Sigma-Point Filtering for Nonlinear Systems with Non-Additive Heavy-Tailed Noise

Filip Tronarp, Roland Hostettler, and Simo Särkkä
Department of Electrical Engineering and Automation
Aalto University, Espoo, Finland
E-Mail: { filip.tronarp, roland.hostettler, simo.sarkka }@aalto.fi

***Abstract*—This paper is concerned with sigma-point methods for filtering in nonlinear systems, where the process and measurement noise are heavy tailed and enter the system non-additively. The problem is approached within the framework of assumed density filtering and the necessary statistics are approximated using sigma-point methods developed for Student's t-distribution. This leads to UKF/CKF-type of filters for Student's t-distribution. Four different sigma-point methods are considered that compute exact expectations of polynomials for orders up to 3, 5, 7, and 9, respectively. The resulting algorithms are evaluated in a simulation example and real data from a pedestrian dead-reckoning experiment. In the simulation experiment the nonlinear Student's t filters are found to be faster in suppressing large errors in the state estimates in comparison to the UKF when filtering in nonlinear Gaussian systems with outliers in process and measurement noise. In the pedestrian dead-reckoning experiment the sigma-point Student's t filter was found to yield better loop closure and path length estimates as well as significantly improved robustness towards extreme accelerometer measurement spikes.**

I. Introduction

The problem of state estimation is an ubiquitous problem in science and engineering. It arises in numerous sensor fusion applications in aerospace industry [1], smartphones [1], [2], health technology, as well as pedestrian dead-reckoning and indoor navigation [3], [4], among others.

In a mathematical sense, the state estimation problems arise in partially observed systems, where the measurements of a stochastic process are a function of the process and external noise. In this article, such state-space models of the following form are considered:

$$\begin{aligned} X_t &= f(X_{t-1}, U_t), \\ Y_t &= h(X_t, V_t), \end{aligned} \tag{1}$$

where $X_t \in \mathbb{R}^{d_X}$ is a latent process with initial distribution $X_0 \sim p(x_0)$, $Y_t \in \mathbb{R}^{d_Y}$ is the measurement of X_t, $U_t \in \mathbb{R}^{d_U}$ is heavy-tailed process noise, $f\colon \mathbb{R}^{d_X \times d_U} \to \mathbb{R}^{d_X}$ is the transition function of the state, $V_t \in \mathbb{R}^{d_V}$ is a heavy-tailed measurement error, and $h\colon \mathbb{R}^{d_X \times d_V} \to \mathbb{R}^{d_Y}$ is the measurement function. Furthermore it is assumed that $\mathrm{C}[X_{t-1}, U_t] = 0$, $\mathrm{C}[V_t, U_t] = 0$, and $\mathrm{C}[X_t, V_t] = 0$, where $\mathrm{C}[\cdot, \cdot]$ denotes the covariance. The state estimation problem is then to find a good estimate (posterior mean) of x_t given the measurement sequence, $y_{1:t}$, that is, $\hat{x}_{t|t} = \mathrm{E}[X_t \mid y_{1:t}]$, and the uncertainty (posterior covariance), $\mathrm{V}[X_t \mid y_{1:t}]$. Moreover, the notation $X_{\tau|t} \triangleq X_\tau | y_{1:t}$ is used for brevity. Furthermore, if $X \sim \mathrm{St}(\mu, \Sigma, \vartheta)$ then

$$\Sigma = \frac{\vartheta - 2}{\vartheta} \mathrm{V}[X], \tag{2}$$

where it's important to note that the covariance only exists for $\vartheta > 2$.

State estimation in systems with heavy-tailed process and measurement noise has already been studied to some degree. Student's t filters are developed for linear systems in [5], where it is also mentioned that their framework can be extended to nonlinear systems in a manner similar to the development of the extended Kalman filter for Gaussian systems. Nonlinear systems with additive noise, Gaussian state transitions, and Student's t observation are studied in [6], [7] where a variational Bayes approach is employed. In [8], an approximating smoother for the case with additive Student's t noise in the process and measurement is developed using variational Bayes. In the context of Gaussian processes in machine learning, Student's t-distributions are also sometimes used [9], [10], and state-space methods for that context have been developed, for example, in [11]. Other methods for robust filtering include, for example ℓ_1 and H_∞ approaches [12], [13].

In contrast, this paper develops sigma-point methods for filtering in nonlinear systems where the process and measurement noise are heavy-tailed, more specifically Student's t-distributed, and enter the system non-additively. The methods use numerical integration algorithms for integrating nonlinear functions with respect to Student's t-distributions, which are instances of the fully symmetric integration formulas presented in [14]. In this scenario the distribution of the state can be matched to a Student's t-density leading to a nonlinear extension of the Student's t filters presented in [5].

II. Sigma-Point Student's t Filtering

Here the filtering problem is approached from the point of view of statistical linear regression (SLR) [15]–[17], where we reformulate a sigma-point filtering approximation as a linearization of the state-space model with modified noise terms to account for the linearization error. Sigma-point filtering refers to numerical integration schemes that approximate the quantities needed to perform filtering by evaluating the function whose expectation is sought at a set of so called sigma-points; the expectation is then computed by a weighted sum [18]. This section proceeds with a discussion on Student's t filtering in the linear case, followed by Student's t statistically linearized

filter and ends with Student's t sigma-point filtering where sigma-point methods for Student's t-distribution are presented along with algorithms for performing approximate statistically linearized filtering in the Student's t filtering framework.

A. Student's t Filtering in Linear Systems

Methods for Student's t filtering in linear systems have already been developed [5]. Though since it's important for the development of Student's t filtering in nonlinear systems the results are presented here as well.

Suppose the system is linear, that is $f(X_{t-1}, U_t) = A^{(f)}X_{t-1} + B^{(f)}U_t$ and $h(X_t, V_t) = A^{(h)}X_t + B^{(h)}V_t$, where $A^{(f)} \in \mathbb{R}^{d_X \times d_X}, B^{(f)} \in \mathbb{R}^{d_X \times d_U}, A^{(h)} \in \mathbb{R}^{d_Y \times d_X}$, and $B^{(h)} \in \mathbb{R}^{d_y \times d_V}$. Furthermore, assume $X_{t-1|t-1}, U_t, V_t$ are jointly Student's t-distributed according to $\mathrm{St}(\mu_{t-1}, \Sigma_{t-1}, \vartheta_{t-1})$, where

$$\mu_{t-1} = \begin{bmatrix} \hat{x}_{t-1|t-1} \\ 0 \\ 0 \end{bmatrix}, \Sigma_{t-1} = \begin{bmatrix} \hat{\Sigma}_{t-1|t-1} & 0 & 0 \\ 0 & \Sigma_{U_t} & 0 \\ 0 & 0 & \Sigma_{V_t} \end{bmatrix}. \quad (3)$$

