimal filtering field and is only applicable to stationary signals. The Kalman filter is an extension of the Wiener-Kolmogorov filter and is applied to linear systems with non-stationary Gaussian signals.

3.1. The Kalman filter (KF)

Rudolf Kalman introduced a new approach to the linear estimation and prediction problem that later became famous as the Kalman filter (KF) [24]. It is an optimal recursive Bayesian filter restricted to the class of linear Gaussian estimation problems. The KF is a generalization of the WF and by using a state transition model, adapts itself to non-stationary signals. It was successfully utilized by NASA in the Lunar and Apollo missions. The KF requires a dynamic model of the system, known control inputs, and measurements containing white noise. Under these strict assumptions, it provides optimal state estimates by recursively predicting the states, estimating the uncertainty of the predicted states, computing a weighted average of the predicted and measured values, and refining the predicted states. There has been a significant amount of research on the KF theory as applied to engineering applications.

A one cycle KF has two main stages: prediction, and update. The prediction step uses the state estimate from the previous time step to produce an estimate at the current time step. This predicted state estimate is also known as the *a priori* state estimate. In the update stage, the current *a priori* prediction is combined with current observation for refining the state estimate into the *a posteriori* state estimate. To formulate the KF process, assume the linearized form of system equations of (1) and (2), given by [4]:

$$\mathbf{x}_{k+1} = \mathbf{F}_k\mathbf{x}_k + \mathbf{G}_k\mathbf{u}_k + \mathbf{w}_k, \tag{16}$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k, \tag{17}$$

The KF process for estimating the system with the above equations may be summarized as [4]:

1. Prediction Step:

- Calculation of the predicted (*a priori*) state and covariance estimates [4]:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{F}_k\hat{\mathbf{x}}_{k|k} + \mathbf{G}_k\mathbf{u}_k, \tag{18}$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k\mathbf{P}_{k|k}\mathbf{F}_k^T + \mathbf{Q}_k. \tag{19}$$

2. Update Step

- Calculation of the measurement error (innovation) and its covariance [4]:

$$\mathbf{e}_{\mathbf{z}_{k+1|k}} = \mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1|k}, \tag{20}$$

$$\mathbf{S}_{k+1} = \mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^T + \mathbf{R}_{k+1}. \tag{21}$$

- Calculation of the optimal Kalman gain [4]:

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^T\mathbf{S}_{k+1}^{-1}. \tag{22}$$

- Calculation of the update (*a posteriori*) state and covariance estimates [4]:

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}\mathbf{e}_{\mathbf{z}_{k+1|k}}, \tag{23}$$

$$\mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1}\mathbf{S}_{k+1}\mathbf{K}_{k+1}^T. \tag{24}$$

Note that **Q** and **R** refer to the process and measurement noise covariance matrices, respectively [4]. Fig. 6 presents a block-diagram scheme of a one cycle Kalman filtering process.

In the KF derivation process, it is assumed that the system model is known and linear, as well as the system and measurement noises being white, and the states have initial conditions with known means and variances [1,4]. However, in real engineering applications, these assumptions are not always preserved or true. The convergence of the KF is dependent on the computer precision and mathematical operations required for calculating matrix inversions [3,25]. The main aspects of the KF method are summarized as follows:

- It provides a real-time optimal recursive estimator in the MMSE sense. It produces unbiased and minimum variance estimates of system states. This illustrates that the expected value of the error between estimates and real states is zero and the expected value of the root-mean-squared of the error is minimum [26].
- It operates like an adaptive low-pass filter and its cut-off frequency is related on the ratio between the system uncertainties and measurement noise, as well as the estimate covariance [26].
- When covariance matrices are symmetric, the recursive computation of the Kalman filtering may diverge which leads to numerical instability in the estimation process. If both the process and the measurement noise covariance matrices are assumed to be very small, the covariance of the estimation error will reduce quickly and it may lead to the numerical instability [14].

3.2. Modifications to the Kalman filter

An important issue with the Kalman filter is its numerical stability. In simulations with small values of the process noise covariance **Q**, the round-off error equation may have a small positive eigenvalue. This makes the numeric form of the state covariance matrix be indefinite, in spite of its true form that is positive-definite. However, positive definite matrixes have a triangular matrix square root $\mathbf{P} = \mathbf{S}.\ \mathbf{S}^T$. Squared-form (or factored-form) derivation helps the estimation filter to guarantee numerical stability [27]. The square-root formulation of the filter is obtained by using three techniques in the linear algebra including QR decomposition, Cholesky factor updating, and efficient least squares [28]. In this context, the covariance matrix is decomposed into factored terms that are propagated forward and updated at each measurement sample time. There exist two main factored-form filters including the Potter's square-root filter and Bierman-Thornton's U-D filter [29]. The *U-D* decomposition form is obtained by $\mathbf{P} = \mathbf{U}.\ \mathbf{D}.\ \mathbf{U}^T$, where **U** is a unit

triangular matrix and **D** is a diagonal matrix. The Bierman-Thornton's U-D filter has similar accuracy to Potter's filter and has less computational cost. It is obtained by using transformation techniques that involve an upper triangle covariance factorization [29]. Grewal et al. have presented a number of different techniques to construct the *U* and *D* matrices and the application of the *U-D* decomposition [3].

Numerical stability of filtering methods may be increased by decomposing the covariance matrix into Cholesky factors, specifically when dealing with finite precision arithmetic [30]. Another way to increase the KF stability is to impose boundaries on the state estimates that are based on the prior knowledge of the system [31]. In this context, upper bounds may be defined on the level of parametric or modeling uncertainties. This provides a bound on the KF that increases stability. Formulations of the *a priori* and the *a posteriori* error covariance may be modified such that they explicitly contain effects of modeling uncertainties. For instance, one may define the *a priori* error covariance matrix $\mathbf{P}_{k+1|k}$, as follows [4]:

$$\mathbf{P}_{k+1|k} = \hat{\mathbf{F}}\mathbf{P}_{k|k}\hat{\mathbf{F}}^T + \tilde{\mathbf{F}}\mathbf{X}_k\tilde{\mathbf{F}}^T + \hat{\mathbf{F}}\mathbf{Y}_{k|k}\hat{\mathbf{F}}^T + \tilde{\mathbf{F}}\mathbf{Y}_{k|k}\tilde{\mathbf{F}}^T + \mathbf{Q}_k, \tag{25}$$

where it contains the modeling error explicitly. Matrix $\mathbf{X}_k$ denotes the mean square value matrix (or a correlation matrix, namely $E\{\mathbf{x}_k\mathbf{x}_k^T\}$), matrix $\mathbf{Y}_{k|k}$ denotes the cross term between the true states $\mathbf{x}_k$and the error $\tilde{\mathbf{x}}_k$, namely $E\{\mathbf{x}_k\tilde{\mathbf{x}}_k^T\}$. The *a posteriori* error covariance matrix may also be defined as [4]:

$$\begin{aligned}\mathbf{P}_{k+1|k+1} = {} & (\mathbf{I} - \mathbf{K}_{k+1}\hat{\mathbf{H}})\mathbf{P}_{k+1|k}(\mathbf{I} - \mathbf{K}_{k+1}\hat{\mathbf{H}})^T + \mathbf{K}_{k+1}\mathbf{R}_{k+1}{\mathbf{K}_{k+1}}^T \\ & + \mathbf{K}_{k+1}\tilde{\mathbf{H}}\mathbf{X}_{k+1}\hat{\mathbf{H}}^T{\mathbf{K}_{k+1}}^T - (\mathbf{I} - \mathbf{K}_{k+1}\hat{\mathbf{H}})\mathbf{Y}_{k+1|k}\tilde{\mathbf{H}}^T{\mathbf{K}_{k+1}}^T \\ & - \mathbf{K}_{k+1}\tilde{\mathbf{H}}{\mathbf{Y}_{k+1|k}}^T(\mathbf{I} - \mathbf{K}_{k+1}\hat{\mathbf{H}}).\end{aligned} \tag{26}$$

Another strategy for increasing the KF stability includes the addition of fictitious process noise and consideration of a fading memory to the KF formulation [32]. Using a fading memory results in neglecting measurements in the distant past and putting more emphasis on the current information. Although this modification leads to a partial loss to the optimality via the new formulation, it helps to improve the robustness and stability of the filter. In this way, the *a priori* state error covariance is restated as [32]:

$$\mathbf{P}_{k+1|k} = \alpha\ \mathbf{F}\mathbf{P}_{k+1|k}\mathbf{F}^T + \mathbf{Q}_k, \tag{27}$$

where α denotes the forgetting factor which is a positive, typically slightly larger than 1 (i.e., $\alpha = 1.01$). Its value is chosen based on how much the past measurements are desired [32].

The KF performance may be improved numerically by introducing the "Joseph form" of the *a posteriori* state error covariance matrix in the following form [3, 32]:

$$\mathbf{P}_{k+1|k+1} = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H})\mathbf{P}_{k+1|k}(\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H})^T + \mathbf{K}_{k+1}\mathbf{R}_{k+1}\mathbf{K}_{k+1}^T. \tag{28}$$

This form is proven to be more stable and robust over the former formulation presented in Eq. (24). Using the Joseph form guarantees that the *a posteriori* error covariance matrix is always symmetric positive definite at the cost of increasing the computational complexity [3, 32]. Another approach that helps to increase the numerical stability of the KF is to force the covariance matrix to be symmetric and to initialize it accordingly [32]. In this context, the *a posteriori* covariance matrix may be restated as follows [32]:

$$\mathbf{P}_{k+1|k+1} = (\mathbf{P}_{k+1|k+1} + {\mathbf{P}_{k+1|k+1}}^T)/2. \tag{29}$$

Another approach to this context is to equalize off-diagonal entries to each other (i.e., $\mathbf{P}_{ij} = \mathbf{P}_{ji}$), or making the eigenvalues of $\mathbf{P}_{k+1|k+1}$ to be positive. Using an appropriate initial value for the covariance improves the filter performance and prevents large or abrupt changes in the covariance in the estimation process [32].

4. Gaussian filters for nonlinear state estimation

For the general case of nonlinear systems, the predicted distribution $p(\mathbf{x}_{k+1}|\mathbf{Z}_k)$ cannot be computed exactly. Therefore, it needs to use some kind of approximations that however sacrifices optimality for computability and searches for a sub-optimal nonlinear filtering approach that is computationally tractable.

In order to approximate the nonlinear filtering process, there are two main approaches including the local approach and the global approach as follows [13]:

1. **Local approach:** In this approach, the distributions are assumed to be Gaussian, and then the *a posteriori* distribution is calculated using a direct numerical approximation in a local sense. This approach leads to several estimation techniques that are based on linearization such as the extended Kalman filter (EKF) [2, 3], and the central difference filter (CDF) [33, 34], or PDF approximation such as the unscented Kalman filter (UKF) [2, 3], quadrature Kalman filter (QKF) [11], and the cubature Kalman filter (CKF) [13]. The locality approach for the filter designation makes the filters to be simple and fast for implementation [13].
2. **Global approach:** In this approach, there are no assumptions pertaining to the *a posteriori* distribution; it is calculated using an indirect numerical approximation in a global sense. This leads to new filtering techniques such as the point-mass filter that uses adaptive grids [35], the Gaussian mixture filter [36], the mixture Kalman filter [12], and the well-known particle filter (PF). The particle filtering technique uses a set of weighted particles to approximate the state *a posteriori* PDF that contains nonlinear and non-Gaussian characteristics. The main disadvantage of estimation techniques categorized in the global approach is their large computational cost that makes them useless for some on-line state estimation applications [13].

Based on the method of approximation, the nonlinear Gaussian filters are classified into different groups. The linearization-based filtering and numerical integration based-filtering are explained in the next.

4.1. Linearization-based filtering: The extended Kalman filter (EKF)

The extended Kalman filter (EKF) is used for estimating states of a nonlinear dynamic system. Local linearization is performed in this method in order to approximate the nonlinearity of the state or measurement model at the operating point and to calculate a corrective gain. The EKF derivation is based on the Taylor series expansion of the nonlinear state (1) and measurement (2) models with linear terms. However, these nonlinear **f** and **h** functions cannot be applied to the covariance term directly, and their Jacobian's must be computed. Similar to the standard KF, the EKF has two main stages of prediction and update, whereas the *a priori* state and measurement estimates are calculated by [32]:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}(\hat{\mathbf{x}}_{k|k},\ \mathbf{u}_k,\ \mathbf{w}_k), \tag{30}$$

$$\hat{\mathbf{z}}_{k+1|k} = \mathbf{h}(\hat{\mathbf{x}}_{k+1|k}). \tag{31}$$

The Jacobian matrices, **F** and **H**, for the nonlinear state and measurement models are calculated by [32]:

$$\mathbf{F}_{k+1} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k+1|k},\mathbf{u}_k}, \quad \mathbf{H}_{k+1} = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k+1|k}}. \tag{32}$$

The main aspects of the EKF estimation technique may be summarized as:

- If the system is highly nonlinear, or a local linearization assumption does not fit the estimation problem well, a large estimation error will be produced and the EKF solution may lead to an estimate that diverges from the true state trajectory.
- Due to linearization, the EKF does not provide optimal state

estimates in the RMSE sense. Additionally, it does not guarantee unbiased state estimates and the calculated error covariance matrix does not necessarily equal to the real error covariance matrix [26].
- EKF's parameters need to be tuned such that the convergence improves. The convergence of the EKF is also dependent on the choice of the initial state estimates [26].

4.2. Numerical integration-based filtering

The main difficulty produced within the EKF derivation is the local linearization at a single point in the state probability densities. In order to ameliorate this difficulty, several techniques were proposed. Some techniques use the higher-order terms of the Taylor series expansion for approximating nonlinearities and lead to the so-called higher-order filters (e.g., second-order filter [1,37]). However, due to some difficulties appearing in calculation of the *Hessian* matrix, these approaches have not been used in the recent state estimation strategies. In order to overcome the main drawbacks of linearization-based approaches for nonlinear state estimation, the estimation filter may be constructed based on the transformation of statistical information. In regards to the computational issues, it is understood that approximating a probability distribution is much easier than approximating an arbitrary nonlinear transformation [37]. It in turn results in using the PDF approximation techniques for solving the integrals of Eqs. (3)–(7). The main basis for the integration-based estimation approach may be summarized in three main steps [37]:

1. Calculating the mean and covariance of a probability density via a set of selected samples
2. Propagating the samples by means of the nonlinear transformation function
3. Determining parameters of the propagated Gaussian approximation from the transformed samples

The Bayesian filtering paradigm is mainly based on calculating Gaussian weighted integrals whose integrands are formulated as: *nonlinear function×Gaussian density*. To provide a general formulation for numerical integration-based filtering, one may define a multi-dimensional weighted integral given by [13]:

$$I = \int_D f(x) w(x) dx, \tag{33}$$

where $f(x)$ is an arbitrary function, $D \subseteq \mathbb{R}^n$ is the region of integration, and $w(x) \geq 0$ is the known weighting function applied for all $x \in D$. In Gaussian filtering, $w(x)$ has a Gaussian distribution and preserves the non-negativity condition in the entire region D. In some cases, it may be extremely difficult to solve the integral (33) analytically. However, the integral I may be approximated using a weighted summation in which a set of points x_i and weights w_i is used as follows [13]:

$$I \approx \sum_{i=1}^{m} w_i f(x_i). \tag{34}$$

There are two main approaches for calculating x_i and w_i, that include the product and non-product rules. These two approaches are described as follows [13]:

1. **Product rules:** In this approach, the quadrature rule is used to calculate the integral (33) numerically [38]. In the case of Gaussian filters, this rule is restated by the Gauss-Hermite quadrature rule, whereas the weighting function $w(x)$ has a Gaussian distribution. The integrand $f(x)$ is then approximated by a polynomial in terms of x, and the Gauss-Hermite quadrature rule is applied to calculate the Gaussian-weighted integral [13]. Julier et al. introduced the Unscented Kalman filter (UKF) [39] based on the unscented transform, as another example of this approach. Furthermore, Ito et al. [40] proposed two different techniques. The first technique is the Gauss-Hermite filter (GHF) formulated based on the Gauss-Hermite quadrature rule and the second technique is the central difference filter (CDF) formulated based on polynomial interpolations.
2. **Non-product rules:** This approach is used to address the dimensionality issue in the product rules approach. In this context, the integrals are numerically solved by selecting sample points from the integration domain and applying the non-product rules. Some of the main non-product rules include the Monte Carlo technique [41], quasi-Monte Carlo technique [42], Lattice rules [43], and sparse grids [44]. The randomized Monte Carlo technique calculates integrals by utilizing a set of equally weighted sample points that are selected randomly. The quasi-Monte Carlo technique and lattice rule use a deterministic approach to produce the sample points from a unit hyper-cubic region [13]. The sparse grids method is a numerical technique used to integrate or interpolate high dimensional functions based on the Smolyak's rule. The sparse grids method searches to find the more important dimensions and put more grid points there [13].

