A student's t filter for heavy tailed process and measurement noise

Michael Roth, Emre Ozkan and Fredrik Gustafsson

Linköping University Post Print

Tweet

N.B.: When citing this work, cite the original article.

A STUDENT'S T FILTER FOR HEAVY TAILED PROCESS AND MEASUREMENT NOISE

Michael Roth, Emre Özkan, Fredrik Gustafsson

Division of Automatic Control, Linköping University, Linköping, Sweden.
E-mail: {roth, emre}@isy.liu.se, fredrik.gustafsson@liu.se

ABSTRACT

We consider the filtering problem in linear state space models with heavy tailed process and measurement noise. Our work is based on Student's *t* distribution, for which we give a number of useful results. The derived filtering algorithm is a generalization of the ubiquitous Kalman filter, and reduces to it as special case. Both Kalman filter and the new algorithm are compared on a challenging tracking example where a maneuvering target is observed in clutter.

Index Terms— Student's *t* distribution, Kalman filter, robustness, outliers

1. INTRODUCTION

The aim of this document is to derive a filter for linear state space models that is robust towards outliers in both process and measurement noise, simple to implement, and applicable to high dimensional problems. Therefore, we investigate Student's *t* distribution [1], which is particularly suited to model the phenomena and noise behavior we want a filter to cope with, for instance: measurement outliers from unreliable sensors, target maneuvers in tracking applications, or errors that occur when working with linearized nonlinear systems.

Filtering in linear state space models is a mature field with the Kalman filter [2,3] in its center for several decades. Apart from being the optimal Bayesian filter for linear Gaussian systems, the Kalman filter is the best (in terms of the mean squared estimation error) linear unbiased estimator [2] for arbitrary noise. Consequently, any achievable performance gain must come from a nonlinear filter. Our starting point is, as can be for the Kalman filter, the evolution of random vectors throughout time. The particular choice of the *t* distribution in conjunction with minor approximate adjustments results in a nonlinear filter which, as it turns out, reduces to the Kalman filter for one choice of parameters, in the same way as the *t* distribution converges to the Gaussian for infinite degrees of freedom. Moreover, the filter is easy to implement and compares to the Kalman filter in terms of computational demand.

Earlier efforts to derive filters based on the *t* distribution were made in [4,5] but did not receive much attention in the signal processing community. Our approach differs considerably from [4,5] in that we include careful approximation steps to overcome the shortcomings of the *t* distribution. Several recent publications address the filtering problem with heavy tailed measurement noise only. Whereas [6] and [7] suggest the use of variational approximations, an optimization approach using ℓ_1 regularization is sketched in [8]. Common to [6–8] is the assumption of well behaved process noise. This is easily violated in realistic scenarios, e.g. tracking of maneuvering targets, and leaves the methods of [6–8] debatable for many applications.

We think that a filter that allows for heavy tailed distributions for noises and states is of substantial practical value. One promising application is approximate nonlinear filtering, where conventional algorithms (EKF and related [2, 9, 10], UKF [11]) involve strong simplifications of the underlying model while keeping the assumption that the state is Gaussian throughout time. Consequences include diverging estimates. The more flexible particle filters [12,13] are usually not applicable in high dimensions. The here suggested filter addresses heavy tailed distributions in a natural manner, while keeping the computational cost low.

The outline of the paper is as follows. Sec. 2 introduces Student's *t* distribution and a number of useful results, based upon which a filtering algorithm is developed in Sec. 3. A simulation study in Sec. 4 is used to show the potential of the derived filter. Concluding remarks are given in Sec. 5.

2. STUDENT'S T DISTRIBUTION

Let $y \in \mathbb{R}^d$ be a zero-mean random vector with Gaussian [14] distribution $\mathcal{N}(0, \Sigma)$ and covariance Σ. Let $w > 0$, $w \in \mathbb{R}$ have a Gamma distribution $\text{Gam}(\frac{\nu}{2}, \frac{\nu}{2})$, in shape/rate parametrization [15]. Then

$$x = \mu + w^{-\frac{1}{2}}y \qquad (1)$$

admits the multivariate *t* distribution [1] $\text{St}(\mu, \Sigma, \nu)$ with mean vector μ, scale matrix Σ, and degrees of freedom (dof) parameter ν. Using $\Delta^2 = (x - \mu)^T \Sigma^{-1}(x - \mu)$ and the Gamma function [16], its probability density function (pdf)

can be written as

$$p(x) = \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(\nu\pi)^{d/2}} \frac{1}{\sqrt{\det(\Sigma)}} \left(1 + \frac{\Delta^2}{\nu}\right)^{-\frac{\nu+2}{2}}. \quad (2)$$

For the remainder of this paper, we abbreviate the above pdf by $\mathrm{St}(x; \mu, \Sigma, \nu)$.

