18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

A New Nonlinear State Estimator Using the Fusion of Multiple Extended Kalman Filters

Zhansheng Duan, Xiaoyun Li
Center for Information Engineering Science Research
College of Electronics and Information Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi 710049, China
Email: zduan@uno.edu

***Abstract*—For linear systems, the optimal filtering is provided by the celebrated Kalman filter. For nonlinear systems, only suboptimal filters can be obtained in general. The Extended Kalman filter (EKF) is such a suboptimal filter. It helped the promotion of the Kalman filter. With the development of more advanced nonlinear filters, however, the EKF is receiving less and less attention because it performs the worst most often. The EKF is based on the first-order Taylor series expansion. Ideally, the ground truth of the state should be picked as the expansion points, which are unfortunately unavailable in estimation problem. Instead, the most recent estimates are used. As a result of this misspecification, the EKF may have degraded performance or even failure. To overcome this, a multiple model extension to the EKF is proposed in this paper. Its key idea is to use multiple probabilistically weighted points to represent the whole state space. Then the linearization about each weighted point will lead to a possible model. Correspondingly, the original nonlinear filtering problem is changed into a variable structure multi-model estimation problem. How to design finite number of probabilistically weighted points to approximate the posterior densities is suggested. Numerical examples show that the proposed extension to the EKF is quite promising when compared to several existing competitive nonlinear filters.**

Keywords: Nonlinear filtering, extended Kalman filter, Taylor series expansion, Gaussian assumption, multiple model estimation, model set design.

I. INTRODUCTION

Due to the introduction of the state-space formulation, the Kalman filter [1] was a breakthrough for the estimation of dynamic systems compared with the Wiener filter. It is the optimal filter for linear systems. However, the assumptions required by the Kalman filter are too stringent. For example, the system has to be linear Gaussian and driven by uncorrelated white noises. Unfortunately these assumptions can not be met in many real applications. A main reason is that most systems in reality are nonlinear. This prevents the direct use of the Kalman filter for nonlinear systems. To still apply the Kalman filter for nonlinear systems, some modifications to it then had to be made. Among all, the EKF is probably the earliest and most well-known one. The key idea of it is to linearize the nonlinear system using the first-order Taylor series expansion (TSE) so that the resultant linearized system fits the assumptions of the Kalman filter. This extension is very natural and a very typical engineering way to handle nonlinear problems.

Research supported in part by National 973 project of China through grant 2013CB329405, National Natural Science Foundation of China through grant 61174138, and the Fundamental Research Funds for the Central Universities of China.

Historically, the EKF helped boost the use of the Kalman filter in practical applications, e.g., navigation, orbit determination, target tracking, weather forecasting, finance, etc. However, performance degradation or even failure were also reported when the EKF was applied in many nonlinear problems. A lot of work had been done on how to improve the performance of the EKF. For example, heuristic ways [2] include increasing the process noise covariance matrix or multiplying the state prediction covariance matrix by a fudge factor (slightly) larger than unity. The second-order EKF uses the Taylor series expansion up to the second order and the iterated EKF iteratively linearizes the system about the estimates from the last iteration.

In the recent years, more and more advanced nonlinear filters were developed, including the Unscented filtering [3], [4], Cubature Kalman filter [5], Divided Difference filters [6] (DD1, DD2), Gaussian-Hermite Quadrature filter [7], [8], linear regression Kalman filter [9], etc. All these nonlinear filters are based on the linear minimum mean-squared error (LMMSE) estimator [10]. Different nonlinear filters just have different ways to approximate the first two moments required by the LMMSE estimator. In the Unscented filtering, the first two moments are approximated by the Unscented transformation which depends on well selected sigma-points. The Divided Difference filters use the first-order and second-order Stirling's interpolation to approximate the required first two moments, which lead to DD1 and DD2, respectively. The Cubature Kalman filter is based on a third-degree spherical-radial cubature rule that provides a set of cubature points. Based on a moment-matching Gaussian assumption, the Gaussian-Hermite Quadrature filter uses Gauss-Hermite quadratures to compute the first-two moments. The linear regression Kalman filter first finds the optimal statistical linearization of the nonlinear system in the LMMSE sense and then applies the Kalman filter as done in the EKF.

Another significant progress on nonlinear filtering in the recent years is the particle filter [11]. The goal of the particle filter is to approximate the posterior density of the estimand

(the quantity to be estimated) by a bunch of randomly generated particles. This is different from the goal of the nonlinear filters introduced above, which all intend to obtain just the first two moments of the posterior density. However, there seems to be an overuse of the power of the particle filter especially when the goal of nonlinear filtering is not to obtain an estimate of the posterior density but just its first two moments. For example, in most target tracking applications, our goal is just to provide estimates of the states of the targets of interest and associated evaluation as to how good the estimates are. That is, we are just interested in the first two moments of the states of the targets. But the particle filter is still used in some tracking applications. This is mainly motivated by the fact that once the density is known, we can easily figure out the associated first two moments. Nevertheless, to obtain the whole density is much harder than just to obtain its first two moments in general.

