Unscented Rauch-Tung-Striebel Smoother

Simo Särkkä*, *Member, IEEE*

***Abstract*— This article considers the application of the unscented transform to optimal smoothing of non-linear state space models. In this article, a new Rauch-Tung-Striebel type form of the fixed-interval unscented Kalman smoother is derived. The new smoother differs from the previously proposed two-filter formulation based unscented Kalman smoother in the sense that it is not based on running two independent filters forward and backward in time. Instead, a separate backward smoothing pass is used, which recursively computes corrections to the forward filtering result. The smoother equations are derived as approximations to the formal Bayesian optimal smoothing equations. The performance of the new smoother is demonstrated with a simulation.**

***Index Terms*— unscented Kalman smoother, Rauch-Tung-Striebel smoother, unscented transform**

I. Introduction

Optimal smoothing in context of state space models refers to statistical (Bayesian) methodology that can be used for computing estimates of the past state history of a time varying system based on the history of noisy measurements obtained from it. Phenomena, which can be modeled as this kind of state space models can be found, for example, in navigation, aerospace engineering, space engineering, remote surveillance, telecommunications, physics, audio signal processing, control engineering, several other fields [1]–[10].

Optimal smoothing is closely related to optimal filtering, which is a class methods for computing estimates of current and future states of the system. The formal solutions to the filtering [1], [11]–[13] and smoothing [14]–[17] problems are well known, and numerous approximation methods have been proposed, for both the filtering [1], [4], [18]–[22] and smoothing [18], [23]–[28] cases.

In this article we shall concentrate on the *unscented transform* [22], [27], [29]–[31], which is a relatively new method for forming Gaussian approximations to random variables that are obtained as non-linear transformations of Gaussian random variables. The unscented transform was first applied to optimal filtering of non-linear discrete-time systems, and the corresponding filter is called the *unscented Kalman filter* [22], [27], [31]. Continuous-time and continuous-discrete-time versions of the filter have been presented in [32].

The unscented transform has also been used for approximating the optimal smoothing solutions of state space models. The unscented Kalman smoother, which is presented in [27] is based on computing a suitable linear combination of two filters, which are ran in forward and backward directions. The backward filter is also a UKF, which uses the inverse of the forward dynamic model as the dynamic model for the filter. This form of optimal smoother can be interpreted as an approximate non-linear extension to the Fraser's two-filter smoother presented in [16].

Simo Särkkä* is with Helsinki University of Technology, P.O. Box 9203. FIN-02015 HUT, Finland. E-mail: ssarkka@lce.hut.fi, Tel: +358 40 757 0730, Fax: +358 9 451 4830.

However, as discussed in [33], computing the inverse of the forward dynamic model and performing the backward filtering using it as the dynamic model does not in general lead to the right result. The general two-filter smoothing equations have been presented in [33], together with Monte Carlo based methods for approximating them.

The Rauch-Tung-Striebel (RTS) smoother [15] differs from the two-filter smoother by that in the RTS smoother the measurements are first processed by the forward filter and then a separate backward smoothing pass is used for obtaining the smoothing solution. In this article, a new unscented transform based optimal smoother is derived, which is of the same form as the Rauch-Tung-Striebel smoother.

A. Problem Formulation

In this article we shall consider a *state space models* of the form

$$\begin{aligned} x_k &= f_{k-1}(x_{k-1}, q_{k-1}) \\ y_k &= h_k(x_k, r_k), \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state, $y_k \in \mathbb{R}^m$ is the measurement at time t_k, $q_{k-1} \sim \mathrm{N}(0, Q_{k-1})$ is the Gaussian process noise, $r_k \sim \mathrm{N}(0, R_k)$ is the Gaussian measurement noise, $f_{k-1}(\cdot)$ is the dynamic model function and $h_k(\cdot)$ is the measurement model function. The time steps k run from 0 to T and at time step 0 there is no measurement, only the prior distribution $x_0 \sim \mathrm{N}(m_0, P_0)$.

