Sigma Point-Based Estimation

Eun-Hwan Shin
Applanix Corporation, A Trimble Company

Abstract

Discussed in this paper are the sigma point-based filter and smoother for a generic discrete-time nonlinear system model. The sigma-point filter is treated as an approximation of the generic nonlinear Kalman filter equations, which extends the chapter on nonlinear Kalman filter theory in Gelb (1974). The implementation of the square-root sigma-point filter is addressed in detail with the necessary numerical tools so that it can easily be used for practical use. The Rauch-Tung-Striebel (RTS) formulation of the sigma-point smoother is derived by combining the statistical linear regression and the optimization criteria given in Rauch et al (1965). The notes can be used as a supplementary reading material in an introductory nonlinear estimation course.

Keywords: sigma point, Kalman filter, smoother, nonlinear estimation

1 Introduction

The analysis and prediction of complex dynamic phenomena and nonlinear phenomena have become very important in various fields of research (Kitagawa and Higuchi, 2001). Navigation is a typical field of nonlinear dynamic systems and in the core of navigation system development lies the problem of estimating the states of a dynamic system. When it comes to state estimation for nonlinear systems, however, there is no single solution available that clearly outperforms all other strategies (Nørgaad et al, 2000).

A classical discussion on nonlinear filters can be found in Gelb (1974), where various ways of dealing with the nonlinear filtering problem are introduced. The first approach is based on the truncated Taylor series expansion of the nonlinear system model. The extended Kalman filter (EKF) and a second-order filter

are derived with this approach. The EKF has been a primary choice over the last two decades or more. However, it can suffer from performance degradation under severe nonlinearity. The second-order filter demonstrates an improved performance but it has higher computational complexity. Another approach is to use statistical approximation; for instance, the statistically linearized Kalman filter (SLKF). Generally the latter approach can achieve higher accuracy but a drawback is that the probability density function (PDF) should be known for the evaluation of the statistical expectation. Therefore, the latter approach could hardly find its applications.

The unscented Kalman filter (UKF) has become an important option since the publication of Julier et al (1995). The UKF is based on the approximation of the PDF using deterministically chosen samples. The UKF has been applied in many fields including the low-cost inertial navigation (Shin, 2005), space attitude estimation (Crassidis and Markley, 2003), stock volatility estimation (Zoeter et al, 2004) and etc. The development of the UKF also inspired other researches. Nørgaad et al (2000) derived the second-order divided difference (DD2) filter based on the central divided difference and showed its similarity to the UKF. Lefebvre et al (2002) argued that the UKF is a special case of the linear regression Kalman filter. Thus, the sigma point Kalman filter (SPKF) has become the term to encompass those filtering methods that are based on deterministically chosen samples.

Although numerous papers on sigma point-based filtering methods have been published, sigma point-based smoothing methods have not been investigated much until recently. Shin (2005) used the forward/backward filtering approach. In some applications, however, the forward system model may not be invertible. In this case, the RTS formulation of the sigma point-based smoothing algorithm is required. Further, the RTS smoother has been the primary choice owing to its simplicity in the implementation. The SLKF or statistical linear regression Kalman filter (SLRKF) may not be suitable for practical use but they have become useful tools to find variants of sigma point-based estimation methods. Using the statistical

linearization technique, Vercauteren and Wang (2005) investigated decentralized sigma-point information filters and Terejanu et al (2008) derived the sigma-point RTS smoother.

Another important issue in practical system development is the square-root filtering. Square-root filters can further improve the numerical quality of the covariance update by propagating the square-root of the state error covariance, S defined as $P = SS^T$ where P is the state error covariance matrix and $\cdot^T$ denotes the transpose of a matrix. In 1970s, the UD-factorization algorithm was introduced due to the heavy computational burden of the square-root implementation; it factors P into an upper-triangular matrix U with 1's along its main diagonal and a diagonal matrix D: $P = UDU^T$ (Haykin, 1996, p 327). With advances in the modern computer technology, however, computational cost is not as serious a factor as it used to be. Further, a Kalman filter using the UD-factorization may suffer from overflow/underflow problems (Stewart and Chapman, 1990). Thus, the square-root implementation will be preferred for utmost quality in the covariance update (Haykin 1996, p. 328}. Although van der Merwe and Wan (2001) discussed that the square-root sigma-point filter was more efficient than the covariance formulation, most of the papers published so far are based on the covariance formulation.

The purpose of this paper is to provide students with a practical introduction on the sigma-point based filter and smoother. The latest developments in this field will be merged into the frame work of classical ones. Mathematical preliminaries are given in Section 2. The generic nonlinear Kalman filter is discussed in Section 3 based on Gelb (1974). Sections 4 to 6 deal with the derivation of the EKF, the SPKF and the SLRKF from the generic Kalman filter, which clarifies the approximations made for each of these filters. The sigma-point RTS smoother is derived in Section 7 based on the statistical linear regression and the optimality criteria given by Rauch et al (1965) using the Gaussian density assumption. With this approach, the derivation of the sigma-point RTS becomes very simple. Finally, some concluding remarks are given in Section 8.

2 Mathematical preliminaries

Basic definitions and theorems required to understand the rest of the paper are summarized here. Three important numerical tools are discussed as well: Givens rotations, Cholesky rank-1 modification methods, and statistical linear regression. The first two are required in building a square-root sigma-point filter

and the third in deriving the sigma-point RTS smoother.

