Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Approximate Gaussian Conjugacy: Parametric Recursive Filtering under Nonlinearity, Multimode, Uncertainty, and Constraint, and Beyond*

Tian-cheng LI †[1], Jin-ya SU[2,4], Wei LIU[3], Juan M. CORCHADO[1]

([1]*School of Sciences, University of Salamanca, Salamanca 37007, Spain*)

([2]*Dept. of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, UK*)

([3]*Dept. of Electronic and Electrical Engineering, University of Sheffield, Sheffield, S1 4ET, UK*)

([4]*Key Laboratory of Measurement and Control of Complex Systems of Engineering (Southeast University), China*)

†E-mail: t.c.li@usal.es; w.liu@sheffield.ac.uk; corchado@usal.es

Received June 11, 2017; Revision accepted Aug. 20, 2017; Crosschecked Aug. 22, 2017

Abstract: Since the landmark work of R. E. Kalman in the 1960s, considerable efforts have been devoted to time series state space models for a large variety of dynamic estimation problems. In particular, parametric filters that seek exact analytical estimates based on closed-form Markov-Bayes recursion, e.g., recursion from a Gaussian or Gaussian mixture (GM) prior to a Gaussian/GM posterior (termed Gaussian conjugacy in this paper), form the backbone for general time series filter design. Due to challenges arising from nonlinearity, multimode (including target maneuver), intractable uncertainties (such as unknown inputs and/or non-Gaussian noises) and constraints (including circular quantities), and so on, new theories, algorithms and technologies are continuously being developed in order to maintain, or approximate to be more precise, such a conjugacy. They have in a large part contributed to the prospective developments of time series parametric filters in the last six decades. This paper reviews the state-of-the-art in distinctive categories and highlights some insights which may otherwise be overlooked. In particular, specific attention is paid to nonlinear systems with very informative observation, multimodal systems including Gaussian mixture posterior and maneuvers, intractable unknown inputs and constraints, to fill the voids in existing reviews/surveys. To go beyond a pure review, we also provide some new thoughts on Markov-free state process modeling and filter evaluation regarding computing speed.

Key words: Kalman filter; Gaussian filter; time series estimation; Bayesian filtering; nonlinear filtering; constrained filtering; Gaussian mixture; maneuver; unknown inputs

http://dx.doi.org/10.1631/FITEE.1000000 **CLC number:**

1 Introduction

Dynamic state estimation is ubiquitous in engineering and of central interest in fields of signal/information processing and control, with a broad range of applications related to detection, positioning, monitoring, tracking, navigation, and robotics. The problem is basically concerned with estimating a latent state that evolves over time from a sequence of noisy observations in the presence of clutter, disturbances and outliers. The rapid development of physical sensors and the ever-increasing proliferation of smartphones, mobile robots and unmanned vehicles have further increased the interest in such problems.

Estimation has a long research history, although it was the Kalman filter (KF) (Kalman,

‡ Corresponding author

* This work is in part supported by the Marie Skłodowska-Curie Individual Fellowship (H2020-MSCA-IF-2015) under Grant no. 709267 and the Open Project Program of Ministry of Education Key Laboratory of Measurement and Control of CSE under Grant MCCSE2017A01

ORCID: Tian-cheng LI, http://orcid.org/0000-0002-0499-5135

1960) that thrived the field and initiated modern estimation study. Historical "giants" of estimation include Gauss (1795) and Legendre (1806) who invented the theory of least square estimation independently which anticipates most of the modern-day approaches to estimation problems, Fisher (1912) who introduced the maximum likelihood method, Kolmogorov (1940) and Wiener (1942) who established the statistical foundation for interpolation & extrapolation and filtering & prediction, and Bode and Shannon (1950) who proposed the state space model, among many others; please refer to retrospective reviews offered by (Sorenson, 1970; Grewal and Andrews, 2014; Singpurwalla *et al.*, 2017). It is the interpretation of the KF from a Bayesian prior to posterior viewpoint (Ho and Lee, 1964; Lindley and Smith, 1972) that opened the floodgate for both statisticians and engineers to advance the state of the art of filtering. Considerable efforts have since been devoted to both linear and nonlinear time series state space models raised in a wide range of applications.

However, for a general nonlinear stochastic process with very few exceptions, approximation has to be resorted to. The approximation can be parametric, non-parametric or a mixture of both. In the non-parametric case, the target probability density function (PDF) can be approximated with Monte Carlo approaches based on random sampling of which the particle filter (PF) (Arulampalam *et al.*, 2002; Cappé *et al.*, 2007; Moral, Piere Del and Doucet, Arnaud, 2014; Bugallo *et al.*, 2017) is the best known, and grid-based approaches (Šimandl *et al.*, 2006; Kalogerias and Petropulu, 2016) based on finite state space partitioning. In the parametric case, the PDF is represented by a family of functions that are fully characterized by certain parameters such as Gaussian approximation (GA) and Gaussian mixture/sum (GM/GS) filters. They are collectively referred to as parametric filters in this paper, of which moment matching to the Bayes prior and posterior is the key. They form the backbone for general time series filter design and are the focus of this survey.

There are several excellent tutorials, surveys and textbooks, primarily in the context of nonlinearity (Nørgaard *et al.*, 2000; Wu *et al.*, 2006; Crassidis *et al.*, 2007; Hendeby, 2008; Šimandl and Duník, 2009; Patwardhan *et al.*, 2012; Stano *et al.*, 2013; Morelande and García-Fernández, 2013; García-Fernández *et al.*, 2015b; Huber, 2015; Duník *et al.*, 2015; Särkkä *et al.*, 2016; Roth *et al.*, 2016; Li *et al.*, 2016b; Afshari *et al.*, 2017) or on some sub-topics such as noise covariance metrics estimation (Duník *et al.*, 2017b) and circular Bayes filtering (Kurz *et al.*, 2016). However, some important parts have not been addressed or only addressed briefly in those reviews, including

- A unifying framework to analyze the common essences of different filters,
- Very informative observation (i.e., the observation noise is insignificant),
- Multimodal system, and
- Intractable uncertainties and constraints (especially the state of the art and classification).

These topics will form the key part of our review, complementing existing works. To minimize overlap with these studies, common contents will not be addressed. We base our review on a transparent and concise framework termed *approximate Gaussian conjugacy* (AGC). That is, all reviewed works arguably aim at maintaining, or approximating to be more precise, *a closed-form Markov-Bayes recursion from a GA/GM prior to a GA/GM posterior*, to deal with the challenges due to nonlinearity, multimode, intractable uncertainty and constraint. By doing so, different efforts are organized along the same line. To go beyond a pure review, we also include discussions on alternatives to hidden Markov model (HMM) and on filter evaluation regarding computing speed, with our new thoughts. All of these strive to give a concise overview of the state of the art as well as shed some light on future development in this area.

The remainder of the paper is organized as follows. Section 2 introduces the three essential concepts behind the sequential Bayesian inference (SBI) consisting of the Markov-Bayes recursion, Cramér-Rao lower bound, and Gaussian conjugacy. Sections 3, 4, 5 and 6 review the challenges and solutions for parametric filtering under nonlinearity, multimode (including GM filtering and target maneuver), intractable uncertainties (including unknown and/or non-Gaussian inputs/noise) and constraints on the state or observation (including circular quantities), respectively. Some new thoughts on HMM and filter evaluation are presented in Section 7 before we conclude in Section 8.

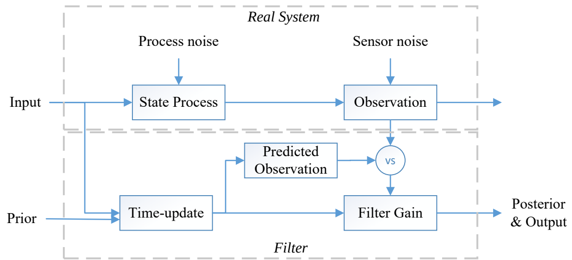


Fig. 1 Block-diagram of the evolution of a recursive filter of the prediction-updating format

2 Basis of SBI

2.1 Markov-Bayes recursion

The time-series (a.k.a. sequential) Bayesian inference is carried out by constructing the posterior PDF of the latent state based the observation series and the a prior model knowledge of the system. Using the posterior distribution, one can make state inference, e.g., finding the value that maximizes the posterior, namely maximum a posteriori (MAP) estimation, or the value that minimizes a cost function, e.g., mean square error (MSE) as is done in the KFs.

To be more specific, in the Markov-Bayes setting, the state process is assumed to follow a first order Markov process (a.k.a, HMM) and the observations are conditionally independent given the states. This leads to a discrete-time state space model (SSM) with additive noises

$$\mathbf{x}_t = f_t(\mathbf{x}_{t-1}) + \mathbf{u}_t + \mathbf{v}_t \quad (1)$$

$$\mathbf{y}_t = h_t(\mathbf{x}_t) + \mathbf{w}_t \quad (2)$$

where $t \in \mathbb{N}$ indicates the time instant, $\mathbf{x}_t \in \mathbb{R}^{d_\mathbf{x}}$, $\mathbf{y}_t \in \mathbb{R}^{d_\mathbf{y}}$ and $\mathbf{u}_t \in \mathbb{R}^{d_\mathbf{x}}$ denote the state, the observation (also called measurement) and the input, respectively, and $\mathbf{v}_t \in \mathbb{R}^{d_\mathbf{x}}$, $\mathbf{w}_t \in \mathbb{R}^{d_\mathbf{y}}$ denote the noises affecting the state function f_t and the observation function h_t, respectively. Except for $\mathbf{y}_t$, all the other quantities can be unknown and if so may need to be estimated jointly with the state $\mathbf{x}_t$. The state process model (1) shall be written in a differential form for the continuous time case. Furthermore, as a rare case, the observation process (2) can also be modeled as a continuous process (Ghoreyshi and Sanger, 2015). However, we will not distinguish this in this paper.

SBI basically consists of one-step forward prediction $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$, filtering $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, and smoothing $p(\mathbf{x}_t|\mathbf{y}_{1:T})$, where $t < T, \mathbf{y}_{1:t} \triangleq \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t\}$. Here we focus on the filtering distribution. The filtering recursion is given by performing prediction and correction recursively. The prediction step combines the previous filtering distribution $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ with the state transition $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1})$ (i.e., Chapman-Kolmogorov equation) as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \quad (3)$$

This forms the prior distribution (called *the prior* hereafter) that is a one-step forecast of the state. Next, given a new observation $\mathbf{y}_t$, the prior will be updated by the Bayes' rule as follows, resulting in the Bayes posterior distribution (called *the posterior* hereafter)

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} \quad (4)$$

where $p(\mathbf{y}_t|\mathbf{x}_t)$ is the likelihood function.

Given the posterior in (4), the expectation of the state $\mathbf{x}_t$ conditioned on all the observations $\mathbf{y}_{1:t}$, namely the expected a posteriori (EAP) estimate, is given by

$$\hat{\mathbf{x}}_t^{\mathrm{EAP}} \triangleq \mathrm{E}\big[\mathbf{x}_t|\mathbf{y}_{1:t}\big] = \int \mathbf{x}_t p(\mathbf{x}_t|\mathbf{y}_{1:t})d\mathbf{x}_t \quad (5)$$

which also gives the minimum MSE (MMSE) estimation, of optimality defined on the second-order statistics. Alternatively, the MAP estimate (García-Fernández and Svensson, 2015) is given by

$$\hat{\mathbf{x}}_t^{\mathrm{MAP}} \triangleq \arg\max_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{y}_{1:t}) \quad (6)$$

Different from the prevailing MSE criterion, it might be of interest to base the lost function on some other criteria, such as the maximum correntropy criterion (MCC) (Liu *et al.*, 2007) which has the advantages to handle impulsive non-Gaussian noise thanks to using higher-order statistics information. Correspondingly, a new class of linear KFs (Wu *et al.*, 2015; Izanloo *et al.*, 2016; Chen *et al.*, 2017), have been developed. More generally, there are cases when robustness (i.e., adaptability to outliers, system errors and disturbances, etc.) is more preferable than optimality which will lead to various robust filtering algorithms; see Section 5.5.

Without loss of generality, one typical iteration process of a recursive filter can be illustrated as shown in Fig. 1. One of the main reasons for the popularity of HMMs is the friendly first order assumption that states are conditionally-independent given the previous state. This facilitates forward-backward inference for model learning and parameter estimation but also severely limits the temporal dependencies that can be modeled. Some alternatives will be presented in Section 7.

2.2 Bayesian Cramér-Rao lower bound

It is theoretically pivotal to derive performance bounds on estimation errors when estimating parameters of interest in a given model, as well as developing estimators to achieve these limits. When the parameters to be estimated are deterministic, a popular approach is to bound the MSE achievable within the class of unbiased estimators. The Cramér-Rao lower bound (CRLB), given by the inverse of the Fisher information matrix, provides the optimum performance for any unbiased estimator of a fixed parameter on the variance of estimation error; see Table 1. However, it is necessary to note that,

Highlight 1. *CRLB limits only the variance of unbiased estimators and lower MSE can be obtained by allowing for a bias in the estimation, while ensuring that the overall estimation error is reduced* (Stoica and Moses, 1990; Eldar, 2008).

Van Trees presented an analogous MSE bound for a random parameter, the posterior CRLB (Van Trees, 1968), which is also referred to as the Bayesian CRLB (BCRLB). Furthermore, an elegant recursive approach was developed (Tichavsky *et al.*, 1998) to calculate the sequential BCRLB based on the posterior distribution for a general discrete-time nonlinear filtering problem that avoids Gaussian assumptions. However, in general, BCRLB has no closed-form expressions in nonlinear systems. As such, a large body of alternative Bayesian bounds has been proposed (Van Trees and Bell, 2007; Zuo *et al.*, 2011; Zheng *et al.*, 2012; Fritsche *et al.*, 2016).

On BCRLB, there are two points worth noting. First, the *unconditional* BCRLB is determined only by the system dynamic model, system observation model and the prior knowledge regarding the system state at the initial time, and is thus independent of any specific realization of the system state. A variant of the CRLB for *constrained estimation* problems was derived in (Gorman and Hero, 1990). Since more information about the parameter is incorporated into the estimator, the constrained CRLB can be lower than the unconstrained version. Some attempts have been made to include the information obtained from observations by incorporating the tracker's information into the calculation of the BCRLB; please refer to (Zuo *et al.*, 2011; Fritsche *et al.*, 2016) and the references therein for details.

Second, in the Bayesian setting, both the state and observation sequences are random quantities on which the CRLB/BCRLB is based. However, in the majority of practical setups particularly in the context of tracking, positioning and localization, only a single state sequence is of interest, such as a trajectory of an aircraft or a ground vehicle. In these situations, the estimator performance shall be evaluated based on the MSE matrix conditioned on a specific state sequence, for which the general BCRLB does not provide a lower bound (Fritsche *et al.*, 2016). Instead, it was shown that

Highlight 2. *"The KF is conditionally biased with a non-zero process noise realization in the given [deterministic] state sequence and is not an efficient estimator in a conditional sense, even in a linear and Gaussian system."*

2.3 Gaussian conjugacy

Some important properties of the Gaussian distribution are notable. Given only the first two moments, the Gaussian distribution makes the least assumptions about the true distribution in the maximum entropy sense and minimizes the Fisher information over the class of distributions with a bounded variance (Kim and Shevlyakov, 2008). As a general example, letting θ denote the parameter vector, $\mathbf{w}$ the noise and $\mathbf{y} = \mathbf{x}_\theta + \mathbf{w}$ the random observation model, we have the following property (Stoica and Babu, 2011; Park *et al.*, 2013).

Highlight 3. *Among all possible distributions of the observation noise* $\mathbf{w}$ *with a fixed covariance matrix, the CRLB for* $\mathbf{x}$ *attains its maximum when* $\mathbf{w}$ *is Gaussian, i.e., the Gaussian scenario is the "worst-case" for estimating* $\mathbf{x}$.

More importantly, the Gaussian variable is self-conjugate. That is, if the likelihood function is Gaussian, choosing a Gaussian/GM prior over the mean will ensure that the posterior distribution is also Gaussian/GM without using any approximation; we refer to this as strict Gaussian conjugacy in this paper. In addition, the inverse Wishart distribution provides a conjugate prior for the covariance matrix of a Gaussian distribution with known mean, termed Gaussian inverse Wishart (GIW). Please refer to (Murphy, 2007) for more conjugate priors related to Gaussian distribution.