The purpose of the *smoothing algorithm* is to find approximations to the smoothing distributions $p(x_k \,|\, y_{1:T})$ for all $k \in \{0, 1, \ldots, T\}$ and in this article the approximations are chosen to be Gaussian:

$$p(x_k \,|\, y_{1:T}) \approx \mathrm{N}(x_k \,|\, m_k^s, P_k^s).$$

B. Bayesian Optimal Filtering and Smoothing Equations

Consider a generic state space model of the form

$$\begin{aligned} x_k &\sim p(x_k \,|\, x_{k-1}) \\ y_k &\sim p(y_k \,|\, x_k), \end{aligned} \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the state, $y_k \in \mathbb{R}^m$ is the measurement, $p(x_k \,|\, x_{k-1})$ is the transition density of the dynamic model and $p(y_k \,|\, x_k)$ is the likelihood density of the measurements. It is easy to see that the model (1) is a special case of this model.

Formally, the filtering distributions of the model are given by the following *optimal filtering equations* [13], [14]:

1) *Prediction step:*

$$p(x_k \,|\, y_{1:k-1}) = \int p(x_k \,|\, x_{k-1})\, p(x_{k-1} \,|\, y_{1:k-1})\, \mathrm{d}x_{k-1}. \tag{3}$$

2) *Update step:*

$$p(x_k \,|\, y_{1:k}) = \frac{p(y_k \,|\, x_k)\, p(x_k \,|\, y_{1:k-1})}{\int p(y_k \,|\, x_k)\, p(x_k \,|\, y_{1:k-1})\, \mathrm{d}x_k}. \tag{4}$$

The *optimal smoothing equations* [14] of the model can be written in two alternative forms [33]:

1) *Two-filter smoother*:

$$p(x_k \mid y_{1:T}) \propto p(x_k \mid y_{1:k})\, p(y_{k+1:T} \mid x_k), \tag{5}$$

where the first term on the right hand side is computed by the optimal filter and the second can be computed with a filter, which runs backwards in time. The unscented Kalman smoother presented in [27] can be seen to be an approximation to this form of smoother.

2) *Forward-backward smoother*:

$$p(x_k \mid y_{1:T}) = p(x_k \mid y_{1:k}) \int \left[\frac{p(x_{k+1} \mid x_k)\, p(x_{k+1} \mid y_{1:T})}{p(x_{k+1} \mid y_{1:k})} \right] \mathrm{d}x_{k+1}, \tag{6}$$

where $p(x_k \mid y_{1:k})$ is the filtering distribution of the time step k and $p(x_{k+1} \mid y_{1:k})$ is the predicted distribution of the time step $k+1$, which can be computed by the first of the optimal filtering equations (3). The smoothing recursion is started from last time step $k = T$ and proceeded backwards in time. The smoothing algorithm presented in this article is based on this forward-backward smoother equation.

The filtering and smoothing equations are only formal in the sense that they rarely can be directly used in practical computations, because they are computationally intractable. For this reason, numerical approximations are required.

II. Main Results

In this section we shall present the *unscented Rauch-Tung-Striebel smoother* or the forward-backward unscented Kalman smoother, which can be used for approximating the smoothing solutions of state space models of the form (1).

A. Unscented RTS Smoother

For the purpose of deriving the Rauch-Tung-Striebel type of smoother, the forward-backward smoothing equation (6) can be divided into the following three steps:

1) Form the joint distribution of x_k and x_{k+1} given $y_{1:k}$:

$$p(x_k, x_{k+1} \mid y_{1:k}) = p(x_{k+1} \mid x_k)\, p(x_k \mid y_{1:k}), \tag{7}$$

where $p(x_k \mid y_{1:k})$ is the filtering distribution of the current time step.

