[8] E. J. Chen and L. H. Lee, "A multi-objective selection procedure of determining a Pareto set," *Comput. Oper. Res.*, vol. 36, pp. 1872–1879, 2009.
[9] L. H. Lee, E. P. Chew, S. Teng, and D. Goldsman, "Finding the non-dominated Pareto set for multi-objective simulation models," *IIE Trans.*, vol. 42, pp. 656–674, 2010.
[10] C. H. Chen, J. Lin, E. Yücesan, and S. E. Chick, "Simulation budget allocation for further enhancing the efficiency of ordinal optimization," *Discrete Event Dyna. Syst.: Theory Appl.*, vol. 10, pp. 251–270, 2000.
[11] C. H. Chen, E. Yücesan, L. Dai, and H. C. Chen, "Efficient computation of optimal budget allocation for discrete event simulation experiment," *IIE Trans.*, vol. 42, pp. 60–70, 2010.
[12] J. Branke, S. E. Chick, and C. Schmidt, "Selecting a selection procedure," *Manage. Sci.*, vol. 53, pp. 1916–1932, 2007.
[13] N. A. Pujowidianto, L. H. Lee, C. H. Chen, and C. M. Yap, "Optimal computing budget for constrained optimization," in *Proc. Winter Simul. Conf.*, 2009, pp. 584–589.
[14] C. H. Chen, "A lower bound for the correct subset-selection probability and its application to discrete event system simulations," *IEEE Trans. Autom. Control*, vol. 41, no. 8, pp. 1227–1231, Aug. 1996.
[15] C. H. Chen and L. H. Lee, *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*. Singapore: World Scientific Publishing, 2011.

Square-Root Sigma-Point Information Filtering

Guoliang Liu, *Member, IEEE*, Florentin Wörgötter, and Irene Markelić

***Abstract*—The sigma-point information filters employ a number of deterministic sigma-points to calculate the mean and covariance of a random variable which undergoes a nonlinear transformation. These sigma-points can be generated by the unscented transform or Stirling's interpolation, which corresponds to the unscented information filter (UIF) and the central difference information filter (CDIF) respectively. In this technical note, we develop the square-root extensions of UIF and CDIF, which have better numerical properties than the original versions, e.g., improved numerical accuracy, double order precision and preservation of symmetry. We also show that the square-root unscented information filter (SRUIF) might lose the positive-definiteness due to the negative Cholesky update, whereas the square-root central difference information filter (SRCDIF) has only positive Cholesky update. Therefore, the SRCDIF is preferable to the SRUIF concerning the numerical stability.**

***Index Terms*—Central difference information filter, multiple sensor fusion, nonlinear estimation, sigma-point filter, square-root filter, unscented information filter.**

I. Introduction

The information filter for nonlinear systems comprises two stages—prediction and update [1], [2]. The state and measurement prediction can be implemented by using finite-sample approximation techniques. For update, the existing nonlinear IF needs the so-called pseudomeasurement matrix H_k [2]. In essence, H_k is the optimal statistical linearization of the nonlinear measurement function in the sense of minimizing the fitting error. This method is also referred to as statistical linear regression [3]. This is the fundamental explanation or derivation for the existing nonlinear information filter framework. Therefore, any finite-sample approximation techniques can be used in the prediction part of nonlinear information filter, such as the *unscented transform* [4] and *Stirling's interpolation* [5]. In the literature, the unscented transform based unscented information filter (UIF) has been proposed by Kim *et al.* [1] and Lee [2]. On the other hand, the Stirling's interpolation based central difference information filter (CDIF) has been proposed by Liu [6]. Since both UIF and CDIF use a number of sigma-points to approximate the true mean and covariance [7], they can be called *sigma-points information filters* (SPIFs). The SPIFs requires the square-root of the covariance to calculate the sigma-points, so that the covariance matrix has to be symmetric and positive definite. As shown in [8], these two properties might be lost due to errors introduced by arithmetic operations performed on finite word-length digital computers, or ill-conditioned nonlinear filtering problems, i.e., near perfect measurements.