With the increasing development of more advanced nonlinear filters, as demonstrated by many comparisons in the literature and practical applications, there seems to be a common view that the EKF is probably the worst among all. However, they are not sufficient to suggest a complete replacement of the EKF by some other nonlinear filters, such as the Unscented filter or Cubature Kalman filter, etc. It is well known that the EKF hinges on the TSE about an expansion point. In a small neighborhood of the expansion point, the TSE does provide a good approximation to the original nonlinear function. However, outside the small neighborhood, the approximation accuracy could be very poor. Ideally, the expansion point should be the ground truth of the system state. Unfortunately the ground truth is not available because it is exactly the quantity to be estimated in all estimation problems. Therefore, in the EKF, we just replace the ground truth by its best estimate. That is why the expansion point for the nonlinear dynamic system is the previously updated estimate and the expansion point for the nonlinear measurement system is the predicted estimate. However, the estimate is just the filter-calculated "best" one and it is not necessarily the truly best one. Even if the estimate is really the best one, there may still be a significant gap between the estimate and the ground truth. So the reason for the performance degradation or even failure of the EKF in some cases is not that the framework of the TSE is problematic. On the contrary, the TSE based linearization is so natural and has found successful use in tons of engineering applications. The performance degradation or even failure should be mainly due to the use of inappropriate expansion points. If we can find better expansion point, the EKF should still be a quite competitive nonlinear filter. So the purpose of this paper is to develop a new better performed nonlinear filter still based on the framework of the TSE.

The performance degradation or even failure of the EKF can also be understood from the angle of decision process. Given the posterior density of the estimand, the EKF just picks its mean as the expansion point. In terms of decision process, this is a hard decision. It eliminates the possibility of the other points in the state space completely although all of them are also possible. As we all know, nonlinear filtering is state estimation for nonlinear systems. The effect of the gap between the hard decision, i.e., the "best" estimate, and the ground truth may accumulate over time. If the hard decision deviates from the ground truth too much at a time instant, there is almost no way for the successive expansion points to be closer to the ground truth. So from this angle, to reduce the decision error, a soft decision is preferred. That is, each point in the state space should be given a probabilistic chance. Unfortunately, the state space is continuous and computationally it is infeasible if we try to enumerate all points probabilistically in the whole state space.

In this paper, a multiple model extension of the EKF is proposed. Its key idea is to use multiple probabilistically weighted points to represent the whole state space so that the linearization about each point will give a possible model. Then the original nonlinear filtering problem is naturally changed into a variable structure multi-model estimation problem. How to design a finite number of probabilistically weighted points to approximate a continuous state space is suggested. Numerical examples show that the proposed framework is quite competitive relative to several existing popular nonlinear filters.

This paper is organized as follows. Sec. II states the problem. Sec. III analyzes the reason for performance degradation or even failure of the EKF in some cases. Sec. IV presents the proposed multi-model extension of the EKF. Sec. V provides suggestions on how to approximate a continuous probability density function (pdf) by a discrete probability mass function (pmf). Sec. VI provides numerical examples to illustrate the effectiveness and efficiency of the proposed nonlinear filter. Sec. VII concludes the paper.

II. Problem formulation

Consider the following typical form of a nonlinear stochastic dynamic system

$$x_{k+1} = f_k(x_k) + w_k$$

which is observed through the following nonlinear measurement model

$$z_k = h_k(x_k) + v_k$$

It is assumed that $\langle w_k \rangle$ and $\langle v_k \rangle$ are both zero-mean white Gaussian noise sequences with covariance Q_k and R_k, respectively, and uncorrelated with each other. Also, $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$ and x_0 is uncorrelated with both $\langle w_k \rangle$ and $\langle v_k \rangle$.

Remark 1: For simplicity, both $\langle w_k \rangle$ and $\langle v_k \rangle$ are assumed to additive. However, all discussions below are also directly applicable to non-additive cases.

For a state filtering problem, our goal is to find the best estimate of the system state x_k given the measurement sequence $z^k = \{z_1, \cdots, z_k\}$ up to the most recent time instant k. It is well known that the optimal estimate of x_k in the sense of minimum mean-squared error (MMSE) is provided by the conditional mean $E[x_k|z^k]$ and its associated MSE matrix $\mathrm{cov}(x_k|z^k)$, which are the first two moments of the posterior

distribution $p(x_k|z^k)$ of x_k. However, due to the nonlinearity of the dynamic system and measurement model, in most cases only some kind of approximations to them can be found in reality. So our goal of state filtering is changed to find a good approximation to $E[x_k|z^k]$ and $\mathrm{cov}(x_k|z^k)$.

III. Performance analysis of the extended Kalman filter

For linear Gaussian systems, the Kalman filter can obtain $E[x_k|z^k]$ and $\mathrm{cov}(x_k|z^k)$ exactly. However, it can not be directly applied to nonlinear systems. The EKF is an extension of the Kalman filter for nonlinear cases. It provides an approximation to $E[x_k|z^k]$ and $\mathrm{cov}(x_k|z^k)$. The idea of it is to use the first-order TSE to approximate the original nonlinear systems so that the Kalman filter can be applied. However, to use the TSE, a key issue is the selection of the expansion point. Ideally, the expansion point should be the ground truth of the system state. That is, ideally one cycle of the EKF (abbreviated as **iEKF** below) should be as follows.