Based on conjugate prior, the Bayes prior and posterior can be computed in a closed form. More precisely, since the Gaussian PDF is uniquely determined by its first moment (mean) and the second moment (covariance), the Gaussian conjugacy will render recursive computations of the Bayes prior and posterior in the simple manner of recursive algebraic computing the mean and covariance of the conditional PDFs, namely moment matching. Such a conjugacy is very engineering-friendly especially when computing time is considered (see our discussion in Section 7.2) and forms the essence for the sequential closed-form recursion.

The strict Gaussian conjugacy, however, requires both the state transition function f_t, and the observation function h_t to be linear , the inputs $\mathbf{u}_t$ and the noises $\mathbf{v}_t$ and $\mathbf{w}_t$ unconditionally/white Gaussian/GM (independent of the state). Then, the optimal, conjugate solution is given by the KF (or a mixture of KFs in case of GM filtering), as shown in Table 2. Any violation of these requirements will lead to a non-Gaussian/GM posterior and destroy the closed-form Gaussian recursion. Also, all the parameters need to be known a priori. These requirements are fastidious and unrealistic in most realistic systems. In order to retain the AGC, approximation has to be applied for easing the challenge from nonlinearity (regarding both functions f_t and h_t), multimodal posterior, and intractable system uncertainties (primarily regarding noises $\mathbf{v}_t$ and $\mathbf{w}_t$ and input $\mathbf{u}_t$) and constraints, which will be addressed in the following Sections 3, 4, 5 and 6, respectively.

3 Nonlinearity

Nonlinearity appearing in the system functions forms a pivotal and explicit challenge to the Gaussian conjugacy simply because a Gaussian distribution after nonlinear transformation will be no more Gaussian. A considerable number of approximation approaches have been developed to account for nonlinearity, which can be primarily classified into two categories, approximating either the nonlinear function or the nonlinear-transformed PDFs. The former, with typical examples of extended KF (EKF), modal KF (Mohammaddadi *et al.*, 2016), divided difference filter (Nørgaard *et al.*, 2000), and Fourier-Hermite KF (Sarmavuori and Särkkä, 2012) seeks functions' approximation using polynomial expansions (e.g. Taylor series, Fourier-Hermite series, Stirling's interpolation, or Modal series). The latter one, with representative examples of unscented KF (UKF) (Julier and Uhlmann, 2004), Gauss-Hermite filter and central difference filter (Ito and Xiong, 2000), cubature KF (CKF) (Arasaratnam and Haykin, 2009; Jia *et al.*, 2013), sparse-grid quadrature filter (Arasaratnam and Haykin, 2008; Jia *et al.*, 2012), stochastic integration filter (Duník *et al.*, 2013), and iterated posterior linearization filter (IPLF) (García-Fernández *et al.*, 2015b; Raitoharju *et al.*, 2017), is based on a set of deterministically chosen weighted sigma points. It was shown that many sigma-point methods can be interpreted as Gaussian quadrature based methods (Särkkä *et al.*, 2016). They calculate the posterior PDF using a direct numerical approximation in a local sense, and are therefore also referred to as the local approach. An alternative to deterministic sampling for approximating an arbitrary PDF is random sampling, e.g., the popular mixture KF (Chen and Liu, 2000), ensemble KF (Evensen, 2003; Roth *et al.*, 2017b), Monte Carlo KF (Song, 2000), and Gaussian/GS PF (Kotecha and Djurić, 2003a,b), which still strives to maintain AGC. This allows asymptotically exact integral evaluation, albeit with much higher computational complexity. These approaches carries out numerical approximation in a global sense like the PF and so are also referred to as the global approach.

All of these GA filters have triggered tremendous further developments. For instance, the UKF has perhaps gained the most approval in the community whereas it may suffer from numerical instability (e.g., may have a negative weight for the center point) (Arasaratnam and Haykin, 2009; Jia *et al.*, 2013), systematic error (Duník *et al.*, 2013), and nonlocal sampling problem for high-dimensional applications (Chang *et al.*, 2013). These problems, together with parameter setting strategies (Straka

Table 1 Fisher information and Cramér-Rao inequality

Given that an unknown (random) parameter $\mathbf{x}$ is observed as $\mathbf{y}$ with likelihood function $p(\mathbf{y}|\mathbf{x})$, the second moment of the partial derivative with respect to $\mathbf{x}$ of the natural logarithm of the likelihood function is called the *Fisher information* for $\mathbf{x}$ contained in $\mathbf{y}$, i.e.,

$$I(\mathbf{x}) \triangleq \mathrm{E}\left[\left(\frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{y}|\mathbf{x})\right)^2 |\mathbf{x}\right] = \int \left(\frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{y}|\mathbf{x})\right)^2 p(\mathbf{y}|\mathbf{x})\, d\mathbf{y} \tag{7}$$

where, for any given value of $\mathbf{x}$, the expression $\mathrm{E}[...|\mathbf{x}]$ denotes the conditional expectation over values for $\mathbf{y}$ with respect to the probability function $p(\mathbf{y}|\mathbf{x})$ given $\mathbf{x}$, and $\frac{\partial}{\partial \mathbf{x}} f$ is the derivative of function f with respect to x. Note that $0 \leq I(\mathbf{x}) < \infty$. For any unbiased estimator $\hat{\mathbf{x}}(\mathbf{y})$, the Cramér-Rao inequality is given by

$$\mathrm{Var}(\hat{\mathbf{x}}(\mathbf{y})) \geq \frac{1}{I(\mathbf{x})} \tag{8}$$

In statistics, it is

$$\mathrm{MSE}(\hat{\mathbf{x}}(\mathbf{y})) \geq \frac{1}{I(\mathbf{x})} \tag{9}$$

or more precisely, $\mathrm{MMSE}(\hat{\mathbf{x}}(\mathbf{y})) = \frac{1}{I(\mathbf{x})}$.

Table 2 Closed-form recursion of Kalman filtering

Given that in (1)-(2) both the state transition function f_t, and the observation function h_t are linear, the input $\mathbf{u}_t$ is known, and the noises $\mathbf{v}_t$ and $\mathbf{w}_t$ are unconditionally Gaussian, the SSM (1)-(2) can be rewritten as

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{v}_t - \mathrm{E}[\mathbf{v}_t] \tag{10}$$

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{w}_t - \mathrm{E}[\mathbf{w}_t] \tag{11}$$

where the input $\mathbf{u}_t$ and the mean of the noise $\mathbf{v}_t$ are included in the function f_t, yielding a new linear transition function $\mathbf{F}$; the remaining part of noise $\mathbf{v}_t$, namely $\mathbf{v}_t - \mathrm{E}[\mathbf{v}_t]$, can be treated as zero-mean noise with covariance $\mathbf{Q}_t = \mathrm{E}\left[\left(\mathbf{v}_t - \mathrm{E}[\mathbf{v}_t]\right)\left(\mathbf{v}_t - \mathrm{E}[\mathbf{v}_t]\right)^{\mathrm{T}}\right]$. Similarly, the mean of the noise $\mathbf{w}_t$ can be integrated into the observation function h_t, leading to a new observation function $\mathbf{H}$ and then the remaining part of noise $\mathbf{w}_t$, namely $\mathbf{w}_t - \mathrm{E}[\mathbf{w}_t]$, can be treated as zero-mean noise with covariance $\mathbf{R}_t = \mathrm{E}\left[\left(\mathbf{w}_t - \mathrm{E}[\mathbf{w}_t]\right)\left(\mathbf{w}_t - \mathrm{E}[\mathbf{w}_t]\right)^{\mathrm{T}}\right]$. For this formulation, the prediction-correction steps of the KF are given as follows.

Prediction (time updating):

$$\hat{\mathbf{x}}_{t|t-1} = \int \mathbf{F}\mathbf{x}_{t-1} \mathcal{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1})\, d\mathbf{x}_{t-1} \tag{12}$$

$$\mathbf{P}_{t|t-1} = \int \mathbf{F}\mathbf{x}_{t-1}(\mathbf{F}\mathbf{x}_{t-1})^{\mathrm{T}} \mathcal{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1})\, d\mathbf{x}_{t-1} - \hat{\mathbf{x}}_{t|t-1}\hat{\mathbf{x}}_{t|t-1}^{\mathrm{T}} + \mathbf{Q}_t \tag{13}$$

where $\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P})$ denotes the Gaussian PDF with mean $\hat{\mathbf{x}}$ and covariance $\mathbf{P}$.

Correction (data updating)

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{G}_t\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right) \tag{14}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{G}_t \mathbf{P}_{yy} \mathbf{G}_t^{\mathrm{T}} \tag{15}$$

where,

$$\mathbf{G}_t = \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1} \tag{16}$$

$$\hat{\mathbf{y}}_{t|t-1} = \int \mathbf{H}\mathbf{x}_t \mathcal{N}\left(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}\right) d\mathbf{x}_t \tag{17}$$

$$\mathbf{P}_{xy} = \int \left(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}\right)\left(\mathbf{H}\mathbf{x}_t - \hat{\mathbf{y}}_{t|t-1}\right)^{\mathrm{T}} \mathcal{N}\left(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}\right) d\mathbf{x}_t \tag{18}$$

$$\mathbf{P}_{yy} = \int \left(\mathbf{H}\mathbf{x}_t - \hat{\mathbf{y}}_{t|t-1}\right)\left(\mathbf{H}\mathbf{x}_t - \hat{\mathbf{y}}_{t|t-1}\right)^{\mathrm{T}} \mathcal{N}\left(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}\right) d\mathbf{x}_t + \mathbf{R}_t \tag{19}$$

et al., 2014; Zhang *et al.*, 2015; Scardua and Cruz, 2017) and constrained filtering (see Section 5), have led to ever-increasing further research developments for deterministic sampling-based filtering. Meanwhile, the degree of nonlinearity or non-Gaussianity has also been well investigated and various measures have been developed (not limited to state estimation); see the review and discussion offered by (Liu and Li, 2015) and (Duník *et al.*, 2016). This provides a principle to select a nonlinear filter from many according to the property of the problem.

In order to better exploit the information about the state from the same measurement sequence, different local filters that extract different portions of the system information can be employed to linearize the same nonlinear functions and the results combined for better accuracy. This is called the *cooperative* local (or Gaussian) filter design approach (Duník *et al.*, 2017a), which resembles the idea of *multiple conversion approach* (Lan and Li, 2017) where jointly utilizes multiple nonlinear filters based on a weighted sum of several sub-functions of the (same) measurement. Accuracy benefit to do so is as foreseeable the challenge to computing complexity; see our discussion in Section 7.2.

While general nonlinear filtering has been well elaborated and reviewed from various viewpoints, we focus on two interesting subtopics.

3.1 Converted measurement filtering

The unconditional noise requirement (that is, the noises are white and independent of the state) may not be met strictly in practice. This relaxation is particularly useful when the state model is linear and Gaussian while the measurement model is nonlinear but can be converted to a linear one, namely injective. Although converting the nonlinear measurement to the state space yields a non-Gaussian uncertainty for sure, the system will become linear enabling the use of a linear filter, namely converted measurement filtering (CMF). It was first introduced in (Lerro and Bar-Shalom, 1993). The state of the art (Liu *et al.*, 2013; Lan and Li, 2015) has demonstrated that proper "uncorrelated conversion" of the nonlinear measurement can make further use of the measurement information and thereby can be used to augment the filter rather than use original measurement only. This leads to an updating protocol that resembles "repeated" use of the measurement by linear combination of the original measurement and the nonlinearly converted measurement. However, it was further pointed out in (García-Fernández *et al.*, 2015a) that CMF works better particularly for informative systems but not for the non-informative system that has large measurement noise variance. Therefore, an interacting mechanism is advocated to switch between an unscented linear CMF and a normal unscented nonlinear filter.

Obviously, nonlinear conversion will lead to (pseudo-measurement) noise that is state dependent and non-Gaussian, even the original noise is state independent and white Gaussian. Therefore, a critical issue involved is to determine the unbiased mean and covariance of the observation noise after converting (Bordonaro *et al.*, 2014; Lan and Li, 2015), entailing correct moment matching. A review and comparison of algebraic approaches for the Gaussian noise related debiasing was delivered in (Bordonaro *et al.*, 2014). To handle originally non-Gaussian noises, Monte Carlo sampling can be used for general conversion (Li *et al.*, 2016). However, we note in many cases the measurement model is non-injective, e.g., a bearing observation of the target in the planar space, preventing CMF unless multiple sensors are used jointly to make the observation (in the form of observation matrix) determined or over determined (Li *et al.*, 2017b).

When the noise is multiplicative, namely dependent on the state, the conversion will need knowledge of the state. For example, a maximum likelihood estimator is used in (Wang *et al.*, 2012) to remove the distance-sensing nonlinearity in case of hybrid additive and multiplicative noises.

3.2 Very informative observation

Dramatically fast and ever-increasing escalation has been seen on computers and sensors including radar, camera, sonar and so on. Incredibly, Moore's law proved accurate for several decades in the semiconductor industry. It is fair to say, what we have today is totally different to that when Kalman invented the KF. Clearly, either super-quality sensors or high-dimensional observations due to the joint use of multiple/massive moderate sensors, are supposed to remarkably benefit our estimation by providing very informative observation (VIO) of the system. Unfortunately, advanced KFs may not always outperform the basic KF in such cases. But instead,

Table 3 A VIO SSM

The state process function and the observation function are given as follows, respectively (Van Der Merwe *et al.*, 2000)

$$x_t = 1 + \sin(0.04\pi t) + 0.5x_t + v_t \tag{20}$$

$$y_t = \begin{cases} 0.2x_t^2 + w_t, & \text{if} t \leq 30 \\ 0.5x_t - 2 + w_t, & \text{if} t > 30 \end{cases} \tag{21}$$

where the process noise u_t is a gamma random variable $\Gamma(3, 2)$ and the observation noise is Gaussian $v_t \sim \mathcal{N}(v; 0, 0.00001)$ and the default simulation length is 60 iteration steps. For this SSM, the observation-only inference (without debiasing) is given as

$$x_t^{\mathrm{O}_2} = \begin{cases} \sqrt{5y_t}, & \text{if } t \leq 30 \quad \text{(biased)} \\ 2y_t + 4, & \text{if } t > 30 \quad \text{(unbiased)} \end{cases} \tag{22}$$

it turns out that (Morelande and García-Fernández, 2013; García-Fernández *et al.*, 2015b):

Highlight 4. *"For sufficiently precise measurements, none of the KF variants, including the KF itself, are based on an accurate approximation of the joint density. Conversely, for imprecise measurements all KF variants accurately approximate the joint density, and therefore the posterior density. Differences between the KF variants become evident for moderately precise measurements."*

Therefore, seeking increasingly accurate approximations of the KF can be of limited benefit in a VIO system. Instead, a SBI filter may just lose to the observation-only (O2) inference that directly converts the observation to the state space (Li *et al.*, 2016) and is immune to any state process modeling error/bias. As shown in Table 2, the basic formulation of the KF omits any bias (whether due to mis-modeling, intractable noises, disturbances or over approximation) propagated in the prior, which is naive at best to be true in real world problems. Indeed, SBI becomes very sensitive to any bias in the prior in a VIO system, which is the key factor leading to the defeat. Please refer to the quantitative results given in Section 4.2 of (Li *et al.*, 2016).

A VIO SSM is given in Table 3, which was originally proposed in (Van Der Merwe *et al.*, 2000) and has since been widespread for filter test. For this example, the computationally extremely fast O2 inference can beat either EKF/UKF, unscented PF and so on by orders of magnitude in both accuracy and computational speed! As such, prominent attention is desired nowadays as sensors are deployed with gradually increased quantity (higher precise) or quality (joint use of massive sensors) (Li *et al.*, 2017b,c), which popularizes VIO in reality. However, we have the following note (Li *et al.*, 2016, 2017b):

Highlight 5. *While the BCRLB sets a best line (in the sense of MMSE) that any unbiased sequential estimator can at maximum achieve, the O2 inference sets the bottom line that any "effective" estimator shall at minimum achieve.*

We would like to further add that "in nonlinear systems, due to inevitable approximations, estimators generally exhibit a bias, and thus the BCRLB actually cannot be achieved by any unbiased estimator." (Fritsche *et al.*, 2016)

As a compromise, iterative algorithms may be applied to repeatedly leverage the informative observation. The first iterated EKF (IEKF) (Jazwinski, 1970, pp. 349-351) implemented the first-order Taylor series expansion (TSE) of the observation function repeatedly for posterior updating to avoid filtering divergence due to the once first-order TSE truncation. It produces a sequence of mean estimates, which was shown (Bell and Cathey, 1993) to be equivalent to the Gauss-Newton (GN) algorithm for computing the MAP estimate. IEKF performs well when the true posterior is close to being Gaussian, but convergence of the GN algorithm is not guaranteed. Furthermore, a generalized iterated KF (Hu *et al.*, 2015) for nonlinear stochastic discrete-time estimation with state-dependent observation noise, adopts the Newton-Raphson iterative optimization steps yielding an approximate MAP estimate of the states. Of high relevance, the IPLF (García-Fernández *et al.*, 2015b; Raitoharju *et al.*, 2017) uses statistical linear regression instead of the first order TSE for better linearization and iterated a posterior estimate updating.

