

# A Linear Extension of Unscented Kalman Filter to Higher-Order Moment-Matching

Jiang Liu<sup>1</sup>, Yujin Wang<sup>1</sup> and Ju Zhang<sup>1</sup>

**Abstract**—This paper addresses the problem of optimal state estimation (OSE) for a wide class of nonlinear time series models. Empirical evidence suggests that the Unscented Kalman Filter (UKF), proposed by Julier and Uhlman, is a promising technique for OSE with satisfactory performance. Unscented Transformation (UT) is the central and vital operation performed in UKF. A crucial point of UT is to construct a  $\sigma$ -set, which consists of points with associated weights capturing the input statistics, e.g., first and second and possibly higher moments. We analyze the standard choice of  $\sigma$ -set and propose a novel method for generating  $\sigma$ -set so as to capture arbitrary higher order input statistics. This method could be considered as a linear extension of UT or UKF, and its computational complexity is the same order as that of the UKF and so EKF. The performance of the algorithm is illustrated by empirical examples. Results show an improvement in accuracy compared to traditional UKF.

## I. INTRODUCTION

Optimal recursive state estimation of nonlinear stochastic dynamic systems with noisy observation data has received considerable research interest for several decades. We will focus on the discrete-time nonlinear stochastic system with the following state space form:

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{v}_k \quad (1a)$$

$$\mathbf{y}_k = \mathbf{g}_k(\mathbf{x}_k) + \mathbf{w}_k \quad (1b)$$

where the vectors  $\mathbf{x}_k \in \mathbb{R}^{n_x}$  and  $\mathbf{y}_k \in \mathbb{R}^{n_y}$  stand for the unobserved state of the system and measurement at time instant  $k$ , respectively,  $\mathbf{f}_k : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_x}$ ,  $\mathbf{g}_k : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_y}$  are known vector functions, and  $\mathbf{w}_k \in \mathbb{R}^{n_x}$ ,  $\mathbf{v}_k \in \mathbb{R}^{n_y}$  are independent state and measurement white noises with zero means and covariance matrices  $\mathbf{P}_k$  and  $\mathbf{Q}_k$  respectively.

Kalman filter (KF) [1] is capable of providing exact solution for (1a) and (1b) when  $\mathbf{f}_k$  and  $\mathbf{g}_k$  are linear functions. However, there is no general closed-form solutions when the transformations  $\mathbf{f}_k$  and  $\mathbf{g}_k$  are nonlinear. Various estimation techniques have been developed for approximating solutions to nonlinear systems. Probably the most widely used estimator for nonlinear systems is the extended Kalman filter (EKF) [2], which applies the Kalman filter to nonlinear systems by simply linearizing the transformations  $\mathbf{f}_k$  and  $\mathbf{g}_k$  through the Jacobian matrices of  $\mathbf{f}_k$  and  $\mathbf{g}_k$  if exist. In

\*This work was partially supported by NSFC 61202131, West Light Foundation of Chinese Academy of Sciences, Chongqing Natural Science Foundation: cstc2012ggB40004, cstc2013jjys40001, cstc2014jcsfgyjs0005 and cstc2014zktjcxxyB0031.

<sup>1</sup>Jiang Liu, Yujin Wang and Ju Zhang are with Chongqing Key Laboratory of Automated Reasoning and Cognition, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences. Emails: liujiang, wangyujin, zhangju@cigit.ac.cn

practice, the EKF has two well-known drawbacks. First, EKF requires the existence of the Jacobian matrices. However, this is not always the case [3], [4]. Second, the EKF linearized approximations can be extremely inaccurate and lead to filter instability [5]. The unscented Kalman filter (UKF) was proposed by Julier and Uhlmann [6], [7] to tackle the drawbacks of EKF, which approximates the states using a careful selection of deterministic samples (so-called  $\sigma$ -points) and associated weights (may be out of interval  $[0, 1]$ ) so as to preserve the moments of the distribution. A method similar to UKF is sequential Monte Carlo filter or particle filter (PF) [8], which approximates the states by random samples (particles) and associated probability weights derived from the posterior distribution. Particle filters need a specification of posterior density while UKF needs only the moments information. Among the Gaussian approximation filters, quadrature rule based Gauss-Hermite quadrature filter (GHQF) [9] and cubature rule based cubature Kalman filter (CKF) [10] have been shown better numerical stability and higher accuracy than UKF. From a computational point of view, the third-degree cubature rule is a special form of unscented transformation [10], namely, they have identical process but different choices of scaling parameters. Empirical works suggest that the UKF and its variations are promising techniques with satisfactory performance.

Recently, we have seen more techniques developed for state estimation of nonlinear stochastic systems, which can approximate the statistics of the process as accurately as possible and achieve acceptable computational performance at the same time. For instance, particle filters employing Monte Carlo methods has demonstrated good accuracy with large number of samples, but its computational complexity increases exponentially with the dimension of the system state. By decomposing the multivariate problem into univariate ones, paper [11] significantly improved computational cost of PF method. The CKF was generalized by [12] with arbitrary degree of accuracy to improve the performance of the third-degree CKF and the UKF so as to achieve performance close to that of the GHQF. There are several variations [13], [14], [15], [16], [17], [18], [19], [20] of UKF that obtain better accuracy than UKF by matching high-order moments. These variations distinguish from each other mainly at the different selections of sigma set about the UT, and their computational costs corresponds to the number of sigma points.

Unlike the existing sigma set generation methods of UKF, this paper proposes a brutal-force selection of sigma set by separating the choices of sigma points and the associated

weights into two steps. It first chooses sigma points. Then it computes the weights through matching the high-order moments. So, it achieves high accuracy once it matches adequate moments.