Time update:

$$\begin{aligned}\hat{x}_{k|k-1} &= E[x_k|z^{k-1}]\\ &\approx f_{k-1}(x^*_{k-1}) + \bar{F}_{k-1}(\hat{x}_{k-1|k-1} - x^*_{k-1})\\ x_k - \hat{x}_{k|k-1} &\approx \bar{F}_{k-1}(x_{k-1} - \hat{x}_{k-1|k-1})\\ P_{k|k-1} &= \mathrm{cov}(x_k|z^{k-1}) \approx \bar{F}_{k-1}P_{k-1|k-1}\bar{F}'_{k-1} + Q_{k-1}\end{aligned}$$

where

$$\bar{F}_{k-1} = \left.\frac{\partial f_{k-1}(x_{k-1})}{\partial x_{k-1}}\right|_{x_{k-1}=x^*_{k-1}}$$

and x^*_{k-1} is the ground truth of x_{k-1}.

Measurement update:

$$\begin{aligned}\hat{z}_{k|k-1} &= E[z_k|z^{k-1}] \approx h_k(x^*_k) + \bar{H}_k(\hat{x}_{k|k-1} - x^*_k)\\ \tilde{z}_{k|k-1} &= z_k - \hat{z}_{k|k-1} \approx \bar{H}_k(x_k - \hat{x}_{k|k-1}) + v_k\\ S_k &= \mathrm{cov}(z_k|z^{k-1}) \approx \bar{H}_k P_{k|k-1}\bar{H}'_k + R_k\\ K_k &= \mathrm{cov}(\tilde{x}_{k|k-1}, \tilde{z}_{k|k-1}) \approx P_{k|k-1}\bar{H}'_k S_k^{-1}\\ \hat{x}_{k|k} &= E[x_k|z^k] \approx \hat{x}_{k|k-1} + K_k(z_k - \hat{z}_{k|k-1})\\ \tilde{x}_{k|k} &\approx x_k - \hat{x}_{k|k-1} - K_k(z_k - \hat{z}_{k|k-1})\\ &= (I - K_k\bar{H}_k)(x_k - \hat{x}_{k|k-1}) - K_k v_k\\ P_{k|k} &= \mathrm{cov}(x_k|z^k) \approx P_{k|k-1} - P_{k|k-1}\bar{H}'_k S_k^{-1}\bar{H}_k P_{k|k-1}\end{aligned}$$

where

$$\bar{H}_k = \left.\frac{\partial h_k(x_k)}{\partial x_k}\right|_{x_k=x^*_k}$$

and x^*_k is the ground truth of x_k.

Unfortunately both x^*_{k-1} and x^*_k are not available because they are exactly the quantities to be estimated. To overcome this, they are replaced by their best estimates in the EKF. One cycle of the EKF can be summarized as follows.

Time update:

$$\begin{aligned}\hat{x}_{k|k-1} &= E[x_k|z^{k-1}] \approx f_{k-1}(\hat{x}_{k-1|k-1}) \qquad (1)\\ P_{k|k-1} &= \mathrm{cov}(x_k|z^{k-1}) \approx F_{k-1}P_{k-1|k-1}F'_{k-1} + Q_{k-1}\end{aligned}$$

where

$$F_{k-1} = \left.\frac{\partial f_{k-1}(x_{k-1})}{\partial x_{k-1}}\right|_{x_{k-1}=\hat{x}_{k-1|k-1}}$$

Measurement update:

$$\begin{aligned}\hat{z}_{k|k-1} &= E[z_k|z^{k-1}] \approx h_k(\hat{x}_{k|k-1}) \qquad (2)\\ S_k &= \mathrm{cov}(z_k|z^{k-1}) \approx H_k P_{k|k-1}H'_k + R_k\\ K_k &\approx P_{k|k-1}H'_k S_k^{-1}\\ \hat{x}_{k|k} &= E[x_k|z^k] \approx \hat{x}_{k|k-1} + K_k(z_k - \hat{z}_{k|k-1})\\ P_{k|k} &= \mathrm{cov}(x_k|z^k) \approx P_{k|k-1} - P_{k|k-1}H'_k S_k^{-1}H_k P_{k|k-1}\end{aligned}$$

where

$$H_k = \left.\frac{\partial h_k(x_k)}{\partial x_k}\right|_{x_k=\hat{x}_{k|k-1}}$$

As can be clearly seen from the above, the key of the EKF is the use of the first-order TSE. Ideally, the expansion point for the nonlinear state transition function $f_{k-1}(x_{k-1})$ should be x^*_{k-1} and the expansion point for the nonlinear measurement function $h_k(x_k)$ should be x^*_k. But due to the unavailability of them, they are replaced by their "best" estimates $\hat{x}_{k-1|k-1}$ and $\hat{x}_{k|k-1}$, respectively. However, on one hand, they are just the filter-calculated "best" estimates, but not necessarily the truly best ones. On the other hand, even if they are the truly best ones, there is still a gap between them and the ground truth due to the non-reducible estimation errors. Depending on the degree of nonlinearity of the system models and the setting up of the used system parameters, this gap may even lead to the collapse (divergence) of the EKF. So inspired by this, unlike the abundant existing developments to nonlinear filtering, our idea is to still use the TSE based framework of EKF but to find better expansion points. It can be clearly seen that the closer the expansion points are to the ground truth, the better they are.

