# SIGMA POINT METHODS IN OPTIMAL SMOOTHING OF NON-LINEAR STOCHASTIC STATE SPACE MODELS

*Simo Särkkä and Jouni Hartikainen*

Aalto University
Department of Biomedical Engineering and Computational Science
Rakentajanaukio 2, 02150 Espoo, Finland.

## ABSTRACT

In this article, we shall show how the sigma-point based approximations that have previously been used in optimal filtering can also be used in optimal smoothing. In particular, we shall consider unscented transformation, Gauss-Hermite quadrature and central differences based optimal smoothers. We briefly present the smoother equations and compare performance of different methods in simulated scenarios.

## 1. INTRODUCTION

This article is concerned with non-linear optimal smoothing, which in this context refers to methodology that can be used for computing estimates of a fixed time interval of states of a non-linear stochastic state space model given the measurements on the same interval. These kind of batch estimates of the whole history of a state trajectory are often useful, for example, in target tracking, navigation, audio signal processing, machine learning, time series analysis and time series prediction [1, 2, 3, 4, 5].

The stochastic state space model model is assumed to be of the form

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k, \end{aligned} \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state, $\mathbf{y}_k \in \mathbb{R}^d$ is the measurement at time step $k$, $\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q}_{k-1})$ is the Gaussian process noise, $\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R}_k)$ is the Gaussian measurement noise, $\mathbf{f}(\cdot)$ is the dynamic model function and $\mathbf{h}(\cdot)$ is the measurement model function. On time step $k = 0$ the distribution of the state is $\mathbf{x}_0 \sim \mathrm{N}(\mathbf{m}_0, \mathbf{P}_0)$, where the mean and covariance are known.

Classically, non-linear optimal smoothing methods have been derived as extensions to optimal filtering methods [6, 7, 8], which are methods for computing the optimal solution of the current state given the measurements up to the current point of time. In particular, the solutions to the linear Gaussian smoothing problems are well known [9, 10, 11] and well documented in many text books (see, e.g., [12, 13, 14, 15]).

The non-linear case has been commonly tackled with Taylor series based approximation methods, that is, with extended Kalman smoothers (EKS) or extended Rauch-Tung-Striebel (ERTS) smoothers. The ERTS approach and the more generalized formalism is also well documented in literature [12, 15, 16, 17, 18, 19]. The unscented transformation (UT) [20, 21] is a relatively recent alternative to the Taylor series approximations, and it was originally applied to non-linear filtering problems under the name unscented Kalman filter (UKF). More recently, the unscented transformation has also been applied to optimal smoothing [22, 23, 24].

The UKF can be interpreted to belong to a wider class of filters called sigma-point filters [25], which also includes other types of filters such as central differences Kalman filter (CDKF), Gauss-Hermite Kalman filter (GHKF) and a few others [26, 27, 28, 29]. In this article we shall use the general Gaussian smoothing equations derived in [30] and show how the sigma-point approximations can also be used in non-linear optimal smoothing.

## 2. SIGMA POINT BASED SMOOTHERS

### 2.1. General Gaussian Optimal Smoothing

Assume that we have applied a Gaussian optimal filter, such as one of the assumed density filters (ADF) proposed in [26, 27] to the state space model in Equations (1). The result of filtering is then given as

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \approx \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_{k|k}, \mathbf{P}_{k|k}),$$

where $\mathbf{m}_{k|k}$ and $\mathbf{P}_{k|k}$ are the mean and covariance computed by the filter. If we follow the unscented RTS smoother

SS thanks Academy of Finland's Centre of Excellence in Computational Complex Systems Research for funding.

