Laplace state space filter with exact inference and moment matching

Julian Neri, Philippe Depalle, Roland Badeau

LAPLACE STATE SPACE FILTER WITH EXACT INFERENCE AND MOMENT MATCHING

Julian Neri[*] *Philippe Depalle*[*] *Roland Badeau*[†]

[*]McGill University, CIRMMT, Montréal, Canada.
[†]LTCI, Télécom Paris, Institut Polytechnique de Paris, France.

ABSTRACT

We present a Bayesian filter for state space models with Laplace-distributed observation noise that is robust to heavy-tailed and outlier-ridden univariate time-series data. We analytically derive a closed-form expression of the exact posterior for a Laplace likelihood conditioned on a Gaussian prior. Posterior statistics are propagated forward in time by a proxy Gaussian density. The forward Kullback-Leibler divergence from the posterior to the Gaussian is minimized by matching their moments. The proposed method supports both linear and non-linear systems, and has a fast recursive structure analogous to the Kalman filter that enables online inference. Results show that the new method outperforms existing approximate inference methods, especially in challenging scenarios where the system's parameters are uncertain.

***Index Terms*—** heavy-tailed noise, Kalman filter, state estimation, Laplace distribution, Bayesian inference

1. INTRODUCTION

State space models are probabilistic time-series models that have proven beneficial in a wide range of disciplines including control systems, audio processing, target tracking, finance, autonomous navigation, and neuroscience [1]. The Kalman filter is an efficient and numerically accurate algorithm for exactly inferring the latent state sequence of a model with Gaussian state and observation noise [2]. However, time-series data often exhibit non-Gaussian noise, consisting of outliers, glint noise [3], sensor failure, and extended periods of drastically increased noise levels [4]. The Kalman filter's performance severely degrades in all of these cases because the Gaussian assumption does not hold. A model with heavy-tailed observation noise is better at representing outliers [5], but does not admit a closed-form recursive filtering solution, rendering exact inference intractable. Although sequential Monte Carlo methods like particle filtering can estimate posterior statistics for arbitrary state space models [6], there is no limit to the number of samples needed to attain a certain degree of estimation quality, they are subject to the curse of dimensionality, and the reliability of their estimates are hard to assess.

Recently, there has been much interest in fast and reliable deterministic estimators for state space models with heavy-tailed observation noise. Deterministic methods typically exploit a recursive structure akin to the Kalman filter for speed and use alternative cost functions, heuristics, or approximate inference at each time step for robust state estimation. Examples include the maximum correntropy filter [7], minimax-based filters [5, 8], or variational inference-based filters [9–12]. In particular, recent research on estimating models with Laplace-distributed observation noise demonstrated that the model is robust to extreme outliers and a variety of heavy-tailed-distributed noises. The Laplace distribution has many applications including glint noise modeling and differential privacy [13]. Existing methods rely on iterative optimization algorithms to approximate the posterior, including convex optimization [14], majorization-minimization [15, 16], Huber cubature filtering [17], and variational inference with Gaussian scale mixtures [10].

In this paper, we propose a new Bayesian filter for state space models with Laplace observation noise that provides fast, high quality estimation. Unlike previous methods, we analytically derive the exact posterior moments for the Laplace state space model and propagate them forward in time through a Gaussian density. Minimizing the forward Kullback-Leibler (KL) divergence between the posterior and the Gaussian is achieved by simply matching their moments. Converting the exact posterior into a Gaussian enables a fast and numerically stable recursive algorithm, analogous to the Kalman filter. The proposed filter is robust to impulsive outliers, heavy-tailed noise, and extended periods of increased noise. Using these locally-exact posterior statistics provides superior results over other approximate inference methods such as particle filtering and variational Bayes. Moreover, the filter elegantly extends to models with non-linear dynamics.

This paper is organized as follows. Section 2 defines the Laplace state space model. Section 3 details the new filtering method. Section 4 extends the method to non-linear dynamics. Results are presented in Section 5. Finally, Section 6 concludes the paper and proposes future research.

