Posterior linearisation filter for non-linear state transformation noises

Matti Raitoharju, Roland Hostettler, Simo Särkkä

***Abstract*—This paper is concerned with discrete time Kalman-type filtering with state transition and measurement noises that may be non-additive or non-linearly transformed. More specifically, we extend the iterative estimation algorithm Posterior Linearization Filter (PLF) for estimation with this kind of noises. The approach solves the prediction and update step simultaneously, which allows to use the PLF iterations to improve the estimation in the non-linear state transition model. The proposed algorithm also produces single step fixed-lag smoothing estimates. We show in examples how the proposed approach can be used with non-Gaussian state transition noises and non-linearly transformed state transition noises.**

***Index Terms*—Nonlinear estimation, Kalman filtering, Fixed-lag smoother, Non-additive noise**

I. Introduction

The estimation of dynamic state from noisy measurements is a problem that arises in many fields of science and engineering, for example, target tracking, audio signal processing, or finance [1], [2]. In the Bayesian framework the system is described using a state transition model and a measurement model. Using these models, we can obtain the posterior probability density function (PDF) that describes the state given measurements available in an optimal manner. The computation of the PDF can be divided into two parts, first prediction with state transition model to obtain prior, then update with measurements to obtain posterior [3].

In general, that is when the model is non-linear or non-Gaussian, the posterior PDF cannot be calculated analytically. In these situations, particle filters are one option [4], but they have often high computational burden. Gaussian filters approximate the posterior PDF using a Gaussian and may be accurate enough when the posterior is unimodal [1], [3]. For non-linear estimation a Gaussian filter may use different ways to linearize non-linear models to obtain the posterior. The General Gaussian Filter (GGF) uses a statistical linearization for linearization [1]. However, the statistical linearization is in general case intractable and approximations have to be used. One of most common ways to make approximate statistical linearization is to use the unscented transform [5], [6]. This filter is called Unscented Kalman Filter (UKF).

The GGF makes the linearization for prediction at the posterior of previous time step and for the update at the prior of the current time step. In [3] it was shown that it is beneficial to make the linearization at the posterior for the measurement update. Because the posterior being unknown, the Posterior Linearization Filter (PLF) approximates the update with respect to the posterior iteratively. However, the prediction in PLF is still done as in GGF.

A smoother uses measurements from future in addition from past and present. Rauch-Tung-Striebel (RTS) smoother is an optimal smoother for linear-Gaussian case [7]. It performs the smoothing so that it first does the filtering and then a backward pass to obtain the smoothed estimates for the whole track. The PLF was extended for smoothing in [8]. Posterior Linearization Smoother (PLS) can be applied only for systems with additive and Gaussian noises. A fixed-lag smoother uses measurements from a fixed number of steps from the future [1].

In this paper, we extend PLF so that the state transition model may have non-additive noises anditerative posterior linearization is used in the prediction. To achieve this we do the prediction and update simultaneously instead of having two separate phases. The algorithm also produces a single step fixed-lag smoothed estimates of the state.

The rest of the paper is organized as follows. In Section III the problem is formulated. Section II presents related work. Section IV shows the proposed approach. Section V contains the examples and Section VI concludes the article.

II. Related work

A. Posterior linearization filter

In [3] PLF was presented and it was shown that the posterior linearization scheme minimizes the expected Kullback-Leibler Divergence (KLD). PLF uses state transition model of form

$$x_i = f_i(x_{i-1}) + e_i, \tag{1}$$

where $f_i(\cdot)$ is called state transition function and e_i is normal distributed with zero mean and covariance R. The measurement model is

$$y_i = h_i(x_i) + \varepsilon_i, \tag{2}$$

where $h_i(\cdot)$ is called measurement function and ε is normal distributed with zero mean and covariance Q.

The prediction in PLF is identical to the prediction of GGF:

$$\mu_{i|i-1} = \int f_i\left(x_i\right) p_{\mathrm{N}}\left(x_i|\mu_{i-1|i-1}, P_{i-1|i-1}\right) \mathrm{d}x_i \tag{3}$$

$$P_{i|i-1} = \int \left(f_i\left(x_i\right) - \mu_{i|i-1}\right)\left(f_i\left(x\right) - \mu_{i|i-1}\right)^T \cdot p_{\mathrm{N}}\left(x_i|\mu_{i-1|i-1}, P_{i-1|i-1}\right) \mathrm{d}x_i + Q, \tag{4}$$

M. Raitoharju and S. Särkkä are with the Department of Electrical Engineering and Automation (EEA), Aalto University, Finland. Email: {matti.raitoharju,simo.sarkka}@aalto.fi, Matti Raitoharju is also with Patria, Finland, Roland Hostettler is with Department of Electrical Engineering, Uppsala University, Sweden. Email: roland.hostettler@angstrom.uu.se

where $p_{\mathrm{N}}(x|\mu_{i|i}, P_{i|i})$ is the PDF of a multivariate normal distribution.