JH thanks Finnish Graduate School in Computational Sciences (FICS) for funding.

derivation in [23, 24] and apply the assumed density approximation instead of the unscented transformation, we arrive at the following equations for the general RTS smoother [30]:

$$\begin{aligned}
\mathbf{m}_{k+1|k} &= \int \mathbf{f}(\mathbf{x}_k)\, \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_{k|k}, \mathbf{P}_{k|k})\, d\mathbf{x}_k \\
\mathbf{P}_{k+1|k} &= \int [\mathbf{f}(\mathbf{x}_k) - \mathbf{m}_{k+1|k}]\, [\mathbf{f}(\mathbf{x}_k) - \mathbf{m}_{k+1|k}]^T \\
&\quad \times \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_{k|k}, \mathbf{P}_{k|k})\, d\mathbf{x}_k + \mathbf{Q}_k \\
\mathbf{C}_{k,k+1} &= \int [\mathbf{x}_k - \mathbf{m}_{k|k}]\, [\mathbf{f}(\mathbf{x}_k) - \mathbf{m}_{k+1|k}]^T \\
&\quad \times \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_{k|k}, \mathbf{P}_{k|k})\, d\mathbf{x}_k \\
\mathbf{G}_k &= \mathbf{C}_{k,k+1}\, \mathbf{P}_{k+1|k}^{-1} \\
\mathbf{m}_{k|T} &= \mathbf{m}_{k|k} + \mathbf{G}_k\, (\mathbf{m}_{k+1|T} - \mathbf{m}_{k+1|k}) \\
\mathbf{P}_{k|T} &= \mathbf{P}_{k|k} + \mathbf{G}_k\, (\mathbf{P}_{k+1|T} - \mathbf{P}_{k+1|k})\, \mathbf{G}_k^T.
\end{aligned} \tag{2}$$

That is, if we start from the filtering result on step $k = T$ and iterate the above equations back to $k = 0$, we get the following approximate smoothing distributions:

$$p(\mathbf{x}_k \,|\, \mathbf{y}_{1:T}) \approx \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_{k|T}, \mathbf{P}_{k|T}).$$

A straight-forward way to approximate the integrals in Equations (2) is by Monte Carlo sampling, as has been done in Monte Carlo Kalman filter (MCKF) [31]. This approach is computationally heavy, but does provide very accurate approximations of the integrals when the number of Monte Carlo samples is high enough. We shall use the Monte Carlo approximations as reference results for evaluation of the sigma-point methods.

### 2.2. Linearization Based Smoother

To illustrate the usage of the Equations (2), we show how the linearization approximation can be used for deriving the extended Rauch-Tung-Striebel smoother. Consider the linear approximation:

$$\mathbf{f}(\mathbf{x}_k) \approx \mathbf{f}(\mathbf{m}_{k|k}) + \mathbf{F}_x(\mathbf{m}_{k|k})\, (\mathbf{x}_k - \mathbf{m}_{k|k}),$$

where $\mathbf{F}_x$ is the Jacobian matrix of $\mathbf{f}$. By substituting this into the integrals in Equations (2), we get the mean as follows:

$$\begin{aligned}
\mathbf{m}_{k+1|k} &= \int \mathbf{f}(\mathbf{x}_k) \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_k, \mathbf{P}_k)\, d\mathbf{x}_k \\
&= \int [\mathbf{f}(\mathbf{m}_k) + \mathbf{F}_x(\mathbf{m}_k)\, (\mathbf{x}_k - \mathbf{m}_k)] \\
&\quad \times \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_k, \mathbf{P}_k)\, d\mathbf{x}_k \\
&= \mathbf{f}(\mathbf{m}_{k|k}).
\end{aligned}$$

Similarly, for the covariance and cross-covariance we get the familiar equations

$$\begin{aligned}
\mathbf{P}_{k+1|k} &= \mathbf{F}_x(\mathbf{m}_{k|k})\, \mathbf{P}_{k|k}\, \mathbf{F}_x^T(\mathbf{m}_{k|k}) + \mathbf{Q}_k \\
\mathbf{C}_{k,k+1} &= \mathbf{P}_{k|k}\, \mathbf{F}_x^T(\mathbf{m}_{k|k}).
\end{aligned}$$

That is, the extended Rauch-Tung-Striebel smoother can be implemented by combining the above equations with the last three Equations in (2). In this article, this simple linearization based method is used as a base line in evaluation of the sigma-point methods.