IV. Multi-model extended Kalman filter

In the EKF, the filter-calculated "best" estimates are selected as the expansion points first, and then the Kalman filter for linear case is applied. This follows a decision then estimation procedure. But why are the filter-calculated "best" estimates selected as the expansion points? This is because they are thoughts as the closest to the ground truth in the state space. For example, the MMSE-optimal or LMMSE-optimal criteria defines different closeness measures. From the point of view of decision procedure, this is a hard decision since only one specific point is selected from the state space. Inspired by the work of maneuvering target tracking, where soft decision based multi-model estimation is now the mainstream because of its prevailing advantage, soft decision should also be preferred for the selection of expansion points for the TSE used in the EKF. In soft decision, the decision is the pmf of all possible decision choices if there are countably many of them. That is, the decision is the possibility of each choice. Whereas in hard decision, the decision is just a single point of the set of all possible decision choices. For example, we can simply pick up the one with the highest probability. From probability theory,

we know that although the probability of an event may be very small, this does not mean that it has no chance to happen. So by also considering the other choices with relatively lower probabilities, soft decision seems to be able to help alleviate the decision errors associated with expansion point across time to certain extent.

First, suppose that $p(x_{k-1}|z^{k-1})$ can be approximated by a discrete pmf

$$P\{x_{k-1} = \hat{x}^i_{k-1}|z^{k-1}\} = \hat{w}^i_{k-1},\ i = 1, \cdots, N$$

where

$$\sum\nolimits_{i=1}^{N} \hat{w}^i_{k-1} = 1,\ \hat{w}^i_{k-1} \geq 0$$

Then by choosing these N points as the expansion points of the first-order TSEs of the state transition model, we will have N linearized state transition models as

$$\hat{\mathcal{M}}^i_{k-1} : x_k = f_{k-1}(\hat{x}^i_{k-1}) + F^i_{k-1}(x_{k-1} - \hat{x}^i_{k-1}) + w_{k-1}$$

where

$$P\{\hat{\mathcal{M}}^i_{k-1}|z^{k-1}\} = \hat{w}^i_{k-1}$$
$$F^i_{k-1} = \left.\frac{\partial f_{k-1}(x_{k-1})}{\partial x_{k-1}}\right|_{x_{k-1}=\hat{x}^i_{k-1}}$$

Furthermore, suppose that $p(x_k|z^{k-1})$ can also be approximated by a discrete pmf

$$P\{x_k = \bar{x}^j_k|z^{k-1}\} = \bar{w}^j_k,\ j = 1, \cdots, M$$

where

$$\sum\nolimits_{j=1}^{M} \bar{w}^j_k = 1,\ \bar{w}^j_k \geq 0$$

Then by choosing these M points as the expansion points of the first-order TSEs of the measurement equation, we will have M linearized measurement models as

$$\bar{\mathcal{M}}^j_k : z_k = h_k(\bar{x}^j_k) + H^j_k(x_k - \bar{x}^j_k) + v_k$$

where

$$P\{\bar{\mathcal{M}}^j_k|z^{k-1}\} = \bar{w}^j_k,\ H^j_k = \left.\frac{\partial h_k(x_k)}{\partial x_k}\right|_{x_k=\bar{x}^j_k}$$

By choosing $\{\hat{\mathcal{M}}^i_{k-1}\}^N_{i=1}$ as the model set for state transition from $k-1$ to k and $\{\bar{\mathcal{M}}_j\}^M_{j=1}$ as the model set for measurement model at k, it can be clearly seen that the original nonlinear filtering problem is now changed into a multi-model estimation problem [12].

Remark 2: Two points should be noted about the new multi-model estimation problem induced from nonlinear filtering. First, there is uncertainty not just in the state transition model, but also in the measurement model. Second, the new multi-model estimation problem is in essence a variable structure multi-model filtering problem [13] since the model set is changing over time.

One cycle of the multi-model extension of the extended Kalman filter can be summarized as follows.

Time update:

Time update purely conditioned on $\hat{\mathcal{M}}^i_{k-1}$ $(i = 1, \cdots, N)$:

$$\begin{aligned}
\hat{x}^i_{k|k-1} &\approx f_{k-1}(\hat{x}^i_{k-1}) + F^i_{k-1}(\hat{x}_{k-1|k-1} - \hat{x}^i_{k-1}) \quad (3)\\
x_k - \hat{x}^i_{k|k-1} &\approx F^i_{k-1}(x_{k-1} - \hat{x}_{k-1|k-1})\\
P^i_{k|k-1} &\approx F^i_{k-1}P_{k-1|k-1}(F^i_{k-1})' + Q_{k-1}
\end{aligned}$$

Remark 3: Note that time update (3) purely conditioned on $\hat{\mathcal{M}}^i_{k-1}$ is different from the time update (1) in the classical EKF in that an additional term $F^i_{k-1}(\hat{x}_{k-1|k-1} - \hat{x}^i_{k-1})$ is added. This extra term can be thought as a correction term to the original prediction step of the EKF.