The simplest technique among numerical integration-based filters is the unscented Kalman filter (UKF) invented by Julier et al. [39]. The unscented transform is used in the UKF to transform statistical information of the probabilistic densities into a predictor-corrector form. Lefebvre et al. described the unscented transform as a statistical linear regression technique that uses the system information at multiple points, in spite of the local linearization (e.g., the EKF) uses information of only one point [37]. More efficient filters are obtained by developing the Gauss-Hermite rule for numerical integration, such as the Quadrature Kalman filter (QKF) [10,11]. In other research, Norgaard et al. [34] invented the Divided Difference filter (DDF) to overcome several difficulties that appear in calculation of the derivatives in the EKF formulation. The DDF approximates the derivatives (e.g., Jacobian/Hessian matrices) and replaces them by the central divided difference. This is performed using the Sterling's polynomial interpolation criterion that makes the DDF a derivative free filter [37].

More recently, Ito et al. presented the mixed Gaussian filter that approximates the conditional probability density of states using a linear combination of multiple Gaussian distributions [40]. In order to update estimates in the mixed Gaussian filter, a Gaussian filter is applied to each Gaussian distribution, when each update is independent from the others and they operate in a parallel manner [40]. Kotecha et al. [45] invented the Gaussian Particle filter (GPF) technique. Since the GPF selects an optimal number of random samples and also benefits from the ability of analytical calculation and transformation of samples, it may be considered as a near-optimal estimation technique. The GPF is an extension to the Gaussian filter and applies the Monte Carlo integration technique to the Bayesian update rule [37]. The main drawback of any random-based sampling method (e.g., the GPF) is its high computational cost that makes it useless for on-line applications [46]. In the subsequent section, some of the main Gaussian filters that use the numerical integration-based approach are reviewed and compared in terms of accuracy, efficiency and complexity.

4.2.1. The unscented Kalman filter (UKF)

The sigma-point Kalman filter (SPKF) [32] is formulated using a weighted statistical linear regression approach that linearizes the nonlinear state model statistically [25,47]. The SPKF method produces a certain number of points called the sigma points from the projected probability distribution of the states. To provide the *a posteriori* estimate of the probability distribution, the sigma points are then mapped through the system's nonlinear model. This strategy makes any linearization unnecessary. Therefore, the calculation of Jacobin matrices is avoided and the accuracy of the state estimation increases considerably [25,48]. In this context, the unscented Kalman filter (UKF) utilizes a deterministic sampling approach, called the unscented transform, to select a minimal set of sample points around the mean.

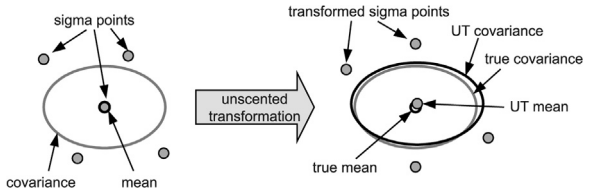


Fig. 7. Schematic of the unscented transformation used in the UKF [124].

They are propagated using the nonlinear functions. It is possible to approximately determine the mean and covariance of the density using the Monte Carlo sampling technique or Taylor series approximation [49]. The UKF can capture the *a posteriori* mean and covariance to the third order for any nonlinearity, and is therefore more accurate than the EKF. Another advantage of the UKF is that there is no need to compute the Jacobian or partial derivatives [39,49]. The UKF has a number of different forms that include the general unscented [32,50], the simplex unscented [32,50,51], and the spherical unscented [32,50,51]. Fig. 7 shows a schematic representation of the unscented transformation. To formulate the UKF, assume an n-dimensional state vector $\mathbf{x}_k$, with a mean $\overline{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ that are approximated by $2n+1$ sigma points. The UKF process is recursive, and is formulated in the prediction and update steps as [49]:

1. The prediction step [49]:

- Calculation of the sigma points as follows [49]:

$$\begin{cases} \boldsymbol{\chi}^0_{k|k} = \overline{\mathbf{x}}_{k|k}, & i = 0 \\ \boldsymbol{\chi}^i_{k|k} = \overline{\mathbf{x}}_{k|k} + (\gamma\sqrt{\mathbf{P}_{k|k}})_i, & i = 1, \ldots, n; \\ \boldsymbol{\chi}^{i+n}_{k|k} = \overline{\mathbf{x}}_{k|k} - (\gamma\sqrt{\mathbf{P}_{k|k}})_i, & i = 1, \ldots, n \end{cases} \tag{35}$$

where the parameter $\gamma = \sqrt{n+\kappa}$ is the associated weight of samples, determined as follows [49]:

$$\begin{cases} w_0 = \kappa/(n+\kappa), & i = 0 \\ w_i = 1/2(n+\kappa), & i = 1, \ldots, 2n' \end{cases} \tag{36}$$

where κ is a scaling factor. Note that $(\sqrt{(n+\kappa)\mathbf{P}_{k+1|k}})_i$ is the ith row or column of the matrix square root of $(n+\kappa)\mathbf{P}_{k+1|k}$. Note that summation of the normalized weights is equal to one.

- Predicting the state mean and covariance by propagating sigma points, as follows [49]:

$$\begin{aligned} \boldsymbol{\chi}^i_{k+1|k} &= f(\boldsymbol{\chi}^i_{k|k}),\ \hat{\mathbf{x}}_{k+1|k} = \sum_{i=0}^{2n} w_i \boldsymbol{\chi}^i_{k+1|k},\ \mathbf{P}_{k+1|k} \\ &= \mathbf{Q}_k + \sum_{i=0}^{2n} w_i [\boldsymbol{\chi}^i_{k|k} - \hat{\mathbf{x}}_{k+1|k}][\boldsymbol{\chi}^i_{k|k} - \hat{\mathbf{x}}_{k+1|k}]^T. \end{aligned} \tag{37}$$

- Calculating the measurement predictions as [49]:

$$\boldsymbol{\xi}^i_{k+1|k} = h(\boldsymbol{\chi}^i_{k+1|k}),\ \hat{\mathbf{z}}_{k+1|k} = \sum_{i=0}^{2n} w_i\, \boldsymbol{\xi}^i_{k+1|k}, \tag{38}$$

2. The update step [49]:

- Calculating the UKF gain [49]:

$$\begin{aligned} \mathbf{P}_{z,k+1|k} &= \sum_{i=0}^{2n} w_i\, [\boldsymbol{\xi}^i_{k+1|k} - \hat{\mathbf{z}}_{k+1|k}][\boldsymbol{\xi}^i_{k+1|k} - \hat{\mathbf{z}}_{k+1|k}]^T,\ \mathbf{P}_{xz,k+1|k} \\ &= \sum_{i=0}^{2n} w_i\, [\boldsymbol{\chi}^i_{k+1|k} - \hat{\mathbf{x}}_{k+1|k}][\boldsymbol{\xi}^i_{k+1|k} - \hat{\mathbf{z}}_{k+1|k}]^T,\ \mathbf{K}_{k+1} \\ &= \mathbf{P}_{xz,k+1|k}\mathbf{P}^{-1}_{z,k+1|k}. \end{aligned} \tag{39}$$

- Calculating the state mean and covariance updates [49]:

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}(\mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1|k}),\ \mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1}\mathbf{P}_z\mathbf{K}^T_{k+1}. \tag{40}$$

The main aspects of the UKF estimation technique are summarized below:

1. The UKF is similar to Monte Carlo methods, as it uses a number of points to estimate the system's mean and covariance. The main difference is that UKF only uses a small number of points that are not generated randomly such that the computational cost decreases. The convergence of the UKF is highly dependent on the choice of sigma points [50].
2. The UKF is better than the EKF in terms of the accuracy and computational cost. Tuning the EKF can be problematic when the Jacobian matrix is not derived easily. Furthermore, the EKF can only handle limited level of nonlinearities.
3. Although the particle filter may account several arbitrary distributions and nonlinearities, it remains computationally expensive. The UKF provides a trade-off between the particle filter and EKF in terms of accuracy and computational cost.

4.2.2. The Gauss-Hermite filter (GHF)

The Gauss-Hermite quadrature rule is the main basis for constructing the Gauss-Hermit filter (GHF). The rule states the weight function is assumed to be Gaussian density with zero mean and unit variance$N(x, 0, 1)$, when the interval of interest is $(-\infty, \infty)$. It is difficult to calculate quadrature points q_i and weights w_i analytically for a nonlinear system. In this context, some appropriate points should be chosen as the quadratic points based on the problem under study. Then after, the weights w_i may be obtained by calculating the moments M_i of the integral for the m number of quadrature points as follows [11]:

$$M_i = \int_a^b x^i \mathbf{W}(\mathbf{x})d\mathbf{x}, \quad for \quad i \in \{0, 1, \ldots, m-1\}, \tag{41}$$

The Vandermonde system of equations is stated as follows [11]:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ q_1 & q_2 & \cdots & q_m \\ \vdots & \vdots & & \vdots \\ q_1^{m-1} & q_2^{m-1} & \cdots & q_m^{m-1} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} M_0 \\ M_1 \\ \vdots \\ M_{m-1} \end{pmatrix}. \tag{42}$$

Thereafter, the set (q_i, w_i) may be used to approximate an integral using the quadrature rule [40]:

$$\int_{\mathbb{R}^n} F(\mathbf{x}) \frac{1}{(2\pi)^{n/2}} e^{-|\mathbf{x}|^2} d\mathbf{x} \sim \sum_{i=1}^{N} w_i F(q_i), \tag{43}$$

where $\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$ are starting values for the mean and covariance of the random variable $\mathbf{x}_0$. More details on the Gauss-Hermite quadrature rule and its applications for numerical integration are provided in [11, 40].

4.2.3. The quadrature Kalman filter (QKF)

The quadrature Kalman filter (QKF) was introduced and implemented by Arasaratnam and Haykin in 2007 [11]. The QKF was firstly formulated for nonlinear systems with an additive Gaussian distribution of the noise. In this formulation, the process and measurement models are linearized by using the statistical linear regression approach that projects the Gaussian density function based on a set of Gauss-Hermite quadrature points [11,52]. The main concept of the new QKF was extended to cover discrete-time nonlinear systems with an additive non-Gaussian distribution of the noise. In this extension, a bank of parallel QKFs referred to as the Gaussian sum-quadrature Kalman filter was used to approximate the *a priori* and *a posteriori* density functions. This approximation was alternatively performed using a finite number of weighted summations of Gaussian distributions, when the weights are calculated from the residuals of the QKFs [11].

Arasaratnam et al. reported that the Gaussian sum-quadrature Kalman filter is more accurate than other nonlinear filtering methods, such as the basic particle filters [11].

In this paper, only the general formulation of the QKF for nonlinear systems with an additive Gaussian distribution of noise is explained. At first the *a priori* and the *a posteriori* error covariance must be factored respectively as: $\mathbf{P}_{k|k} = \sqrt{\mathbf{P}_{k|k}}\sqrt{\mathbf{P}_{k|k}}^T$ and $\mathbf{P}_{k+1|k} = \sqrt{\mathbf{P}_{k+1|k}}\sqrt{\mathbf{P}_{k+1|k}}^T$. The QKF may be summarized in two steps that include the prediction and update steps as follows [11]:

1. The prediction step [11]:

- Calculation of the quadrature points $\{\mathbf{X}_{l,k|k}\}_{l=1}^m$ and their predicted values $\{\mathbf{X}^*_{l,k+1|k}\}_{l=1}^m$ for states as [11]:

$$\mathbf{X}_{l,k|k} = \sqrt{\mathbf{P}_{k|k}}\boldsymbol{\xi}_l + \hat{\mathbf{x}}_{k|k},\ \mathbf{X}^*_{l,k|k} = f(\mathbf{X}_{l,k+1|k}, \mathbf{u}_k). \tag{44}$$

- Calculation of the predicted state and covariance estimates, respectively as follows [11]:

$$\hat{\mathbf{x}}_{k+1|k} = \sum_{l=1}^{m} \omega_l \mathbf{X}^*_{\mathbf{l},k+1|k},\ \mathbf{P}_{k+1|k} = \sum_{l=1}^{m} \omega_l \mathbf{X}^*_{l,k+1|k}(\mathbf{X}^*_{l,k+1|k})^T - \hat{\mathbf{x}}_{k|k-1}(\hat{\mathbf{x}}_{k|k-1})^T + \mathbf{Q}_k. \tag{45}$$

- Calculation of the predicted quadrature points $\{\mathbf{Z}_{l,k+1|k}\}_{l=1}^m$ for measurement as well as the predicted measurement vector, respectively as follows [11]:

$$\mathbf{Z}_{l,k+1|k} = h(\mathbf{X}_{l,k+1|k}, \mathbf{u}_k),\ \hat{\mathbf{z}}_{k+1|k} = \sum_{l+1}^{m} \omega_l \mathbf{Z}_{l,k+1|k}. \tag{46}$$

- Evaluation of the predicted error covariance matrix and cross covariance matrix [11]:

$$\mathbf{P}_{zz,k+1|k} = \mathbf{R}_{k+1} + \sum_{l=1}^{m} \omega_l \mathbf{Z}_{\mathbf{l},k+1|k}\mathbf{Z}^T_{\mathbf{l},k+1|k} - \hat{\mathbf{z}}_{\mathbf{l},k+1|k}\hat{\mathbf{z}}^T_{\mathbf{l},k+1|k},\ \mathbf{P}_{xz,k+1|k} = \sum_{l=1}^{m} \omega_l \mathbf{X}_{l,k+1|k}\mathbf{Z}^T_{\mathbf{l},k+1|k} - \hat{\mathbf{x}}_{k|k-1}\hat{\mathbf{z}}^T_{\mathbf{l},k+1|k}. \tag{47}$$

2. The update step [11]:

- Evaluation of the QKF gain as [11]:

$$\mathbf{K}_{k+1} = \mathbf{P}_{xz,k+1|k}\mathbf{P}^{-1}_{zz,k+1|k}. \tag{48}$$

- Calculation of the update state and covariance estimates, respectively as [11]:

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}(\mathbf{z}_{k+1} - \mathbf{z}_{k+1|k}),\quad \mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1}\mathbf{P}_{zz,k+1|k}\mathbf{K}^T_{k+1}. \tag{49}$$

The main aspects of the QKF estimation technique may be summarized as:

1. If the *a priori* mean is far from the *a posteriori* mean, the EKF will fail to make accurate estimates. Since the QKF needs only to calculate some functions and not the derivatives of $f(\square)$ and $h(\square)$, it may be applied to non-smooth and non-analytical systems [11].
2. The QKF is able to estimate systems with correlated or non-additive Gaussian process and measurement noise, by adding terms to the state vector and relevant covariance [11].
3. The main disadvantage of the QKF is evident when applied to high dimensional systems, especially when the state vector size is greater than six. In high dimensional systems, the QKF's error covariance matrix may diverge from its nominal value [11].
4. Another disadvantage of the QKF is evident when applied to estimate systems. When applied to estimate systems with a limited word length for a long period of time, the round-off errors will accumulate and the QKF's accuracy may decrease. This may even cause numerical instability for the QKF in some cases [11].

4.2.4. The Cubature Kalman filter (CKF)

The Cubature Kalman filter (CKF) is a nonlinear state estimation technique for large-dimensional systems. It was invented and implemented by Arasaratnam and Haykin in 2009 [13]. The CKF formulation is based on a cubature transformation [53] that makes it possible to numerically calculate the Gaussian-weighted integrals for nonlinear Bayesian filtering. In order to produce a set of cubature points that will be later mapped through the state transition model, a third-degree spherical-radial cubature rule is used [13]. The cubature transformation overcomes the divergence and dimensionality issues that are the main issues with running the EKF, UKF or QKF estimation techniques. Furthermore, the CKF provides more accurate state estimates for nonlinear systems subjected to white Gaussian noise [54]. The cubature transformation helps the CKF to reduce the computational difficulties of calculating conditional density for some solvable multi-dimensional integrals.