### 2.3. Unscented Approximation

Unscented transformation (UT) [21] is a relatively recent method for approximating general transformations of random variables. As shown in [27], UT can be interpreted as an approximation to integrals of the form appearing in Equations (2). The underlying idea in the UT approximation is that it is easier to approximate the distribution of random variable under a non-linear transformation than the non-linear transformation itself [21]. In UT this is accomplished by using a set of sigma points, which capture the statistics of the original distribution exactly. The moments of the transformed distribution are then computed by feeding the points through the transformation and by computing the moments from the transformed sigma points.

In UT the approximations to the mean, covariance and cross-covariance are formed as

$$\begin{aligned}
\mathbf{m}_{k+1|k} &= \sum_{i=0}^{2n} W_i^{(m)}\, \mathbf{f}(x_k^{(i)}) \\
\mathbf{P}_{k+1|k} &= \sum_{i=0}^{2n} W_i^{(c)}\, [\mathbf{f}(\mathbf{x}_k^{(i)}) - \mathbf{m}_{k+1|k}] \\
&\quad \times [\mathbf{f}(\mathbf{x}_k^{(i)}) - \mathbf{m}_{k+1|k}]^T + \mathbf{Q}_k \\
\mathbf{C}_{k,k+1} &= \sum_{i=0}^{2n} W_i^{(c)}\, [\mathbf{x}_k^{(i)} - \mathbf{m}_{k|k}]\, [\mathbf{f}(\mathbf{x}_k^{(i)}) - \mathbf{m}_{k+1|k}]^T
\end{aligned}$$

where the sigma-points $\mathbf{x}_k^{(i)}$ are the columns of the Cholesky factor $\mathbf{S}$ of $\mathbf{P}_{k|k}$ such that $\mathbf{P}_{k|k} = \mathbf{S}\, \mathbf{S}^T$:

$$\begin{aligned}
\mathbf{x}_k^{(0)} &= \mathbf{m}_{k|k} \\
\mathbf{x}_k^{(i)} &= \mathbf{m}_{k|k} + [\sqrt{(n+\lambda)}]\mathbf{S}_i, \qquad i = 1, \ldots, n \\
\mathbf{x}_k^{(i)} &= \mathbf{m}_{k|k} - [\sqrt{(n+\lambda)}]\mathbf{S}_{i-n}, \quad i = n+1, \ldots, 2n.
\end{aligned}$$

The constant weights are defined as

$$
\begin{aligned}
W_0^{(m)} &= \lambda/(n+\lambda) \\
W_0^{(c)} &= \lambda/(n+\lambda) + (1-\alpha^2+\beta) \\
W_i^{(m)} &= 1/\{(2(n+\lambda))\}, \quad i = 1,\ldots,2n \\
W_i^{(c)} &= 1/\{(2(n+\lambda))\}, \quad i = 1,\ldots,2n,
\end{aligned}
$$

where the scaling parameter $\lambda$ is defined as $\lambda = \alpha^2(n+\kappa) - n$, and $\alpha$, $\beta$ and $\kappa$ are adjustable parameters of the transformation.

The number of sigma points in UT is always $2n+1$, where $n$ is the state dimensionality and thus in that sense the computational complexity grows linearly in the number of state dimensions.

### 2.4. Gauss-Hermite Approximation

Gauss-Hermite (GH) quadrature is a well known general method for approximating integrals of functions over one-dimensional Gaussian functions. As shown, for example, in [26] the method can be extended for computation of integrals over multidimensional Gaussian distributions. When applied to integrals in Equations (2), with $m$th order GH, we get the following approximations:

$$
\begin{aligned}
\mathbf{m}_{k+1|k} &= \sum_{i=1}^{N} W_i \mathbf{f}(\mathbf{x}_k^{(i)}) \\
\mathbf{P}_{k+1|k} &= \sum_{i=1}^{N} W_i [\mathbf{f}(\mathbf{x}_k^{(i)}) - \mathbf{m}_{k+1|k}] \\
&\qquad \times [\mathbf{f}(\mathbf{x}_k^{(i)}) - \mathbf{m}_{k+1|k}]^T + \mathbf{Q}_k \\
\mathbf{C}_{k,k+1} &= \sum_{i=1}^{N} W_i [\mathbf{x}_k^{(i)} - \mathbf{m}_{k|k}][\mathbf{f}(\mathbf{x}_k^{(i)}) - \mathbf{m}_{k+1|k}]^T,
\end{aligned}
$$

where $\mathbf{x}_k^{(i)}$ are the $m^n$ quadrature points, and $W_i$ are their associated weights, which are formed as products of the one-dimensional Gauss-Hermite weights.