2) Compute the conditional distribution of x_k given x_{k+1} and $y_{1:k}$ by conditioning the joint distribution of x_k and x_{k+1} to x_{k+1}

$$p(x_k \mid x_{k+1}, y_{1:k}) = \frac{p(x_k, x_{k+1} \mid y_{1:k})}{p(x_{k+1} \mid y_{1:k})}, \tag{8}$$

where the denominator term is given as

$$p(x_{k+1} \mid y_{1:k}) = \int p(x_{k+1} \mid x_k)\, p(x_k \mid y_{1:k})\, \mathrm{d}x_k. \tag{9}$$

But now, due to the Markov properties of the state space model we have $p(x_k \mid x_{k+1}, y_{1:T}) = p(x_k \mid x_{k+1}, y_{1:k})$ and thus it follows that

$$p(x_k \mid x_{k+1}, y_{1:T}) = \frac{p(x_k, x_{k+1} \mid y_{1:k})}{p(x_{k+1} \mid y_{1:k})}. \tag{10}$$

3) The joint distribution of x_k and x_{k+1} given $y_{1:T}$ can be now computed as

$$p(x_k, x_{k+1} \mid y_{1:T}) = p(x_k \mid x_{k+1}, y_{1:T})\, p(x_{k+1} \mid y_{1:T}), \tag{11}$$

where $p(x_{k+1} \mid y_{1:T})$ is the smoothing distribution of the next time step. The smoothing distribution of x_k is given by marginalizing the joint distribution over x_{k+1}:

$$p(x_k \mid y_{1:T}) = \int p(x_k \mid x_{k+1}, y_{1:T})\, p(x_{k+1} \mid y_{1:T})\, \mathrm{d}x_{k+1}. \tag{12}$$

Assume that the (approximate) mean and covariance of the filtering distributions

$$p(x_k \mid y_{1:k}) \approx \mathrm{N}(x_k \mid m_k, P_k),$$

for the model (1) have been computed by the unscented Kalman filter or a similar method. Further assume that the smoothing distribution of time step $k+1$ is known and Gaussian

$$p(x_{k+1} \mid y_{1:T}) \approx \mathrm{N}(x_{k+1} \mid m^s_{k+1}, P^s_{k+1}).$$

An unscented transform based approximation to the optimal smoothing solution can be derived as follows:

1) Generate unscented transform based Gaussian approximation to the joint distribution of x_k and x_{k+1}, that is, to the equation (7):

$$\begin{pmatrix} x_k \\ x_{k+1} \end{pmatrix} \mid y_{1:k} \sim \mathrm{N}\left(\begin{pmatrix} m_k \\ m^-_{k+1} \end{pmatrix}, \begin{pmatrix} P_k & C_{k+1} \\ C^T_{k+1} & P^-_{k+1} \end{pmatrix} \right), \tag{13}$$

This can be done by concatenating the state and process noise to a new augmented random variable $\tilde{x}_k = (x^T_k \; q^T_k)^T$, which then has the distribution

$$\tilde{x}_k \mid y_{1:k} \sim \mathrm{N}\left(\begin{pmatrix} m_k \\ 0 \end{pmatrix}, \begin{pmatrix} P_k & 0 \\ 0 & Q_k \end{pmatrix} \right).$$

It is now easy to use the unscented transform for forming a Gaussian approximation to the joint distribution of $\tilde{x}_k = (x^T_k \; q^T_k)^T$ and $x_{k+1} = f_k(x_k, q_k)$.

2) Because the distribution (13) is Gaussian, the conditioning in equation (8) or (10) can be obtained by the computation rules of Gaussian distributions and this conditional distribution is again Gaussian. This results in the approximation

$$x_k \mid y_{1:T} \sim \mathrm{N}(m'_{k+1}, P'_{k+1}),$$

where

$$\begin{aligned} D_k &= C_{k+1}\, [P^-_{k+1}]^{-1} \\ m'_{k+1} &= m_k + D_k (x_{k+1} - m^-_{k+1}) \\ P'_{k+1} &= P_k - D_k\, P^-_{k+1}\, D^T_k. \end{aligned}$$