Then the random variables, $X_{t|t-1}, V_t$ are jointly Student's t distributed with the parameters

$$\mu'_t = \begin{bmatrix} A^{(f)}\hat{x}_{t-1|t-1} \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{x}_{t|t-1} \\ 0 \end{bmatrix}, \quad (4a)$$

$$\Sigma'_t = \begin{bmatrix} A^{(f)}\hat{\Sigma}_{t-1|t-1}(A^{(f)})^T + B^{(f)}\Sigma_{U_t}(B^{(f)})^T & 0 \\ 0 & \Sigma_{V_t} \end{bmatrix} = \begin{bmatrix} \hat{\Sigma}_{t|t-1} & 0 \\ 0 & \Sigma_{V_t} \end{bmatrix} \quad (4b)$$

and degrees of freedom parameter, $\vartheta'_t = \vartheta_{t-1}$, unchanged [19]. Going from (3) to (4) corresponds to the prediction step in a Student's t filter [5]. Now, the joint distribution of $X_{t|t-1}, Y_t$ is similarly given by a Student's t-distribution with parameters

$$\mu_t = \begin{bmatrix} \hat{x}_{t|t-1} \\ A^{(h)}\hat{x}_{t|t-1} \end{bmatrix}, \quad (5a)$$

$$\Sigma_t = \begin{bmatrix} \hat{\Sigma}_{t|t-1} & \hat{\Sigma}_{t|t-1}(A^{(h)})^T \\ A^{(h)}\hat{\Sigma}_{t|t-1} & A^{(h)}\hat{\Sigma}_{t|t-1}(A^{(h)})^T + B^{(h)}\Sigma_{V_t}(B^{(h)})^T \end{bmatrix} = \begin{bmatrix} \hat{\Sigma}_{t|t-1} & \hat{\Sigma}_{t|t-1,Y_t} \\ \hat{\Sigma}_{Y_t,t|t-1} & \Sigma_{Y_t} \end{bmatrix}. \quad (5b)$$

Hence, the update step is then given by [19]

$$\vartheta_t = \vartheta'_t + d_Y, \quad (6a)$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \hat{\Sigma}_{t|t-1,Y_t}\Sigma_{Y_t}^{-1}(y_t - A^{(h)}\hat{x}_{t|t-1}), \quad (6b)$$

$$\hat{\Sigma}_{t|t} = \frac{\vartheta'_t + \Delta^2_{y,t}}{\vartheta'_t + d_Y}(\hat{\Sigma}_{t|t-1} - \hat{\Sigma}_{t|t-1,Y_t}\Sigma_{Y_t}^{-1}\hat{\Sigma}^T_{t|t-1,Y_t}), \quad (6c)$$

where

$$\Delta^2_{y,t} = (y_t - A^{(h)}\hat{x}_{t|t-1})^T\Sigma_{Y_t}^{-1}(y_t - A^{(h)}\hat{x}_{t|t-1}).$$

This provides an exact filter that converges to the ordinary Kalman filter since the degrees of freedom parameter, $\vartheta_t \to \infty$ as $t \to \infty$. Though, it was proposed in [5] that assuming $X_{t-1|t-1}$ and U_t are marginally Student's t, the prediction step can be carried out by approximating their joint density by a Student's t with degrees of freedom parameter $\min(\vartheta_{t-1}, \vartheta_{U_t})$ and rescaled matrices $\hat{\Sigma}_{t-1|t-1}$ and Σ_{U_t} in order to preserve the covariances of the marginal distributions. The update is then carried out by assuming $X_{t|t-1}$ and V_t are marginally Student's t and their joint distribution is approximated in the same manner with degrees of freedom parameter $\min(\vartheta_{V_t}, \min(\vartheta_{t-1}, \vartheta_{U_t}))$ and the matrices $\hat{\Sigma}_{t|t-1}$ and Σ_{V_t} are rescaled to preserve covariances.

B. Statistically Linearized Student's t Filter

Since both the Gaussian distribution and the Student's t-distribution are closed under linear transformation the nonlinear filtering problem in either case is suitably solved by linearizing the transition function, $f(X, U)$, as well as the measurement function, $h(X, V)$. Under the assumption that $X_{t-1}, U_t, V_t \mid y_{1:t-1}$ are jointly Gaussian, the ordinary Kalman update may be employed. Though one can also assume that $X_{t-1}, U_t, V_t \mid y_{1:t-1}$ are jointly Student's t which enables the Student's t prediction and update.

Assume that X and W are uncorrelated, with $\mathrm{E}[X] = \mu_X$, $\mathrm{E}[W] = 0$, $\mathrm{C}[X, X] = \Sigma_X$, and $\mathrm{C}[W, W] = \Sigma_W$. Then, for an arbitrary nonlinear function $g(X, W)$, approximations of the following form are sought:

$$g(X, W) \approx AX + BW + D + E, \quad (7)$$

where $A \in \mathbb{R}^{d \times d_X}$, $B \in \mathbb{R}^{d \times d_W}$, $D \in \mathbb{R}^d$, $E \in \mathbb{R}^d$, and E is a zero-mean random variable with covariance matrix $\mathrm{V}[E]$. Note that if $g = f$ then $W = U$ and $d = d_X$, and if $g = h$ then $W = V$ and $d = d_Y$. The statistical linear regression (SLR) solution to this problem is given as [15]–[17]

$$A = \mathrm{C}[g, X]\,\mathrm{V}[X]^{-1}, \quad (8a)$$

$$B = \mathrm{C}[g, W]\,\mathrm{V}[W]^{-1}, \quad (8b)$$

$$D = \mathrm{E}[g] - A\mathrm{E}[X], \quad (8c)$$

$$\mathrm{V}[E] = \mathrm{V}[g] - A\mathrm{V}[X]A^T - B\mathrm{V}[W]B^T. \quad (8d)$$

In the Gaussian case this can be interpreted as minimizing the Kullback-Leibler divergence of the approximate joint density of $g(X, W)$ and X implied by Equations (8a) to (8d) and the true joint density (see [17]).