The CKF uses the cubature rule to numerically approximate n-dimensional integrals such that [13]:

$$\int_{\mathbb{R}^n} f(x)N(\mathbf{x}, \overline{\mathbf{x}}, \mathbf{P}^{xx})\,d\mathbf{x} \approx \frac{1}{2n} f(\overline{\mathbf{x}} + \sqrt{\mathbf{P}^{xx}}\xi_i), \tag{50}$$

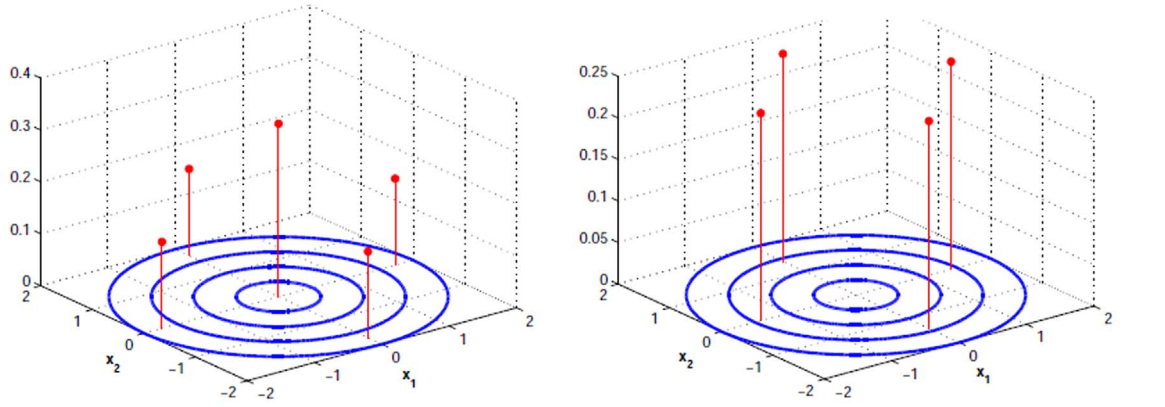
where the covariance is factorized as $\mathbf{P}^{xx} = \sqrt{\mathbf{P}^{xx}}\sqrt{\mathbf{P}^{xx}}^T$ and a set of $2n$ cubature points is given by:

$$\xi_i = \begin{cases} \sqrt{n}\ e_i, & i = 1, 2, \ldots, n \\ -\sqrt{n}\ e_{i-n}, & i = n+1, n+2, \ldots, 2n \end{cases}; \tag{51}$$

where $e_i \in \mathbb{R}^n$ represents the i^{th} elementary column vector. Arasaratnam et al. proposed the third degree cubature rule to approximate polynomial integrands [13]. The main structure of the CKF is similar to the UKF, but they are based on a thoroughly different set of deterministic points that provide weights for Gaussian integrals. The UKF utilizes the unscented transform to weight the selected sigma point set, whereas the CKF utilizes the cubature rule to provide weights for cubature point set. Fig. 8 presents a comparison of the point set distributions for the UKF and CKF estimation techniques. As illustrated by Fig. 8, the location and the height of each point represent the sample point and its weights respectively. The main advantage of using the cubature-point set made by the CKF over the sigma-point set made by UKF, is to increase the filter stability as well as its numerical accuracy. Note that to increase the accuracy of the CKF, Jia et al. have introduced a new family of CKFs with arbitrary degrees of accuracy that calculate the spherical and radial integrals [55]. The described third-degree CKF is a special example of this family. The accuracy and performance of the high-order CKFs is similar to the Gauss-Hermite filter (GHF). In order to achieve $(2m+1)^{th}$ degree of accuracy, the number of points that are required for the cubature transform increases by the dimension n polynomially. Since the computational complexity of CKFs is a polynomial function of the point's dimension, CKFs are more efficient than the GHF computationally [55].

The main advantages of the CKF over other estimation methods are summarized [13]:

1. Note that the cubature rule is a derivative-free transformation and hence, it removes the difficulties that may appear in the calculation of the Jacobian and Hessian of systems with complicated nonlinearities. This derivative-free characteristic allows writing the prepackaged computer programs [13].
2. The cubature rule involves $2n$ cubature points, where n is the number of state variables. Hence, $2n$ functional computations are required at each iteration cycle. The computational complexity is linearly changing with the state vector dimension n and this makes the CKF effective for estimating high dimensional systems [13].



(a) Sigma point set for the UKF (b) Third degree spherical-radial cubature point set for CKF

Fig. 8. Comparison of the two-dimensional point set distribution in UKF and CKF [13].

3. Negative weight in the CKF formulation prevents the factorization of the covariance matrix in a squared form. The CKF formulation of the filter guarantees that the sample weights are positive definite, hence the squared form of the CKF is always available [13].

4.2.5. The Monte Carlo Kalman filter (MCKF)

In the Monte Carlo Kalman filter (MCKF), the Monte Carlo numerical integration technique is used to approximate the expected values in the integral forms. In this way, N_s samples are drawn from the state Gaussian distribution $N(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{P}^{xx})$, where $\{\mathbf{x}^{(i)}, \ i = 1, \ldots, N_s\}$ is a set of random samples (or particles) with weights $\{\mathbf{w}^{(i)} = 1/N_s, \ i = 1, \ldots, N_s\}$. The state distribution is approximated by the Monte Carlo technique as [56]:

$$N(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{P}^{xx}) \simeq \sum_{i=1}^{N_s} \mathbf{w}^{(i)} \delta(\mathbf{x} - \mathbf{x}^{(i)}), \tag{52}$$

where N_s is the number of samples and δ is the Dirac function. Note that the probability density near the sample $\mathbf{x}^{(i)}$ is obtained by the density of points in a region around $\mathbf{x}^{(i)}$. If $N_s \to \infty$, the approximation of the integral will converge to its true value. The MCKF is constructed based on approximating the predicted values of state, measurement, and their covariance through the Monte Carlo numerical integration technique. The MSKF is extensively explained in [56]. The main points regarding the MCKF are as:

1. The Monte Carlo integration rule is similar to the quadrature integration rule presented previously. The two rules are similar, however, a difference between the two rules exist. In the quadrature rule, the sample points are selected at fixed intervals, while in the Monte Carlo rule they are selected randomly [56].
2. In the Monte Carlo integration process, the variance of state estimation is proportional to $1/N_s$ which means for a simulation study with 10^4 samples, the error in variance is equal to 1%. Since the numeric integration of the Monte Carlo process is recursive, it may result in increased error and the filter's divergence [57, 58].
3. The computational cost of the MCKF is independent of the number of dimensions of the integrands. The GHF computational cost is proportional to M^n and therefore, by increasing the system dimension, growth occurs rapidly. Hence, for such cases with a large dimension, the MCKF is more popular than the GHF [56, 57].

5. Adaptive filtering for state estimation

The previous state estimation techniques are all formulated under the assumption that statistics of the input and measurement noise and system parameters are known. However, in real applications, there is often some degree of uncertainty or inaccuracy in the values of physical parameters, initial conditions, or noise characteristics. Applying the filter without any modification to such cases degrades the optimality of the estimation method and increases the state estimation error. In order to alleviate such effects, one solution is to estimate the uncertain parameters and noise statistics during the filtering process and then augment an adaptation mechanism to the filter. This mechanism is referred to as an adaptive filter, which tunes the filter gain based on the parametric variations or noise statistics. Note that adaptation is considered into the filtering process such that robustness against statistical variation of parameters increases. Adaptation does not affect optimality of the filter with respect to a specific statistical model [59].

There are two main approaches for adaptive state estimation including the adaptive filter design based on gain adaptation, and the Multiple Models (MM) approach. In the adaptive filter with gain adaptation approach, there is only one model of the system and some techniques are augmented to estimate the state and known parameters recursively based on statistic properties of noise and uncertainties. In contrast, in the MM approach, several state-space models are used to cover all operating regimes of the system. Each model presents a particular operating regime of the system under certain conditions. The state and covariance estimates are calculated as a weighted summation of each filter output.

5.1. Adaptive filtering with gain adaptation

There are three main approaches for adapting the filter gain that include: 1- the adaptive filter with noise tuning [60,61], 2- the adaptive filter with parameter tuning [60–62], and 3- the joint filtering of states and parameters [62–64]. In the adaptive filter with noise tuning approach, whenever the filter starts to diverge, some techniques are applied for tuning the levels of measurement noise and or modeling uncertainties. The main symptom of the filter divergence is the characteristics of the measurement error (innovation) vector. By starting divergence, the measurement error vector is not white and its covariance does not match with the predicted measurement error covariance. In the adaptive filter with parameter tuning approach, some off-line techniques are used to estimate the system and noise parameters based on a batch of measurements. For instance, the expectation-maximization technique may be directly used to calculate the Maximum Likelihood Estimate (MLE) [65]. The success of this approach depends on a number of factors such as the number of uncertain parameters, the magnitude of uncertainty, the functional dependence of the system outputs on uncertain parameters, and the quality of measurements [65].

In the joint filtering of states and parameters, the system's unknown parameters are considered as new states. Hence, the new state vector contains the former states and unknown parameters and

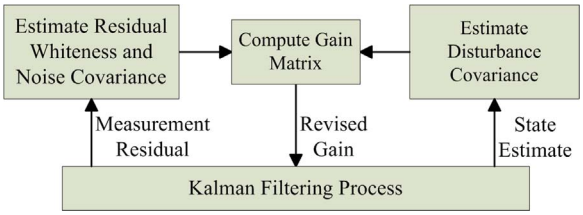


Fig. 9. Schematic of the adaptive Kalman filter (AKF) estimation process [60].

used to estimate both the states and unknown parameters. There are several techniques that may combine with estimation filters (e.g. the EKF [62], the UKF [63], or the particle filter [64]) and tune their gain to jointly estimate the unknown states and parameters. However, this approach is not efficient in some situations and may cause numerical instability. Following sections describe the adaptive filter with noise tuning [60], the adaptive filter with parameter tuning [66], and the joint filtering of states and parameters [62].

5.1.1. The adaptive Kalman filter (AKF) with noise tuning

The Kalman filter extracts all the available information from the measurement vector $\mathbf{z}_k$, where the measurement error $\mathbf{e}_\mathbf{z}$ is zero-mean white noise. Following the Kalman filter formulation, assuming $E\{\mathbf{x}_k\mathbf{v}_k{}^T\} = E\{\hat{\mathbf{x}}_k\mathbf{v}_k{}^T\} = 0$, $\mathbf{e}_{\mathbf{z},k|k}$ and its covariance $\mathbf{S}_k$ are calculated using Eqs. (20) and (21), respectively. In this context, if $\mathbf{e}_{\mathbf{z},k|k}$ is not white or its corresponding covariance does not equal to Eq. (21), the filter is not optimal. This issue may be due to a number of factors such as the input disturbance, sensor and instrumentation noise, modeling and parametric uncertainties, and the case in which the measurement error covariance is different from the one generated by $\mathbf{Q}$ and $\mathbf{R}$. In order to check whiteness of the measurement error signal $\mathbf{e}_\mathbf{z}$, a whiteness test may be defined in terms of the autocorrelation function $\mathbf{C}_k$ of the measurement error $\mathbf{e}_\mathbf{z}$, as follows [60]:

$$\mathbf{C}_k = \frac{1}{N}\sum_{i=k}^{N} \mathbf{e}_{\mathbf{z}_i}\mathbf{e}^T_{\mathbf{z}_{i-k}}, \tag{53}$$

where N is the number of samples. In this context, if the measurement error is white, $\mathbf{C}_k$ should be non-zero only when k is zero. Assuming that the cross-correlation of the measurement elements is negligible, the whiteness test can only be applied on diagonal elements of $\mathbf{C}_k$ [60].

The adaptive Kalman filter (AKF) with noise tuning may be used to estimate systems with a color measurement error. It applies an off-line adaptive estimation procedure to estimate the process noise covariance $\mathbf{Q}$, and the measurement noise covariance $\mathbf{R}$, as follows [60,65]:

1. Estimate bias $\bar{\mathbf{v}}_k$ of the measurement error (innovation) and its covariance $\hat{\mathbf{S}}_k$ as [60]:

$$\bar{\mathbf{v}}_k = \frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_k,\ \hat{\mathbf{S}}_k = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{v}_k - \bar{\mathbf{v}})(\mathbf{v}_k - \bar{\mathbf{v}})^T. \tag{54}$$

2. Estimate $\mathbf{R}$ based on the above relations as [60]:

$$\hat{\mathbf{R}}_{k+1} = \hat{\mathbf{S}}_{k+1} - \mathbf{H}\left(\frac{1}{N}\sum_{k=1}^{N}\mathbf{P}_{k+1|k}\right)\mathbf{H}^T. \tag{55}$$

The state estimation error is calculated by: $\mathbf{e}_{\mathbf{x},k+1|k} \triangleq \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k} = \mathbf{F}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}) + \mathbf{w}_k$. Hence, the state error covariance matrix $\mathbf{Q}$ is calculated by [60]:

$$\mathbf{Q}_{k+1} = \mathbf{F}\,\mathbf{cov}(\mathbf{e}_{\mathbf{x},k|k})\,\mathbf{F}^T - \mathbf{cov}(\mathbf{e}_{\mathbf{x},k+1|k}). \tag{56}$$

Note that $\mathbf{cov}(\mathbf{e}_{\mathbf{x},k|k})$ may be approximated using $\mathbf{P}_{k|k}$, whereas $\mathbf{cov}(\mathbf{e}_{\mathbf{x},k+1|k})$ cannot be approximated by $\mathbf{P}_{k+1|k}$ through the AKF. It is because if the AKF approximates $\mathbf{P}_{k+1|k}$ such that $\mathbf{P}_{k+1|k} = \mathbf{F}\mathbf{P}_{k|k}\mathbf{F} + \mathbf{Q}_k$, then it leads to $\mathbf{Q}$ with a wrong covariance. To overcome this issue, $\mathbf{x}_{k+1}$ needs to initially be approximated by $\hat{\mathbf{x}}_{k+1|k+1}$, and then, the following equation is used [60]:

$$\mathbf{e}_{\mathbf{x},k+1|k} \approx \hat{\mathbf{x}}_{k+1|k+1} - \hat{\mathbf{x}}_{k+1|k} = \mathbf{d}_{k+1}. \tag{57}$$

3. Estimate $\mathbf{cov}(\mathbf{e}_{\mathbf{x},k+1|k})$ as the empirical covariance of $\mathbf{d}_k$, as follows [60]:

$$\overline{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{d}_k,\ \mathbf{P}_{k+1|k} = \frac{1}{N-1}\sum_{k=1}^{N}(\mathbf{d}_k - \overline{\mathbf{X}})(\mathbf{d}_k - \overline{\mathbf{X}})^T. \tag{58}$$

4. Estimate $\mathbf{Q}$ using the following relation [60]:

$$\hat{\mathbf{Q}}_{k+1} = \mathbf{P}_{k+1|k} - \mathbf{F}\left(\frac{1}{N}\sum_{k=1}^{N}\mathbf{P}_{k+1|k+1}\right)\mathbf{F}^T. \tag{59}$$

The AKF is now formulated using the estimated values of $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ within the Kalman filter formulation. Fig. 9 presents a block-diagram scheme of the AKF process.

5.1.2. The adaptive filter with parameter tuning

Zhong Ji and Brown [66] have introduced an adaptive filter technique based on tuning the unknown parameters over time. In this method, the system dynamics are described using two sets of equations that include the state model and the parameter model. Thereafter, two EKFs, named as the state filter and the parameter filter, are run in parallel to estimate values of states and parameters respectively. The system dynamics is initially described in state-space, as follows [66]:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) + \mathbf{w}_k,\ \mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \boldsymbol{\theta}) + \mathbf{v}_k, \tag{60}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the vector of state variables, and $\boldsymbol{\theta} \in \mathbb{R}^p$ is the vector of unknown parameters. The state filter operates similar to the EKF method, as described in Section 4.1. It predicts the states and their covariance using the state and measurement models and then updates them by calculating the gain.