3) If the smoothing distribution of the next time step is assumed to be known and Gaussian

$$p(x_{k+1} \mid y_{1:T}) \approx \mathrm{N}(x_{k+1} \mid m^s_{k+1}, P^s_{k+1}),$$

then the distribution (11) is

$$\begin{pmatrix} x_k \\ x_{k+1} \end{pmatrix} \mid y_{1:T} \sim \mathrm{N}(m''_{k+1}, P''_{k+1}),$$

where

$$m''_{k+1} = \begin{pmatrix} m_k + D_k (x_{k+1} - m^-_{k+1}) \\ m^s_{k+1} \end{pmatrix}$$
$$P''_{k+1} = \begin{pmatrix} D_k\, P^s_{k+1}\, D^T_k + P'_{k+1} & D_k\, P^s_{k+1} \\ P^s_{k+1}\, D^T_k & P^s_{k+1} \end{pmatrix}.$$

Marginalizing over x_{k+1} leads to the Gaussian approximation to the smoothing distribution of the step k:

$$x_k \mid y_{1:T} \sim \mathrm{N}(m^s_k, P^s_k),$$

where

$$\begin{aligned} m^s_k &= m_k + D_k \left[m^s_{k+1} - m^-_{k+1} \right] \\ P^s_k &= P_k + D_k \left[P^s_{k+1} - P^-_{k+1} \right] D^T_k. \end{aligned} \tag{14}$$

In summary, a single step of the *unscented RTS smoother* can be now performed as follows:

1) Form the matrix of sigma points of the augmented random variable $\tilde{x}_k = (x^T_k\ q^T_k)^T$

$$\tilde{X}_k = \begin{pmatrix} \tilde{m}_k & \cdots & \tilde{m}_k \end{pmatrix} + \sqrt{c} \begin{pmatrix} 0 & \sqrt{\tilde{P}_k} & -\sqrt{\tilde{P}_k} \end{pmatrix}.$$

where

$$\tilde{m}_k = \begin{pmatrix} m_k \\ 0 \end{pmatrix} \qquad \tilde{P}_k = \begin{pmatrix} P_k & 0 \\ 0 & Q_k \end{pmatrix}.$$

2) Propagate the sigma points through the dynamic model:

$$\tilde{X}^-_{k+1,i} = f_k(\tilde{X}^x_{k,i}, \tilde{X}^q_{k,i}), \quad i = 1 \ldots 2n+1,$$

where $\tilde{X}^x_{k,i}$ and $\tilde{X}^q_{k,i}$ denote the parts of the augmented sigma point i, which correspond to x_k and q_k, respectively.

3) Compute the predicted mean m^-_{k+1}, the predicted covariance P^-_{k+1} and the cross-covariance C_{k+1}:

$$\begin{aligned} m^-_{k+1} &= \sum_i W^{(m)}_{i-1}\, \tilde{X}^-_{k+1,i} \\ P^-_{k+1} &= \sum_i W^{(c)}_{i-1}\, (\tilde{X}^-_{k+1,i} - m^-_{k+1})\, (\tilde{X}^-_{k+1,i} - m^-_{k+1})^T \\ C_{k+1} &= \sum_i W^{(c)}_{i-1}\, (\tilde{X}^x_{k,i} - m_k)\, (\tilde{X}^-_{k+1,i} - m^-_{k+1})^T, \end{aligned}$$

where the definitions of the weights $W^{(m)}_i$ and $W^{(c)}_i$ are the same as in [32].

4) Compute the smoother gain D_k, the smoothed mean m^s_k and the covariance P^s_k:

$$\begin{aligned} D_k &= C_{k+1}\, [P^-_{k+1}]^{-1} \\ m^s_k &= m_k + D_k \left[m^s_{k+1} - m^-_{k+1} \right] \\ P^s_k &= P_k + D_k \left[P^s_{k+1} - P^-_{k+1} \right] D^T_k. \end{aligned}$$

The above procedure is a recursion, which can be used for computing the smoothing distribution of step k from the smoothing distribution of time step $k+1$. Because the smoothing distribution and filtering distribution of the last time step T are the same, we have $m^s_T = m_T$, $P^s_T = P_T$, and thus the recursion can be used for computing the smoothing distributions of all time steps by starting from the last step $k = T$ and proceeding backwards to the initial step $k = 0$.