The following notation is used throughout the paper: bold uppercase denotes a matrix; bold lowercase denotes a vector; $\mathbf{x}_{1:n}$ denotes the set $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$; $\mathbb{E}[\mathbf{x}]$ is the expected value of $\mathbf{x}$; $\text{cov}[\mathbf{x}, \mathbf{y}] = \mathbb{E}\left[\mathbf{x}\mathbf{y}^\mathsf{T}\right] - \mathbb{E}[\mathbf{x}]\,\mathbb{E}[\mathbf{y}]^\mathsf{T}$ is the covariance of $\mathbf{x}$ and $\mathbf{y}$; $\mathbf{I}$ is the identity matrix; $|.|$ is the absolute value; $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate Gaussian density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathcal{L}(x|a, b)$ is a Laplace density with mean a and scale b [18].

2. PROBABILISTIC MODEL

State space models assume that a sequence of observable M-dimensional data $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$ are generated from a latent variable sequence of D-dimensional states $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ whose probabilistic dynamics are governed by a first-order Markov chain. The joint probability for a state space model is

$$p(\mathbf{Y}, \mathbf{X}) = p(\mathbf{y}_1|\mathbf{x}_1)p(\mathbf{x}_1)\textstyle\prod_{n=2}^{N} p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1})\,, \quad (1)$$

where $p(\mathbf{y}_n|\mathbf{x}_n)$ is the emission probability, $p(\mathbf{x}_n|\mathbf{x}_{n-1})$ is the transition probability, and $p(\mathbf{x}_1)$ is the initial state's prior probability.

State space models can assume either linear or non-linear dynamics. For linear models, states are transformed over adjacent

*Thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN- 2018-05662) for funding.

times by $D \times D$ system dynamics matrix $\mathbf{A}$, and to the observable space by $M \times D$ output matrix $\mathbf{C}$. For non-linear models, states are transformed over adjacent times by function $h(\mathbf{x}_n)$ and to the observable space by function $g(\mathbf{x}_n)$.

The Laplace state space model assumes that univariate ($M = 1$) observation y_n is corrupted by zero-mean Laplace-distributed noise with scale R, while state $\mathbf{x}_n$ is corrupted by zero-mean Gaussian noise with $D \times D$ covariance matrix $\mathbf{Q}$. Output vector $\mathbf{c}$ is $1 \times D$. The initial state has $D \times 1$ mean $\mathbf{m}_0$ and $D \times D$ covariance $\mathbf{P}_0$. The initial, transition, and emission probabilities for the linear model are, respectively,

$$p(\mathbf{x}_1) = \mathcal{N}(\mathbf{x}_1 | \mathbf{m}_0, \mathbf{P}_0)\,, \tag{2}$$

$$p(\mathbf{x}_n | \mathbf{x}_{n-1}) = \mathcal{N}(\mathbf{x}_n | \mathbf{A}\mathbf{x}_{n-1}, \mathbf{Q})\,, \tag{3}$$

$$p(y_n | \mathbf{x}_n) = \mathcal{L}(y_n | \mathbf{c}\mathbf{x}_n, R) \tag{4}$$

$$= \frac{1}{2R} \exp\left(-\frac{|y_n - \mathbf{c}\mathbf{x}_n|}{R}\right). \tag{5}$$

The Laplace distribution's variance is $2R^2$ [18]. The linear model is addressed first, then extended to support non-linear dynamics.

3. PROPOSED METHOD

Filtering in state space models refers to inferring the marginal posterior probability of state $\mathbf{x}_n$ given every observation up to time n,

$$p(\mathbf{x}_n | y_{1:n}) = \frac{p(y_n|\mathbf{x}_n)p(\mathbf{x}_n|y_{1:n-1})}{p(y_n|y_{1:n-1})}. \tag{6}$$

This involves computing the predictive distribution $p(\mathbf{x}_n|y_{1:n-1})$ and the marginal likelihood $p(y_n|y_{1:n-1})$, then updating the posterior $p(\mathbf{x}_n|y_{1:n})$. For linear Gaussian state space models this is an exact inference algorithm, called the Kalman filter [1, 2, 19]. However, for any model with non-Gaussian noise exact inference over all times is not analytically tractable and thus requires approximations.