The t distribution reduces to the Gaussian as ν tends to infinity, thus includes it as a special case. A number of convenient properties are shared by both and can be easily derived. As for the pdf of affine mappings of t variables [1] we have that $z = Ax + b$, with appropriate A and b, admits

$$p(z) = \mathrm{St}(z; A\mu + b, A\Sigma A^T, \nu). \quad (3)$$

The dof parameter remains unaltered. Turning to random vectors $x_1 \in \mathbb{R}^{d_1}$ and $x_2 \in \mathbb{R}^{d_2}$ that are jointly t distributed with

$$p(x_1, x_2) = \mathrm{St}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}, \nu\right), \quad (4)$$

the marginal pdf of x_1 can be computed by applying a linear transformation with $A = \begin{bmatrix} I & 0 \end{bmatrix}$ to (4):

$$p(x_1) = \mathrm{St}(x_1; \mu_1, \Sigma_{11}, \nu). \quad (5)$$

From (4) and (5), and considerably more effort, follows the conditional pdf

$$p(x_1|x_2) = \mathrm{St}(x_1; \mu_{1|2}, \Sigma_{1|2}, \nu_{1|2}), \quad (6a)$$

with the parameters

$$\nu_{1|2} = \nu + d_2, \quad (6b)$$

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \quad (6c)$$

$$\Sigma_{1|2} = \frac{\nu + \Delta_2^2}{\nu + d_2}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T) \quad (6d)$$

and $\Delta_2^2 = (x_2 - \mu_2)^T \Sigma_{22}^{-1}(x_2 - \mu_2)$. The above expression (6a) is again a t pdf. This result is in accordance with literature: [17] gives the expression without further details, [15] lists the results and asks for a derivation in the exercises. The monograph [1] misleadingly states that x_1 given x_2 is in general not t distributed. In order to clarify this statement, derivations of all given results are provided in a companion document [18] to this paper. It can be seen that the conditional mean (6c) corresponds to the Gaussian conditional mean [14]. The matrix parameter (6d) is a scaled version of the conditional covariance in the Gaussian case, which is recovered as ν tends to infinity. In contrast to the Gaussian, $\Sigma_{1|2}$ depends on x_2. Also, the dof parameter in (6b) increases.

The covariance of a t random variable [1] is $\frac{\nu}{\nu-2}\Sigma$ for $\nu > 2$. Furthermore, the moments of the t distribution are very much influenced by the tail behavior. Therefore, a t pdf with $\nu = 3$ appears more peaked around the mean than a Gaussian with identical covariance. Turning to independence

and uncorrelatedness, we investigate (4) with $\Sigma_{12} = 0$. Using the covariance result we know that x_1 and x_2 are uncorrelated. A closer look at $p(x_1, x_2)$, however, reveals that the joint pdf does not factor. Hence, uncorrelated random variables that are jointly t distributed are in general not independent, in contrast to the Gaussian case. Quadratic forms $x^T \Sigma^{-1} x/d$ follow an F-distribution [15] if $p(x) = \mathrm{St}(x; 0, \Sigma, \nu)$, a fact which can be used to define probability regions and gates in target tracking (similar to the use of the chi-squared distribution in the Gaussian case).