More implementations for iterated/repeated observation (or its conversion) updating have been realized on different Gaussian filters, e.g., (Zhan and Wan, 2007; Zanetti, 2012; Steinbring and Hanebeck, 2014). This has a very close connection to the concept of progressive Bayes (Hanebeck *et al.*, 2003), which strives to apply the Bayes updating in a progressive manner, and the aforementioned uncorrelated augmentation (Liu *et al.*, 2013; Lan and Li, 2015, 2017). In fact, the idea of emphasizing the observation when it is very informative has also inspired the development of random sampling based filters such as annealed/unscented PFs (Van Der Merwe *et al.*, 2000; Godsill and Clapp, 2001), particle flow filter (Daum and Huang, 2010) (and relevantly Gaussian flow (Nurminen *et al.*, 2017)) and feedback PF (Yang *et al.*, 2016) and some (re)sampling approaches (Li *et al.*, 2015) (Li *et al.*, 2015). In recent years, there has been a burgeoning passion and interest in applying similar ideas and techniques to Bayesian filtering for informative systems; see also (Mitter and Newton, 2003; Ma and Coleman, 2011) for other attempts.

However, a rigorous criterion to determine the optimal number of observation updating iterations seems still missing. In existing works, the convergence is primarily identified by monitoring the Kalman gain as compared with a specified ad-hoc threshold. More importantly, when the observation is not so informative, it turns out to be a bad idea to emphasize on the observation, as quantitatively demonstrated in (Li *et al.*, 2016). Therefore, particular caution should be exercised.

4 Multimode

4.1 Gaussian mixture

Based on the Wiener approximation theorem, any distribution can be expressed as, or approximated sufficiently well by, a finite sum of known Gaussian distributions, called GM. Mixture distribution may arise from stochastically switched Gaussian systems (such as the maneuvering dynamics as addressed in the next subsection), systems with multimodal state (e.g., concurrent multiple targets), multimodal observation (e.g., radar observations often exhibit bimodal properties due to secondary radar reflections), or systems with long-tailed stochastic behavior or noise (see Section 5 for intractable uncertainties), to name a few.

The posterior (4) in the manner of a GM of M components can be written as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \sum_{i=1}^{M} \omega_i \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_t^{(i)}, \mathbf{P}_t^{(i)}) \qquad (23)$$

where $\omega_i > 0$ is the weight of the ith Gaussian component which satisfies $\sum_{i=1}^{M} \omega_i = 1$ in general but not in the finite set statistics-based multi-target intensity cases (Mahler, 2014; Vo and Ma, 2006).

Assuming that the noise sequences have a uniformly convergent series expression in terms of known Gaussian distributions, a number of Gaussian terms with known moments can be used to develop a MMSE filtering algorithm, namely Gaussian mixture filtering (GMF) (Sorenson and Alspach, 1971; Faubel *et al.*, 2009; Ali-Loytty, 2010). Each Gaussian component may be updated based on different nonlinear filter updating rules. For linear dynamic systems with GM noises, GMF provides the MMSE state estimate by tracking the GM posterior. The analytic lower and upper MMSE bounds of linear dynamic systems with GM noise statistics were analyzed in (Pishdad and Labeau, 2015). It has been shown that for highly multimodal GM noise distributions, the bounds and the MMSE will converge and relevant statistics like mean or covariance can be derived in a closed form. In addition, to take system constraints into account, projection based GM-UKF (Ishihara and Yamakita, 2009), GMF (Duník *et al.*, 2010) and density truncation based GM-UKF) (Straka *et al.*, 2012) have been developed. Constrained filtering will be addressed separately in Section 6.

Obviously, the mixture size lies in the core of the trade-off between computing efficiency and filter accuracy. Many either sophisticated or straightforward algorithms have been proposed for adapting/reducing the number of components in the GM. For an adaptive GM, two different approaches have been proposed: adapting the weight of each Gaussian component by minimizing the propagation error committed in the GM approximation (Ito and

Xiong, 2000; Terejanu *et al.*, 2011) and splitting the Gaussian components during the propagation based on nonlinearity induced distortion (DeMars *et al.*, 2013). Both require online optimizations which, however, will add to the overall computational cost. Instead, straightforward mixture reduction (MR) is more practically useful, which is typically realized in the manner of GM *merging* and *pruning*.

The first systematic GM merging scheme was established in (Salmond, 1990), which is perhaps still the most widely used protocol (Faubel *et al.*, 2009; Ali-Loytty, 2010) and can be interpreted as a special type of conservative fusion of components (Reece and Roberts, 2010). A survey has been provided in (Crouse *et al.*, 2011) for more advanced MR solutions with comparison including West's algorithm (West, 1993), Runnalls' Kullback-Leibler reduction algorithm (Runnalls, 2007), constraint optimized weight adaptation (Chen *et al.*, 2010), and MR via clustering (Schieferdecker and Huber, 2009).

A general principle for MR is to minimize the discrepancy between the original and the reduced mixtures, for which two typical metrics are integral square error (ISE) and Kullback-Leiber divergence (KLD). The KLD of the GM-PDF before MR $p(\mathbf{x})$ from that of after MR $q(\mathbf{x})$, denoted $D_{\mathrm{KL}}(p||q)$, is an asymmetric measure of the information lost when $q(\mathbf{x})$ is used to approximate $p(\mathbf{x})$, which is given as

$$\begin{aligned} D_{\mathrm{KL}}(p||q) &\triangleq \int p(\mathbf{x})\ln\frac{p(\mathbf{x})}{q(\mathbf{x})}d\mathbf{x} \\ &= \int p(\mathbf{x})\ln p(\mathbf{x})d\mathbf{x} - \int p(\mathbf{x})\ln q(\mathbf{x})d\mathbf{x} \end{aligned} \tag{24}$$

As the first term completely relies on the PDF before MR, minimizing the KLD in (24) is equivalent to maximizing the second $\int p(\mathbf{x})\ln q(\mathbf{x})d\mathbf{x}$. The KLD is not a distance since it is not symmetric. The ISE is a nonparametric distance which is given as, e.g., between $p(\mathbf{x})$ and $q(\mathbf{x})$,

$$D_{\mathrm{ISE}}(p||q) \triangleq \int (p(\mathbf{x}) - q(\mathbf{x}))^2 d\mathbf{x} \tag{25}$$

The ISE approach was first proposed for MR in the context of multiple hypothesis tracking in (Williams and Maybeck, 2006), which inspired further development (Chen *et al.*, 2010) and the normalized ISE (Petrucci, 2005). One distinctive feature of the method is the availability of exact analytical expressions for GMs. However, the cost function (25) is a complicated multimodal function with many local minima; hence gradient-based methods cannot guarantee convergence to the global minimum, unless the initialization point happens to be close to the global minimum (Williams and Maybeck, 2006). In contrast, the Kullback-Leibler reduction method (Runnalls, 2007) minimizes an upper bound on the KLD between the original mixture and the reduced mixture, which appears to perform better in terms of slimming the GM, and has led to several further developments (Schieferdecker and Huber, 2009; Ardeshiri *et al.*, 2015). Furthermore, a model order reduction procedure is proposed in (Raitoharju *et al.*, 2017) for minimizing the KLD of the reduced order density from the original density.

In contrast to the above MR schemes that gradually reduce the mixture to a desired size via merging and pruning, the algorithm given in (Huber and Hanebeck, 2008) gradually adds new components to a mixture starting from a single component. This method however could be beaten in terms of the ISE by simpler approaches based on clustering (Schieferdecker and Huber, 2009). Note that, MR is closely connected to many multi-hypothesis based approaches such as multi-hypothesis tracker (Reece and Roberts, 2010) and GIW mixtures (Granström and Orguner, 2012). It has been applied to distributed information fusion for consensus (Li *et al.*, 2017c,b), beyond the original centralized filtering.

4.2 Maneuver

Maneuver is an important concept particularly in the context of target tracking, which generally refers to time varying target dynamical mode/model. Maneuvering target tracking (MTT) is essentially a hybrid estimation problem consisting of continuous-state (base state) estimation and discrete-state (mode) decision. The basic framework to describe the maneuvering state dynamics is the so-called *jump Markov system* (JMS), in which the target dynamical model switches/jumps from one HMM to another. Simply put, there are two primary types of MTT methods: the single-model (SM) method (also called decision-based method (Li and Jilkov, 2002; Zhou and Frank, 1996) and the multiple-model (MM) method (Li and Jilkov, 2005). In the former, the filter is adaptive and operated on the basis of the model selected by the model decision process and consequently the hybrid estimation problem is solved by combining state estimation with explicit model decision. In this regard, timely detection of the target maneuver, namely the model adaptation of the filter, is key (Ru *et al.*, 2009). Once it fails to do so and a wrong model is used, the performance of the filter will degrade significantly.

A much simpler adaptive filter for MTT is given by handling maneuvers and random process noises jointly by a white, colored or heavy tailed noise process (Zhou and Frank, 1996; Gordon *et al.*, 2003; Ru *et al.*, 2009; Guo *et al.*, 2015). This will allow converting the MTT problem into that of state estimation in presence of non-stationary process noise with unknown statistics; further discussion on uncertainties will be presented in next section. This approach primarily applies to insignificant maneuver.

For the sake of *"not putting all the eggs in one basket"*, the MM method employs a bank of maneuver models to describe the time-varying motion and runs a bank of elemental filters based on these models, each being associated with a probability. The final estimate is given by the weighted result of these sub-filters. The most representative MM method is the interacting multi-model (IMM) algorithm and variable-structure IMM estimators (Li and Bar-Shalom, 1996; Li and Jilkov, 2005; Lan *et al.*, 2013; Granström *et al.*, 2015). The number of models in the former is fixed, whereas in the latter it can be adaptively selected from a broad set of candidate models. However, operating multiple models in parallel can be computationally very costly, but still it can be insufficient when the real model parameters vary in a continuous space (Xu *et al.*, 2016), or oppositely, too many models become as bad as too few models (Li and Bar-Shalom, 1996).

In either way, model decision/adaption delay is inevitable (Fan *et al.*, 2010), and it behaves as the delay of maneuver detection in the SM methods and as the time of probability convergence to the true model in the MM methods.

Highlight 6. *Many adaptive-model approaches proposed for MTT may show superiority when the target indeed maneuvers but perform disappointingly or even significantly worse than those without using an adaptive model, when there is actually no maneuver. We call this over-reaction due to adaptability.*

To combat the model decision delay and over-reaction, a novel solution (Li *et al.*, 2017a) is to characterize the target motion by a continuous-time trajectory function as in (30) and thereby formulate the MTT problem as an optimization problem with the goal of finding a trajectory function best fitting the sensor data, e.g., in the sense of least squares of the fitting error; see Section 7.1. The fitting approach needs neither ad-hoc maneuver detection nor multiple model design and therefore is computationally reliable and fast. It is particularly applicable to a class of smoothly maneuvering targets such as passenger aircraft, ships, trains and buses, in which no abrupt and significant change should occur for the passengers' safety and most often, the carrier moves on a predefined smooth route.

5 Intractable uncertainty

5.1 Classification of uncertainties

Besides the system functions f_t and h_t which are often considered deterministic, either known or unknown, there are three key variables whose statistics need to be specified properly for setting up a filter, including the control input $\mathbf{u}_t$ (which can be considered either deterministic or stochastic), the state process noise $\mathbf{v}_t$ and the observation noise $\mathbf{w}_t$. All of these contribute to the uncertainty of the system, the core of the stochastic process. On the one hand, if their statistics are unknown, they have to be estimated concurrently with the hidden states using available sensor observations, referred to as *simultaneous state and parameter estimation* or *adaptive filtering*. This is a challenging task since in many cases direct observation of certain parameters is very expensive or difficult if not impossible (Ghahremani and Kamwa, 2011) or the observation itself contains significant intractable uncertainties such as outlier, clutter and misdetection, to be explained below. On the other hand, they may conflict with the unconditionally Gaussian system requirement, for which proper remedies have to be taken for AGC.

In the most common case, the observation function h_t is given a priori. However, it is also often that the position of the sensor is unknown (and time varying) and needs to be estimated simultaneously with that of the target. This is often referred to as *joint sensor localization and target tracking*; see, e.g., (Guo *et al.*, 2016). Beside the maneuvering model, there are various specific problems where only a part of the parameters involved in the system function vary and need to be estimated, such as resistance in motor systems, aerodynamic parameters in UAV, etc. Unlike discrete maneuvers, these parameters may not change in a jump manner but in the continuous space. They are more generally related to system identification, out the scope of our survey.

An emerging tool for non-parametric state space modeling called *Gaussian process* (GP) regression (Rasmussen and Williams, 2005), which represents the unknown system function by a random function and infers the posterior distribution of the function from data, is very different from the PDF approximation addressed in this paper. GP gains increasing importance in machine learning (Rasmussen and Williams, 2005), robotics (Ko and Fox, 2009), signal processing (Deisenroth *et al.*, 2012; Frigola-Alcade, 2015; Särkkä *et al.*, 2016), etc. when it is difficult to find an accurate parametric form of the system function. It is interesting to recognize that the GP can be broadly classified into our AGC framework, i.e. from a GP prior to a GP posterior, to accommodate more general likelihood functions.

Overall, the major intractable uncertainties involved for an adaptive filter design based on SSM can be classified as shown in Fig. 2. To avoid distracting our attention on filters under AGC, we will leave aside the uncertainty issues caused by the (either partly or entirely) unknown system functions or abnormal observation data in this paper. What follows will focus on estimation or approximation of the statistics of inputs $\mathbf{u}_t$, state process noise $\mathbf{w}_t$ and observation noise $\mathbf{w}_t$ when they are either unknown or non-Gaussian/correlated.

5.2 Unknown input

The models and/or models' parameters may deviate from their nominal values by an unknown constant or time-varying bias, which are called unknown inputs (UIs). The corresponding filtering problem in the presence of UI is termed UI filtering (UIF). To note, the UI may appear in both state dynamics and measurement models, although we only model inputs in the state dynamic model in (1). Based on different assumptions made on UI in algorithm design, the existing UIF algorithms can be broadly categorized into the following three main classes.

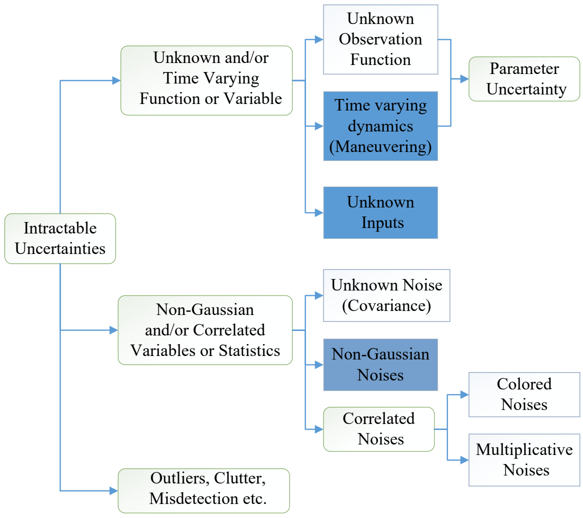


Fig. 2 Classification of major uncertainties involved in SSM. Only the highlighted three sub-topics are accommodated in this review while the others are either well addressed in existing surveys (e.g., unknown noise covariance estimation in (Duník *et al.*, 2017b)), rare cases (e.g., unknown observation function) or does not form the key theme of our review (e.g., correlated noises, outliers, clutter, misdetection, etc.)