The rest of this article is organized as follows. Section II reviews the original UKF and related work. In Section III, we analyze the selection of sigma points and present the new algorithm for sigma points selection. We illustrate the new algorithm and give experimental results in Section IV. Finally, Section V summarizes the current work and discusses the possible future work.

## II. UNSCENTED KALMAN FILTER

The UKF algorithm is a local filter structure [18], wherein the states and measurements defined in equations (1a) and (1b) are recursively updated using the UT, and the filtering is optimized by the Kalman gain.

### A. Unscented Transformation

Consider the transformation of some random variable  $\mathbf{x}$  through a known nonlinear function

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \quad (2)$$

with known mean  $\bar{\mathbf{x}}$  and covariance matrix  $\mathbf{P}_{\mathbf{x}}$ , the objective is to calculate the mean  $\bar{\mathbf{y}}$  and covariance  $\mathbf{P}_{\mathbf{y}}$  of  $\mathbf{y}$ .

Based upon the principle that it is easier to approximate a probability distribution than an arbitrary function, the UT approximates  $\bar{\mathbf{y}}$  and  $\mathbf{P}_{\mathbf{y}}$  by two steps in what follows.

*Step I:* We approximate the distribution of random variable  $\mathbf{x}$  by deterministically selecting samples, so-called sigma points,  $\{\mathcal{X}_i\}$  with associated weights  $\{W_i\}$  such that

$$\sum_i W_i = 1 \quad (3a)$$

$$\bar{\mathbf{x}} = \sum_i W_i \mathcal{X}_i \quad (3b)$$

$$\mathbf{P}_{\mathbf{x}} = \sum_i W_i (\mathcal{X}_i - \bar{\mathbf{x}})(\mathcal{X}_i - \bar{\mathbf{x}})^T \quad (3c)$$

where (3a) is used for an unbiased estimate. The set  $\{\mathcal{X}_i, W_i\}$  is called  $\sigma$ -set. The standard UT [7] employed the following symmetrically distributed sigma points and weights

$$\mathcal{X}_0 = \bar{\mathbf{x}}, \quad W_0 = \frac{\kappa}{n_{\mathbf{x}} + \kappa} \quad (4a)$$

$$\mathcal{X}_i = \bar{\mathbf{x}} + (\sqrt{(n_{\mathbf{x}} + \kappa)\mathbf{P}_{\mathbf{x}}})_i, \quad W_i = \frac{1}{2(n_{\mathbf{x}} + \kappa)} \quad (4b)$$

$$\mathcal{X}_j = \bar{\mathbf{x}} - (\sqrt{(n_{\mathbf{x}} + \kappa)\mathbf{P}_{\mathbf{x}}})_i, \quad W_j = \frac{1}{2(n_{\mathbf{x}} + \kappa)} \quad (4c)$$

where  $i = 1, \dots, n_{\mathbf{x}}, j = n_{\mathbf{x}} + i$ , the term  $(\sqrt{(n_{\mathbf{x}} + \kappa)\mathbf{P}_{\mathbf{x}}})_i$  denote the  $i$ -th column of the square root of matrix  $(n_{\mathbf{x}} + \kappa)\mathbf{P}_{\mathbf{x}}$  and  $\kappa$  is the scaling parameter that scales the third and higher order terms. If  $n_{\mathbf{x}} + \kappa = 3$  it is possible to match marginal fourth moments when  $\mathbf{x}$  is Gaussian.

*Step II:* Compute the transformed sigma points

$$\mathcal{Y}_i = \mathbf{h}(\mathcal{X}_i).$$

Then use  $\{\mathcal{Y}_i\}$  and  $\{W_i\}$  to approximate the mean and covariance of  $\mathbf{y}$  as follows

$$\hat{\mathbf{y}} = \sum_i W_i \mathcal{Y}_i \quad (5a)$$

$$\hat{\mathbf{P}}_{\mathbf{y}} = \sum_i W_i (\mathcal{Y}_i - \hat{\mathbf{y}})(\mathcal{Y}_i - \hat{\mathbf{y}})^T \quad (5b)$$

Note that  $\hat{\mathbf{y}}$  and  $\hat{\mathbf{P}}_{\mathbf{y}}$  are only approximations of the true mean  $\bar{\mathbf{y}}$  and  $\mathbf{P}_{\mathbf{y}}$  respectively. The reduced UT [21] can improve the complexity by less sigma points. The scaled [13], [18] and high order UT [14], [17], [19] can improve the accuracy by using additional free parameters.

### B. Unscented Kalman Filter

Incorporating with minimum mean square error method, the UT (4a-5b) can be used to update the state estimation of the system defined as (1a) and (1b). The resulting UKF [7], [18] can be summarized as follows.

*Step 1:* Start with the instant  $k = 0$ , define a priori initial condition  $\hat{\mathbf{x}}_{0|0} = E(\mathbf{x}_0) = \bar{\mathbf{x}}_0$  and  $\mathbf{P}_{0|0} = cov[\mathbf{x}_0] = \mathbf{P}_0$ .