Remarkably, we can analytically derive the exact marginal posterior for the Laplace state space model at time n per equation (6) given that the posterior from time $n-1$ is approximated by a Gaussian $\tilde{p}$ with mean $\boldsymbol{\mu}_{n-1}$ and covariance $\mathbf{V}_{n-1}$. Our exposition of the proposed filter begins with the predictive distribution.

3.1. Predictive distribution

The predictive distribution is given by

$$p(\mathbf{x}_n|y_{1:n-1}) = \int p(\mathbf{x}_n|\mathbf{x}_{n-1})\tilde{p}(\mathbf{x}_{n-1}|y_{1:n-1})d\mathbf{x}_{n-1}$$
$$= \mathcal{N}(\mathbf{x}_n|\mathbf{m}_{n-1}, \mathbf{P}_{n-1})\,, \tag{7}$$

where the predictive mean and covariance matrix are, respectively,

$$\mathbf{m}_{n-1} = \mathbf{A}\boldsymbol{\mu}_{n-1}\,, \tag{8}$$

$$\mathbf{P}_{n-1} = \mathbf{A}\mathbf{V}_{n-1}\mathbf{A}^\mathsf{T} + \mathbf{Q}\,. \tag{9}$$

3.2. Marginal likelihood

The marginal likelihood is the normalizing factor in the denominator of equation (6). It is found by integrating out $\mathbf{x}_n$ from the joint distribution. Crucially, we can actually solve the integral analytically. In doing so, we get the following closed-form expression.

$$p(y_n|y_{1:n-1}) = \int p(y_n|\mathbf{x}_n)p(\mathbf{x}_n|y_{1:n-1})d\mathbf{x}_n$$
$$= \int \mathcal{L}(y_n|\mathbf{c}\mathbf{x}_n, R)\mathcal{N}(\mathbf{x}_n|\mathbf{m}_{n-1}, \mathbf{P}_{n-1})d\mathbf{x}_n$$
$$= \frac{\Phi_n^{(-)} + \Phi_n^{(+)}}{4R} \exp\left(-\frac{\tilde{y}_n^2}{2S_n}\right), \tag{10}$$

where the residual is $\tilde{y}_n = y_n - \widehat{y}_n$ and

$$\widehat{y}_n = \mathbf{c}\mathbf{m}_{n-1}\,, \tag{11}$$

$$S_n = \mathbf{c}\mathbf{P}_{n-1}\mathbf{c}^\mathsf{T}\,, \tag{12}$$

$$\Phi_n^{(\pm)} = \operatorname{erfcx}\left(\frac{\sqrt{S_n}}{\sqrt{2R^2}} \pm \frac{\tilde{y}_n}{\sqrt{2S_n}}\right). \tag{13}$$

The scaled complementary error function $\operatorname{erfcx}(x) = e^{x^2}\operatorname{erfc}(x)$ is available in many programming languages. It avoids underflow or overflow errors associated with directly computing the product of e^{x^2} and the complementary error function $\operatorname{erfc}(x)$ as defined in [20].

Observation y_n has the following mean and covariance with respect to the marginal likelihood $p(y_n|y_{1:n-1})$:

$$\mathbb{E}\left[y_n\right] = \widehat{y}_n\,, \tag{14}$$

$$\operatorname{cov}[y_n, y_n] = S_n + 2R^2. \tag{15}$$

For the purpose of Bayesian model comparison, the model evidence is approximately $p(\mathbf{Y}) \approx \prod_{n=1}^{N} p(y_n|y_{1:n-1})$.