Definition 1. For a random variable, x, the expected value is defined as

$$\hat{x} = E[x] = \int_x xp(x)dx , \qquad (1)$$

where $p(\cdot)$ denotes the probability density function.

Definition 2. The Mahalanobis distance squared for a random vector $\delta x = x - \hat{x}$ with the covariance P is defined as follows:

$$\left\| \delta x \right\|_{P^{-1}}^2 = \delta x^T P^{-1} \delta x , \qquad (2)$$

where $\cdot^{-1}$ denotes the inverse of a matrix.

Definition 3. For an $n{\times}n$ matrix, A, the trace is defined as

$$tr(A) = \sum_{i=1}^{n} a_{ii} , \qquad (3)$$

where a_{ii} is the ith diagonal element of A.

Definition 4. A matrix, A, is orthogonal if $AA^T{=}I$, where I is an identity matrix. Hence, $A^T{=}A^{-1}$.

Theorem 1. Let A be an $n{\times}r$ and B an $r{\times}n$ matrix. Then, $tr(AB) = tr(BA)$ (Koch, 1988, p. 47).

Theorem 2. Let x be an $n{\times}1$ vector and A a symmetrical $n{\times}n$ matrix. Then, $\partial(x^T Ax)/\partial x = 2Ax$ (Koch, 1988, p. 82).

Theorem 3. For a square matrix A, $tr(A){=}tr(A^T)$.

Theorem 4. Let A be an $m{\times}n$ and B an $n{\times}m$ matrix. Then, $\partial tr(AB)/\partial A = B^T$. Using this and Theorem 3, one can show that $\partial tr(BA^T)/\partial A = B$. Further,

$$\partial tr(ABA^T)/\partial A = A(B + B^T) .$$

Lemma 1. The matrix inversion lemma, also called the Woodbury matrix identity, is a useful tool in the linear systems analysis:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}DA^{-1}B)^{-1}DA^{-1}, (4)$$

where A, B, C and D are matrices of appropriate sizes.

2.1 Minimum-variance estimation

The minimum-variance estimator is an estimator that seeks to find the optimal state, $\hat{x}$, such that the following functional is minimized using the measurement z (Gelb, 1974, p. 104):

$$J(\hat{x}) = \int_x (\hat{x} - x)(\hat{x} - x)^T p(x \mid z) dx, \tag{5}$$

where $p(x|z)$ denotes the PDF of x conditioned on z. The solution to this problem is found from:

$$\hat{x} = \int_x x p(x \mid z) dx = E[x \mid z], \tag{6}$$

which is the conditional mean estimate.

2.2 Triangularization by Givens rotations

Matrix triangularization methods, often called QR-factorization methods, were originally developed to solve least-squares problems and became essential tools in building a square-root Kalman filter (Grewal and Andrews, 2001, p. 229). Only Givens rotations will be introduced here for brevity; refer to Grewal and Andrews (2001) for other triangularization methods.

Let $A=\{a_{ij}\}$ be an $n \times (r+n)$ matrix. Then, a $(r+n) \times (r+n)$ orthogonal matrix Θ can be post-multiplied to yield

$$A\Theta = \begin{bmatrix} 0 & S \end{bmatrix}, \tag{7}$$

where S is an $n \times n$ upper triangular matrix. A in this setting is often called the prearray. Θ can be obtained by successive Givens rotations:

$$\Theta = \prod_{\substack{i=n,n-1,\dots,1 \\ j=1,\dots,r+i-1}} \Theta_{ij}, \tag{8}$$

where Θ_{ij} is the Givens rotation matrix that annihilates a_{ij} for $i = n, n-1, \dots, 1$ and $j = 1, 2, \dots, r+i-1$. Thus, Θ_{ij} will take the following form; all diagonal elements are one except for $\theta_{jj} = \theta_{r+i,r+i} = \cos\theta$ and all off-diagonal elements are zero except for $\theta_{j,r+i} = -\theta_{r+i,j} = \sin\theta$, where θ_{km} is the kmth element of Θ_{ij}. Hence, we can solve

$$a_{ij} \cos\theta - a_{i,r+i} \sin\theta = 0, \tag{9}$$

$$\sin^2\theta + \cos^2\theta = 1, \tag{10}$$

to yield

$$\sin\theta = a_{ij} / \sqrt{a_{i,r+i}^2 + a_{ij}^2}, \tag{11}$$

$$\cos\theta = a_{i,r+i} / \sqrt{a_{i,r+i}^2 + a_{ij}^2}, \tag{12}$$

As only the jth and the $(r+i)$th column of the prearray are affected by the multiplication of Θ_{ij}, an efficient in-place triangularization routine can be written. The following codes, written in GNU Octave script, can be used for the in-place triangularization, a corrected version of the script in Grewal and Andrews (2001, p. 234):

```
for i = n:-1:1,
    for j = 1:r+i-1,
        rho = sqrt(A(i,r+i)^2 + A(i,j)^2);
        s = A(i,j)/rho;
        c = A(i,r+i)/rho;
        for k = 1:i,
            x = c * A(k,j) - s * A(k,r+i);
            A(k,r+i) = s * A(k,j) + c * A(k,r+i);
            A(k,j) = x;
        endfor
    endfor
endfor
```