The linear Student's t filter equations can be applied provided that the quantities in Equations (8a) to (8d) are available and the additional assumption that

$$X_{t-1|t-1}, U_t, V_t, E_t^{(f)}, E_t^{(h)} \mid y_{1:t-1} \sim \mathrm{St}(\mu_{t-1}, \Sigma_{t-1}, \vartheta_{t-1}),$$

where

$$\mu_{t-1} = \begin{bmatrix} \hat{x}_{t-1|t-1} & 0 & 0 & 0 & 0 \end{bmatrix}^T,$$

$$\Sigma_{t-1} = \frac{\vartheta_{t-1} - 2}{\vartheta_{t-1}}\,\mathrm{blkdiag}\Big(\mathrm{V}[X_{t-1|t-1}], \mathrm{V}[U_t], \ldots, \mathrm{V}[V_t], \mathrm{V}[E_t^{(f)}], \mathrm{V}[E_t^{(h)}]\Big).$$

The prediction step can then be carried out according to the following

$$\begin{bmatrix} X_{t|t-1} \\ V_t \\ E_t^{(h)} \end{bmatrix} = \begin{bmatrix} A_t^{(f)} & B_t^{(f)} & 0 & \mathrm{I} & 0 \\ 0 & 0 & \mathrm{I} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} X_{t-1|t-1} \\ U_t \\ V_t \\ E_t^{(f)} \\ E_t^{(h)} \end{bmatrix} + \begin{bmatrix} D_t^{(f)} \\ 0 \\ 0 \end{bmatrix}.$$

$X_{t|t-1}, V_t$ and $E_t^{(h)}$ are now jointly Student's t-distributed with parameters

$$\mu_t' = \begin{bmatrix} \hat{x}_{t|t-1} \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_t' = \begin{bmatrix} \hat{\Sigma}_{t|t-1} & 0 & 0 \\ 0 & \Sigma_V & 0 \\ 0 & 0 & \Sigma_{E_t^{(h)}} \end{bmatrix}. \quad (9)$$

The joint distribution of $X_{t|t-1}, Y_t$ is then readily retrieved by the following linear transformation

$$\begin{bmatrix} X_{t|t-1} \\ Y_t \end{bmatrix} = \begin{bmatrix} \mathrm{I} & 0 & 0 \\ A_t^{(h)} & B_t^{(h)} & \mathrm{I} \end{bmatrix} \begin{bmatrix} X_{t|t-1} \\ V_t \\ E_t^{(h)} \end{bmatrix} + \begin{bmatrix} 0 \\ D_t^{(h)} \end{bmatrix}, \quad (10)$$

where the parameters of the joint distribution are given by

$$\mu_t = \begin{bmatrix} \hat{x}_{t|t-1} \\ \mathrm{E}[h(X_{t|t-1}, V_t)] \end{bmatrix}, \quad (11)$$

$$\Sigma_t = \begin{bmatrix} \hat{\Sigma}_{t|t-1} & \hat{\Sigma}_{t|t-1}(A_t^{(h)})^T \\ A_t^{(h)}\hat{\Sigma}_{t|t-1} & \frac{\vartheta_{t-1}-2}{\vartheta_{t-1}}\mathrm{V}[h(X_{t|t-1}, V_t)] \end{bmatrix}. \quad (12)$$

Finally, the parameters of the distribution of the state are then readily updated by equation (6).

C. Student's t Sigma-Point Filter

The expectations present in Equations (8a) to (8d) are generally not tractable. Though, in the framework of Gaussian filtering, they can be approximated by, for example, the use of numerical integration methods [14] that give the correct integrals for polynomials of up to order 3, 5, 7, or 9 [20], [21] or by using various other sigma-point methods [18]. However, the integration methods developed in [14] are not only applicable to Gaussian distributions, but for any fully symmetric distribution. As a matter of fact, the Student's t-distribution is also fully symmetric, among other distributions.

The sigma-point method for computing the expectation of a non-linear function, $\tilde{g}(X)$ with respect to Student's t-distribution takes the following form

$$\begin{aligned} \mathrm{E}[\tilde{g}(X)] &= \int \tilde{g}(x)\mathrm{St}(x; \mu_X, \Sigma_X, \vartheta_X)\mathrm{d}x \\ &= \int \tilde{g}(\mu_X + \Sigma_X^{-\frac{1}{2}}x')\mathrm{St}(x'; 0, \mathrm{I}, \vartheta_X)\mathrm{d}x' \\ &\approx \sum w_j \tilde{g}(\mu_X + \Sigma_X^{-\frac{1}{2}}\mathcal{X}_j). \end{aligned} \quad (13)$$

In general, Student's t sigma-point methods differ from the Gaussian sigma-point methods in the choice of the weights, w_j, and the sigma-points, $\mathcal{X}_j$.

Student's t sigma-point method of degree 3 is given by the following weights [14]

$$w_j = \begin{cases} \frac{\kappa}{d_X+\kappa} & j = 0 \\ \frac{1}{2(d_X+\kappa)} & j = 1, \ldots, 2d_X \end{cases} \quad (14)$$

and the sigma-points

$$\mathcal{X}_j = \begin{cases} 0 & j = 0 \\ s\mathrm{e}_j & j = 1, \ldots, d_X \\ -s\mathrm{e}_{j-d_X} & j = d_X+1, \ldots, 2d_X \end{cases} \quad (15)$$

where

$$s = \left(\frac{\vartheta_X}{\vartheta_X - 2}(d_X + \kappa)\right)^{\frac{1}{2}}. \quad (16)$$

The unit vectors, e_j, denote basis vectors from the canonical basis and the quantity, κ, is a free parameter in the algorithm. Clearly, Student's t sigma-point method of degree 3 is exactly the same as the unscaled unscented transform [22]. The factor $\frac{\vartheta_X}{\vartheta_X-2}$ in the coordinate scaling is due to the fact that the covariance is $\mathrm{V}[X'] = \frac{\vartheta_X}{\vartheta_X-2}\mathrm{I}$. That is, if the integral in equation (13) would have been transformed into an expectation with regards to the variable

$$X'' = \left(\frac{\vartheta_X}{\vartheta_X - 2}\Sigma_X\right)^{\frac{1}{2}}(x - \mu_X), \quad (17)$$

the unscaled unscented transform would have been recovered.

For Student's sigma point method of degree 5 the weights are computed as [14]

$$w_j = \begin{cases} 1 - d_X\left(\frac{\mathrm{I}_2}{\mathrm{I}_4}\right)^2 & j = 0 \\ \frac{1}{2}\left(\frac{\mathrm{I}_2}{\mathrm{I}_4}\right)^2(\mathrm{I}_4 - (d_X - 1)\mathrm{I}_{22}) & j = 1 \ldots, 2d_X \\ \frac{1}{4}\left(\frac{\mathrm{I}_2}{\mathrm{I}_4}\right)^2 \mathrm{I}_{22} & j \in \mathcal{J}_{[s,s]} \end{cases}, \quad (18)$$