5.2.1 Noise interpretation of UI

This approach is simply modeling the UI by a zero mean Gaussian noise with a usually large, stationary or time-varying (Liang *et al.*, 2004), covariance. An expectation-maximization (EM) based iterative optimization framework has been proposed for joint state estimation and parameter identification, which treated unknown covariances as missing data in (Bavdekar *et al.*, 2011). However, in more general cases, this assumption is often violated, which may have an adverse effect on the filtering performance or suffer from instability (divergence) (Azam *et al.*, 2015). This is because the UI is usually a non-stationary process (i.e. signal with an arbitrary type and magnitude) and cannot be well captured by a stationary and zero-mean random noise (Ghahremani and Kamwa, 2011).

5.2.2 Known UI dynamics

In this category, the UI is assumed to (approximately) follow known dynamics with an unknown initial condition. This approach can accommodate several types of UIs such as unknown constant, ramp, polynomials in time, sinusoids, or their combinations

(Su *et al.*, 2016). A common approach is to augment the UI (or the state of its dynamics) into the state variable, resulting in an augmented system for which conventional filters can be adopted. This is termed augmented KF (AKF) (Mayne, 1963) and has shown particular supremacy in fault diagnosis (Su and Chen, 2017). To reduce the computation cost of the AKF, a two-stage KF was proposed in (Friedland, 1969), which decouples the AKF into a state sub-filter and an UI sub-filter. It was further extended and optimized in (Hsieh, 2000) and generalized to the optimal multi-stage KF in (Chen and Hsieh, 2000). Inspired by (Bavdekar *et al.*, 2011), an EM optimization scheme for joint state estimation and parameter identification has been proposed in (Lan *et al.*, 2013) for stochastic systems with UIs in both the process and measurement models.

However, the augmented system is generally nonlinear even though the original one is linear. Moreover, one possible drawback of AKF is that a mean estimation error (or bias) may appear when the assumed UI dynamics is not satisfied, e.g., abrupt maneuver in target tracking (Bogler, 1987), fast time-varying disturbances in disturbance observer based control (Kim and Rew, 2013). An intuitive solution is to choose an appropriate covariance matrix for the noise term in UI dynamics, which reveals the confidence placed on the utilized UI model (Azam *et al.*, 2015), for a trade-off between estimation bias and accuracy due to stochastic errors.

5.2.3 Unknown UI dynamics

In this category, no particular dynamics is assumed on the UI. The original work of this kind (Kitanidis, 1987) solved the problem of state filtering in the presence of UI using the minimum variance unbiased estimation (i.e. minimizing the trace of state error covariance matrix under the unbiased algebra constraint). Various properties for the developed filters have been successively investigated, including the existence condition (Darouach and Zasadzinski, 1997), asymptotic stability (Fang and Callafon, 2012), and global optimality (Cheng *et al.*, 2009). Later, this approach was further extended to the case with direct feed-through of UI (Cheng *et al.*, 2009), and simultaneous input and state filtering including recursive three-step filter (RTSF) (Gillijns and Moor, 2007; Hsieh, 2009), and filtering with partial information on the input (Su *et al.*, 2015b). Recently, its relationship with the classical KF has also been rigorously established in (Li, 2013; Su *et al.*, 2015a) in terms of existence, optimality and asymptotic stability by assuming that the inputs are available at an aggregate level.

In comparison to AKF, this approach could lead to unbiased estimation while it is more sensitive to sensor noise due to the lack of prior UI dynamics information. Another point worth mentioning is the existence condition. A necessary condition of AKF is the detectability of augmented matrix pair, while strong detectability is usually required in approaches without information of UI dynamics (Yong *et al.*, 2016), which are slightly stricter.

Recent work is more focused on how to accommodate prior information on UI or unknown parameters so that both state and UI filtering performance can be improved. For example, amplitude constraint and equality constraint are considered in fault diagnosis (Simon and Simon, 2006) and in traffic management (Su *et al.*, 2015b; Li, 2013), respectively. It should be highlighted that the extra information on UI stems from the experience or knowledge of the designers. A better alternative is to learn from massive historical data. To this end, clustering and classification are exploited in (Yi *et al.*, 2016) to model vehicle acceleration for a better situation awareness performance. Another open problem comes from hybrid UIs such as an linear combination of dynamic, random, and deterministic UIs (Liang *et al.*, 2008) or more challengingly, different UI switching.

5.3 Unknown noise

There is also a large amount of literature on noise (covariance) estimation in both the state and observation equations. Interested readers can refer to a cutting-edge and comprehensive survey on this topic in (Duník *et al.*, 2017b). However, a remarkable result which appears recently (Ristic *et al.*, 2017) states that

Highlight 7. *The theoretically best achievable second order error performance, namely the CRLB, in target state estimation is independent of knowledge (or the lack of it) of the observation noise variance.*

This is in accordance with the results in (Djurić and Miguez, 2002) which demonstrates that the noise covariances are unnecessary in estimation, as they can be integrated out. More surprisingly, it was shown that the filters which do not use the true value of observation noise variance but instead estimate it online can achieve the theoretical bound, while the CKF, which is using the true value of the Gaussian observation noise variance, cannot. An explanation for this is that the filters that estimate the observation noise variance online are able to distinguish the accurate from inaccurate bearing observations and adapt their Kalman gains accordingly, resulting in overall more accurate tracking performance. This finding is interesting as it raises a puzzle: *is it a real advantage if the filter knows the true observation noise statistics?*

5.4 Non-Gaussian or non-white noise: heavy tail, correlation, and dependence

Gaussian distribution is simply incompetent to model outliers (because of clutter, impulsive noise, glint noise, or unreliable sensors etc.), skewness, heavy tails and bounded support. In addition to the aforementioned GM, a pragmatic way to approach outlier and skewed observation noise is to assume heavy-tailed noise (also called glint noise), for which elliptically contoured distributions, such as Student's t-distribution (Girón and Rojano, 1994; Tipping and Lawrence, 2005; Loxam and Drummond, 2008; Aravkin *et al.*, 2012; Piché *et al.*, 2012; Roth *et al.*, 2013; Nurminen *et al.*, 2015) and Lévy distribution (Sornette and Ide, 2001; Gordon *et al.*, 2003) turn out to be helpful.

The Student's t-distribution has been demonstrated to be less sensitive to outliers than the Gaussian distribution, thereby enjoying better robustness while retaining the minimum variance optimality of the KF. Either the process noise or the observation noise can be modeled as Student's t distribution (Aravkin *et al.*, 2012) while the latter takes a majority in literature. Based on Student's t observation noise assumption, the Bayesian filtering and smoothing recursions are developed for linear systems in (Piché *et al.*, 2012; Roth *et al.*, 2017a) based on which different parametric filters can be implemented. Student's t mixture filter has also been developed in (Loxam and Drummond, 2008).

While both Student's t distribution and the Gaussian distribution belong to the family of elliptically contoured distributions, the Gaussian approximation to the posterior PDF is more reasonable than the Student's t approximation with a fixed DOF (degree of freedom) parameter for the case of moderate contaminated process and observation noises (Huang *et al.*, 2017). In this sense, GM might be a better alternative (Bilik and Tabrikian, 2010), given proper MR-management. For an t-distributed observation noise with heavy tails, while the CRLB significantly underestimates the optimal MSE, the KF has significantly larger MSE (Piché, 2016).

There are actually at least two other intractable uncertainties leading to non-white noise, such as colored noise due to noise correlation in the time direction (Wang *et al.*, 2015) and multiplicative noises due to their dependence on the state (Spinello and Stilwell, 2010; Wang *et al.*, 2014; Agamennoni and Nebot, 2014; Liu, 2015; Huang *et al.*, 2015, 2016). Noise correlation could occur at the same time instant or one time step apart (or more complicated multiple time steps apart). Interested readers can refer to the provided references.

5.5 Robust filtering

Another filtering optimality is regarding the adaptability against a class of more significant uncertainties such as clutter, disturbances/outliers and misdetection, termed robust filtering. These uncertainties can be classified as "abnormal noise" to the system, which are unfortunately too "strong" to be effectively handled by the aforementioned maneuvering/adaptive model, noise estimation methods or heavy-tailed/correlated noise modeling approaches. Instead, robust filtering technologies such as Huber's M (maximum-likelihood-type)-estimation that can detect clutter in either state processes or observations (Koch and Yang, 1998; Yang *et al.*, 2001; Zhang *et al.*, 2016) or the H-infinity/H-∞ filter (Simon, 2006) that can handle arbitrary (unknown) noise of bounded energy, are required.

A filter is called robust if the actual error variances guarantee a minimal upper bound for all admissible uncertainties. This variant research theme was stimulated by the increased interest in robust control theory and has received a lot of attention in 1990s and early 2000s with the development of convex optimization. Some robust Gaussian filters have been reviewed in (Afshari *et al.*, 2017; Simon, 2006). Recent attention on robust filtering turns to sensor network and practical considerations such as missing data , and communication delay (Dong *et al.*, 2010), etc. It is out of the focus of this review, but we have the following observation to highlight the core difference between robust filtering and MMSE filtering.

Highlight 8. *Robust filtering is much more related to robustness with respect to statistical variations than it is to optimality with respect to a specified statistical model. Typically, the worst case estimation error rather than the MSE needs to be minimized in a robust filter. As a result, robustness is usually achieved by sacrificing the performance in terms of other criteria such as MSE and computing efficiency.*

6 Constraint

There are two basic types of constraints: physical constraints reflecting limits to physical state variables, such as positivity of mass or pressure and limitation of speed or angle; and design constraints which represent desired operating limits, such as technological limitations or geometric considerations of the system. The constraint can be either on the state or on the observation, and can be given either in equality or inequality. For example, an equality constraint between the state variables can be written as a function

$$C(\mathbf{x}_t) = \mathbf{0} \tag{26}$$

A constraint like this can be taken into account at different stages during the process of filtering, corresponding to three different classes of constrained estimation, in a bottom-up order:

- System modeling stage - modify the model,
- Filter updating stage - modify the filter, and
- Estimate output stage - modify the estimates.

6.1 Equality and inequality

6.1.1 Constrained system modeling

When the equality is defined between dimensions of the state in (26), the state can be converted to a lower-dimensional unconstrained state by representing part of the state vector as a linear function of the remaining part as governed by the equality constraint (Wen and Durrant-Whyte, 1992) . The dimension reduction can also be achieved through null space decomposition (Hewett *et al.*, 2010), in which an orthogonal factorization is used to decompose the constrained state estimation problem into stochastic and deterministic components, which are then solved separately. In contrast, the equality constraint can also be appended to the observation equation by creating an additional deterministic pseudo-observation (Tahk and Speyer, 1990; Duan and Li, 2013) from the constraint (26) as follows

$$\mathbf{y}_t = C(\mathbf{x}_t) \tag{27}$$

with the observation always treated as of mean $\mathbf{0}$ and variance $\mathbf{0}$.

The pseudo-observation model will increase the observation dimension and thereby increase the size of the matrix that needs to be inverted in the Kalman gain computation. It will also lead to a singular covariance matrix, which may cause numerical problems. More importantly, in (27), the state is not guaranteed to obey the constraint value, inappropriate for strict mathematical constraints.

6.1.2 Constrained estimation process

Instead of modifying the system models that will either increase or reduce the problem dimensions, an alternative systematic approach is to take into account the constraint during the filtering process, e.g., designing equality constrained dynamic systems based on which the filter estimate satisfies the constraints automatically (Duan and Li, 2015; Xu *et al.*, 2013), in order to provide constrained point estimates, together with constrained covariance matrices in some cases. As a representative example, the moving horizon estimation (MHE) filter minimizes the mean square error while satisfying the constraint (Ishihara and Yamakita, 2009). However, it is computationally intensive for larger horizons and nonlinearities in the observation equation or constraint.

It is important to note, under the constrained dynamics, the state process noise is state dependent in general (Duan and Li, 2015). Simply, the Gaussian distribution has an infinite tail, which does not hold in limited/constrained state spaces.

6.1.3 Constrained estimates

If neither the system models nor the filters are modified to accommodate the constraint, the last thing that can be done is to adjust the final estimate(s) produced by the unconstrained filter based on unconstrained system models. This can be done in two ways, either projecting the state space outside the constraint into the constrained area or truncating the unconstrained conditional PDF of the state so that only the part residing in the constrained area is preserved and the remainder is set to zero.

The method in (Ko and Bitmead, 2007; Julier and LaViola, 2007; Kandepu *et al.*, 2008) projects the unconstrained estimate onto the constraint subspace by a projection function $p(\mathbf{x}_t)$ satisfying (cf.(26))

$$C(p(\mathbf{x}_t)) = \mathbf{0} \tag{28}$$

for all values of $\mathbf{x}_t$.

The simplest projection approach is called *clipping*, which moves point estimates lying outside the constrained region to the boundary (Kandepu *et al.*, 2008). Furthermore, for curve road tracking, a nonlinear projection method based on second order TSE is realized in (Yang and Blasch, 2009), gaining higher accuracy than the first order linearization (Wang *et al.*, 2002). In (Ko and Bitmead, 2007), the projected KF is extended from discrete-time to continuous-time and from linear constraints to nonlinear constraints. In (Julier and LaViola, 2007), the projection method is utilized twice: one to constrain the entire distribution and the other to constrain the statistics of the distribution. Simon (2010) has analyzed three different ways that the KF solution can be projected onto the state constraint surface.

Instead of revising the point-estimate with respect to the constraint, it is more theoretically sound to modify the conditional PDF of the state estimate, typically the first two moments of the PDF. This is referred to as *the truncation approach*, in which the shape of the conditional PDF within the constrained region is preserved. This provides generally high-quality estimates with moderate computational demands (Teixeira *et al.*, 2010). In this manner, linear (Simon, 2006) and nonlinear inequality constraints (Straka *et al.*, 2012) were considered, respectively.

Nonlinear equality constraints differ from the linear case due to two sources of errors: truncation errors because of nonlinear transformation of the PDF and base point errors because the filter linearizes around the estimated value of the state rather than the true value (Julier and LaViola, 2007; Geeter *et al.*, 1997). To overcome these difficulties, the so-called *smoothly constrained KF* is proposed (Geeter *et al.*, 1997), which transforms hard constraints into soft ones and provides an exponential weighting term that progressively tightens the constraints.

Although the pseudo-observation and projection methods share the same property which allows projecting the state estimate to the constraint surface, they are qualitatively different (Julier and LaViola, 2007). The pseudo-observation method uses the KF's linear update rule and therefore is linear and its parameters are chosen to minimize the MSE estimate. The projection method can utilize any projection operator consistent with the constraint. However, if this operator takes no account of the covariance matrix, it can actually cause the covariance to increase (Geeter *et al.*, 1997). Illustrations of both approaches can be found in (Julier and LaViola, 2007).

6.2 Circular statistics

Circular estimation is involved when the state or the observation is subject to periodic quantities such as angle, orientation, or direction, which exists in an enormous number of periodic phenomena. The shifted Rayleigh filter (Clark *et al.*, 2007) is a moment matching algorithm that exploits the essential structure of the nonlinearities present in bearings-only tracking and generates the exact posterior given a Gaussian prior. Instead of suboptimal constrained filtering that treats the periodic character as a constraint, the more reliable and systematic solution shall be based on circular/directional statistics; please refer to (Kurz *et al.*, 2016) for an excellent survey on circular Bayes filtering.

A straightforward projection of the standard 1D Gaussian distribution to the circular state space is wrapping the Gaussian distribution around the unit circle and adding up all probabilities wrapped to the same point (as illustrated in Fig. 3), namely the wrapped normal (WN) distribution, of which the PDF can be immediately given as

$$p_{\mathrm{WN}}(\theta; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} e^{-\frac{(\theta-\mu+2k\pi)^2}{2\sigma^2}} \tag{29}$$

where the circular variable $\theta \in [0, 2\pi)$, $k \in \mathbb{N}$ and parameters for location $\mu \in [0, 2\pi)$ and for concentration $\sigma > 0$ which resemble the mean and standard deviation of the corresponding Gaussian distribution, respectively.