The parameter filter moreover predicts the parameters and their covariance, as follows [66]:

$$\hat{\boldsymbol{\theta}}_{k+1|k} = \hat{\boldsymbol{\theta}}_{k|k},\ \mathbf{P}_{\boldsymbol{\theta},k+1|k} = \lambda^{-1}\mathbf{P}_{\boldsymbol{\theta},k|k}, \tag{61}$$

where λ is the forgetting factor and is selected such that $0 < \lambda \leq 1$. The parameter filter's gain is given by:

$$\mathbf{e}_{\mathbf{z},k+1|k} = \mathbf{z}_{k+1} - \mathbf{h}(\hat{\mathbf{x}}_{k+1|k}, \hat{\boldsymbol{\theta}}_{k+1|k}),\ \mathbf{K}_{\boldsymbol{\theta},k+1} = \mathbf{P}_{\boldsymbol{\theta},k+1|k}\,\mathbf{H}_{\boldsymbol{\theta},k+1}{}^T(\mathbf{H}_{\boldsymbol{\theta},k+1}\mathbf{P}_{\boldsymbol{\theta},k+1|k}\mathbf{H}_{\boldsymbol{\theta},k+1}{}^T + \mathbf{R}_{\boldsymbol{\theta},k+1})^{-1}. \tag{62}$$

Note that $\mathbf{H}_\boldsymbol{\theta}$ is the parameter measurement matrix and is given by [66]:

$$\mathbf{H}_{\boldsymbol{\theta},k+1} = -\frac{\partial \mathbf{e}_{\mathbf{z},k+1|k}}{\partial \hat{\boldsymbol{\theta}}_{k+1|k}} = \mathbf{H}_{k+1}\frac{\partial \hat{\mathbf{x}}_{k+1|k}}{\partial \hat{\boldsymbol{\theta}}_{k+1|k}}\bigg|_{\hat{\boldsymbol{\theta}}_{k+1|k}}, \tag{63}$$

Finally, the parameter filter updates the predicted parameters and their covariance, as follows [66]:

$$\hat{\boldsymbol{\theta}}_{k+1|k+1} = \hat{\boldsymbol{\theta}}_{k+1|k} + \mathbf{K}_{\boldsymbol{\theta},k+1}\mathbf{e}_{\mathbf{z},k+1|k},\ \mathbf{P}_{\boldsymbol{\theta},k+1|k+1} = (I - \mathbf{K}_{\boldsymbol{\theta},k+1}\mathbf{H}_{\boldsymbol{\theta},k+1})\mathbf{P}_{\boldsymbol{\theta},k+1|k}. \tag{64}$$

5.1.3. The adaptive filter with joint estimation of states and parameters

Sun et al. [62] have introduced a method for joint filtering of states and parameters that applies to systems with Eq. (60). In this method, uncertain parameters of the model are considered as additional states and the EKF method is used to estimate the augmented state vector. The augmented state vector is hence given by: $\mathbf{X} = [\mathbf{x}^T, \boldsymbol{\theta}^T]^T$, where $\mathbf{X} \in \mathbb{R}^{n+p}$. The augmented state equation is restated as [62]:

$$\mathbf{X}_{k+1} = \begin{bmatrix} \mathbf{x}_{k+1} \\ \boldsymbol{\theta}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) \\ \boldsymbol{\theta}_k \end{bmatrix} + \begin{bmatrix} \mathbf{w}_k \\ \boldsymbol{\eta}_k \end{bmatrix} = \mathbf{g}(\mathbf{X}_k, \mathbf{u}_k) + \boldsymbol{\xi}_k, \tag{65}$$

where $\boldsymbol{\eta}_k$ is the uncorrelated Gaussian noise on uncertain parameters with the covariance matrix $\boldsymbol{\Phi}_k$. The measurement model corresponding to the augmented state vector is given by [62]:

$$\mathbf{z}_k = \mathbf{h}(\mathbf{X}_k, \mathbf{u}_k) + \mathbf{v}_k. \tag{66}$$

In order to use the EKF method for estimating the augmented state vector, Jakobian matrices of for the state and measurement models are respectively calculated by [62]:

$$\mathbf{F}_k = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial \mathbf{x}^T} & \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^T} \\ 0 & \mathbf{I} \end{bmatrix}_{\hat{\mathbf{x}}_{k|k}, \hat{\boldsymbol{\theta}}_{k|k}}, \quad \mathbf{H}_{k+1} = \begin{bmatrix} \frac{\partial \mathbf{h}}{\partial \mathbf{x}^T} & \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}^T} \end{bmatrix}_{\hat{\mathbf{x}}_{k+1|k}, \hat{\boldsymbol{\theta}}_{k+1|k}, \mathbf{u}_{k+1}}. \tag{67}$$

Thereafter, the EKF is applied to the augmented state vector in two steps including prediction and update. Note that the process noise covariance matrix is equal to: $\begin{bmatrix} \mathbf{Q} & 0 \\ 0 & \boldsymbol{\Phi} \end{bmatrix}$.

5.2. The multiple models (MM) adaptive filtering

An interesting approach for the modeling and estimation of complex nonlinear systems is to describe the system by a finite number of possible operating regimes. Such systems are generally classified as hybrid dynamical systems. A hybrid dynamical system is defined as a system that contains two types of time-varying elements [4]. The first type referred to as state variables, includes elements that vary with time. The second type includes elements that only transit from one operational mode to another. This is referred to as the mode or modal state. Note that the state variables only describe the systems dynamic behavior, while the mode states represent a possible system's regime among a finite number of possible operating regimes. These formulations are often referred to as the Markovian jump or hybrid estimation phenomenon [4]. Further studies regarding the above phenomena are found in [4,67–70].

The multiple models (MM) approach is the most well-known approach used to describe a hybrid dynamic system in which a set of models is considered that covers all of the possible operating regimes. The first generation of the MM algorithms were produced by Magill [69] and Lainiotis [71,72], and were widely implemented and promoted by several researchers. These researchers included Maybeck [73–77], Bar-Shalom [4,78–80], Rong Li [80–83], and a number of other prominent researches. In the MM approach, it is assumed that the dynamic system operates according to one of a finite number of models, each corresponding to a particular operating regime. These models may differ in their mathematical structure or in their noise and uncertainty characteristics. The structural differences include dimensions of state variables, system inputs, and outputs. Noise and uncertainties may differ in the level or their probability distributions and can be represented as an additive or multiplicative term [2,4]. MM filters are generally classified into two categories: static and dynamic. In the static MM method, the system follows a fixed operating mode and no switching occurred from one mode to another during the estimation process. In contrast, the dynamic MM estimator switches from one mode to another automatically in order to provide the most accurate estimate of the operating situation [4].

Tugnait presented a survey on suboptimal MM methods for discrete dynamic systems with abruptly changing structures [68]. He stated that the transition probability matrix is available and the abrupt changes can be modeled by finite states of a Markov chain. Note that partitioning a complex system into several simpler systems is a powerful approach, and results in parallel processing of the original system that provides optimal and robust solutions [84–86]. Some studies have been performed on the MM algorithm when the transition probability matrix is unknown [68,87]. Other studies contain situations that model the system's switching structure through a semi-Markov chain [88,89]. Rong Li et al. presented a MM technique for the purpose of noise identification, which is applicable for both stationary and non-stationary noise with rapidly or slowly varying statistics [80,82,90]. Since the performance of the MM estimation strategy is directly related to the model sets selection, the primary difficulty in the implementation of MM methods lies in the correct identification of the model set. It has been proven that the use of too many models (i.e., over-designing the solution) may have as bad an effect on the solution as the use of too few models (i.e., under-designing) [83].

To formulate a MM filter, assume a general form of state and measurement models, respectively [4]:

$$\mathbf{x}_{k+1} = \mathbf{F}_{m,k}\mathbf{x}_k + \mathbf{G}_{m,k}\mathbf{u}_{m,k} + \boldsymbol{\Gamma}_{m,k}\mathbf{w}_{m,k}, \; \mathbf{z}_k = \mathbf{H}_{m,k}\mathbf{x}_k + \mathbf{v}_{m,k}. \tag{68}$$

Where m and M denote the current model and the set of all possible modes respectively. The event that model m_i is operating at time k is presented as: $M_{i,k} = \{m_k = m_i\}$. It is assumed that the system model sequence is a homogenous Markov chain with transition probabilities calculated as follows [4]:

$$P\{m_{j,k+1} | m_{i,k}\} \pi_{ij,k}; \quad \forall\, i, j \in M, \tag{69}$$

where π_{ij} is the Markovian transition probability from mode i to mode j, where $\sum_{j=1}^{r} \pi_{ij,k} = 1$. [4]. The mode probabilities are updated at each new measurement and the resulting weights are used to estimate the state. Fig. 10 shows a block-diagram scheme of one cycle of a static MM filter.

As long as each mode sequence is matched to a filter, the number of filters required for the state estimation process will grow exponentially. In order to avoid this numerical problem, suboptimal techniques should be considered. A simple technique for obtaining a suboptimal solution is to keep the N samples of histories with the largest

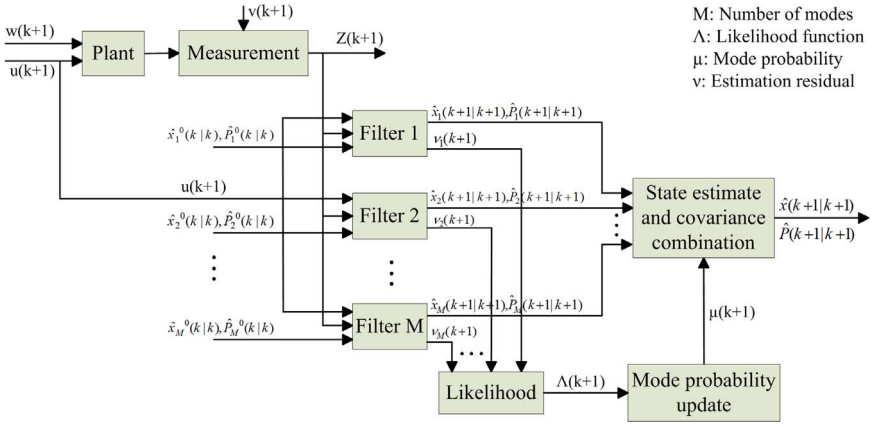


Fig. 10. Block-diagram scheme of one cycle of a static MM filter.

probabilities, ignore the rest, and renormalize the selected N probabilities in a way their summation will equal to unity. Within this approach, there are three methods: the 1st-order Generalized Pseudo Bayesian (GPB1), the 2nd-order version (GPB2), and the Interacting Multiple Model (IMM) strategy. In the GPB1 method, only the possible models in the last sampling period are taken into account. The algorithm will only need to run r parallel filters to formulate the best estimate. The GPB2 method uses the last two sampling periods, and hence r^2 filters are required. The IMM algorithm is computationally more efficient than the GPB1 and GPB2 algorithms [4]. For the IMM strategy, with r hypotheses, each filter utilizes a different weighted combination of the previous model conditioned estimates. This model condition is referred to as the mixed initial condition. Based on this, there is an interaction between different possible modes of the system at each period of time. In addition to the reduction of the computational cost of the IMM filter, the accuracy of the overall estimate and the convergence rate is increased significantly [4].

5.2.1. The interacting multiple model (IMM) filter

The interacting multiple model (IMM) algorithm is the most popular type of MM filter, and is capable of estimating the system state variables among several switching modes. Bloom is among the first researchers to propose the IMM algorithm with a suitable compromise between the performance and complexity in MM systems [91]. Its computational cost is close to other methods such as those with small quadratic components, while its performance is similar to GPB2 [79]. The IMM filter is able to estimate the state of a dynamic system that operates in several operating modes (regimes), which can 'switch' from one mode to another. In this form, multiple state equations are used to describe each of the operating regimes. These regimes are typically referred to as linear models, where each model captures a particular operating point of a general nonlinear time-varying model. A Markov transition matrix is then used to determine the probability that the system is in one of the operating regimes [4]. Fig. 11 shows a block diagram representation of one cycle of an IMM estimation filter.

In order to formulate the IMM filter, let a hybrid linear system of (68) be used to describe a nonlinear dynamic system. The IMM filter consists of three steps that are briefly described as [81]:

1. Interaction Step: The mixing probabilities are calculated, which refer to the probability of an event when mode m_i was in effect at time k-1, given that the mode m_j is in effect at time k conditioned on Z^{k-1}. In this step, the initial state estimate and covariance estimate are calculated from r different Kalman (or other) filters corresponding to the r different operation modes [81].

2. Filtering Step: In this step, mode-matched filtering is performed. The likelihood function associated to each of the r filters is also computed. Any estimation method or filter may be used during this step; however, the most commonly implemented method is the Kalman filter. The mixed initial state and covariance are used as inputs to the Kalman (or other) filter matched to mode m_j. The filtering step starts by predicting the state and covariance of each mode, calculating the measurement error, the filter gain, and finally updating the state and covariance estimates as well as the mode probability [81].

3. Combination Step: In this step the overall state mean and covariance are estimated suing a weighted summation of the state and covariance estimate from each individual mode [81].

6. Robust filtering for state estimation

In this section, the three main approaches used for robust state estimation of stochastic systems are reviewed. The first approach is the robust Kalman filter [92] in which some techniques are used to robustify the KF given norm-bounded noise and uncertainties. The second approach, referred to as the variable structure filtering (VSF), was firstly introduced by Habibi and his co-researchers [93]. The VSF is a model-based filter and benefits from the robustness characteristic of the variable structure systems. The VSF-type filters provide robust state estimates against a large amount of structural and parametric uncertainties. The third approach, referred to as H_∞ filtering, had been introduced by Zames in 1980 [94]. The H_∞ filtering approach focuses on the worst-case energy gain designation that produces estimation error with small energies for all small disturbance energies. These three approaches are explained in the next.

6.1. The robust Kalman filter (RKF)

The main idea of the KF designation is minimizing the estimation error covariance. However, the KF is only accurate when there are small amounts of uncertainties and noise in the process model, initial conditions and measurements. There are a large number of publications that describe the robust Kalman filter (RKF) techniques. Xie et al.

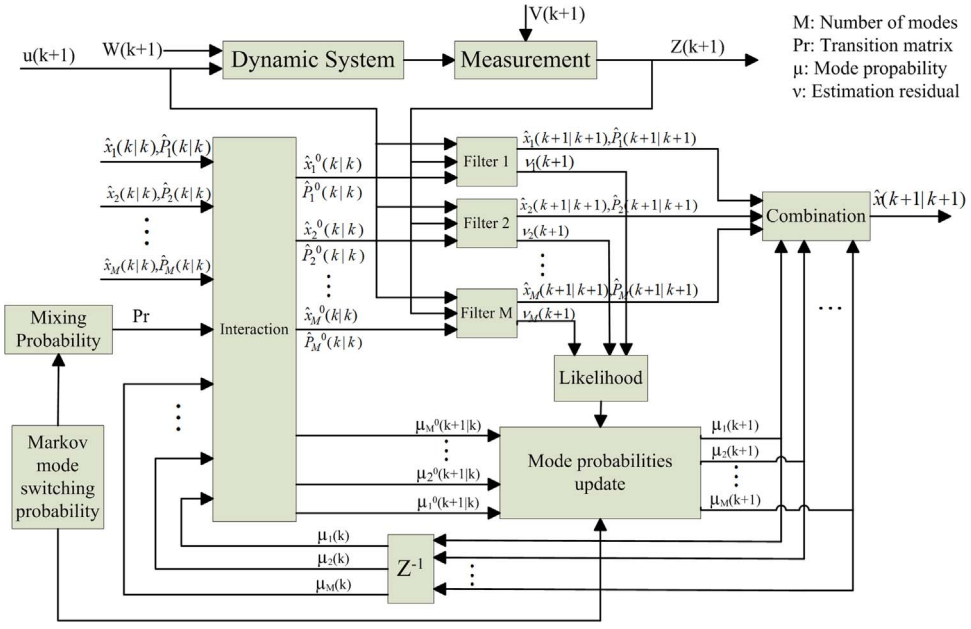


Fig. 11. Block-diagram scheme of the interacting multiple model (IMM) strategy.

have proposed a RKF technique for linear systems subjected to norm-bounded parametric uncertainty in the state and measurement matrices [95]. Masreliez et al. have introduced a robust Bayesian estimator that can operate under two different scenarios [96]. The first situation is when the state x is Gaussian and the measurement z is non-Gaussian (heavy-tailed). The second scenario is when the state is non-Gaussian (heavy-tailed) and the measurement z is Gaussian [96]. Furthermore, Hsieh has proposed a RKF technique that is insensitive to unknown inputs [97]. This filter is an alternative to the Kitanidis's unbiased minimum variance filter [98]. Wang et al. also introduced a RKF algorithm as applied to linear systems with stochastic parametric uncertainties [99]. This method is designed to minimize an upper bound of the mean square estimation error at each step. Benavoli et al. also designed a RKF by considering the uncertainty characterizations in terms of coherent lower previsions [100].

