Unscented Kalman Filtering for Additive Noise Case: Augmented versus Nonaugmented

Yuanxin Wu, Dewen Hu, *Member, IEEE*, Meiping Wu, and Xiaoping Hu

***Abstract*—This paper concerns the unscented Kalman filtering (UKF) for the nonlinear dynamic systems with additive process and measurement noises. It is widely accepted for such a case that the system state needs not to be augmented with noise vectors and the resultant nonaugmented UKF yields similar, if not the same, results to the augmented UKF. In this letter, we find that under the condition of $n+\kappa=\mathrm{const}$, the basic difference between them is that the augmented UKF draws a sigma set only once within a filtering recursion, while the nonaugmented UKF has to redraw a new set of sigma points to incorporate the effect of additive process noise. This difference generally favors the augmented UKF in that the odd-order moment information is partly captured by the nonlinearly transformed sigma points and propagated throughout the recursion. The simulation results agree well with the analyses.**

***Index Terms*—Dynamic system, unscented Kalman filtering, unscented transformation.**

I. Introduction

In light of the intuition that to approximate a probability distribution is easier than to approximate an arbitrary nonlinear transformation, Julier and Uhlmann [1], [2] invented the unscented transformation (UT) to make probabilistic inference. Eliminating the cumbersome derivation and evaluation of Jacobian/Hessian matrices, the UT-based unscented Kalman filter (UKF) is much easier to implement and performs better than the EKF. The original UKF was first formulated in its augmented form (1)–(3). It is believed that for the special (but often found) case, where process and measurement noises are additive, the computational complexity can be reduced by using the nonaugmented form, which presumably yields similar results [3], if not the same. The nonaugmented UKF has been accepted and employed to analyze the practical systems [4], [5]. In this letter, we will show that this assumption is not quite correct and that the nonaugmented UKF usage can lead to noticeable losses in accuracy. The contents are organized as follows. Section II shows the conditionally equivalent relationship between the nonaugmented and augmented UTs. This will facilitate the discussions about the UKF, which is essentially a natural extension of the UT to recursive estimation. Section III analyzes and compares the nonaugmented and augmented UKFs. Section IV examines a representative example in the signal processing community to support our findings, and the conclusions are drawn in Section V.

Manuscript received September 22, 2004; revised November 27, 2004. This work was supported in part by the National Natural Science Foundation of China under Grants 60374006, 60234030, 60171003, and 30370416, the Distinguished Young Scholars Fund of China under Grant 60225015, and the Ministry of Education of China TRAPOYT Project. An early version of this paper appears in *American Control Conference*, Portland, OR, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Vitor H. Nascimento.

The authors are with the Laboratory of Inertial Technology, Department of Automatic Control, College of Mechatronics and Automation, National University of Defense Technology, Changsha, 410073, China (e-mail: yuanx_wu@hotmail.com; dwhu@nudt.edu.cn).

Digital Object Identifier 10.1109/LSP.2005.845592

II. Unscented Transformation

Consider a one-step nonlinear transformation with additive noise

$$y=f(x)+w \tag{1}$$

where x is an $n\times 1$ random vector with mean $\hat{x}$ and covariance P_x, and w is an $m\times 1$ zero-mean noise vector with covariance Q that is uncorrelated with x. The problem is to calculate the mean $\hat{y}$ and covariance P_y of y. Note that here, neither x nor w is restricted to be Gaussian as long as their mean and covariance are given. Equation (1) can be reformulated through the state augmentation method as

$$y^a=f^a(x^a) \tag{2}$$

where the augmented random vector is $x^a=[x^T\ w^T]^T$, and the new nonlinear transformation is defined as $f^a(x^a)=f^a([x^T\ w^T]^T)=f(x)+w$. The problem now is to calculate the mean $\hat{y}^a$ and the covariance P_{y^a} of y^a.

A. Nonaugmented UT

1) The random vector x is approximated by $2n+1$ symmetric sigma points

$$\begin{aligned}
\chi_0&=\hat{x} & W_0&=\frac{\kappa}{(n+\kappa)} & &\\
\chi_i&=\hat{x}+\left(\sqrt{(n+\kappa)P_x}\right)_i & W_i&=\frac{1}{2(n+\kappa)}, & i&=1,\dots,n\\
\chi_{i+n}&=\hat{x}-\left(\sqrt{(n+\kappa)P_x}\right)_i & W_{i+n}&=\frac{1}{2(n+\kappa)} & &
\end{aligned} \tag{3}$$

where $(\sqrt{P})_i$ is the ith column of the matrix square root of P, and W_i is the weight associated with the ith sigma point. The scalar κ is a scaling parameter that is usually set to 0 or $3-n$, [2], [3]. Note that if κ is set to 0, the sigma points and their weights will be related to n, the dimension of x. $\kappa=3-n$ is selected so that the fourth-order moment information is mostly captured in the true Gaussian case [2]. In general, other choices of κ would lead to better or worse results, depending on specific characteristics of the integrand [6].