*Step 2:* With respect to the last measurement  $\mathbf{y}_k$ , the optimal estimation of  $\mathbf{x}_k$  is

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}) \quad (6a)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{\mathbf{y},k|k-1} \mathbf{K}_k^T \quad (6b)$$

where  $\mathbf{K}_k = \mathbf{P}_{\mathbf{x}\mathbf{y},k|k-1} (\mathbf{P}_{\mathbf{y},k|k-1})^{-1}$  is the filter gain and

$$\hat{\mathbf{y}}_{k|k-1} = E[\mathbf{g}_k(\mathbf{x}_k) | \mathbf{y}^{k-1}] \quad (7a)$$

$$\mathbf{P}_{\mathbf{y},k|k-1} = E[(\mathbf{g}_k(\mathbf{x}_k) - \hat{\mathbf{y}}_{k|k-1}) \times (\mathbf{g}_k(\mathbf{x}_k) - \hat{\mathbf{y}}_{k|k-1})^T | \mathbf{y}^{k-1}] + \mathbf{Q}_k \quad (7b)$$

$$\mathbf{P}_{\mathbf{x}\mathbf{y},k|k-1} = E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{g}_k(\mathbf{x}_k) - \hat{\mathbf{y}}_{k|k-1})^T | \mathbf{y}^{k-1}] \quad (7c)$$

wherein (7a)-(7c) are computed using (5a-5b) and

$$\hat{\mathbf{P}}_{\mathbf{x}\mathbf{y}} = \sum_i W_i (\mathcal{X}_i - \hat{\mathbf{x}})(\mathcal{Y}_i - \hat{\mathbf{y}})^T \quad (8)$$

by considering  $\mathbf{P}_{\mathbf{x}}$  as  $\mathbf{P}_{k|k-1}$ ,  $\mathbf{h}(\cdot)$  as  $\mathbf{g}(\cdot)$ ,  $\hat{\mathbf{y}}$  as  $\hat{\mathbf{y}}_{k|k-1}$ ,  $\mathbf{P}_{\mathbf{y}}$  as  $\mathbf{P}_{\mathbf{y},k|k-1}$ ,  $\hat{\mathbf{x}}$  as  $\hat{\mathbf{x}}_{k|k-1}$  and  $\mathbf{P}_{\mathbf{x}\mathbf{y}}$  as  $\mathbf{P}_{\mathbf{x}\mathbf{y},k|k-1}$ .

*Step 3:* The predictive statistic are computed according to

$$\hat{\mathbf{x}}_{k+1|k} = E[\mathbf{f}_k(\mathbf{x}_k) | \mathbf{y}^k] \quad (9a)$$

$$\mathbf{P}_{k+1|k} = E[(\mathbf{f}_k(\mathbf{x}_k) - \hat{\mathbf{x}}_{k+1|k}) \times (\mathbf{f}_k(\mathbf{x}_k) - \hat{\mathbf{x}}_{k+1|k})^T | \mathbf{y}^k] + \mathbf{P}_k \quad (9b)$$

which are computed according to (5a-5b) by considering  $\hat{\mathbf{x}}$  as  $\hat{\mathbf{x}}_{k|k}$ ,  $\mathbf{P}_{\mathbf{x}}$  as  $\mathbf{P}_{k|k}$ ,  $\mathbf{h}(\cdot)$  as  $\mathbf{f}(\cdot)$ ,  $\hat{\mathbf{y}}$  as  $\hat{\mathbf{x}}_{k+1|k}$ , and  $\mathbf{P}_{\mathbf{y}}$  as  $\mathbf{P}_{k+1|k}$ .

Let  $k = k + 1$ . The algorithm loops at steps 2 and 3.

So, the most costly computations in UKF are the approximations of mean and covariance by the UT formulated by (3a-5b). The accuracy and complexity of UKF are substantially determined by the UT. As the error can propagate with the recursive estimation of state, it is natural to enhance the accuracy of UT. The next section will describe a new method for UT and a corresponding extension of UKF.

### III. HIGH-ORDER MOMENTS UT AND UKF

The mean and covariance matrix play a vital role for the selection of sigma points at UKF and its all variations. In fact, the mean  $\bar{\mathbf{x}}$  and the square roots  $\{(\sqrt{\mathbf{P}_{\mathbf{x}}})_i\}$  of covariance matrix of a random variable  $\mathbf{x}$  specify the geometric characteristic of the probability distribution of  $\mathbf{x}$ . That is, the center of distribution is at  $\bar{\mathbf{x}}$ , and the deviation of distribution is described by points  $\{(\sqrt{\mathbf{P}_{\mathbf{x}}})_i\}$ . The sigma points and weights in (4a-4c) form a solution to (3a-3c). This solution could be obtained from three different approaches as follows.

*Approach I:* The weights  $\{W_i\}$  are fixed as in (4a-4c). We take sigma points of form  $\{\mathcal{X}_i = \bar{\mathbf{x}} + (\sqrt{\alpha_i \mathbf{P}_{\mathbf{x}}})_i\}$ . Solving the equation system consisting of (3a) and

$$\bar{\mathbf{x}} = \sum_i W_i (\bar{\mathbf{x}} + (\sqrt{\alpha_i \mathbf{P}_{\mathbf{x}}})_i) \quad (10a)$$

$$\mathbf{P}_{\mathbf{x}} = \sum_i W_i (\sqrt{\alpha_i \mathbf{P}_{\mathbf{x}}})_i (\sqrt{\alpha_i \mathbf{P}_{\mathbf{x}}})_i^T \quad (10b)$$

obtains that the sigma points  $\{\mathcal{X}_i\}$  with  $\alpha_i = \sqrt{n_{\mathbf{x}} + \kappa}$  for all  $i$  associated the corresponding weights is a solution.

*Approach II:* The sigma points are fixed as  $\{\mathcal{X}_i = \bar{\mathbf{x}} + (\sqrt{\alpha_i \mathbf{P}_{\mathbf{x}}})_i\}$  with known  $\{\alpha_i\}$ . With these known  $\{\mathcal{X}_i\}$  and under the assumption  $\alpha_i = (n_{\mathbf{x}} + \kappa)$  for all  $i$  for some  $\kappa$ , solving the equation system consisting of (3a-3c) and

$$W_i = W_j, \quad i, j > 0 \quad (11)$$

obtains the weights  $\{W_i\}$  as same as ones in (4a-4c) with associated the fixed sigma points is a solution.