3. STUDENT'S T FILTER

In this section, we derive a filter that is fully based on the t distribution, i.e. both measurement and process noise distribution have heavy tails. The filtering pdf $p(x_k|y_{1:k})$ is a t density at each time k. First, we present an exact solution under somewhat restrictive modeling assumptions. In a second stage, the algorithm is adjusted to cope with more general scenarios. We consider linear state space systems with state difference and measurement equation given by

$$x_{k+1} = F_k x_k + v_k, \quad (7a)$$

$$y_k = H_k x_k + e_k. \quad (7b)$$

Here, $x_k \in \mathbb{R}^{d_x}$ is the state at time k, $y_k \in \mathbb{R}^{d_y}$ the measurement, v_k and e_k process and measurement noise of appropriate dimension, respectively. The system matrices F_k and H_k are known, and the initial state x_0 and noises are marginally t distributed. F_k and H_k might be obtained by linearization of a nonlinear state space model, which discloses how the algorithm can be applied to nonlinear problems.

3.1. Exact filter

Time update: Assume that the noise is such that for each k we have the following joint density:

$$p(x_k, v_k|y_{1:k}) = \mathrm{St}\left(\begin{bmatrix} x_k \\ v_k \end{bmatrix}; \begin{bmatrix} \hat{x}_{k|k} \\ 0 \end{bmatrix}, \begin{bmatrix} P_{k|k} & 0 \\ 0 & Q_k \end{bmatrix}, \eta_k\right). \quad (8)$$

Here, the state x_k and v_k are uncorrelated, in general not independent, which is in accordance with the assumptions that are used to derive the Kalman filter [3] from a least-squares perspective. The parameters of the predictive pdf $p(x_{k+1}|y_{1:k})$ can be computed by a linear transformation and subsequent marginalization:

$$\hat{x}_{k+1|k} = F_k \hat{x}_{k|k}, \quad (9a)$$

$$P_{k+1|k} = Q_k + F_k P_{k|k} F_k^T. \quad (9b)$$

The dof parameter is η_k retained, and we obviously recover the Kalman filter time update.

Measurement update: Assume that the predicted state and the measurement noise have the joint pdf

$$p(x_k, e_k|y_{1:k-1})$$
$$= \text{St}\left(\begin{bmatrix} x_k \\ e_k \end{bmatrix}; \begin{bmatrix} \hat{x}_{k|k-1} \\ 0 \end{bmatrix}, \begin{bmatrix} P_{k|k-1} & 0 \\ 0 & R_k \end{bmatrix}, \eta_{k-1}\right). \quad (10)$$

A linear transformation then gives the joint density of state and measurement

$$p(x_k, y_k|y_{1:k-1}) = \quad (11)$$
$$\text{St}\left(\begin{bmatrix} x_k \\ y_k \end{bmatrix}; \begin{bmatrix} \hat{x}_{k|k-1} \\ H_k\hat{x}_{k|k-1} \end{bmatrix}, \begin{bmatrix} P_{k|k-1} & P_{k|k-1}H_k^T \\ H_kP_{k|k-1} & S_k \end{bmatrix}, \eta_{k-1}\right)$$

with $S_k = H_kP_{k|k-1}H_k^T + R_k$. The conditional pdf of state given measurements $p(x_k|y_{1:k}) = \text{St}(x_k; \hat{x}_{k|k}, P_{k|k}, \eta_k)$ can be obtained using (6a). The updated parameters are

$$\eta_k = \eta_{k-1} + d_y, \quad (12a)$$
$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + P_{k|k-1}H_k^TS_k^{-1}(y_k - H_k\hat{x}_{k|k-1}), \quad (12b)$$
$$P_{k|k} = \frac{\eta_{k-1} + \Delta_{y,k}^2}{\eta_{k-1} + d_y}(P_{k|k-1} - P_{k|k-1}H_k^TS_k^{-1}H_kP_{k|k-1}) \quad (12c)$$

where $\Delta_{y,k}^2 = (y_k - H_k\hat{x}_{k|k-1})^TS_k^{-1}(y_k - H_k\hat{x}_{k|k-1})$. The matrix parameter $P_{k|k}$ is a scaled version of the covariance in the Kalman filter, and depends nonlinearly on the observation. The actual filter consists of iterating through equations (9) and (12), initialized with appropriate $\hat{x}_0$, P_0, and η_0.

Because of the assumed noise properties in (8) and (10) we were able to apply the results of Sec. 2 to obtain an exact filter. The resulting estimate (12b) is optimal in several ways [2], being the mean, the mode, and a median of the conditional distribution. However, the parameter η_k increases in each filter iteration and both state and noise distributions converge to Gaussians. That is, after few steps we obtain a Kalman filter and lose the heavy tailed properties we initially wanted to retain.