2) Instantiate each point through the function to yield a set of transformed sigma points

$$\gamma_i=f(\chi_i). \tag{4}$$

3) The mean $\hat{y}$ is given by the weighted average of the transformed points

$$\hat{y} = \sum_{i=0}^{2n} W_i \gamma_i. \tag{5}$$

4) The covariance P_y is the weighted outer product of the transformed points plus the noise covariance

$$P_y = \sum_{i=0}^{2n} W_i (\gamma_i - \hat{y})(\gamma_i - \hat{y})^T + Q. \tag{6}$$

B. Augmented UT

1) The augmented random vector x^a is approximated by $2(n+m)+1$ symmetric sigma points in (7), shown at the bottom of the page, where the weight W_i^a and scalar κ^a are counterparts of W_i and κ in (3). Note that $\hat{x}^a = [\hat{x}^T \; 0_{1\times m}]^T$ and

$$\begin{aligned}(\sqrt{P_{x^a}})_i &= \left(\sqrt{\begin{bmatrix} P_x & 0_{n\times m} \\ 0_{m\times n} & Q \end{bmatrix}}\right)_i \\ &= \begin{cases} \begin{bmatrix} (\sqrt{P_x})_i \\ 0_{m\times 1} \end{bmatrix}, & i = 1,\ldots,n \\ \begin{bmatrix} 0_{n\times 1} \\ (\sqrt{Q})_{i-n} \end{bmatrix}, & i = n+1,\ldots,n+m. \end{cases}\end{aligned} \tag{8}$$

Substituting (8) into (7) yields (9), shown at the bottom of the page.

2) Instantiate each point through the new function to yield a set of transformed sigma points

$$\gamma_i^a = f^a\left(\chi_i^a\right). \tag{10}$$

3) The mean $\hat{y}^a$ is the weighted average of the transformed points

$$\hat{y}^a = \sum_{i=0}^{2(n+m)} W_i^a \gamma_i^a. \tag{11}$$

Substituting (9) and (10) yields

$$\begin{aligned}\hat{y}^a &= W_0^a f^a\left(\chi_0^a\right) + \sum_{i=1}^{n} W_i^a \left[f^a\left(\chi_i^a\right) + f^a\left(\chi_{i+n}^a\right)\right] \\ &\quad + \sum_{i=1}^{m} W_i^a \left[f^a\left(\chi_{i+2n}^a\right) + f^a\left(\chi_{i+2n+m}^a\right)\right] \\ &= \frac{m+\kappa^a}{n+m+\kappa^a} f(\hat{x}) + \frac{1}{2(n+m+\kappa^a)} \\ &\quad \times \sum_{i=1}^{n} \left[f\left(\hat{x} + \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i\right)\right. \\ &\quad \left. + f\left(\hat{x} - \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i\right)\right].\end{aligned} \tag{12}$$

As compared with (5), the following equation

$$\begin{aligned}\hat{y}^a \equiv \hat{y} &= \sum_{i=0}^{2n} W_i \gamma_i \\ &= \frac{\kappa}{n+\kappa} f(\hat{x}) + \frac{1}{2(n+\kappa)} \\ &\quad \times \sum_{i=1}^{n} \left[f\left(\hat{x} + \left(\sqrt{(n+\kappa)P_x}\right)_i\right)\right. \\ &\quad \left. + f\left(\hat{x} - \left(\sqrt{(n+\kappa)P_x}\right)_i\right)\right]\end{aligned} \tag{13}$$

is satisfied if and only if

$$n + \kappa = n + m + \kappa^a \triangleq C. \tag{14}$$

That is to say, the sums $n+\kappa$ and $n+m+\kappa^a$ are identical and independent of the state dimension.