The combined multi-model prediction is then:

$$\begin{aligned}
\hat{x}_{k|k-1} &= \sum\nolimits_{i=1}^{N} \hat{w}^i_{k-1}\hat{x}^i_{k|k-1}\\
P_{k|k-1} &= \sum\nolimits_{i=1}^{N} \hat{w}^i_{k-1}P^i_{k|k-1}\\
&\quad + \sum\nolimits_{i=1}^{N} \hat{w}^i_{k-1}(\hat{x}^i_{k|k-1} - \hat{x}_{k|k-1})(\hat{x}^i_{k|k-1} - \hat{x}_{k|k-1})'
\end{aligned}$$

Measurement update:

Measurement update purely conditioned on $\bar{\mathcal{M}}^j_k$ $(j = 1, \cdots, M)$:

$$\begin{aligned}
\hat{z}^j_{k|k-1} &\approx h_k(\bar{x}^j_k) + H^j_k(\hat{x}_{k|k-1} - \bar{x}^j_k) \quad (4)\\
\tilde{z}^j_{k|k-1} &= z_k - \hat{z}^j_{k|k-1} \approx H^j_k(x_k - \hat{x}_{k|k-1}) + v_k\\
S^j_{k|k-1} &\approx H^j_k P_{k|k-1}(H^j_k)' + R_k\\
K^j_k &= \mathrm{cov}(\tilde{x}_{k|k-1}, \tilde{z}^j_{k|k-1}) \approx P_{k|k-1}(H^j_k)'(S^j_{k|k-1})^{-1}\\
\hat{x}^j_{k|k} &\approx \hat{x}_{k|k-1} + K^j_k(z_k - \hat{z}^j_{k|k-1})\\
\tilde{x}^j_{k|k} &\approx x_k - \hat{x}_{k|k-1} - K^j_k(z_k - \hat{z}^j_{k|k-1})\\
&= (I - K^j_k H^j_k)(x_k - \hat{x}_{k|k-1}) - K^j_k v_k\\
P^j_{k|k} &\approx (I - K^j_k H^j_k)P_{k|k-1}(I - K^j_k H^j_k)' + K^j_k R_k (K^j_k)'
\end{aligned}$$

Remark 4: Similarly note that measurement prediction (4) purely conditioned on $\bar{\mathcal{M}}^j_k$ is also different from the measurement prediction (2) in the classical EKF in that an additional term $H^j_k(\hat{x}_{k|k-1} - \bar{x}^j_k)$ is added. This extra term can also be thought as a correction term to the original measurement prediction step of the EKF.

According to the Bayes' rule, the posterior model probability can be obtained as

$$\begin{aligned}
\breve{w}^j_k &= P\{\bar{\mathcal{M}}_j|z^k\}\\
&= \frac{p(z_k|\bar{\mathcal{M}}_j)\bar{w}^j_k}{\sum_{l=1}^{M} p(z_k|\bar{\mathcal{M}}_l)\bar{w}^l_k}
\end{aligned}$$

where

$$p(z_k|\bar{\mathcal{M}}_j) \approx \mathcal{N}(\tilde{z}^j_{k|k-1}; 0, S^j_{k|k-1})$$

Finally the combined multi-model update is then:

$$\begin{aligned}
\hat{x}_{k|k} &= \sum\nolimits_{j=1}^{M} \breve{w}^j_k \hat{x}^j_{k|k}\\
P_{k|k-1} &= \sum\nolimits_{j=1}^{M} \breve{w}^j_k P^j_{k|k}\\
&\quad + \sum\nolimits_{j=1}^{M} \breve{w}^j_k (\hat{x}^j_{k|k} - \hat{x}_{k|k})(\hat{x}^j_{k|k} - \hat{x}_{k|k-1})'
\end{aligned}$$

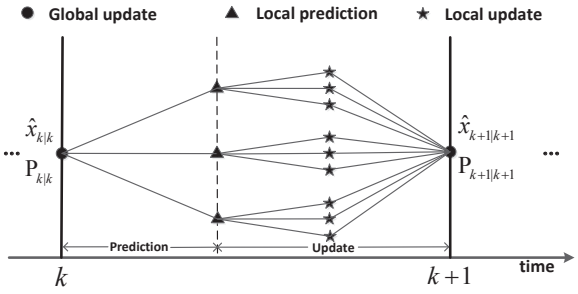

Figure 1. Once cycle of multi-model extended Kalman filter for $N = 3$

Remark 5: In the above, the M measurement models (expansion points) can be simply generated as follows. From each local predicted density with the first two moments $\hat{x}^i_{k|k-1}$ and $P^i_{k|k-1}$ $(i = 1, \cdots, N)$, we can again generate N mass points similarly and use them in the first-order TSEs of the measurement model. Then altogether we will have $M = N^2$ measurement models. One cycle of this procedure is illustrated in Fig. 1.