3.2. Suggested heavy tailed filter

The requested noise conditions of the previous section, equations (8) and (10), are hardly ever met in filtering problems. Instead, the state and noises in (7) do not share common dof parameters, i.e.

$$p(x_0) = \text{St}(x_0; 0, P_0, \eta_0), \quad (13a)$$
$$p(v_k) = \text{St}(v_k; 0, Q_k, \gamma_k), \quad (13b)$$
$$p(e_k) = \text{St}(e_k; 0, R_k, \delta_k) \quad (13c)$$

with $\eta_0 \neq \gamma_k \neq \delta_k$ in general. The resulting conflict is here illustrated for a time update. Suppose at time k, we have

$$p(x_k|y_{1:k}) = \text{St}(x_k; \hat{x}_{k|k}, P_{k|k}, \eta_k), \quad (14)$$

and $p(v_k)$ given by (13b). Assuming independence of x_k and v_k yields a product for the joint pdf $p(x_k, v_k|y_{1:k})$, which is in general not t (not even elliptically contoured), and thus puts an early end to our filtering efforts. If $p(v_k)$ and $p(x_k|y_{1:k})$, however, shared a common dof parameter, say $\widetilde{\eta}_k$, we could get back to (8) by assuming that v_k and x_k are not correlated. We therefore suggest the approximation

$$p(x_k, v_k|y_{1:k}) \approx \text{St}\left(\begin{bmatrix} x_k \\ v_k \end{bmatrix}; \begin{bmatrix} \hat{x}_{k|k} \\ 0 \end{bmatrix}, \begin{bmatrix} \widetilde{P}_{k|k} & 0 \\ 0 & \widetilde{Q}_k \end{bmatrix}, \widetilde{\eta}_k\right) \quad (15)$$

which requires choosing $\widetilde{P}_{k|k}$, $\widetilde{Q}_k$, and $\widetilde{\eta}_k$. The merits of the approximation are substantial in that we can now apply the time update (9) of the exact filter. In a similar fashion, we approximate

$$p(x_k, e_k|y_{1:k-1})$$
$$\approx \text{St}\left(\begin{bmatrix} x_k \\ e_k \end{bmatrix}; \begin{bmatrix} \hat{x}_{k|k-1} \\ 0 \end{bmatrix}, \begin{bmatrix} \widetilde{P}_{k|k-1} & 0 \\ 0 & \widetilde{R}_k \end{bmatrix}, \overline{\eta}_{k-1}\right) \quad (16)$$

in the measurement update, which takes us back to (10). Hence, we can apply the measurement update (12) of the exact filter.

In summary, the two approximation steps (15) and (16) extend the exact filter of Sec. 3.1, and provide convenient closed form solutions for time and measurement update. While the actual approximations remain to be specified at this stage, we can already state that the filter deviates from an exact Bayesian solution. Nevertheless, such approximation efforts are common practice in many filters. Examples include the reduction of mixture components in Gaussian mixture filters [2] as well as the inherent Gaussian assumption in many nonlinear filters [9–11]. In our case, the approximations are mild in the sense that the approximated marginal densities remain t pdfs, yet with possibly altered dof.

We can adjust the filter such that at each time k the filtering pdf $p(x_k|y_{1:k})$ is heavy tailed, instead of converging to a Gaussian. This fact, as discussed in Sec. 1, makes the filter particularly suited for many applications.

The approximation steps (15) and (16) can be cast into the following generic problem: Given that z_1 and z_2 are uncorrelated with $p(z_1) = \text{St}(z_1; \mu_1, \Sigma_1, \nu_1)$ and $p(z_2) = \text{St}(z_2; \mu_2, \Sigma_2, \nu_2)$, approximate the joint density $p(z_1, z_2)$ as t pdf by choosing common dof $\widetilde{\nu}$ and the matrix parameter as $\text{blkdiag}(\widetilde{\Sigma}_1, \widetilde{\Sigma}_2)$. One approach is to first select $\widetilde{\nu}$. Among more systematic choices, an ad-hoc solution that preserves the heaviest tails among $p(z_1)$ and $p(z_2)$ is $\widetilde{\nu} = \min(\nu_1, \nu_2)$. Each marginal pdf is then considered in a second stage, i.e. $p(z_1)$ is approximated by $q(z_1) = \text{St}(z_1; \mu_1, \widetilde{\Sigma}_1, \widetilde{\nu})$. Reasoning that qualitative relations between $p(z_1)$ and $q(z_1)$ should be preserved, it suffices to assume that $\widetilde{\Sigma}_1 = c_1\Sigma_1$ with a scalar factor c_1. Again, numerous alternatives are conceivable, e.g. moment matching, fitting probability regions, fitting $p(z_1)$ to $q(z_1)$ on an ellipsoid. It should be noted that