Note that in univariate case when $m = 3$ the number of points is the same as in UT and the methods can indeed be made equal by a suitable choice of UT parameters [26]. When $m > 3$, the number of points is different also in single dimension, but of course, GH is potentially more accurate than UT. The disadvantage of GH is that the number of sigma points grows exponentially with the number of dimensions, whereas the number of sigma points in UT is a linear function of state dimensionality.

### 2.5. Central Difference Approximation

In central difference (CD) method [26] the idea is to fit a quadratic function to the non-linearity such the values at certain preselected points around a central point match exactly. In practice, we form the approximation as follows:

$$
\mathbf{f}(\mathbf{x}_k) \approx \mathbf{f}(\mathbf{m}_{k|k}) + \sum_{i=1}^{n} \tilde{\mathbf{F}}_{s,i}\, \mathbf{s}_{k,i} + \frac{1}{2} \sum_{i=1}^{n} \tilde{\mathbf{F}}_{ss,i}\, \mathbf{s}_{k,i}^2,
$$

where we have made the variable change $\mathbf{x}_k = \mathbf{m}_{k|k} + \mathbf{S}\,\mathbf{s}_k$. The matrices $\tilde{\mathbf{F}}_s$ and $\tilde{\mathbf{F}}_{ss}$ are the central difference approximations for the Jacobian $\mathbf{F}_s = \partial \mathbf{f}(\mathbf{x}_k)/\partial \mathbf{s}$ and diagonal Hessian $\mathbf{F}_{ss} = \partial^2 \mathbf{f}(\mathbf{x}_k)/\partial \mathbf{s}^2$ evaluated at $\mathbf{m}_{k|k}$. The columns are given by

$$
\begin{aligned}
\tilde{\mathbf{F}}_{s,i} &= \frac{\mathbf{f}(\mathbf{m}_{k|k} + h\,\mathbf{S}\,\mathbf{e}_i) - \mathbf{f}(\mathbf{m}_{k|k} - h\,\mathbf{S}\,\mathbf{e}_i)}{2h} \\
\tilde{\mathbf{F}}_{ss,i} &= \frac{\mathbf{f}(\mathbf{m}_{k|k} + h\mathbf{S}\mathbf{e}_i) - 2\mathbf{f}(\mathbf{m}_{k|k}) + \mathbf{f}(\mathbf{m}_{k|k} - h\mathbf{S}\mathbf{e}_i)}{h^2},
\end{aligned}
$$

where $h$ is the step size of the approximation and $\mathbf{e}_i$ the unit vector in the direction of the coordinate axis $i$. Thus, the approximations for the mean, covariance and cross-covariance are given as

$$
\begin{aligned}
\mathbf{m}_{k+1|k} &= \mathbf{f}(\mathbf{m}_{k|k}) + \frac{1}{2} \sum_{i=1}^{n} \tilde{\mathbf{F}}_{ss,i} \\
\mathbf{P}_{k+1|k} &= \sum_{i=1}^{n} \tilde{\mathbf{F}}_{s,i} \tilde{\mathbf{F}}_{s,i}^T + \frac{1}{2} \sum_{i=1}^{n} \tilde{\mathbf{F}}_{ss,i} \tilde{\mathbf{F}}_{ss,i}^T + \mathbf{Q}_k \\
\mathbf{C}_{k,k+1} &= \mathbf{S}\,\tilde{\mathbf{F}}_s^T.
\end{aligned}
$$

Note that central difference approximation coincides with the linearization approximation (see, Section 2.2), if the diagonal second-order correction term is dropped out and the central difference approximations are replaced with exact derivatives.