3.3. Marginal posterior

Substituting equations (5), (7), and (10) into (6) gives the marginal posterior. Finally, state $\mathbf{x}_n$ has the following mean and covariance with respect to the marginal posterior $p(\mathbf{x}_n|y_{1:n})$:

$$\mathbb{E}\left[\mathbf{x}_n\right] = \mathbf{m}_{n-1} + \mathbf{k}_n\delta_n\,, \tag{16}$$

$$\operatorname{cov}[\mathbf{x}_n, \mathbf{x}_n] = \mathbf{P}_{n-1} + \mathbf{k}_n\mathbf{k}_n^\mathsf{T}\Delta_n\,, \tag{17}$$

where we have defined

$$\mathbf{k}_n = \mathbf{P}_{n-1}\mathbf{c}^\mathsf{T}R^{-1}\,, \tag{18}$$

$$\delta_n = \frac{\Phi_n^{(-)} - \Phi_n^{(+)}}{\Phi_n^{(-)} + \Phi_n^{(+)}}\,, \tag{19}$$

$$\Delta_n = \frac{1}{\Phi_n^{(-)} + \Phi_n^{(+)}}\left(\frac{4\Phi_n^{(-)}\Phi_n^{(+)}}{\Phi_n^{(-)} + \Phi_n^{(+)}} - \sqrt{\frac{8R^2}{\pi S_n}}\right). \tag{20}$$

Studying the behavior of δ_n and Δ_n is helpful in understanding how this model is robust to outliers and heavy-tailed noise. Figure 1 shows the values of δ_n and Δ_n as functions of the residual, and for different ratios of latent noise variance $\mathbf{P}_{n-1}$ to observed noise scale R. Since Δ_n is proportional to R, the normalized value $\Delta_n R^{-1}$ is shown. Both functions are symmetric about $\tilde{y}_n = 0$, and non-linear. The shape of δ_n and Δ_n depends on $\mathbf{c}$, $\mathbf{P}_{n-1}$, and R, and generally controls the model's sensitivity to outliers.

The function δ_n is "S"-shaped, being approximately linear near $\tilde{y}_n = 0$, and being compressed to limits as the residual tends to positive or negative infinity. The function Δ_n is bell-shaped for larger variances (a loose model), and is approximately a box function for variances close to zero (a tight model). As $\tilde{y}_n$ goes to $+\infty$, $\delta_n = 1$, and $\Delta_n = 0$. As $\tilde{y}_n$ goes to $-\infty$, $\delta_n = -1$, and $\Delta_n = 0$. Considering equations (16) and (17), when the difference between the prediction and the data is large (and thus tending the residual towards positive or negative infinity), the expected latent state covariance remains as the predicted value $\mathbf{P}_{n-1}$, while the mean is $\mathbf{m}_{n-1} + \mathbf{k}_n$ (when $\tilde{y}_n = +\infty$) or $\mathbf{m}_{n-1} - \mathbf{k}_n$ (when $\tilde{y}_n = -\infty$).

For the Kalman filter, δ_n is simply equal to the residual $\tilde{y}_n$ and is thus a line with slope altered by the Kalman gain. Consequently, outliers severely affect the Kalman filter because they pull the expected value of the state far from the prediction. Using the Laplace observation density, there is a limit to an observation's influence on the state update ($\mathbf{m}_{n-1} \pm \mathbf{k}_n$).