*Approach III:* The sigma points are parameterized by  $\{\mathcal{X}_i = \bar{\mathbf{x}} + (\sqrt{\alpha_i \mathbf{P}_{\mathbf{x}}})_i\}$ . With the special form  $\{\mathcal{X}_i\}$  and unknown weights  $\{W_i\}$ , under the assumption that all  $\alpha_i$  are same, solving the equation system consisting of (3a), (10a), (10b) and (11) also obtains the solution (4a-4c) with  $\alpha_i = \sqrt{n_{\mathbf{x}} + \kappa}$  for any  $\kappa > -n_{\mathbf{x}}$ .

In the Approaches I and II,  $\alpha_i$  could be regarded as a scaling parameter with respect to the basis  $(\sqrt{\mathbf{P}_{\mathbf{x}}})_i$ , which describe the deviation degree of sigma points relative to mean based upon the covariance. In some sense, the role of  $\alpha_i$  is identical to one of the radius parameter  $r$  in the CKF [10]. The formula (11) suggests an equal weight for all sigma points deviated from mean. This requirement is not necessary in general. For example, almost all  $W_i$  have different values in reduced sigma point filter [21]. Note that the Approach II only needs to solve a linear system about variables  $\{W_i\}$ , and thence it has relatively low computational cost.

The standard UKF [6] can be obtained by any one of these approaches, which approximated the distribution of  $\mathbf{x}$  by matching only first two moments. The scaled UKF [13], [18] could be derived by Approach II considering a uniform deviation  $\mathcal{X}'_i = \mathcal{X}_0 + \beta(\mathcal{X}_i - \mathcal{X}_0)$  to a fixed sigma point  $\mathcal{X}_0$  with  $\{\mathcal{X}_i = \bar{\mathbf{x}} + (\sqrt{\alpha_i \mathbf{P}_{\mathbf{x}}})_i\}$  such that all  $\alpha_i$  are same, which took approximation by matching some 4th order moments. The method in [20] is similar to the scaled UKF, but the approximation therein exploited the benefit of 4th marginal moments through employing sigma points

and weights obtained by closed form rather than numerical optimization, while [19] has a similar story by utilizing the moments containing the first 4th marginal and mixed third order moments. The high order UKF [14] could be derived by Approach III with more sigma points and different weights for different sigma points, wherein it matched marginal moments of  $\mathbf{x}$  up to 8th order for the sake of accuracy.

From the experiments in the foregoing literature and the references therein, we may have the following principle: *The more moments are matched, the more accuracy is achieved by the approximation.* Based upon the principle, we now propose a new sigma points selection scheme, so-called *linear high-order moments UT (LUT)*, so as to match arbitrary order known moments.

#### A. Linear High-Order Moments UT

Given statistics of a random variable  $\mathbf{x}$ : mean  $\bar{\mathbf{x}}$ , covariance matrix  $\mathbf{P}_{\mathbf{x}}$  and high order central moments  $m_1^{(3)}(\mathbf{x})$ ,  $m_2^{(3)}(\mathbf{x})$ ,  $\dots$ ,  $m_{n_{\mathbf{x}}}^{(3)}(\mathbf{x})$ ,  $\dots$ ,  $m_1^{(\ell)}(\mathbf{x})$ ,  $\dots$ ,  $m_{n_{\mathbf{x}}}^{(\ell)}(\mathbf{x})$  for some  $\ell > 2$ , where  $m_k^{(l)}(\mathbf{x}) = E[(\bar{x}_k - x_k)^l]$  denoted by  $m_k^{(l)}$  for simplicity, for  $2 < l \leq \ell$  and  $1 \leq k \leq n_{\mathbf{x}}$ , without loss of generality, assume these moments have different absolute values, the sigma points and weights are computed as follows.

1. Choose sigma points:  $\mathcal{X}_0 = \bar{\mathbf{x}}$ ,
  - $\mathcal{X}_i = \bar{\mathbf{x}} + (\sqrt{\alpha \mathbf{P}_{\mathbf{x}}})_i$  for  $1 \leq i \leq n_{\mathbf{x}}$ ,
  - $\mathcal{X}_i = \bar{\mathbf{x}} - (\sqrt{\alpha \mathbf{P}_{\mathbf{x}}})_{i-n_{\mathbf{x}}}$  for  $n_{\mathbf{x}} < i \leq 2n_{\mathbf{x}}$ ,
  - $\mathcal{X}_i^{lj} = \bar{\mathbf{x}} + (\sqrt{|m_j^{(l)}| \mathbf{P}_{\mathbf{x}}})_i$ ,  $1 \leq i \leq n_{\mathbf{x}}$ ,  $2 < l \leq \ell$ , and  $1 \leq j \leq n_{\mathbf{x}}$ ,
  - $\mathcal{X}_i^{lj} = \bar{\mathbf{x}} - (\sqrt{|m_j^{(l)}| \mathbf{P}_{\mathbf{x}}})_{i-n_{\mathbf{x}}}$ ,  $n_{\mathbf{x}} < i \leq 2n_{\mathbf{x}}$ ,  $2 < l \leq \ell$ , and  $1 \leq j \leq n_{\mathbf{x}}$ ,

where  $\alpha > 0$  is an auxiliary parameter in order to get a positive solution to the weights.