The number of sigma points required by CD method is also linear function of number of dimensions as in UT. Thus the computational complexity of CD is also approximately the same as of UT.

## 3. SIMULATIONS

### 3.1. Univariate Non-stationary Growth Model

As the first simulation example we consider the univariate non-stationary growth model discussed, for example, in [31, 32, 33], which is particularly challenging for Gaussian approximation based filters and smoothers since it is highly non-linear and bimodal.

The dynamic and measurement models are given by

$$
\begin{aligned}
x_k &= \alpha_d\, x_{k-1} + \beta_d\, \frac{x_{k-1}}{1 + x_{k-1}^2} \\
&\qquad + \gamma_d\, \cos(1.2(k-1)) + q_k \\
y_k &= \frac{x_k^2}{20} + r_k.
\end{aligned}
$$

| Filter | RMSE | Smoother | RMSE |
|---|---|---|---|
| EKF | 10.6 (0.06) | ERTS | 9.33 (0.04) |
| UKF | 7.33 (0.02) | URTS | 7.05 (0.02) |
| GHKF | 7.14 (0.02) | GHRTS | 6.77 (0.02) |
| CDKF | 7.14 (0.02) | CDRTS | 6.77 (0.02) |
| MCKF | 5.89 (0.02) | MCRTS | 4.86 (0.02) |

**Table 1**. Univariate Non-stationary Growth Model: mean RMSE values of tested filters and smoothers over 1000 simulations (standard errors in parentheses).

The parameters of the dynamics are $\alpha_d = 0.5$, $\beta_d = 25$ and $\gamma_d = 8$, and noises are distributed as $q_k \sim \mathrm{N}(0, 1)$ and $r_k \sim \mathrm{N}(0, 1)$. The prior distribution for the state was set to $x_0 \sim \mathrm{N}(0.1, 1)$ and the true state was initialized to $x_0 = 0.1$. The system was simulated 1000 times for 400 time steps, and the following filters and smoothers were tested:

- EKF & ERTS: standard extended Kalman filter and RTS smoother with closed form derivatives.
- UKF & URTS: unscented Kalman filter and unscented RTS smoother with transformation parameters $\alpha = 1$, $\beta = 0$ and $\kappa = 2$ (which corresponds to the suggested parameterization in [20]).
- GHKF & GHRTS: Gauss-Hermite Kalman filter and RTS smoother with $m = 3$.
- CDKF & CDRTS: Central difference Kalman filter and RTS smoother with step size $h = \sqrt{3}$.
- MCKF & MCRTS: Monte Carlo Kalman filter and RTS smoother with $N = 10000$ Monte Carlo samples.

Table 1 lists the root mean squared error (RMSE) values for the tested filters and smoothers, which shows that Monte Carlo approximation works best with this model, and Gauss-Hermite and central difference based methods have equal performance. The UT based results are slightly worse and the errors of EKF/ERTS are much higher than of the others.

### 3.2. Tracking of Maneuvering Target

As the second 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. The same simulation setup was also used in [30], and very similar simulation setup was used in [29] for assessing the performance of the cubature Kalman filter (CKF).

The non-linear dynamic model of the coordinated turn model [1] is

$$x_k = \begin{pmatrix} 1 & \frac{\sin(\omega\Delta t)}{\omega} & 0 & -\left(\frac{1-\cos(\omega\Delta t)}{\omega}\right) & 0 \\ 0 & \cos(\omega\Delta t) & 0 & -\sin(\omega\Delta t) & 0 \\ 0 & \frac{1-\cos(\omega\Delta t)}{\omega} & 1 & \frac{\sin(\omega\Delta t)}{\omega} & 0 \\ 0 & \sin(\omega\Delta t) & 0 & \cos(\omega\Delta t) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} x_{k-1} + q_{k-1}, \tag{3}$$

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 [29].

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

$$\theta_k^i = \arctan\left(\frac{y_k - s_y^i}{x_k - s_x^i}\right) + r_k^i, \tag{4}$$

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. The parameters of the methods were the same as in previous section, except that the UT parameters were $\alpha = 1$, $\beta = 0$ and $\kappa = 3 - n$ (corresponds to [20]).

