

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



Review:

Approximate Gaussian conjugacy: parametric recursive filtering under nonlinearity, multimodality, uncertainty, and constraint, and beyond*

Tian-cheng LI^{†1}, Jin-ya SU², Wei LIU³, Juan M. CORCHADO¹

(¹School of Sciences, University of Salamanca, Salamanca 37007, Spain)

(²Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, UK)

(³Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 4ET, UK)

E-mail: t.c.li@usal.es; J.Su2@lboro.ac.uk; w.liu@sheffield.ac.uk; corchado@usal.es

Received June 11, 2017; Revision accepted Sept. 22, 2017; Crosschecked Nov. 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 analytical estimates based on a 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 a general time series filter design. Due to challenges arising from nonlinearity, multimodality (including target maneuver), intractable uncertainties (such as unknown inputs and/or non-Gaussian noises) and constraints (including circular quantities), etc., new theories, algorithms, and technologies have been developed continuously to maintain such a conjugacy, or to approximate it as close as possible. They had contributed in large part to the prospective developments of time series parametric filters in the last six decades. In this paper, we review the state of the art in distinctive categories and highlight some insights that may otherwise be easily overlooked. In particular, specific attention is paid to nonlinear systems with an informative observation, multimodal systems including Gaussian mixture posterior and maneuvers, and intractable unknown inputs and constraints, to fill some gaps in existing reviews and surveys. In addition, we provide some new thoughts on alternatives to the first-order Markov transition model and on filter evaluation with regard to computing complexity.

Key words: Kalman filter; Gaussian filter; Time series estimation; Bayesian filtering; Nonlinear filtering; Constrained filtering; Gaussian mixture; Maneuver; Unknown inputs

<https://doi.org/10.1631/FITEE.1700379>

CLC number: TP391

1 Introduction

Dynamic state estimation, which is basically concerned with estimating a latent state that evolves over time from a sequence of observations in the

presence of noise, clutter, and disturbances, is of central interest in fields of signal/information processing and control. It has a broad range of applications related to detection, positioning, monitoring, tracking, navigation, and robotics. The rapid development of sensors and ever-increasing proliferation of smart-phones, mobile robots, and unmanned vehicles have further increased the interest in the topic.

Estimation has a long research history, although it was the Kalman filter (KF) (Kalman, 1960) that thrived the field and initiated modern estimation

[†] Corresponding author

* Project supported by the Marie Skłodowska-Curie Individual Fellowship (H2020-MSCA-IF-2015) (No. 709267) and the Open Project Program of Ministry of Education Key Laboratory of Measurement and Control of Complex Systems of Engineering, Southeast University, China (No. MCCSE2017A01)

© ORCID: Tian-cheng LI, <http://orcid.org/0000-0002-0499-5135>
 © Zhejiang University and Springer-Verlag GmbH Germany 2017

study. Historical ‘giants’ of estimation include Gauss and Legendre who independently invented the theory of least square estimation in 1795 and 1806, respectively, which anticipates most of the modern-day approaches to estimation problems, Fisher who introduced the maximum likelihood method in 1912, Kolmogorov and Wiener who established the statistical foundation for interpolation and extrapolation, filtering and prediction in 1940 and 1942, respectively, and Bode and Shannon who proposed the state-space model among many others in 1950; please refer to retrospective reviews offered by Sorenson (1970), Grewal and Andrews (2014), and Singpurwalla *et al.* (2017). It was the interpretation of 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 in a wide range of realms.

However, for a general nonlinear stochastic process with 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, such as the particle filter (PF) (Arulampalam *et al.*, 2002; Cappé *et al.*, 2007; del Moral and Arnaud, 2014; Bugallo *et al.*, 2017), and grid-based approaches (Gerstner and Griebel, 1998; Šimandl *et al.*, 2006; Kalogierias and Petropulu, 2016) based on a finite discrete state space. In the parametric case, PDF is represented by a family of functions that are fully characterized by certain parameters such as Gaussian approximation (GA) and Gaussian mixture (GM) 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 have been many excellent tutorials, surveys, and textbooks, primarily in the context of non-linearity (Nørgaard *et al.*, 2000; Wu *et al.*, 2006; Crassidis *et al.*, 2007; Hendeby, 2008; Šimandl and Duník, 2009; Li and Jilkov, 2012; Patwardhan *et al.*, 2012; Morelande and García-Fernández, 2013; Stano *et al.*, 2013; Duník *et al.*, 2015; García-Fernández

and Svensson, 2015; Huber, 2015; Roth *et al.*, 2016; Särkkä *et al.*, 2016; 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 issues have not been addressed or only addressed briefly, including: (1) a unifying framework to analyze the common essences of different filters; (2) very informative observation systems (i.e., observation noise is insignificant); (3) the classification of multimodal systems, intractable uncertainties, and constraints.

These issues will form the key part of our review, complementing the existing work. To minimize overlap with these studies, common contents will not be addressed. A comprehensive overview is still high impossible. Instead, we base our review on a transparent and concise framework termed ‘approximate Gaussian conjugacy (AGC)’. That is, all reviewed work arguably aims 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, multimodality, 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 the first-order hidden Markov model (HMM) and on filter evaluation regarding computing speed, with our new thoughts. All of these strive to give a concise albeit admittedly subjective overview of the state of the art, highlight several significant issues that can easily be ignored, and shed some light on the future research trend.

2 Basis of sequential Bayesian inference

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 on the observation series and a prior model knowledge of the system. Using the posterior distribution, one can make state inference, typically 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)).

To be more specific, the dynamic state estimation problem cast in measurable state space \mathcal{X} and observation space \mathcal{Y} can be formulated in 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 $\mathbf{x}_t, \mathbf{u}_t \in \mathcal{X}$ denote the state and the input, respectively, $\mathbf{y}_t \in \mathcal{Y}$ denotes the observation, and $\mathbf{v}_t \in \mathcal{X}, \mathbf{w}_t \in \mathcal{Y}$ denote the additive noises affecting state function f_t and observation function h_t at time instant $t \in \mathbb{N}$, respectively.

Note that state process model (1) shall be written in a differential form for the continuous time case, so does observation function (2) in a rare case (Ghoreyshi and Sanger, 2015).

We consider a process $\{(\mathbf{x}_t, \mathbf{y}_t) | t \geq 1\}$, where $\{\mathbf{x}_t | t \geq 1\}$ is a first-order HMM/Markov chain on \mathcal{X} , and each observation $\mathbf{y}_t \in \mathcal{Y}$ is conditionally independent of the rest of the process given \mathbf{x}_t . This reads (1) $p(\mathbf{x}_{0:t}) = p(\mathbf{x}_0) \prod_{k=1}^t p(\mathbf{x}_k | \mathbf{x}_{k-1})$ and (2) $p(\mathbf{y}_{0:t} | \mathbf{x}_{0:t}) = \prod_{k=0}^t p(\mathbf{y}_k | \mathbf{x}_k)$. As such, the filtering posterior is given by performing prediction and correction recursively. The prediction step combines the previous filtering distribution $p(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1})$ with state transition $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}_{0:t-1})$, as

$$p(\mathbf{x}_t | \mathbf{y}_{0:t-1}) = \int p(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1}) p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}_{0:t-1}) d\mathbf{x}_{t-1}. \quad (3)$$

This one-step forecast (a.k.a. the Chapman-Kolmogorov equation) forms the prior distribution (called ‘the prior’ hereafter). Next, given a new observation \mathbf{y}_t , the prior will be updated by the Bayes rule, resulting in the Bayes posterior distribution (called ‘the posterior’ hereafter), i.e.,

$$p(\mathbf{x}_t | \mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{0:t-1})}{\int p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{0:t-1}) d\mathbf{x}_t}, \quad (4)$$

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

Given the posterior in Eq. (4), the expected a posteriori (EAP) estimate of state \mathbf{x}_t conditioned on all observations $\mathbf{y}_{0:t}$ is given by

$$\hat{\mathbf{x}}_t^{\text{EAP}} \triangleq E[\mathbf{x}_t | \mathbf{y}_{0:t}] = \int \mathbf{x}_t p(\mathbf{x}_t | \mathbf{y}_{0:t}) d\mathbf{x}_t, \quad (5)$$

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

$$\hat{\mathbf{x}}_t^{\text{MAP}} \triangleq \arg \max_{\mathbf{x}_t} p(\mathbf{x}_t | \mathbf{y}_{0:t}). \quad (6)$$

Different from the prevalent 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 advantages in handling impulsive non-Gaussian noises, thanks to using higher-order statistics information. Correspondingly, a new class of linear KFs (Chen and Principe, 2012; Wu et al., 2015; Chen et al., 2017) have been developed. Generally, there are cases where robustness (i.e., adaptability to outliers, system errors, and disturbances) is preferable to optimality, leading to various robust filtering algorithms (see Section 5.5).

Without loss of generality, one typical iteration process of a recursive filter can be illustrated 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.

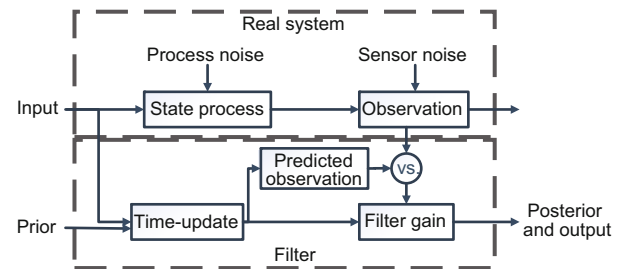


Fig. 1 Block-diagram of the evolution of a recursive filter of the prediction-updating format

2.2 Bayesian Cramér-Rao lower bound

It is theoretically pivotal to derive the performance bounds on estimation errors when estimating parameters of interest in a given model, and developing estimators to achieve these limits. When the parameters to be estimated are deterministic, a popular approach is to bound 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 (Appendix A). However, it is important to note that:

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

van Trees (1968) presented an analogous MSE bound for a random parameter, the posterior CRLB, which is also referred to as the ‘Bayesian CRLB (BCRLB)’. An elegant recursive approach was developed by 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 by only the system dynamic model, system observation model, and the prior knowledge regarding the system state at the initial time. It is thus independent of any specific realization of the system state. However, for constrained estimation problems, the corresponding constrained CRLB (Gorman and Hero, 1990) can be lower than the unconstrained version, thanks to the additional constraint information about the parameter. Some attempts have been made to include the information obtained from observations by incorporating the tracker’s information into the calculation of 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 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, such as a trajectory of an aircraft or a ground vehicle, is of interest. 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 KF is biased conditionally with a non-zero process noise realization in the (deterministic) state sequence and is not an efficient estimator in a conditional sense, even in a linear 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, by denoting θ as the parameter vector, w as the noise, and $y = x_\theta + w$ as the random observation model, we have the following property (Stoica and Babu, 2011; Park *et al.*, 2013):

Highlight 3 Among all possible distributions of observation noise w with a fixed covariance matrix, the CRLB for x attains its maximum when w is Gaussian; i.e., the Gaussian scenario is the ‘worst case’ for estimating 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. Please refer to Murphy (2007) for more conjugate priors related to Gaussian distribution. For example, the inverse Wishart distribution provides a conjugate prior for the covariance matrix of a Gaussian distribution with a known mean.