Bertsekas et al. presented the set-valued approach for state estimation that is based on defining ellipsoids around state estimates that are consistent with the measurement data [92,101]. Note that the centers of ellipsoids are assumed to be the estimated states. In this context, there are several recursive algorithms to account for uncertain models, particularly the one proposed by Savkin et al. [102]. The Guaranteed-cost design is another important approach in which the filter is designed by preserving an upper bound on the variance of the state estimation error. This approach is mostly applied to quadratically stable systems in the steady-state phase of the operation [92,95]. In this section, the Sayed's robust Kalman filtering technique is reviewed as a general framework for linear robust state estimation.

Assume an uncertain dynamic model with bounded uncertainties in the state model such that [92]:

$$\mathbf{x}_{k+1} = (\mathbf{F}_k + \delta\mathbf{F}_k)\mathbf{x}_k + (\mathbf{G}_k + \delta\mathbf{G}_k)\mathbf{u}_k, \tag{70}$$

where $\delta\mathbf{F}_k$ and $\delta\mathbf{G}_k$ denote small uncertainties in state and control matrices. It is assumed that uncertainties in $\mathbf{F}_k$ and $\mathbf{G}_k$ satisfy the following equality [92]:

$$[\delta\mathbf{F}_k \quad \delta\mathbf{G}_k] = \mathbf{M}_k\boldsymbol{\Delta}_k[\mathbf{E}_{\mathbf{f},k} \quad \mathbf{E}_{\mathbf{g},k}], \tag{71}$$

where $\{\mathbf{M}, \mathbf{E_f}, \mathbf{E_g}\}$ are known matrices, and $\boldsymbol{\Delta}_k$ is selected such that $\|\boldsymbol{\Delta}_k\| \leq 1$. The *a posteriori* state estimate $\hat{\mathbf{x}}_{k+1|k}$ is obtained from the *a priori* state estimate $\hat{\mathbf{x}}_{k|k}$ by solving the following criterion [92]:

$$\min_{\{\mathbf{x}_k,\mathbf{u}_k\}} \max_{\delta\mathbf{F}_k\delta\mathbf{G}_k} \left\{ \|\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\|^2_{\mathbf{P}^{-1}_{k|k}} + \|\mathbf{u}_k\|^2_{\mathbf{Q}_k^{-1}} + \|\mathbf{z}_{k+1} - \mathbf{H}\mathbf{x}_{k+1}\|^2_{\mathbf{R}^{-1}_{k+1}} \right\}. \tag{72}$$

Sayed has presented a solution $\{\hat{\mathbf{x}}_{k+1|k} - \hat{\mathbf{u}}_{k+1|k}\}$ to the above problem by solving the corresponding set of equations. In order to follow his solution, it is assumed that $\mathbf{x}_0$, $\mathbf{u}_k$, and $\mathbf{v}_k$ are uncorrelated zero-mean white stochastic processes with following variances [92]:

$$E\left(\begin{bmatrix}\mathbf{x}_0\\ \mathbf{u}_i\\ \mathbf{v}_i\end{bmatrix}\begin{bmatrix}\mathbf{x}_0\\ \mathbf{u}_j\\ \mathbf{v}_j\end{bmatrix}^T\right) = \begin{bmatrix}\boldsymbol{\Pi}_0 & 0 & 0\\ 0 & \mathbf{Q}_i\delta_{ij} & 0\\ 0 & 0 & \mathbf{R}_i\delta_{ij}\end{bmatrix}, \tag{73}$$

where $\boldsymbol{\Pi}_0$, $\mathbf{Q}_i$, and $\mathbf{R}_i$ are given weighting matrices and are assumed to be positive-definite. δ_{ij} denotes the *Kronecker delta* and is equal to one when $i = j$ and equal to zero otherwise. The initial conditions for the filter are set to $\hat{\mathbf{x}}_{0|0} = \mathbf{P}_{0|0}\mathbf{H}_0^T\mathbf{R}_0{}^{-1}\mathbf{z}_0$ and $\mathbf{P}_{0|0} = (\boldsymbol{\Pi}_0^{-1} + \mathbf{H}_0{}^T\mathbf{R}_0{}^{-1}\mathbf{H}_0)^{-1}$, alternatively. The RKF presented by Sayed may be summarized into the following three steps [92]:

Step 1: If $\mathbf{H}_{i+1}\mathbf{M}_i = 0$, then put $\hat{\lambda}_i = 0$. If not, the construct the cost function $\mathbf{G}(\lambda)$, as presented by (74), and determine $\hat{\lambda}_i$ by minimizing $\mathbf{G}(\lambda)$ [92].

$$\mathbf{G}(\lambda) = \|\mathbf{x}(\lambda)\|^2_{\mathbf{Q}} + \lambda\|\mathbf{E_a}\mathbf{x}(\lambda) - \mathbf{E_b}\|^2 + \|\mathbf{A}\mathbf{x}(\lambda) - \mathbf{b}\|^2_{\mathbf{w}(\lambda)}. \tag{74}$$

Where $\mathbf{x} = col\{\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}, \mathbf{u}_k\}$, $\mathbf{b} = \mathbf{y}_{k+1} - \mathbf{H}_{k+1}\mathbf{F}_k\hat{\mathbf{x}}_{k|k}$, $\mathbf{A} = \mathbf{H}_{k+1}[\mathbf{F}_k \quad \mathbf{G}_k]$, $\mathbf{Q} = \mathbf{P}_{k|k}{}^{-1} \oplus \mathbf{Q}_{k+1}{}^{-1}$, $\mathbf{W} = \mathbf{R}_{k-1}{}^{-1}$, $\mathbf{H} = \mathbf{H}_{k+1}\mathbf{M}_k$, $\mathbf{E_a} = [E\{\mathbf{F_k}\} \quad E\{\mathbf{G_k}\}]$, and $\mathbf{E_b} = -E\{\mathbf{F_k}\}\hat{\mathbf{x}}_{k|k}$. Moreover, the non-negative scalar parameter is determined from the optimization problem, and functions $\mathbf{W}(\lambda)$, $\mathbf{Q}(\lambda)$, and $\mathbf{x}(\lambda)$ are given by [92]:

$$\mathbf{W}(\lambda) \triangleq \mathbf{W} + \mathbf{W}\mathbf{H}(\lambda\mathbf{I} - \mathbf{H}^T\mathbf{W}\mathbf{H})^{+}\mathbf{H}^T\mathbf{W},\ \mathbf{Q}(\lambda) \triangleq \mathbf{Q} + \lambda\mathbf{E_a}{}^T\mathbf{E_a},\ \mathbf{x}(\lambda) \triangleq [\mathbf{Q}(\lambda) + \mathbf{A}^T\mathbf{W}\mathbf{A}]^{-1}[\mathbf{A}^T\mathbf{W}(\lambda)\mathbf{b} + \lambda\mathbf{E_a}{}^T\mathbf{E_b}]. \tag{75}$$

Step 2: Replace matrices $\mathbf{Q}_k$, $\mathbf{R}_{k+1}$, $\mathbf{P}_{k|k}$, $\mathbf{G}_k$, and $\mathbf{F}_k$, from the KF method by following matrices [92]:

$$\begin{aligned}\hat{\mathbf{Q}}_k^{-1} &= \mathbf{Q}_k^{-1} + \hat{\lambda}_k E_{g,k}^T[\mathbf{I} + \hat{\lambda}_k\mathbf{E}_{f,k}\mathbf{P}_{k|k}\mathbf{E}_{f,k}^T]^{-1}\mathbf{E}_{g,k},\ \hat{\mathbf{R}}_{k+1}\\ &= \mathbf{R}_{k+1} - \hat{\lambda}_k{}^{-1}\mathbf{H}_{k+1}\mathbf{M}_k\mathbf{M}_k{}^T\mathbf{H}_{k+1}^T,\ \hat{\mathbf{P}}_{k|k}\\ &= \mathbf{P}_{k|k} - \mathbf{P}_{k|k}E_{f,k}^T[\hat{\lambda}_k{}^{-1}\mathbf{I} + \mathbf{E}_{f,k}\mathbf{P}_{k|k}\mathbf{E}_{f,k}^T]^{-1}\mathbf{E}_{f,k}\mathbf{P}_{k|k},\ \hat{\mathbf{G}}_k\\ &= \mathbf{G}_k - \hat{\lambda}_k\mathbf{F}_k\hat{\mathbf{P}}_{k|k}\mathbf{E}_{f,k}{}^T\mathbf{E}_{g,k},\ \hat{\mathbf{F}}_k = (\mathbf{F}_k - \hat{\lambda}_k\hat{\mathbf{G}}_k\hat{\mathbf{Q}}_k\mathbf{E}_{g,k}^T\mathbf{E}_{f,k})(\mathbf{I} - \hat{\lambda}_k\hat{\mathbf{P}}_{k|k}\mathbf{E}_{f,k}^T\mathbf{E}_{f,k}).\end{aligned} \tag{76}$$

If $\hat{\lambda}_k = 0$, then it will be obtained that $\hat{\mathbf{Q}}_k = \mathbf{Q}_k$, $\hat{\mathbf{R}}_{k+1} = \mathbf{R}_{k+1}$, $\hat{\mathbf{P}}_{k|k} = \mathbf{P}_{k|k}$, $\hat{\mathbf{G}}_k = \mathbf{G}_k$, and $\hat{\mathbf{F}}_k = \mathbf{F}_k$[46].

Step 3: Estimate the state and the state error covariance using the KF formulations and the above matrices.

6.2. The variable structure filtering (VSF)

The variable structure filter (VSF) is a model-based state estimation strategy that was introduced and implemented by Habibi et al. [93]. The VSF-type filters use the variable structure system's concept to preserve stability given bounded parametric uncertainties. Thus, the main objective is to increase stability and convergence of the filter for situations with higher degrees of modeling or parametric uncertainties. In such situations the performance of common estimation techniques such as Kalman-type filters may degrade significantly. The degradation occurs as a result of filter instability, inappropriate definition of initial conditions, modeling uncertainties and measurement noise. The VSF-type filtering and its newer extension (e.g., the Smooth Variable Structure Filter (SVSF) [103] uses the robustness property of variable structure systems that results in stability within an upper bound for uncertainties and noise levels.

In variable structure systems, the control input often contains a discontinuous term, called the sliding variable s, that is defined as a function of the state variable in the following form [103]:

$$u(x,t) = \begin{cases} u^+(x,t) & if \quad s(x) > 0\\ u^-(x,t) & if \quad s(x) < 0\end{cases}, \tag{77}$$

where $u^+(x,t)$ and $u^-(x,t)$ are continuous functions. Following the variable structure theory, the VSF's gain contains a discontinuous corrective term that preserves stability given bounded noise and uncertainties. It refines the *a priori* state estimates into the *a posteriori* state estimates. In order to formulate the VSF method, the sliding variable is defined as: $\mathbf{S}_k = \boldsymbol{\Lambda}\mathbf{e}_{\mathbf{z}_{k|k}}$, where $\mathbf{e}_{\mathbf{z}_{k|k}} = \mathbf{z}_k - \hat{\mathbf{z}}_k$ is the estimation error and $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ is a diagonal matrix with constant positive elements. Here, the objective is to eliminate the sliding variable S and satisfy the sliding condition given by $\mathbf{S} = 0$. This condition eliminates the estimation error even in highly uncertain noisy situations.

In this section, the first generation of the VSF-type filtering, namely the VSF method [93], is described. Following this, a more efficient version of the VSF, referred to as the smooth variable structure filter (SVSF) [103] is described in details. Newer versions of this filter such as the SVSF with covariance derivation [25, 104], and the SVSF with a variable boundary layer (SVSF-VBL) [25, 105].

6.2.1. The variable structure filtering (VSF)

Habibi et al. introduced the simplest generation of the VSF in 2003 [93]. In order to implement the VSF for state estimation, the system must be completely observable. One cycle of the VSF method, as it applies to a linear system of Eqs. (16) and (17), contains following steps [93]:

1. Prediction Step:

- Calculation of *a priori* state and measurement estimates [93]:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{F}_k\hat{\mathbf{x}}_{k|k} + \mathbf{G}_k\mathbf{u}_k,\ \hat{\mathbf{z}}_k = \mathbf{H}_k\mathbf{x}_k. \tag{78}$$

2. Update Step:

- Calculation of the VSF's corrective gain that is stated as [93]:

$$\begin{aligned}\mathbf{K}_{k+1} = {} & \mathbf{F}_{k+1}{}^{-1}\mathbf{H}^{+}(\|\mathbf{H}\ \mathbf{F}_{k+1}|\{\Upsilon|\mathbf{H}^{+}\|\ \mathbf{e}_{\mathbf{z}_{k+1|k}}| + |\mathbf{F}_{k+1}{}^{-1}\mathbf{H}^{+}\xi_{\max}\ \mathbf{z}_{k+1}| \\ & + [|\mathbf{H}^{+}| + |\mathbf{F}_{k+1}{}^{-1}\mathbf{H}^{+}|(\xi_{\max} + \mathbf{I})] \times \mathbf{V}_{\max} \\ & + |\mathbf{F}_{k+1}{}^{-1}\mathbf{H}^{+}\delta_{\max}\ \mathbf{u}_{k+1}| + (|\mathbf{F}_{k+1}{}^{-1}| + |\mathbf{F}_{k+1}{}^{-1}\mathbf{H}^{+}\widetilde{\mathbf{H}}_{\max}|)\mathbf{W}_{\max}\}| \\ & \circ \mathrm{sgn}(\mathbf{e}_{\mathbf{z}_{k+1|k}})),\end{aligned} \tag{79}$$

where sgn is the signum function, $\circ$ is the Schur product, and + is the pseudo-inverse transform. Υ is a diagonal matrix with positive elements that contain the convergence rate γ_{ii} for measurement z_i. Moreover, $\mathbf{V}_{\max}$ and $\mathbf{W}_{\max}$ denote upper bounds for the measurement and process noises, and $\boldsymbol{\xi}_{\max}$, $\boldsymbol{\delta}_{\max}$, and $\widetilde{\mathbf{H}}_{\max}$ denote the upper bound for small uncertainties in the state model $\mathbf{F}$, control matrix $\mathbf{G}$, measurement matric $\mathbf{H}$, respectively [93].

- Refine the *a priori* state estimate into the *a posteriori* state estimate [93]:

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}. \tag{80}$$

The discontinuous formulation of $\mathbf{K}$ produces high frequency chattering that degrades the estimation performance. In order to reduce these unwanted effects, the smoothing boundary layer concept may be considered. Utilizing the smoothing boundary layer concept, outside the smoothing layer $\boldsymbol{\psi}$ the signum function may be applied to guarantee stability, when inside the layer a saturation function is applied to approximate the signum function and eliminate high frequency chattering [93]:

$$sat\,(e_{z_i,k+1|k}/\psi_i) = \begin{cases} 1, & e_{z_i,k+1|k}/\psi_i > 1 \\ e_{z_i,k+1|k}\ /\ \psi_i & -1 \le e_{z_i,k+1|k}/\psi_i \le 1, \\ -1, & e_{z_i,k+1|k}/\psi_i \le -1 \end{cases} \tag{81}$$

The width of the boundary layer indicates the level of uncertainties in the estimation process. However, in order to remove the chattering completely, the width of the smoothing boundary layer should be sufficiently large. However, increasing the smoothing layer's width decreases the average level of accuracy in state estimates and hence, there needs to be a compromise between the level of uncertainties and the VSF's performance [93]. Stability of the VSF is proven based on the Lyapunov's second law of stability [93]. Habibi has also presented the derivation of the VSF corrective gain with explicit consideration of modeling uncertainties [93]. In subsequent research, Habibi introduced the extended variable structure filter [106] that is an extension to the VSF applied to nonlinear systems.

6.2.2. The smooth variable structure filter (SVSF)

The smooth variable structure filter (SVSF) is a more advanced generation of the variable structure filters, introduced and implemented by Habibi in 2007 [103]. Similar to the VSF concept, the SVSF is a model-based robust state estimation method that can be used to estimate state variables of smooth nonlinear dynamic systems. It has an inherent switching action that guarantees convergence of the state estimates to within a region of the real values. The switching characteristic of the SVSF is due to the variable structure formulation of the discontinuous gain, which provides robustness to bounded uncertainties. Most filters only provide the estimation error (filter innovation) and its covariance as measures of performance. The SVSF also provides another indicator that is linked to modeling uncertainties [103].

