On Gaussian Optimal Smoothing of Non-Linear State Space Models

Simo Särkkä, *Member, IEEE* and Jouni Hartikainen

***Abstract*— In this note we shall present a new Gaussian approximation based framework for approximate optimal smoothing of non-linear stochastic state space models. The approximation framework can be used for efficiently solving non-linear fixed-interval, fixed-point and fixed-lag optimal smoothing problems. We shall also numerically compare accuracies of approximations, which are based on Taylor series expansion, unscented transformation, central differences and Gauss-Hermite quadrature.**

***Index Terms*— non-linear Rauch-Tung-Striebel smoothing, Gaussian assumed density smoothing, non-linear optimal smoothing, Bayesian smoothing**

I. Introduction

In this note we shall consider fixed-point, fixed-lag and fixed-interval optimal smoothing of *non-linear stochastic state space models* of the form

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

where $x_k \in \mathbb{R}^n$ is the state, $y_k \in \mathbb{R}^m$ is the measurement at time step 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, $f(\cdot)$ is the dynamic model function and $h(\cdot)$ 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 $x_0 \sim \mathrm{N}(m_0, P_0)$.

A. Gaussian Optimal Filtering

Gaussian optimal filtering or assumed density filtering (ADF) with Gaussian assumption (see, e.g., [1]) is a well known method, where the filtering distributions are assumed to be approximately Gaussian. In the method, the formal Bayesian filtering equations are then computed assuming that the distributions are indeed Gaussian. The ADF framework is similar to the classical non-linear filtering and smoothing formalism (see, e.g., [2]–[4]) in the sense that both are based on approximations of the first two moments of the distributions. However, the ADF framework is a bit more general than the classical formalism.

Based on the ADF method, in the article [5], Ito and Xiong presented a generalized numerical integration based framework for Gaussian filtering of state space models of the form (1), and framework was further analyzed and extended by Wu et al. in [6]. The framework is similar to the unscented transformation based UKF filtering formalism of Julier et al. [7], [8], but has been generalized to contain, for example, Gauss-Hermite quadrature and central differences based filters aside with the UKF and EKF. It is also possible to use other numerical integration methods instead of the mentioned, such as Cubature rules [9] or Monte Carlo integration [10].

In ADF, the Gaussian approximation to the filtering distribution is of the form

$$p(x_k \mid y_{1:k}) \approx \mathrm{N}(x_k \mid m_{k|k}, P_{k|k}), \tag{2}$$

where $\mathrm{N}(x_k \mid m_{k|k}, P_{k|k})$ denotes the multivariate Gaussian probability density with mean $m_{k|k}$ and covariance $P_{k|k}$.

The approximation is formed by first approximating the *prediction step* with the following integrals for the mean and covariance:

$$\begin{aligned} m_{k|k-1} &= \int f(x_{k-1})\, \mathrm{N}(x_{k-1} \mid m_{k-1|k-1}, P_{k-1|k-1})\, dx_{k-1} \\ P_{k|k-1} &= \int [f(x_{k-1}) - m_{k|k-1}]\, [f(x_{k-1}) - m_{k|k-1}]^T \\ &\quad \times \mathrm{N}(x_{k-1} \mid m_{k-1|k-1}, P_{k-1|k-1})\, dx_{k-1} + Q_{k-1}. \end{aligned} \tag{3}$$

On the *update step* we first compute the measurement mean, prediction covariance and cross-covariance by approximating the integrals

$$\begin{aligned} \hat{z} &= \int h(x_k)\, \mathrm{N}(x_k \mid m_{k|k-1}, P_{k|k-1})\, dx_k \\ P_{zz} &= \int [h(x_k) - \hat{z}]\, [h(x_k) - \hat{z}]^T \\ &\quad \times \mathrm{N}(x_k \mid m_{k|k-1}, P_{k|k-1})\, dx_k \\ P_{xz} &= \int [x_k - m_{k|k-1}]\, [h(x_k) - \hat{z}]^T \\ &\quad \times \mathrm{N}(x_k \mid m_{k|k-1}, P_{k|k-1})\, dx_k. \end{aligned} \tag{4}$$