In this technical note, we propose to use square-root forms for both UIF and CDIF, which have shown improved numerical characteristics compared to their regular forms. Here we call them square-root unscented information filter (SRUIF) and square-root central difference information filter (SRCDIF) respectively. The square-root filters predict and update the square-root covariance instead of the full covariance. In this way, the square-root filters achieve better numerical characteristics than the regular ones, e.g., improved numerical accuracy, double order precision and preservation of symmetry [8]. The first square-root filter was developed by Potter [9] and was used in the Apollo manned mission [10]. Since then, many square-root extensions of conventional filters have been introduced and analyzed. Our work was inspired by Van der Merwe [11] who proposed square-root forms of sigma-point Kalman filters. Here we introduce the square-root extensions of UIF and CDIF and their numerical advantages for solving nonlinear state estimation.

This technical note is organized as follows. First, we briefly review the CDIF algorithm for nonlinear estimation in Section II, and then we introduce the SRCDIF and SRUIF in Section III and Section IV, respectively. Simulation results of a reentry vehicle tracking problem are presented and discussed in Section V. Finally, the work is concluded in Section VI.

II. Central Difference Information Filter

The CDIF employs Stirling's interpolation to generate the sigma-points, which can be further used to estimate the mean and covariance of the system state. In general, the CDIF algorithm includes two steps: prediction and measurement update which are described below. In case of multiple sensors, a global information fusion form of the measurement update can be derived [6].

A. Prediction

Here we consider the discrete-time nonlinear dynamic system

$$x_k = F(x_{k-1}, v) \tag{1}$$

where x_k is the state vector of the system at time step k, and $v \sim \mathcal{N}(\overline{v}, R_v)$ is Gaussian noise.

Manuscript received August 25, 2011; revised November 30, 2011, and February 06, 2012; accepted March 27, 2012. Date of publication April 05, 2012; date of current version October 24, 2012. This work was supported by the Bernstein Focus Neurotechnology (BFNT) Göttingen. Recommended by Associate Editor L. Schenato.

The authors are with Bernstein Center for Computational Neuroscience, III Physikalisches Institut—Biophysik, University of Göttingen, 37077 Germany (e-mail: liu@physik3.gwdg.de; worgott@physik3.gwdg.de; irene@physik3.gwdg.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2012.2193708

First, the state vector is augmented with the noise variable and the corresponding augmented covariance matrix is derived by

$$x_{k-1}^{a_v} = \begin{bmatrix} x_{k-1} \\ \overline{v} \end{bmatrix}, \quad P_{k-1}^{a_v} = \begin{bmatrix} P_{k-1} & 0 \\ 0 & R_v \end{bmatrix}. \tag{2}$$

A symmetric set of $2L+1$ sigma points is generated

$$\mathcal{X}_{i,k-1}^{a_v} = \begin{cases} x_{k-1}^{a_v}, & i = 0 \\ x_{k-1}^{a_v} + \left(h\sqrt{P_{k-1}^{a_v}}\right)_i, & i = 1, \cdots, L \\ x_{k-1}^{a_v} - \left(h\sqrt{P_{k-1}^{a_v}}\right)_i, & i = L+1, \cdots, 2L \end{cases} \tag{3}$$

where h is a scaling parameter and L is the dimension of the state $x_{k-1}^{a_v}$. The subscript i indicates the i_{th} column of the matrix. Each sigma point $\mathcal{X}_{i,k-1}^{a_v}$ contains the state and noise variable components

$$\mathcal{X}_{i,k-1}^{a_v} = \begin{bmatrix} \mathcal{X}_{i,k-1}^{x} \\ \mathcal{X}_{i,k-1}^{v} \end{bmatrix}. \tag{4}$$

These sigma points are further passed through the nonlinear function (1), such that the predicted sigma points for the discrete time k are derived as

$$\mathcal{X}_{i,k|k-1}^{x} = F\left(\mathcal{X}_{i,k-1}^{x}, \mathcal{X}_{i,k-1}^{v}\right). \tag{5}$$