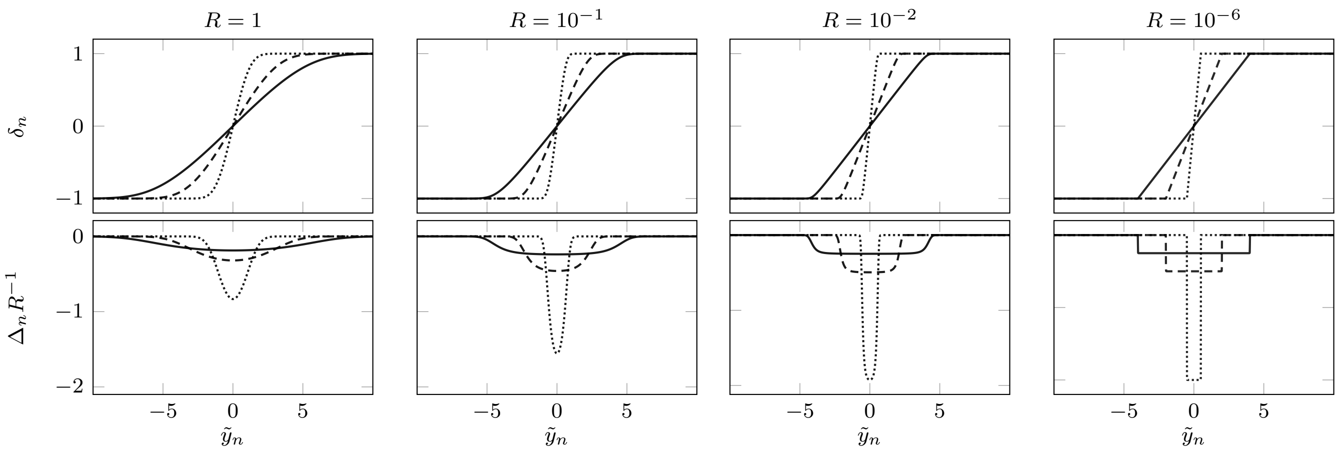

Fig. 1. Plots of δ_n and $\Delta_n R^{-1}$. The dotted, dashed, and solid lines correspond to $\mathbf{P}_{n-1} = R/2, 2R$, and $4R$, respectively, and $\mathbf{c} = 1$.

3.4. Moment matching

Now we turn to the problem of approximating the posterior with a Gaussian density $\tilde{p}$ so that the filter can be implemented recursively. To do so, we minimize the forward KL divergence [19, 21]

$$\mathrm{KL}\left(p(\mathbf{x}_n|y_{1:n})\|\tilde{p}(\mathbf{x}_n|y_{1:n})\right) , \tag{21}$$

where $\tilde{p}(\mathbf{x}_n|\mathbf{y}_{1:n}) = \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_n, \mathbf{V}_n)$. The forward KL divergence is minimized when the moments of $\tilde{p}$ match p. Therefore, we set the mean and covariance of the Gaussian equal to the mean and covariance of the posterior from equations (16) and (17):

$$\boldsymbol{\mu}_n = \mathbb{E}\left[\mathbf{x}_n\right] , \tag{22}$$

$$\mathbf{V}_n = \mathrm{cov}[\mathbf{x}_n, \mathbf{x}_n] . \tag{23}$$

This completes the filter for time n. At time $n + 1$, the filter begins again at the prediction step with equations (8) and (9).

4. NON-LINEAR MODEL

The proposed method elegantly extends to non-linear models. In this case, a state is transformed by a non-linear function $h(\mathbf{x}_{n-1})$ then output to the observable space by a non-linear function $g(\mathbf{x}_n)$. Adopting a local linearization approach similar to [22], we linearize $h(.)$ around the previous state estimate $\boldsymbol{\mu}_{n-1}$ and $g(.)$ around the predicted mean $\mathbf{m}_{n-1}$. Considering the results of the previous section, the predicted state, predicted observation, system dynamics matrix, and output matrix are approximated by, respectively,

$$\mathbf{m}_{n-1} = h(\boldsymbol{\mu}_{n-1}) , \qquad \widehat{y}_n = g(\mathbf{m}_{n-1}) , \tag{24}$$

$$\mathbf{A}_n = \frac{dh}{d\mathbf{x}_{n-1}}\bigg|_{\boldsymbol{\mu}_{n-1}} , \qquad \mathbf{c}_n = \frac{dg}{d\mathbf{x}_n}\bigg|_{\mathbf{m}_{n-1}} , \tag{25}$$

where $\mathbf{A}_n$ and $\mathbf{c}_n$ are Jacobian matrices. On the contrary, variational inference-based approaches are violated by this kind of local linearization and require the monitoring of convergence or iterative optimization of the variational lower bound [23].

5. RESULTS

Experiments were conducted to test the performance of the proposed method at estimating linear and non-linear models. Performance was evaluated according to the root mean square error (RMSE),

$$\mathrm{RMSE} = \sqrt{(\mathbb{E}\left[\mathbf{x}_n\right] - \mathbf{x}_n)^{\mathsf{T}}(\mathbb{E}\left[\mathbf{x}_n\right] - \mathbf{x}_n)} , \tag{26}$$

where $\mathbf{x}_n$ is the true state and $\mathbb{E}\left[\mathbf{x}_n\right]$ is the state estimate. In all experiments, the RMSE was averaged over 1000 Monte Carlo runs.