The mean and covariance in the Gaussian approximation (2) can be then computed as

$$\begin{aligned} L_k &= P_{xz}\, (R_k + P_{zz})^{-1} \\ m_{k|k} &= m_{k|k-1} + L_k\, [y_k - \hat{z}] \\ P_{k|k} &= P_{k|k-1} - L_k\, P_{xz}^T. \end{aligned} \tag{5}$$

II. Main Results

In this note, we shall generalize the unscented transformation based smoothing formalism presented by Särkkä in [11], [12] in the same sense as Ito and Xiong, and Wu et al. [5], [6] have generalized the formalism of Julier at al. [7]. In addition to that we shall also show how the resulting methodology can be used for approximating non-linear fixed-point and fixed-lag smoothing solutions aside with the fixed-interval (Rauch-Tung-Striebel) smoothing solution. In Section II-D we shall generalize the results to models with non-additive process and measurement noises.

In mathematical terms we shall form the following Gaussian approximations to the smoothing solutions:

- *Fixed-interval smoothing solution*:

  $$p(x_k \mid y_{1:T}) \approx \mathrm{N}(x_k \mid m_{k|T}, P_{k|T}),$$

  where the interval length T is fixed and $k = 1, \ldots, T$.

Manuscript received April 26, 2010; revised April 26, 2010.

Simo Särkkä* is with Aalto University, P.O. Box 12200. FI-00076 AALTO, Finland. E-mail: ssarkka@lce.hut.fi, Tel: +358 50 512 4393, Fax: +358 9 470 24833.

Jouni Hartikainen is with Aalto University, P.O. Box 12200. FI-00076 AALTO, Finland. E-mail: jmjharti@lce.hut.fi, Tel: +358 50 512 4271, Fax: +358 9 470 24833.

- *Fixed point smoothing solution*:

$$p(x_j \mid y_{1:k}) \approx \mathrm{N}(x_j \mid m_{j|k}, P_{n|k}),$$

where the point index j is fixed and k is increasing.
- *Fixed lag smoothing solution*:

$$p(x_{k-n} \mid y_{1:k}) \approx \mathrm{N}(x_{k-n} \mid m_{k-n|k}, P_{k-n|k}),$$

where the lag length n is fixed and k is increasing.

A. General Fixed-Interval Smoother Equations

First assume that the means and covariances of the approximate filtering distributions

$$p(x_k \mid y_{1:k}) \approx \mathrm{N}(x_k \mid m_{k|k}, P_{k|k}),$$

for the model (1) have been computed by some Gaussian approximation based filter falling into the class of approximate filters presented in [5], [6]. Further assume that the smoothing distribution of time step $k+1$ is approximately Gaussian

$$p(x_{k+1} \mid y_{1:T}) \approx \mathrm{N}(x_{k+1} \mid m_{k+1|T}, P_{k+1|T}).$$

Due to the Markov properties the state, x_k is independent of $y_{k+1:T}$ given x_{k+1}, and thus we have (cf. [12]):

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

By using the *Bayes' rule* the distribution of x_k given x_{k+1} and $y_{1:T}$ can be expressed as

$$\begin{aligned} p(x_k \mid x_{k+1}, y_{1:T}) &= p(x_k \mid x_{k+1}, y_{1:k}) \\ &= \frac{p(x_k, x_{k+1} \mid y_{1:k})}{p(x_{k+1} \mid y_{1:k})}. \end{aligned} \tag{6}$$

That is, Gaussian approximation to $p(x_k \mid x_{k+1}, y_{1:T})$ can be formed as follows:

1) Form a Gaussian approximation to the joint distribution $p(x_k, x_{k+1} \mid y_{1:k})$.
2) Compute Gaussian approximation to the distribution $p(x_k \mid x_{k+1}, y_{1:T})$ by conditioning the joint distribution above to x_{k+1} with the well known computation rules of Gaussian distributions.