certain concepts, e.g. minimization of the Kullback-Leibler divergence [14], are difficult to apply in the approximation step. This is due to the complicated form that the expressions take for t pdfs. An extensive discussion of options is beyond the scope of this paper.

4. SIMULATION STUDY

We consider the problem of tracking a maneuvering target that is observed in clutter. The target moves according to a constant velocity model [19] and its position is measured. The state space matrices and nominal parameters are given by

$$F = \begin{bmatrix} I_2 & TI_2 \\ 0 & I_2 \end{bmatrix}, H = \begin{bmatrix} I_2 & 0 \end{bmatrix}, Q = \begin{bmatrix} \frac{T^3}{3}I_2 & \frac{T^2}{2}I_2 \\ \frac{T^2}{2}I_2 & TI_2 \end{bmatrix} q,$$

$R = rI_2$, $q = 1$, $r = 100$, and $T = 0.5$. We consider 500 time steps for each filtering experiment. Outlier corrupted process and measurement noise are generated according to

$$v_k \sim \begin{cases} \mathcal{N}(0, Q), & \text{w.p. } 0.95 \\ \mathcal{N}(0, 1000Q), & \text{w.p. } 0.05 \end{cases} \quad (17a)$$

$$e_k \sim \begin{cases} \mathcal{N}(0, R), & \text{w.p. } 0.9 \\ \mathcal{N}(0, 100R), & \text{w.p. } 0.1 \end{cases} \quad (17b)$$

i.e. v_k and e_k are most of the times generated from a Gaussian with nominal covariance Q or R. About five percent of the process and ten percent of the measurement noise values are drawn from Gaussians with severely increased covariance. Obviously, there is a mismatch between this example and the assumed distributions in Sec. 3.2. We specifically choose this configuration to show how the filter performs under modeling uncertainty, with merely the nominal Q and R at hand. Simulated position trajectories exhibit abrupt turns and the target speeds contain large steps.

It should be noted that the above tracking problem is not trivially solved by tuning of noise parameters in a Kalman filter because both noise sources contain outliers. Also, the solutions given in [6–8] do not account for such conditions and may perform worse than a reasonably tuned Kalman filter.

We here compare three filters that are utilized to recover target velocity and position: KF – a Kalman filter with the nominal noise parameters Q and R. KF2 – a Kalman filter that is aware of the true covariances of the mixture distributions in (17), and hence the best linear filter [2] in the mean squared error sense. TF – the filter of Sec. 3.2, which is merely aware of the nominal noise covariances but assumes that state and both noises admit t pdfs with dof parameter 3. The necessary approximation steps are carried out by taking the minimum among two dof parameters and adjusting the matrix parameters such that the 80 percent probability regions are preserved. The noises are adjusted accordingly, e.g. $\mathcal{N}(0, Q)$ is replaced by $\text{St}(0, \widetilde{Q}, 3)$ with the same 80 percent probability region.
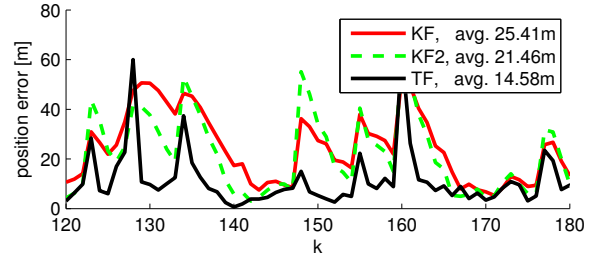
Fig. 1. Position error over time steps. Excerpt of one realization. All filters are subject to spikes in the position error. TF is the fastest to decrease the error again. Also shown is the average position error of the entire realization.