V. Model-set Design

In the above, to ease the discussion of the framework of the multi-model Extended Kalman filter, we just assume that both model sets $\{\hat{\mathcal{M}}^i_{k-1}\}_{i=1}^N$ and $\{\bar{\mathcal{M}}_j\}_{j=1}^N$ are given. However, how can we have them for nonlinear filtering problems. This is a model-set design problem and will be discussed next.

Model-set design problem is a critical component of multi-model estimation problem. However, most existing work focus on multi-model filtering algorithm development with known model-set. The existing work on systematic model-set design is very scarce. By introducing the concept of random model, three classes of general methods for optimal design of model sets—by minimizing distribution mismatch, minimizing modal distance, and moment matching, respectively—are proposed in [14]. All of them can be applied to our multi-model extended Kalman filter. For illustrative purpose, only the first class will be considered—by minimizing distribution mismatch—in this paper.

To apply the model-set design method using the concept of random model, we need to determine the distribution of the random model first. For nonlinear state estimation, given the updated estimate $\hat{x}_{k-1|k-1}$ and $P_{k-1|k-1}$ at $k-1$, similar to [7], we can simply assume that the posterior pdf is Gaussian,

$$p(x_{k-1}|z^{k-1}) \approx \mathcal{N}(x_{k-1}; \hat{x}_{k-1|k-1}, P_{k-1|k-1})$$

Then by model-set design, it is simply meant to find a discrete pmf to approximate this Gaussian distribution. That is, in essence, the design problem is just how to approximate a continuous distribution by a discrete distribution. Certainly we want the approximation error—the mismatch (difference) between the continuous distribution and the discrete distribution—to be as small as possible. So one natural idea is to set up an optimization objective function based on the error between the two distribution functions and then minimize it. In the following, we will provide some suggestions on how to achieve this approximation.

For the ease of discussion below, we assume that the quantity of interest is a scalar, the cdf (cumulative distribution function) of the scalar-valued continuous random variable is $F_c(x)$ and the cdf of a candidate discrete random variable is $F_d(x)$.

In [14], it is found that for a predetermined tolerance parameter ϵ, which puts a constraint on the candidate cdf $F_d(x)$

$$|F_c(x) - F_d(x)| < \epsilon, \ \forall x$$

the minimum number of mass points needed for $F_d(x)$ is simply $N = \lceil 1/2\epsilon \rceil$ = smallest integer not smaller than $1/2\epsilon$.

If the error of approximating $F_c(x)$ by $F_d(x)$ is measured by the *Kolmogorov-Smirnov distance* between two distributions

$$d(F_c(x), F_d(x)) = \sup_{x \in R} |F_c(x) - F_d(x)|$$

i.e., using this distance as the objective function for minimization, the approximation of a continuous distribution by a discrete distribution will be changed into the following min-max optimization problem

$$F^*_d(x) = \arg \inf_{F_d(x)} \sup_{x \in R} |F_c(x) - F_d(x)|$$

under the constraint that the number of mass point of $F_d(x)$ should be N. It seems that this is a functional optimization problem with respect to $F_d(x)$, which is not easy in general. Due to the fact that a discrete distribution is characterized by the number of mass points, the locations of the mass points and the probability at each mass point, however, the functional optimization problem over $F_d(x)$ can be changed into a much easier parameter optimization problem.

The optimal pmf for the above min-max problem is

$$P\{x = x^*_i\} = \frac{1}{N}, \ i = 1, \cdots, N$$
$$x^*_i = F_c^{-1}(\frac{i - 1/2}{N})$$

where $F_c^{-1}(\cdot)$ is the inverse function of $F_c(\cdot)$.

In [15], [16], the number of mass points N of $F_d(x)$ also needs to be predetermined. Unlike the *Kolmogorov-Smirnov distance* used above, the difference between $F_c(x)$ and $F_d(x)$ is measured by the *weighted Cramér-von Mises distance*

$$d(F_c(x), F_d(x)) = \int_{-\infty}^{+\infty} g(x)(F_c(x) - F_d(x))^2 dx$$

where $g(x)$ is a nonnegative weighting function and it is selected in such a way that only some portions, problem dependent, of $F_c(x)$ are approximated with high accuracy.

Correspondingly, the optimization problem is

$$F^*_d(x) = \arg \min_{F_d(x)} \int_{-\infty}^{+\infty} g(x)(F_c(x) - F_d(x))^2 dx$$

under the constraint that the number of mass point of $F_d(x)$ should be N.

When the N mass points are equally distributed and $g(x) = 1$, it is proved that the locations of the mass points are

$$x_i^* = F_c^{-1}(\frac{i-1/2}{N}),\ i = 1, \cdots, N$$

which are exactly the same as the locations of the mass points when the *Kolmogorov-Smirnov distance* is used.

For a scalar-valued continuous random variable x following an arbitrary distribution, we can use the above criteria to approximate it and find the corresponding pmf. For the vector case, unfortunately, we do not have such a nice approximation as in the above. However, we can still assume that x follows a Gaussian distribution with certain mean $\bar{x}$ and covariance matrix C_x. The extension of the above result for scalar case to multivariate Gaussian case is discussed next.