PLF updates the state using a measurement model iteratively, using previous estimate as a point for statistical linearization. An iteration of the update is:

$$\hat{y}_i = \int h_i(x_i)\, p_{\mathrm{N}}\left(x_i|\mu_i^L, P_i^L\right)\mathrm{d}x_i \tag{5}$$

$$\Phi = \int \left(h_i(x_i) - \hat{y}_i\right)\left(h_i(x_i) - \hat{y}_i\right)^T \cdot p_{\mathrm{N}}\left(x_i|\mu_i^L, P_i^L\right)\mathrm{d}x \tag{6}$$

$$\Psi = \int \left(x_i - \mu_{i|i-1}\right)\left(h_i(x_i) - \hat{y}_i\right)^T \cdot p_{\mathrm{N}}\left(x_i|\mu_i^L, P_i^L\right)\mathrm{d}x_i \tag{7}$$

$$A = \Psi^T\left(P_i^L\right)^{-1} \tag{8}$$

$$b = \hat{y} - A\mu_i^L \tag{9}$$

$$\Omega = \Psi - AP_i^L A^T \tag{10}$$

$$S = AP_{i|i-1}A^T + R \tag{11}$$

$$K = P_{i|i-1}A^T S^{-1} \tag{12}$$

$$\mu_i^{L+1} = \mu_{i|i-1} + K\left(y - A\mu_{i|i-1} - b\right) \tag{13}$$

$$P_i^{L+1} = P_{i|i-1} - KSK^T, \tag{14}$$

where μ_k^L and P_k^L define the mean and covariance of the distribution that is used for the next statistical linearization. The algorithm is iterative and the mean (13) and covariance (14) are used as the linearization parameters for next iteration. The final μ_k^L and P_k^L are used as the parameters of the posterior. If the iteration is done only once, the result is identical to the result of GGF. The integrals in equations (3)-(7) can be approximated with a suitable method, such as the unscented transform [6].

B. Nonlinearly transformed measurement noises

In [9], [10] a measurement equation had a form

$$y_i = h_i(x_i) + g_i(\varepsilon_i). \tag{15}$$

The nonlinear transformation $g_i(\cdot)$ is chosen so that if ε_i has standard normal distribution, then $g_i(\varepsilon_i)$ has desired distribution. With a non-iterative Kalman filter extension that linearizes at prior the noise part $g(\varepsilon_i)$ in (15) has always a same linearization and could be replaced with a Gaussian. With PLF the linearization changes between each iteration and the PLF can be used with transforming functions that effectively allow to use non-Gaussian noises with a Kalman type filter. The formulation in [10] uses an augmented state, where the measurement noise variables ε_i are augmented with the state dimensions in the update to obtain posterior estimate and the posterior linearization for those too. In the augmented state formulation, the posterior covariance may be singular, which causes problems in computation of the statistical linearization of the PLF. To avoid this a small diagonal covariance was added for P_i^L to make it always nonsingular.

In [9] it was shown that the $g_i(\cdot)$ can be chosen to be

$$g_i(\varepsilon_i) = F^{-1}(\Phi(\varepsilon_i)), \tag{16}$$

where Φ is the Cumulative Distribution Function (CDF) of a standard normal distribution and F^{-1} is the inverse CDF of the desired distribution. In [10] was also presented a method to determine function $g_i(\cdot)$ from samples. Function $g_i(\cdot)$ was built using piecewise cubic Hermite interpolating polynomials. This removes the step to find a analytical function and specific distribution for the noise.

III. Problem formulation

We use more general state transition model than in (1):

$$x_i = f_i(x_{i-1}, e_i) \tag{17}$$

and more general measurement model than in (2) and (15):

$$y_i = h_i(x_i, \varepsilon_i). \tag{18}$$

We assume that e_i is normal distributed with mean μ_{e_i} and covariance Q_i and ε is normal distributed with mean μ_{ε_i} and covariance R_i.