Table 2 lists the errors 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 (except MCKF). 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 practically the same.

## 4. DISCUSSION

The classification to sigma-point methods in [25] is based on interpreting the methods as special cases of (weighted) statistical linear regression [34], and within this interpretation the unscented transformation, central differences and Gauss-Hermite based methods described in this article can certainly be counted as sigma-point methods, but others exist also. For example, the divided difference filter (DDF) [28] and the cubature Kalman filter (CKF) [29] can be interpreted as sigma-point filters and the corresponding smoothers could be implemented analogously to the smoothers in this article. The Gaussian process approximation based filter

| Method | pos. RMSE | vel. RMSE | $\omega$ RMSE |
|---|---|---|---|
| EKF | 194.7 (9.6) | 184.6 (9.9) | 0.0820 (0.007) |
| UKF | 67.1 (0.4) | 62.3 (0.4) | 0.0281 (0.0001) |
| CDKF | 65.5 (0.3) | 60.5 (0.4) | 0.0281 (0.0001) |
| GHKF | 63.7 (0.3) | 58.0 (0.4) | 0.0280 (0.0001) |
| MCKF | 63.8 (0.3) | 58.1 (0.4) | 0.0280 (0.0001) |
| ERTS | 150.0 (8.2) | 142.8 (8.5) | 0.0766 (0.008) |
| URTS | 41.6 (0.7) | 30.9 (0.9) | 0.0124 (0.0005) |
| CDRTS | 39.8 (0.7) | 29.4 (0.9) | 0.0121 (0.0005) |
| GHRTS | 37.5 (0.7) | 27.8 (0.9) | 0.0119 (0.0005) |
| MCRTS | 37.7 (0.7) | 27.8 (0.9) | 0.0120 (0.0005) |

**Table 2**. Tracking of Maneuvering Target: mean RMSE values of position, velocity and turn rate estimates for the tested filters and smoothers over 1000 simulations (standard errors in parentheses).

presented in [35] could also be included, because it can be interpreted to use the Bayes-Hermite quadrature [36] as the underlying approximation method.

As discussed in [25], statistical linearization is closely related to sigma-point approximations. However, it is important to note that the statistical linear regression [34] which is the basis of sigma-point framework [25] is not exactly equivalent to statistical linearization [19]. In the original statistical linearization [19] the idea is to form statistically linearized approximation

$$\mathbf{f}(\mathbf{x}_k) = \mathbf{b} + \mathbf{A}\,(\mathbf{x}_k - \mathbf{m}_{k|k}),$$

by minimizing the following kind of error function with respect to the parameters $\mathbf{A}$ and $\mathbf{b}$:

$$\begin{aligned}\mathrm{MSE}(\mathbf{b},\mathbf{A}) = \mathrm{E}[&(\mathbf{f}(\mathbf{x}_k) - \mathbf{b} - \mathbf{A}\,(\mathbf{x}_k - \mathbf{m}_{k|k}))^T \\ &\times (\mathbf{f}(\mathbf{x}_k) - \mathbf{b} - \mathbf{A}\,(\mathbf{x}_k - \mathbf{m}_{k|k})],\end{aligned}$$

where the expectation is over $\mathbf{x}_k \sim \mathrm{N}(\mathbf{m}_{k|k}, \mathbf{P}_{k|k})$. The result is:

$$\begin{aligned}\mathbf{b} &= \mathrm{E}[\mathbf{f}(\mathbf{x}_k)] \\ \mathbf{A} &= \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\,(\mathbf{x}_k - \mathbf{m}_{k|k})^T]\,\mathbf{P}_{k|k}^{-1}.\end{aligned}$$

By substituting this into the Equations (2), we get the statistically linearized RTS (SLRTS) smoother

$$\begin{aligned}\mathbf{m}_{k+1|k} &= \mathrm{E}[\mathbf{f}(\mathbf{x}_k)] \\ \mathbf{P}_{k+1|k} &= \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\,(\mathbf{x}_k - \mathbf{m}_{k|k})^T]\,\mathbf{P}_{k|k}^{-1} \\ &\quad \times \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\,(\mathbf{x}_k - \mathbf{m}_{k|k})^T]^T + Q_k \\ \mathbf{C}_{k,k+1} &= \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\,(\mathbf{x}_k - \mathbf{m}_{k|k})^T]^T.\end{aligned}$$

Thus the sigma-point methods presented in the article can be considered as discrete approximations to these equations.