2.3 Cholesky rank-1 modifications

Let A be an $n \times n$ positive definite matrix. Then, A has a Cholesky factor, S, such that $SS^T = A$, where S is an upper-triangular matrix. The rank-one update problem is to obtain an updated Cholesky factor, $\tilde{S}$, such that

$$\tilde{S}\tilde{S}^T = SS^T + xx^T, \tag{13}$$

where x is an $n \times 1$ vector. The solution is obtained in the following way:

$$\begin{bmatrix} x & S \end{bmatrix} \Theta = \begin{bmatrix} 0 & \tilde{S} \end{bmatrix}, \tag{14}$$

where Θ is an orthogonal matrix annihilating the first column of the pre-array, which can be computed by successive Givens rotations. The rank-one downdate problem is to obtain a downdated Cholesky factor, $\tilde{S}$, such that:

$$\tilde{S}\tilde{S}^T = SS^T - xx^T, \tag{15}$$

Let a be the solution vector of the linear system $Sa = x$ and let

$$\alpha = \sqrt{1 - a^T a}. \tag{16}$$

Then the solution of the rank-one downdate problem can be obtained as follows (Dongarra et al, 1979):

$$\begin{bmatrix} \alpha & a^T \\ 0 & S \end{bmatrix} \Theta = \begin{bmatrix} 1 & 0 \\ x & \tilde{S} \end{bmatrix}, \tag{17}$$

where Θ is an orthogonal matrix that annihilates the (1,2)-entry of the pre-array.

2.4 Statistical linear regression

A nonlinear vector function $f(x)$ of a random vector variable, x, can be linearized as follows:

$$f(x) = Ax + b + e, \tag{18}$$

where A and b are respectively a matrix and a vector to be determined and e is the linearization error. If we want to have an unbiased linearization,

$$\hat{e} = \hat{f}(x) - A\hat{x} - b = 0. \tag{19}$$

Thus,

$$b = \hat{f}(x) - A\hat{x} \tag{20}$$

and substituting the above into Eq. (18) yields

$$e = A\delta x - \delta f, \tag{21}$$

where $\delta x = \hat{x} - x$ and $\delta f = \hat{f}(x) - f(x)$. Then, the weighted least squares solution of A can be obtained by minimizing the following cost function:

$$J(A) = E[e^T S e], \tag{22}$$

where S is an arbitrary positive semi-definite matrix. Thus, substituting Eq. (21) into the following

$$\begin{aligned} \frac{\partial J(A)}{\partial A} &= \frac{\partial}{\partial A} E[tr(e^T S e)] \\ &= \frac{\partial}{\partial A} E[tr(See^T)] = 0 \end{aligned} \tag{23}$$

and using the theorems in Section 2, we can write

$$A = P_{fx} P_{xx}^{-1}, \tag{24}$$

where $P_{xx} = E[\delta x \delta x^T]$ and $P_{fx} = E[\delta f \delta x^T]$. Therefore, the linearization error covariance can be written as

$$P_{ee} = E[ee^T] = P_{ff} - A P_{xx} A^T, \tag{25}$$

where $P_{ff} = E[\delta f \delta f^T]$.

3 Generic nonlinear Kalman filter

Let the discrete-time nonlinear system transition model be given as follows:

$$x_k = f(x_{k-1}, k-1) + G_k w_k, \tag{26}$$

where $f(\cdot)$ is a nonlinear vector function; x_k, w_k and G_k are the system state vector, the system noise vector and the noise input mapping matrix at time t_k, respectively. It is assumed that $E[w_k] = 0$, $E[x_i w_k^T] = 0$ for all i, k and $E[w_i w_k^T] = Q_k \delta_{ik}$, where δ_{ik} is the Kronecker delta function. The system is observed by the following nonlinear measurement model:

$$z_k = h(x_k) + v_k, \tag{27}$$

where $h(\cdot)$ is a nonlinear vector function; z_k is the measurement vector, v_k is the measurement noise vector with $E[v_k] = 0$, $E[v_i v_k^T] = R_k \delta_{ik}$ and $E[x_k v_i^T] = 0$ for all i, k. Eqs. (26) and (27) can be applied to various real-world problems.

The Kalman filter is a minimum variance estimator interested in propagating the first two moments of the PDF, mean and covariance, recursively. Let us assume that the estimates at time t_{k-1} are given as follows:

$$\hat{x}_{k-1|k-1} = E[x_{k-1} \mid z_1, ..., z_{k-1}], \tag{28}$$

$$P_{k-1|k-1} = E[\delta x_{k-1|k-1} \delta x_{k-1|k-1}^T], \tag{29}$$

$$\delta x_{k-1|k-1} = \hat{x}_{k-1|k-1} - x_{k-1}. \tag{30}$$

The prediction stage, also called the time update, extrapolates the state to the next time epoch using the system transition model, Eq. (26):

$$\hat{x}_{k|k-1} = E[f(x_{k-1}, k-1) \mid z_1, ..., z_{k-1}], \tag{31}$$

$$P_{k|k-1} = E[\delta x_{k|k-1} \delta x_{k|k-1}^T], \tag{32}$$

$$\delta x_{k|k-1} = \hat{x}_{k|k-1} - x_k. \tag{33}$$

Then the two moments are updated using the measurements taken at time t_k. The update equations will be obtained following Gelb (1974, p. 185-187). Let the updated state vector be written as a linear function of measurements:

$$\hat{x}_{k|k} = a_k + K_k z_k, \tag{34}$$

where a_k and K_k are a vector and the filter gain matrix, respectively, to be determined. Substituting Eqs. (33), (30) and (27) into Eq. (34) yields

$$\delta x_{k|k} = \delta x_{k|k-1} - \hat{x}_{k|k-1} + a_k + K_k[h(x_k) + v_k], \tag{35}$$