III. Illustrative Example

A. Re-entry Vehicle Tracking

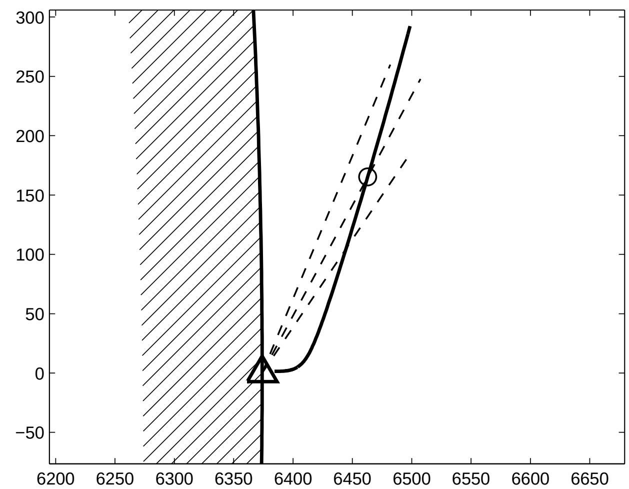

Fig. 1. In reentry target tracking problem radar is used for tracking a space vehicle, which enters the atmosphere at a very high speed.

As an example we consider a simulated re-entry tracking problem (see, Figure 1), where a radar is used for tracking a space vehicle, which enters the atmosphere at a very high speed. The problem was used for demonstrating the performance of UKF in [31], [34].

In the simulation the parameters were selected to be the same as used in [31], [34] and the following methods were tested:

- *EKF:* Extended Kalman filter, where first order Taylor series expansions of the dynamic and measurements models were used for approximating the non-linearities. Note that unlike in [31], [34] the analytical derivatives were used in the EKF, not finite difference approximations.
- *UKF:* Unscented Kalman filter, where the unscented transform was used for approximating the non-linearities.
- *EKS:* Extended Kalman smoother, where the results of forward and backward EKFs where combined to get the smoothed result.
- *UKS:* Unscented Kalman smoother, where the results of forward and backward UKFs were combined to get the smoothed result. The approximate backward dynamic model was obtained by changing the direction of time in the continuous-time dynamic model and making Euler discretized approximation to it.
- *URTSS:* Unscented Rauch-Tung-Striebel smoother proposed in this article.

The results of 1000 Monte Carlo simulations are shown in Table I. The results of the EKF and UKF are almost identical in RMSE sense. The result of EKS is not available, because

TABLE I
MEANS AND STANDARD DEVIATIONS OF RMSE VALUES OF THE POSITION IN 1000 MONTE CARLO RUNS OF THE RE-ENTRY TRACKING PROBLEM.

Method	*E[RMSE]*	*STD[RMSE]*
EKF	0.0083	0.0007
UKF	0.0083	0.0007
EKS	N/A	N/A
UKS	0.0044	0.0005
URTSS	0.0044	0.0005

with most of the data sets the matrix computations became so ill-conditioned that the smoother diverged. This is because the Taylor series based approximation does not work well for the inverse dynamic model. The UKS and URTSS give results that are almost identical and superior to the filters.

This simulation shows that the performance of the URTSS is the same as of UKS even though UKS has the additional information on the inverse of the dynamic model function. The linearization based EKS is completely inapplicable to this non-linear model.

IV. CONCLUSION

In this article, a new Rauch-Tung-Striebel type of unscented Kalman smoother for computing approximate optimal smoothing solutions of non-linear state space models has been proposed. Unlike the previously proposed unscented Kalman smoother, the new smoother is not based on combining results of two unscented Kalman filters running forward and backward in time. Instead, a separate backward smoothing pass is used for computing suitable corrections to the forward filtering result in order to obtain the smoothing solution. The performance of the smoother has been demonstrated and compared to other approaches with a numerical simulation.

ACKNOWLEDGMENT

The author would like to thank Aki Vehtari, Jouko Lampinen and Jouni Hartikainen for helpful discussion and comments.