Given an n-dimensional Gaussian distribution $p(x) = \mathcal{N}(x; \bar{x}, C_x)$, due to the symmetry and positive-definiteness of C_x, it follows from the singular value decomposition that

$$C_x = U\Sigma U'$$

where U is a unitary matrix, $\Sigma = \text{diag}\{\sigma_1, \cdots, \sigma_n\}$ and σ_i's, $i = 1, \cdots, n$, are the singular values of C_x. Also, if $y \sim \mathcal{N}(y; 0_{n\times 1}, I_{n\times n})$, then we have

$$U\Sigma^{1/2}y + \bar{x} \sim \mathcal{N}(\bar{x}, C_x)$$

So a simple way to approximate $\mathcal{N}(x; \bar{x}, C_x)$ by a discrete pmf can be summarized by the following three steps.

Step 1: Generate n identical scalar pmfs, each of which uses the same N mass points to approximate a standard normal distribution using the criteria above.

Step 2: Combine the mass points of all n pmfs to have a discrete approximation to $\mathcal{N}(0_{n\times 1}, I_{n\times n})$ by N^n mass points, each of which happens with equal probability $1/N^n$.

Step 3: Transform all the N^n mass points from the last step through a linear transformation $L(y_i) = U\Sigma^{1/2}y_i + \bar{x}$ one by one to have a discrete approximation to $\mathcal{N}(\bar{x}, C_x)$.

In the above extension to multivariate case, the key is to approximate a single standard normal distribution which has the following cdf

$$F_c(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Unfortunately we do not have an analytical form for its inverse function $F_c^{-1}(x)$. To save computational load, we can work out some look-up tables for different numbers of mass points in advance. For example, the look-up tables for $N = 3, 5, 7$ are as follows.

Table I
MASS POINTS OF STANDARD NORMAL DISTRIBUTION FOR $N = 3$

-0.9674	0	0.9674

Table II
MASS POINTS OF STANDARD NORMAL DISTRIBUTION FOR $N = 5$

-1.2816	-0.5244	0	0.5244	1.2816

Table III
MASS POINTS OF STANDARD NORMAL DISTRIBUTION FOR $N = 7$

-1.4652	-0.7916	-0.3661	0	0.3661	0.7916	1.4652

The comparison between the cdf of standard normal distribution and its approximation using different numbers of mass points is shown in Fig. 2. It can be clearly seen that the use of different numbers of mass points does provide reasonable approximations to the cdf of standard normal distribution. Also, the more mass points are used, the better the approximation accuracy is.

VI. Illustrative Examples

To illustrate the effectiveness and efficiency of the proposed multi-model extension of the classical EKF, consider the following nonlinear dynamic system

$$x_k = \frac{1}{2}x_{k-1} + \frac{25x_{k-1}}{1+x_{k-1}^2} + 8\cos\left[1.2\left(k-1\right)\right] + w_{k-1}$$

which is observed as

$$z_k = \frac{1}{20}x_k^2 + v_k$$

as an example. It is known that

$$x_0 \sim \mathcal{N}\left(0.1, 2\right),\ w_k \sim \mathcal{N}\left(0, 1\right),\ v_k \sim \mathcal{N}\left(0, 1\right)$$

This scalar nonlinear system works as a benchmark testing example in many existing work on nonlinear filtering [11], [17].

Next we compare the filtering performance of the classical EKF, the UF (unscented filtering), the PF (Bootstrap particle filter with 100 particles), the MMEKF (Multi-model Extended Kalman filter), and the iEKF (the classical EKF with expansion

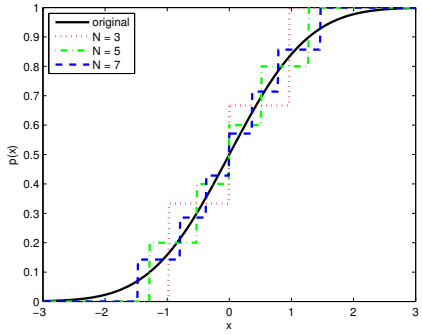

Figure 2. Comparision between cdf of standard normal distribution and its approximations

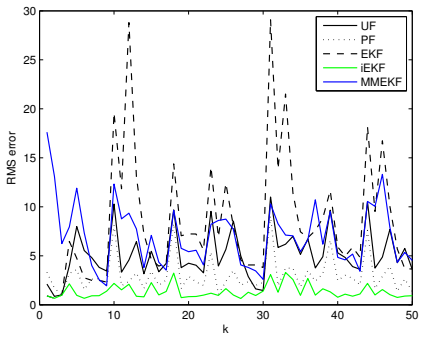

Figure 3. RMS error for $N = 3$

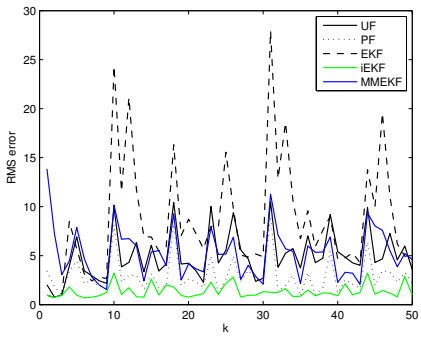

Figure 4. RMS error for $N = 5$

point chosen to be the ground truth). Since the expansion point of the TSE of the iEKF is the ground truth, it is not realizable in reality. However, it can still work as a lower bound for TSE based Extended Kalman filters for performance comparison purpose due to its use of the ideal expansion point.