Based on conjugate prior, the Bayes prior and posterior can be computed in a closed form. More precisely, since the Gaussian PDF is determined uniquely 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 of 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 Section 7.2), and forms the essence for sequential closed-form recursion.

The strict Gaussian conjugacy, however, requires both state transition function f_t and observation function h_t be linear, and inputs u_t and noises

\mathbf{v}_t and \mathbf{w}_t be unconditionally/white Gaussian/GM (independent of the state). Then, the optimal, conjugate solution is given by KF (or a mixture of KFs in case of GM filtering), as shown in Appendix B. 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. To retain AGC, approximation has to be applied for easing the challenge from nonlinearity (regarding both functions f_t and h_t), multimodal posterior, 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 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. These approaches 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 (Mohammad-dadi *et al.*, 2017), divided difference filter (Nørgaard *et al.*, 2000; Wang *et al.*, 2017), 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, 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 by Särkkä *et al.* (2016) that many sigma-point methods can be interpreted as Gaussian quadrature based methods. They

calculated the posterior PDF in a local sense; therefore, the methods are also referred to as the local numerical approximation 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/GM PF (Kotecha and Djurić, 2003a; 2003b)), which still strives to maintain AGC. This allows asymptotically exact integral evaluation, albeit with much higher computational complexity. Like PF, these approaches are referred to as the global numerical approximation approach.

All of these GA filters have triggered tremendous further developments. For instance, UKF has perhaps gained the most approval, 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); refer to Adurthi *et al.* (2017). These problems, together with parameter setting strategies (Straka *et al.*, 2014; Zhang *et al.*, 2015; Scardua and da Cruz, 2017) and constrained filtering (see Section 5), have led to ever-increasing further developments for deterministic sampling-based filtering. Meanwhile, various measures of the degree of nonlinearity/non-Gaussianity have been developed (not limited to state estimation); see the review 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.

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 a 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), which jointly uses multiple nonlinear filters based on a weighted sum of several sub-functions of the (same) measurement (each sub-function corresponds to one filter).

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 (i.e., the noises are white and independent of the state) may not be fulfilled strictly in practice. This relaxation is particularly useful when the state model is linear and Gaussian when the measurement model is nonlinear but can be converted to a linear (namely ‘injective’) one. Such a linear-dynamic nonlinear-observation system is very common in target tracking and robot positioning realms. 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 by Lerro and Bar-Shalom (1993). Obviously, nonlinear conversion will lead to a (pseudo-measurement) noise which 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 conversion (Bordonaro *et al.*, 2014; Lan and Li, 2015), entailing correct moment matching.

A review of algebraic approaches for the Gaussian noise related debiasing was delivered by Bordonaro *et al.* (2014). To handle originally non-Gaussian noises, Monte Carlo sampling can be used for general conversion (Li *et al.*, 2016a). A recent work of Bordonaro *et al.* (2017) converted range, bearing, and range rate collaboratively to Cartesian position and velocity, permitting the use of CMF with a poor angle accuracy. When the noise is multiplicative (namely ‘dependent’) on the state, the conversion will need knowledge of the state. For example, a maximum likelihood estimator was used by Wang *et al.* (2012) to remove the distance-sensing nonlinearity in case of hybrid additive and multiplicative noises. However, we note that 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.*, 2017a).

The state of the art (Liu *et al.*, 2013; Lan and Li, 2015) has demonstrated that proper ‘uncorrelated conversion’ of the nonlinear measurement can mine more information from the measurement information

for better filtering accuracy, compared to the original measurement. This leads to an updating protocol which is based on linear combination of the original measurement and its uncorrelated conversions. However, it was further pointed out by García-Fernández *et al.* (2015a) that CMF works better, particularly for informative systems but not for non-informative systems that have a large measurement noise variance. Therefore, an interacting mechanism is advocated to switch between an unscented linear CMF and a normal unscented nonlinear filter.

3.2 Very informative observation

Dramatically fast and ever-increasing escalation has been seen on computers and sensors including radar, camera, and sonar. It is fair to say, what we have today is totally different from that when Kalman invented the KF. Either high-precision sensors or high-dimensional observations due to the joint use of multiple/massive moderate sensors are supposed to remarkably benefit our estimation by providing a very informative observation (VIO). Unfortunately, advanced KFs may not always outperform the basic KF in such cases. Instead, 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 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 among the KF variants become evident for moderately precise measurements.

Therefore, seeking increasingly accurate AGC approximations can be of limited benefit in a VIO system. Instead, an SBI filter may just lose to the observation-only (O2) inference that directly converts the observation to the state space (Li *et al.*, 2016a), equal to using a uniform/non-informative prior. It is important to note that the default formulation of most filters omits the bias propagated in the prior by taking unbiasedness as granted, which is naive at best to be true in the real world. Indeed, the bias (due to either mis-modeling or over/improper approximation) in the prior is the key leading to the defeat of a filter, especially in a VIO system.

A VIO SSM is given in Appendix C. It was first proposed by van der Merwe *et al.* (2000) and has

since been widespread for filter test. However, on it the simple O2 inference can beat EKF/UKF, unscented PF, etc., by orders of magnitude in terms of both accuracy and computational speed. In particular, prominent attention is desired as sensors are deployed with a gradually increasing quantity (with higher precision) or quality (joint use of massive sensors) (Li *et al.*, 2017a; 2018a) nowadays, popularizing VIO in reality. Therefore, we have the following note (Li *et al.*, 2016a; 2017a):

Highlight 5 While 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 worst achieve.

As a compromise, iterative algorithms may be applied to repeatedly leverage the informative observation. The first iterated EKF (IEKF) (Jazwinski, 1970) implements the first-order Taylor series expansion (TSE) of the observation function repeatedly for posterior updating to avoid filtering divergence due to the one-time first-order TSE truncation. IEKF produces a sequence of mean estimates. It was shown in 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; however, 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. With a high relevance, IPLF (García-Fernández *et al.*, 2015b; Raitoharju *et al.*, 2017) uses statistical linear regression instead of the first-order TSE for a better linearization, and iterates a posterior estimate updating.

More implementations for iterated/repeated observation (or its conversion) updating have been realized on different Gaussian filters (Zhan and Wan, 2007; Zanetti, 2012; Steinbring and Hanebeck, 2014; Huang *et al.*, 2016b). These have a close connection to the concepts of progressive correction (Oudjane and Musso, 2000) and progressive Bayes (Hanebeck *et al.*, 2003), both of which strive to apply Bayes updating in a progressive manner and 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 feedback PF (Yang *et al.*, 2016), and some (re)sampling approaches (Li *et al.*, 2015a; 2015b). There has been a burgeoning passion in applying data-driven techniques to enhance filtering in VIO systems; refer to Mitter and Newton (2003), Ma and Coleman (2011), and Nurminen *et al.* (2017) for other attempts. Parameter learning for VIO systems was also studied in Svensson *et al.* (2017).

These data-driven approaches essentially weaken the impact of the prior and converge to the O2 inference. A rigorous criterion on the optimal trade-off between the prior and the data in forming the posterior in these approaches seems still missing. In the existing work, the convergence has been identified primarily by monitoring the Kalman gain as compared with a specified ad-hoc threshold.

However, when the observation is not so informative, it turns out to be a bad idea to emphasize the observation, as quantitatively demonstrated in Li *et al.* (2016a). Therefore, particular caution should be exercised.

4 Multimodality

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 Section 4.2), 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, to name a few.

The posterior Eq. (4) in the manner of a GM of M_t components at time t can be written as

$$p(\mathbf{x}_t | \mathbf{y}_{0:t}) = \sum_{i=1}^{M_t} \omega_t^{(i)} \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_t^{(i)}, \mathbf{P}_t^{(i)}), \quad (7)$$

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

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 an 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 MMSE will converge, and relevant statistics such as mean or covariance can be derived in a closed form. In addition, taking 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 sophisticated or straightforward algorithms have been proposed for adapting/reducing the number of components in 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 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 on-line optimizations, which, however, will add to the overall computational cost. Instead, mixture reduction (MR) is more practically useful. It is typically realized in the manner of GM merging and pruning.

The first systematic GM merging scheme established by Salmond (1990) is perhaps still the most widely used protocol (Faubel *et al.*, 2009; Ali-Loytty, 2010) due to its computing simplicity and provable efficiency in practice. In fact, it is

deemed a type of ‘conservative’ fusion (Reece and Roberts, 2010) and its covariance-fusion part can be further optimized for a smaller trace, leading to the so-called ‘optimal mixture reduction (OMR)’ (Li *et al.*, 2018b). To be more specific, for a GM $\sum_{i=1}^{M_t} \omega_t^{(i)} \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_t^{(i)}, \mathbf{P}_t^{(i)})$, the OMR scheme fuses its components into a single weighted Gaussian component $\omega_{\text{OMR}} \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{\text{OMR}}, \mathbf{P}_{\text{OMR}})$ as follows:

$$\omega_{\text{OMR}} = \sum_{i=1}^{M_t} \omega_t^{(i)}, \quad (8)$$

$$\hat{\mathbf{x}}_{\text{OMR}} = \frac{\sum_{i=1}^{M_t} \omega_t^{(i)} \hat{\mathbf{x}}_t^{(i)}}{\omega_{\text{OMR}}}, \quad (9)$$

$$\mathbf{P}_{\text{OMR}} = \underset{\tilde{\mathbf{P}}_t^{(i)}}{\text{argmintr}} \left(\tilde{\mathbf{P}}_t^{(i)} \right), \quad (10)$$

where, to ensure conservativeness, the adjusted covariance matrix is given by

$$\tilde{\mathbf{P}}_t^{(i)} = \mathbf{P}_t^{(i)} + \left(\hat{\mathbf{x}}_{\text{OMR}} - \hat{\mathbf{x}}_t^{(i)} \right) \left(\hat{\mathbf{x}}_{\text{OMR}} - \hat{\mathbf{x}}_t^{(i)} \right)^{\text{T}}. \quad (11)$$