Filtering problems with (18) could be solved with the PLF presented in [10], but (17) would not get any help from using PLF as it does the prediction as GGF. PLS has different linearization points for the state transition model between iterations, but it is only for additive Gaussian noises.

IV. Proposed approach

In [10] the measurement noise ε_i was augmented with the state so that both were estimated in iterations. In this paper we will also augment the state transition noise in the augmented state $\hat{x}_i$

$$\hat{x}_i = \begin{bmatrix} x_{i-1} \\ e_i \\ \varepsilon_i \end{bmatrix}. \tag{19}$$

We propose to do the filtering taking the state transition noise also into the state and instead of making GGF predict – update procedure we make the update at using a measurement that contains the state transition function

$$y_i = \hat{h}_i(\hat{x}_i) = h_i(f_i(x_{i-1}, e_i), \varepsilon_i). \tag{20}$$

This augmented state can be updated with the version of PLF that was introduced in [10]. The iterations involving the augmented state (19) and measurement model (20) produce the $\hat{x}_i|y_{1:i}$, but we have to note that the first elements of the augmented state are $x_{i-1}|y_{1:i}$, which is not the posterior of the time instance i, but a single step fixed-lag smoothed state. One interpretation for the proposed algorithm is that we make the linearization for the prior in the smoothed posterior of the previous state to obtain more accurate linearization. The mean and covariance of the posterior $x_i|y_{1:i}$ can be then obtained integrals

$$\mu_{x_{i|i}} = \int f_i(x,e) p_N\left(\begin{bmatrix} x \\ e \end{bmatrix} \middle| \begin{bmatrix} \mu_{x_{i-1|i}} \\ \mu_{e_{i|i}} \end{bmatrix}, \operatorname{cov}\begin{bmatrix} x_{i-1|i} \\ e_{i|i} \end{bmatrix}\right)\mathrm{d}x \tag{21}$$

$$\operatorname{cov} x_{i|i} = \int \left(f_i(x,e) - \mu_{x_{i|i}}\right)\left(f_i(x,e) - \mu_{x_{i|i}}\right)^T \times p_N\left(\begin{bmatrix} x \\ e \end{bmatrix} \middle| \begin{bmatrix} \mu_{x_{i-1|i}} \\ \mu_{e_{i|i}} \end{bmatrix}, \operatorname{cov}\begin{bmatrix} x_{i-1|i} \\ e_{i|i} \end{bmatrix}\right)\mathrm{d}x, \tag{22}$$

which can be approximated using, for example, UKF. These moments can be then used for initialization of the next time step.

A. Algorithm

In [9] the augmented state with nonlinearly transformed (15) was solved using a variation of Recursive Update Kalman Filter (RUKF) [11], [12], however there were noted some situations, where the recursive update did not produce satisfactory results. In [10] a modification of [3] was used. In [10] it was noted that there were sometimes numerical problems caused by singular covariance matrices of the augmented state. The matrices are made nonsingular by adding a diagonal covariance matrix to the covariance that is used for computing the statistical linearization. For this there is a parameter α for which we use value 10^{-3}. There is also variable β that limits the step lengths if we are estimating with noises with unit variance that occur when we are doing estimation for situations presented in Section II-B. This improves stability especially when the noise models are empirically determined as the noise terms make at most one sigma steps. Algorithm 1 shows the main general algorithm and Algorithm 2 shows the iterations within PLF updates. We will use 10 iterations ($L_{\max}$) in our examples.

Algorithm 1: Filtering for non-additive models

Input: Prior mean μ_0 and covariance P_0
for $i = 1 : i_{\max}$ **do**
  Build augmented mean $\hat{\mu}_i = \begin{bmatrix} \mu_{i-1} \\ \mu_{e_i} \\ \mu\varepsilon_i \end{bmatrix}$
  Build augmented covariance
  $\hat{P}_i = \begin{bmatrix} P_{i-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & R_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & Q_i \end{bmatrix}$
  Build augmented measurement function $\hat{h}(\hat{x})$
  Update augmented state using Algorithm 2
  (Store single-step fixed-lag estimates if necessary)
  Use (21) to obtain posterior mean μ_i
  Use (22) to obtain posterior covariance P_i
end