## 5. CONCLUSION

In this article we have shown how the same sigma point approximations that have previously been used in sigma-point Kalman filtering can also be used in optimal smoothing. We have shown how the unscented transformation, Gauss-Hermite quadrature and central differences based smoothers can be derived from the general Gaussian non-linear smoothing equations and demonstrated their performance in simulated applications.

## 6. REFERENCES


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

[2] W. Fong, S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing with application to audio signal enhancement," *IEEE Transactions on Signal Processing*, vol. 50(2), pp. 438–449, February 2002.

[3] Z. Ghahramani and S. T. Roweis, "Learning nonlinear dynamical systems using an EM algorithm," in *Neural Information Processing Systems 11 (NIPS'98)*, 1999, pp. 431–437.

[4] 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.

[5] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-Blackwellized particle filter for multiple target tracking," *Information Fusion Journal*, vol. 8(1), pp. 2–15, 2007.

[6] 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.

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

[8] M. S. Grewal and A. P. Andrews, *Kalman Filtering, Theory and Practice Using MATLAB*, Wiley Interscience, 2001.

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

[10] 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.

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

[12] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communications and Control*, McGraw-Hill Book Company, 1971.

[13] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.

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

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

[16] H. Cox, "On the estimation of state variables and parameters for noisy dynamic systems," *IEEE Transactions on Automatic Control*, vol. 9(1), pp. 5–12, January 1964.

[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] J. S. Meditch, "A survey of data smoothing for linear and nonlinear dynamic systems," *Automatica*, vol. 9, pp. 151–162, 1973.

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

[20] S. J. Julier, J. K. Uhlmann, and Hugh F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proceedings of the 1995 American Control, Conference, Seattle, Washington*, 1995, pp. 1628–1632.

[21] 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.

[22] E. A. Wan and R. van der Merwe, "The unscented Kalman filter," in *Kalman Filtering and Neural Networks*, Simon Haykin, Ed., chapter 7. Wiley, 2001.

[23] S. Särkkä, *Recursive Bayesian Inference on Stochastic Differential Equations*, Doctoral dissertation, Helsinki University of Technology, 2006.

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

[25] R. van der Merwe and E. Wan, "Sigma-point Kalman filters for probabilistic inference in dynamic state-space models," in *Proceedings of the Workshop on Advances in Machine Learning*, 2003.

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

[27] 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.

[28] M. Nørgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36(11), pp. 1627 – 1638, 2000.

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

[30] S. Särkkä and J. Hartikainen, "On Gaussian optimal smoothing of non-linear state space models," *IEEE Transactions on Automatic Control*, 2010, (to appear).

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

[32] G. Kitagawa, "Non-gaussian state-space modeling of nonstationary time series," *Journal of the American Statistical Association*, vol. 82, pp. 1032–1064, 1987.

[33] Y. Wu, D. Hu, M. Wu, and X. Hu, "Unscented Kalman filtering for additive noise case: Augmented versus nonaugmented," *IEEE Signal Processing Letters*, vol. 12(5), pp. 357–360, 2005.

[34] T. Lefebvre, H. Bruyninckx, and J. De Schuller, "Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators" [and authors' reply]," *IEEE Transactions on Automatic Control*, vol. 47(8), pp. 1406–1409, 2002.

[35] M. P. Deisenroth, M. F. Huber, and U. D. Hanebeck, "Analytic moment-based Gaussian process filtering," in *Proceedings of the 26th International Conference on Machine Learning*, 2009.

[36] A. O'Hagan, "Bayes-Hermite quadrature," *Journal of Statistical Planning and Inference*, vol. 29, pp. 245–260, 1991.