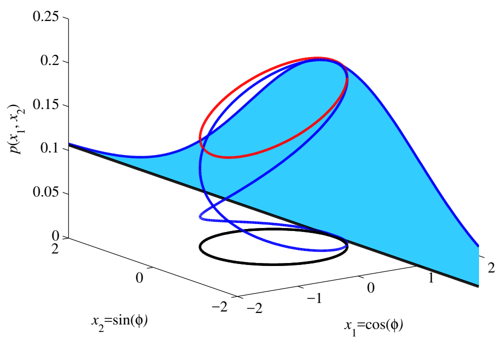


Fig. 3 A WN distribution (red) is obtained by wrapping a Gaussian distribution (blue) with $\mu = 0$ and $\sigma = 2$ around the unit circle (black) centralized around origin, where $[x_1, x_2]^{\mathrm{T}}$ gives the position on the circle.

7 New thoughts

7.1 Limitations of HMM and alternatives

Despite their popularity, HMMs are believed to be poor for modeling speech due to the restrictive conditional independence assumption that the latent state is assumed to be Markovian, i.e., the conditional density of $\mathbf{x}_t$ given the past state $\mathbf{x}_{1:t-1}$, depends only on $\mathbf{x}_{t-1}$. To overcome the limitation, there are two popular approaches. The first is to introduce additional latent variables that allow more complex inter-state dependencies to be modeled, such as factor analyzed HMM, switching linear dynamical systems (Rosti and Gales, 2003), and segmental models (Ostendorf *et al.*, 1996). The second is to relax the assumption that observations are conditionally independent given the current state by introducing explicit dependencies between observations such as buried Markov models (Bilmes, 1999), mixed memory models (Saul and Jordan, 1999), trajectory-HMM (Tokuda *et al.*, 2004), and conditional Markov chains (Bielecki *et al.*, 2017), to name a few.

Different from the stochastic modeling of the state process, Judd etc. presented a series of non-sequential/optimization based estimation and forecasting methods, particularly in the area of chaotic systems and weather forecasting applications, e.g., (Judd and Stemler, 2009; Smith *et al.*, 2010; Judd, 2015), avoiding the use of state transition noise $\mathbf{v}_t$ in (1). However, it seems that, research in this field is very little interacted with the mainstream Bayesian inference that plays a dominating role in signal processing and information fusion. In fact, similar deterministic Markov models have been applied in the noise reduction methods (Kostelich and Schreiber, 1993), MHE (Michalska and Mayne, 1995), and the GN filter (Nadjiasngar and Inggs, 2013). Interestingly, Judd's shadowing filter yields more reliable and even more accurate performance than the Bayesian filters when the nonlinearity is significant, but the noise is largely observational (Judd and Stemler, 2009), or when the objects do not display any significant random motions at the length and the time scales of interest (Judd, 2015). The GN filter that models the state transition by a deterministic differential equation is proven to be Cramér-Rao consistent (yielding minimum variance) (Morrison, 2012). These approaches emphasize the deterministic part of the system and frame the estimation problem as optimization, which has the advantage of dealing with constraints.

Highlight 9. *The standard structure of recursive filtering is based on infinite impulse response (IIR), namely all the observations prior to the present time have effect on the state estimate at present time and therefore the filter suffers from legacy errors.*

As such, once an error is made, whether due to erroneous modeling, outliers or too much approximation, it can hardly be removed. To combat this, several Kalman-like FIR (finite impulse response) estimators have been proposed, e.g., (Kwon *et al.*, 1999; Liang *et al.*, 2004; Zhao *et al.*, 2016a,b), which have been proven to be superior to the standard KF in certain cases such as the noise covariances and initial conditions are not known exactly and noise is not white. The FIR filter shares the similar idea to MHE on limiting the use of legacy information.

Moreover, particularly in the context of target tracking, positioning and localization, HMM utilizes only certain structural information in a specific and exact manner. It is unclear how to properly use some important but fuzzy/linguistic information such as a context that "the trajectory is smooth" or "the target moves close to a straight line". This type of information is analogous to the aforementioned soft constraint (Simon, 2010) but the difference is also obvious: soft constraints are usually referred to the constant that is strictly defined as in (26), (27), (28) but does not need to be strictly satisfied while the fuzzy/linguistic information addressed here can not be quantitatively defined even.

Given these considerations, Li *et al.* (2017a) proposes to use a trajectory function to replace the HMM for describing the state function, i.e.,

$$\mathbf{x}_t = f(t) \tag{30}$$

where $f(t)$ is a deterministic trajectory function of time t (FoT) defined in the state-time domain.

Considering that any trajectory can be represented by a FoT to an arbitrary accuracy, formulation (30) is quite general and versatile. Now, the state estimation problem is reformulated as an trajectory function estimation problem, which is finding a deterministic trajectory that best explains the time series observations in the underlying time-window $[k_1, k_2]$ that may move forward or extend-in-size with time, conditioned on a priori model information. Once the FoT estimate $F(t)$ is obtained, the state at any time t in the effective fitting time window (EFTW) $[K_1, K_2]$ (that does not have to be an integer) can be estimated, namely,

$$\hat{\mathbf{x}}_t = F(t), \forall t \in [K_1, K_2], \tag{31}$$

where EFTW $[K_1, K_2]$ at least covers the sampling time window $[k_1, k_2]$, namely $K_1 \leq k_1, k_2 \leq K_2$.

To incorporate any model information such as that "the target is free falling" or "the trajectory starts from a known position", the trajectory function may be more precisely specified as $F(t; C_k) \in \mathfrak{F}$ where $\mathfrak{F}$ is a finite set of specific functions, such as *a set of polynomials of no more than 3-order* and C_k is the parameter set to be estimated at discrete filtering time instant k (when new sensor data arrive), both of which shall reflect the a priori model information and fully determine the FoT at discrete time instant k. To be more precise, one may define a penalty factor $\Omega(C_k)$ on the model fitting error as a measure of the disagreement of the fitting function to the model constraint a priori, e.g.,

$$\Omega(C_k) \triangleq \| F(t_0; C_k) - \mathbf{x}_0 \| \tag{32}$$

to measure the mismatch between the fitting trajectory and known state $\mathbf{x}_0$ at time t_0 given a priori, where $\| \mathbf{a} - \mathbf{b} \|$ is a measure of the distance between $\mathbf{a} \in \mathbf{R}^{D_\mathbf{y}}$ and $\mathbf{b} \in \mathbf{R}^{D_\mathbf{y}}$ such as the square error.

Then, combining the observation function (2), a priori constraint (32) and the trajectory FoT (30) leads to an optimization problem for minimizing both the data fitting error and the model fitting error, which can be written as follows

$$\underset{F(t;C_k)}{\operatorname{argmin}} \sum_{t=k_1}^{k_2} \| \mathbf{y}_t - h_t(F(t; C_k), \bar{\mathbf{v}}_t) \| + \lambda\Omega(C_k), \tag{33}$$

where $\bar{\mathbf{v}}_t$ is an average to compensate for the observation error that can be specified as the noise mean $E(\mathbf{v}_t)$ if known or otherwise as zero to assume zero-mean noise and $\lambda > 0$ controls the trade-off between the data fitting error and the model fitting error.

As an advantage to the HMM, the FoT motion model (30) does not only ease restrictive independence assumption among time series states but also relaxes the chronological, uniform-incoming requirement posed on the observation series. As such, neither missing detection/delayed data, nor irregular sensor revisit frequency will be so challenging as in a Markov-Bayes estimator (Li *et al.*, 2017a). More importantly, the fitting framework accommodates poor prior information on the target dynamics (while it can handle smooth target maneuver) or even on the sensor observation statistics. However, how to obtain the statistical property of the estimate in these situations is still an open problem.

7.2 Filter evaluation: on computing speed

So far, we have fully omitted the computing speed of different estimators, which however is the key in many real word applications. To set up a filter, we must be clear that the affordable filtering iteration interval is determined by the duration between adjacent observations. That is, the filter updating speed must be higher or at least equal to that of the sensor revisit speed; otherwise, some sensor data will be missed/delayed.

When the filter updating speed is much faster than the sensor revisit speed, there will be some idle time at each filter iteration before the next sensor data arrives. This time can be used for additional computation such as smoothing the estimate series made so far (Li *et al.*, 2016a) by revising preceding estimates including the estimate that has just been made. Or more straightforwardly, adjust the filter a priori to properly include more computation (such as using higher order polynomial expansions or a larger number of sampling data-points or jointly exploiting multiple filters for cooperation) to reduce the idle time while obtaining better estimation.

On the opposite, when the sensor revisit rate is higher than the filtering iteration rate, or high enough to always provide newest observation, such as described by a continuous-time observation model (Ghoreyshi and Sanger, 2015), it will be another story. In such a situation, a faster updating filter has the advantage of making use of more sensor data, and suffering from smaller state transition uncertainty. For example, in real time visual tracking based on a high speed video stream, the video can be divided to a sufficient number of frames. The more frames used, the less difference between successive frames. Both more frames and less process noise can help track the content in the video. All of these will very likely lead to the conclusion that a faster filter has a better estimation performance.

Unfortunately, computing speed is often treated as a pure engineering issue and is overlooked by theoretical scientists. Instead, different filters are usually compared and evaluated based on the same simulation/ experiment setting such as using the same sensor data series and state process noises, disregarding the real filter updating rate. These pure simulations may be beyond reproach, but the indication makes sense only in very limited real world scenarios. Otherwise, whether the sensor revisit rate is high or low, it is unfair to force a computationally faster filter to wait [for a slower filter to have the same updating rate for comparison]. It should always update as fast as possible for maximally and timely utilizing more sensor data if possible, or carry out additional calculation such as smoothing to improve its estimation (before new sensor data arrive). In either way, we assert clearly that:

Highlight 10. *Computing speed matters!*

Disregarding this key issue may lead to endlessly seeking complicated modeling and/or filtering strategies for fantastically better result, which may never come true in reality. To illustrate this, we consider one case involved in sampling-based filters. In a common simulation setup as addressed above (i.e., setting all parameters disregarding the computing speed of the filter), more samples tend to yield better estimation accuracy almost for sure. This, however, cannot be guaranteed at all in reality since further increasing the number of samples will increase the computational load, slow down the filtering iteration, and therefore increase the state transition interval and the corresponding process noise. Even some sensor data may be missed when the filter updating rate turns to be slower than the sensor revisit rate. Finally, it may reduce the estimation accuracy more than it can improve. This fact will overturn the simulation indication in which more samples will almost always offer better estimation. Bearing this in mind, it is not always a good idea to develop computationally complicated filters, not only because it will definitely cost more computation resources, but also because it may lead to worse estimation accuracy.

8 Conclusion

The state-of-the-art time series parametric filters have been reviewed in four major categories, including nonlinearity (especially very informative nonlinear systems), multimode (including GM filtering and MTT), intractable uncertainties (including unknown and non-Gaussian inputs/noise) and constraints. We pointed out that a key concept behind these works is approximate Gaussian conjugacy. A few important points have been given in highlights, as well as some of our thoughts on hidden Markov modeling and practical filter evaluation. To avoid unnecessary overlap with existing review/surveys, several important topics such as noise covariance estimation approaches, correlated/independent noise-based filters, and circular statistics-based filters were not touched in this work.

Instead of addressing any applications of these filters, we put our focus on the common and general theory and algorithm design. However, we note that, efficient filter design should be based on the specific problem characteristics and requirement, e.g., estimation in robotics can be very different to that of fault diagnosis and that of target tracking.

In addition, to avoid over-wide discussion, another two major subfields regarding time series parametric filtering were not addressed either, including:

- Sensor network related distributed fusion and filtering, in the presence of imperfect sensor data such as unknown correlation, communication delay, packet loss, etc., and

- Finite set statistics (Mahler, 2014) based multi-target filtering, especially regarding multiple-sensor multiple-target scenario in the presence of misdetection and false alarm.

Both topics are closely related and have gained increasing interest. For example, finite set statistics provide a promising platform for dealing with random clutter/false alarms and misdetection as well as a random, time-varying, number of targets. Recently, an interesting connection of it to the conventional multi-target filters is presented in (Streit, 2017). Sensor network based distributed fusion is also highly tangled with imperfect sensor data such as missing data (like misdetection), false data (like false alarm) and moving sensory platforms (causing sensor uncertainty). In particular, the rapid development of sensors and their joint deployment, e.g., large scale wireless sensor networks, provide a foundation for new paradigms to address the challenges that arise in harsh environments. As a consequence, the signal processing community has showed increasing interest in novel data fusion/mining methods such as clustering, data fitting, and model learning, including the mentioned GP regression, for incorporating statistical filtering techniques to gain substantial performance enhancement. Hopefully and if in any way possible, we plan to address these issues in our second part of the series of survey.

References

Afshari, H., Gadsden, S., Habibi, S., 2017. Gaussian filters for parameter and state estimation: A general review of theory and recent trends. *Signal Processing*, **135**:218238.
http://dx.doi.org/10.1016/j.sigpro.2017.01.001

Agamennoni, G., Nebot, E.M., 2014. Robust estimation in non-linear state-space models with state-dependent noise. *IEEE Trans. Signal Process.*, **62**(8):2165-2175.
http://dx.doi.org/10.1109/TSP.2014.2305636

Ali-Loytty, S.S., 2010. Box Gaussian mixture filter . *IEEE Trans. Autom. Control*, **55**(9):2165-2169.
http://dx.doi.org/10.1109/TAC.2010.2051486

Arasaratnam, I., Haykin, S., 2008. Square-root quadrature Kalman filtering. *IEEE Trans. Signal Process.*, **56**(6):2589-2593.
http://dx.doi.org/10.1109/TSP.2007.914964

Arasaratnam, I., Haykin, S., 2009. Cubature Kalman filters. *IEEE Trans. Autom. Control*, **54**(6):1254-1269.
http://dx.doi.org/10.1109/TAC.2009.2019800

Aravkin, A., Burke, J.V., Pillonetto, G., 2012. Robust and trend-following Kalman smoothers using student's t. *IFAC Proceedings Volumes*, **45**(16):12151220 (16th IFAC Symposium on System Identification).
http://dx.doi.org/10.3182/20120711-3-BE-2027.00283

Ardeshiri, T., Granström, K., Ozkan, E., *et al.*, 2015. Greedy reduction algorithms for mixtures of exponential family. *IEEE Signal Process. Lett.*, **22**(6):676-680.
http://dx.doi.org/10.1109/LSP.2014.2367154

Arulampalam, M.S., Maskell, S., Gordon, N., *et al.*, 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, **50**(2):174-188.
http://dx.doi.org/10.1109/78.978374

Azam, S.E., Chatzi, E., Papadimitriou, C., 2015. A dual Kalman filter approach for state estimation via output-only acceleration measurements. *Mech Syst Signal Process.*, **60-61**:866-886.
http://dx.doi.org/10.1016/j.ymssp.2015.02.001

Bavdekar, V.A., Deshpande, A.P., Patwardhan, S.C., 2011. Identification of process and measurement noise covariance for state and parameter estimation using extended Kalman filter. *J. Process Control*, **21**(4):585601.
http://dx.doi.org/10.1016/j.jprocont.2011.01.001

Bell, B.M., Cathey, F.W., 1993. The iterated Kalman filter update as a Gauss-newton method. *IEEE Trans. Autom. Control*, **38**(2):294-297.
http://dx.doi.org/10.1109/9.250476

Bielecki, T.R., Jakubowski, J., Niewęgłowski, M., 2017. Conditional Markov chains: Properties, construction and structured dependence. *Stoch Process Their Appl.*, **127**(4):1125-1170.
http://dx.doi.org/10.1016/j.spa.2016.07.010

Bilik, I., Tabrikian, J., 2010. MMSE-based filtering in presence of non-Gaussian system and measurement noise. *IEEE Trans. Aerosp. Electron. Syst.*, **46**(3):1153-1170.
http://dx.doi.org/10.1109/TAES.2010.5545180

Bilmes, J.A., 1999. Buried Markov models for speech recognition. 1999 IEEE Int Conf Acoust Speech Signal Process.. Proceedings. ICASSP99 (Cat. No.99CH36258), **2**:713-716 vol.2.
http://dx.doi.org/10.1109/ICASSP.1999.759766

Bogler, P.L., 1987. Tracking a maneuvering target using input estimation. *IEEE Trans. Aerosp. Electron. Syst.*, **23**(3):298-310.
http://dx.doi.org/10.1109/TAES.1987.310826

Bordonaro, S., Willett, P., Bar-Shalom, Y., 2014. Decorrelated unbiased converted measurement Kalman filter. *IEEE Trans. Aerosp. Electron. Syst.*, **50**(2):1431-1444.
http://dx.doi.org/10.1109/TAES.2014.120563

Bugallo, M.F., Elvira, V., Martino, L., *et al.*, 2017. Adaptive importance sampling: The past, the present, and the future. *IEEE Signal Process. Mag.*, **34**(4):60-79.
http://dx.doi.org/10.1109/MSP.2017.2699226

Kalman, R., 1960. A new approach to linear filtering and prediction problems. *J. basic Engineering*, **82**(1):35-45.