Fig. 12 shows the main concept of the SVSF method. It presents the system state trajectory, estimated state trajectory, and existence subspace versus time. To start the estimation process, an initial value is selected for the state estimation process based on a prior knowledge of the systems. Then after, the estimated state is pushed towards a neighborhood of the system's true value referred to as the existence subspace. Once the value enters into the existence subspace, the estimated state is forced into switching along the system state trajectory via the SVSF gain. The estimated state trajectory remains within the existing subspace that has a width proportional to modeling uncertainties, measurement noise, and disturbances. There has been a great amount of research to improve the SVSF method and prepare it as a useful tool for FDI applications [25,93,103,106]. The SVSF method differs with the VSF method in the derivation approach and the corrective gain formulation. The derivation of the VSF' gain is based on the explicit consideration of the upper bounds for modeling uncertainties and measurement noise. The derivation of the SVSF's gain is based on introducing a positive definite Lyapanov candidate that contains squared value of the estimation error as: $V_k = |e_{z_k|k}|$. Stability is then achieved by proving that the negative definiteness of the Lyapunov difference. It is proven that the SVSF process is stable and convergent if [103]:

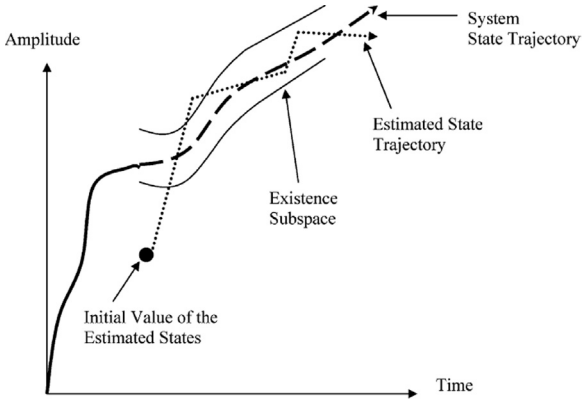


Fig. 12. Representation of the SVSF estimation concept [25].

$$|e_{z_k|k}| < |e_{z_{k-1|k-1}}|. \tag{82}$$

The SVSF estimation process has the same steps as the VSF process, but its corrective gain formulation is different. For a linear system with one measurement corresponding to each of the state variables, the SVSF's gain is stated as [103]:

$$\mathbf{K}_{k+1} = \mathbf{H}^{-1}(|\mathbf{e}_{\mathbf{z}_{k+1|k}}| + \boldsymbol{\gamma}\,|\mathbf{e}_{\mathbf{z}_{k-1|k-1}}|) \circ sat\,(\boldsymbol{\psi}^{-1}\mathbf{e}_{\mathbf{z}_{k+1|k}}), \tag{83}$$

where $\circ$ denotes the Schur product (element-by-element multiplication), $\boldsymbol{\gamma}$ is the convergence rate matrix that is diagonal and is defined such that $0 < \gamma_{ii} < 1$. Moreover, $\boldsymbol{\psi}^{-1}$ is a diagonal matrix denote the constant smoothing boundary layer widths [103]. The saturation function is defined as same as Eq. (81) [103]. It is proven that the corrective gain of (83) pushes the estimated states across the switching hyper plane and guarantees stability. By adopting the Luenberger observer into the SVSF method, the SVSF method may be applied to systems with fewer measurements than states [103].

Note that there are two different boundary layers in the SVSF concept including the existence layer, and the smoothing layer. The existence layer is referred to as the neighborhood of the estimated state trajectory in which the stability of the estimation process is ensured. The width of the existence layer varies in time as a function of the modeling uncertainties. Although the width of the existence layer is unknown, it is possible to obtain an upper boundary β for it. The smoothing boundary layer is defined to approximate the sign function in the corrective gain formulation and filter out chattering. Its width is known as ψ and outside this layer the sign function is applied to achieve the stability, while inside the smoothing layer the discontinuity of $\mathbf{K}$ is interpolated by the saturation function to provide smooth state

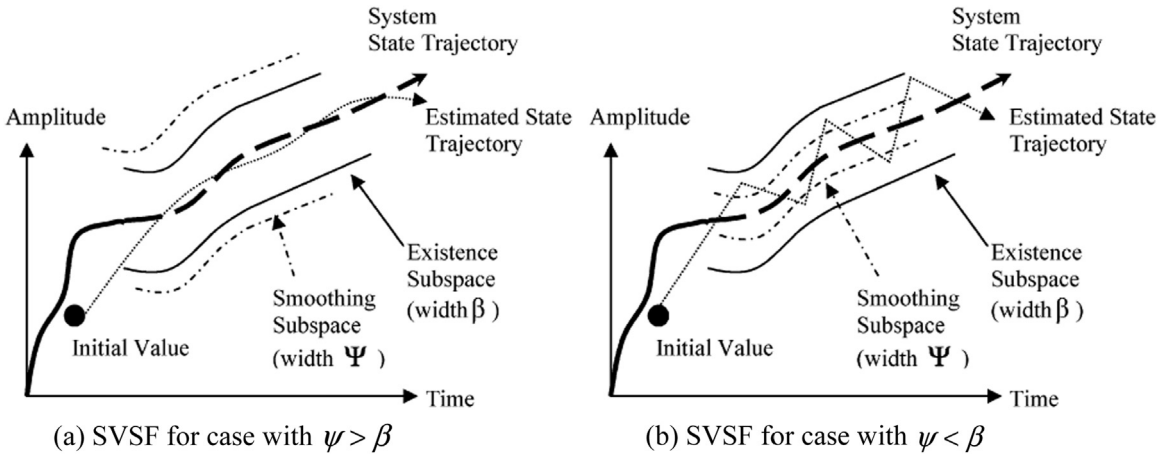


Fig. 13. Effect of the smoothing boundary layer width ψ on the SVSF performance [103].

estimates. As presented in Fig. 13 (a), when the smoothing layer width is larger than the existence layer width $\psi > \beta$, chattering is filtered out. Otherwise as presented in Fig. 13 (b), if the smoothing layer width is smaller than the existence layer width $\psi < \beta$, then the smoothing layer will be ineffective and chattering will appear [103].

Generally speaking, the filter gain construction is the main difference between the KF and SVSF. The KF gain depends on the *a priori* and the *a posteriori* measurements error values, whereas the SVSF gain depends on the smoothing boundary layer widths, convergence rate $\boldsymbol{\gamma}$, and the measurement matrix $\mathbf{H}$ [103]. A significant amount of research has been conducted to improve the SVSF's performance. Gadsden et al. combined the SVSF with other filters such as the PF [107], the cubature Kalman filter (CKF) [108], and the IMM filter [109]. New research also concentrated on the derivation of a state error covariance term for the SVSF [104], formulating a continuous-time form of the SVSF [110], and defining an optimal smoothing boundary layer [105]. Further details and developments on the SVSF may be found in [25], and [111]. The main features of the SVSF that make it a unique and attractive tool for state estimation may be summarized as follows:

a) It provides robustness and guarantees stability within a predefined boundary layer for bounded uncertainties and noise levels [103].
b) Other estimation techniques such as the KF, UKF, CKF, and PF provide the innovation and the error covariance as measures of performance. However, the SVSF also provides a secondary indicator of performance based on the chattering function [111], which explicitly relates to uncertainties and modeling errors [103].

6.2.3. The SVSF with a variable boundary layer (SVSF-VBL)

The former version of the SVSF is introduced when the width of the smoothing boundary layer remains constant. As discussed, the width of the smoothing boundary layer is selected based on available knowledge of the upper bound of modeling uncertainties and maximum levels of measurement noise and parametric errors. However, considering a constant width for this layer is a conservative choice that decreases the accuracy of state estimations. A more efficient smoothing boundary layer may be obtained when its width is changing as a function of uncertainty and noise levels. Gadsden introduced the state error covariance matrix for the SVSF and then used it to derive an optimal time-varying width for the smoothing boundary layer [112]. The calculation process of the error covariance matrix is similar to that of Kalman filtering [104]. The key idea for specifying the boundary layer width matrix $\boldsymbol{\Psi}$ is to take the partial derivative of the *a posteriori* error covariance matrix with respect to ψ. This idea is similar to calculating an optimal gain for the Kalman filter. This leads to an optimal formulation of the SVSF that optimizing the diagonal entries of the state error covariance matrix. Hence, a time-varying smoothing boundary layer for the SVSF method is calculated by [112]:

$$\frac{\partial[trace(\mathbf{P}_{k+1|k+1})]}{\partial \boldsymbol{\Psi}} = 0. \tag{84}$$

It is proven that the optimal time-varying smoothing layer for the SVSF leads to the well-known Kalman filter solution for linear systems. Following this, Gadsden proposed a method entitled the SVSF-VBL. It is a combination of the SVSF and KF. In this method, the SVSF guarantees stability for estimates that are outside the smoothing boundary layer and provides optimality for estimates inside the boundary layer [112]. Fig. 14 (a) presents the SVSF-VBL concept. In the SVSF method, the smoothing boundary layer width is equal to the limit. It results in the loss of optimality demonstrated as the difference between the limit and the optimal boundary layer. However, the SVSF-VBL (KF) gain should be applied to provide efficient estimates. Fig. 14 (b) presents the SVSF-VBL concept for estimating systems with high

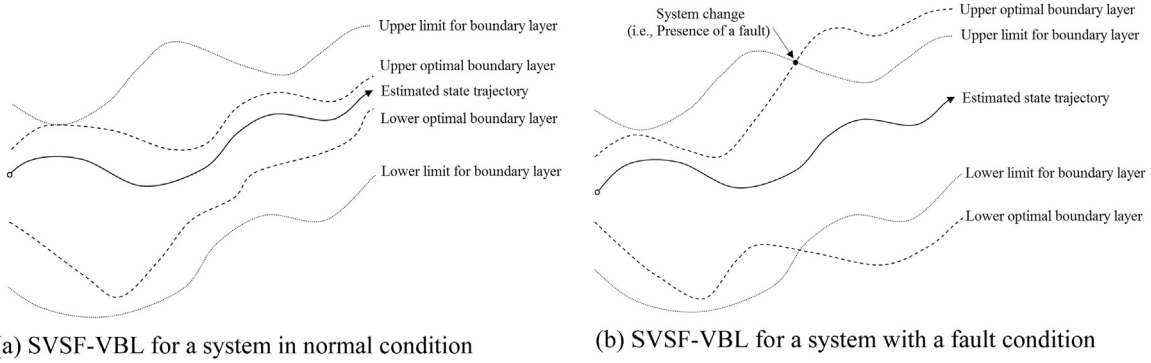


Fig. 14. Main concept of SVSF-VBL for a system in normal and faulty conditions [112].

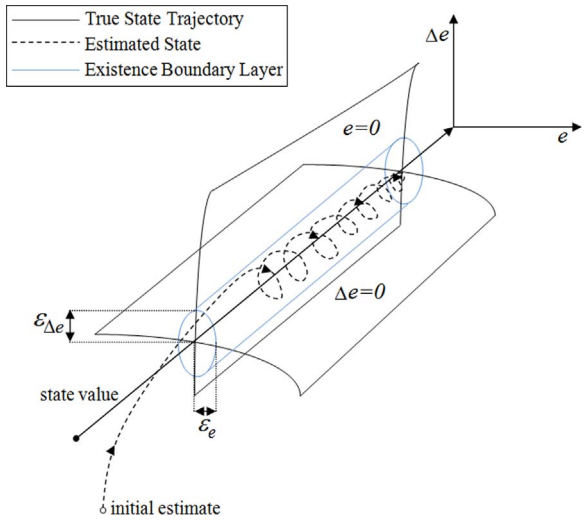


Fig. 15. Main concept of the dynamic 2nd-order filter for state estimation [113].

amount of uncertainties such as a system with a fault condition. In this case, the optimal smoothing boundary layer is larger than the limit enforced by the SVSF method. Hence, the SVSF-VBL gain is made equal to the SVSF gain to use its robust characteristic and guarantee stability in uncertain conditions [112]. Inside the limit, the SVSF-VBL optimal boundary layer is used. One cycle of the SVSF-VBL state estimation technique contains the following steps [112]:

1. Prediction Step [112]:

- Calculation of the *a priori* state and covariance estimates using (18) and (19) respectively.
- Derivation of the *a priori* measurement error using (20).

2. Update Step [112]:

- Calculation of the combined error vector as:

$$\mathbf{E}_{k+1} = |\mathbf{e}_{\mathbf{z}_{k+1|k}}| + \boldsymbol{\gamma}|\mathbf{e}_{\mathbf{z}_{k|k}}|. \tag{85}$$

- Derivation of the smoothing boundary layer matrix given by [112]:

$$\boldsymbol{\psi}_{k+1} = (\mathbf{E}_{k+1}{}^{-1}\mathbf{H}\,\mathbf{P}_{k+1|k}\mathbf{H}^T\mathbf{S}_{k+1}{}^{-1})^{-1}. \tag{86}$$

- Calculation of the SVSF-VBL gain given by [112]:

$$\mathbf{K}_{k+1} = \mathbf{H}^{-1}\,\mathbf{E}_{k+1}\,\boldsymbol{\psi}_{k+1}{}^{-1}. \tag{87}$$

- Refinement of the *a priori* state and covariance estimates into the *a posteriori* estimates that are respectively obtained using (23) and (24) [112].

6.2.4. The dynamic second-order filter (dynamic 2nd-order filter)

The basic idea of the dynamic 2nd-order filter [113] is based on the second-order sliding mode concept in which chattering is suppressed by means of preserving the second sliding condition. Afshari used this concept to design the dynamic 2nd-order filter in 2014. He implemented this method for several real-world applications such as fault detection and diagnosis [114–118], battery management systems [119], and vehicle tracking [120, 121]. The dynamic 2nd-order filter provides a powerful tool for estimating state variables of highly uncertain dynamic systems while the measurement error (innovation) and its time-difference are pushed towards zero. It is important to note that although adding a smoothing boundary layer ψ to the gain formulation of the SVSF-type filtering can decrease chattering, it compromises accuracy and robustness. The smoothing layer interpolates the discontinuous corrective action within a small neighborhood of the switching surface and hence, the real sliding motion doesn't occur. To overcome this problem, the dynamic 2nd-order filter is designed to avoid chattering by preserving a robustness criterion that results in elimination of the measurement error and its difference over time [113].

In order to formulate the dynamic 2nd-order filter, a dynamic manifold $\boldsymbol{\sigma}_k \in \mathbb{R}^{m\times 1}$ is defined as a linear combination of the measurement error and its time-difference. The dynamic manifold is given by [113]:

$$\boldsymbol{\sigma}_k = \Delta\mathbf{e}_{\mathbf{z}_{k|k}} + \mathbf{C}\mathbf{e}_{\mathbf{z}_{k|k}}, \tag{88}$$

where $\mathbf{e}_{\mathbf{z}_{k|k}}$ denotes the measurement error (innovation sequence), $\Delta\mathbf{e}_{\mathbf{z}_{k|k}}$ denotes its time-difference, and $\mathbf{C} \in \mathbb{R}^{m\times m}$ represents the cut-off frequency coefficient matrix. Note that $\mathbf{C} = \mathbf{diag}(c_{ii})$ is a diagonal matrix and its entries are defined such that $c_{ii} > 0$. Introducing the cut-off frequency coefficient into the filter formulation provides the dynamic 2nd-order filter as a second-order filter with an adjustable bandwidth. It is available to tune the cut-off frequency in order to remove the noise, chattering, and any high-frequency dynamics from state estimates. Stability of the dynamic 2nd-order filter was proven using the Lyapunov's second law of stability, whereas the Lyapunov function was defined in discrete-time, as follows:

$$V_k = \boldsymbol{\sigma}_k\boldsymbol{\sigma}_k{}^T = (\Delta\mathbf{e}_{\mathbf{z}_{k|k}} + \mathbf{C}\mathbf{e}_{\mathbf{z}_{k|k}})(\Delta\mathbf{e}_{\mathbf{z}_{k|k}} + \mathbf{C}\mathbf{e}_{\mathbf{z}_{k|k}})^T. \tag{89}$$

Note that satisfaction of the Lyapunov function (89) results in decreasing the measurement error and its difference over time [113]. Fig. 15 illustrates the main concept of the dynamic 2nd-order filter. As shown, the existence boundary layer for each measurement is a cylindrical form subspace with a center line at $\sigma_k = \Delta e_{z_{k|k}} + c_{ii}e_{z_{k|k}}$ [113].