The only difference to Salmond’s approach is on the fused covariance as $\mathbf{P}_{\text{Salmond}} = \frac{\sum_{i=1}^{M_t} \omega_t^{(i)} \tilde{\mathbf{P}}_t^{(i)}}{\omega_{\text{OMR}}}$, which has a larger trace than Eq. (10). A more general principle for MR is to minimize the discrepancy between the original and the reduced mixtures (Crouse *et al.*, 2011), for which two typical metrics are the integral square error (ISE) and Kullback-Leiber divergence (KLD). The KLD of GM-PDF before MR $p(\mathbf{x})$ from that after MR $q(\mathbf{x})$, denoted by $D_{\text{KL}}(p||q)$, is an asymmetric measure of the information lost when $q(\mathbf{x})$ is used to approximate $p(\mathbf{x})$, which reads

$$\begin{aligned} D_{\text{KL}}(p||q) &\triangleq \int p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \\ &= \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} - \int p(\mathbf{x}) \ln q(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (12)$$

As the first term completely relies on PDF before MR, minimizing KLD in Eq. (12) is equivalent to maximizing the second $\int p(\mathbf{x}) \ln q(\mathbf{x}) d\mathbf{x}$. As a non-parametric distance, the ISE between $p(\mathbf{x})$ and $q(\mathbf{x})$ reads

$$D_{\text{ISE}}(p||q) \triangleq \int (p(\mathbf{x}) - q(\mathbf{x}))^2 d\mathbf{x}. \quad (13)$$

The ISE approach was first proposed for MR in the context of multiple hypothesis tracking in Williams and Maybeck (2006). It has 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, cost function (13) 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. It appears to perform better in terms of slimming GM, and has led to several further developments (Schieferdecker and Huber, 2009; Ardeshiri *et al.*, 2015; Raitoharju *et al.*, 2017).

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 ISE by simpler approaches based on clustering (Schieferdecker and Huber, 2009).

MR is actually a key part of many multi-hypothesis based approaches such as multi-hypothesis tracker (Reece and Roberts, 2010). It has also been applied to distributed information fusion for consensus (Li *et al.*, 2018b).

4.2 Maneuver

Maneuver is an important concept particularly in the context of target tracking. It 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. A straightforward solution to MTT is given by handling maneuvers and random process noises jointly by a white, colored, or heavy tailed noise process (Gordon *et al.*, 2003; Ru *et al.*, 2009; Guo *et al.*, 2015). This allows converting the MTT problem into that of state estimation in presence of non-stationary process noise with unknown statistics. This extended process noise approach primarily applies to insignificant maneuver.

The prevalent, considered-standard 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 JMS methods: the decision-based single-model (SM) method (Zhou and Frank, 1996; Li and Jilkov, 2002) and the multiple-model (MM) method (Li and Jilkov, 2005). In the former, the filter is adaptive and is operated on the basis of the model selected during the decision process, and consequently the hybrid estimation problem is solved by combining state estimation with an 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.

Considering ‘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 results of these sub-filters. The most representative MM method is the interacting MM (IMM) algorithm and variable-structure IMM estimators (Li and Bar-Shalom, 1996; Li and Jilkov, 2005; Lan J *et al.*, 2013; Granström *et al.*, 2015). An idea similar to IMM has also been developed in PFs, e.g., in Martino *et al.* (2017). The number of models in IMM is fixed, whereas in variable-structure IMM it can be selected adaptively from a broad set of candidate models. Operating multiple models in parallel can be very computationally costly; however, it can still 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.

In either way, model decision/adaption delay is inevitable (Fan *et al.*, 2011). 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 performs 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 these problems, the target motion can be described by a continuous-time trajectory function as in Eq. (18), and thereby the MTT problem can be formulated as an optimization problem to find a trajectory function best fitting the sensor data, e.g., in the sense of least squares of the fitting error (in Section 7.1). The fitting approach needs neither ad-hoc maneuver detection nor MM design and is therefore computationally reliable and fast (Li *et al.*, 2017d). It is particularly well-suited to a class of smoothly maneuvering targets such as passenger aircraft, ships, and trains, where no abrupt and significant maneuvering 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 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 control input \mathbf{u}_t (which can be considered either deterministic or stochastic), state process noise \mathbf{v}_t , and observation noise \mathbf{w}_t . All of these constitute the uncertainty of the system and the core of the stochastic process. On 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, observation function h_t is given a priori. However, it is also often that the position of the sensor is unknown (and time-varying) or there is a sensor bias, and the position of the sensor needs to be estimated simultaneously with that

of the target. This is often referred to as joint sensor localization/registration and target tracking, e.g., in Guo *et al.* (2016). Besides 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 and aerodynamic parameters in UAV. Unlike discrete maneuvers, these parameters may not change in a jump manner but in the continuous space. They are 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 (either transition function f_t or observation function h_t) by a random function (namely, GP SSM) and infers the posterior distribution of the function from data, is very different from the PDF approximation addressed in this paper. A GP is a distribution over functions. It is fully specified by a mean and a covariance function encoding basic structural assumptions of the class of functions to be modeled, e.g., smoothness and periodicity. GP gains an increasing importance in machine learning (Rasmussen and Williams, 2005), state estimation (Deisenroth *et al.*, 2012; Frigola-Alcade, 2015; Särkkä *et al.*, 2016), parameter/model learning processing (Wang *et al.*, 2008; Ko and Fox, 2009), etc., when it is difficult to find an accurate parametric form of the system function. It is interesting to recognize that 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{v}_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

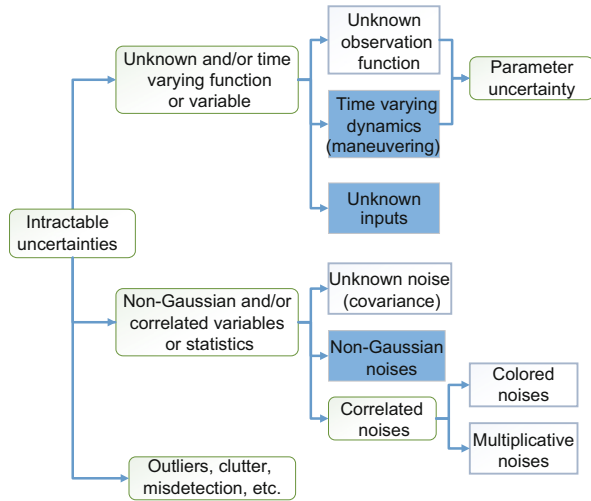


Fig. 2 Classification of major uncertainties involved in the state-space model (SSM)

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

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)’. UI may appear in both state dynamics and measurement models (including sensor bias), although we model only inputs in the state dynamic model in Eq. (1). Based on a priori assumptions made on UI, existing UIF algorithms can be broadly categorized into three main classes.

5.2.1 Noise interpretation of the unknown input

This approach is simply modeling UI by a zero-mean Gaussian noise with a usually large, stationary, or time-varying (Liang *et al.*, 2004) covariance. However, this assumption is often violated, leading to an adverse filtering performance such as instability (Azam *et al.*, 2015). This is because UI is typically 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.

5.2.2 Known unknown input dynamics

In this category, UI is approximately known dynamics and unknown initialization. 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 UI (or the state of

its dynamics) into the state variable, resulting in an augmented system for which conventional filters can be adopted, namely augmented KF (AKF) (Mayne, 1963; Su and Chen, 2017). To reduce the computation cost of AKF, Friedland (1969) proposed a two-stage KF to decouple AKF into a state sub-filter and a UI sub-filter. It was further extended and optimized by Hsieh (2000) and generalized to the optimal multi-stage KF by Chen and Hsieh (2000). An expectation-maximization (EM) based iterative optimization framework, which treats unknown covariances as missing data, was proposed by Bavdekar *et al.* (2011) for joint state estimation and parameter identification, and similarly by Lan H *et al.* (2013) for stochastic systems with UIs in both the process and measurement models.

Note that the augmented system is typically nonlinear even though the original one is linear. Also, a mean estimation error (or bias) may appear when the assumed UI dynamics is not fulfilled due to any mismatch, e.g., abrupt maneuver in target tracking (Bogler, 1987) and fast time-varying disturbances in disturbance observer based control (Kim and Rew, 2013). To combat this, an appropriate covariance matrix for the noise term in UI dynamics is the key for a trade-off between estimation bias and accuracy (Azam *et al.*, 2015).

5.2.3 Unknown input dynamics

In this category, no specific dynamics is assumed on UI. The original work of this kind (Kitanidis, 1987) is based on the MSE unbiased estimation (i.e., minimizing the trace of the state error covariance matrix under the unbiased algebra constraint). Various properties for the developed filters have been investigated successively, including the existence condition (Darouach and Zasadzinski, 1997), asymptotic stability (Fang and de Callafon, 2012), and global optimality (Cheng *et al.*, 2009). Later, this approach has been extended to the case with direct feed-through of UI (Cheng *et al.*, 2009), 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 by Li (2013) and 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 a priori 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 the state and UI filtering performance can be improved. For example, amplitude and equality constraints are considered in fault diagnosis and traffic management (Li, 2013; Su *et al.*, 2015b), 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 were exploited by Yi *et al.* (2016) to model vehicle acceleration for a better situation awareness performance. Another open problem comes from hybrid UIs, such as a linear combination of dynamic, random, and deterministic UIs (Liang *et al.*, 2008) or to be more challenging, different UI switching.

5.3 Unknown noise

There is a large body of literature on noise covariance estimation in both state and observation equations. Interested readers can refer to a cutting-edge, comprehensive survey offered by Duník *et al.* (2017b). A remarkable result which appeared recently (Ristic *et al.*, 2017) states that:

Highlight 7 The theoretically best achievable second-order error performance, namely 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 demonstrated 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 uses 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 bearing observations from inaccurate ones and adapt their Kalman gains accordingly, resulting in an overall better 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 in modeling outliers (because of clutter, impulsive noise, glint noise, unreliable sensors, etc.), skewness, heavy tails, and bounded support. In addition to the aforementioned GM, a pragmatic way to approach outliers 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 a better robustness while retaining the minimum variance optimality of 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 the literature. Based on Student’s t observation noise assumption, the Bayesian filtering and smoothing recursions were developed for linear systems in Piché *et al.* (2012) and Roth *et al.* (2017a), based on which different parametric filters can be implemented. Student’s t -mixture filter was also developed by 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 degree of freedom (DOF) 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 a proper MR-management. For a t -distributed

observation noise with heavy tails, while CRLB significantly underestimates the optimal MSE, KF has a significantly larger MSE (Piché, 2016).