2. Assign weights to sigma points according to *the principle that the sigma points with the same scalar with respect to  $\sqrt{\mathbf{P}_{\mathbf{x}}}$  have the same weight*, as follows:  $W_0$  to  $\mathcal{X}_0$ ,  $W_1$  to  $\mathcal{X}_i$  for all  $1 \leq i \leq 2n_{\mathbf{x}}$ ,  $W_{lj}$  to  $\mathcal{X}_i^{lj}$  for  $1 \leq i \leq 2n_{\mathbf{x}}$ ,  $2 < l \leq \ell$ , and  $1 \leq j \leq n_{\mathbf{x}}$ .
3. Match the moments by the following constraints

$$\bar{\mathbf{x}} = W_0 \mathcal{X}_0 + \sum_i W_1 \mathcal{X}_i + \sum_l \sum_j \sum_i W_{lj} \mathcal{X}_i^{lj} \quad (12a)$$

$$\mathbf{P}_{\mathbf{x}} = \sum_i W_1 (\mathcal{X}_i - \bar{\mathbf{x}})(\mathcal{X}_i - \bar{\mathbf{x}})^T + \sum_l \sum_j \sum_i W_{lj} (\mathcal{X}_i^{lj} - \bar{\mathbf{x}})(\mathcal{X}_i^{lj} - \bar{\mathbf{x}})^T \quad (12b)$$

$$m_{\beta}^{(\mu)} = \sum_i W_1 ((\mathcal{X}_i)_{\beta} - (\bar{\mathbf{x}})_{\beta})^{\mu} + \sum_l \sum_j \sum_i W_{lj} ((\mathcal{X}_i^{lj})_{\beta} - (\bar{\mathbf{x}})_{\beta})^{\mu} \quad (12c)$$

where  $(\cdot)_{\beta}$  denotes the  $\beta$ th element of a vector,  $1 \leq \beta, j \leq n_{\mathbf{x}}$ ,  $2 < \mu, l \leq \ell$  and  $1 \leq i \leq 2n_{\mathbf{x}}$ . Similarly, we may put the restraint (3a) for an unbiased estimate.

Alternatively, we may consider matching the average  $l$ th moment like [17] in place of (12c).

- Solve the linear system from step 3 about  $W_0, W_1, \{W_{l_j}\}$ . Its solution together with the sigma points in the step 1 provides us an approximation of  $\mathbf{x}$ .

Note that the standard UT chooses sigma points derived from mean and covariance to match the first two moments. Thus, it is natural to choose sigma points derived from high-order moments to match high-order moments, herein each high-order moment  $m_j^{(l)}$  chooses  $2n_x$  many sigma points  $\mathcal{X}_i^{lj}$ . For the sake of low computational cost, here we prefer to employ the Approach II. By the choice of  $\sigma$ -set, (12a) and (12b) are reduced, respectively, to

$$W_0 + 2n_x W_1 + \sum_l \sum_j 2n_x W_{l_j} = 1 \quad (13a)$$

$$\alpha W_1 + \sum_l \sum_j |m_j^{(l)}| W_{l_j} = 1 \quad (13b)$$

If we consider only matching even order moments, namely,  $u$  of (12c) is even, then all coefficients of weights are nonnegative. So it always can obtain positive weights as  $\alpha$  is properly selected.

Anyway, the Approaches I and III can also be applied to matching high-order moments. The usage of Approach I needs a proper selection of weights, which is discussed in [22]. These Approaches require additional scaling parameters  $\beta_{lj}$  associated with sigma points  $\mathcal{X}_i^{lj}$ . The resulting system as seen in step 3, wherein  $\beta_{lj}$  takes place of  $m_j^{(l)}$  in the corresponding (13b), is not linear anymore.

### B. Linear High-Order Moments UKF

For the sake of applying LUT to filtering, we need also to estimate the high-order moments in addition to mean and covariance matrix at each time instant, in comparison to UKF. Certainly, the higher order moments of the randoms  $\mathbf{v}_k$  and  $\mathbf{w}_k$  are also required. The *linear high order moment UKF* (LUKF) works by employing LUT to update its moments information as follows.

- In LUKF, the means and covariances of  $\mathbf{x}_{k+1}$  are updated similar to (6a-9b) in UKF except that the  $\sigma$ -set is selected by LUT described as previous section.
- LUKF utilizes higher order moments to compute its  $\sigma$ -set at each time instant  $k$ . As mean is approximated by UKF, it needs a model for describing a more general solution to optimal recursive state estimation which can provide a recursive relationship between two successive states and also be consistent with UKF. Here, we adopt

$$\tilde{\mathbf{x}}_{k+1} = \mathbf{f}_k(\tilde{\mathbf{x}}_k) + \mathbf{v}_k + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - (\mathbf{g}_{k+1}(\mathbf{f}_k(\tilde{\mathbf{x}}_k) + \mathbf{v}_k) + \mathbf{w}_{k+1})) \quad (14)$$

as the optimal estimation model, where  $\mathbf{K}_{k+1}$  is the filter gain obtained by KF method. It is easy to see that the mean and covariance computed by (14) are as same as ones computed by UKF. Thus, this model is consistent with UKF.