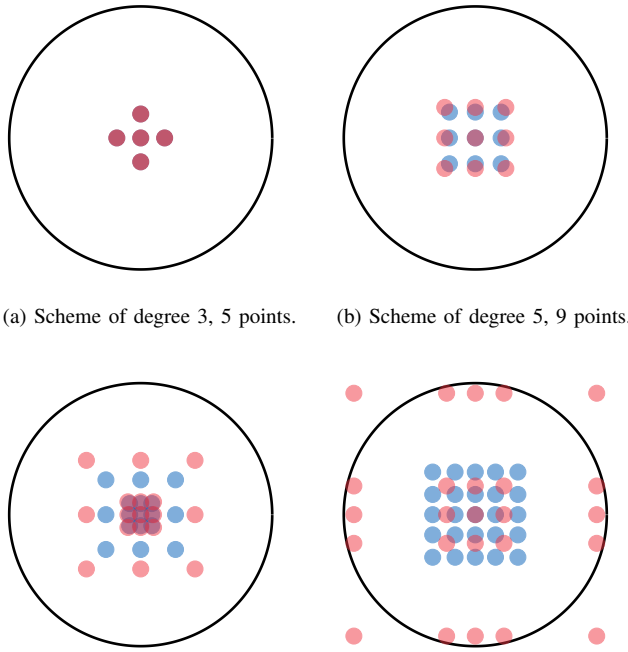
where

$$\mathcal{J}_{[s,s]} = \left\{2d_X + 1, \ldots, 2d_X + \frac{2d_X!}{(d_X-2)!}\right\}$$

and the sigma points are given by

$$\mathcal{X}_j = \begin{cases} 0 & j = 0 \\ s\mathrm{e}_j & j = 1, \ldots, d_X \\ -s\mathrm{e}_{j-d_X} & j = d_X+1, \ldots, 2d_X \\ \mathcal{S}_{[s,s]}^j & j \in \mathcal{J}_{[s,s]} \end{cases} \quad (19)$$

and $\mathcal{S}_{[s,s]}^j$ is the j-th element of the set of points $\mathcal{S}_{[s,s]}$ generated by all possible combinations of assigning the pairs $(s, s), (-s, s), (s, -s)$, and $(-s, -s)$ to the coordinate indices of a vector of length d_X such that no duplicate vectors are



(a) Scheme of degree 3, 5 points. (b) Scheme of degree 5, 9 points.

(c) Scheme of degree 7, 17 points. (d) Scheme of degree 9, 25 points.

Fig. 1. Placement of sigma-points for Student's t, $\vartheta = 9$, (red), Gaussian (blue) and a circle of radius 10 (black). (a) Sigma-point method of degree 3, (b) sigma-point method of degree 5, (c) sigma-point method of degree 7, and (d) sigma point method of degree 9. It is clear that Student's sigma-point method prefers to place the points further from origin.

generated. The quantities I_2, I_4, I_{22}, and s are given by

$$\mathrm{I}_2 = \frac{\vartheta_X}{\vartheta_X - 2}, \tag{20a}$$

$$\mathrm{I}_4 = \frac{3\vartheta_X^2}{(\vartheta_X - 2)(\vartheta_X - 4)}, \tag{20b}$$

$$\mathrm{I}_{22} = \frac{\vartheta_X^2}{(\vartheta_X - 2)(\vartheta_X - 4)}, \tag{20c}$$

$$s = \left(\frac{\mathrm{I}_2}{\mathrm{I}_4}\right)^{\frac{1}{2}}. \tag{20d}$$

The formulas for generating sigma-points and their corresponding weights are a lot more involved for methods of higher degrees and will not be listed here, instead, refer to [14]. Some useful formulas for computing the integrals in [14] when the weight function is a Student's t density are provided in [23].

Typically, the sigma-points of Student's t-distribution are placed further away from origin than for the Gaussian case, except for the schemes of degree 3 where both the Gaussian and Student's t method produce the same numerical scheme. Though when $\vartheta_X \to \infty$ the placement of the Gaussian sigma-point methods is recovered. This is illustrated in Fig. 1 where the sigma-points of a Gaussian and a Student's t-distribution, both with identity covariance, are shown.

It is also important to note that since sigma-point methods rely on the computation of higher order moments, the degree of freedom parameter, ϑ_X, puts a restriction on the degree of the sigma-point scheme. More specifically if the sigma-point scheme is of order $2p + 1$ it is necessary for the degrees of freedom to abide by $2p < \vartheta$ in order to have a mathematically sound scheme. Furthermore, the degrees of freedom put a restriction on what functions have an expectation. Namely, $\tilde{g}(\mu_X + \Sigma_X^{-\frac{1}{2}} x)$ can not grow faster than $||x||^{\vartheta}$.

Algorithm 1 Student's t Sigma-Point Predictor I

Input: $\hat{x}_{t|t}, \hat{\Sigma}_{t|t}, \Sigma_{U_{t+1}}, \vartheta_t$
Output: $\hat{x}_{t+1|t}, \hat{\Sigma}_{t+1|t}, \vartheta_{t+1}$
Approximate $\mathrm{E}[f(X_{t+1|t}, U_{t+1})]$ and $\mathrm{V}[f(X_{t+1|t}, U_{t+1})]$, with a Student's t sigma-point method.
$\hat{x}_{t+1|t} \leftarrow \mathrm{E}[f(X_{t+1|t}, U_{t+1})]$
$\hat{\Sigma}_{t+1|t} \leftarrow \frac{\vartheta_t - 2}{\vartheta_t} \mathrm{V}[f(X_{t+1|t}, U_{t+1})]$

Algorithm 2 Student's t Sigma-Point Predictor II

Input: $\hat{x}_{t|t}, \hat{\Sigma}_{t|t}, \Sigma_{U_{t+1}}, \Sigma_{V_{t+1}}, \vartheta_t, \vartheta_{U_{t+1}}, \vartheta_{V_{t+1}}$
Output: $\hat{x}_{t+1|t}, \hat{\Sigma}_{t+1|t}, \vartheta_{t+1}$
$\vartheta_t \leftarrow \min(\vartheta_t, \vartheta_{U_{t+1}}, \vartheta_{V_{t+1}})$
Rescale $\hat{\Sigma}_{t|t}, \Sigma_{U_{t+1}}, \Sigma_{V_{t+1}}$ so that the covariances are the same after degree of freedom change.
Approximate $\mathrm{E}[f(X_{t+1|t}, U_{t+1})]$ and $\mathrm{V}[f(X_{t+1|t}, U_{t+1})]$, with a Student's t sigma-point method.
$\hat{x}_{t+1|t} \leftarrow \mathrm{E}[f(X_{t+1|t}, U_{t+1})]$
$\hat{\Sigma}_{t+1|t} \leftarrow \frac{\vartheta_t - 2}{\vartheta_t} \mathrm{V}[f(X_{t+1|t}, U_{t+1})]$