A typical realization is presented in Figure 1 where we show position errors [m] over time. All filters are subject to inevitable spikes in the position error, which either stem from a measurement outlier or a target maneuver. KF requires longer than its competitors to decrease the error. The optimal linear KF2 provides improved position estimates but is aware of the true noise statistics. TF merely knows the nominal Q and R, but still yields an improvement over KF2. It adjusts quickly after a peak in the position error, which can be explained by the observation dependent update of $P_{k|k}$ in (12). TF appears especially beneficial for severely maneuvering targets. Also, an averaged position error [m] of that particular run is displayed in Figure 1. The errors in estimated target speed are comparable for all filters, yet KF2 outperforms KF and TF. The above behavior is confirmed in 1000 Monte Carlo runs. The averaged errors (over time steps and MC runs) are: KF 23.8, KF2 20.0, TF 14.5, for position [m], and KF 11.5, KF2 10.8, TF 11.5 for speed [$\frac{\text{m}}{\text{sec}}$]. The qualitative relations are also preserved in 1000 MC runs with randomized Q and R, i.e. $q = 10^s$ with uniform $s \sim \mathcal{U}(-2, 3)$, and $r = 10^t$ with $t \sim \mathcal{U}(-1, 2)$. Here, TF outperforms KF and KF2 in position and velocity error: KF 7.5, KF2 6.3, TF 5.0, for position [m], and KF 13.5, KF2 13.6, TF 12.9, for speed [$\frac{\text{m}}{\text{sec}}$].

5. CONCLUSION

Based on the t distribution, we derived a filter that is a non-linear generalization of the Kalman filter, yet as simple to implement. Its potential application lies in problems with heavy tailed process and measurement noise, e.g. filtering using linearized nonlinear state space models or maneuvering target tracking in clutter. We showed that the new filter outperforms the best linear estimator in a challenging tracking scenario, even though it does not know the underlying noise statistics. It performs particularly well for severely maneuvering targets. We have thereby shown that the new filter provides robustness towards outliers in process and measurement noise, and performs well under model mismatch.

6. REFERENCES

[1] Samuel Kotz and Saralees Nadarajah, *Multivariate t Distributions and Their Applications*, Cambridge University Press, Feb. 2004.

[2] Brian D.O. Anderson and John B. Moore, *Optimal Filtering*, Prentice Hall, June 1979.

[3] Thomas Kailath, Ali H. Sayed, and Babak Hassibi, *Linear Estimation*, Prentice Hall, Apr. 2000.

[4] Richard J. Meinhold and Nozer D. Singpurwalla, "Robustification of Kalman filter models," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 479–486, June 1989.

[5] F.J. Girón and J.C. Rojano, "Bayesian Kalman filtering with elliptically contoured errors," *Biometrika*, vol. 81, no. 2, pp. 390–395, June 1994.

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

[7] Robert Piché, Simo Sarkka, and Jouni Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *Proceedings of MLSP*, Sept. 2012.

[8] Jacob Mattingley and Stephen Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 50 –61, May 2010.

[9] Magnus Nørgaard, Niels K. Poulsen, and Ole Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627–1638, Nov. 2000.

[10] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[11] S.J. Julier, J.K. Uhlmann, and H.F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *American Control Conference, 1995. Proceedings of the*, 1995, vol. 3, pp. 1628–1632 vol.3.

[12] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[13] F. Gustafsson, "Particle filter theory and practice with positioning applications," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 25, no. 7, pp. 53–82, 2010.

[14] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, Aug. 2006.

[15] Morris H. DeGroot, *Optimal Statistical Decisions*, Wiley-Interscience, WCL edition, Apr. 2004.

[16] Milton Abramowitz and Irene A Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55, Dover publications, 1965.

[17] Gary Koop, *Bayesian Econometrics*, Wiley-Interscience, July 2003.

[18] Michael Roth, "On the multivariate t distribution," Tech. Rep., Division of automatic control, Linköping university, Apr. 2013.

[19] Yaakov Bar-Shalom, X. Rong Li, and Thiagalingam Kirubarajan, *Estimation with Applications to Tracking and Navigation*, Wiley-Interscience, June 2001.