Finally, the first two moments of the predicted state vector are obtained by the weighted sum of the transformed sigma points:

$$\hat{x}_k = \sum_{i=0}^{2L} w_i^m \mathcal{X}_{i,k|k-1}^{x} \tag{6}$$

$$\hat{P}_k = \sum_{i=1}^{L} w_i^{(c_1)} \alpha_i \alpha_i^T + \sum_{i=1}^{L} w_i^{(c_2)} \beta_i \beta_i^T \tag{7}$$

where $\alpha_i = \mathcal{X}_{i,k|k-1}^{x} - \mathcal{X}_{i+L,k|k-1}^{x}$ and $\beta_i = \mathcal{X}_{i,k|k-1}^{x} + \mathcal{X}_{i+L,k|k-1}^{x} - 2\mathcal{X}_{0,k|k-1}^{x}$. The corresponding weights for the mean and covariance are defined as

$$\begin{aligned} w_0^{(m)} &= \frac{h^2 - L}{h^2} \\ w_i^{(m)} &= \frac{1}{2h^2}, \\ w_i^{(c1)} &= \frac{1}{4h^2}, \\ w_i^{(c2)} &= \frac{h^2 - 1}{4h^4}, \quad i = 1, \cdots, 2L \end{aligned} \tag{8}$$

where $h \geq 1$ is the scalar central difference step size. If the random variables obey a Gaussian distribution, the optimal value of h is $\sqrt{3}$ [11].

As stated in [12], the information matrix and information vector are the dual of the mean and covariance, so that the predicted information matrix $\hat{Y}_k$ and the information vector $\hat{y}_k$ are derived as

$$\hat{y}_k = \hat{Y}_k \hat{x}_k \tag{9}$$

$$\hat{Y}_k = (\hat{P}_k)^{-1}. \tag{10}$$

B. Measurement Update

The measurement function of the nonlinear system is defined as

$$z_k = H(x_k) + n \tag{11}$$

where z_k is the measurement and $n \sim \mathcal{N}(\overline{n}, R_n)$ is the Gaussian noise of the measurement.

The sigma points used for the measurement update are derived as:

$$\mathcal{X}_{i,k|k-1} = \begin{cases} \hat{x}_k, & i = 0 \\ \hat{x}_k + \left(h\sqrt{\hat{P}_k}\right)_i, & i = 1, \cdots, L \\ \hat{x}_k - \left(h\sqrt{\hat{P}_k}\right)_i, & i = L+1, \cdots, 2L \end{cases} \tag{12}$$

The predicted measurement points are obtained by transforming the sigma points through (11)

$$\mathcal{Z}_{i,k|k-1} = H(\mathcal{X}_{i,k|k-1}). \tag{13}$$

Furthermore, the mean and cross-covariance are derived by

$$\hat{z}_k = \sum_{i=0}^{2L} w_i^m \mathcal{Z}_{i,k|k-1} \tag{14}$$

$$\hat{P}_{x_k z_k} = \sqrt{w_1^{(c1)} \hat{P}_k} (\mathcal{Z}_{1:L} - \mathcal{Z}_{L+1:2L})^T. \tag{15}$$

Finally, the measurement update of the information vector and the information matrix are derived as

$$y_k = \hat{y}_k + \phi_k \tag{16}$$

$$Y_k = \hat{Y}_k + \Phi_k \tag{17}$$

where ϕ_k and Φ_k are information contribution terms for the information vector and matrix respectively, which can be derived by

$$\phi_k = \hat{Y}_k \hat{P}_{x_k z_k} R_n^{-1} \left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T \hat{y}_k\right) \tag{18}$$

$$\Phi_k = \hat{Y}_k \hat{P}_{x_k z_k} R_n^{-1} \left(\hat{P}_{x_k z_k}\right)^T (\hat{Y}_k)^T. \tag{19}$$

The derivation of (18) and (19) is given in [1] and [2].