Algorithm 2: PLF for augmented states

for $L = 1 : L_{\max}$ **do**
  $\hat{y} = \int h_k(x)\, p_{\mathrm{N}}\left(x|\hat{\mu}^L_{i,\tau}, \hat{P}^L_{i,\tau} + \alpha\, \mathrm{diag}(\hat{P}^L_{i,\tau})\right) \mathrm{d}x$
  $\Phi = \int (h_k(x) - \hat{y}_k)(h_k(x) - \hat{y}_k)^T \cdot p_{\mathrm{N}}\left(x|\hat{\mu}^L_{i,\tau}, \hat{P}^L_{i,\tau} + \alpha\, \mathrm{diag}(\hat{P}^L_{i,\tau})\right) \mathrm{d}x$
  $\Psi = \int (x - \hat{\mu}^L_{i,\tau})(h_k(x) - \hat{y}_k)^T \cdot p_{\mathrm{N}}\left(x|\hat{\mu}^L_{i,\tau}, \hat{P}^{L+1}_{i,\tau} + \alpha\, \mathrm{diag}(\hat{P}^{L+1}_{i,\tau})\right) \mathrm{d}x$
  $A = \Psi^T \left(\hat{P}^L_{i,\tau} + \alpha\, \mathrm{diag}(\hat{P}^L_{i,\tau})\right)^{-1}$
  $b = \hat{y} - A\mu^L_k$
  $\Omega = \Psi - A(\hat{P}^L_{i,\tau} + \alpha\, \mathrm{diag}(\hat{P}^L_{i,\tau}))A^T$
  $S = A\hat{P}^0_{i,\tau}A^T$
  $K = \hat{P}^0_{i,\tau}A^T S^{-1}$
  $\Delta = K\left(y - A\hat{\mu}^0_{i,\tau} - b\right)$
  if *We are estimating transformed noises with unit variance as in Section II-B* **then**
    $\beta = \min\left(\frac{1}{\max \Delta_{[n+1:\hat{n}]}}, 1\right)$
  else
    $\beta = 1$
  end
  $\hat{\mu}^{L+1}_{i,\tau} = (1-\beta)\hat{\mu}^L_{i,\tau} + \beta\left(\hat{\mu}^0_{i,\tau} + \Delta\right)$
  $\hat{P}^{L+1}_{i,\tau} = \hat{P}^0_{i,\tau} - KSK^T$
end

V. Examples

A. Student-t state transition model

In this example, we show how the proposed algorithm can handle a situation with a non-Gaussian state transition noise. We consider a state transition model with Student-t distribution and a single time step update. Compared to normal distribution Student-t has heavier tails. The state is unidimensional, x_0 has standard normal distribution and the state transition model is

$$x_1 = x_0 + e, \tag{23}$$

where e is Student-t distributed with 3 degrees of freedom and scaled with $\frac{1}{\sqrt{3}}$ so that it has unit variance. Thus we use transformation function (16) with F^{-1} being the inverse CDF of the Student-t distribution. Measurement model is

$$y = x + \varepsilon, \tag{24}$$

where ε has standard normal distribution. The combined measurement is then

$$y = x_0 + F^{-1}(\Phi(\hat{e})) + \varepsilon, \tag{25}$$

where $\hat{e} \sim (0, 1)$. The measurement value is 10. We solve this one step problem, with the proposed algorithm and with linear Kalman filter, assuming that the prior $x_0 + e$ has normal distribution with same mean and variance as the true prior ($\mathrm{N}(0, 2)$) and then the true posterior by computing the PDF in a dense grid. Top part of Figure 1 shows the distributions before update and the measurement likelihood. The measurement is clearly far away from the prior mean. Bottom part of Figure 1 shows the posteriors computed in the three different ways described above. Because the state transition model has heavier tails than the measurement, the measurement dominates the posterior. Figure 1 shows how the posterior of the proposed algorithm that takes the heavy-tailness of the prior into account has posterior distribution close to the true posterior unlike the Kalman posterior.