The dynamic 2nd-order filtering process is similar to the SVSF method [103], but the corrective gain formulation is different. It is stated as [113]:

$$\mathbf{K}_{k+1} = \mathbf{H}^{-1}[\mathbf{e}_{\mathbf{z}_{k+1|k}} - (\boldsymbol{\gamma} + \boldsymbol{\Lambda})\mathbf{e}_{\mathbf{z}_{k|k}} + \boldsymbol{\Lambda}\mathbf{e}_{\mathbf{z}_{k-1|k-1}}], \tag{90}$$

where $\boldsymbol{\Lambda} = \mathbf{diag}(\lambda_{ii}) \in \mathbb{R}^{m\times m}$ denotes the cut-off frequency matrix such that $0 < \lambda_{ii} < 1$. It was proven that the dynamic 2nd-order filter with the corrective gain of Eq. (90) satisfies the stability condition (89) [113]. The corrective gain (90) represents a second-order Markov process providing higher degrees of smoothness in the state estimated trajectories. It is because the corrective gain applies separate constraints to the measurement error and its difference that suppress both of them in finite time. The 2nd-order formulation of gain provides higher amount of information from the past and this allows the dynamic 2nd-order filter to update estimates at time *k+1* based on the information available from step *k* and *k-1*. This improves the filter's performance in terms of accuracy, smoothness and robustness without the need for approximation [113]. Fig. 16 shows the concept of decreasing the error by means of the dynamic 2nd-order filter.

6.3. The H_∞ Filtering

The first systematic approach into the robustness concept was firstly introduced by Zames in 1980 [94]. He presented the H_∞ theory for design and implementation of robust controllers that are insensitive to modeling uncertainties, and lack of statistical knowledge of inputs. The H_∞ theory may be considered as an extension to the linear quadratic Gaussian (LQG) theory introduced in 1960s [122]. The LQG design was performed based on a perfect model of the system and complete knowledge of input statistics. In contrast to the LQG concept, the H_∞ method was proposed to negate the necessity of a perfect model or complete knowledge of the input statistics. The H_∞ theory is designed based on tracking the energy of signal for the worst possible values of modeling uncertainties w and measurement noise v [122] (Table 1 and 2).

In order to clarify the H_∞ concept, one may define a measure of how good the estimator is as:

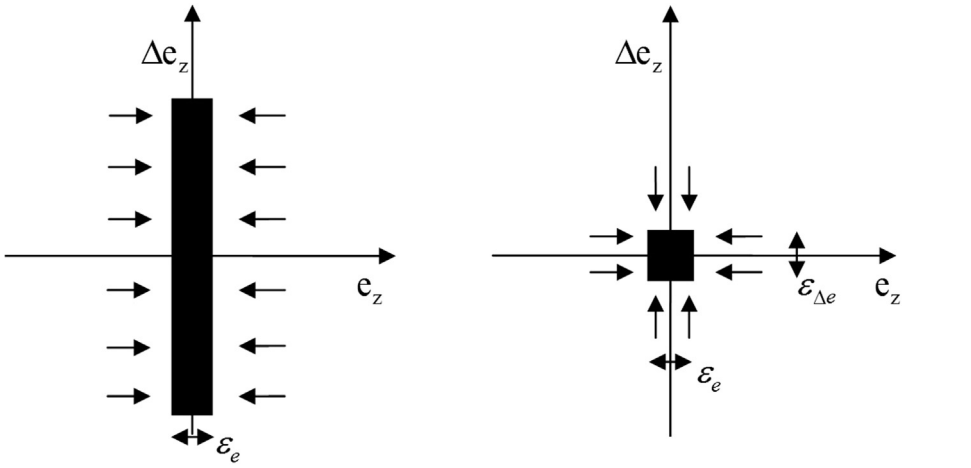


Fig. 16. Decreasing the error by means of the dynamic 2nd-order filter [113].

$$\min_{\hat{\mathbf{x}}} \max_{\mathbf{w},\mathbf{v}} J \tag{91}$$

where **w** and **v** are the process and the measurement noise that try to degrade the state estimates. The main objective of the H_∞ filtering is to provide state estimates by minimizing the worst possible effect of **w** and **v** on the estimation error. The cost function J may be defined as [122]:

$$J = \frac{ave\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|_{\mathbf{Q}}}{ave\|\mathbf{w}_k\|_{\mathbf{W}} + ave\|\mathbf{v}_k\|_{\mathbf{V}}}, \tag{92}$$

where **Q**, **W**, and **V**, each denotes the weighting matrix corresponding to a parameter, when the averages are calculated on the weighted norms overall time steps k. Note that minimizing the cost function (92) means that the H_∞ filter tries to calculate the state estimates $\hat{\mathbf{x}}_k$ to be as close to $\mathbf{x}_k$ as possible, when noise terms make function J large. It is too difficult to mathematically find a solution for the described problem. It is possible to solve the problem for $J < 1/\theta$, when θ is a constant parameter, and called the performance bound. It is chosen by the designer and its value depends on the case under study. Satisfying the condition $J < 1/\theta$ through the H_∞ filter, it is not important how large the magnitudes of noise terms **w** and **v** are. The H_∞ filter ensures that the ratio of the estimation error to noise will always remain less than $1/\theta$ [122].

In order to formulate the H_∞ filter recursively, based on the game theory approach [123], let assume a linear stochastic system of Eqs. (16) and (17). The system is moreover subjected to [32]:

$$\mathbf{z}_k = \mathbf{L}_k \mathbf{x}_k, \tag{93}$$

Table 2
Symbols and math operations with their definitions.

Symbol	Definition	Symbol	Definition
$\bar{\square}$	Mean value	**c**	Autocorrelation function
$\hat{\square}$	Estimated value	**e**	Error
$\tilde{\square}$	Error (residual) value	**i**	Information content
$\square^T$	Transpose operator	**k**	Sample time
F	State matrix	**m**	Mode number in the MM filtering
G	Control matrix	**p(□)**	Probability distribution function
H	Measurement matrix	**r**	Residual vector
I	Identity matrix	**sgn(□)**	Signum function
N	Sample size	**u**	Control vector
N[□]	Normal distribution	**v**	Measurement noise
E{□}	Expected value	**w**	Process noise
$\mathbf{P_{xx}}$**(□)**	Covariance	**x**	State vector
$\mathbf{P_{xy}}$**(□)**	Cross covariance	**z**	Measurement vector
$\mathbf{P_{k|k}}$	Conditional Probability	**Θ**	H_∞ performance bound
Pr[□]	Probability function	**π**	Markovian transition probability
Q	Process noise covariance matrix	**μ**	Mixing probability in the IMM filter
R	Measurement noise covariance matrix	**σ**	Variance
S	Residual covariance matrix	**γ**	Weight for the UKF computations
T	Sample rate	**ξ**	Sigma points of the UKF
Tr	Trace of a matrix	**δ**	Kronecker delta
K	Filter's gain	**ψ**	Smoothing boundary layer

Table 1
Abbreviations with their complete descriptions.

Abbreviation	Description	Abbreviation	Description
CDF	Central Difference Filter	**MLE**	Maximum Likelihood Estimator
CKF	Cubature Kalman Filter	**MM**	Multiple Models
DDF	Divided Difference Filter	**MMSE**	Minimum Mean Square Error
EHA	Electro-Hydraulic Actuator	**MSE**	Mean Square Error
EKF	Extended Kalman Filter	**PDF**	Probability Density Function
GHF	Gauss-Hermite Filter	**PF**	Particle Filter
GPB	Generalized Pseudo Bayesian	**QKF**	Quadrature Kalman Filter
H_∞	H-infinity Filter	**RKF**	Robust Kalman Filter
IMM	Interacting Multiple Model	**SPKF**	Sigma-Point Kalman Filter
KF	Kalman Filter	**SVSF**	Smooth Variable Structure Filter
LQG	Linear Quadratic Gaussian	**SVSF-VBL**	SVSF with a Variable Boundary Layer
MAP	Maximum A Posteriori	**UKF**	Unscented Kalman Filter
MCKF	Monte Carlo Kalman Filter	**VSF**	Variable Structure Filter
MKF	Mixture Kalman Filter	**WF**	Wiener-Kolmogorov filter

where the objective is to estimate $\mathbf{z}_k$ as a linear combination of the state. Note that $\mathbf{L}_k$ is a full rank weighting matrix, and in the case of directly estimating the states, it is set to $\mathbf{L}_k = \mathbf{I}$. Similar to (92), the cost function J for this system of equation is defined as [32]:

$$J = \frac{\sum_{k=0}^{N-1} \|\mathbf{z}_k - \hat{\mathbf{z}}_k\|^2_{\mathbf{S}_k}}{\|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|^2_{\mathbf{P}_0^{-1}} + \sum_{k=0}^{N-1} \left(\|\mathbf{w}_k\|^2_{\mathbf{Q}_k^{-1}} + \|\mathbf{v}_k\|^2_{\mathbf{R}_k^{-1}} \right)}, \tag{94}$$

where $\mathbf{P}_0$, $\mathbf{Q}_k$, $\mathbf{R}_k$, and $\mathbf{S}_k$ are symmetric and positive definite matrices selected by the designer based on the case under study. As discussed, the cost function J should be enforced to be less than $1/\theta$. The gain for the H_∞ estimation process may be calculated as [32]:

$$\bar{\mathbf{S}}_k = \mathbf{L}_k{}^T \mathbf{S}_k \mathbf{L}_k,\ \mathbf{K}_k = \mathbf{P}_k[\mathbf{I} - \theta\, \bar{\mathbf{S}}_k \mathbf{P}_k + \mathbf{H}_k{}^T \mathbf{R}_k{}^{-1} \mathbf{H}_k \mathbf{P}_k]^{-1} \mathbf{H}_k \mathbf{R}_k{}^{-1}. \tag{95}$$

The state vector and the state error covariance matrix may be predicted as [32]:

$$\hat{\mathbf{x}}_{k+1} = \mathbf{F}_k \hat{\mathbf{x}}_k + \mathbf{F}_k \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k),\ \mathbf{P}_{k+1} = \mathbf{F}_k \mathbf{P}_k [\mathbf{I} - \theta\, \bar{\mathbf{S}}_k \mathbf{P}_k + \mathbf{H}_k{}^T \mathbf{R}_k{}^{-1} \mathbf{H}_k \mathbf{P}_k]^{-1} \mathbf{F}_k{}^{-1} + \mathbf{Q}_k. \tag{96}$$

Note that the following condition needs to be satisfied during the state estimation process [32]:

$$\mathbf{P}_k{}^{-1} - \theta\, \bar{\mathbf{S}}_k + \mathbf{H}_k{}^T \mathbf{R}_k{}^{-1} \mathbf{H}_k > 0. \tag{97}$$

The following points are addressed by comparing the H_∞ filter with the Kalman filter [32]:

- In the H_∞ filter, matrices $\mathbf{P}_0$, $\mathbf{Q}_k$, and $\mathbf{R}_k$ are design parameters chosen by the designer based on the prior knowledge of noise, uncertainties, and the initial error. The noise and uncertainties may be nonzero mean. In the Kalman filter, noise, uncertainties and the error must be zero mean, when $\mathbf{Q}_k$, $\mathbf{R}_k$, and $\mathbf{P}_0$ are their corresponding covariance [32].
- One may assume that $\mathbf{L}_k = \mathbf{S}_k = \mathbf{I}$ in the H_∞ filter formulation. If the performance bound is set as $\theta = 0$ for estimation, then the H_∞ filter reduces to the Kalman filter. It means that the Kalman filter is a min-max filter, when the performance bound is set to $\theta = \infty$. Hence, the H_∞ filter may be considered as a robust version of the Kalman filter, but it is not optimal in the MMSE sense [32].
- The Kalman filter may become more robust by increasing $\mathbf{Q}_k$ artificially, which enlarges the covariance $\mathbf{P}_{k+1}$ and gain $\mathbf{K}_k$, alternatively. Similarly, by subtracting the term $\theta\, \bar{\mathbf{S}}_k \mathbf{P}_k$ from the H_∞ gain, it makes $\mathbf{P}_{k+1}$ and $\mathbf{K}_k$ larger. It intuitively results in increasing robustness of the H_∞ filter [32].

7. Conclusion

In this paper, an exhaustive survey of Gaussian filters for the state estimation task was provided and recent trends and developments were discussed in detail. The state estimation task was described based on the well-known Bayesian paradigm. In order to obtain a general framework for the Gaussian filter, the estimation paradigm was regenerated under the Gaussian assumption of process and measurement noise. The main Gaussian filters, presented in the literature, were then classified into several groups. Classification was based on certain characteristics that included linearity or nonlinearity of the process model, numerical integration techniques used for the state's PDF propagation, and methods for providing robustness or adaptive characteristics. The main issue common to all of the discussed filters is centered on how to properly extract the states from uncertain, inaccurate, and noisy measurements.

References

[1] A. Gelb, Applied Optimal Estimation, MIT Press, Cambridge, MA, 1974.
[2] B. Ristic, S. Arulampalam, N. Gordon, Beyond the Kalman Filter: Particle Filters for Tracking Applications, Artech House, Boston, 2004.
[3] M.S. Grewal, A.P. Andrews, Kalman Filtering: Theory and Practice Using MATLAB, Second ed, John Wiley & Sons, INC, New York, 2001.
[4] Y. Bar-Shalom, X. Rong Li, T. Kirubarajan, Estimation With Application to Tracking and Navigation, John Wiley and Sons, INC, New York, 2004.
[5] S. Haykin, Adaptive Filter Theory, Third ed, Prentic-Hall, Upper Saddle River, 1996.
[6] N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series, MIT Press, Cambridge, Massachusetts, 1964.
[7] A. Kolmogorov, Interpolation and Extrapolation of Stationary Random Sequences, Kluwer Academic Publishers, Dordrecht, 1941.
[8] H.W. Sorenson, Least-Squares Estimation: from Gauss to Kalman, IEEE Spectr. (1970) 63–68.
[9] R. Kalman, R. Bucy, New Results in Linear Filtering and Prediction Theory (March)ASME J. Basic Eng. 83 (1961) 95–108 (March).
[10] I. Arasaratnam, S. Haykin, Squre-root quadrature Kalman filtering, IEEE Trans. Signal Process. 95 (6) (2008) 953–977.
[11] I. Arasaratnam, S. Haykin, R.J. Elliott, Discete-time nonlinear filtering algorithms using gauss-hermite quadrature, Proc. IEEE 95 (5) (2007) 953–977.
[12] R. Chen, J. Liu, Mixture Kalman Filters, J. R. Stat. Soc., Ser. B 62 (3) (2000) 493–508.
[13] I. Arasaratnam, S. Haykin, Cubature Kalman filters, IEEE Trans. Autom. Control 54 (6) (2009) 1254–1269.
[14] B.D.O. Anderson, J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ, 1979.
[15] L.S. Wang, Y.T. Chiang, F.R. Chang, Filtering method for nonlinear systems with constraints, IEEE Proc., Control Theory Appl. 149 (6) (2002) 525–531.
[16] Y.C. Ho, R.C. Lee, A Bayesian approach to problems in stochastic estimation and control, IEEE Trans. Autom. Control 9 (4) (1964) 333–339.
[17] R.L. Stratonovich, On the theory of optimal non-linear filtering of random functions, Theory Probab. its Appl. 4 (1959) 223–225.
[18] R.L. Stratonovich, Application of the Markov processes theory to optimal filtering, Radio Eng. Electron. Phys. 5 (11) (1960) 1–19.
[19] T. Kailath, An innovation approach to least-squares estimation, Paert I: linear filtering in additive white noise, IEEE Trans. Autom. Control 13 (6) (1968) 646–655.
[20] Y. Chinniah, R. Burton, S.R. Habibi, Failure monitoring in a high performance hydrostatic actuation system using the extended Kalman filter, Mechatronics 16 (10) (2006) 643–653.
[21] N.J. Gordon, D.J. Salmond, A.F. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proc., Part F: Radar Sonar Navig. 140 (2) (1993) 107–113.
[22] X. Shao, B. Haung, J.M. Lee, Constrained Bayesian state estimation – a comparative study and a new particle filter based approach, J. Process Control 20 (2010) 143–157.
[23] T. Kailath, A view of three decades of linear filterng theory, IEEE Trans. Inf. Theory 20 (2) (1974) 146–181.
[24] R.E. Kalman, A new approach to linear filtering and prediction problems, ASME Trans. Basic Eng. 82 (1960) 35–45.
[25] S. A. Gadsden, "Smooth variable structure filter: theory and applications," 2011.
[26] V. Awasthi, K. Raj, A survey on the algorithms of Kalman filter and its variants in state estimation, VSRD Tech. Non-Tech. J. 2 (2) (2011) 73–88.
[27] N.A. Carlson, Fast triangular formulation of the square root filter, AIAA J. 11 (9) (1973) 1259–1265.
[28] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes: The Art of Scientific Computing, Vols. , Numerical Recipes: the, Cambridge University Press, New York, 2007.
[29] C.L. Thornton, G.J. Bierman, Givens transformation techniques for Kalman filtering, Acta Astronaut. 4 (7–8) (1977) 847–863.
[30] A. Andrews, A square root formulation of the Kalman covariance equations, AIAA J. 6 (1968) 1165–1166.
[31] M. Verhaegen, P.V. Dooren, Numerical aspects of different Kalman filter implementations, IEEE Trans. Autom. Control 31 (10) (1986) 907–917.
[32] D. Simon, Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches, Wiley-Interscience, New York, NY, 2006.
[33] T.S. Schei, A finite difference method for linearizing in nonlinear estimation algoeithms, Automatica 33 (11) (1997) 2051–2058.
[34] M. Norgaard, N.K. Poulsen, O. Ravn, New developments in state estimation of nonlinear systems, Automatica 36 (2000) 1627–1638.
[35] M. Simandl, J. Kralovec, T. Soderstorm, Advanced point-mass method for nonlinear state estimation, Automatica 42 (7) (2006) 1133–1145.
[36] D.L. Alspach, H.W. Sorenson, Nonlinear Bayesian estimation using using Gaussian sum approximation, IEEE Trans. Autom. Control 17 (4) (1972) 439–448.
[37] Y. Wu, M. Wu, X. Hu, A numerical-integration perspective on Gaussian filters, IEEE Trnasactions Signal Process. 54 (8) (2006) 2910–2921.
[38] A.H. Stroud, Gaussian Quadrature Formulas, New Jercy, Prentice Hall, 1966.
[39] S.J. Julier, J.K. Uhlmann, Unscented filtering and nonlinear estimation, Proc. IEEE 92 (3) (2004) 401–422.
[40] K. Ito, K. Xiong, Gaussian filters for nonlinear filtering problem, IEEE Trans. Autom. Control 45 (5) (2000) 910–927.
[41] J.S. Liu, Monte Carlo Strategies in Scientific Computing, Springer, 2001.
[42] H. Niederreiter, Quasi-Monte Carlo methods and pseudo-random numbers, Bull. Am. Math. Soc. 84 (1978) 957–1041.
[43] A.S.J.I.H. Sloan, Lattice Methods for Multiple Integration, Oxford University Press, Oxford, 1994.