C. Global Information Fusion

For multiple sensor fusion, if the measurement noises between the sensors are uncorrelated, the measurement update for information fusion is simply expressed as a linear combination of the local information contribution terms [13]:

$$y_k = \hat{y}_k + \sum_{i=1}^{N} \phi_{i,k} \tag{20}$$

$$Y_k = \hat{Y}_k + \sum_{i=1}^{N} \Phi_{i,k} \tag{21}$$

where N is the number of sensors. (20) and (21) show the main advantage of the information filters, which is the efficient measurement update. This superiority makes information filters more suitable for multiple sensor fusion than the Kalman filters. Note that the information matrix Y_k is the inverse of the covariance matrix P_k as shown in (10). When there is no prior information concerning the initial state, the Kalman filters have difficulties to cope with this situation since P_k is infinite. However, the information filters can deal with this special situation well, because $Y_k = (P_k)^{-1} = 0$. Other comparisons between information filters and Kalman filters can be found in [12].

III. Square-Root Central Difference Information Filter

The CDIF requires the square-root of the covariance to calculate the sigma-points in each discrete time update and measurement update, as shown in (3) and (12). The square-root operation is computationally expensive and demands that the covariance matrix must be positive semi-definite. To avoid the square-root operation and improve the numerical stability, we introduce the square-root central difference information filter (SRCDIF).

The square root form has important numerical advantages over the regular one: First, since the square-root of the covariance matrix is directly available, the SRCDIF saves computational cost for generating the sigma-points. Second, the numerical accuracy is improved because the condition number of the square root of the covariance matrix is only half of the covariance matrix [12]. Third, the square-root filters can achieve twice the effective precision of the regular forms [14]. Fourth, the symmetry and nonnegative properties of the covariance matrix are kept [12].

A. SRCDIF for State Estimation

The SRCDIF benefits from three powerful matrix factorization techniques: *QR decomposition*, *Cholesky factor update* and *efficient least squares*. In the following, we will use *qr*, *chol*, *cholupdate* to refer to the *QR decomposition*, *Cholesky decomposition*, and *Cholesky factor update*, respectively.

Algorithm 1 SRCDIF for state estimation

- Initialization:
  $x_0 = E(x)$, $S_{x0} = chol\{E((x-x_0)(x-x_0)^T)\}$, $S_v = \sqrt{R_v}$ and $S_n = \sqrt{R_n}$.
- For $k = 1, \cdots, \infty$:
  1) Generate sigma points for prediction:

$$x_{k-1}^{a_v} = \begin{bmatrix} x_{k-1} \\ \overline{v} \end{bmatrix}, \quad S_{k-1}^{a_v} = \begin{bmatrix} S_{x_{k-1}} & 0 \\ 0 & S_v \end{bmatrix} \tag{22}$$

$$\mathcal{X}_{k-1}^{a_v} = \begin{bmatrix} x_{k-1}^{a_v} & x_{k-1}^{a_v} + hS_{k-1}^{a_v} & x_{k-1}^{a_v} - hS_{k-1}^{a_v} \end{bmatrix}. \tag{23}$$

  2) Prediction equations:

$$\mathcal{X}_{k|k-1}^{x} = F\left(\mathcal{X}_{k-1}^{x}, \mathcal{X}_{k-1}^{v}, u_{k-1}\right) \tag{24}$$

$$\hat{x}_k = \sum_{i=0}^{2L} w_i^{(m)} \mathcal{X}_{i,k|k-1}^{x} \tag{25}$$

$$A = \sqrt{w_1^{(c_1)}} \left(\mathcal{X}_{1:L,k|k-1}^{x} - \mathcal{X}_{L+1:2L,k|k-1}^{x}\right) \tag{26}$$

$$B = \sqrt{w_1^{(c_2)}} \left(\mathcal{X}_{1:L,k|k-1}^{x} + \mathcal{X}_{L+1:2L,k|k-1}^{x} - 2\mathcal{X}_{0,k|k-1}^{x}\right) \tag{27}$$