The Monte Carlo simulation results over 200 runs are shown in Figs. 3, 4 and 5 for the cases $N = 3, 5, 7$, respectively.

It can be clearly seen that the classical EKF performs the worst among all filters. Also, its RMS error curves have the biggest jumps among all filters. So it is really necessary to have it replaced by some better performed nonlinear filters for the system considered. However, the iEKF performs the best among all filters. It even beats the PF and can definitely works as a lower bound for the RMS error. Also its error curves fluctuate the least among all filters, which is strongly desired. The best performance of the iEKF and its least fluctuations further indicates that the TSE based framework of the EKF is still quite promising to lead to better performed nonlinear filters if the expansion points are well selected. And what makes the EKF perform the worst should be mainly coming from the misspecification of the expansion point.

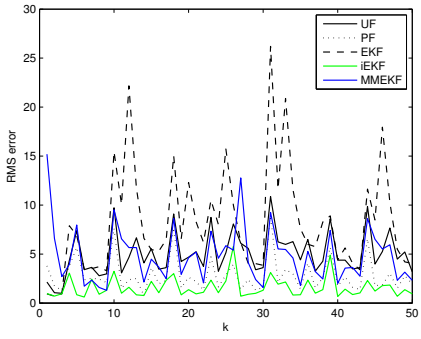

Figure 5. RMS error for $N = 7$

Just from Fig. 3, it can be seen that the performance of the classical EKF can be significantly improved by simply considering two more expansion points around the filter-calculated estimate. So the multi-model extension of the EKF is really attractive. For the above scalar nonlinear system considered, the number of sigma points used by the UF is 3. When the MMEKF also uses 3 mass points, its performance is just a little poorer than that of the UF. However, as can be seen from Fig. 4, when the MMEKF uses 5 mass points, it is already very hard to tell which one between the MMEKF and the UF performs better. If more mass points (7 mass points) are used, the MMEKF can beat the UF very easily as can be seen from Fig. 5. It may be unfair to say so because the MMEKF used more mass points. Nevertheless the way to increase the mass points used by the MMEKF is very systematic and easy. Whereas to increase the number of sigma points used by the UF is very hard and needs ingenious design skill usually.

VII. Conclusions

The malfunction of the EKF is mainly because of the misspecification of the expansion point of the TSE but not the TSE itself. Motivated by this, it is demonstrated in this paper that the TSE based framework of the EKF is still quite promising in leading to better performed nonlinear filters. This is achieved through a multi-model extension to the EKF. In the new extension, multiple probabilistically weighted expansion points are used instead of just the filtered calculated "best" estimate. By taking first-order TSEs at multiple expansion points, multiple linearized dynamic and measurement models are obtained. The original nonlinear filtering problem is then naturally changed into a variable structure multi-model estimation problem. Suggestions on how to design the probabilistically weighted discrete points to approximate the whole state space is provided. Numerical examples show that performance of the proposed multi-model extension to the EKF is quite promising when compared with several existing competitive nonlinear filters.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[2] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. Wiley Interscience Co., 2001.

[3] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new method for nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, March 2000.
[4] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, March 2004.
[5] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.
[6] M. Norgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627–1638, November 2000.
[7] K. Ito and K. Q. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.
[8] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-time nonlinear filtering algorithm using gauss-hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, May 2007.
[9] T. Lefebvre, H. Bruyninckx, and J. D. Schutter, "Comment on "a new method for nonlinear transformation of means and covariances in filters and estimators"," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1406–1408, August 2002.
[10] X. R. Li and V. P. Jilkov, "A survey of maneuvering target tracking: Approximation techniques for nonlinear filtering," in *Proceedings of 2004 SPIE Conference on Signal and Data Processing of Small Targets*, vol. 5428, San Diego, CA, April 2004, pp. 537–550.
[11] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, February 2002.
[12] X. R. Li and V. P. Jilkov, "Survery of maneuvering target tracking. part V: Multiple-model methods," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1255–1321, October 2005.
[13] X. R. Li and Y. Bar-Shalom, "Multiple-model estimation with variable structure," *IEEE Transactions on Automatic Control*, vol. 41, no. 4, pp. 478–493, April 1996.
[14] X. R. Li, Z. L. Zhao, and X. B. Li, "General model-set design methods for multiple-model approach," *IEEE Transactions on Automatic Control*, vol. 50, no. 9, pp. 1260–1276, September 2005.
[15] O. C. Schrempf, D. Brunn, and U. D. Hanebeck, "Dirac mixture density approximation based on minimization of the weighted cramer-von mises distance," in *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Heidelberg, Germany, September 2006, pp. 512–517.
[16] ——, "Density approximation based on dirac mixtures with regard to nonlinear estimation and filtering," in *Proceedings of the 2006 IEEE Conference on Decision and Control*, San Diego, CA, December 2006.
[17] D. Simon, *Optimal State Estimation*. Wiley Interscience Co., 2006.