Cappé, O., Godsill, S.J., Moulines, E., 2007. An overview of existing methods and recent advances in sequential Monte Carlo. *Proc. IEEE*, **95**(5):899-924.
http://dx.doi.org/10.1109/JPROC.2007.893250

Chang, L., Hu, B., Li, A., *et al.*, 2013. Transformed unscented Kalman filter. *IEEE Trans. Autom. Control*, **58**(1):252-257.
http://dx.doi.org/10.1109/TAC.2012.2204830

Chen, B., Liu, X., Zhao, H., *et al.*, 2017. Maximum correntropy Kalman filter. *Automatica*, **76**:7077.
http://dx.doi.org/10.1016/j.automatica.2016.10.004

Chen, F.C., Hsieh, C.S., 2000. Optimal multistage Kalman estimators. *IEEE Trans. Autom. Control*, **45**(11):2182-2188.
http://dx.doi.org/10.1109/9.887678

Chen, H.D., Chang, K.C., Smith, C., 2010. Constraint optimized weight adaptation for Gaussian mixture reduction. Proc. SPIE, **7697**:76970N-76970N-10. http://dx.doi.org/10.1117/12.851993

Chen, R., Liu, J.S., 2000. Mixture Kalman filters. *J R Stat Soc Series B Stat Methodol.*, **62**(3):493-508. http://dx.doi.org/10.1111/1467-9868.00246

Cheng, Y., Ye, H., Wang, Y., *et al.*, 2009. Unbiased minimum-variance state estimation for linear systems with unknown input. *Automatica*, **45**(2):485491. http://dx.doi.org/10.1016/j.automatica.2008.08.009

Clark, J.M.C., Vinter, R.B., Yaqoob, M.M., 2007. Shifted rayleigh filter: a new algorithm for bearings-only tracking. *IEEE Trans. Aerosp. Electron. Syst.*, **43**(4):1373-1384. http://dx.doi.org/10.1109/TAES.2007.4441745

Crassidis, J.L., Markley, F.L., Cheng, Y., 2007. Survey of nonlinear attitude estimation methods. *J. Guid. Control Dyn.*, **30**(1):12-28.

Crouse, D.F., Willett, P., Pattipati, K., *et al.*, 2011. A look at Gaussian mixture reduction algorithms. 14th Int. Conf. Inf. Fusion, p.1-8.

Darouach, M., Zasadzinski, M., 1997. Unbiased minimum variance estimation for systems with unknown exogenous inputs. *Automatica*, **33**(4):717719. http://dx.doi.org/10.1016/S0005-1098(96)00217-8

Daum, F., Huang, J., 2010. Generalized particle flow for nonlinear filters. Proc. SPIE, **7698**:76980I-76980I-12. http://dx.doi.org/10.1117/12.839421

Deisenroth, M.P., Turner, R.D., Huber, M.F., *et al.*, 2012. Robust filtering and smoothing with Gaussian processes. *IEEE Trans. Autom. Control*, **57**(7):1865-1871. http://dx.doi.org/10.1109/TAC.2011.2179426

DeMars, K.J., Bishop, R.H., Jah, M.K., 2013. Entropy-based approach for uncertainty propagation of nonlinear dynamical systems. *J. Guid. Control Dyn.*, **36**(4):1047-1057. http://dx.doi.org/10.2514/1.58987

Djurić, P.M., Miguez, J., 2002. Sequential particle filtering in the presence of additive Gaussian noise with unknown parameters. 2002 IEEE Int Conf Acoust Speech Signal Process., **2**:II-1621-II-1624. http://dx.doi.org/10.1109/ICASSP.2002.5744928

Dong, H., Wang, Z., Gao, H., 2010. Robust h∞ filtering for a class of nonlinear networked systems with multiple stochastic communication delays and packet dropouts. *IEEE Trans. Signal Process.*, **58**(4):1957-1966. http://dx.doi.org/10.1109/TSP.2009.2038965

Duan, Z., Li, X.R., 2013. The role of pseudo measurements in equality-constrained state estimation. *IEEE Trans. Aerosp. Electron. Syst.*, **49**(3):1654-1666. http://dx.doi.org/10.1109/TAES.2013.6558010

Duan, Z., Li, X.R., 2015. Analysis, design, and estimation of linear equality-constrained dynamic systems. *IEEE Trans. Aerosp. Electron. Syst.*, **51**(4):2732-2746. http://dx.doi.org/10.1109/TAES.2015.140441

Duník, J., Šimandl, M., Straka, O., 2010. Multiple-model filtering with multiple constraints. Proc. 2010 Amer. Control Conf., p.6858-6863. http://dx.doi.org/10.1109/ACC.2010.5531573

Duník, J., Straka, O., Šimandl, M., 2013. Stochastic integration filter. *IEEE Trans. Autom. Control*, **58**(6):1561-1566. http://dx.doi.org/10.1109/TAC.2013.2258494

Duník, J., Straka, O., Šimandl, M., *et al.*, 2015. Random-point-based filters: analysis and comparison in target tracking. *IEEE Trans. Aerosp. Electron. Syst.*, **51**(2):1403-1421. http://dx.doi.org/10.1109/TAES.2014.130136

Duník, J., Straka, O., Mallick, M., *et al.*, 2016. Survey of nonlinearity and non-Gaussianity measures for state estimation. 19th Int. Conf. Inf. Fusion, p.1845-1852.

Duník, J., Straka, O., Ajgl, J., *et al.*, 2017a. From competitive to cooperative filter design. Proc. 20th Int. Conf. Inf. Fusion, p.235-243. http://dx.doi.org/10.23919/ICIF.2017.8009652

Duník, J., Straka, O., Kost, O., *et al.*, 2017b. Noise covariance matrices in state-space models: A survey and comparison of estimation methods-part I. *Int J Adapt Control Signal Process.*, . http://dx.doi.org/10.1002/acs.2783

Eldar, Y.C., 2008. Rethinking biased estimation: Improving maximum likelihood and the Cramer-Rao bound. *Found. Trends Signal Process.*, **1**(4):305-449. http://dx.doi.org/10.1561/2000000008

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**(4):343-367. http://dx.doi.org/10.1007/s10236-003-0036-9

Fan, X., Fan, G., Havlicek, J.P., 2010. Generative model for maneuvering target tracking. *IEEE Trans. Aerosp. Electron. Syst.*, **46**(2):635-655. http://dx.doi.org/10.1109/TAES.2010.5461646

Fang, H., de Callafon, R.A., 2012. On the asymptotic stability of minimum-variance unbiased input and state estimation. *Automatica*, **48**(12):31833186. http://dx.doi.org/10.1016/j.automatica.2012.08.039

Faubel, F., McDonough, J., Klakow, D., 2009. The split and merge unscented Gaussian mixture filter. *IEEE Signal Process. Lett.*, **16**(9):786-789. http://dx.doi.org/10.1109/LSP.2009.2024859

Friedland, B., 1969. Treatment of bias in recursive filtering. *IEEE Trans. Autom. Control*, **14**(4):359-367. http://dx.doi.org/10.1109/TAC.1969.1099223

Frigola-Alcade, R., 2015. Bayesian time series learning with Gaussian processes. University of Cambridge.

Fritsche, C., Orguner, U., Gustafsson, F., 2016. On parametric lower bounds for discrete-time filtering. 2016 IEEE Int Conf Acoust Speech Signal Process., p.4338-4342. http://dx.doi.org/10.1109/ICASSP.2016.7472496

García-Fernández, A.F., Svensson, L., 2015. Gaussian map filtering using Kalman optimization. *IEEE Trans. Autom. Control*, **60**(5):1336-1349. http://dx.doi.org/10.1109/TAC.2014.2372909

García-Fernández, A.F., Morelande, M.R., Grajal, J., *et al.*, 2015a. Adaptive unscented Gaussian likelihood approximation filter. *Automatica*, **54**:166175. http://dx.doi.org/10.1016/j.automatica.2015.02.005

García-Fernández, A.F., Svensson, L., Morelande, M.R., *et al.*, 2015b. Posterior linearization filter: Principles and implementation using sigma points. *IEEE Trans. Signal Process.*, **63**(20):5561-5573. http://dx.doi.org/10.1109/TSP.2015.2454485

Geeter, J.D., Brussel, H.V., Schutter, J.D., *et al.*, 1997. A smoothly constrained Kalman filter. *IEEE Trans. Pattern Anal. Mach. Intell.*, **19**(10):1171-1177. http://dx.doi.org/10.1109/34.625129

Ghahremani, E., Kamwa, I., 2011. Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements. *IEEE Trans. Power Syst.*, **26**(4):2556-2566. http://dx.doi.org/10.1109/TPWRS.2011.2145396

Ghoreyshi, A., Sanger, T.D., 2015. A nonlinear stochastic filter for continuous-time state estimation. *IEEE Trans. Autom. Control*, **60**(8):2161-2165. http://dx.doi.org/10.1109/TAC.2015.2409910

Gillijns, S., Moor, B.D., 2007. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, **43**(5):934937. http://dx.doi.org/10.1016/j.automatica.2006.11.016

Girón, F.J., Rojano, J.C., 1994. Bayesian Kalman filtering with elliptically contoured errors. *Biometrika*, **81**(2):390-395.

Godsill, S., Clapp, T., 2001. Improvement Strategies for Monte Carlo Particle Filters. *In*: Doucet, A., de Freitas, N., Gordon, N. (Eds.), Sequential Monte Carlo Methods in Practice. Springer, New York, NY, p.139-158. Sequential Monte Carlo Methods in Practice. http://dx.doi.org/10.1007/978-1-4757-3437-9_7

Gordon, N., Percival, J., Robinson, M., 2003. The Kalman-Lévy filter and heavy-tailed models for tracking manoeuvring targets. Proc. 6th Int. Conf. Inf. Fusion, **2**:1024-1031. http://dx.doi.org/10.1109/ICIF.2003.177351

Gorman, J.D., Hero, A.O., 1990. Lower bounds for parametric estimation with constraints. *IEEE Trans. Inf. Theory*, **36**(6):1285-1301. http://dx.doi.org/10.1109/18.59929

Granström, K., Orguner, U., 2012. On the reduction of Gaussian inverse wishart mixtures. 15th Int. Conf. Inf. Fusion, p.2162-2169.

Granström, K., Willett, P., Bar-Shalom, Y., 2015. Systematic approach to imm mixing for unequal dimension states. *IEEE Trans. Aerosp. Electron. Syst.*, **51**(4):2975-2986. http://dx.doi.org/10.1109/TAES.2015.150015

Grewal, M.S., Andrews, A.P., 2014. Kalman Filtering: Theory and Practice with MATLAB. Wiley-IEEE Press.

Guo, Y., Tharmarasa, R., Rajan, S., *et al.*, 2016. Passive tracking in heavy clutter with sensor location uncertainty. *IEEE Trans. Aerosp. Electron. Syst.*, **52**(4):1536-1554. http://dx.doi.org/10.1109/TAES.2016.140820

Guo, Y., Fan, K., Peng, D., *et al.*, 2015. A modified variable rate particle filter for maneuvering target tracking. *Front Inform Technol Electron Eng*, **16**(11):985-994. http://dx.doi.org/10.1631/FITEE.1500149

Hanebeck, U.D., Briechle, K., Rauh, A., 2003. Progressive Bayes: a new framework for nonlinear state estimation. Proc. SPIE, **5099**:256-267. http://dx.doi.org/10.1117/12.487806

Hendeby, G., 2008. Performance and implementation aspects of nonlinear filtering. Institutionen för systemteknik.

Hewett, R.J., Heath, M.T., Butala, M.D., *et al.*, 2010. A robust null space method for linear equality constrained state estimation. *IEEE Trans. Signal Process.*, **58**(8):3961-3971. http://dx.doi.org/10.1109/TSP.2010.2048901

Ho, Y., Lee, R., 1964. A Bayesian approach to problems in stochastic estimation and control. *IEEE Trans. Autom. Control*, **9**(4):333-339. http://dx.doi.org/10.1109/TAC.1964.1105763

Hsieh, C.S., 2000. Robust two-stage Kalman filters for systems with unknown inputs. *IEEE Trans. Autom. Control*, **45**(12):2374-2378. http://dx.doi.org/10.1109/9.895577

Hsieh, C.S., 2009. Extension of unbiased minimum-variance input and state estimation for systems with unknown inputs. *Automatica*, **45**(9):21492153. http://dx.doi.org/10.1016/j.automatica.2009.05.004

Hu, X., Bao, M., Zhang, X.P., *et al.*, 2015. Generalized iterated Kalman filter and its performance evaluation. *IEEE Trans. Signal Process.*, **63**(12):3204-3217. http://dx.doi.org/10.1109/TSP.2015.2423266

Huang, Y., Zhang, Y., Li, N., *et al.*, 2017. A novel robust student's t-based Kalman filter. *IEEE Trans. Aerosp. Electron. Syst.*, **53**(3):1545-1554. http://dx.doi.org/10.1109/TAES.2017.2651684

Huang, Y., Zhang, Y., Wang, X., *et al.*, 2015. Gaussian filter for nonlinear systems with correlated noises at the same epoch. *Automatica*, **60**:122126. http://dx.doi.org/10.1016/j.automatica.2015.06.035

Huang, Y., Zhang, Y., Li, N., *et al.*, 2016. Design of Gaussian approximate filter and smoother for nonlinear systems with correlated noises at one epoch apart. *Circuits Syst. Signal Process.*, **35**(11):3981-4008. http://dx.doi.org/10.1007/s00034-016-0256-0

Huber, M.F., Hanebeck, U.D., 2008. Progressive Gaussian mixture reduction. 11th Int. Conf. Inf. Fusion, p.1-8.

Huber, M., 2015. Nonlinear Gaussian Filtering: Theory, Algorithms, and Applications. KIT Scientific Publishing.

Ishihara, S., Yamakita, M., 2009. Constrained state estimation for nonlinear systems with non-Gaussian noise. Proc. 48h IEEE Conf. Decision Control, p.1279-1284. http://dx.doi.org/10.1109/CDC.2009.5399627

Ito, K., Xiong, K., 2000. Gaussian filters for nonlinear filtering problems. *IEEE Trans. Autom. Control*, **45**(5):910-927. http://dx.doi.org/10.1109/9.855552

Izanloo, R., Fakoorian, S.A., Yazdi, H.S., *et al.*, 2016. Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise. 2016 CISS, p.500-505. http://dx.doi.org/10.1109/CISS.2016.7460553

Jazwinski, A.H., 1970. Mathematics in science and engineering. *Stochastic Processes and Filtering Theory*, **64**.

Jia, B., Xin, M., Cheng, Y., 2012. Sparse-grid quadrature nonlinear filtering. *Automatica*, **48**(2):327-341. http://dx.doi.org/10.1016/j.automatica.2011.08.057

Jia, B., Xin, M., Cheng, Y., 2013. High-degree cubature Kalman filter. *Automatica*, **49**(2):510-518. http://dx.doi.org/10.1016/j.automatica.2012.11.014

Judd, K., Stemler, T., 2009. Failures of sequential Bayesian filters and the successes of shadowing filters in tracking of nonlinear deterministic and stochastic systems. *Phys. Rev. E*, **79**(6).
http://dx.doi.org/10.1103/PhysRevE.79.066206

Judd, K., 2015. Tracking an object with unknown accelerations using a shadowing filter. *arXiv:1502.07743 [cs.SY]*, **abs/1502.07743**.