$$\hat{S}_{x_k} = qr\left\{[A \quad B]\right\} \tag{28}$$

$$\hat{y}_k = \hat{S}_{x_k}^{-T}\left(\hat{S}_{x_k}^{-1}\hat{x}_k\right) \tag{29}$$

$$\hat{S}_{y_k} = qr\left\{\hat{S}_{x_k}^{-1} I\right\}. \tag{30}$$

  3) Generate sigma points for measurement update:

$$\mathcal{X}_{k|k-1} = \left[\hat{x}_k \quad \hat{x}_k + h\hat{S}_{x_k} \quad \hat{x}_k - h\hat{S}_{x_k}\right]. \tag{31}$$

  4) Measurement update equations:

$$\mathcal{Z}_{k|k-1} = H(\mathcal{X}_{k|k-1}) \tag{32}$$

$$\hat{z}_k = \sum_{i=0}^{2L} w_i^{(m)} \mathcal{Z}_{i,k|k-1} \tag{33}$$

$$\hat{P}_{x_k z_k} = \sqrt{w_1^{c_1}} \hat{S}_{x_k} [\mathcal{Z}_{1:L,k|k-1} - \mathcal{Z}_{L+1:2L,k|k-1}]^T \tag{34}$$

$$U = \hat{S}_{x_k}^{-T}\left(\hat{S}_{x_k}^{-1}\hat{P}_{x_k z_k}\right) S_n^{-T} \tag{35}$$

$$y_k = \hat{y}_k + US_n^{-1}\left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T \hat{y}_k\right) \tag{36}$$

$$S_{y_k} = cholupdate\left\{\hat{S}_{y_k}, U, +\right\}. \tag{37}$$

- *QR decomposition*. In the CDIF, the square-root of the covariance matrix S is derived by *Cholesky decomposition* on P: $S = chol(P)^T$ where S is a lower triangular matrix and fulfills $P = SS^T$. If we know $P = AA^T$, the square-root factor S can be directly calculated from A by *QR decomposition*: $S = qr(A)^T$. If the matrix $A \in \mathbb{R}^{L\times N}$, then the computational complexity of a *QR decomposition* is $\mathcal{O}(NL^2)$.
- *Cholesky factor update*. If the original update of the covariance matrix is $P \pm uu^T$ and S is the Cholesky factor, then the rank 1 update of S is $S = cholupdate(S, u, \pm)$ where u is the update vector and $\pm$ means the positive $(+)$ or negative $(-)$ update. The positive update is usually numerical stable, but the negative update may destroy the positive definite property of S [14], [15]. If u is a matrix, we can update each column of u one by one in a loop. For each column vector, the computational complexity is $\mathcal{O}(L^2)$. This procedure can alternatively be implemented as $S = qr([S \pm u]^T)$ using *QR decomposition* without the loop updates.
- *Efficient least squares*. The least-squares solution for the linear equation $Px = b$ can be solved efficiently using forward and back substitution if the Cholesky factor S is known and satisfies $P = SS^T$. For example, we can solve the linear equation $Px = b$ by $x = S^{-T}(S^{-1}b)$. This operation has computational complexity $\mathcal{O}(L^2)$.

The whole process is shown in Algorithm 1, where h is the scaling parameter, L is the dimension of the state, R_v and R_n are the process noise covariance and observation noise covariance, respectively, $w_i^{(m)}$ and $w_i^{(c)}$ are weights calculated in (8), and I is the identity matrix.