- The higher order moments  $\{m_\beta^{(\mu)}(\mathbf{x}_{k+1})\}$  are approximated by using the estimation  $\tilde{\mathbf{x}}_{k+1|k+1}$  and model (14) by the below formula.

$$m_\beta^{(\mu)}(\mathbf{x}_{k+1}) = E[(\tilde{\mathbf{x}}_{k+1} - \hat{\mathbf{x}}_{k+1})_\beta^\mu] \quad (15)$$

In (15),  $\mathbf{v}_k$  and  $\mathbf{w}_{k+1}$  are prior knowledge about noises, moments of  $\tilde{\mathbf{x}}_k$  are approximated recursively up to time instant  $k$ . The random variable  $\tilde{\mathbf{y}}_k := \mathbf{f}_k(\tilde{\mathbf{x}}_k) + \mathbf{v}_k$  is a function of independent  $\mathbf{x}_k$  and  $\mathbf{v}_k$ . We estimate the moments of  $\tilde{\mathbf{y}}_k$  using LUT and augmented UKF proposed in [23] as following. To capture the distribution of random vector  $(\mathbf{x}_k^T, \mathbf{v}_k^T)$ , choose the sigma points

$$\mathcal{X}_0 = \begin{pmatrix} \bar{\mathbf{x}}_k \\ \mathbf{0} \end{pmatrix} \text{ and}$$

$$\mathcal{X}_i^{\mathbf{x}} = \begin{pmatrix} \mathcal{X}_i \\ \mathbf{0} \end{pmatrix}, 1 \leq i \leq 2n_x \quad (16a)$$

$$\mathcal{X}_i^{\mathbf{v}} = \begin{pmatrix} \bar{\mathbf{x}}_k \\ (\sqrt{\mathbf{P}_k})_i \end{pmatrix}, 1 \leq i \leq n_x \quad (16b)$$

$$\mathcal{X}_i^{\mathbf{v}} = \begin{pmatrix} \bar{\mathbf{x}}_k \\ -(\sqrt{\mathbf{P}_k})_{i-n_x} \end{pmatrix}, n_x < i \leq 2n_x \quad (16c)$$

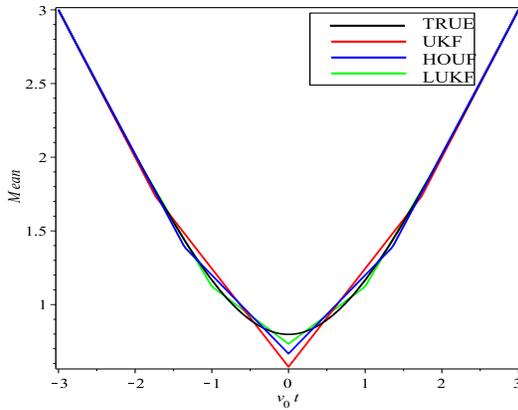
$$\mathcal{X}_i^{lj, \mathbf{x}} = \begin{pmatrix} \mathcal{X}_i^{lj} \\ \mathbf{0} \end{pmatrix}, 1 \leq j \leq n_x, 1 \leq i \leq 2n_x \quad (16d)$$

$$\mathcal{X}_i^{lj, \mathbf{v}} = \begin{pmatrix} \bar{\mathbf{x}} \\ (\sqrt{|m_j^{(l)}| \mathbf{P}_k})_i \end{pmatrix}, 1 \leq i \leq n_x \quad (16e)$$

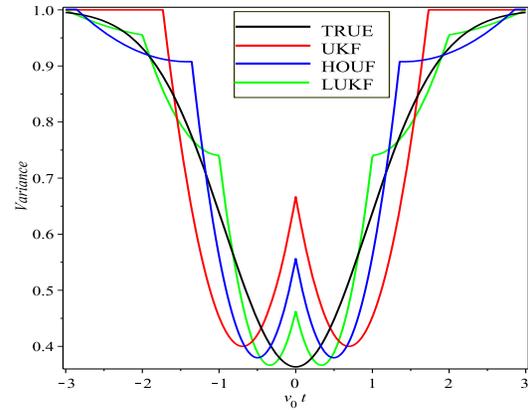
$$\mathcal{X}_i^{lj, \mathbf{v}} = \begin{pmatrix} \bar{\mathbf{x}} \\ -(\sqrt{|m_j^{(l)}| \mathbf{P}_k})_{i-n_x} \end{pmatrix}, n_x < i \leq 2n_x \quad (16f)$$

where  $\mathbf{0}$  is the zero vector of dimension  $n_v$ ,  $\mathcal{X}_i$  and  $\mathcal{X}_i^{lj}$  are sigma points selected the same as in LUT by considering  $\mathbf{P}_x$  as  $\mathbf{P}_{\mathbf{x}_k}$  and  $m_j^{(l)}$  as  $m_j^{(l)}(\mathbf{x}_k)$  or  $m_j^{(l)}(\mathbf{v}_k)$ ,  $1 \leq \beta \leq n_x$  and  $l, \mu > 2$ . Assign weights  $W_0$  to  $\mathcal{X}_0^{\mathbf{v}}$ ,  $W_1$  to  $\mathcal{X}_i^{\mathbf{x}}$  and  $\mathcal{X}_i^{\mathbf{v}}$ ,  $W_{lj, \mathbf{x}}$  to  $\mathcal{X}_i^{lj, \mathbf{x}}$  and  $\mathcal{X}_i^{lj, \mathbf{v}}$ . Now, we apply LUT to such sigma points and then get a constraint on weights the same as (13a-13b). Solving the resulted linear system, the corresponding  $\sigma$ -set can capture the high-order moments of  $(\mathbf{x}_k^T, \mathbf{v}_k^T)$ . Then we make use of the transformed sigma points through  $\mathbf{f}(\tilde{\mathbf{x}}_k) + \mathbf{v}_k$  and their wights to estimate the moments of  $\tilde{\mathbf{y}}_k$ . With a similar process, we can estimate  $(\tilde{\mathbf{y}}_k - \mathbf{K}_{k+1} \mathbf{g}_{k+1}(\tilde{\mathbf{y}}_k)) + \mathbf{K}_{k+1} \mathbf{w}_{k+1}$  by using sigma points of random vector  $(\tilde{\mathbf{y}}_k^T, \mathbf{w}_{k+1}^T)$  selected by LUT similar to (16a-16f), since  $\mathbf{w}_{k+1}$  is independent of  $\tilde{\mathbf{y}}_k$ .