[44] A.M.G.T. Gerstner, Numerical integration using sparse grids, Numer. Algorithms 18 (4) (1998) 209–232.
[45] J.H.K.A.P.A. Djuric, Gaussian particle filtering, IEEE Trans. Signal Process. 51 (10) (2003) 2592–2601.
[46] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 59 (2) (2002) 174–188.
[47] R. Van der Merwe, Sigma-point Kalman filters for probabilistic inference in dynamic state space models, 2004.
[48] J. Ambadan, Y. Tang, Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems, J. Atmos. Sci. 66 (2) (2009).
[49] E. A. Wan and R. Van der Merwe, "The unscented Kalman filter for nonlinear estimation," in: Proceedings of the IEEE Symposium on Adaptive Systems for Signal Processing, Communications, and Control, Lake Louise, Canada, 2000.
[50] S. Julier and J. K. Uhlmann, "A general method for approximating nonlinear transformations of probability distributions," Transformation, Robotics Research Group, Oxford University, vol. 4, no. 7, pp. 1–27, 1996.
[51] S. Julier, "The spherical simplex unscented transformation," in: Proceedings of the American Conference, 2003.
[52] S.A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, Sov. Math. Dokl. 4 (1963) 240–243.
[53] S.L. Sobolev, Formulas of mechanical cubature on the surface of a sphere, Sib. Math. (Russ.) vol. Z (3) (1981) 769–796.
[54] R. Cools, Constructing Cubature Formulas: the Science Behind the Art 6, Cambridge University Press, 1997, pp. 1–54 Acta Numerica.
[55] B. Jia, M. Xin, Y. Cheng, High-degree cubature Kalman filter, Automatica 49 (2013) 510–518.
[56] A. J. Haug, A tutorial on Bayesian estimation and tracking techniques allicable to nonlinear and non-Gaussian processes, McLean, Virginia, 2005.
[57] A. Doucet, S. Godsill, C. Andrieu, On sequential monte carlo sampling methods for Bayesian filtering, Stat. Comput. 10 (2000) 197–208.
[58] A. Doucet, J.F.G. de Freitas, N.J. Gordon, Sequential Monte Carlo Methods in Practice, Springer-Verlag, New York, 2001.
[59] B. Hassibi, On the Robustness of LMS Filters, USD, Pasadena, CA, 2003.
[60] R.F. Stengel, Optimal Control and Estimation, Wiley, Mineola, 1994.
[61] S. Haykin, Kalman filtering and Neural Network, John Wiley & Sons Inc, New York, USA, 2001.
[62] X. Sun, L. Jin, M. Xiong, Extended Kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks, Plos One J. 3 (11) (2008) 1–13.
[63] A. Rahimi, H. Alighanbari, Enhanced adaptive unscented Kalman filter for reaction wheels, IEEE Trans. Aerosp. Electron. Syst. 51 (2) (2015) 1568–1575.
[64] X. Yang, K. Xing, K. Shi, Q. Pan, Joint state and parameter estimation in particle filtering and stochastic optimization, J. Control Theory Appl. 6 (2) (2008) 215–220.
[65] N. Shimkin, Estimation and identification in dynamical systems, 2009.
[66] Ji Zhong, M. Brown, Joint state and parameter estimation for biochemical dynamic pathways with iterative extended Kalman filter: comparison with dual state and parameter estimation, Autom. Control Syst. J. 2 (2009) 69–77.
[67] M.W. Hofbar, B.C. Williams, Haybrid estimation of complex systems, IEEE Trans. Syst., Man, Cybern. 34 (5) (2004) 2178–2191.
[68] J.K. Tugnait, Detection and estimation for abrupt changing systems, Automatica 18 (5) (1982) 607–615.
[69] D.T. Magill, Optimal adaptive estimation of sampled stochastic processes, IEEE Trans. Autom. Control 10 (5) (1965) 434–439.
[70] X. Rong Li, Hybrid Estimation TechniqueControl and Dynamic Systems, Academic Press, New York, 1996, pp. 213–287.
[71] D.G. Lainiotis, Partitioning: a unifying framework for adaptive systems, I: estimation, Proc. IEEE 64 (8) (1976) 1126–1143.
[72] D. Lainiotis, Optimal adaptive estimation: structure and parameter adaptation, IEEE Trans. Autom. Control 16 (2) (1971) 160–170.
[73] P.S. Maybeck, Application of multiple model adaptive algorithms to reconfigurable flight control, Control, Dyn. Syst. 52 (1992) 291–320.
[74] P.S. Maybeck, Multiple model adaptive algorithms for detecting and compensating sensor and actuator/surface failures in aircraft flight control systems, Int. J. Robust. Nonlinear Control 9 (1999) 1051–1070.
[75] P.S. Maybeck, Stochastic Models, Estimation and Control Vol. II, Academic Press, New York, 1982.
[76] P.S. Maybeck, Stochastic Models, Estimation, and Control Vol. I, Academic Press, New York, 1979.
[77] M.J. Stepaniak, P.S. Maybeck, Model-based control redistribution applied to the Vista F-16, IEEE Trans. Aerosp. Electron. Syst. 34 (4) (1998) 1249–1260.
[78] H.A.P. Bloom, Y. Bar-Shalom, The interacting multiple model algorithm for systems with Markovian switching coefficients,", IEEE Trans. Autom. Control 33 (8) (1988) 780–783.
[79] E. Mazor, A. Averbuch, Y. Bar-Shalom, J. Dayan, Interacting multiple model method in target tracking: a survey, IEEE Trans. Aerosp. Electron. Syst. 34 (1) (1998) 103–123.
[80] X. Rong Li, Y. Bar-Shalom, A recursive multiple model approach to noise identification, IEEE Trans. Aerosp. Electron. Syst. 30 (3) (1994) 671–684 (X. Rong Li, and Y. Bar-Shalom, Vol. 30, No. 3, , 1994).
[81] X. Rong Li, Y. Bar-Shalom, Performance prediction of the interacting multiple model algorithm, IEEE Trans. Aerosp. Electron. Syst. 29 (3) (1993) 755–771.
[82] X. Rong Li and Y. Zhang, Multiple Model estimation with variable structure, part V: likely-model set algorithm, in: Proceedings of IEEE Transactions on Aerospace and Electronic Systems, vol. 36, no. 2, pp. 448–466.
[83] X. Rong Li, Z. Zhao, X.B. Li, General model-set design methods for multiple Model approach, IEEE Trans. Autom. Control 50 (9) (2005) 1260–1276 (108) (X. Rong Li, Z. Zhao, and X.B. Li).
[84] J. Luo, A. Bixby, L. Qiap and M. Kawamoto, An interacting multiple-model approach to model-based prognostics, in: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Washington, D.C., United States, 2003.
[85] C.J. Masrelize, R.D. Martin, Robust bayesian estimation for the linear model and robustifying the Kalman filter, IEEE Trans. Autom. Control 22 (3) (1977) 361–371.
[86] T.Y. Sung, Y.H. Hu, Parallel VLSI implementation of the Kalman filter, IEEE Trans. Aerosp. Electron. Syst. 23 (2) (1987) 215–224.
[87] M. Millnert, An approach to recursive identification of abruptly changing systems, in: Proceedings of the 19th IEEE Conference on Decision and Control, Albuquerque, New Mexico, United States, 1980.
[88] R.L. Moose, An adaptive state estimation solution to the maneuvering target problem, IEEE Trans. Autom. Control 20 (3) (1975) 359–362.
[89] J. Goutsias, M. Mendel, Optimal simultaneous detection and estimation of filtered discrete semi-Markov chains (J. Goutsias, and M. Mendel)IEEE Trans. Inf. Theory 34 (3) (1988) 551–568 (J. Goutsias, and M. Mendel).
[90] X. Rong Li, Multiple-model estimation with variable structure-Part II: model-set adaptation, IEEE Trans. Autom. Control 45 (11) (2000) 2047–2060.
[91] H. A. P. Bloom, An efficient filter for abruptly changing systems, in: Proceedings of the 23rd Conference on Decision and Control, Las Vegas, NV, United States, 1984.
[92] A.H. Sayed, A framework for state-space estimation with uncertain models, IEEE Trans. Autom. Control 46 (7) (2001) 998–1013.
[93] S.R. Habibi, R. Burton, The variable structure filter, ASME J. Dyn. Syst., Meas. Control 125 (3) (2003) 287–293.
[94] G. Zames, Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms and approximate inverses, IEEE Trans. Autom. Control 26 (1981) 301–320.
[95] L. Xie, C. Soh, C.E. Souza, Robust Kalman filtering for uncertain discrete-time systems, IEEE Trans. Autom. Control 39 (6) (1994) 1310–1314.
[96] C.J. Masreliez, R.D. Martin, Robust bayesian estimation for the linear model and robustifying the Kalman filter, IEEE Trans. Autom. Control 22 (3) (1977) 361–371.
[97] C.S. Hsieh, Robust two-stage Kalman filtering for systems with unknown inputs, IEEE Trans. Autom. Control 45 (12) (2000) 2347–2378.
[98] P.K. Kitanidis, Unbiased minimum-variance linear state estimation, Automatica 23 (6) (1987) 775–778.
[99] F. Wang, V. Balakrishnan, Robust Kalman filters for linear time-varying systems with stochastic parameter uncertainties, IEEE Trans. Signal Process. 50 (4) (2002) 803–813.
[100] A. Benavoli, M. Zaffalon, E. Miranda, Robust filtering through coherent lower previsions, IEEE Trans. Autom. Control 56 (7) (2011) 1567–1581.
[101] D. Bertsekas, I. Rhodes, Recursive state estimation for a set-membership description of uncertainty, IEEE Trans. Autom. Control 16 (2) (1971) 117–128.
[102] A.V. Savkin, I.R. Petersen, Robust state estimation and model validation for discrete-time uncertain systems with a deterministic description of noise and uncertainty, Automatica 34 (2) (1998) 271–274.
[103] S.R. Habibi, The smooth variable structure filter,", Proc. IEEE 95 (5) (2007) 1026–1059.
[104] S. A. Gadsden and S. R. Habibi, A new form of the smooth variable structure filter with a covariance derivation, in: Proceedings of the IEEE Conference on Decision and Control, Atalanta, United States, 2010.
[105] 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, San Francisco, United States, 2011.
[106] S.R. Habibi, The extended variable structure filter, ASME J. Dyn. Syst., Meas. Control 128 (2) (2006) 341–351.
[107] S. A. Gadsden, D. Dunne, S. R. Habibi and T. Kirubarajan, "Combined particle and smooth variable structure filtering for nonlinear estimation problems," in: Proceedings of the 14th International Conference on Information Fusion, Chicago, United States, 2011.
[108] S. A. Gadsden, M. Al-Shabi and S. R. Habibi, Estimation of an electro-hydrostatic actuator using a combined cubature Kalman and smooth variable structure filter, in: Proceedings of the ASME International Mechanical Engineering Congress and Exposition, Vancouver, Canada, 2010.
[109] S. A. Gadsden, K. McCullough and S. R. Habibi, Fault detection and diagnosis of an electro-hydrostatic actuator using a novel interacting multiple model approach, in: Proceedings of the American Control Conference, San Francisco, United States, 2011.
[110] S. A. Gadsden, M. El Sayed and S. R. Habibi, The continuous-time smooth variable structure filter, in: Proceedings of the 23rd Canadian Congress of Applied Mechanics, Vancouver, Canada, 2011.
[111] M. A. Al-Shabi, "The genral toeplitz/observability smooth variable structure filter, " 2011.
[112] S.A. Gadsden, S.R. Habibi, A new robust filtering strategy for linear systems, J. Dyn. Syst., Meas., Control 135 (1) (2013) 014503.
[113] H. H. Afshari, D. Al-Ani and S. R. Habibi, State estimation of a faulty actuator using in: Proceedings of the second-order smooth variable structure filter (the 2nd-order SVSF), in 28th IEEE Canadian Conference on Electrical and Computer Engineering, Halifax, Canada, 2015.
[114] H.H. Afshari, S.A. Gadsen, S. Habibi, Robust fault diagnosis of an electro-hydrostatic actuator using the novel optimal second-order SVSF and IMM strategies, Int. J. Fluid Power 15 (3) (2014) 181–196.

[115] H. H. Afshari, S. A. Gadsden and S. Habibi, Condition monitoring of an electro-hydrostatic actuator using the dynamic, in: Proceedings of the 2nd-order smooth variable structure filter, in ASME International Design Engineering Technical Conferences & Computers and Information, Boston, Massachusetts, USA, 2015.
[116] H. H. Afshari, D. Al-Ani and S. Habibi, Robustness analysis of some state estimation methods with an explicit consideration of modeling uncertainties, in: Proceedings of the 28th IEEE Canadian Conference on Electrical and Computer Engineering, Halifax, NS, Canada, 2015.
[117] A. Zanj, H.H. Afshari, Dynamic analysis of a complex pneumatic valve using pseudo-Bondgraph simulation technique, ASME J. Dyn. Syst., Meas., Control 135 (2013) 034502.
[118] A. Zanj, H.H. Afshari, Dynamic analysis of a complex pneumatic valve using pseudo-bondgraph simulation technique, ASME J. Dyn. Syst., Meas., Control 135 (3) (2013).
[119] H. H. Afshari, M. Attari, R. Ahmed, M. Farag and S. Habibi, Modeling, parameterization, and state of charge estimation of Li-Ion cells using a circuit model, in: Proceedings of the 2016 ITEC/IEEE Transportation Electrification Conference and Expo., Dearborn, Michigan, USA, 2016.
[120] M. Attari, H. H. Afshari and S. Habibi, Maneuvering car tracking using the interacting multiple model and the dynamic, in: Proceedings of the 2nd-order SVSF Method, in ASME International Design Engineering Technical Conferences & Computers and Information in Engineering, Boston, Massachusetts, USA, 2015.
[121] H.H. Afshari, D. Al-Ani, S. Habibi, A new adaptive control scheme based on the interacting multiple model (IMM) estimation, J. Mech. Sci. Technol. 30 (6) (2016) 2759–2767.
[122] D. Simon, From here to infinity, Embed. Syst. Program. 14 (11) (2000) 20–32.
[123] R. Banavar, A game theoretic approach to linear dynamic estimation, Austin, Texas, USA, 1992.
[124] G. Steiner, D. Watzenig, C. Magel, U. Baumgaetner, Statistical robust design using the unscented transformation, Int. J. Comput. Math. Electr. Electron. Eng. 24 (2) (2005) 606–619.