There are actually at least two other intractable uncertainties leading to non-white noises, such as colored noises 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; Agamennoni and Nebot, 2014; Wang *et al.*, 2014; Huang *et al.*, 2015; Liu, 2015; Huang *et al.*, 2016a). 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, to name a few.

A filter is called robust if the actual error variances guarantee a minimum upper bound for all admissible uncertainties. This variant research theme has been 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 Simon (2006) and Afshari *et al.* (2017). Recent attention on robust filtering turns to the sensor network and practical considerations such as missing data, communication delay (Dong *et al.*, 2010), and distributed fusion (Qi *et al.*, 2014). It is out of the focus of this review; however, we have the following observation to highlight the core differences 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 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 Constraints

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 on either the state or the observation. While a few constraints are completely quantifiable in statistics, the most are typically given as contextual/‘soft’ information (Snidaro *et al.*, 2015) that has to be converted to an equality or inequality for use in the filter. For example, an equality constraint between the state variables can be written as a function:

$$C(\mathbf{x}_t) = \mathbf{0}. \quad (14)$$

The constraint can be taken into account at different inference stages, corresponding to three different types of strategies for constraining, i.e., in a bottom-up order: (1) modeling stage, (2) filtering stage, and (3) output stage.

6.1 Equality and inequality

6.1.1 Constrained system modeling

When the equality is defined between dimensions of the state in Eq. (14), 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 constraint (14) as follows:

$$\mathbf{y}_t = C(\mathbf{x}_t), \quad (15)$$

with the pseudo-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 Kalman gain computation. It will also lead to a singular covariance matrix, which may cause numerical problems. More importantly, in Eq. (15), the state is not guaranteed to obey the constraint, 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 constraints during the filtering process, e.g., designing equality constrained dynamic systems based on which the filter estimate fulfills the constraints automatically (Xu *et al.*, 2013; Duan and Li, 2015), 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 that, 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 methods proposed by Julier and LaViola (2007), Ko and Bitmead (2007), and Kandepu *et al.* (2008) project the unconstrained estimate onto the constraint subspace by a projection function $p(\mathbf{x}_t)$ satisfying (refer to Eq. (14))

$$C(p(\mathbf{x}_t)) = \mathbf{0} \quad (16)$$

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). In Ko and Bitmead (2007), the projected KF was extended from discrete time to continuous time and from linear constraints to nonlinear constraints. In Julier and LaViola (2007), the projection method was used twice: one to constrain the entire distribution and the other the statistics of the distribution. Simon (2010) analyzed three different ways by which 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 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 (Straka *et al.*, 2012) inequality constraints were considered.

Nonlinear equality constraints differ from the linear case due to two sources of errors: truncation errors because of nonlinear transformation of 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). To overcome these deficiencies, the second-order TSE was used by Yang and Blasch (2009) to gain a higher accuracy than the first-order linearization, and the so-called ‘smoothly constrained KF’ was 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. The former

uses KF's linear update rule. Therefore, it is linear and its parameters are chosen to minimize the MSE estimate. The latter can use any projection operator consistent with the constraint. 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. It 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 one-dimensional (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 (Fig. 3), namely the 'wrapped normal (WN) distribution', of which the PDF can be immediately given as

$$p_{\text{WN}}(\theta; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(\theta - \mu + 2k\pi)^2}{2\sigma^2}\right), \quad (17)$$

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 ($\sigma > 0$) resemble the mean and standard deviations of the corresponding Gaussian distribution, respectively.

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, including the Markovian state $p(\mathbf{x}_{0:t}) = p(\mathbf{x}_0) \prod_{k=1}^t p(\mathbf{x}_k | \mathbf{x}_{k-1})$ and conditional independence of observations $p(\mathbf{y}_{0:t} | \mathbf{x}_{0:t}) = \prod_{k=0}^t p(\mathbf{y}_k | \mathbf{x}_k)$.

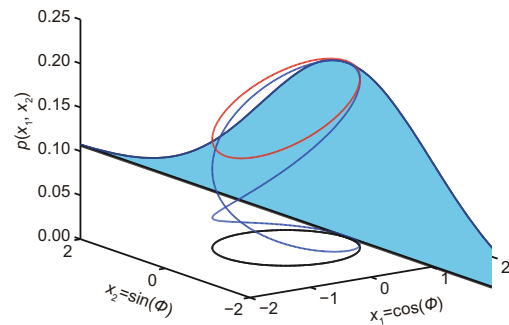


Fig. 3 A wrapped normal distribution (red) is obtained by wrapping a Gaussian distribution (blue) with $\mu = 0$ and $\sigma = 2$ around the unit circle (black) centralized around origin, where $[x_1, x_2]^T$ gives the position on the circle (References to color refer to the online version of this figure)

Extensions have been sought to break through either limitation. 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 allows explicit dependencies between observations such as buried Markov models (Bilmes, 1999), mixed memory models (Saul and Jordan, 1999), trajectory-HMM (Zen *et al.*, 2007), and conditional Markov chains (Bielecki *et al.*, 2017), to name a few.

Different from the stochastic modeling of the state process, a series of non-sequential/optimization based estimation and forecasting methods, particularly in the area of chaotic systems and weather forecasting applications, have been presented (Judd and Stemler, 2009; Smith *et al.*, 2010; Judd, 2015) to avoid the use of state transition noise \mathbf{v}_t in Eq. (1). In fact, similar deterministic Markov models have been applied in noise reduction methods (Kostelich and Schreiber, 1993), MHE (Michalska and Mayne, 1995), and the GN filter (Nadjiasngar and Inggis, 2013). Interestingly, Judd's shadowing filter yields more reliable and even more accurate results than the Bayesian filters when nonlinearity is significant while the noise is largely observational (Judd and Stemler, 2009), or when the objects do not display any significant random motion 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 advantages in 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 an effect on the state estimate at the present time. Therefore, the filter suffers from legacy errors.

As such, once a bias is made, due to whether erroneous modeling, outliers, or too much approximation, it can hardly be removed. Critically, the filter can diverge (namely deviate dramatically from the true signal) (Carrassi *et al.*, 2017) due to the accumulation of underestimated errors. To combat this, several Kalman-like finite impulse response (FIR) estimators have been proposed (Kwon *et al.*, 1999; Liang *et al.*, 2004; Zhao *et al.*, 2016a; 2016b), and proven to be superior to the standard KF in certain cases, such as when the noise covariances and initial conditions are not known exactly and the noise is not white. The FIR filter has a similar idea to MHE on limiting the use of legacy information.

Moreover, particularly in the context of target tracking, positioning, and localization, it is not so clear how to optimally use some important but fuzzy information such as a context ‘the trajectory is smooth’ or ‘the trajectory passes closely to \mathbf{x}_0 at time t_0 ’. This type of information is akin to the aforementioned soft constraint (Simon, 2010); however, the difference is obvious: soft constraints are usually referred to as a condition that is exactly defined as in Eqs. (14)–(16) but does not need to be fulfilled strictly while the fuzzy linguistic information addressed here prevents quantitative definition like so.

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

$$\mathbf{x}_t = f(t), \quad (18)$$

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 an FoT to an arbitrary accuracy, formulation (18) is quite general and versatile. Now, the state estimation problem is reformulated as a 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 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

$$\hat{\mathbf{x}}_t = F(t), \quad \forall t \in [K_1, K_2], \quad (19)$$

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

To incorporate any model information, the trajectory function may be more precisely specified as $F(t; C_k) \in \mathfrak{F}$, where \mathfrak{F} limits itself to a finite set of specific functions, such as a set of polynomials of no more than three orders, and C_k is the parameter set to be estimated at discrete time instant k (when new sensor data arrive). 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.,

$$\Omega(C_k) \triangleq \|F(t_0; C_k) - \mathbf{x}_0\|, \quad (20)$$

which is 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} and \mathbf{b} such as the square error.

Then, combining observation function (2), prior constraint (20), and trajectory FoT Eq. (18) leads to an optimization problem for jointly minimizing the data and the model fitting errors, i.e.,

$$\operatorname{argmin}_{F(t; C_k)} \sum_{t=k_1}^{k_2} \|\mathbf{y}_t - h_t(F(t; C_k), \bar{\mathbf{v}}_t)\|_{w_t} + \lambda \Omega(C_k), \quad (21)$$

where:

1. $\lambda > 0$ controls the trade-off between the data fitting error and the model fitting error.

2. w_t is the weight assigned on the data at time t to account for the time-varying uncertainty, e.g., according to the covariance of \mathbf{v}_t if known. That is, in the LS sense $\|\mathbf{e}\|_{w_t} := \mathbf{e}^T (\operatorname{Cov}[\mathbf{v}_t])^{-1} \mathbf{e}$ is a Mahalanobis distance. Alternatively, a scalar fading factor can also be considered in the weight design, such as $\|\mathbf{e}\|_{w_t} := \beta^{k-t} \|\mathbf{e}\|$ ($0 < \beta < 1$), to emphasize the newest data by assigning lower weights to history data.

3. $\bar{\mathbf{v}}_t$ is a parameter to compensate for the observation error (if anything is known) and can be

specified as the noise mean $E[\mathbf{v}_t]$ if known or otherwise as zero by assuming the sensor unbiased.

4. As default, $k_2 = k$, ensuring that the newest observation data are used while k_1 can either be fixed (i.e., the length of the time window $[k_1, k_2]$ will increase with that of k_2) or move with k_2 (namely, a sliding time window).

One may observe that the key difference between formulation (21) and the Markov–Bayes optimization is that the former defines the ‘model error’ more flexibly. As an advantage, the FoT motion model (18) not only eases the 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.*, 2017c). More importantly, the fitting framework accommodates poor prior information on the target dynamics 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 omitted the computing complexity of different estimators, which, however, is key in many real-world applications. To set up a filter, we must be clear that the affordable filter iteration interval is determined by the duration between adjacent observations. That is, the filter updating speed must be higher or at least equal to 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 smooth the estimate series made so far (Li *et al.*, 2016b) 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 a better estimation. Here, we note that employing more complicated computation or even more information does not always mean a better accuracy benefit—recall the VIO

example given in Appendix C. Interestingly, similar effect of ‘less-is-more’ appears in cognitive science (Gigerenzer and Brighton, 2009).