Here we shall use the ADF approach and form the approximation by matching the first two moments of the distribution $p(x_k, x_{k+1} \mid y_{1:k})$. Note that because the distribution is independent of the future measurements $y_{k+1:T}$, the approximation depends only on the filtering distributions and dynamic model.

If we assume that the filtering distribution is Gaussian $p(x_k \mid y_{1:k}) = \mathrm{N}(x_k \mid m_{k|k}, P_{k|k})$, then the approximation can be obtained as follows. First compute the Gaussian integrals

$$\begin{aligned} m_{k+1|k} &= \int f(x_k)\, \mathrm{N}(x_k \mid m_{k|k}, P_{k|k})\, dx_k \\ P_{k+1|k} &= \int [f(x_k) - m_{k+1|k}]\, [f(x_k) - m_{k+1|k}]^T \\ &\quad \times \mathrm{N}(x_k \mid m_{k|k}, P_{k|k})\, dx_k + Q_k \\ C_{k,k+1} &= \int [x_k - m_{k|k}]\, [f(x_k) - m_{k+1|k}]^T \\ &\quad \times \mathrm{N}(x_k \mid m_{k|k}, P_{k|k})\, dx_k. \end{aligned} \tag{7}$$

Note that the terms $m_{k+1|k}$ and $P_{k+1|k}$ are simply the predicted mean and covariance of the corresponding Gaussian filter and the terms $C_{k,k+1}$ could be easily computed and stored during the filtering. The integrals above can be approximated using analogous numerical integration or analytical approximation schemes as in the filtering case [5], [6], [9], [10].

The Gaussian approximation to the joint distribution can be then formed as:

$$p(x_k, x_{k+1} \mid y_{1:k}) \approx \mathrm{N}\left(\begin{pmatrix} m_{k|k} \\ m_{k+1|k} \end{pmatrix}, \begin{pmatrix} P_{k|k} & C_{k,k+1} \\ C_{k,k+1}^T & P_{k+1|k} \end{pmatrix} \right). \tag{8}$$

The division of the distributions in the equation (6) is just the definition of conditioning on x_{k+1} and thus by the computation rules of Gaussian distributions we get

$$\begin{aligned} p(x_k \mid x_{k+1}, y_{1:T}) &= p(x_k \mid x_{k+1}, y_{1:k}) \\ &= \mathrm{N}(x_k \mid m', P'), \end{aligned} \tag{9}$$

where

$$\begin{aligned} m' &= m_{k|k} + C_{k,k+1}\, P_{k+1|k}^{-1} (x_{k+1} - m_{k+1|k}) \\ P' &= P_{k|k} - C_{k,k+1}\, P_{k+1|k}^{-1}\, C_{k,k+1}^T. \end{aligned} \tag{10}$$

If we define an additional temporary variable, the smoother gain, as

$$G_k = C_{k,k+1}\, P_{k+1|k}^{-1} \tag{11}$$

then the equations can be written as:

$$\begin{aligned} m' &= m_{k|k} + G_k\, (x_{k+1} - m_{k+1|k}) \\ P' &= P_{k|k} - G_k\, P_{k+1|k}\, G_k^T. \end{aligned} \tag{12}$$