$$\begin{aligned}\chi_0^a &= \hat{x}^a & W_0^a &= \frac{\kappa^a}{(n+m+\kappa^a)} \\ \chi_i^a &= \hat{x}^a + \left(\sqrt{(n+m+\kappa^a)P_{x^a}}\right)_i & W_i^a &= \frac{1}{2(n+m+\kappa^a)}, \qquad i = 1,\ldots,n+m \\ \chi_{i+n+m}^a &= \hat{x}^a - \left(\sqrt{(n+m+\kappa^a)P_{x^a}}\right)_i & W_{i+n+m}^a &= \frac{1}{2(n+m+\kappa^a)}\end{aligned} \tag{7}$$

$$\begin{aligned}\chi_0^a &= \begin{bmatrix} \hat{x} \\ 0_{m\times 1} \end{bmatrix} & W_0^a &= \frac{\kappa^a}{(n+m+\kappa^a)} \\ \chi_i^a &= \begin{bmatrix} \hat{x} + \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i \\ 0_{m\times 1} \end{bmatrix} & W_i^a &= \frac{1}{2(n+m+\kappa^a)} \\ \chi_{i+n}^a &= \begin{bmatrix} \hat{x} - \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i \\ 0_{m\times 1} \end{bmatrix} & W_{i+n}^a &= \frac{1}{2(n+m+\kappa^a)}, \qquad i = 1,\ldots,n \quad j = 1,\ldots,m. \\ \chi_{j+2n}^a &= \begin{bmatrix} \hat{x} \\ \left(\sqrt{(n+m+\kappa^a)Q}\right)_j \end{bmatrix} & W_{j+2n}^a &= \frac{1}{2(n+m+\kappa^a)} \\ \chi_{j+2n+m}^a &= \begin{bmatrix} \hat{x} \\ -\left(\sqrt{(n+m+\kappa^a)Q}\right)_j \end{bmatrix} & W_{j+2n+m}^a &= \frac{1}{2(n+m+\kappa^a)}\end{aligned} \tag{9}$$

4) The covariance P_y^a is the weighted outer product of the transformed points

$$\begin{aligned} P_y^a &= \sum_{i=0}^{2(n+m)} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T \\ &= \sum_{i=0}^{2n} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T \\ &\quad + \sum_{i=2n+1}^{2(n+m)} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T. \end{aligned} \tag{15}$$

With (14) assumed, the first term on the right side becomes

$$\begin{aligned} \sum_{i=0}^{2n} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T &= \frac{C-m-n}{C}(f(\hat{x}) - \hat{y}) \\ \times (f(\hat{x}) - \hat{y})^T &+ \sum_{i=1}^{2n} W_i (f(\chi_i) - \hat{y})(f(\chi_i) - \hat{y})^T \end{aligned} \tag{16}$$

and the second term is

$$\begin{aligned} &\sum_{i=2n+1}^{2(n+m)} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T \\ &= \frac{1}{2C}\sum_{i=1}^{m} 2\left\{[f(\hat{x}) - \bar{y}][f(\hat{x}) - \bar{y}]^T + C(\sqrt{Q})_i(\sqrt{Q})_i^T\right\} \\ &= \frac{m}{C}(f(\hat{x}) - \hat{y})(f(\hat{x}) - \hat{y})^T + Q. \end{aligned} \tag{17}$$

Therefore

$$P_y^a = \sum_{i=0}^{2n} W_i (f(\chi_i) - \hat{y})(f(\chi_i) - \hat{y})^T + Q \equiv P_y. \tag{18}$$

In summary, the UT is configurable with the free parameter κ. Except the condition that (14) is satisfied, the nonaugmented UT will be different from the augmented UT. Coincidently, the common version of the UT selects $C = 3$, [2], [7] which leads to equivalence of the nonaugmented and augmented UTs. In this case, $\kappa = 3 - n$ and $\kappa^a = 3 - n - m$. For other versions, such as the simplex UT [2, App. III] and the UT that uses $\kappa = 0$ [2, Sec. III] and [8], it will be another story. The discussions in the next section are made about the common UT.

III. UKF

The UKF is a straightforward extension of the UT to the recursive estimation [1], [2]. For the sake of brevity, we prefer to treat the UT as a "black box" rather than to get involved in details again, as in Section II.