As we want to have an unbiased estimate, $E[\delta x_{k|k}] = E[\delta x_{k|k-1}] = 0$. Hence,

$$a_k = \hat{x}_{k|k-1} - K_k \hat{h}(x_k), \tag{36}$$

which when substituted into Eq. (34) yields

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - \hat{h}(x_k)]. \tag{37}$$

Also, the estimation error can be written as

$$\delta\tilde{x}_{k|k} = \delta\tilde{x}_{k|k-1} + K_k[h(x_k) - \hat{h}(x_k)] + K_k v_k. \tag{38}$$

Assuming that $P_{k|k}$ is independent of z_k, we obtain

$$\begin{aligned}
P_{k|k} &= P_{k|k-1} + K_k[\delta h_k \delta h_k^T]K_k^T \\
&\quad - E[\delta x_{k|k-1}\delta h_k^T]K_k^T - K_k E[\delta h_k \delta x_{k|k-1}^T] \\
&\quad + K_k R_k K_k^T
\end{aligned} \tag{39}$$

where $\delta h_k = \hat{h}(x_k) - h(x_k)$. The minimum variance estimate is obtained by minimizing the cost function $J(K_k) = tr(P_{k|k})$. Thus, solving $\partial J(K_k)/\partial K_k = 0$ yields the optimal filter gain matrix:

$$K_k = E[\delta x_{k|k-1}\delta h_k^T]\left\{E[\delta h_k \delta h_k^T] + R_k\right\}^{-1}. \tag{40}$$

By substituting Eq. (40) into Eq. (39), the updated error covariance can be written as follows:

$$\begin{aligned}
P_{k|k} &= P_{k|k-1} - K_k E[\delta h_k \delta x_{k|k-1}^T] \\
&= P_{k|k-1} - K_k \left\{E[\delta h_k \delta h_k^T] + R_k\right\}K_k^T.
\end{aligned} \tag{41}$$

4 Extended Kalman filter

The EKF is based on the approximation of the nonlinear function. By applying the Taylor series expansion and taking up to the first order terms, the nonlinear function in Eq. (26) can be written as follows:

$$f(x_{k-1}, k-1) \approx f(\hat{x}_{k-1|k-1}, k-1) - \Phi_{k|k-1}\delta x_{k|k-1}, \tag{42}$$

$$\Phi_{k|k-1} = \left.\frac{\partial f(x_{k-1}, k-1)}{\partial x_{k-1}}\right|_{x_{k-1}=\hat{x}_{k-1|k-1}}. \tag{43}$$

The prediction equations of the EKF can be obtained by substituting Eq. (42) into Eqs. (31) and (32) as follows:

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, k-1), \tag{44}$$

$$P_{k|k-1} = \Phi_{k|k-1}P_{k-1|k-1}\Phi_{k|k-1}^T + G_k Q_k G_k^T. \tag{45}$$

Similarly, the nonlinear function $h(x_k)$ in Eq. (27) can be approximated as follows:

$$h(x_k) \approx h(\hat{x}_{k|k-1}) - H_k \delta x_{k|k-1}, \tag{46}$$

$$H_k = \left.\frac{\partial h(x_k)}{\partial x_k}\right|_{x_k=\hat{x}_{k|k-1}}. \tag{47}$$

The update equations are obtained by substituting Eq. (46) to Eqs. (37), (40) and (41) as follows:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - h(\hat{x}_{k|k-1})], \tag{48}$$

$$K_k = P_{k|k-1}H_k^T(H_k P_{k|k-1}H_k^T + R_k)^{-1}, \tag{49}$$

$$P_{k|k} = (I - K_k H_k)P_{k|k-1}. \tag{50}$$

5. Sigma-point Kalman filter

In a SPKF, the PDF is approximated by a set of sigma points, χ_i's, and their associated weights, w_i's. As a Gaussian PDF can completely be described by the mean and the covariance, the sigma points are selected such that the following conditions hold:

$$\sum_{i=0}^{L-1} w_i = 1, \tag{51}$$

$$\sum_{i=0}^{L-1} w_i \chi_i = \hat{x}, \tag{52}$$

$$\sum_{i=0}^{L-1} w_i(\chi_i - \hat{x})(\chi_i - \hat{x})^T = P, \tag{53}$$

where L is the number of sigma points.