5.1. Linear system results

Signals generated from linear systems were used as observations for experiments designed to gage the performance of the proposed filter and how it compares to existing filters: a Kalman Filter (KF), a variational Bayes (VB)-based filter that represents the Laplace distribution as a Gaussian scale mixture (*e.g.* [10]), and a bootstrap particle filter (PF) [24] that assumes Laplace noise. PF used 100 samples from the Laplace state space model and was approximately 100 times slower than the other methods. More samples would improve the estimation quality but further increase computation time.

Each observation was generated according to first-order linear state space dynamics. The true latent state $\mathbf{x}_n$ was two-dimensional ($D = 2$) and oscillated according to the system dynamics matrix

$$\mathbf{A} = \begin{bmatrix} \cos(2\pi fT) & -\sin(2\pi fT) \\ \sin(2\pi fT) & \cos(2\pi fT) \end{bmatrix} , \tag{27}$$

where the frequency was randomly set in the range $f \in [0, \frac{4Fs}{N}]$ Hz, the sampling rate was $Fs = 8$ kHz, and the sampling period was $T = 1/Fs$ seconds. Elements of the output matrix were independently sampled from a zero-mean unit-variance Gaussian distribution: $c_j \sim \mathcal{N}(0, 1)$, for $j \in [1\,..\,D]$. The rows of the output matrix were normalized such that $|\sum_j c_j| = 1$. The latent noise covariance matrix was diagonal $\mathbf{Q} = \mathbf{I}\sigma^2$ with variance $\sigma^2 = 10^{-4}$.

The first experiment involved observations that were corrupted by Laplace-distributed noise with a variance of 10^{-2}. Each method inferred the latent state sequence using the correct model parameters, but assumed an initial state prior covariance matrix of $\mathbf{P}_0 = \mathbf{I}$. Figure 2a shows how the RMSE evolved for each method. The proposed filter quickly settled to the lowest RMSE value.

The second experiment involved signals with severe outliers. The outliers were sampled randomly from a Gaussian distribution with a variance of 1, spaced over time at an average rate of 1 in 10 samples. Otherwise, the signal had light Gaussian noise with a variance of 10^{-5}. Each filter assumed that the observed noise variance

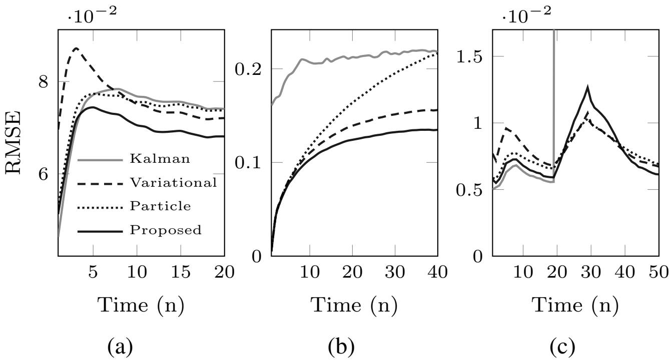

Fig. 2. Linear filtering results given a signal with: (a) Laplace noise, (b) Gaussian outliers, and (c) Gaussian noise of variance 1 for $20 \leq n \leq 30$, and 10^{-4} otherwise.

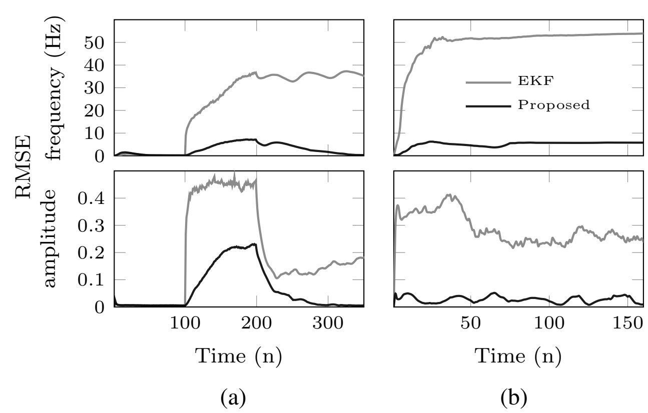

Fig. 3. Non-linear filtering (tracking) of sinusoidal model parameters given a signal with: (a) increased noise in the range $100 \leq n \leq 200$, and (b) outliers.

was 10^{-4}. Figure 2b shows the results. As expected, the Kalman filter is severely affected by outliers. The three filters that assume Laplace noise are more robust to the outliers. Specifically, the proposed method consistently exhibited the lowest RMSE.