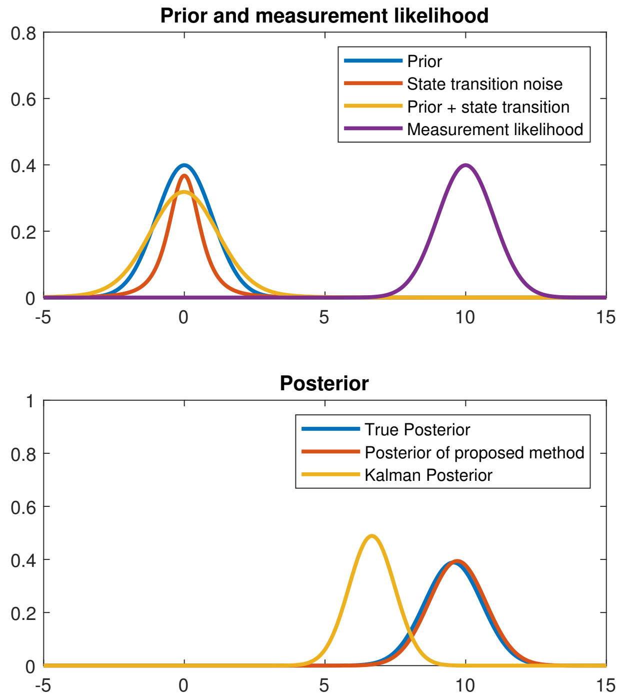

Fig. 1. Probabilities associated with the example 1

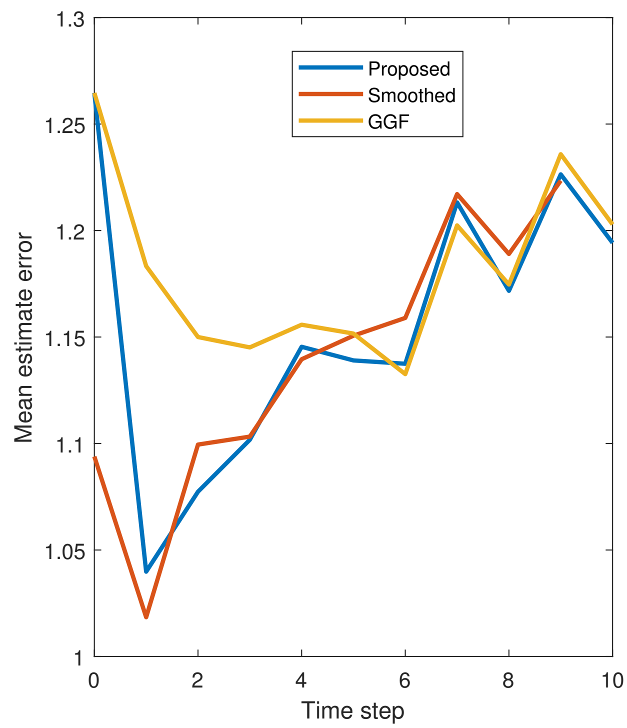

Fig. 2. Probabilities associated with the example 1

B. Nonlinear motion model

In this example, we consider a filtering example with state transition model

$$x_i = x_{i-i} + \|x_{i-i}\| \, e_i, \tag{26}$$

where $e_i \sim \mathrm{N}(0, I)$ and thus the magnitude of state transition noise is directly proportional to the norm of the state. The state contains 2 variables and the measurement model is linear

$$y_i = x_i + \varepsilon_i, \tag{27}$$

where $\varepsilon_i \sim \mathrm{N}(0, I)$. Because the measurement model is linear, there are no benefit of using PLF over GGF. We simulated 1000 tracks of length 20 time steps using prior mean $\mathbf{0}$ and diagonal unit covariance and used UKF for approximating integrals in the algorithms.

Figure 2 shows the mean estimation errors of the means at each time step of the filtering and also the smoothed estimates. Initial prior at time step 0 is identical for both filtering algorithms, but the smoother improves it. Towards the end of the track the state has moved away from the origin and the measurement variance is much smaller than the variance of state transition variance and so the state estimate is close to the measurement, but in the beginning of the track as the state is close to origin and $\|x_{i-i}\| \, e_i$ has small variance compared to R_i the proposed algorithm outperforms GGF and the smoother is even more accurate, as expected.

C. Accelerometer positioning

In this example, we consider a one dimensional tracking using an accelerometer. The state transition model is

$$\begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} = \begin{bmatrix} x_{i-1,1} + x_{i-1,2} + \frac{x_{i,3}}{2} \\ x_{i-1,2} + x_{i-1,3} \\ x_{i-1,3} + e_i \end{bmatrix}, \tag{28}$$

where e_i is Student-t distributed. The prior is $\mathrm{N}(0, 10^{-6} I)$. The used measurement model is an accelerometer measurement

$$y_i = x_{i,3} + \varepsilon_i, \tag{29}$$

where $\varepsilon \sim \mathrm{N}(0, 1)$.