Algorithm 3 Student's t Sigma-Point Updater

Input: $y_{t+1}, \hat{x}_{t+1|t}, \hat{\Sigma}_{t+1|t}, \Sigma_{V_{t+1}}, \vartheta_t$
Output: $\hat{x}_{t+1|t+1}, \hat{\Sigma}_{t+1|t+1}, \vartheta_{t+1}$
Approximate $\mathrm{E}[Y_{t+1}]$, $\mathrm{V}[Y_{t+1}]$, and $\mathrm{C}[X_{t+1|t}, Y_{t+1}]$ with a Student's t sigma-point method.
$z_{t+1} \leftarrow y_{t+1} - \mathrm{E}[Y_{t+1}]$
$\Sigma_{Y_{t+1}} \leftarrow \frac{\vartheta_t - 2}{\vartheta_t} \mathrm{V}[Y_{t+1}]$
$K_{t+1} \leftarrow \frac{\vartheta_t - 2}{\vartheta} \mathrm{C}[X_{t+1|t}, Y_{t+1}] \Sigma_{Y_t}^{-1}$
$\Delta_{y,t+1} \leftarrow z_{t+1}^T \Sigma_{Y_{t+1}}^{-1} z_{t+1}$
$\vartheta_{t+1} \leftarrow \vartheta_t + d_Y$
$\hat{x}_{t+1|t+1} \leftarrow \hat{x}_{t+1|t} + K_{t+1} z_{t+1}$
$\hat{\Sigma}_{t+1|t+1} \leftarrow \frac{\vartheta_t + \Delta_{y,t+1}^2}{\vartheta_t + d_Y} (\hat{\Sigma}_{t+1|t} - K_{t+1} \Sigma_{Y_{t+1}} K_{t+1}^T)$

The algorithm for the Student's t filter in the case where it converges to the unscented Kalman filter is constructed by cycling through the predictor in Algorithm 1 and the updater in Algorithm 3. A Student's t filter with preserved heavy tail behavior is then retrieved by employing Algorithm 2 as a predictor instead. The filter resulting from Algorithm 1 converges to the UKF hence its' only advantage is better supression of outliers prior to asymptotics kicking in. On the other hand the filter resulting from Algorithm 2 will preserve low degrees of freedom indefinitely and is therefore more suitable in the case where outliers are expected at any

time during the filtering procedure. Note that the approach for preserved heavy tail filtering here differs from the one in [5] though only when $\vartheta_{U_t} > \vartheta_{V_t}$. Nevertheless if it is desired to do predictions with high degrees of freedom and updates with low degrees of freedom only small adjustments to Algorithm 2 and Algorithm 3 are required to achieve this. It is also important to note that since the covariance update in Algorithm 3 depends on the squared Mahlanobis distance between measurement and predicted measurement, the filter may behave differently than the UKF even for fairly high values of the degrees of freedom parameter. To clarify, the filters' behaviors are only similar when $\Delta^2_{y,t+1} \approx d_Y$. That is, if the squared Mahalanobis distance between measurement and measurement mean significantly deviates from this relationship, then there is either a downscaling or an upscaling of the filter covariance in comparison to the UKF.

III. Experimental Results

The practical applicability of the proposed Student's t sigma-point filter is evaluated using a simulated example first and then in a pedestrian dead-reckoning (PDR) application using an inertial measurement unit (IMU) where high peaks in the measured IMU signals are a common problem. In both examples, the sigma-point filter is compared to the standard UKF as well as the extendet Student's t filter [5].

A. Simulated Example

The following system is considered

$$\begin{bmatrix} X^{(1)}_{t+1} \\ X^{(2)}_{t+1} \end{bmatrix} = \begin{bmatrix} 1 - \frac{0.1}{1+||X_t||} & \frac{1}{1+||X_t||} \\ -\frac{1}{1+||X_t||} & 1 - \frac{0.1}{1+||X_t||} \end{bmatrix} \begin{bmatrix} X^{(1)}_t \\ X^{(2)}_t \end{bmatrix} + U_{t+1} \tag{21}$$

$$Y_{t+1} = \begin{bmatrix} (1 + V^{(1)}_{t+1}) X^{(1)}_{t+1} \\ (1 + V^{(2)}_{t+1}) X^{(2)}_{t+1} \end{bmatrix}, \tag{22}$$

where the noise processes are governed by the following probability laws,

$$U_{t+1} \sim 0.95\mathcal{N}(u; 0, 0.01\mathrm{I}) + 0.05\mathcal{N}(u; 0, 5\mathrm{I}) \tag{23}$$

$$V_{t+1} \sim 0.9\mathcal{N}(v; 0, 0.01\mathrm{I}) + 0.1\mathcal{N}(v; 0, 5\mathrm{I}). \tag{24}$$

The system is simulated 5000 times for 250 time steps. Student's t sigma-point filter is compared against the unscented Kalman filter and the extended Student's t filter. The sigma-point filters used the scheme of degree 3 and the free parameter was set to $\kappa = -1$. The degrees of freedom parameter was set to $\vartheta = 4$ for Student's t filters. The performance of the filters are compared in terms of execution time (ET), mean norm error (MeNE) and maximum norm error (MaNE). Where mean norm error and max norm error refers to the time average and the maximum value of the sequence $\{||\hat{x}_t - x_t||\}_{t=1}^{250}$, respetively. x_t denote the realised value of X_t and $\hat{x}_t$ refers to the filter estimate of x_t. In order to evaluate Student's t filters for some moderately large degrees of freedom parameter the preceeding experiment is run again but with $\vartheta = 50$. The simulations were carried out in Matlab on a computer with an Intel Xeon E3-1231 v3 CPU 4×3.4 GHz and 16 GB of RAM.

TABLE I
Comparison between UKF, ESTF, and SPSTF in terms of mean and max norm errors in 5000 Monte Carlo simulations with trajectories of 250 samples, $\vartheta = 4$.

MeNE	UKF	ESTF	SPSTF
2.5%-th percentile	6.2890	4.1578	4.2166
50%-th percentile	12.7680	8.6331	8.7770
97.5%-th percentile	25.6030	59.9242	19.8164

MaNE	UKF	ESTF	SPSTF
2.5%-th percentile	41.1887	58.3000	67.9382
50%-th percentile	66.7930	189.7000	212.7570
97.5%-th percentile	162.7787	3569.4000	639.4494

TABLE II
Comparison between UKF, ESTF, and SPSTF in terms of mean and max norm errors in 5000 Monte Carlo simulations with trajectories of 250 samples, $\vartheta = 50$.