The joint distribution of x_k and x_{k+1} given all the data is now

$$\begin{aligned} p(x_{k+1}, x_k \mid y_{1:T}) &= p(x_k \mid x_{k+1}, y_{1:T})\, p(x_{k+1} \mid y_{1:T}) \\ &= \mathrm{N}\left(\begin{pmatrix} x_{k+1} \\ x_k \end{pmatrix} \,\middle|\, m'', P'' \right) \end{aligned} \tag{13}$$

where

$$\begin{aligned} m'' &= \begin{pmatrix} m_{k+1|T} \\ m_{k|k} + G_k\, (m_{k+1|T} - m_{k+1|k}) \end{pmatrix} \\ P'' &= \begin{pmatrix} P_{k+1|T} & P_{k+1|T}\, G_k^T \\ G_k\, P_{k+1|T} & P_{k|k} + G_k\, (P_{k+1|T} - P_{k+1|k})\, G_k^T \end{pmatrix}. \end{aligned} \tag{14}$$

The marginal distribution of x_k is then

$$p(x_k \mid y_{1:T}) = \mathrm{N}(x_k \mid m_{k|T}, P_{k|T}), \tag{15}$$

where

$$\begin{aligned} m_{k|T} &= m_{k|k} + G_k\, (m_{k+1|T} - m_{k+1|k}) \\ P_{k|T} &= P_{k|k} + G_k\, (P_{k+1|T} - P_{k+1|k})\, G_k^T. \end{aligned} \tag{16}$$

Thus the *general fixed-interval smoother* can be implemented by starting from the filtering solution $m_{T|T}, P_{T|T}$ and by performing the following computations on each time step $k = T-1, T-2, \ldots, 0$:

1) *Prediction:* Compute the predicted mean $m_{k+1|k}$, predicted covariance $P_{k+1|k}$ and cross-covariance $C_{k,k+1}$ from the filtering results using the equations (7).

2) *Smoothing:* Compute the smoothing gain G_k, the smoother mean $m_{k|T}$ and the covariance $P_{k|T}$ from the equations

$$\begin{aligned} G_k &= C_{k,k+1}\, P_{k+1|k}^{-1} \\ m_{k|T} &= m_{k|k} + G_k\,(m_{k+1|T} - m_{k+1|k}) \\ P_{k|T} &= P_{k|k} + G_k\,(P_{k+1|T} - P_{k+1|k})\, G_k^T. \end{aligned} \tag{17}$$

B. General Fixed-Point Smoother Equations

The general fixed-interval smoother described in the previous section has the property that given the gain sequence, we only need linear operations for performing the smoothing, and in this sense, the smoothing is a completely *linear operation*. The only non-linear operations in the smoother are in the approximations of the integrals (7). However, these operations are performed to the filtering results and thus we can compute the smoothing gain sequence G_k from the filtering results in causal manner. Because of these properties we may now derive a fixed-point smoother using similar methods as have been used for deriving the linear fixed-point smoother from the linear Rauch-Tung-Striebel smoother in [13].

The result is that the *general fixed-point smoother* algorithm for smoothing the time point j can be implemented by performing the following operations on each time step $k = 1, 2, 3, \ldots$:

1) *Gain computation:* Compute the predicted mean $m_{k|k-1}$, predicted covariance $P_{k|k-1}$ and cross-covariance $C_{k-1,k}$ from the filtering results using the equations (7). Then compute the gain from the equation

$$G_{k-1} = C_{k-1,k}\, P_{k|k-1}^{-1}. \tag{18}$$

2) *Fixed-point smoothing:*
   a) If $k < j$, just store the filtering result.
   b) If $k = j$, set $B_{j|j} = I$. The fixed-point smoothed mean and covariance on step j are equal to the filtered mean and covariance $m_{j|j}$ and $P_{j|j}$.
   c) If $k > j$, compute the smoothing gain and the fixed-point smoother mean and covariance:

$$\begin{aligned} B_{j|k} &= B_{j|k-1} G_{k-1} \\ m_{j|k} &= m_{j|k-1} + B_{j|k}[m_{k|k} - m_{k|k-1}] \\ P_{j|k} &= P_{j|k-1} + B_{j|k}[P_{k|k} - P_{k|k-1}] B_{j|k}^T. \end{aligned} \tag{19}$$

Because only constant number of computations is needed on each time step, the algorithm can be easily implemented in real time.