In the prediction step, the *Cholesky factor* $\hat{S}_{x_k}$ is updated using *QR decomposition* on the weighted sigma points. This step replaces the $\hat{P}_k$ update in (7) and has the complexity $\mathcal{O}(L^3)$. The information vector $\hat{y}_k = \hat{P}_k^{-1}\hat{x}_{x_k} = \hat{S}_{x_k}^{-T}(\hat{S}_{x_k}^{-1}\hat{x}_k)$ is derived using *efficient least squares* in (29). Because $\hat{S}_{x_k}$ is a square and triangular matrix, we can directly use back-substitution for solving $\hat{y}_k$ without the need for matrix inversion. The back substitution only requires $\mathcal{O}(L^2)$. Next is the calculation of the square-root information matrix $\hat{S}_{y_k}$ in (30). This step requires a *QR decomposition* since $\hat{S}_{y_k}$ is a upper triangular matrix and $\hat{S}_{x_k}$ is a lower triangular matrix. $\hat{S}_{y_k}$ and $\hat{S}_{x_k}$ meet $\hat{S}_{y_k}^T\hat{S}_{y_k} = \hat{S}_{x_k}^{-T}\hat{S}_{x_k}^{-1}$. To avoid the inversion, here we use *efficient least squares* to solve $\hat{S}_{x_k}^{-1}$ as $\hat{S}_{x_k}^{-1}I$, where I is the identity matrix.

In the measurement update step, the updated information vector in (36) is derived from (16) as follows:

$$\begin{aligned} y_k &= \hat{y}_k + \hat{Y}_k\hat{P}_{x_k z_k}R_n^{-1}\left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T\hat{y}_k\right) \\ &= \hat{y}_k + \hat{P}_k^{-1}\hat{P}_{x_k z_k}R_n^{-1}\left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T\hat{y}_k\right) \\ &= \hat{y}_k + \left(\hat{S}_{x_k}\hat{S}_{x_k}^T\right)^{-1}\hat{P}_{x_k z_k}\left(S_n S_n^T\right)^{-1}\left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T\hat{y}_k\right) \\ &= \hat{y}_k + \hat{S}_{x_k}^{-T}\left(\hat{S}_{x_k}^{-1}\hat{P}_{x_k z_k}\right)S_n^{-T}S_n^{-1}\left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T\hat{y}_k\right) \\ &= \hat{y}_k + US_n^{-1}\left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T\hat{y}_k\right) \end{aligned} \tag{38}$$

where $U = \hat{S}_{x_k}^{-T}\left(\hat{S}_{x_k}^{-1}\hat{P}_{x_k z_k}\right)S_n^{-T}$ as shown in (35). Since $\hat{S}_{x_k}$ and S_n are square and triangular matrices, y_k can be calulated using *efficient least squares* without the matrix inverse. The updated information matrix in (17) can be rewritten as

$$\begin{aligned} Y_k &= \hat{Y}_k + \hat{Y}_k\hat{P}_{x_k z_k}R_n^{-1}\left(\hat{P}_{x_k z_k}\right)^T(\hat{Y}_k)^T \\ &= \hat{Y}_k + \hat{Y}_k\hat{P}_{x_k z_k}S_n^{-T}S_n^{-1}\left(\hat{Y}_k\hat{P}_{x_k z_k}\right)^T \\ &= \hat{Y}_k + \hat{Y}_k\hat{P}_{x_k z_k}S_n^{-T}\left(\hat{Y}_k\hat{P}_{x_k z_k}S_n^{-T}\right)^T \\ &= \hat{Y}_k + UU^T. \end{aligned} \tag{39}$$

Because $\hat{S}_{y_k}$ is the Cholesky factor of the information matrix $\hat{Y}_k$, the updated Cholesky factor S_{y_k} of Y_k can be derived using the Cholesky update. If the observation dimension is M, the updated square-root information matrix S_{y_k} is calculated in (37) by applying an M-sequential Cholesky update to $\hat{S}_{y_k}$. The columns of matrix U are update vectors. This sequential Cholesky update only requires $\mathcal{O}(L^2M)$.