MeNE	UKF	ESTF	SPSTF
2.5%-th percentile	6.3708	4.7380	4.7355
50%-th percentile	12.8018	9.2261	9.2109
97.5%-th percentile	24.4484	26.2806	19.7414

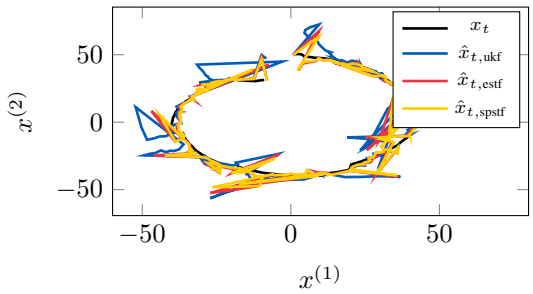
MaNE	UKF	ESTF	SPSTF
2.5%-th percentile	40.7023	49.4582	49.4700
50%-th percentile	66.5971	133.3444	133.4799
97.5%-th percentile	155.1950	708.7380	452.3309

B. Pedestrian Dead-Reckoning Example

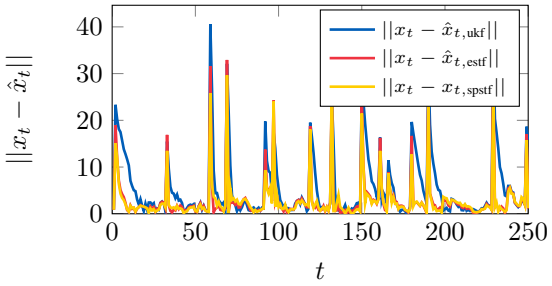
PDR has many different applications in localization and tracking, for example for tracking the movements of firefighters in and around accident sites. The predominant approach to solving this problem is to use strapdown inertial navigation using IMUs, attached to, among others, the users legs or embedded in their shoes [24]. Due to the nature of the IMU, an often encountered problem is high peaks in the measurement signals during the step-down phase of the gait cycle [4]. These spikes can cause errors in the PDR system as they violate the Gaussian assumption of the measurement and process noise. Hence, a filter based on heavy-tailed noise ought to be more robust in these situations.

1) Dynamic Model and Experiment Setup: PDR commonly makes use of indirect Kalman filtering [3], that is, the measured accelerations and rotational velocities are treated as control inputs to the system and then directly integrated. This, however, has the disadvantage that sensor biases and noise cause the integrator to drift. In order to mitigate that, known conditions of the gait cycle are used to generate pseudo-measurements of the integration error. One such state in particular is the stationary phase where the foot is still on the ground, that is, its velocity is zero. Comparing this known velocity to the integrated velocity yields the so called zero-velocity updates (ZUPT) that can be used to adjust the integrator. The discretized integrator is then given by

$$\begin{bmatrix} p_t \\ v_t \\ q_t \end{bmatrix} = \begin{bmatrix} I_3 & \Delta t I_3 & 0 \\ 0 & I_3 & 0 \\ 0 & 0 & \Omega(\omega^M_t) \end{bmatrix} \begin{bmatrix} p_{t-1} \\ v_{t-1} \\ q_{t-1} \end{bmatrix} + \begin{bmatrix} \frac{(\Delta t)^2}{2} \\ \Delta t \\ 0 \end{bmatrix} a_t \tag{25}$$

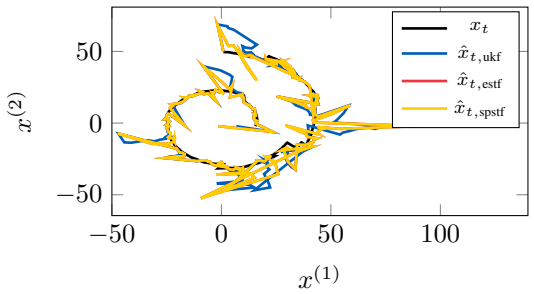


(a) A trajectory, $(x_t^{(1)}, x_t^{(2)})$, of 250 samples and the filter estimates. The degreees of freedom in Student's t filters were set to $\vartheta = 4$.

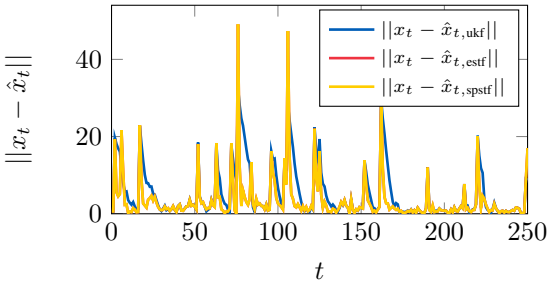


(b) The instantaneous norm of the error in state estimates, the degrees of freedom in Student's t filters were set to $\vartheta = 4$.

Fig. 2. Visualization of the performance of the unscented Kalman filter (UKF), the extended Student's t filter (estf), and the sigma-point Student's t filter (spstf). (a) A sample trajectory of length 250 with the filter estimates of the trajectory, (b) The norm of the error in state estimates of the filters. Student's t filters used a degrees of freedom parameter, $\vartheta = 4$. Student's t filters are significantly faster in supressing large errors compared to the UKF.



(a) A trajectory, $(x_t^{(1)}, x_t^{(2)})$, of 250 samples and the filter estimates. The degreees of freedom in Studnet's t filters were set to $\vartheta = 50$.



(b) The instantaneous norm of the error in state estimates, the degrees of freedom in Student's t filters were set to $\vartheta = 50$.

Fig. 3. Visualization of the performance of the unscented Kalman filter (UKF), the extended Student's t filter (estf), and the sigma-point Student's t filter (spstf). (a) A sample trajectory of length 250 with the filter estimates of the trajectory, (b) The norm of the error in state estimates of the filters. Student's t filters used a degrees of freedom parameter, $\vartheta = 50$. Student's t filters are notably faster in supressing large errors compared to the UKF.

where p_t is the position of the person in the reference coordinate frame at time instant t and v_t its velocity. q_t is the rotation quaternion from the IMU coordinate frame to the reference coordinate frame and Δt is the time since the last integrator update. a_t^M and ω_t^M denote the acceleration and gyroscope measurements both in the IMU coordinate frame, respectively, while a_t denotes the acceleration in the reference coordinate frame and is given by

$$a_t = q_t \circ (a_t^M - g) \circ q_t^*.$$

Here, $\circ$ denotes the quaternion product. Furthermore, the sub-matrix $\Omega(\omega_t^M)$ in (25) is given by

$$\Omega(\omega_t^M) = \cos\left(\frac{|\omega_t^M|\Delta t}{2}\right) I_4 + \begin{bmatrix} 0 & -\omega_t^{M,x} & -\omega_t^{M,y} & -\omega_t^{M,z} \\ \omega_t^{M,x} & 0 & \omega_t^{M,z} & -\omega_t^{M,y} \\ \omega_t^{M,y} & -\omega_t^{M,z} & 0 & \omega_t^{M,x} \\ \omega_t^{M,z} & \omega_t^{M,y} & -\omega_t^{M,x} & 0 \end{bmatrix} \frac{\sin\left(\frac{|\omega_t^M|\Delta t}{2}\right)}{|\omega_t^M|}$$

Finally, the error measurement can be described as

$$y_{t_k} = v_{ref} - v_{t_k} + r_{t_k} = -v_{t_k} + r_{t_k} \tag{26}$$

since the reference velocity during the stationary phase is assumed to be $v_{ref} = 0$. Furthermore t_k denotes the stationary phase's detection time and r_{t_k} is additive measurement noise. For a more thorough introduction to PDR dynamics and the indirect Kalman filtering methods commonly used therein, please refer to [3] or [4].