Julier, S.J., LaViola, J.J., 2007. On Kalman filtering with nonlinear equality constraints. *IEEE Trans. Signal Process.*, **55**(6):2774-2784.
http://dx.doi.org/10.1109/TSP.2007.893949

Julier, S.J., Uhlmann, J.K., 2004. Unscented filtering and nonlinear estimation. *Proc. IEEE*, **92**(3):401-422.
http://dx.doi.org/10.1109/JPROC.2003.823141

Kalogerias, D.S., Petropulu, A.P., 2016. Grid based nonlinear filtering revisited: Recursive estimation asymptotic optimality. *IEEE Trans. Signal Process.*, **64**(16):4244-4259.
http://dx.doi.org/10.1109/TSP.2016.2557311

Kandepu, R., Foss, B., Imsland, L., 2008. Applying the unscented Kalman filter for nonlinear state estimation. *J. Process Control*, **18**(7-8):753-768.
http://dx.doi.org/10.1016/j.jprocont.2007.11.004

Kim, K., Shevlyakov, G., 2008. Why Gaussianity? *IEEE Signal Process. Mag.*, **25**(2):102-113.
http://dx.doi.org/10.1109/MSP.2007.913700

Kim, K.S., Rew, K.H., 2013. Reduced order disturbance observer for discrete-time linear systems. *Automatica*, **49**(4):968975.
http://dx.doi.org/10.1016/j.automatica.2013.01.014

Kitanidis, P.K., 1987. Unbiased minimum-variance linear state estimation. *Automatica*, **23**(6):775778.
http://dx.doi.org/10.1016/0005-1098(87)90037-9

Ko, J., Fox, D., 2009. GP-Bayes filters: Bayesian filtering using Gaussian process prediction and observation models. *Auton Robots*, **27**(1):75-90.
http://dx.doi.org/10.1007/s10514-009-9119-x

Ko, S., Bitmead, R.R., 2007. State estimation for linear systems with state equality constraints. *Automatica*, **43**(8):13631368.
http://dx.doi.org/10.1016/j.automatica.2007.01.017

Koch, K., Yang, Y., 1998. Robust Kalman filter for rank deficient observation models. *J. Geodesy*, **72**(7-8):436-441.
http://dx.doi.org/10.1007/s001900050183

Kostelich, E., Schreiber, T., 1993. Noise-reduction in Chaotic time-series data: a survey of common methods. *Phys. Rev. E*, **48**(3):1752-1763.
http://dx.doi.org/10.1103/PhysRevE.48.1752

Kotecha, J.H., Djurić, P.M., 2003a. Gaussian particle filtering. *IEEE Trans. Signal Process.*, **51**(10):2592-2601.
http://dx.doi.org/10.1109/TSP.2003.816758

Kotecha, J.H., Djurić, P.M., 2003b. Gaussian sum particle filtering. *IEEE Trans. Signal Process.*, **51**(10):2602-2612.
http://dx.doi.org/10.1109/TSP.2003.816754

Kurz, G., Gilitschenski, I., Hanebeck, U.D., 2016. Recursive Bayesian filtering in circular state spaces. *IEEE Aerosp. Electron. Syst. Mag.*, **31**(3):70-87.
http://dx.doi.org/10.1109/MAES.2016.150083

Kwon, W.H., Kim, P.S., Park, P., 1999. A receding horizon Kalman FIR filter for discrete time-invariant systems. *IEEE Trans. Autom. Control*, **44**(9):1787-1791.
http://dx.doi.org/10.1109/9.788554

Lan, H., Liang, Y., Yang, F., *et al.*, 2013. Joint estimation and identification for stochastic systems with unknown inputs. *IET Control Theory Applications*, **7**(10):1377-1386.
http://dx.doi.org/10.1049/iet-cta.2013.0996

Lan, J., Li, X.R., 2015. Nonlinear estimation by lMMSE-based estimation with optimized uncorrelated augmentation. *IEEE Trans. Signal Process.*, **63**(16):4270-4283.
http://dx.doi.org/10.1109/TSP.2015.2437834

Lan, J., Li, X.R., 2017. Multiple conversions of measurements for nonlinear estimation. *IEEE Trans. Signal Process.*, **65**(18):4956-4970.
http://dx.doi.org/10.1109/TSP.2017.2716901

Lan, J., Li, X.R., Jilkov, V.P., *et al.*, 2013. Second-order Markov chain based multiple-model algorithm for maneuvering target tracking. *IEEE Trans. Aerosp. Electron. Syst.*, **49**(1):3-19.
http://dx.doi.org/10.1109/TAES.2013.6404088

Lerro, D., Bar-Shalom, Y., 1993. Tracking with debiased consistent converted measurements versus EKF. *IEEE Trans. Aerosp. Electron. Syst.*, **29**(3):1015-1022.
http://dx.doi.org/10.1109/7.220948

Li, B., 2013. State estimation with partially observed inputs: A unified Kalman filtering approach. *Automatica*, **49**(3):816820.
http://dx.doi.org/10.1016/j.automatica.2012.12.007

Li, T., Bolić, M., Djurić, P.M., 2015. Resampling methods for particle filtering: Classification, implementation, and strategies. *IEEE Signal Process. Mag.*, **32**(3):70-86.
http://dx.doi.org/10.1109/MSP.2014.2330626

Li, T., Prieto, J., Corchado, J.M., 2016a. Fitting for smoothing: A methodology for continuous-time target track estimation. 2016 IPIN, p.1-8.
http://dx.doi.org/10.1109/IPIN.2016.7743582

Li, T., Prieto, J., Corchado, J.M., 2016b. A short revisit of nonlinear Gaussian filters: State-of-the-art and some concerns. 2016 IEEE ICUWB, p.1-4.
http://dx.doi.org/10.1109/ICUWB.2016.7790515

Li, T., Corchado, J., Chen, H., *et al.*, 2017a. Track a smoothly maneuvering target based on trajectory estimation. Proc. 20th Int. Conf. Inf. Fusion, p.800-807.
http://dx.doi.org/10.23919/ICIF.2017.8009731

Li, T., Corchado, J., Sun, S., *et al.*, 2017b. Novel conservative fusion of Gaussian mixtures for distributed PHD fusion. *IEEE Trans. Signal Inf. Process. Net.*, (In revision).

Li, T., Corchado, J., Sun, S., 2017c. On generalized covariance intersection for distributed phd filtering and a simple but better alternative. Proc. 20th Int. Conf. Inf. Fusion, p.808-815.
http://dx.doi.org/10.23919/ICIF.2017.8009732

Li, T., Villarrubia, G., Sun, S., *et al.*, 2015. Resampling methods for particle filtering: identical distribution, a new method, and comparable study. *Front Inform Technol Electron Eng*, **16**(11):969-984.
http://dx.doi.org/10.1631/FITEE.1500199

Li, T., Corchado, J.M., Bajo, J., *et al.*, 2016. Effectiveness of Bayesian filters: An information fusion perspective. *Information Sciences*, **329**:670689.
http://dx.doi.org/10.1016/j.ins.2015.09.041

Li, T., Chen, H., Sun, S., *et al.*, 2017a. Joint Smoothing, Tracking, and Forecasting Based on Continuous-Time Target Trajectory Fitting. *ArXiv e-prints*, (arXiv:1708.02196 [stat.AP]).

Li, T., Corchado, J.M., Sun, S., *et al.*, 2017b. Clustering for filtering: Multi-object detection and estimation using multiple/massive sensors. *Information Sciences*, **388-389**:172190.
http://dx.doi.org/10.1016/j.ins.2017.01.028

Li, T., la Prieta Pintado, F.D., Corchado, J.M., *et al.*, 2017c. Multi-source homogeneous data clustering for multi-target detection from cluttered background with misdetection. *Applied Soft Computing*, (in press).
http://dx.doi.org/10.1016/j.asoc.2017.07.012

Li, X.R., Jilkov, V.P., 2002. Survey of maneuvering target tracking: decision-based methods. **4728**:511-534.
http://dx.doi.org/10.1117/12.478535

Li, X.R., Jilkov, V.P., 2005. Survey of maneuvering target tracking. part v. multiple-model methods. *IEEE Trans. Aerosp. Electron. Syst.*, **41**(4):1255-1321.
http://dx.doi.org/10.1109/TAES.2005.1561886

Li, X.R., Bar-Shalom, Y., 1996. Multiple-model estimation with variable structure. *IEEE Trans. Autom. Control*, **41**(4):478-493.
http://dx.doi.org/10.1109/9.489270

Liang, Y., Zhou, D., Zhang, L., *et al.*, 2008. Adaptive filtering for stochastic systems with generalized disturbance inputs. *IEEE Signal Process. Lett.*, **15**:645-648.
http://dx.doi.org/10.1109/LSP.2008.2002707

Liang, Y., An, D.X., Zhou, D.H., *et al.*, 2004. A finite-horizon adaptive Kalman filter for linear systems with unknown disturbances. *Signal Processing*, **84**(11):21752194.
http://dx.doi.org/10.1016/j.sigpro.2004.06.021

Lindley, D.V., Smith, A.F.M., 1972. Bayes estimates for the linear model. *J. R. Stat. Soc. Series B*, **34**(1):1-41.

Liu, W., Pokharel, P.P., Principe, J.C., 2007. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.*, **55**(11):5286-5298.
http://dx.doi.org/10.1109/TSP.2007.896065

Liu, W., 2015. Optimal estimation for discrete-time linear systems in the presence of multiplicative and time-correlated additive measurement noises. *IEEE Trans. Signal Process.*, **63**(17):4583-4593.
http://dx.doi.org/10.1109/TSP.2015.2447491

Liu, Y., Li, X.R., 2015. Measure of nonlinearity for estimation. *IEEE Trans. Signal Process.*, **63**(9):2377-2388.
http://dx.doi.org/10.1109/TSP.2015.2405495

Liu, Y., Li, X.R., Chen, H., 2013. Generalized linear minimum mean-square error estimation with application to space-object tracking. 2013 Asilomar Conference on Signals, Systems and Computers, p.2133-2137.
http://dx.doi.org/10.1109/ACSSC.2013.6810685

Loxam, J., Drummond, T., 2008. Student-t mixture filter for robust, real-time visual tracking. Proc. ECCV 2008, Marseille, France, October 12-18, 2008, Berlin, Heidelberg, p.372-385.
http://dx.doi.org/10.1007/978-3-540-88690-7_28

Ma, R., Coleman, T.P., 2011. Generalizing the posterior matching scheme to higher dimensions via optimal transportation. 2011 49th Annual Allerton Conference on Communication, Control, and Computing, p.96-102.
http://dx.doi.org/10.1109/Allerton.2011.6120155

Mahler, R., 2014. Advances in Statistical Multisource-Multitarget Information Fusion. Artech House.

Mayne, D.Q., 1963. Optimal non-stationary estimation of the parameters of a linear system with Gaussian inputs âĂă. *J. Electron. Control*, **14**(1):101-112.
http://dx.doi.org/10.1080/00207216308937480

Michalska, H., Mayne, D.Q., 1995. Moving horizon observers and observer-based control. *IEEE Trans. Autom. Control*, **40**(6):995-1006.
http://dx.doi.org/10.1109/9.388677

Mitter, S.K., Newton, N.J., 2003. A variational approach to nonlinear estimation. *SIAM J. Control Optim.*, **42**(5):1813-1833.
http://dx.doi.org/10.1137/S0363012901393894

Mohammaddadi, G., Pariz, N., Karimpour, A., 2016. Modal Kalman filter. *Asian J. Control*, .
http://dx.doi.org/10.1002/asjc.1425

Moral, Piere Del, Doucet, Arnaud, 2014. Particle methods: An introduction with applications. *ESAIM: Proc.*, **44**:1-46.
http://dx.doi.org/10.1051/proc/201444001

Morelande, M.R., García-Fernández, A.F., 2013. Analysis of Kalman filter approximations for nonlinear measurements. *IEEE Trans. Signal Process.*, **61**(22):5477-5484.
http://dx.doi.org/10.1109/TSP.2013.2279367

Morrison, N., 2012. Tracking Filter Engineering: The Gauss-Newton and Polynomial Filters. Institution of Engineering and Technology.
http://dx.doi.org/10.1049/PBRA023E

Murphy, K.P., 2007. Conjugate Bayesian analysis of the Gaussian distribution.

Nadjiasngar, R., Inggs, M., 2013. Gauss-Newton filtering incorporating Levenberg-Marquardt methods for tracking. *Digital Signal Process.*, **23**(5):1662-1667.
http://dx.doi.org/10.1016/j.dsp.2012.12.005

Nørgaard, M., Poulsen, N.K., Ravn, O., 2000. New developments in state estimation for nonlinear systems. *Automatica*, **36**(11):16271638.
http://dx.doi.org/10.1016/S0005-1098(00)00089-3

Nurminen, H., Ardeshiri, T., Piché, R., *et al.*, 2015. Robust inference for state-space models with skewed measurement noise. *IEEE Signal Process. Lett.*, **22**(11):1898-1902.
http://dx.doi.org/10.1109/LSP.2015.2437456

Nurminen, H., Piché, R., Godsill, S., 2017. Gaussian flow sigma point filter for nonlinear Gaussian state-space models. Proc. 20th Int. Conf. Inf. Fusion, p.445-452.
http://dx.doi.org/10.23919/ICIF.2017.8009682

Ostendorf, M., Digalakis, V.V., Kimball, O.A., 1996. From HMM's to segment models: a unified view of stochastic modeling for speech recognition. *IEEE Trans. Speech Audio Process.*, **4**(5):360-378.
http://dx.doi.org/10.1109/89.536930

Park, S., Serpedin, E., Qaraqe, K., 2013. Gaussian assumption: The least favorable but the most useful [lecture notes]. *IEEE Signal Process. Mag.*, **30**(3):183-186.
http://dx.doi.org/10.1109/MSP.2013.2238691

Patwardhan, S.C., Narasimhan, S., Jagadeesan, P., *et al.*, 2012. Nonlinear Bayesian state estimation: A review of recent developments. *Control Eng Pract.*, **20**(10):933953.
http://dx.doi.org/10.1016/j.conengprac.2012.04.003

Petrucci, D.J., 2005. Gaussian Mixture Reduction for Bayesian Target Tracking in Clutter. Air Force Institute of Technology.

Piché, R., Särkkä, S., Hartikainen, J., 2012. Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution. 2012 IEEE Int. Workshop on Machine Learning for Signal Processing, p.1-6.
http://dx.doi.org/10.1109/MLSP.2012.6349794

Piché, R., 2016. Cramer-Rao lower bound for linear filtering with t-distributed measurement noise. 19th Int. Conf. Inf. Fusion, p.536-540.

Pishdad, L., Labeau, F., 2015. Analytic MMSE bounds in linear dynamic systems with Gaussian mixture noise statistics. *arXiv:1506.07603*, .

Raitoharju, M., Svensson, L., García-Fernández, Á.F., *et al.*, 2017. Damped Posterior Linearization Filter. *arXiv:1704.01113 [math.OC]*, .

Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian processes for machine learning. Cambridge: MIT Press.

Reece, S., Roberts, S., 2010. Generalised covariance union: A unified approach to hypothesis merging in tracking. *IEEE Trans. Aerosp. Electron. Syst.*, **46**(1):207-221.
http://dx.doi.org/10.1109/TAES.2010.5417157

Ristic, B., Wang, X., Arulampalam, S., 2017. Target motion analysis with unknown measurement noise variance. Proc. 20th Int. Conf. Inf. Fusion, p.1663-1670.
http://dx.doi.org/10.23919/ICIF.2017.8009853

Rosti, A.V.I., Gales, M.J.F., 2003. Switching linear dynamical systems for speech recognition. University of Cambridge, Department of Engineering.

Roth, M., Özkan, E., Gustafsson, F., 2013. A student's t filter for heavy tailed process and measurement noise. 2013 IEEE Int Conf Acoust Speech Signal Process., p.5770-5774.
http://dx.doi.org/10.1109/ICASSP.2013.6638770

Roth, M., Ardeshiri, T., Özkan, E., *et al.*, 2017a. Robust Bayesian Filtering and Smoothing Using Student's t Distribution. *ArXiv:1703.02428[stat.ME]*, .

Roth, M., Hendeby, G., Fritsche, C., *et al.*, 2017b. The Ensemble Kalman Filter: A Signal Processing Perspective. *ArXiv e-prints*, .

Roth, M., Hendeby, G., Gustafsson, F., 2016. Nonlinear Kalman filters explained: A tutorial on moment computations and sigma point methods. *J. Adv. Inf. Fusion*, **11**(1):47-70.