Let $\chi_{k-1|k-1,i}$'s be the sigma points computed from the updated states, $\hat{x}_{k-1|k-1}$, and the updated covariance, $P_{k-1|k-1}$. Each of the sigma points can be transformed through Eq. (26) as follows:

$$\chi_{k|k-1,i} = f(\chi_{k-1|k-1,i}, k-1). \tag{54}$$

Then, Eqs. (31) and (32) can be approximated as follows:

$$\hat{x}_{k|k-1} = \sum_{i=0}^{L-1} w_i \chi_{k|k-1,i}, \tag{55}$$

$$P_{k|k-1} = \sum_{i=0}^{L-1} w_i \Delta\chi_{k|k-1,i}\Delta\chi_{k|k-1,i}^T, \tag{56}$$

where $\Delta\chi_{k|k-1,i} = \chi_{k|k-1,i} - \hat{x}_{k|k-1}$. Similarly, $\chi_{k|k-1,i}$ can be transformed through the measurement function, Eq. (27) as follows:

$$Z_{k,i} = h(\chi_{k|k-1,i}). \tag{57}$$

Then, the following approximations can be applied:

$$E[h(x_k)] \approx \hat{z}_{k|k-1} = \sum_{i=0}^{L-1} w_i Z_{k,i}, \tag{58}$$

$$E[\delta x_{k|k-1}\delta h_k] \approx P_k^{xh} = \sum_{i=0}^{L-1} w_i \Delta\chi_{k|k-1,i}\Delta Z_{k,i}, \tag{59}$$

$$E[\delta h_k \delta h_k^T] \approx P_k^{hh} = \sum_{i=0}^{L-1} w_i \Delta Z_{k,i}\Delta Z_{k,i}^T, \tag{60}$$

where $\Delta Z_{k,i} = Z_{k,i} - \hat{z}_{k|k-1}$. Substituting the above into Eqs. (37), (40), and (41) yields the sigma point Kalman filter update equations:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - \hat{z}_{k|k-1}), \tag{61}$$

$$K_k = P_k^{xh}(P_k^{hh} + R_k)^{-1}, \tag{62}$$

$$P_{k|k} = P_{k|k-1} - K_k(P_k^{hh} + R_k)K_k^T, \tag{63}$$

5.1 Unscented Kalman filter
If the sigma points are obtained by the unscented transformation (UT), the resulting Kalman filter is called the Unscented Kalman Filter (UKF). The first proposal of the UT done by Julier and Uhlmann (1996) yielded $L = 2n + 1$ sigma points, where n is the number of the states:

$$\chi_i = \begin{cases} \hat{x} & i = 0 \\ \hat{x} + \sqrt{n+\kappa}\sigma_i & i = 1,...,n \\ \hat{x} - \sqrt{n+\kappa}\sigma_{i-n} & i = n+1,...,2n \end{cases} \tag{64a}$$

$$w_i = \begin{cases} \kappa/(n+\kappa) & i = 0 \\ 1/2(n+\kappa) & i = 1,...,2n \end{cases} \tag{64b}$$

where κ is a scaling parameter for the fourth and higher moments of the PDF; and σ_i is the ith column of the square-root covariance matrix, S, that can be computed through the Cholesky decomposition:

$$SS^T = P. \tag{65}$$

κ is usually set to 3-n and thus $w_0 < 0$ if $n > 3$.

5.2 Square-root SPKF
The covariance formulation has the risk of overflow or underflow due to the finite numerical precision and it also requires repeated computation of the Cholesky decomposition, of which the computational cost is very high. These problems can be avoided if the square-root

covariance matrix is propagated by the Kalman filter directly. Let us first discuss the Kalman prediction of the square-root covariance matrix. If $w_i \geq 0$ for all $i = 0, 1, ..., L-1$, Eq. (56) can be rewritten as follows:

$$P_{k|k-1} = [s_0 \quad s_1 \quad ... \quad s_{L-1} \quad G_k Q_k^{1/2}]$$
$$\times [s_0 \quad s_1 \quad ... \quad s_{L-1} \quad G_k Q_k^{1/2}]^T \tag{66}$$

where $Q_k^{1/2}Q_k^{T/2} = Q_k$ and $s_i = \sqrt{w_i}\Delta\chi_{k|k-1,i}$ for $i = 0, 1, ..., L-1$. Hence, $S_{k|k-1}$ can be obtained as follows:

$$[s_0 \quad s_1 \quad ... \quad s_{L-1} \quad G_k Q_k^{1/2}]\Theta_{1k} = [0 \quad S_{k|k-1}] \tag{67}$$

where Θ_{1k} is an orthogonal matrix that annihilates the first L columns of the pre-array. If $w_0 < 0$, the following computation is executed first:

$$[s_1 \quad ... \quad s_{L-1} \quad G_k Q_k^{1/2}]\tilde{\Theta}_{1k} = [0 \quad \tilde{S}_{k|k-1}] \tag{68}$$

where $\tilde{\Theta}_{1k}$ is an orthogonal matrix that annihilates the first L-1 columns of the pre-array. Then, the Cholesky rank-1 downdate algorithm, discussed in Section 2.6, is applied:

$$S_{k|k-1}S_{k|k-1}^T = \tilde{S}_{k|k-1}\tilde{S}_{k|k-1}^T - s_0 s_0^T, \tag{69}$$

where $s_0 = \sqrt{|w_0|}(\chi_{k|k-1,0} - \hat{x}_{k|k-1})$.