The experimental data was gathered as follows. An LG Nexus 5 smartphone was tightly strapped to a test subject's lower leg just above the ankle. The test subject was then asked to first stand still for about 5 s and then casually walk along a predefined path (a loop starting and ending at the same position) while the IMU data was gathered at a sampling rate of 200 Hz for both, the accelerometer and gyroscope.

The parameters used for evaluation were

$$Q = \begin{bmatrix} 2I_3 & 0 \\ 0 & I_3 \end{bmatrix} \quad \text{and} \quad R = 1 \times 10^{-3} I_3$$

for the process noise and the pseudo measurement noise covariances, respectively. The degrees of freedom for the t distributions were chosen to be 3 in order to preserve as heavy tails as possible. Furthermore, the initial position and velocity was set to zero (corresponding to standing still at the origin of the reference coordinate system) and the initial orientation was determined from the stationary data in the beginning of the measurement sequence. Finally, for the detection of the stationary phase, the generalized maximum likelihood ratio test shown in [25] is used, where the noise measurement noise

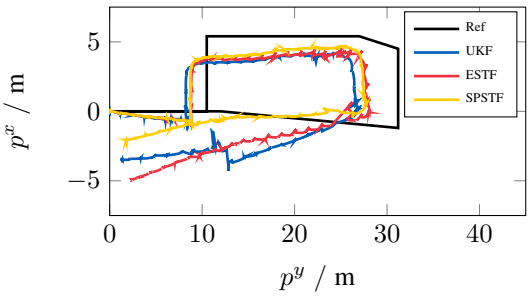


Fig. 4. Estimated trajectories using the standard UKF, the extended Student's t filter (ESTF), and the proposed sigma-point Student's t filter together with the reference trajectory.

covariances were set to

$$\Sigma_a = 1 \times 10^{-3} \operatorname{diag}\left(\begin{bmatrix}5 & 1 & 6\end{bmatrix}\right),$$
$$\Sigma_\omega = 1 \times 10^{-3} \operatorname{diag}\left(\begin{bmatrix}0.2 & 0.1 & 0.3\end{bmatrix}\right)$$

for the accelerometer and gyroscope, respectively. The detection threshold was set to $\gamma = 6 \times 10^3$ and the test data length was set to $N = 10$ samples.

2) Results and Discussion: Fig. 4 illustrates the estimated trajectories for all three filters together with the predefined path. As it can be seen, the estimated trajectories match the predefined path in shape but are slightly off in the y-direction, especially after the third turn (the starting point is $(0, 0)$). The complete trajectory length is 73.6 m while the estimated trajectory lengths are 84.3 m for the UKF, 93.7 m for the ESTF, and 78.5 m for the SPSTF. Furthermore, the offset at the end of the loop (loop closure error) is 4.6 m for the UKF, 5.4 m for the ESTF, and 3.9 m for the SPSTF. This indicates that the SPSTF is slightly more accurate in both, travelled distance and loop closure error. Also, it is interesting to note that the ESTF actually performs worse than the UKF, suggesting that the linearization approach is inferior to the sigma-point approach in general for this problem.

The individual components of the position and speed estimates are depicted in Fig. 5 and Fig. 6. No major difference between the UKF and the SPSTF can be observed in the position estimates except for the diverging behavior in $\hat{p}^x(t)$ around $t = 60\,\mathrm{s}$ (Fig. 5). Again, the ESTF exhibits the most deviating behavior, especially in the $\hat{p}^z(t)$ component. More interestingly, an important difference can be seen in the plots for the speed estimates in Fig. 6. Comparing the estimates for the UKF and the SPSTF, it can be seen that the UKF suffers from larger peaks. This is very prominently visible in the plot for $\hat{v}^y(t)$ around $t = 20\,\mathrm{s} \ldots 40\,\mathrm{s}$. Note that during that time, the direction of motion is primarily in the positive y-direction. Hence, the negative peaks estimated by the UKF are likely not true motions of the foot (moving the foot backwards) but rather caused by the spikes in the acceleration measurements when putting the foot on the ground. Hence, the results indicate that the SPSTF is more robust toward these kind of extreme value measurements since it uses a heavy-tailed distribution as the underlying noise model. Again, the behavior of the ESTF is

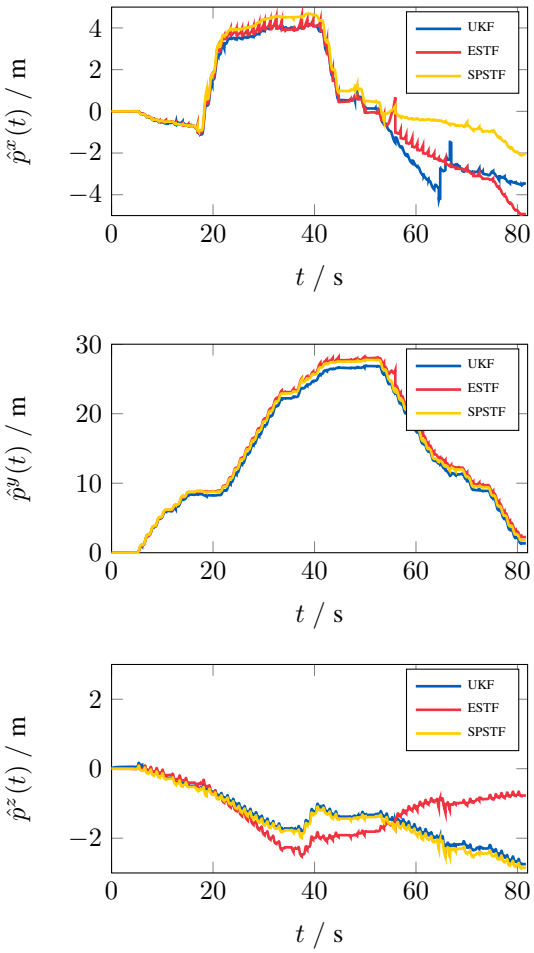


Fig. 5. Estimated position components using the standard UKF, the extended Student's t filter (ESTF), and the proposed sigma-point Student's t filter (SPSTF).

different from the sigma-point based filters in the sense that it has significant peaks in the $\hat{v}^x(t)$ and $\hat{v}^y(t)$ components, which are not present in the sigma-point based filters.