On the opposite, when the sensor revisit rate is higher than the filtering iteration rate, or high enough to always provide newest observation, it will be another story. In such a situation, a faster updating filter has the advantage of using 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 into a sufficient number of frames. The more frames used, the fewer differences between successive frames. Both more frames and fewer process noises 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 iteration rate, 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. In general, no matter 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 be updated as fast as possible for maximally and timely using more sensor data if possible, or additional calculation should be carried out, such as smoothing to improve its estimation (before new sensor data arrive). In either way, we assert that:

Highlight 10 (Computing speed matters) Disregarding this key issue may lead to endlessly seeking complicated modeling and/or filtering strategies for a 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 a 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, reduce the filtering iteration speed, and therefore increase the state transition interval and

the corresponding process noises. Even some sensor data may be missed when the filter updating rate turns out to be smaller than the sensor revisit rate. Finally, it may reduce the estimation accuracy more than it can improve. This fact will overturn the simulation indication. Bearing this in mind, it is not always a good idea to develop complicated filters, because it not only is computationally costly, but also may lead to no accuracy gain.

8 Conclusions and final remarks

Advances in time series parametric filters have been reviewed in four major categories, including nonlinearity (especially VIO nonlinear systems), multimodality (including GM filtering and MTT), intractable uncertainties (including unknown and non-Gaussian inputs/noise), and constraints. We pointed out that the key concept behind the work is AGC. A few important points have been given in highlights, as well as some of our thoughts on HMM and practical filter evaluation. To avoid overlap with existing review/surveys, several important topics such as noise estimation and circular statistics-based filters were not touched.

Instead of addressing some applications of these filters, we put our focus on the common and general theories and algorithm designs. However, we noted that efficient filter design should be based on the specific problem characteristics and requirements; e.g., estimation in robotics is very different to that in geosciences, and the problem of fault diagnosis is very different to that of target tracking. However, one thing is for sure: VIO plays progressively more important roles in all realms due to the revolutionary development of sensors and their massive deployment.

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

1. sensor network related distributed fusion and Bayesian filtering (Li et al., 2017b), in the presence of imperfect sensor data such as correlation and communication delay;

2. finite set statistics (Mahler, 2014) based multi-target filtering, especially regarding multi-sensor multi-target scenarios in the presence of mis-detection and false alarm.

These two topics are closely related and have gained increasing interest. 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 starts to manifest increasing eagerness in novel data fusion/mining methods such as clustering, data fitting, and model learning, including the mentioned GP regression, for incorporating advanced statistical tools and rich sensor data to gain a substantial performance enhancement.

Acknowledgements

T. Li would like to acknowledge Prof. Yu-chi (Larry) Ho with Harvard University for his high patience and generous encouragement shown in repeated discussion and comments on the topics involved in Sections 3.2 and 7.2 of this paper in the last several years since 2013.

References

- Adurthi, N., Singla, P., Singh, T., 2017. Conjugate unscented transformation: applications to estimation and control. *J. Dyn. Syst. Meas. Contr.*, **140**(3):030907. <https://doi.org/10.1115/1.4037783>
- Afshari, H., Gadsden, S., Habibi, S., 2017. Gaussian filters for parameter and state estimation: a general review of theory and recent trends. *Signal Process.*, **135**:218-238. <https://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. <https://doi.org/10.1109/TSP.2014.2305636>
- Ali-Loytty, S.S., 2010. Box Gaussian mixture filter. *IEEE Trans. Autom. Contr.*, **55**(9):2165-2169. <https://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. <https://doi.org/10.1109/TSP.2007.914964>
- Arasaratnam, I., Haykin, S., 2009. Cubature Kalman filters. *IEEE Trans. Autom. Contr.*, **54**(6):1254-1269. <https://doi.org/10.1109/TAC.2009.2019800>
- Aravkin, A., Burke, J.V., Pilonetto, G., 2012. Robust and trend-following Kalman smoothers using Student's *t*. *IFAC Proc. Vol.*, **45**(16):1215-1220. <https://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. <https://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. <https://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.
<https://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 Contr.*, **21**(4):585-601.
<https://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. Contr.*, **38**(2):294-297.
<https://doi.org/10.1109/9.250476>
- Bielecki, T.R., Jakubowski, J., Nieweglowski, M., 2017. Conditional Markov chains: properties, construction and structured dependence. *Stoch. Process. Their Appl.*, **127**(4):1125-1170.
<https://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.
<https://doi.org/10.1109/TAES.2010.5545180>
- Bilmes, J.A., 1999. Buried Markov models for speech recognition. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, p.713-716.
<https://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.
<https://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.
<https://doi.org/10.1109/TAES.2014.120563>
- Bordonaro, S., Willett, P., Bar-Shalom, Y., 2017. Consistent linear tracker with converted range, bearing and range rate measurements. *IEEE Trans. Aerosp. Electron. Syst.*, **53**(6):3135-3149.
<https://doi.org/10.1109/TAES.2017.2730980>
- 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.
<https://doi.org/10.1109/MSP.2017.2699226>
- 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.
<https://doi.org/10.1109/JPROC.2007.893250>
- Carrassi, A., Bocquet, M., Bertino, L., et al., 2017. Data assimilation in the geosciences—an overview on methods, issues and perspectives. arXiv:1709.02798.
<http://arxiv.org/abs/1709.02798>
- Chang, L., Hu, B., Li, A., et al., 2013. Transformed unscented Kalman filter. *IEEE Trans. Autom. Contr.*, **58**(1):252-257.
<https://doi.org/10.1109/TAC.2012.2204830>
- Chen, B., Principe, J.C., 2012. Maximum correntropy estimation is a smoothed MAP estimation. *IEEE Signal Process. Lett.*, **19**(8):491-494.
<https://doi.org/10.1109/LSP.2012.2204435>
- Chen, B., Liu, X., Zhao, H., et al., 2017. Maximum correntropy Kalman filter. *Automatica*, **76**:70-77.
<https://doi.org/10.1016/j.automatica.2016.10.004>
- Chen, F.C., Hsieh, C.S., 2000. Optimal multistage Kalman estimators. *IEEE Trans. Autom. Contr.*, **45**(11):2182-2188.
<https://doi.org/10.1109/9.887678>
- Chen, H.D., Chang, K.C., Smith, C., 2010. Constraint optimized weight adaptation for Gaussian mixture reduction. *SPIE*, **7697**:76970N.
<https://doi.org/10.1117/12.851993>
- Chen, R., Liu, J.S., 2000. Mixture Kalman filters. *J. R. Stat. Soc. Ser. B*, **62**(3):493-508.
<https://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):485-491.
<https://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.
<https://doi.org/10.1109/TAES.2007.4441745>
- Crassidis, J.L., Markley, F.L., Cheng, Y., 2007. Survey of nonlinear attitude estimation methods. *J. Guid. Contr. Dyn.*, **30**(1):12-28.
<https://doi.org/10.2514/1.22452>
- Crouse, D.F., Willett, P., Pattipati, K., et al., 2011. A look at Gaussian mixture reduction algorithms. 14th Int. Conf. on Information Fusion, p.1-8.
- Darouach, M., Zasadzinski, M., 1997. Unbiased minimum variance estimation for systems with unknown exogenous inputs. *Automatica*, **33**(4):717-719.
[https://doi.org/10.1016/S0005-1098\(96\)00217-8](https://doi.org/10.1016/S0005-1098(96)00217-8)
- Daum, F., Huang, J., 2010. Generalized particle flow for nonlinear filters. *SPIE*, **7698**:76980I.
<https://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. Contr.*, **57**(7):1865-1871.
<https://doi.org/10.1109/TAC.2011.2179426>
- del Moral, P., Arnaud, D., 2014. Particle methods: an introduction with applications. *Proc. ESAIM*, **44**:1-46.
<https://doi.org/10.1051/proc/201444001>
- DeMars, K.J., Bishop, R.H., Jah, M.K., 2013. Entropy-based approach for uncertainty propagation of nonlinear dynamical systems. *J. Guid. Contr. Dyn.*, **36**(4):1047-1057.
<https://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. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, p.1621-1624.
<https://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.
<https://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.
<https://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. <https://doi.org/10.1109/TAES.2015.140441>
- Duník, J., Šimandl, M., Straka, O., 2010. Multiple-model filtering with multiple constraints. Proc. American Control Conf., p.6858-6863. <https://doi.org/10.1109/ACC.2010.5531573>
- Duník, J., Straka, O., Šimandl, M., 2013. Stochastic integration filter. *IEEE Trans. Autom. Contr.*, **58**(6):1561-1566. <https://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. <https://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. on Information Fusion, p.1845-1852.
- Duník, J., Straka, O., Ajgl, J., et al., 2017a. From competitive to cooperative filter design. Proc. 20th Int. Conf. on Information Fusion, p.235-243. <https://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. Contr. Signal Process.*, **31**(11):1505-1543. <https://doi.org/10.1002/acs.2783>
- Eldar, Y.C., 2008. Rethinking biased estimation: improving maximum likelihood and the Cramér-Rao bound. *Found. Trends Signal Process.*, **1**(4):305-449. <https://doi.org/10.1561/2000000008>
- Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**(4):343-367. <https://doi.org/10.1007/s10236-003-0036-9>
- Fan, H., Zhu, Y., Fu, Q., 2011. Impact of mode decision delay on estimation error for maneuvering target interception. *IEEE Trans. Aerosp. Electron. Syst.*, **47**(1):702-711. <https://doi.org/10.1109/TAES.2011.5705700>
- Fang, H., de Callafon, R.A., 2012. On the asymptotic stability of minimum-variance unbiased input and state estimation. *Automatica*, **48**(12):3183-3186. <https://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. <https://doi.org/10.1109/LSP.2009.2024859>
- Friedland, B., 1969. Treatment of bias in recursive filtering. *IEEE Trans. Autom. Contr.*, **14**(4):359-367. <https://doi.org/10.1109/TAC.1969.1099223>
- Frigola-Alcade, R., 2015. Bayesian Time Series Learning with Gaussian Processes. PhD Thesis, University of Cambridge, Cambridge, UK.
- Fritsche, C., Orguner, U., Gustafsson, F., 2016. On parametric lower bounds for discrete-time filtering. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, p.4338-4342. <https://doi.org/10.1109/ICASSP.2016.7472496>
- García-Fernández, A.F., Svensson, L., 2015. Gaussian map filtering using Kalman optimization. *IEEE Trans. Autom. Contr.*, **60**(5):1336-1349. <https://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**:166-175. <https://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. <https://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. Patt. Anal. Mach. Intell.*, **19**(10):1171-1177. <https://doi.org/10.1109/34.625129>
- Gerstner, T., Griebel, M., 1998. Numerical integration using sparse grids. *Numer. Algor.*, **18**(3):209-232. <https://doi.org/10.1023/A:1019129717644>
- Gahremani, 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. <https://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. Contr.*, **60**(8):2161-2165. <https://doi.org/10.1109/TAC.2015.2409910>
- Gigerenzer, G., Brighton, H., 2009. Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.*, **1**(1):107-143. <https://doi.org/10.1111/j.1756-8765.2008.01006.x>
- Gillijns, S., Moor, B.D., 2007. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, **43**(5):934-937. <https://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, USA. https://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. on Information Fusion, p.1024-1031. <https://doi.org/10.1109/ICIF.2003.177351>
- Gorman, J.D., Hero, A.O., 1990. Lower bounds for parametric estimation with constraints. *IEEE Trans. Inform. Theory*, **36**(6):1285-1301. <https://doi.org/10.1109/18.59929>
- 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. <https://doi.org/10.1109/TAES.2015.150015>
- Grewal, M.S., Andrews, A.P., 2014. *Kalman Filtering: Theory and Practice with MATLAB*. Wiley-IEEE Press, New York, USA.