As the same strategy can be applied to the Kalman update equations, for brevity we will just discuss only about the case $w_i > 0$ for all $i = 0, 1, ..., L$-1. Writing square-root measurement update equations for the case of having $w_i < 0$ is left as an exercise for students. Let S_k^{hh} be an upper triangular matrix such that $S_k^{hh}(S_k^{hh})^T = P_k^{hh}$. Then, S_k^{hh} can be computed as follows:

$$[\varsigma_0 \quad \varsigma_1 \quad ... \quad \varsigma_{L-1}]\Theta_{2k} = [0 \quad S_k^{hh}], \tag{70}$$

where $\varsigma_i = \sqrt{w_i}(Z_{k,i} - \hat{z}_{k|k-1})$ and Θ_{2k} is an orthogonal matrix that annihilates the first L-n columns of the pre-array. Let U_k and $R_k^{1/2}$ be upper triangular matrices such that $R_k^{1/2}R_k^{T/2} = R_k$ and $U_k U_k^T = P_k^{hh} + R_k$. Then, U_k can be computed as follows:

$$[R_k^{1/2} \quad S_k^{hh}]\Theta_{3k} = [0 \quad U_k], \tag{71}$$

where Θ_{3k} is an orthogonal matrix that annihilates the (1,1) entry of the pre-array. The Kalman gain matrix in Eq. (62) is now computed as follows:

$$K_k = P_k^{xh} U_k^{-T} U_k^{-1}. \tag{72}$$

Finally, the updated square-root covariance matrix, $S_{k|k}$, can be obtained by successive application of the Cholesky rank-1 downdate algorithm as follows:

$$S_{k|k}^i (S_{k|k}^i)^T = S_{k|k}^{i-1} (S_{k|k}^{i-1})^T - \eta_i \eta_i^T, i = 1,...,n, \tag{73}$$

where η_i denotes the *i*th column of $K_k U_k$; $S_{k|k}^0$ is equivalent to $S_{k|k-1}$ and $S_{k|k}^n$ becomes $S_{k|k}$.

6 Statistical linear regression Kalman filter

The results in Section 2.4 shall be used to develop the SLRKF. Considering Eq. (18), the nonlinear function in the system transition model, Eq. (26), can be statistically linearized as follows:

$$f(x_{k-1}, k-1) = A_{f_k} x_{k-1} + b_{f_k} + e_{f_k}, \tag{74}$$

where A_{f_k} and B_{f_k} are parameters to be determined; and e_{f_k} is the linearization error. Given the measurements available up to time t_{k-1}, $\hat{x}_{k-1|k-1}$ and $P_{k-1|k-1}$ can be substituted into Eqs. (24), (20) and (25) to yield

$$A_{f_k} = E[\delta f_{k|k-1} \delta x_{k-1|k-1}^T] P_{k-1|k-1}^{-1}, \tag{75}$$

$$b_{f_k} = \hat{f}_{k|k-1} - A_{f_k} \hat{x}_{k-1|k-1}, \tag{76}$$

$$P_{ef_k} = E[e_{f_k} e_{f_k}^T] = P_{f_k} - A_{f_k} P_{k-1|k-1} A_{f_k}^T, \tag{77}$$

where

$$\hat{f}_{k|k-1} = E[f(x_{k-1}, k-1) \mid z_1,...,z_{k-1}],$$

$$\delta f_{k|k-1} = \hat{f}_{k|k-1} - f(x_{k-1}, k-1),$$

$$P_{f_k} = E[\delta f_{k|k-1} \delta f_{k|k-1}^T].$$

Thus, the prediction stage can be written as follows:

$$\hat{x}_{k|k-1} = \hat{f}_{k|k-1} = A_{f_k} \hat{x}_{k-1|k-1} + b_{f_k}, \tag{78}$$

$$\begin{aligned} P_{k|k-1} &= P_{f_k} + G_k Q_k G_k^T \\ &= A_{f_k} P_{k-1|k-1} A_{f_k}^T + P_{ef_k} + G_k Q_k G_k^T \end{aligned}. \tag{79}$$

Similarly, the measurement function $h(x_k)$ in Eq. (27) can be statistically linearized as follows:

$$h(x_k) = A_{h_k} x_k + b_{h_k} + e_{h_k}, \tag{80}$$

where e_{h_k} is also the linearization error and using the measurements available up to time *tk*-1 and with $\hat{h}_{k|k-1} = E[h(x_k) \mid z_1,...,z_{k-1}]$

$$A_{h_k} = E[\delta h_{k|k-1} \delta x_{k|k-1}^T] P_{k|k-1}^{-1}, \tag{81}$$

$$b_{h_k} = \hat{h}_{k|k-1} - A_{h_k} \hat{x}_{k|k-1}, \tag{82}$$

$$\begin{aligned} P_{eh_k} &= E[e_{h_k} e_{h_k}^T] \\ &= E[\delta h_{k|k-1} \delta h_{k|k-1}^T] - A_{h_k} P_{k|k-1} A_{h_k}^T \end{aligned}, \tag{83}$$

$$\delta h_{k|k-1} = \hat{h}_{k|k-1} - h(x_k) = A_{h_k} \delta x_{k|k-1} - e_{h_k}. \tag{84}$$

Substituting Eqs. (81) through (84) into Eqs. (37), (40) and (41) yields

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - A_{h_k} \hat{x}_{k|k-1} - b_{h_k}), \tag{85}$$

$$K_k = P_{k|k-1} A_{h_k}^T (A_{h_k} P_{k|k-1} A_{h_k}^T + P_{eh_k} + R_k)^{-1} \tag{86}$$

$$P_{k|k} = P_{k|k-1} - K_k(A_{h_k} P_{k|k-1} A_{h_k}^T + P_{eh_k} + R_k)K_k^T. \tag{87}$$

One can easily verify that the statistical linear regression Kalman filter is equivalent to the sigma point Kalman filter if the statistical expectation operator is replaced by the weighted sum of sigma points.