We made two tests, in the first one the e_i had 3 degrees of freedom and variance 3. In the second test, we used 2 degrees of freedom for the Student-t distribution. This distribution does not have finite variance. As the models are linear, we used the linear Kalman filter as reference. For the 3 degrees of freedom we used $Q = 3$ so that the Kalman filter has the correct variance. For the second case, as the variance is not defined, we tested different variances $\{1, 10, 100, 1000, 10000\}$.

We simulated the 10 step tracks 10000 times and results for mean errors for each state variable are given in Table I

TABLE I
RESULTS OF STUDENT-t ACCELEROMETER TEST WITH 3 DEGREES OF FREEDOM

Method	x_1	x_2	x_3
Proposed	11.68	2.30	0.68
Kalman	11.90	2.32	0.70

TABLE II
RESULTS OF STUDENT-t ACCELEROMETER TEST WITH 2 DEGREES OF FREEDOM

Method	x_1	x_2	x_3
Proposed	11.88	2.35	0.69
Kalman $Q = 1$	14.16	2.65	0.87
Kalman $Q = 10$	12.37	2.40	0.75
Kalman $Q = 100$	12.57	2.40	0.79
Kalman $Q = 1000$	12.61	2.43	0.79
Kalman $Q = 10000$	12.61	2.43	0.79

and in Table II. From results we can see that with 3 degrees of freedom the results obtained were slightly better with the proposed method than with Kalman filter that assumed the noise Gaussian instead of Student-t. While making this test we noticed that with large prior variance the methods produced even more similar results. When the state transition noise had 2 degrees of freedom the proposed method outperformed the Kalman filter with all tested variances.

VI. CONCLUSIONS AND FUTURE WORK

We proposed a modification to the PLF that uses the state transition and measurement functions simultaneously to improve the estimation accuracy in situations where the state transition noise is not additive. In our examples, we showed how the proposed method can be used with non-Gaussian state transition noise, how the proposed method outperformed method in literature in situation where the state noise was dependent on the state and how the proposed algorithm is applicable to inertial positioning when state transition model noise is not Gaussian. We also showed how the algorithm produces a single time step fixed lag smoothed estimates.

As future work, the algorithm could be extended to work with multiple time steps in a similar manner as accumulated state density filter [13], [14]. In that case, we would be able to obtain a smoothed estimates from multiple time steps.

REFERENCES

[1] S. Särkkä, *Bayesian Filtering and Smoothing*. United Kingdom: Cambridge University Press, 2013.

[2] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, 2001.

[3] A. F. García-Fernández, 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.

[4] M. 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, 2002.

[5] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[6] E. Wan and R. Van der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. Adaptive Systems for Signal Process., Commun., and Control Symp.. AS-SPCC.*, 2000, pp. 153–158, doi:10.1109/ASSPCC.2000.882463.

[7] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA journal*, vol. 3, no. 8, pp. 1445–1450, 1965.

[8] A. F. García-Fernández, L. Svensson, and S. Särkkä, "Iterated posterior linearization smoother," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 2056–2063, 2017.

[9] M. Raitoharju, R. Piché, and H. Nurminen, "A systematic approach for kalman-type filtering with non-Gaussian noises," in *2016 19th International Conference on Information Fusion (FUSION)*. IEEE, 2016, pp. 1853–1858.

[10] M. Raitoharju, H. Nurminen, D. Cilden-Guler, and S. Särkkä, "Kalman filtering with empirical noise models," in *2021 International Conference on Localization and GNSS (ICL-GNSS)*, 2021, pp. 1–7.

[11] R. Zanetti, "Recursive update filtering for nonlinear estimation," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1481–1490, June 2012, doi:10.1109/TAC.2011.2178334.

[12] Y. Huang, Y. Zhang, N. Li, and L. Zhao, "Design of sigma-point Kalman filter with recursive updated measurement," *Circuits, Systems, and Signal Process.*, pp. 1–16, August 2015, doi:10.1007/s00034-015-0137-y.

[13] W. Koch and F. Govaers, "On accumulated state densities with applications to out-of-sequence measurement processing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2766–2778, 2011.

[14] F. Govaers and W. Koch, "Generalized solution to smoothing and out-of-sequence processing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 1739–1748, 2014.