C. General Fixed-Lag Smoother Equations

It is also possible to derive a general fixed-lag smoother by using a similar procedure as in the previous section. However, this approach will lead to a numerically unstable algorithm as will be seen shortly. Following similar procedure as in [13], we arrive in the following equations:

$$\begin{aligned} m_{k-n|k} &= m_{k-n|k-n-1} \\ &\quad + G_{k-n-1}^{-1}[m_{k-n-1|k-1} - m_{k-n-1|k-n-1}] \\ &\quad + B_{k-n|k}[m_{k|k} - m_{k|k-1}] \\ P_{k-n|k} &= P_{k-n|k-n-1} \\ &\quad + G_{k-n-1}^{-1}[P_{k-n-1|k-1} - P_{k-n-1|k-n-1}] G_{k-n-1}^{-T} \\ &\quad + B_{k-n|k}[P_{k|k} - P_{k|k-1}] B_{k-n|k}^T. \end{aligned} \tag{20}$$

The equations above can be, *in principle*, used for recursively computing the fixed-lag smoothing solution. The number of computations does not depend on the lag length. This solution can be seen to be of the same form as the fixed-lag smoother given in [4], [13], [14]. Unfortunately, it has been shown [15] that this form of smoother is *numerically unstable* and thus not usable in practice.

In [16], [17] stable algorithms for optimal fixed-lag smoothing are derived by augmenting the n lagged states to a Kalman filter. This approach ensures the stability of the algorithm. Using certain simplifications it is possible to reduce the computations, and this is also possible when certain types extended Kalman filters are used [16], [17]. Unfortunately, such simplifications cannot be done in more general case and, for example, when the unscented transformation [7], [8] or a quadrature rule [5] is used, the required amount of computations becomes high, because the Cholesky factorization of the whole joint covariance of the n lagged states would be needed in the computations.

Another possibility, which is employed here, is to take advantage of the fact that Rauch-Tung-Striebel smoother equations are numerically stable and can be used for fixed-lag smoothing. The fixed-lag smoothing can be efficiently implemented by taking into account that the gain sequence needs to be evaluated only once, and the same gains can be used in different smoothers operating on different intervals. Thus the *general fixed-lag smoother* can be implemented by performing the following on each time step $k = 1, 2, 3, \ldots$:

1) *Gain computation:* During the Gaussian filter prediction step compute and store the predicted mean $m_{k|k-1}$, predicted covariance $P_{k|k-1}$ and cross-covariance $C_{k-1,k}$ defined in equations (7). Also compute and store the smoothing gain

$$G_{k-1} = C_{k-1,k}\, P_{k|k-1}^{-1}. \tag{21}$$

2) *Fixed-lag smoothing:* Using the stored gain sequence, compute the smoothing solutions for steps $j = k - n, \ldots, k$ using the following backward recursion, starting from the filtering solution on step $j = k$:

$$\begin{aligned} m_{j|k} &= m_{j|j} + G_j\,\left[m_{j+1|k} - m_{j+1|j}\right] \\ P_{j|k} &= P_{j|j} + G_j\,\left[P_{j+1|k} - P_{j+1|j}\right]\, G_j^T. \end{aligned} \tag{22}$$

The required number of computations per time step grows linearly with the length of lag. Thus the computational requirements are comparable to algorithms presented in [16], [17]. The algorithm defined in equations (20) would be computationally more efficient, but as already stated, it would be numerically unstable.