B. Square-Root CDIF for Multiple Sensor Fusion

In the case where information from multiple sensors is available, i.e., $N > 1$, we can fuse this using the Square-Root CDIF. For the i_{th} sensor, the information contribution for the information vector is

$$\phi_{i,k} = US_n^{-1}\left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T \hat{y}_k\right) \tag{40}$$

where U is defined in (35). The information contribution for the square-root information matrix is

$$S_{i,\phi_k} = U. \tag{41}$$

The final estimated result is derived by:

$$y_k = \hat{y}_k + \sum_{i=0}^{N} \phi_{i,k} \tag{42}$$

$$S_{y_k} = cholupdate\left\{\hat{S}_{y_k}, [S_{1,\phi_k}\ S_{2,\phi_k}\ \cdots\ S_{N,\phi_k}], +\right\}. \tag{43}$$

IV. Square-Root Unscented Information Filter

Algorithm 2 SRUIF for state estimation

- Initialization:
  $x_0 = E(x)$, $S_{x0} = chol\{E((x - x_0)(x - x_0)^T)\}$, $S_v = \sqrt{R_v}$, and $S_n = \sqrt{R_n}$.
- For $k = 1, \cdots, \infty$:
  1) Generate sigma points for prediction:

$$x_{k-1}^{a_v} = \begin{bmatrix} x_{k-1} \\ \overline{v} \end{bmatrix}, \quad S_{k-1}^{a_v} = \begin{bmatrix} S_{x_{k-1}} & 0 \\ 0 & S_v \end{bmatrix} \tag{44}$$

$$\mathcal{X}_{k-1}^{a_v} = \begin{bmatrix} x_{k-1}^{a_v} & x_{k-1}^{a_v} + \gamma S_{k-1}^{a_v} & x_{k-1}^{a_v} - \gamma S_{k-1}^{a_v} \end{bmatrix}. \tag{45}$$

  2) Prediction equations:

$$\mathcal{X}_{k|k-1}^{x} = F\left(\mathcal{X}_{k-1}^{x}, \mathcal{X}_{k-1}^{v}, u_{k-1}\right) \tag{46}$$

$$\hat{x}_k = \sum_{i=0}^{2L} w_i^{(m)} \mathcal{X}_{i,k|k-1}^{x} \tag{47}$$

$$\hat{S}_{x_k} = qr\left\{\sqrt{w_1^{(c)}}\left(\mathcal{X}_{1:2L,k|k-1}^{x} - \hat{x}_k\right)\right\} \tag{48}$$

$$C = \sqrt{w_0^{(c)}}\left(\mathcal{X}_0^{x} - \hat{x}_k\right) \tag{49}$$

$$\hat{S}_{x_k} = cholupdate\left\{\hat{S}_{x_k}, C, sign\left\{w_0^{(c)}\right\}\right\} \tag{50}$$

$$\hat{y}_k = \hat{S}_{x_k}^{-T}\left(\hat{S}_{x_k}^{-1}\hat{x}_k\right) \tag{51}$$

$$\hat{S}_{y_k} = qr\left\{\hat{S}_{x_k}^{-1} I\right\}. \tag{52}$$

  3) Generate sigma points for measurement update:

$$\mathcal{X}_{k|k-1} = \left[\hat{x}_k \quad \hat{x}_k + \gamma\hat{S}_{x_k} \quad \hat{x}_k - \gamma\hat{S}_{x_k}\right]. \tag{53}$$

  4) Measurement update equations:

$$\mathcal{Z}_{k|k-1} = H(\mathcal{X}_{k|k-1}) \tag{54}$$

$$\hat{z}_k = \sum_{i=0}^{2L} w_i^{(m)} \mathcal{Z}_{i,k|k-1} \tag{55}$$

$$\hat{P}_{x_k z_k} = \sum_{i=0}^{2L} w_i^{(c)}\left[\mathcal{X}_{i,k|k-1} - \hat{x}_k^-\right]\left[\mathcal{Z}_{i,k|k-1} - \hat{z}_k\right]^T \tag{56}$$

$$U = \hat{S}_{x_k}^{-T}\left(\hat{S}_{x_k}^{-1}\hat{P}_{x_k z_k}\right) S_n^{-T} \tag{57}$$

$$y_k = \hat{y}_k + US_n^{-1}\left(z_k - \hat{z}_k + \hat{P}_{x_k z_k}^T \hat{y}_k\right) \tag{58}$$

$$S_{y_k} = cholupdate\left\{\hat{S}_{y_k}, U, +\right\}. \tag{59}$$

In this section we consider the square-root implementation of the UIF. Because the UIF uses the *unscented transform* to calculate the sigma points, the architecture of the Square-Root unscented information filter (SRUIF) has few differences from the SRCDIF. As mentioned in Section III, the main techniques behind the square-root form estimators are: *QR decomposition*, *Cholesky factor update* and *efficient least squares*. We show how to use these in the SRUIF in the following.