The third experiment was designed to evaluate each filter's reaction to a sudden and lasting increase in the observation noise variance. Observations were corrupted by Gaussian noise with variance of 1 for $20 \leq n \leq 30$, and 10^{-4} otherwise. Each filter was given the correct system dynamics and observed noise variance equal to 10^{-4}. RMSE plots for this experiment are shown in Figure 2c. The RSME for the Kalman filter was lowest until $n = 20$, since it is the optimal estimator for signals corrupted by Gaussian noise, but grew rapidly at $n = 20$ (out of view) because the Kalman filter closely followed the increased noise. The three filters that assumed Laplace noise were only lightly affected by the period of increased noise. VB was least affected because the auxiliary variable of the Gaussian scale mixture distribution acts as a time-varying weight on R. The proposed filter exhibited an RMSE close to the Kalman filter before the period of increased noise, good performance during it, and the lowest RMSE afterwards.

5.2. Non-linear system results

A non-linear filtering problem was given to the proposed method and compared with the extended Kalman filter (EKF) [1]. The problem involved the estimation of a sinusoid's instantaneous amplitude a_n, frequency f_n in Hz, and phase ϕ_n in radians according to the sinusoidal model, which assumes that a univariate observation y_n is given by

$$\phi_n = \phi_{n-1} + 2\pi T f_n \,, \tag{28}$$

$$y_n = a_n \sin(\phi_n) + \varepsilon_n \,, \tag{29}$$

where ε_n is Gaussian noise.

While there are different ways to design a non-linear state space model for amplitude and frequency estimation [25], we used linear latent dynamics and transformed the state through a non-linear function to coincide with equations (28) and (29). State $\mathbf{x}_n = \begin{bmatrix}\phi_n, & f_n, & a_n\end{bmatrix}^\mathsf{T}$ evolved linearly according to $h(.)$ and was transformed by a non-linear function $g(.)$ to generate the univariate observation y_n:

$$h(\mathbf{x}_{n-1}) = \begin{bmatrix}\phi_{n-1} + 2\pi T f_{n-1}, & f_{n-1}, & a_{n-1}\end{bmatrix}^\mathsf{T} \,, \tag{30}$$

$$g(\mathbf{x}_n) = a_n \sin(\phi_n). \tag{31}$$

Observations were generated from the model with initial values set to $\phi_0 = 0$, $f_0 = 300$ Hz, and $a_0 = 1$, and with latent noise variances set to 10^{-9} for ϕ_n, 10^{-1} for f_n, and 10^{-5} for a_n. Each filter inferred the latent state sequence using the aforementioned model. VB was excluded from this experiment because it did not reliably extend to the non-linear model. PF with 100 particles tracked similarly to the proposed method, but took around 100 times longer.

The first test involved data that was corrupted by Gaussian noise with a variance of 10^{-3} for $100 \leq n \leq 200$, and 0 otherwise. RMSE plots in Figure 3a show that EKF failed to estimate the latent state accurately during and after the noise. The frequency and amplitude estimates did not return to reasonable estimates after the noise had vanished. Conversely, the proposed filter was stable through the period of increased noise. Afterwards, it returned to more accurate tracking of the frequency and amplitude.

The second test involved outlier-ridden data. The outliers were sampled randomly from a Gaussian distribution with a variance of 1, and were spaced over time at an average rate of 1 in 20 samples. Otherwise, the signal had light Gaussian noise with a variance of 10^{-5}. RMSE plots in Figure 3b show that the impulsive outliers degraded the tracking accuracy of EKF. On the other hand, the proposed method filtered the outliers well, exhibiting a low RMSE for both frequency and amplitude estimations.