IV. Conclusion

A sigma-point method for filtering nonlinear systems where Student's t-distributed process and measurement noise enter the system non-additively was developed in an analogous manner to the UKF. Student's t sigma-point method differs from the Gaussian one in that it prefers to place the sigma-points further away from the origin, reflecting the heavier tail.

The resulting algorithm was compared to an extended Student's t filter and the UKF in a simulation experiment where Student's t filters were found to be faster in suppressing large errors in the state estimates in comparison to the UKF though they may momentarily produce very large errors, the extended filter in particular. All the same, Student's t filters were found to be superior to the UKF in sense of mean norm error. The sigma-point Student's t filter was also compared to the UKF in a PDR experiment where Student's t filter was found to outperform the UKF in terms of error in the estimated

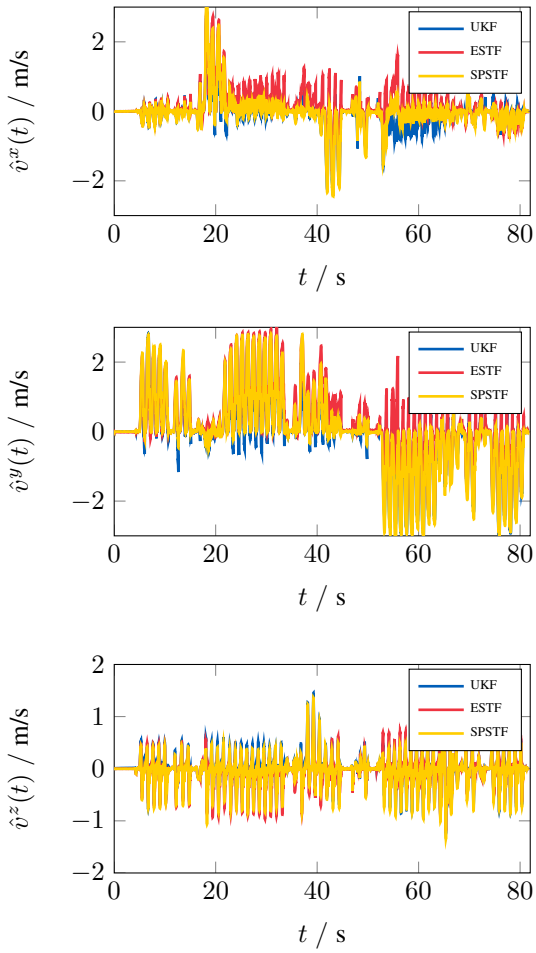


Fig. 6. Estimated speed components using the standard UKF, the extended Student's t filter (ESTF),and the proposed sigma-point Student's t filter (SPSTF).

trajectory length, loop-closure error, and it giving more realistic velocity estimates.

Future work will include a more in-depth study of the stability properties of the Student's t filter in relation to the Kalman filters. That is, to investigate the question whether it is possible to use a Student's t filter in cases where Kalman filters fail to converge. Another open aspect is the choice of degrees of freedom and how it affects the behavior of the proposed filter. A possible extension is to use an ensemble of filters using different degrees of freedom or considering the degrees of freedom as a random variable itself. Furthermore, investigating marginalized particle filters using Student's t filters in place of Kalman filters for the marginalized states.

Acknowledgment

Financial support by the Academy of Finland under the grant no. #295080 (CrowdSLAM) is hereby gratefully acknowledged.

References

[1] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House Radar Library, 1999.

[2] M. Li, B. H. Kim, and A. I. Mourikis, "Real-time motion tracking on a cellphone using inertial sensing and a rolling-shutter camera," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, May 2013, pp. 4712–4719.

[3] D. H. Titterton and J. L. Weston, *Strapdown Inertial Navigation Technology*. The Institution of Electrical Engineers, 2004.

[4] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *IEEE Computer Graphics and Applications*, vol. 25, no. 6, pp. 38–46, Nov 2005.

[5] M. Roth, E. Özkan, and F. Gustafsson, "A Student's t filter for heavy tailed process and measurement noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 5770–5774.

[6] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *IEEE International Workshop on Machine Learning for Signal Processing*, 2012, pp. 1–6.

[7] G. Agamennoni, J. I. Nieto, and E. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5024–5037, 2012.

[8] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 468–472, April 2016.

[9] P. Jylänki, J. Vanhatalo, and A. Vehtari, "Robust Gaussian process regression with a Student-t likelihood," *Journal of Machine Learning Research*, vol. 12, pp. 3227–3257, 2011.

[10] A. Shah, A. G. Wilson, and Z. Ghahramani, "Student-t processes as alternatives to Gaussian processes," in *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, ser. JMLR W&CP, vol. 33, 2014, pp. 877–885.

[11] A. Solin and S. Särkkä, "State space methods for efficient inference in Student-t process regression," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2015.

[12] A. Y. Aravkin, B. M. Bell, J. V. Burke, and G. Pillonetto, "An ℓ_1-Laplace robust Kalman smoother," *IEEE Transactions on Automatic Control*, vol. 56, no. 12, pp. 2898–2911, 2011.

[13] D. Simon, *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*. John Wiley & Sons, 2006.

[14] J. McNamee and F. Stenger, "Construction of fully symmetric numerical integration formulas," *Numerische Mathematik*, vol. 10, pp. 327–344, 1967.

[15] 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" [with authors' reply]," *Automatic Control, IEEE Transactions on*, vol. 47, no. 8, pp. 1406–1409, 2002.

[16] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-time nonlinear filtering algorithms using Gauss–Hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[17] A. F. Garcia-Fernandez, L. Svensson, M. R. Morelande, and S. Särkkä, "Posterior linearization filter: Principles and implementation using sigma points," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5561–5573, 2015.

[18] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.

[19] M. Roth, "On the multivariate t distribution," Division of Automatic Control, Linköping University, Tech. Rep. LiTH-ISY-R-3059, 2013.

[20] J. Kokkala, A. Solin, and S. Särkkä, "Sigma-point filtering and smoothing based parameter estimation in nonlinear dynamic systems," *Journal of Advances in Information Fusion*, to appear.

[21] S. Särkkä, J. Hartikainen, L. Svensson, and F. Sandblom, "On the relation between Gaussian process quadratures and sigma-point methods," *Journal of Advances in Information Fusion*, to appear.

[22] 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.

[23] S. Kotz and S. Nadarajah, *Multivariate t-distributions and their applications*. Cambridge University Press, 2004.

[24] J.-O. Nilsson, A. K. Gupta, and P. Händel, "Foot-mounted inertial navigation made easy," in *Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on*, Oct 2014, pp. 24–29.

[25] I. Skog, P. Händel, J.-O. Nilsson, and J. Rantakokko, "Zero-velocity detection – an algorithm evaluation," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 11, pp. 2657–2666, Nov 2010.