7 Sigma-point RTS smoother

Given the smoothed state at time t_k, $\hat{x}_{k|N}$, the smoothed state at time t_{k-1}, $\hat{x}_{k-1|N}$, can be obtained by minimizing the following combined Mahalanobis distance squared:

$$\begin{aligned} J(x_{k-1}) &= \left\| \hat{x}_{k|N} - A_{f_k} x_{k-1} - b_{f_k} \right\|_{\tilde{Q}_k^{-1}}^2 \\ &\quad + \left\| \hat{x}_{k-1|k-1} - x_{k-1} \right\|_{P_{k-1|k-1}^{-1}}^2 \end{aligned} \tag{88}$$

$$\begin{aligned} \tilde{Q}_k &= G_k Q_k G_k^T + P_{ef_k} \\ &= P_{k|k-1} - A_{f_k} P_{k-1|k-1} A_{f_k}^T \end{aligned} \tag{89}$$

which states that one-step prediction using $\hat{x}_{k-1|N}$ should agree with $\hat{x}_{k|N}$ within the uncertainty in the

system transition model and at the same time $\hat{x}_{k-1|N}$ should agree with $\hat{x}_{k-1|k-1}$ within the uncertainty of the forward updated solution. Eq. (88) is analogous to the result derived in Rauch et al. (1965) based upon the maximum-likelihood criteria.

Thus, $\hat{x}_{k-1|N}$ is the solution of $\partial J(x_{k-1})/\partial x_{k-1} = 0$:

$$
\begin{aligned}
0 &= A_{f_k}^T \tilde{Q}_k^{-1}(A_{f_k} x_{k-1} + b_{f_k} - \hat{x}_{k|N}) \\
&+ P_{k-1|k-1}^{-1}(x_{k-1} - \hat{x}_{k-1|k-1})
\end{aligned} \quad (90)
$$

So, using the above and Eq. (76), we can write the solution as follows:

$$
\begin{aligned}
\hat{x}_{k-1|N} &= (A_{f_k}^T \tilde{Q}_k^{-1} A_{f_k} + P_{k-1|k-1}^{-1})^{-1} \\
&\quad \times \left[P_{k-1|k-1}^{-1} \hat{x}_{k-1|k-1} + A_{f_k}^T \tilde{Q}_k^{-1}(\hat{x}_{k|N} - b_{f_k}) \right] \quad (91) \\
&= \hat{x}_{k-1|k-1} + K_{k-1}^s (\hat{x}_{k|N} - \hat{x}_{k|k-1})
\end{aligned}
$$

$$
K_{k-1}^s = (A_{f_k}^T \tilde{Q}_k^{-1} A_{f_k} + P_{k-1|k-1}^{-1})^{-1} A_{f_k}^T \tilde{Q}_k^{-1} \quad (92)
$$

where K_{k-1}^s is called the smoothing gain matrix.

Using the matrix inversion lemma and Eq. (89), we can write

$$
\begin{aligned}
(A_{f_k}^T \tilde{Q}_k^{-1} A_{f_k} + P_{k-1|k-1}^{-1})^{-1} &= P_{k-1|k-1} \\
&- P_{k-1|k-1} A_{f_k}^T (\tilde{Q}_k + A_{f_k} P_{k-1|k-1} A_{f_k}^T)^{-1} A_{f_k} P_{k-1|k-1} \quad (93) \\
&= P_{k-1|k-1} - P_{k-1|k-1} A_{f_k}^T P_{k|k-1}^{-1} A_{f_k} P_{k-1|k-1}
\end{aligned}
$$

Substituting the above into Eq. (92) and using Eq. (75), the smoothing gain matrix can be rewritten as follows:

$$
\begin{aligned}
K_{k-1}^s &= P_{k-1|k-1} A_{f_k}^T P_{k|k-1}^{-1}(P_{k|k-1} - A_{f_k} P_{k-1|k-1} A_{f_k}^T) \tilde{Q}_k^{-1} \\
&= P_{k-1|k-1} A_{f_k}^T P_{k|k-1}^{-1} \quad (94) \\
&= E[\delta x_{k-1|k-1} \delta f_{k|k-1}^T] P_{k|k-1}^{-1}
\end{aligned}
$$

If $E[\delta x_{k-1|k-1} \delta f_{k|k-1}^T]$ is evaluated using the sigma points, the resulting solution becomes that of the sigma point smoother:

$$
E[\delta x_{k-1|k-1} \delta f_{k|k-1}^T] = \sum_{i=0}^{L-1} w_i \Delta \chi_{k-1|k-1,i} \Delta \chi_{k|k-1,i}^T . \quad (95)
$$

Finally, the smoothed covariance is computed as follows:

$$
P_{k-1|N} = P_{k-1|k-1} + K_{k-1}^s [P_{k|N} - P_{k|k-1}](K_{k-1}^s)^T . \quad (96)
$$

8 Remarks

Latest developments in sigma point filter and smoother theories are summarized in this paper. It may be clearly identified that all the Kalman filters discussed here try to linearize the given nonlinear system model; the difference is in how they linearize. All the numerical building blocks are discussed in detail so that the readers can readily apply the theory to any practical examples; especially, the square-root formulation is emphasized. It is highly recommended for the students to investigate how the filters and smoothers work by simulations.