D. Non-Additive Noise Models

The idea of the assumed Gaussian approximation and moment matching can also be generalized to more general non-additive state space models of the form

$$\begin{aligned} x_k &= f(x_{k-1}, q_{k-1}) \\ y_k &= h(x_k, r_k), \end{aligned} \tag{23}$$

where q_{k-1} and r_k are Gaussian. The integrals in equations (7) should be then replaced with the more general expressions

$$\begin{aligned} m_{k+1|k} &= \int f(x_k, q_k) \, \mathrm{N}(x_k \mid m_{k|k}, P_{k|k}) \\ &\quad \times \mathrm{N}(q_k \mid 0, Q_k) \, dx_k \, dq_k \\ P_{k+1|k} &= \int [f(x_k, q_k) - m_{k+1|k}] \, [f(x_k, q_k) - m_{k+1|k}]^T \\ &\quad \times \mathrm{N}(x_k \mid m_{k|k}, P_{k|k}) \, \mathrm{N}(q_k \mid 0, Q_k) \, dx_k \, dq_k \\ C_{k,k+1} &= \int [x_k - m_{k|k}] \, [f(x_k, q_k) - m_{k+1|k}]^T \\ &\quad \times \mathrm{N}(x_k \mid m_{k|k}, P_{k|k}) \, \mathrm{N}(q_k \mid 0, Q_k) \, dx_k \, dq_k. \end{aligned} \tag{24}$$

Note that in the case of unscented transformation this approach is equivalent to the augmentation approach commonly used in unscented Kalman filtering [8], [18], [19]. The approach was also used in the unscented Rauch-Tung-Striebel smoother presented in [12]. The distribution of random variable q_k could also well be non-Gaussian and the only change would be to replace the Gaussian distributions $\mathrm{N}(q_k \mid 0, Q_k)$ with the non-Gaussian ones.

III. Simulation

A. Tracking of Maneuvering Target

As a simulation example we consider the problem of tracking a target in two dimensional space executing a maneuvering turn with unknown and time-varying turn rate. This is very typical setup in target tracking applications, and is useful for testing non-linear filters and smoothers since the dynamics are non-linear due to the unknown turn rate. Recently, very similar simulation setup was used in [9] for assessing the Cubature Kalman filter.

The dynamic model of the coordinated turn model [20] is

$$x_k = \begin{pmatrix} 1 & \frac{\sin(\omega \Delta t)}{\omega} & 0 & -\left(\frac{1-\cos(\omega \Delta t)}{\omega}\right) & 0 \\ 0 & \cos(\omega \Delta t) & 0 & -\sin(\omega \Delta t) & 0 \\ 0 & \frac{1-\cos(\omega \Delta t)}{\omega} & 1 & \frac{\sin(\omega \Delta t)}{\omega} & 0 \\ 0 & \sin(\omega \Delta t) & 0 & \cos(\omega \Delta t) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} x_{k-1} + q_{k-1}, \tag{25}$$

where the state of the target is $x = (x_1, \dot{x}_1, x_2, \dot{x}_2, \omega)$, and x_1, x_2 are the coordinates and $\dot{x}_1, \dot{x}_2$ are the velocities in two dimensional space. The process noise parameters used in the simulation were the same as in [9].

In the simulation setup we have four sensors measuring the angles θ between the target and the sensors. The non-linear measurement model for sensor i can be written as

$$\theta_k^i = \arctan\left(\frac{y_k - s_y^i}{x_k - s_x^i}\right) + r_k^i, \tag{26}$$

TABLE I
Tracking of Maneuvering Target: RMSE values of position, velocity and turn rate estimates for the tested filters and smoothers over 1000 simulations.

Filter	RMSE(pos)	RMSE(vel)	RMSE(ω)
EKF	264.6	248.2	0.1184
UKF	85.7	74.9	0.0375
CDKF	84.1	72.9	0.0374
GHKF	82.3	70.1	0.0374
Smoother	**RMSE(pos)**	**RMSE(vel)**	**RMSE(ω)**
ERTS	208.4	201.6	0.1132
URTS	60.8	35.9	0.0152
CDRTS	58.7	34.1	0.0148
GHRTS	56.3	32.1	0.0145

where (s_x^i, s_y^i) is the position of the sensor i in two dimensions, and $r_k^i \sim N(0, \sigma_\theta^2)$ is the measurement noise. The measurement noise in the angular measurement was assumed to be $\sigma_\theta = \sqrt{5}$ mrad. The target trajectory and measurements were simulated 1000 times for 100 time steps by drawing the initial state randomly from the prior on each simulation run. By using the simulated trajectories and measurements we assessed the performance of following filters and smoothers:

- EKF & ERTS: extended Kalman filter and RTS smoother.
- UKF & URTS: unscented Kalman filter and RTS smoother with transformation parameters $\alpha = 1$, $\beta = 0$ and $\kappa = 3 - n$.
- CDKF & CDRTS: Central difference Kalman filter and RTS smoother with step size $h_{cd} = \sqrt{3}$.
- GHKF & GHRTS: Gauss-Hermite Kalman filter and RTS smoother with $m = 3$ quadrature points.

Table I lists the RMSE values for the tested filters and smoothers, which shows that the errors of UKF, CDKF and GHKF are quite similar, but GHKF gives slightly smaller error than the other filters. The EKF errors are a couple of times higher than of the other methods. The effect of smoothing is similar with all the methods, that is, the smoothing simply reduces the estimation error, but the ordering of the methods stays the same.

IV. Conclusion

In this note, we have presented a new framework for approximate Gaussian optimal smoothing of non-linear state space models. The framework is based on generalizing the unscented transformation based smoothing formalism presented by Särkkä in [11], [12] in the same sense as Ito and Xiong, and Wu et al. [5], [6] generalized the filtering formalism of Julier et al. [7]. We have also shown how the resulting methodology can be used for approximating non-linear fixed-point and fixed-lag smoothing solutions aside with the fixed-interval (Rauch-Tung-Striebel) smoothing solution. A simulated example demonstrating the performance of the proposed smoothers has also been presented.

Acknowledgment

The authors like to thank Aki Vehtari and Jouko Lampinen for helpful discussions and support.

REFERENCES

[1] P. Maybeck, *Stochastic Models, Estimation and Control, Volume 2*. Academic Press, 1982.

[2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[3] A. Gelb, Ed., *Applied Optimal Estimation*. The MIT Press, 1974.

[4] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*. Chapman & Hall/CRC, 2004.

[5] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45(5), pp. 910–927, 2000.

[6] Y. Wu, D. Hu, M. Wu, and X. Hu, "A numerical-integration perspective on Gaussian filters," *IEEE Transactions on Signal Processing*, vol. 54(8), pp. 2910–2921, 2006.

[7] 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.

[8] ——, "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.

[9] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54(6), pp. 1254–1269, 2009.

[10] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51(10), 2003.

[11] S. Särkkä, "Recursive Bayesian inference on stochastic differential equations," Doctoral dissertation, Helsinki University of Technology, 2006.

[12] ——, "Unscented Rauch-Tung-Striebel smoother," *IEEE Transactions on Automatic Control*, vol. 53(3), pp. 845–849, 2008.

[13] J. Meditch, *Stochastic Optimal Linear Estimation and Control*. McGraw-Hill, 1969.

[14] H. E. Rauch, "Solutions to the linear smoothing problem," *IEEE Transactions on Automatic Control*, vol. AC-8, pp. 371–372, 1963.

[15] C. N. Kelly and B. D. O. Anderson, "On the stability of fixed-lag smoothing algorithms," *Journal of Franklin Institute*, vol. 291(4), pp. 271–281, April 1971.

[16] J. B. Moore, "Discrete-time fixed-lag smoothing algorithms," *Automatica*, vol. 9(2), pp. 163–174, March 1973.

[17] J. Moore and P. Tam, "Fixed-lag smoothing of nonlinear systems with discrete measurement," *Information Sciences*, vol. 6, pp. 151–160, 1973.

[18] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92(3), pp. 401–422, March 2004.

[19] 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.

[20] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. Wiley Interscience, 2001.