- 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. <https://doi.org/10.1631/FITEE.1500149>
- 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. <https://doi.org/10.1109/TAES.2016.140820>
- Hanebeck, U.D., Briehle, K., Rauh, A., 2003. Progressive Bayes: a new framework for nonlinear state estimation. *SPIE*, **5099**:256-267. <https://doi.org/10.1117/12.487806>
- Hendeby, G., 2008. Performance and Implementation Aspects of Nonlinear Filtering. PhD Thesis, Linköping University, Linköping, Sweden.
- 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. <https://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. Contr.*, **9**(4):333-339. <https://doi.org/10.1109/TAC.1964.1105763>
- Hsieh, C.S., 2009. Extension of unbiased minimum-variance input and state estimation for systems with unknown inputs. *Automatica*, **45**(9):2149-2153. <https://doi.org/10.1016/j.automat.2009.05.004>
- Hsieh, C.S., 2000. Robust two-stage Kalman filters for systems with unknown inputs. *IEEE Trans. Autom. Contr.*, **45**(12):2374-2378. <https://doi.org/10.1109/9.895577>
- 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. <https://doi.org/10.1109/TSP.2015.2423266>
- Huang, Y., Zhang, Y., Wang, X., et al., 2015. Gaussian filter for nonlinear systems with correlated noises at the same epoch. *Automatica*, **60**:122-126. <https://doi.org/10.1016/j.automat.2015.06.035>
- Huang, Y., Zhang, Y., Li, N., et al., 2016a. Design of Gaussian approximate filter and smoother for nonlinear systems with correlated noises at one epoch apart. *Circ. Syst. Signal Process.*, **35**(11):3981-4008. <https://doi.org/10.1007/S00034-016-0256-0>
- Huang, Y., Zhang, Y., Li, N., et al., 2016b. Design of Sigma-point Kalman filter with recursive updated measurement. *Circ. Syst. Signal Process.*, **35**(5):1767-1782. <https://doi.org/10.1007/s00034-015-0137-y>
- 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. <https://doi.org/10.1109/TAES.2017.2651684>
- Huber, M.F., 2015. Nonlinear Gaussian Filtering: Theory, Algorithms, and Applications. KIT Scientific Publishing, Karlsruhe, Germany.
- Huber, M.F., Hanebeck, U.D., 2008. Progressive Gaussian mixture reduction. 11th Int. Conf. on Information Fusion, p.1-8.
- Ishihara, S., Yamakita, M., 2009. Constrained state estimation for nonlinear systems with non-Gaussian noise. 48th IEEE Conf. on Decision Control, p.1279-1284. <https://doi.org/10.1109/CDC.2009.5399627>
- Ito, K., Xiong, K., 2000. Gaussian filters for nonlinear filtering problems. *IEEE Trans. Autom. Contr.*, **45**(5):910-927. <https://doi.org/10.1109/9.855552>
- Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York, USA, p.349-351.
- Jia, B., Xin, M., Cheng, Y., 2012. Sparse-grid quadrature nonlinear filtering. *Automatica*, **48**(2):327-341. <https://doi.org/10.1016/j.automat.2011.08.057>
- Jia, B., Xin, M., Cheng, Y., 2013. High-degree cubature Kalman filter. *Automatica*, **49**(2):510-518. <https://doi.org/10.1016/j.automat.2012.11.014>
- Judd, K., 2015. Tracking an object with unknown accelerations using a shadowing filter. arXiv:1502.07743. <http://arxiv.org/abs/1502.07743>
- 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):066206. <https://doi.org/10.1103/PhysRevE.79.066206>
- Julier, S.J., LaViola, J.J., 2007. On Kalman filtering with nonlinear equality constraints. *IEEE Trans. Signal Process.*, **55**(6):2774-2784. <https://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. <https://doi.org/10.1109/JPROC.2003.823141>
- Kalman, R., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**(1):35-45. <https://doi.org/10.1115/1.3662552>
- Kalogieras, D.S., Petropulu, A.P., 2016. Grid based nonlinear filtering revisited: recursive estimation asymptotic optimality. *IEEE Trans. Signal Process.*, **64**(16):4244-4259. <https://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 Contr.*, **18**(7-8):753-768. <https://doi.org/10.1016/j.jprocont.2007.11.004>
- Kim, K.S., Rew, K.H., 2013. Reduced order disturbance observer for discrete-time linear systems. *Automatica*, **49**(4):968-975. <https://doi.org/10.1016/j.automat.2013.01.014>
- Kim, K., Shevlyakov, G., 2008. Why Gaussianity? *IEEE Signal Process. Mag.*, **25**(2):102-113. <https://doi.org/10.1109/MSP.2007.913700>
- Kitanidis, P.K., 1987. Unbiased minimum-variance linear state estimation. *Automatica*, **23**(6):775-778. [https://doi.org/10.1016/0005-1098\(87\)90037-9](https://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. <https://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):1363-1368. <https://doi.org/10.1016/j.automat.2007.01.017>
- Koch, K.R., Yang, Y., 1998. Robust Kalman filter for rank deficient observation models. *J. Geod.*, **72**(7-8):436-441. <https://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. <https://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. <https://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. <https://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. <https://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. Contr.*, **44**(9):1787-1791. <https://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 Contr. Theory Appl.*, **7**(10):1377-1386. <https://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. <https://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. <https://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. <https://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. <https://doi.org/10.1109/7.220948>
- Li, B., 2013. State estimation with partially observed inputs: a unified Kalman filtering approach. *Automatica*, **49**(3):816-820. <https://doi.org/10.1016/j.automatica.2012.12.007>
- Li, T., Bolić, M., Djurić, P.M., 2015a. Resampling methods for particle filtering: classification, implementation, and strategies. *IEEE Signal Process. Mag.*, **32**(3):70-86. <https://doi.org/10.1109/MSP.2014.2330626>
- Li, T., Villarrubia, G., Sun, S., et al., 2015b. Resampling methods for particle filtering: identical distribution, a new method, and comparable study. *Front. Inform. Technol. Electron. Eng.*, **16**(11):969-984. <https://doi.org/10.1631/FITEE.1500199>
- Li, T., Corchado, J.M., Bajo, J., et al., 2016a. Effectiveness of Bayesian filters: an information fusion perspective. *Inform. Sci.*, **329**:670-689. <https://doi.org/10.1016/j.ins.2015.09.041>
- Li, T., Prieto, J., Corchado, J.M., 2016b. Fitting for smoothing: a methodology for continuous-time target track estimation. *Int. Conf. on Indoor Positioning and Indoor Navigation*, p.1-8. <https://doi.org/10.1109/IPIN.2016.7743582>
- Li, T., Corchado, J.M., Sun, S., et al., 2017a. Clustering for filtering: multi-object detection and estimation using multiple/massive sensors. *Inform. Sci.*, **388-389**:172-190. <https://doi.org/10.1016/j.ins.2017.01.028>
- Li, T., Corchado, J., Prieto, J., 2017b. Convergence of distributed flooding and its application for distributed Bayesian filtering. *IEEE Trans. Signal Inform. Process. Netw.*, **3**(3):580-591. <https://doi.org/10.1109/TSIPN.2016.2631944>
- Li, T., Chen, H., Sun, S., et al., 2017c. Joint smoothing, tracking, and forecasting based on continuous-time target trajectory fitting. *arXiv:1708.02196*. <http://arxiv.org/abs/1708.02196>
- Li, T., Corchado, J., Chen, H., et al., 2017d. Track a smoothly maneuvering target based on trajectory estimation. *Proc. 20th Int. Conf. on Information Fusion*, p.800-807. <https://doi.org/10.23919/ICIF.2017.8009731>
- Li, T., la Prieta Pintado, F.D., Corchado, J.M., et al., 2018a. Multi-source homogeneous data clustering for multi-target detection from cluttered background with mis-detection. *Appl. Soft Comput.*, **60**:436-446. <https://doi.org/10.1016/j.asoc.2017.07.012>
- Li, T., Corchado, J., Sun, S., et al., 2018b. Partial consensus and conservative fusion of Gaussian mixtures for distributed PHD fusion. *arXiv:1711.10783*. <http://arxiv.org/abs/1711.10783>
- Li, X.R., Bar-Shalom, Y., 1996. Multiple-model estimation with variable structure. *IEEE Trans. Autom. Contr.*, **41**(4):478-493. <https://doi.org/10.1109/9.489270>
- Li, X.R., Jilkov, V.P., 2002. Survey of maneuvering target tracking: decision-based methods. *SPIE*, **4728**:511-534. <https://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. <https://doi.org/10.1109/TAES.2005.1561886>
- Li, X.R., Jilkov, V.P., 2012. A survey of maneuvering target tracking, Part VIc: approximate nonlinear density filtering in discrete time. *SPIE*, **8393**:83930V. <https://doi.org/10.1117/12.921508>
- Liang, Y., An, D.X., Zhou, D.H., et al., 2004. A finite-horizon adaptive Kalman filter for linear systems with unknown disturbances. *Signal Process.*, **84**(11):2175-2194. <https://doi.org/10.1016/j.sigpro.2004.06.021>
- Liang, Y., Zhou, D.H., Zhang, L., et al., 2008. Adaptive filtering for stochastic systems with generalized disturbance inputs. *IEEE Signal Process. Lett.*, **15**:645-648. <https://doi.org/10.1109/LSP.2008.2002707>
- Lindley, D.V., Smith, A.F.M., 1972. Bayes estimates for the linear model. *J. R. Stat. Soc. Ser. B*, **34**(1):1-41.
- 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. <https://doi.org/10.1109/TSP.2015.2447491>
- 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. <https://doi.org/10.1109/TSP.2007.896065>
- Liu, Y., Li, X.R., 2015. Measure of nonlinearity for estimation. *IEEE Trans. Signal Process.*, **63**(9):2377-2388. <https://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. *Asilomar Conf. on Signals, Systems, and Computers*, p.2133-2137.
<https://doi.org/10.1109/ACSSC.2013.6810685>
- Loxam, J., Drummond, T., 2008. Student-*t* mixture filter for robust, real-time visual tracking. *European Conf. on Computer Vision*, p.372-385.
https://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. *49th Annual Allerton Conf. on Communication, Control, and Computing*, p.96-102.
<https://doi.org/10.1109/Allerton.2011.6120155>
- Mahler, R., 2014. *Advances in Statistical Multisource-Multitarget Information Fusion*. Artech House, Norwood, USA.
- Martino, L., Read, J., Elvira, V., et al., 2017. Cooperative parallel particle filters for online model selection and applications to urban mobility. *Dig. Signal Process.*, **60**:172-185.
<https://doi.org/10.1016/j.dsp.2016.09.011>
- Mayne, D.Q., 1963. Optimal non-stationary estimation of the parameters of a linear system with Gaussian inputs. *J. Electron. Contr.*, **14**(1):101-112.
<https://doi.org/10.1080/00207216308937480>
- Michalska, H., Mayne, D.Q., 1995. Moving horizon observers and observer-based control. *IEEE Trans. Autom. Contr.*, **40**(6):995-1006.
<https://doi.org/10.1109/9.388677>
- Mitter, S.K., Newton, N.J., 2003. A variational approach to nonlinear estimation. *SIAM J. Contr. Optim.*, **42**(5):1813-1833.
<https://doi.org/10.1137/S0363012901393894>
- Mohammaddadi, G., Pariz, N., Karimpour, A., 2017. Modal Kalman filter. *Asian J. Contr.*, **19**(2):728-738.
<https://doi.org/10.1002/asjc.1425>
- Morelande, M.R., García-Fernández, A.F., 2013. Analysis of Kalman filter approximations for nonlinear measurements. *IEEE Trans. Signal Process.*, **61**(22):5477-5484.
<https://doi.org/10.1109/TSP.2013.2279367>
- Morrison, N., 2012. *Tracking Filter Engineering: the Gauss-Newton and Polynomial Filters*. IET, London, UK.
<https://doi.org/10.1049/PBRA023E>
- Murphy, K.P., 2007. *Conjugate Bayesian Analysis of the Gaussian Distribution*. Technical Report, University of British Columbia, Vancouver, Canada.
- Nadjiasngar, R., Inggs, M., 2013. Gauss-Newton filtering incorporating Levenberg-Marquardt methods for tracking. *Dig. Signal Process.*, **23**(5):1662-1667.
<https://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):1627-1638.
[https://doi.org/10.1016/S0005-1098\(00\)00089-3](https://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.
<https://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. on Information Fusion*, p.1-8.
<https://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.
<https://doi.org/10.1109/89.536930>
- Oudjane, N., Musso, C., 2000. Progressive correction for regularized particle filters. *3rd Int. Conf. on Information Fusion: THB2/10*.
<https://doi.org/10.1109/IFIC.2000.859873>
- 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.
<https://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. *Contr. Eng. Pract.*, **20**(10):933-953.
<https://doi.org/10.1016/j.conengprac.2012.04.003>
- Petrucci, D.J., 2005. *Gaussian Mixture Reduction for Bayesian Target Tracking in Clutter*. BiblioScholar, Sydney, Australia.
- Piché, R., 2016. Cramér-Rao lower bound for linear filtering with *t*-distributed measurement noise. *19th Int. Conf. on Information Fusion*, p.536-540.
- Piché, R., Särkkä, S., Hartikainen, J., 2012. Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-*t* distribution. *IEEE Int. Workshop on Machine Learning for Signal Processing*, p.1-6.
<https://doi.org/10.1109/MLSP.2012.6349794>
- Pishdad, L., Labeau, F., 2015. Analytic MMSE bounds in linear dynamic systems with Gaussian mixture noise statistics. *arXiv:1505.01765*.
<http://arxiv.org/abs/1506.07603>
- Qi, W., Zhang, P., Deng, Z., 2014. Robust weighted fusion Kalman filters for multisensor time-varying systems with uncertain noise variances. *Signal Process.*, **99**: 185-200.
<https://doi.org/10.1016/j.sigpro.2013.12.013>
- Raitoharju, M., Svensson, L., García-Fernández, Á.F., et al., 2017. Damped posterior linearization filter. *arXiv:1704.01113*. <http://arxiv.org/abs/1704.01113>
- Rasmussen, C.E., Williams, C.K.I., 2005. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, USA.
- 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.
<https://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. on Information Fusion*, p.1663-1670.
<https://doi.org/10.23919/ICIF.2017.8009853>
- Rosti, A.V.I., Gales, M.J.F., 2003. *Switching linear dynamical systems for speech recognition*. UK Speech Meeting.