The prediction of the resulting UKF, whether based on the nonaugmented UT or the augmented UT, consists of two concatenated UTs: one for the process function (the first UT) directly followed by the other for the measurement function (the second UT). The second UT for the measurement function makes a difference between the nonaugmented UKF and augmented UKF. For the nonaugmented UKF, readers are referred to the addition of the noise covariance in (6). The transformed sigma points in the first UT only reflect the statistical information in $f(x)$ and do not consider the effect of the process noise at all [see the derivations from (3)–(5)]. So, just before the second UT, the nonaugmented UKF has to[1] redraw a new set of sigma points to incorporate the effect of additive process noise ([3], Table 7.3.2). Regarding the augmented UKF, the transformed sigma points in the first UT for the process function can be retained and then propagated through the measurement equation ([2], Step 5 of Fig. 7). By doing so, the computation of redrawing sigma points is spared, and more importantly, the odd-moment information is partly captured[2] and well propagated throughout one filtering recursion. In contrast, because of having to use the redrawn symmetric sigma points, the nonaugmented UKF is unable to propagate odd-moment information. Although the transformed sigma points of the first UT do capture the odd-moment information, the indispensable regeneration of a new sigma set for the second UT interrupts its propagation. Expectably, the augmented UKF would be identical to the nonaugmented UKF if the odd-moment information was intentionally abandoned through redrawing a new sigma set in the second UT.

Referring to Section II, the principle of the UT is to capture the first two moments of the random vector (x or x^a) via a set of sigma points. However, it should be made clear that with another set of sigma points capturing extra statistical information other than mean and covariance instead, the UT would hopefully yield better results. This is a natural conclusion that can be readily deduced from the Monte Carlo method [10], which represents a distribution by a collection of samples from that distribution. Therefore, the difference in drawing sigma points between the two versions of UKF generally favors the augmented UKF in that the extra odd-order moment information is partly captured by the nonlinearly transformed sigma points in the first UT and propagated throughout the whole recursion.

IV. Example

Both nonaugmented and augmented UKFs are applied to the univariate nonstationary growth model (UNGM) [11]. The discrete-time dynamic system equation for this model can be written as

$$\begin{aligned} x_n &= 0.5x_{n-1} + 25\frac{x_{n-1}}{1 + x_{n-1}^2} + 8\cos(1.2(n-1)) + u_n \\ y_n &= \frac{x_n^2}{20} + v_n, \qquad n = 1, \ldots, N \end{aligned} \tag{19}$$

where the process noise u_n and measurement noise v_n are both Gaussian noises with zero mean and unity variance. The reference data were generated using $x_0 = 0.1$ and $N = 500$. The bimodality makes this problem more difficult to address using conventional methods.[3]

The initial conditions were $\hat{x}_0 = 0, P_0 = 1$. The performance of the two UKFs was compared using the mean squared error (MSE) defined by $\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_n - \hat{x}_n)^2$.

[1]Unfortunately, the sigma points redrawing was neglected in tackling practical systems [4], [5].

[2]Formally proving that the transformed sigma points can capture significant odd-moment information would, in general, be rather involved. In [9], we presented a one-dimensional example and examined the third-order moment information (skew).

[3]Admittedly, the UKF's performance for this example may be unsatisfactory since a Gaussian approximation is implicitly made to the posterior distribution. However, it did not hinder the comparison of two versions of UKFs hereafter.

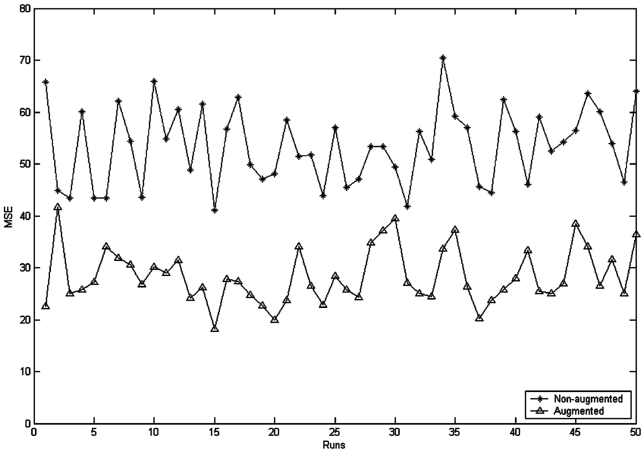

Fig. 1. MSEs across 50 random runs.