For instance, a second-order system model is taken from Grewal and Andrews (2001, p. 185). Let $x_k^T = [x_1(t_k) \quad x_2(t_k) \quad x_3(t_k)]$ be the state vector and $z_k = z(t_k)$ the measurement. x_1, x_2, and x_3 are the displacement, the rate, and the damping coefficient, respectively. Then, the system and measurement model are given as follows:

$$
x_1^k = x_1^{k-1} + \Delta t_k x_2^{k-1}
$$

$$
x_2^k = -25 x_1^{k-1} + (1 - 10 \Delta t_k x_3^{k-1}) x_2^{k-1} + (12 + w_k) \Delta t_k
$$

$$
x_3^k = x_3^{k-1}
$$

$$
z_k = x_1^k + v_k
$$

where $x_i^k = x_i(t_k)$ and $\Delta t_k = t_k - t_{k-1}$ is the time interval. The initial values and the noise statistics are given as follows: $\hat{x}_{0|0} = 0$, $P_{0|0} = 2I$, $Q_k = 4.47$, $R_k = 0.01$. For the given models, do the followings:

- Generate the true values with $x_3 = 0.1$ and $w_k = v_k = 0$.

- Generate the noise w_k and v_k following the given statistics.

- Build and run the EKF, the UKF, and the sigma-point RTS smoother for the system and measurement model with the noise added.

- Compare the estimation errors and the covariance outputs from the filters and the smoother.

- Repeat the same experiments by gradually increasing the errors in the initial estimates.

References

Crassidis, J.L. and Markley, F.L. (2003) *Unscented filtering for spacecraft attitude estimation,* Journal of Guidance, Control, and Dynamics, 26(4):536-542.

Dongarra, J.J., Moler, C.B., Bunch, J.R. and Stewart, G.W. (1979) *LINPACK Users' Guide*, SIAM, Philadelphia.

Gelb, A., editor (1974) *Applied Optimal Estimation*, The MIT Press.

Grewal, M.S. and Andrews, A.P. (2001) *Kalman Filtering: Theory and Practice using MATLAB®*. John Wiley & Sons, Inc.

Haykin, S. (1996) *Adaptive Filter Theory*. Prentice Hall, Upper Saddle River, New Jersey, 3rd ed.

Julier, S.J. and Uhlmann, J.K. (1996) *A General Method for Approximating Nonlinear Transformations of Probability Distribution*. Technical report, Department of Engineering Science, University of Oxford, UK.

Julier, S.J., Uhlmann, J.K. and Durrant-Whyte, H.F. (1995) *A New Approach for Filtering Nonlinear Systems*. Proceedings of the American Control Conference, pp. 1628-1632.

Kitagawa, G. and Higuchi, T. (2001), *Preface: Special Issue on Nonlinear Non-Gaussian Models and Related Filtering Methods,* Annals of the Institute of Statistical Methematics, 53(1):iii.

Koch, K.-R. (1988) *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer-Verlag, Berlin, Heidelberg.

Lefebvre, T., Bruyninckx, H. and de Schutter, J. (2002) *Comment on "A new method for the nonlinear transformation of means and covariances in filters and estimators."* IEEE Transactions on Automatic Control, 47(8):1406-1408.

Nørgaard, M., Poulsen, N.K., and Ravn, O. (2000) *New developments in state estimation of nonlinear systems*, Automatica, 36:1672-1638.

Rauch, H.E., Tung, F. and Striebel, C.T. (1965) *Maximum likelihood estimates of linear dynamic systems*. AIAA Journal, 3(8):1445-1450.

Shin, E.H. (2005) *Estimation Techniques for Low-Cost Inertial Navigation*. Ph.D. thesis, Department of Geomatics Engineering, The University of Calgary, Calgary, Alberta, Canada.

Stewart, R.W. and Chapman, R. (1990) *Fast Stable Kalman Filter Algorithms Utilising the Square Root*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Albuquerque, New Mexico, pp. 1815-1818.

Terejanu, G., Singh, T. and Scott, P.D. (2007) *Unscented Kalman filter/smoother for a CBRN Puff-based dispersion model*. IEEE, 30(1):78-86.

Van der Merwe, R. and Wan, E.C. (2001) *The Square-Root Unscented Kalman Filter for State and Parameter-Estimation*. International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, Utah.

Vercauteren, T. and Wang, X. (2005) *Decentralized sigma-point information filters for target tracking in collaborative sensor networks*. IEEE Transactions on Signal Processing, 53(8):2997-3009.

Zoeter, O., Ypma, A. and Heskes, T. (2004) *Improved Unscented Kalman Smoothing for Stock Volitility Estimation*. IEEE Workshop on Machine Learning for Signal Processing, pp. 143-152.

Biography

Eun-Hwan Shin is a navigation analyst at Applanix Corporation. His research interests are on developing low-cost integrated navigation systems. He received a B.Sc. and a M.Sc. from the Seoul National University in 1992 and 1994, respectively. He also received a M.Sc. and a Ph.D. degree in Position, Location and Navigation from the University of Calgary in 2001 and 2005, respectively. He received a best-paper-in-track award in IEEE PLANS Conference in 2004.