6. CONCLUSION

In this paper, we introduced a new Bayesian filter for state space models with Laplace-distributed univariate time-series data. We derived the exact posterior moments for a joint distribution composed of a Laplace likelihood and a Gaussian prior. Then, we showed how these moments can be propagated forward in time by a proxy Gaussian density. Our approach led to a filtering algorithm that is simple to implement for both linear and non-linear dynamical systems and has the same low computational complexity as the Kalman filter. Its estimation quality is better than the variational Bayes approach and particle filter. We illustrated this quality through a variety of experiments. Regarding future work, we will first extend the new filter so that it supports multivariate Laplace-distributed observations. Then we will devise a smoother for approximating the marginal posterior given the entire data sequence, which will enable the automatic learning of the Laplace state space model's parameters.

7. REFERENCES

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

[2] R. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[3] G. Hewer, R. Martin, and J. Zeh, "Robust preprocessing for Kalman filtering of glint noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 23, no. 1, pp. 120–128, Jan. 1987.

[4] R. Pearson, "Outliers in process modeling and identification," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 55–63, Jan. 2002.

[5] I. Schick and S. Mitter, "Robust recursive estimation in the presence of heavy-tailed observation noise," *The Annals of Statistics*, vol. 22, no. 2, pp. 1045–1080, 1994.

[6] 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, Feb. 2002.

[7] B. Chen, X. Liu, H. Zhao, and J. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017.

[8] C. Masreliez and R. Martin, "Robust Bayesian estimation for the linear model and robustifying the Kalman filter," *IEEE Transactions on Automatic Control*, vol. 22, pp. 361–371, June 1977.

[9] Y. Huang, Y. Zhang, Y. Zhao, and J. Chambers, "A novel robust Gaussian–Student's t mixture distribution based Kalman filter," *IEEE Transactions on Signal Processing*, vol. 67, no. 13, pp. 3606–3614, July 2019.

[10] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian, and J. Chambers, "Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 10, pp. 2082–2096, Oct. 2017.

[11] G. Agamennoni, J. 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, Oct. 2012.

[12] J. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, USA, Nov. 2007, pp. 1514–1519.

[13] S. Kotz, T. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations: A revisit with applications to communications, economics, engineering, and finance*, Springer Science+Business Media, LLC, 2001.

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

[15] H. Wang, H. Li, W. Zhang, and H. Wang, "Laplace ℓ_1 robust Kalman filter based on majorization minimization," in *20th International Conference on Information Fusion*, Xi'an, China, July 2017.

[16] H. Wang, H. Li, W. Zhang, J. Zuo, and H. Wang, "Laplace ℓ_1 robust Kalman smoother based on majorization minimization," in *AIAA Scitech 2019 Forum*, San Diego, USA, 2019.

[17] L. Cao, D. Qiao, and X. Chen, "Laplace ℓ_1 Huber based cubature Kalman filter for attitude estimation of small satellite," *Acta Astronautica*, vol. 148, pp. 48–56, July 2018.

[18] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*, John Wiley & Sons, Inc., 4th edition, 2011.

[19] C. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag New York, 2006.

[20] E. Ng and M. Geller, "A table of integrals of the error functions," *Journal of research of the National Bureau of Standards - B. Mathematical Sciences*, vol. 73B, no. 1, pp. 1–20, Jan.-Mar. 1969.

[21] T. Minka, "Divergence measures and message passing," Tech. Rep., Microsoft Research, 2005.

[22] P. Dash, G. Panda, A. Pradhan, A. Routray, and B. Duttagupta, "An extended complex Kalman filter for frequency measurement of distorted signals," in *2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings*, Jan. 2000, vol. 3.

[23] M. Chappell, A. Groves, B. Whitcher, and M. Woolrich, "Variational Bayesian inference for a nonlinear forward model," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 223–236, Jan. 2009.

[24] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F - Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, April 1993.

[25] A. Routray, A. Pradhan, and K. Rao, "A novel Kalman filter for frequency estimation of distorted signals in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 3, pp. 469–479, June 2002.