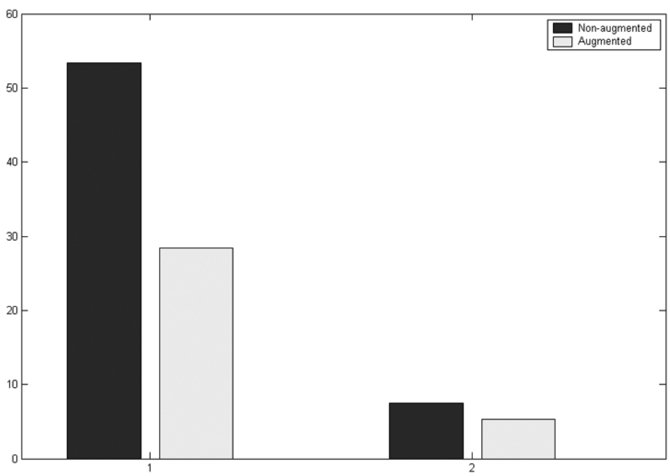

Fig. 2. Mean and standard variance of MSEs.

A large number of computer runs were carried out. Fig. 1 plots the MSEs for 50 random runs. Fig. 2 plots the mean and standard variance of the MSEs. The nonaugmented UKF's MSE is nearly twice as much as that of the augmented UKF, and the standard variance is also evidently smaller for the augmented UKF. As discussed above, the superiorities of the augmented UKF are mainly owed to its capability in capturing and propagating odd-moment information throughout one filtering recursion. The augmented UKF with an unnecessary sigma set regeneration inserted was also simulated and yielded, as expected, the same results as the nonaugmented UKF.

The computation time of the nonaugmented UKF is half of that of the augmented UKF in our simulation. One of the main computational disadvantages of the augmented UKF is that there are more sigma points that have to be propagated through the nonlinear process and measurement equations. Referring to (3) and (9), however, the arithmetic operations for the last extra $2m$ sigma points in (9) can be efficiently reduced using the results of the first sigma point χ_0^a. On the other hand, for the dynamic system with uncorrelated process and measurement noises, in the above example, for instance, the computation complexity can be further lowered by calculating low-dimensional matrix square roots instead. Equipped with the optimized implementation, the augmented UKF promises to yield better performance with comparable computational expense.

V. Conclusions

In this letter, we have analyzed and compared two alternative versions of UT-based filters for the nonlinear dynamic system with additive noises: the nonaugmented UKF and the augmented UKF. We proved that the nonaugmented UT is identical to the augmented counterpart only if $n + \kappa = \mathrm{const}$ is satisfied. We pointed out that the basic difference between the augmented and nonaugmented UKFs is that the former draws sigma points only once in a recursion, while the latter has to redraw a new set of sigma points to incorporate the effect of additive process noise. This difference generally favors the augmented UKF in that the odd-order moment information is captured by the transformed sigma points and well propagated within one recursion. On the other hand, if a new (but unnecessary) set of sigma points were redrawn in the augmented UKF, it would be identical to and yield exactly the same results as the nonaugmented UKF. The simulation results of a representative example agree well with our conclusions.

Acknowledgment

The authors would like to acknowledge the associate editor and anonymous reviewers for their insightful comments, which considerably improved this paper. The first author would also like to thank Dr. S. J. Julier for his inspiring discussions on certain related topics.

References

[1] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

[2] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[3] E. A. Wan and R. v. d. Merwe, "The unscented Kalman filter," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. New York: Wiley, 2001, ch. 7, pp. 221–280.

[4] A. Farina, B. Ristic, and D. Benvenuti, "Tracking a ballistic target: comparison of several nonlinear filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 854–867, Jul. 2002.

[5] J. L. Crassidis and F. L. Markley, "Unscented filtering for spacecraft attitude estimation," *J. Guid. Control Dyn.*, vol. 26, pp. 536–542, 2003.

[6] U. N. Lerner, "Hybrid Bayesian networks for reasoning about complex systems," Ph.D. dissertation, Stanford Univ., Stanford, CA, 2002.

[7] R. v. d. Merwe, "Sigma-point Kalman filters for probabilistic inference in dynamic state-space models," electrical and computer engineering Ph.D. dissertation, Oregon Health Sciences Univ., Portland, OR, 2004.

[8] J. K. Uhlmann, "Simultanenous map building and localization for real time applications," Transfer thesis, Univ. Oxford, Oxford, U.K., 1994.

[9] Y. Wu, D. Hu, M. Wu, and X. Hu. Unscented Kalman kiltering for additive noise case: Augmented versus nonaugmented (longer version) [Online]. Available: http://yuanxinwu.vip.sina.com/publications/UT%20–%20Augmented%20vs.%20Non-augmented.pdf.2004

[10] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[11] J. H. Kotecha and P. A. Djuric, "Gaussian particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.