Most computational cost of LUKF is at the computation of the square root of a matrix, which together with the square roots of moments fixes sigma points. The weights are computed through a linear system that is derived from matching high-order moments. In the LUKF, solving linear system is the additional work in comparison with UKF. Therefore, LUKF is a linear extension of UKF with higher accuracy.



(a) True means and estimates: UKF, HOUF vs. LUKF



(b) True variances and estimates: UKF, HOUF vs. LUKF

Fig. 1: True Values and Estimates of BPM

#### IV. SIMULATION STUDY

In this section, we apply LUKF to study the barrier particle motion model (BPM) [14] and the univariate nonstationary growth model (UNGM) [24] which is very popular in econometrics.

##### A. Experimental Results About BPM

The BPM describes that a particle at the initial position  $\mathbf{x}_0$  with a velocity of  $v_0$  can return when it hits the barrier. The distance from the barrier is modeled by

$$\mathbf{y} = |v_0 t + \mathbf{x}_0| \quad (17)$$

Assuming the initial position  $\mathbf{x}_0$  is a standard normal distribution, then the analytical solutions to the mean  $\tilde{\mathbf{y}}$  and covariance  $\mathbf{P}_y$  of the random variable  $\mathbf{y}$  are

$$\tilde{\mathbf{y}} = v_0 t \operatorname{erf}\left(\frac{v_0 t}{\sqrt{2}}\right) + \sqrt{\frac{2}{\pi}} e^{-\frac{v_0^2 t^2}{2}} \quad (18a)$$

$$\mathbf{P}_y = 1 + v_0^2 t^2 - \tilde{\mathbf{y}}^2 \quad (18b)$$

The performance of UKF, HOUF [14] and LUKF was compared using the absolute error (AE) defined by

$$AE = |\theta - \hat{\theta}| \quad (19)$$

where  $\theta$  is the true value and  $\hat{\theta}$  is an estimate. For a time varying mean  $v_0 t$  within the interval  $[-3, 3]$ , we carry out the UKF, the HOUF matching moments up to 8th-order and the LUKF matching moments up to 6th-order with respect to  $\mathbf{x}_0$ . Fig.1a and Fig.1b show the estimate results about means and variances. It is evident that LUKF has better performance on accuracy than HOUF, which can be verified by their AEs. It is easy to understand that LUKF has better performance than UKF since it uses more sigma points and matches more high-order moments. The reason about better performance than HOUF is not completely clear. It is partially because that LUKF has freedom to select important sigma points.

##### B. Experimental Results About UNGM

The discrete-time dynamic system equation of UNGM can be written as

$$\mathbf{x}_n = \frac{\mathbf{x}_{n-1}}{2} + \frac{25\mathbf{x}_{n-1}}{1 + \mathbf{x}_{n-1}^2} + 8 \cos \frac{6(n-1)}{5} + \mathbf{u}_n \quad (20a)$$

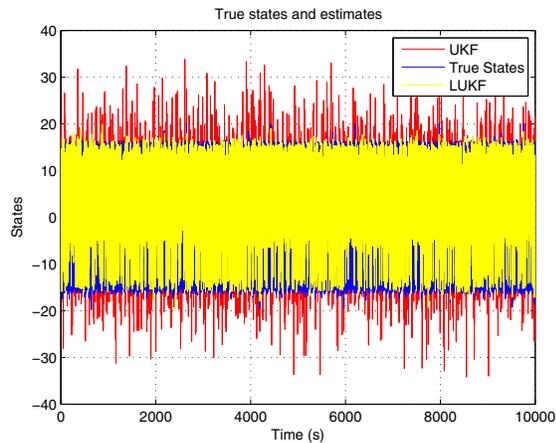
$$\mathbf{y}_n = \frac{\mathbf{x}_n^2}{20} + \mathbf{v}_n, \quad n = 1, \dots, N \quad (20b)$$

where the independent noises  $\mathbf{u}_n$  and  $\mathbf{v}_n$  are both Gaussian with zero mean and unit covariance. The reference data were generated using  $\mathbf{x}_0 = 0.1$  and  $N = 10000$ . The performances of UKF and LUKF were also compared using the AE (19). The LUKFs in Fig.2a and Fig.2b match the moments of  $\mathbf{u}_n, \mathbf{v}_n$  up to 6th-order and 8th-order respectively. Wherein, we pick the additional parameter  $\alpha = 0.9$ . The simulations showed that LUKF possesses a visible stability of numerical estimation, where the error vibrates within a narrow band. Their AEs showed that the LUKF performs noticeably better on the accuracy over a large number of data. Comparing their errors, we can see that the more higher order moments are matched, the more accuracy is achieved. As to the computational complexity of LUKF, the experiment in Fig.2b spends only about 3.1 second for 8th-order LUKF and about 2.6 second for UKF, on my PC<sup>1</sup> using MATLAB 7.10.0 (R2010a). So the practical complexity is applicable.