- Roth, M., Özkan, E., Gustafsson, F., 2013. A Student's t filter for heavy tailed process and measurement noise. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, p.5770-5774. <https://doi.org/10.1109/ICASSP.2013.6638770>
- Roth, M., Hendeby, G., Gustafsson, F., 2016. Nonlinear Kalman filters explained: a tutorial on moment computations and sigma point methods. *J. Adv. Inform. Fus.*, **11**(1):47-70.
- Roth, M., Ardeshiri, T., Özkan, E., et al., 2017a. Robust Bayesian filtering and smoothing using student's t distribution. arXiv:1703.02428. <http://arxiv.org/abs/1703.02428>
- Roth, M., Hendeby, G., Fritsche, C., et al., 2017b. The ensemble Kalman filter: a signal processing perspective. arXiv:1702.08061. <http://arxiv.org/abs/1702.08061>
- 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. <https://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. <https://doi.org/10.1109/TAES.2007.4383588>
- Salmond, D.J., 1990. Mixture reduction algorithms for target tracking in clutter. *SPIE*, **1305**:434-445. <https://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. Inform. Fus.*, **11**(1):31-46. <https://doi.org/10.1016/j.automatica.2014.08.030>
- Sarmavuori, J., Särkkä, S., 2012. Fourier-Hermite Kalman filter. *IEEE Trans. Autom. Contr.*, **57**(6):1511-1515. <https://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. *Mach. Learn.*, **37**(1):75-87. <https://doi.org/10.1023/A:1007649326333>
- Scardua, L.A., da Cruz, J.J., 2017. Complete offline tuning of the unscented Kalman filter. *Automatica*, **80**:54-61. <https://doi.org/10.1016/j.automatica.2017.01.008>
- Schieferdecker, D., Huber, M.F., 2009. Gaussian mixture reduction via clustering. 12th Int. Conf. on Information Fusion, p.1536-1543.
- Šimandl, M., Duník, J., 2009. Derivative-free estimation methods: new results and performance analysis. *Automatica*, **45**(7):1749-1757. <https://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):1133-1145. <https://doi.org/10.1016/j.automatica.2006.03.010>
- Simon, D., 2006. Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. John Wiley & Sons, New York, USA.
- Simon, D., 2010. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Contr. Theory Appl.*, **4**(8):1303-1318. <https://doi.org/10.1049/iet-cta.2009.0032>
- Singpurwalla, N.D., Polson, N.G., Soyer, R., 2017. From least squares to signal processing and particle filtering. *Technometrics*, **2017**:1-15. <https://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. <https://doi.org/10.1016/j.physleta.2010.04.032>
- Snidaro, L., García, J., Llinas, J., 2015. Context-based information fusion: a survey and discussion. *Inform. Fus.*, **25**(Supplement C):16-31. <https://doi.org/10.1016/j.inffus.2015.01.002>
- Song, P., 2000. Monte Carlo Kalman filter and smoothing for multivariate discrete state space models. *Can. J. Statist.*, **28**(3):641-652. <https://doi.org/10.2307/3315971>
- Sorenson, H.W., 1970. Least-squares estimation: from Gauss to Kalman. *IEEE Spectr.*, **7**(7):63-68. <https://doi.org/10.1109/MSPEC.1970.5213471>
- Sorenson, H., Alspach, D., 1971. Recursive Bayesian estimation using Gaussian sums. *Automatica*, **7**(4):465-479. [https://doi.org/10.1016/0005-1098\(71\)90097-5](https://doi.org/10.1016/0005-1098(71)90097-5)
- Sornette, D., Ide, K., 2001. The Kalman-Lévy filter. *Phys. D*, **151**(2-4):142174. [https://doi.org/10.1016/S0167-2789\(01\)00228-7](https://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. Contr.*, **55**(6):1358-1366. <https://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. <https://doi.org/10.1109/TSMCC.2012.2230254>
- Steinbring, J., Hanebeck, U.D., 2014. Progressive Gaussian filtering using explicit likelihoods. 17th Int. Conf. on Information Fusion, p.1-8.
- Stoica, P., Babu, P., 2011. The Gaussian data assumption leads to the largest Cramér-Rao bound [lecture notes]. *IEEE Signal Process. Mag.*, **28**(3):132-133. <https://doi.org/10.1109/MSP.2011.940411>
- Stoica, P., Moses, R.L., 1990. On biased estimators and the unbiased Cramér-Rao lower bound. *Signal Process.*, **21**(4): 349350. [https://doi.org/10.1016/0165-1684\(90\)90104-7](https://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):273-286. <https://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):2657-2664. <https://doi.org/10.1016/j.automatica.2014.08.030>
- Su, J., Chen, W.H., 2017. Model-based fault diagnosis system verification using reachability analysis. *IEEE Trans. Syst. Man Cybern. Syst.*, **99**:1-10. <https://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**:149-154. <https://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. Contr. Eng.*, **3**(1):445-452. <https://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. Meas. Contr.*, **138**(9):091013. <https://doi.org/10.1115/1.4033631>
- Svensson, A., Schön, T.B., Lindsten, F., 2017. Learning of state-space models with highly informative observations: a tempered sequential Monte Carlo solution. arXiv:1702.01618. <http://arxiv.org/abs/1702.01618>
- Tahk, M., Speyer, J.L., 1990. Target tracking problems subject to kinematic constraints. *IEEE Trans. Autom. Contr.*, **35**(3):324-326. <https://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 Contr.*, **20**(1):45-57. <https://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. Contr.*, **56**(9):2151-2156. <https://doi.org/10.1109/TAC.2011.2141550>
- Tichavsky, P., Muravchik, C.H., Nehorai, A., 1998. Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. *IEEE Trans. Signal Process.*, **46**(5):1386-1396. <https://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):123-141. <https://doi.org/10.1016/j.neucom.2005.02.016>
- van der Merwe, R., Doucet, A., de Freitas, N., et al., 2000. The unscented particle filter. Proc. NIPS, p.563-569.
- van Trees, H.L., 1968. Detection, Estimation and Modulation Theory. Wiley, New York, USA.
- van Trees, H.L., Bell, K.L., 2007. Bayesian bounds for parameter estimation and nonlinear filtering/tracking. *IET Radar Sonar Navig.*, **3**(3):285-286. <https://doi.org/10.1049/iet-rsn:20099030>
- Vo, B.N., Ma, W.K., 2006. The Gaussian mixture probability hypothesis density filter. *IEEE Trans. Signal Process.*, **54**(11):4091-4104. <https://doi.org/10.1109/TSP.2006.881190>
- Wang, J.M., Fleet, D.J., Hertzmann, A., 2008. Gaussian process dynamical models for human motion. *IEEE Trans. Patt. Anal. Mach. Intell.*, **30**(2):283-298. <https://doi.org/10.1109/TPAMI.2007.1167>
- 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. Mob. Comput.*, **11**(4):567-576. <https://doi.org/10.1109/TMC.2011.59>
- 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**:1011-1024. <https://doi.org/10.1016/j.amc.2013.12.168>
- Wang, X., Liang, Y., Pan, Q., et al., 2015. Nonlinear Gaussian smoothers with colored measurement noise. *IEEE Trans. Autom. Contr.*, **60**(3):870-876. <https://doi.org/10.1109/TAC.2014.2337991>
- Wang, X., Song, B., Liang, Y., et al., 2017. EM-based adaptive divided difference filter for nonlinear system with multiplicative parameter. *Int. J. Robust Nonl. Contr.*, **27**(13):2167-2197. <https://doi.org/10.1002/rnc.3674>
- Wen, W., Durrant-Whyte, H.F., 1992. Model-based multi-sensor data fusion. Proc. IEEE Int. Conf. on Robotics and Automation, p.1720-1726. <https://doi.org/10.1109/ROBOT.1992.220130>
- Williams, J.L., Maybeck, P.S., 2006. Cost-function-based hypothesis control techniques for multiple hypothesis tracking. *Math. Comput. Model.*, **43**(9-10):976-989. <https://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. <https://doi.org/10.1109/TSP.2006.875389>
- Wu, Z., Shi, J., Zhang, X., et al., 2015. Kernel recursive maximum correntropy. *Signal Process.*, **117**:11-16. <https://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. <https://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. <https://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. <https://doi.org/10.1109/TAES.2009.4805264>
- Yang, T., Laugesen, R.S., Mehta, P.G., et al., 2016. Multi-variable feedback particle filter. *Automatica*, **71**:10-23. <https://doi.org/10.1016/j.automatica.2016.04.019>
- Yang, Y., He, H., Xu, G., 2001. Adaptively robust filtering for kinematic geodetic positioning. *J. Geod.*, **75**(2-3):109-116. <https://doi.org/10.1007/s001900000157>
- Yi, D., Su, J., Liu, C., et al., 2016. Data-driven situation awareness algorithm for vehicle lane change. 19th IEEE Int. Conf. on Intelligent Transportation Systems, p.998-1003. <https://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. *Automatica*, **63**:321-329. <https://doi.org/10.1016/j.automatica.2015.10.040>
- Zanetti, R., 2012. Recursive update filtering for nonlinear estimation. *IEEE Trans. Autom. Contr.*, **57**(6):1481-1490. <https://doi.org/10.1109/TAC.2011.2178334>
- Zen, H., Tokuda, K., Kitamura, T., 2007. Reformulating the HMM as a trajectory model by imposing explicit relationships between static and dynamic feature vector sequences. *Comput. Speech Lang.*, **21**(1):153-173. <https://doi.org/10.1016/J.CSL.2006.01.002>