REFERENCES

[1] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
[2] D. H. Titterton and J. L. Weston, *Strapdown Inertial Navigation Technology*. Peter Pregrinus Ltd., 1997.
[3] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global Positioning Systems, Inertial Navigation and Integration*. Wiley Interscience, 2001.
[4] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. Wiley Interscience, 2001.
[5] J. D. Murray, *Mathematical Biology*. Springer, 1993, vol. 19.
[6] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part I.* John Wiley & Sons, New York, 1968.
[7] ——, *Detection, Estimation, and Modulation Theory Part II.* John Wiley & Sons, New York, 1971.
[8] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, ser. Applied mathematical Sciences. Springer, 2005, no. 160.
[9] S. J. Godsill and P. J. Rayner, *Digital Audio Restoration: A Statistical Model Based Approach*. Springer-Verlag, 1998.
[10] R. F. Stengel, *Optimal Control and Estimation*. Dover Publications, Inc., 1994.
[11] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, March 1960.
[12] R. E. Kalman and R. S. Bucy, “New results in linear filtering and prediction theory,” *Transactions of the ASME, Journal of Basic Engineering*, vol. 83, pp. 95–108, March 1961.
[13] Y. C. Ho and R. C. K. Lee, “A Bayesian approach to problems in stochastic estimation and control,” *IEEE Transactions on Automatic Control*, vol. 9, pp. 333–339, 1964.
[14] R. C. K. Lee, *Optimal Estimation, Identification and Control*. M.I.T. Press, 1964.
[15] H. E. Rauch, F. Tung, and C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3(8), pp. 1445–1450, 1965.
[16] D. Fraser and J. Potter, “The optimum linear smoother as a combination of two optimum linear filters,” *IEEE Transactions on Automatic Control*, vol. 14(4), pp. 387–390, 1969.
[17] C. T. Leondes, J. B. Peller, and E. B. Stear, “Nonlinear smoothing theory,” *IEEE Transactions on Systems Science and Cybernetics*, vol. 6(1), January 1970.
[18] A. Gelb, Ed., *Applied Optimal Estimation*. The MIT Press, 1974.
[19] P. Maybeck, *Stochastic Models, Estimation and Control, Volume 1*. Academic Press, 1979.
[20] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001.
[21] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter*. Artech House, 2004.
[22] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, “A new method for the nonlinear transformation of means and covariances in filters and estimators,” *IEEE Transactions on Automatic Control*, vol. 45(3), pp. 477–482, March 2000.
[23] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communications and Control*. McGraw-Hill Book Company, 1971.
[24] P. Maybeck, *Stochastic Models, Estimation and Control, Volume 2*. Academic Press, 1982.
[25] F. L. Lewis, *Optimal Estimation with an Introduction to Stochastic Control Theory*. John Wiley & Sons, 1986.
[26] G. Kitagawa, “Monte Carlo filter and smoother for non-Gaussian nonlinear state space models,” *Journal of Computational and Graphical Statistics*, vol. 5, pp. 1–25, 1996.
[27] E. A. Wan and R. van der Merwe, “The unscented Kalman filter,” in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. Wiley, 2001, ch. 7.
[28] S. J. Godsill, A. Doucet, and M. West, “Monte Carlo smoothing for nonlinear time series,” *Journal of the American Statistical Association*, vol. 99(465), pp. 156–168, 2004.
[29] S. J. Julier and J. K. Uhlmann, “A general method of approximating nonlinear transformations of probability distributions,” Robotics Research Group, Department of Engineering Science, University of Oxford, Tech. Rep., 1995.
[30] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, “A new approach for filtering nonlinear systems,” in *Proceedings of the 1995 American Control, Conference, Seattle, Washington*, 1995, pp. 1628–1632.
[31] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92(3), pp. 401–422, March 2004.
[32] S. Särkkä, “On unscented Kalman filtering for state estimation of continuous-time nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 52(9), pp. 1631–1641, 2007.
[33] M. Klaas, M. Briers, N. de Freitas, A. Doucet, S. Maskell, and D. Lang, “Fast particle smoothing: If I had a million particles,” in *Proceedings of ICML 2006*, 2006.
[34] S. J. Julier and J. K. Uhlmann, “Corrections to unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92(12), pp. 1958–1958, December 2004.