The SRUIF is shown in Algorithm 2, where $\gamma = \sqrt{(\lambda + L)}$ in (45) is the composite scaling parameter, $\lambda = \alpha^2(L + \kappa) - L$, α and κ are scaling parameters that determine how far the sigma points spread from the mean value [11], L is the dimension of the state, R_v and R_n are process noise covariance and observation noise covariance respectively, $w_i^{(m)}$ and $w_i^{(c)}$ are weights calculated by $w_0^m = \lambda/L + \lambda$, $w_0^c = (\lambda/L + \lambda) + (1 - \alpha^2 + \beta)$, $w_i^m = w_i^c = 1/2(L + \lambda)\ i = 1, \cdots, 2L$, and $sign\{\ \}$ in (50) is the *signum* function.

We compare the SRUIF in Algorithm 2 to the SRCDIF in Algorithm 1. First, the SRUIF uses the *unscented transform* to calculate the sigma points in (45) and (53), where the scaling parameter becomes $\gamma = \sqrt{(\lambda + L)}$ and $\lambda = \alpha^2(L + \kappa) - L$. In contrast to only one scaling parameter h used in the SRCDIF, the SRUIF depends on three parameters λ, α and κ. Second, since the weight $w_0^{(c)}$ might be negative, we need an additional *cholupdate* to update the *Cholesky factor* $\hat{S}_{x_k}$ in (50), whereas the SRCDIF does not need this step since all weights used for the covariance update are positive. As we mentioned in Section III-A, the negative update might destroy the positive definite property of the Cholesky factor, such that the SRCDIF is preferable to the SRUIF concerning the numerical stability. Finally, for multiple sensor fusion, the SRUIF is equivalent to the SRCDIF in (42) and (43).

V. Experiments

In this paper, two individual experiments are demonstrated. The first experiment uses a normal noisy measurement to show the performance and computational cost of the proposed filters. In contrast, the second experiment utilizes a near perfect measurement to illustrate the improved numerical characteristics of the proposed square-root filters. Here we consider a classic space-vehicle reentry tracking problem: a high speed vehicle is tracked by radars located on the surface of the earth as shown in Fig. 1(a). The state vector of the filter consists of the position (x_1 and x_2), the velocity (x_3 and x_4) and a parameter related to the aerodynamic force (x_5). As described in [4], [16], the vehicle state dynamics for the discrete case are given by

$$\begin{aligned}
x_1(k+1) &= x_1(k) + \Delta t x_3(k) \\
x_2(k+1) &= x_2(k) + \Delta t x_4(k) \\
x_3(k+1) &= x_3(k) + \Delta t\left(D(k)x_3(k) + G(k)x_1(k)\right) + w_1(k) \\
x_4(k+1) &= x_4(k) + \Delta t\left(D(k)x_4(k) + G(k)x_2(k)\right) + w_2(k) \\
x_5(k+1) &= x_5(k) + \Delta t w_3(k)
\end{aligned} \tag{60}$$

where $w_1(k)$, $w_2(k)$, $w_3(k)$ are Gaussian process noises, $D(k)$ is the drag-related force, $G(k)$ is the gravity-related force, and $\Delta t = 0.1$ s is the sampling time. The force terms are given by

$$D(k) = \beta(k)V(k)exp\left\{\frac{R_0 - R(k)}{H_0}\right\}$$

$$G(k) = -\frac{Gm_0}{R^3(}$$