- Zhan, R., Wan, J., 2007. Iterated unscented Kalman filter for passive target tracking. *IEEE Trans. Aerosp. Electron. Syst.*, **43**(3):1155-1163.
https://doi.org/10.1109/TAES.2007.4383605
- Zhang, C., Zhi, R., Li, T., et al., 2016. Adaptive M-estimation for robust cubature Kalman filtering. *Sensor Signal Processing for Defence*, p.114-118.
https://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 Process.*, **114**:112-116.
https://doi.org/10.1016/j.sigpro.2015.02.022
- Zhao, S., Shmaliy, Y.S., Liu, F., 2016a. Fast Kalman-like optimal unbiased FIR filtering with applications. *IEEE Trans. Signal Process.*, **64**(9):2284-2297.
https://doi.org/10.1109/TSP.2016.2516960
- Zhao, S., Shmaliy, Y.S., Liu, F., et al., 2016b. Unbiased, optimal, and in-between: the trade-off in discrete finite impulse response filtering. *IET Signal Process.*, **10**(4): 325-334.
https://doi.org/10.1049/iet-spr.2015.0360
- Zheng, Y., Ozdemir, O., Niu, R., et al., 2012. New conditional posterior Cramér-Rao lower bounds for nonlinear sequential Bayesian estimation. *IEEE Trans. Signal Process.*, **60**(10):5549-5556.
https://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. Contr.*, **65**(2):295-307.
https://doi.org/10.1080/00207179608921698
- Zuo, L., Niu, R., Varshney, P.K., 2011. Conditional posterior Cramér-Rao lower bounds for nonlinear sequential Bayesian estimation. *IEEE Trans. Signal Process.*, **59**(1):1-14.
https://doi.org/10.1109/TSP.2010.2080268

Appendix A: 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.,

$$I(\mathbf{x}) \triangleq E \left[\left(\frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) \right)^2 \middle| \mathbf{x} \right] \\ = \int \left(\frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) \right)^2 p(\mathbf{y}|\mathbf{x}) d\mathbf{y}, \quad (\text{A1})$$

where, for any \mathbf{x} , $E[\cdot|\mathbf{x}]$ denotes the conditional expectation over \mathbf{y} with respect to 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 \mathbf{x} . Note that $0 \leq I(\mathbf{x}) < \infty$. For any unbiased estimator $\hat{\mathbf{x}}(\mathbf{y})$, the Cramér-Rao

inequality reads

$$\text{Var}(\hat{\mathbf{x}}(\mathbf{y})) \geq \frac{1}{I(\mathbf{x})}. \quad (\text{A2})$$

In statistics, it is

$$\text{MSE}(\hat{\mathbf{x}}(\mathbf{y})) \geq \frac{1}{I(\mathbf{x})}, \quad (\text{A3})$$

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

Appendix B: Closed-form recursion of Kalman filtering

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

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{v}_t - E[\mathbf{v}_t], \quad (\text{B1})$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{w}_t - E[\mathbf{w}_t], \quad (\text{B2})$$

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

Prediction (time updating):

$$\hat{\mathbf{x}}_{t|t-1} = \int \mathbf{F}_t \mathbf{x}_{t-1} \mathcal{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1}) d\mathbf{x}_{t-1}, \quad (\text{B3})$$

$$\mathbf{P}_{t|t-1} = \int \mathbf{F}_t \mathbf{x}_{t-1} (\mathbf{F}_t \mathbf{x}_{t-1})^T \mathcal{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1}) d\mathbf{x}_{t-1} \\ - \hat{\mathbf{x}}_{t|t-1} \hat{\mathbf{x}}_{t|t-1}^T + \mathbf{Q}_t, \quad (\text{B4})$$

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):

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{G}_t (\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}), \quad (\text{B5})$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{G}_t \mathbf{P}_{yy} \mathbf{G}_t^T, \quad (\text{B6})$$

where

$$\mathbf{G}_t = \mathbf{P}_{xy} \mathbf{P}_{yy}^{-1}, \quad (\text{B7})$$

$$\hat{\mathbf{y}}_{t|t-1} = \int \mathbf{H}_t \mathbf{x}_t \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}) d\mathbf{x}_t, \quad (\text{B8})$$

$$\mathbf{P}_{xy} = \int (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{H}_t \mathbf{x}_t - \hat{\mathbf{y}}_{t|t-1})^T \cdot \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}) d\mathbf{x}_t, \quad (\text{B9})$$

$$\mathbf{P}_{yy} = \int (\mathbf{H}_t \mathbf{x}_t - \hat{\mathbf{y}}_{t|t-1})(\mathbf{H}_t \mathbf{x}_t - \hat{\mathbf{y}}_{t|t-1})^T \cdot \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}) d\mathbf{x}_t + \mathbf{R}_t. \quad (\text{B10})$$

Appendix C: A very informative observation state space model

The SSM is given as follows (van der Merwe *et al.*, 2000):

$$x_t = 1 + \sin(0.04\pi t) + 0.5x_t + v_t, \quad (\text{C1})$$

$$y_t = \begin{cases} 0.2x_t^2 + w_t, & t \leq 30, \\ 0.5x_t - 2 + w_t, & t > 30, \end{cases} \quad (\text{C2})$$

where process noise v_t is a Gamma random variable $\Gamma(3, 2)$, the observation noise is Gaussian $w_t \sim \mathcal{N}(v; 0, 10^{-5})$, and the default simulation length is 60 iteration steps.



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 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).

He has been on the Editorial Board of peer-reviewed journals: *Advances in Distributed Computing and Artificial Intelligence Journal* (2016–) and *Frontiers of Information Technology & Electronic Engineering* (2017–), and on the organizing and/or technical program committee of several international symposiums and conferences, including FUSION 2014–2017, DCAI 2015–2017, and ACM-SAC 2015–2017.

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.