Ru, J., Jilkov, V.P., Li, X.R., *et al.*, 2009. Detection of target maneuver onset. *IEEE Trans. Aerosp. Electron. Syst.*, **45**(2):536-554.
http://dx.doi.org/10.1109/TAES.2009.5089540

Runnalls, A.R., 2007. Kullback-leibler approach to Gaussian mixture reduction. *IEEE Trans. Aerosp. Electron. Syst.*, **43**(3):989-999.
http://dx.doi.org/10.1109/TAES.2007.4383588

Salmond, D.J., 1990. Mixture reduction algorithms for target tracking in clutter. Proc. SPIE, **1305**:434-445.
http://dx.doi.org/10.1117/12.21610

Särkkä, S., Hartikainen, J., Svensson, L., *et al.*, 2016. On the relation between Gaussian process quadratures and sigma-point methods. *J. Adv. Inf. Fusion*, **11**(1):31-46.
http://dx.doi.org/10.1016/j.automatica.2014.08.030

Sarmavuori, J., Särkkä, S., 2012. Fourier-hermite Kalman filter. *IEEE Trans. Autom. Control*, **57**(6):1511-1515.
http://dx.doi.org/10.1109/TAC.2011.2174667

Saul, L.K., Jordan, M.I., 1999. Mixed memory Markov models: Decomposing complex stochastic processes as mixtures of simpler ones. *Machine Learning*, **37**(1):75-87.
http://dx.doi.org/10.1023/A:1007649326333

Scardua, L.A., da Cruz, J.J., 2017. Complete offline tuning of the unscented Kalman filter. *Automatica*, **80**:5461.
http://dx.doi.org/10.1016/j.automatica.2017.01.008

Schieferdecker, D., Huber, M.F., 2009. Gaussian mixture reduction via clustering. 2009 12th Int. Conf. Inf. Fusion, p.1536-1543.

Simon, D., Simon, D.L., 2006. Kalman filtering with inequality constraints for turbofan engine health estimation. *IEE P. Control Theory Appl.*, **153**(3):371-378.
http://dx.doi.org/10.1049/ip-cta:20050074

Simon, D., 2006. Optimal state estimation: Kalman, H infinity, and nonlinear approaches. John Wiley & Sons.

Simon, D., 2010. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory Appl.*, **4**(8):1303-1318.

Singpurwalla, N.D., Polson, N.G., Soyer, R., 2017. From least squares to signal processing and particle filtering. *Technometrics*, .
http://dx.doi.org/10.1080/00401706.2017.1341341

Smith, L.A., Cuellar, M.C., Du, H., *et al.*, 2010. Exploiting dynamical coherence: A geometric approach to parameter estimation in nonlinear models. *Phys. Lett. A*, **374**(26):2618-2623.
http://dx.doi.org/10.1016/j.physleta.2010.04.032

Song, P., 2000. Monte Carlo Kalman filter and smoothing for multivariate discrete state space models. *Canad. J. Statist.*, **28**(3):641-652.
http://dx.doi.org/10.2307/3315971

Sorenson, H.W., 1970. Least-squares estimation: from Gauss to Kalman. *IEEE Spectrum*, **7**(7):63-68.
http://dx.doi.org/10.1109/MSPEC.1970.5213471

Sorenson, H., Alspach, D., 1971. Recursive Bayesian estimation using Gaussian sums. *Automatica*, **7**(4):465479.
http://dx.doi.org/10.1016/0005-1098(71)90097-5

Sornette, D., Ide, K., 2001. The Kalman-lévy filter. *Phys. D*, **151**(2-4):142174.
http://dx.doi.org/10.1016/S0167-2789(01)00228-7

Spinello, D., Stilwell, D.J., 2010. Nonlinear estimation with state-dependent Gaussian observation noise. *IEEE Trans. Autom. Control*, **55**(6):1358-1366.
http://dx.doi.org/10.1109/TAC.2010.2042006

Stano, P., Lendek, Z., Braaksma, J., *et al.*, 2013. Parametric Bayesian filters for nonlinear stochastic dynamical systems: A survey. *IEEE Trans. Cybern.*, **43**(6):1607-1624.
http://dx.doi.org/10.1109/TSMCC.2012.2230254

Steinbring, J., Hanebeck, U.D., 2014. Progressive Gaussian filtering using explicit likelihoods. 17th Int. Conf. Inf. Fusion, p.1-8.

Stoica, P., Babu, P., 2011. The Gaussian data assumption leads to the largest Cramer-Rao bound [lecture notes]. *IEEE Signal Process. Mag.*, **28**(3):132-133.
http://dx.doi.org/10.1109/MSP.2011.940411

Stoica, P., Moses, R.L., 1990. On biased estimators and the unbiased Cramer-Rao lower bound. *Signal Processing*, **21**(4):349350.
http://dx.doi.org/10.1016/0165-1684(90)90104-7

Straka, O., Duník, J., Šimandl, M., 2012. Truncation nonlinear filters for state estimation with nonlinear inequality constraints. *Automatica*, **48**(2):273286.
http://dx.doi.org/10.1016/j.automatica.2011.11.002

Straka, O., Duník, J., Šimandl, M., 2014. Unscented Kalman filter with advanced adaptation of scaling parameter. *Automatica*, **50**(10):26572664.
http://dx.doi.org/10.1016/j.automatica.2014.08.030

Streit, R., 2017. How I learned to stop worrying about a thousand and one filters and love analytic combinatorics. 2017 IEEE Aero. Conf., p.1-21.
http://dx.doi.org/10.1109/AERO.2017.7943644

Su, J., Chen, W.H., 2017. Model-based fault diagnosis system verification using reachability analysis. *IEEE Trans. Syst., Man, Cybern., Syst.*, **PP**(99):1-10.
http://dx.doi.org/10.1109/TSMC.2017.2710132

Su, J., Li, B., Chen, W.H., 2015a. On existence, optimality and asymptotic stability of the Kalman filter with partially observed inputs. *Automatica*, **53**:149154.
http://dx.doi.org/10.1016/j.automatica.2014.12.044

Su, J., Li, B., Chen, W.H., 2015b. Simultaneous state and input estimation with partial information on the inputs. *Syst. Sci. Control Eng.*, **3**(1):445-452.
http://dx.doi.org/10.1080/21642583.2015.1082512

Su, J., Chen, W.H., Yang, J., 2016. On Relationship Between Time-Domain and Frequency-Domain Disturbance Observers and Its Applications. *J Dyn Syst T ASME*, **138**(9).
http://dx.doi.org/10.1115/1.4033631

Tahk, M., Speyer, J.L., 1990. Target tracking problems subject to kinematic constraints. *IEEE Trans. Autom. Control*, **35**(3):324-326.
http://dx.doi.org/10.1109/9.50348

Teixeira, B.O., Tôrres, L.A., Aguirre, L.A., *et al.*, 2010. On unscented Kalman filtering with state interval constraints. *J. Process Control*, **20**(1):4557.
http://dx.doi.org/10.1016/j.jprocont.2009.10.007

Terejanu, G., Singla, P., Singh, T., *et al.*, 2011. Adaptive Gaussian sum filter for nonlinear Bayesian estimation. *IEEE Trans. Autom. Control*, **56**(9):2151-2156.
http://dx.doi.org/10.1109/TAC.2011.2141550

Tichavsky, P., Muravchik, C.H., Nehorai, A., 1998. Posterior Cramer-Rao bounds for discrete-time nonlinear filtering. *IEEE Trans. Signal Process.*, **46**(5):1386-1396.
http://dx.doi.org/10.1109/78.668800

Tipping, M.E., Lawrence, N.D., 2005. Variational inference for student-t models: Robust Bayesian interpolation and generalised component analysis. *Neurocomputing*, **69**(1-3):123141.
http://dx.doi.org/10.1016/j.neucom.2005.02.016

Tokuda, K., Zen, H., Kitamura, T., 2004. Reformulating the hmm as a trajectory model. Proc. Beyond HMM-Workshop on statistical modeling approach for speech recognition.

Van Der Merwe, R., Doucet, A., De Freitas, N., *et al.*, 2000. The unscented particle filter. Proc. 13th Int. Conf. NIPS, Cambridge, MA, USA, p.563-569.

Van Trees, H.L., Bell, K.L., 2007. Bayesian bounds for parameter estimation and nonlinear filtering/tracking. *AMC*, **10**:12.
http://dx.doi.org/10.1049/iet-rsn:20099030

Van Trees, H.L., 1968. Detection, estimation and modulation theory. *J. Wiley. New York.*, **59**:1968.

Vo, B.N., Ma, W.K., 2006. The Gaussian mixture probability hypothesis density filter. *IEEE Trans. Signal Process.*, **54**(11):4091-4104.
http://dx.doi.org/10.1109/TSP.2006.881190

Šimandl, M., Duník, J., 2009. Derivative-free estimation methods: New results and performance analysis. *Automatica*, **45**(7):17491757.
http://dx.doi.org/10.1016/j.automatica.2009.03.008

Šimandl, M., Královec, J., Söderström, T., 2006. Advanced point-mass method for nonlinear state estimation. *Automatica*, **42**(7):11331145.
http://dx.doi.org/10.1016/j.automatica.2006.03.010

Wang, L.S., Chiang, Y.T., Chang, F.R., 2002. Filtering method for nonlinear systems with constraints. *IEE P-CONTR THEOR AP*, **149**(6):525-531.
http://dx.doi.org/10.1049/ip-cta:20020799

Wang, X., Fu, M., Zhang, H., 2012. Target tracking in wireless sensor networks based on the combination of KF and MLE using distance measurements. *IEEE Trans. Mobile Comput.*, **11**(4):567-576.
http://dx.doi.org/10.1109/TMC.2011.59

Wang, X., Liang, Y., Pan, Q., *et al.*, 2015. Nonlinear Gaussian smoothers with colored measurement noise. *IEEE Trans. Autom. Control*, **60**(3):870-876.
http://dx.doi.org/10.1109/TAC.2014.2337991

Wang, X., Liang, Y., Pan, Q., *et al.*, 2014. Design and implementation of Gaussian filter for nonlinear system with randomly delayed measurements and correlated noises. *Appl. Math. Comput.*, **232**:10111024.
http://dx.doi.org/10.1016/j.amc.2013.12.168

Wen, W., Durrant-Whyte, H.F., 1992. Model-based multisensor data fusion. Proceedings 1992 IEEE ICRA, p.1720-1726 vol.2.
http://dx.doi.org/10.1109/ROBOT.1992.220130

West, M., 1993. Approximating posterior distributions by mixture. *J. R. Stat. Soc. Series B*, :409-422.

Williams, J.L., Maybeck, P.S., 2006. Cost-function-based hypothesis control techniques for multiple hypothesis tracking. *Math Comput Model.*, **43**(9-10):976989.
http://dx.doi.org/10.1016/j.mcm.2005.05.022

Wu, Y., Hu, D., Wu, M., *et al.*, 2006. A numerical-integration perspective on Gaussian filters. *IEEE Trans. Signal Process.*, **54**(8):2910-2921.
http://dx.doi.org/10.1109/TSP.2006.875389

Wu, Z., Shi, J., Zhang, X., *et al.*, 2015. Kernel recursive maximum correntropy. *Signal Processing*, **117**:1116.
http://dx.doi.org/10.1016/j.sigpro.2015.04.024

Xu, L., Li, X.R., Duan, Z., *et al.*, 2013. Modeling and state estimation for dynamic systems with linear equality constraints. *IEEE Trans. Signal Process.*, **61**(11):2927-2939.
http://dx.doi.org/10.1109/TSP.2013.2255045

Xu, L., Li, X.R., Duan, Z., 2016. Hybrid grid multiple-model estimation with application to maneuvering target tracking. *IEEE Trans. Aerosp. Electron. Syst.*, **52**(1):122-136.
http://dx.doi.org/10.1109/TAES.2015.140423

Yang, C., Blasch, E., 2009. Kalman filtering with nonlinear state constraints. *IEEE Trans. Aerosp. Electron. Syst.*, **45**(1):70-84.
http://dx.doi.org/10.1109/TAES.2009.4805264

Yang, T., Laugesen, R.S., Mehta, P.G., *et al.*, 2016. Multivariable feedback particle filter. *Automatica*, **71**:1023.
http://dx.doi.org/10.1016/j.automatica.2016.04.019

Yang, Y., He, H., Xu, G., 2001. Adaptively robust filtering for kinematic geodetic positioning. *J. Geodesy*, **75**(2-3):109-116.
http://dx.doi.org/10.1007/s001900000157

Yi, D., Su, J., Liu, C., *et al.*, 2016. Data-driven situation awareness algorithm for vehicle lane change. 2016 IEEE 19th ITSC, p.998-1003.
http://dx.doi.org/10.1109/ITSC.2016.7795677

Yong, S.Z., Zhu, M., Frazzoli, E., 2016. A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems. *Automtica*, **63**:321-329.
http://dx.doi.org/10.1016/j.automatica.2015.10.040

Zanetti, R., 2012. Recursive update filtering for nonlinear estimation. *IEEE Trans. Autom. Control*, **57**(6):1481-1490.
http://dx.doi.org/10.1109/TAC.2011.2178334

Zhan, R., Wan, J., 2007. Iterated unscented Kalman filter for passive target tracking. *IEEE Trans. Aerosp. Electron. Syst.*, **43**(3):1155-1163.
http://dx.doi.org/10.1109/TAES.2007.4383605

Zhang, C., Zhi, R., Li, T., *et al.*, 2016. Adaptive M-estimation for robust cubature Kalman filtering. 2016 Sensor Signal Processing for Defence (SSPD), p.114-118.
http://dx.doi.org/10.1109/SSPD.2016.7590586

Zhang, Y., Huang, Y., Li, N., *et al.*, 2015. Embedded cubature Kalman filter with adaptive setting of free parameter. *Signal Processing*, **114**:112-116.
http://dx.doi.org/10.1016/j.sigpro.2015.02.022

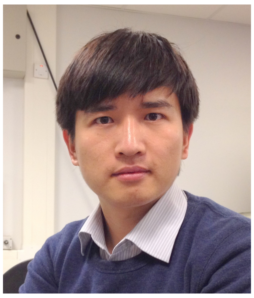
Zhao, S., Shmaliy, Y.S., Liu, F., *et al.*, 2016a. Unbiased, optimal, and in-betweens: the trade-off in discrete finite impulse response filtering. *IET Signal Process.*, **10**(4):325-334.
http://dx.doi.org/10.1049/iet-spr.2015.0360

Zhao, S., Shmaliy, Y.S., Liu, F., 2016b. Fast Kalman-like optimal unbiased FIR filtering with applications. *IEEE Trans. Signal Process.*, **64**(9):2284-2297.
http://dx.doi.org/10.1109/TSP.2016.2516960

Zheng, Y., Ozdemir, O., Niu, R., *et al.*, 2012. New conditional posterior Cramer-Rao lower bounds for nonlinear sequential Bayesian estimation. *IEEE Trans. Signal Process.*, **60**(10):5549-5556.
http://dx.doi.org/10.1109/TSP.2012.2205686

Zhou, D.H., Frank, P.M., 1996. Strong tracking filtering of nonlinear time-varying stochastic systems with coloured noise: application to parameter estimation and empirical robustness analysis. *Int. J. Control*, **65**(2):295-307.
http://dx.doi.org/10.1080/00207179608921698

Zuo, L., Niu, R., Varshney, P.K., 2011. Conditional posterior Cramer-Rao lower bounds for nonlinear sequential Bayesian estimation. *IEEE Trans. Signal Process.*, **59**(1):1-14.
http://dx.doi.org/10.1109/TSP.2010.2080268



Dr. Tian-cheng LI, first author of this invited review, received his Bachelor's degree in Mechanical and Electrical Engineering, with a minor degree in Business Administration, from Harbin Engineering University (China), in 2008, his first PhD in Electrical and Electronic Engineering from London South Bank University (United Kingdom), in 2013 and his second Doctoral degree in Mechatronics Engineering from Northwestern Polytechnical University (China), in 2015. He has been a Research Associate with the BISITE Research Group, University of Salamanca (Spain) since June 2014 and presently, he holds a Marie Skłodowska-Curie Individual Fellowship (H2020-MSCA-IF-2015) with the same Host and the Partner Vienna University of Technology (Austria).

His research interests lie in the general area of statistical signal processing and distributed information fusion, with particular interest in novel sensor data clustering, fitting and mining algorithms for advanced Multi-sensor Multi-object Detection, Tracking and Forecasting.