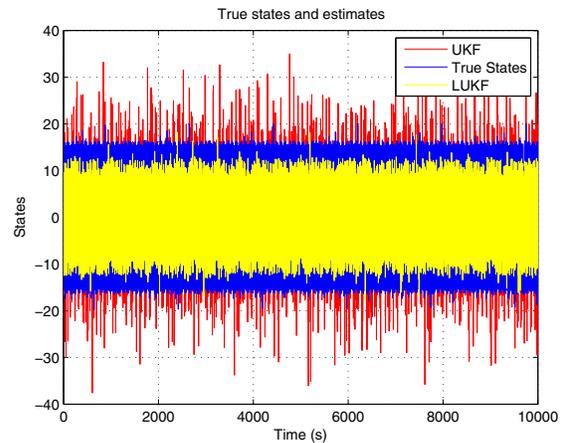
#### V. CONCLUSION

This article presents a new algorithm for optimal recursive state estimation of nonlinear stochastic dynamic systems with noisy observation data. Comparing with the standard UKF, the new algorithm can match more high-order moments, and hence it could provide more accuracy. In comparison with CKF, the new algorithm makes use of general ellipsoids instead of spheres in CKF. And, unlike CKF that needs density distribution to compute the weights, we only need moments information here. The traditional high order UKFs

<sup>1</sup>It is with Intel(R) Core(TM) i3-2100 CPU @3.10GHz, 1.82GB RAM and 32bitOS.



(a) True states and estimates: UKF vs. 6th-order-LUKF



(b) True states and estimates: UKF vs. 8th-order-LUKF

Fig. 2: True Values and Estimates of UNGM

mainly code the high-order moments information into at most two scalar coefficients of sigma points generated by covariance matrix, but the LUKF provides more freedom to such sigma points by considering more sigma points. Several traditional higher-order UT methods need the closed-form solution of a nonlinear system for performance improvement. The LUT can achieve improvement by solving a linear system with lower computational cost. The experimental results illustrates this advantage empirically. Currently, we are investigating how to achieve these goals by less sigma points, and finding a proper principle for determining the scalar parameter  $\alpha$  in (13b).

#### REFERENCES

- [1] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. of Basic Engineering*, no. 82 (Series D), pp. 35–45, 1960.
- [2] H. Sorenson, *Kalman Filtering: Theory and Application*. Piscataway NJ, USA: IEEE Press, 1985.
- [3] J. Kuchar and L. Yang, "A review of conflict detection and resolution modeling methods," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, pp. 179–189, 2000.
- [4] S. Julier and J. Uhlmann, "A counter example to the theory of simultaneous localization and map building," in *Proc. IEEE Int. Conf. on Rob. and Autom.*, vol. 4, 2001, pp. 4238–4243.
- [5] D. Lerro and Y. Bar-Shalom, "Tracking with debiased consistent converted measurements versus ekf," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 3, pp. 1015–1022, Jul 1993.
- [6] S. Julier and J. Uhlmann, "New extension of the kalman filter to nonlinear systems," in *Proc. SPIE*, vol. 3068, 1997, pp. 182–193.
- [7] —, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar 2004.
- [8] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Trans. Sig. Proc.*, vol. 50, no. 2, pp. 174–188, Feb 2002.
- [9] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Autom. Control*, vol. 45, no. 5, pp. 910–927, May 2000.
- [10] I. Arasaratnam and S. Haykin, "Cubature kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.
- [11] K. Høyland, M. Kaut, and S. Wallace, "A heuristic for moment-matching scenario generation," *Comput. Optim. Appl.*, vol. 24, no. 2-3, pp. 169–185, Feb. 2003.
- [12] B. Jia, M. Xin, and Y. Cheng, "The high-degree cubature kalman filter," in *51st IEEE Conf. Decision and Control*, Dec 2012, pp. 4095–4100.
- [13] S. Julier, "The scaled unscented transformation," in *Proc. Am. Control Conf.*, vol. 6, 2002, pp. 4555–4559.
- [14] D. Tenne and T. Singh, "The higher order unscented filter," in *Proc. Am. Control Conf.*, vol. 3, June 2003, pp. 2441–2446.
- [15] P. Date, L. Jalen, and R. Mamon, "A new algorithm for latent state estimation in non-linear time series models," *J. Appl. Math. Comp.*, vol. 203, no. 1, pp. 224 – 232, 2008.
- [16] M. Huber and U. Hanebeck, "Gaussian filter based on deterministic sampling for high quality nonlinear estimation," in *Proc. 17th IFAC World Congress*, 2008, pp. 13 527–13 532.
- [17] K. Ponomareva, P. Date, and Z. Wang, "The higher order unscented filter," in *Proc. MTNS, Budapest*, 2010, pp. 1609–1613.
- [18] J. Dunik, M. Simandl, and O. Straka, "Unscented kalman filter: Aspects and adaptive setting of scaling parameter," *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2411–2416, Sept 2012.
- [19] O. Grothe, "A higher order correlation unscented kalman filter," *J. Appl. Math. Comp.*, vol. 219, no. 17, pp. 9033 – 9042, 2013.
- [20] K. Ponomareva and P. Date, "Higher order sigma point filter: A new heuristic for nonlinear time series filtering," *J. Appl. Math. Comp.*, vol. 221, pp. 662 – 671, 2013.
- [21] S. Julier and J. Uhlmann, "Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations," in *Proc. Am. Control Conf.*, vol. 2, 2002, pp. 887–892 vol.2.
- [22] Y. Wang, J. Liu, W. Yang, and J. Zhang, "A heuristic for sigma set selection of UKF," *Proc. IEEE Int. Conf. on Sig. Proc.*, 2014, to be published.
- [23] Y. Wu, D. Hu, M. Wu, and X. Hu, "Unscented kalman filtering for additive noise case: augmented vs. non-augmented," in *Proc. Am. Control Conf.*, vol. 6, June 2005, pp. 4051–4055.
- [24] J. Kotecha and P. Djuric, "Gaussian particle filtering," *IEEE Trans. Sig. Proc.*, vol. 51, no. 10, pp. 2592–2601, Oct 2003.