

# UPDATES IN BAYESIAN FILTERING BY CONTINUOUS PROJECTIONS ON A MANIFOLD OF DENSITIES

*Filip Tronarp and Simo Särkkä*

Aalto University, Department of Electrical Engineering and Automation, Finland

## ABSTRACT

In this paper, we develop a novel method for approximate continuous-discrete Bayesian filtering. The projection filtering framework is exploited to develop accurate approximations of posterior distributions within parametric classes of probability distributions. This is done by formulating an ordinary differential equation for the posterior distribution that has the prior as initial value and hits the exact posterior after a unit of time. Particular emphasis is put on exponential families, especially the Gaussian family of densities. Experimental results demonstrate the efficacy and flexibility of the method.

**Index Terms**— Projection filtering, continuous-discrete filtering, non-linear filtering, Bayesian state estimation.

## 1. INTRODUCTION

Filtering in stochastic differential equations (SDEs) [1] is an important problem in signal processing [2, 3, 4] and finance [5]. Most commonly, the SDE is measured in discrete time, hence the model is given by:

$$dX(t) = \mu(X(t), t) dt + \sigma(X(t), t) dB(t), \quad (1a)$$

$$Y(t_n) | X(t_n) \sim f(y(t_n) | X(t_n)), \quad (1b)$$

where  $\mu$  is the drift function,  $\sigma$  is the diffusion coefficient,  $B(t)$  is a vector of standard Brownian motions, and  $f(y(t_n) | X(t_n))$  is measurement model in form of the conditional probability density for  $Y(t_n) | X(t_n)$ .

If we let  $\mathcal{Y}(t) = \{y(\tau_n) : \tau_n \leq t\}$ , then the filtering problem is to compute the sequence of filtering densities  $f(x(t_n) | \mathcal{Y}(t_n))$ . When the system is linear and Gaussian, the exact filter is given by the Kalman filter [6]. For non-linear systems with Gaussian excitations, a classical approximation method is the extended Kalman filter [7], which uses Taylor series to linearise the system. It can be extended to numerical integration approaches [8, 9, 10, 11, 12]. The filter update in the aforementioned methods can be improved by iterative linearisation techniques [13, 14]. However, this requires conditional Gaussian models or specific conditions on the moments of  $Y(t_n)$  conditioned on  $X(t_n)$  [15].

Another line of work are the projection filters [16, 17], where the Kushner–Stratonovich equation is projected onto a parametric manifold of densities using differential geometric method [18], smoothing is also possible [19]. However, these require either continuous-time measurement or that the probability densities in the parametric manifold are conjugate prior to the measurement in Equation (1).

The contribution of this paper is to extend the projection filtering methodology to discrete-time measurements with arbitrary likelihoods. This is done by constructing a smooth map from prior to posterior, similar to tempering in Monte Carlo methods [20], giving a differential equation characterisation of the posterior. The methodology of [17] can then be used to project the posterior onto a manifold of probability densities. Emphasis is put on exponential families and in particular Gaussian projections. The method is demonstrated in a tracking example using  $\ell_1$  likelihoods and in a stochastic volatility model, where it is shown to outperform other common approaches.

### 1.1. Projections Onto A Parametric Probability Manifold

Our approach is based on differential geometric view on statistics [18, 17], that we review briefly here. Let  $\mathcal{P}$  be a set of probability densities on  $\mathcal{X} \subset \mathbb{R}^d$  and consider a parametrised subset of  $\mathcal{P}$ ,  $\mathcal{P}_\Theta$ ,  $\theta \in \Theta \subset \mathbb{R}^m$ . Note that if  $p \in \mathcal{P}$  then  $p^{1/2}$  is square integrable,  $p^{1/2} \in \mathcal{L}^2$ , and  $\mathcal{P}_\Theta$  can thus be identified as a sub-manifold of  $\mathcal{L}^2 \supset \mathcal{L}_\Theta^2$ . For  $p_\theta \in \mathcal{P}_\Theta$ , the tangent space at  $\theta$  is given by  $\{\partial_{\theta_i} p_\theta^{1/2}\}_{i=1}^m$ . Furthermore, if  $v = \frac{1}{2} p_\theta^{1/2} u \in \mathcal{L}^2$ , then its projection onto the tangent space of  $\mathcal{L}_\Theta^2$  at  $\theta$  is given by [17, Lemma 2.1]

$$\Pi_\theta v = \frac{1}{2} \mathbb{E}_\theta \left[ u \nabla_\theta^T \log p_\theta \right] g^{-1}(\theta) (\nabla_\theta \log p_\theta) p_\theta^{1/2}, \quad (2)$$

where  $g(\theta)$  is the Fisher information matrix of  $p_\theta$ . Suppose that a smooth curve  $p_\tau^{1/2} \in \mathcal{L}^2$ ,  $\tau \in [0, 1]$  is given by

$$\partial_\tau p_\tau^{1/2} = \mathcal{A}(p_\tau^{1/2}), \quad p_0^{1/2} = p_{\theta_0}^{1/2} \in \mathcal{L}_\Theta^2, \quad (3)$$

where  $\mathcal{A}$  is some operator. The projection of  $p_\tau^{1/2}$  onto  $\mathcal{L}_\Theta^2$  is then defined as [17]

$$\partial_\tau \hat{p}_{\theta(\tau)}^{1/2} = \Pi_{\theta(\tau)} \circ \mathcal{A}(\hat{p}_{\theta(\tau)}^{1/2}), \quad \hat{p}_{\theta(0)}^{1/2} = p_{\theta_0}^{1/2}, \quad (4)$$

Funding from Aalto ELEC Doctoral School and Academy of Finland (project 313708) is gratefully acknowledged.

where  $\circ$  denotes operator composition.

It is worth noting that the projection approximation in this form can only be used for continuous-time flows of probability densities. For example, by setting  $\mathcal{A}$  to be the operator corresponding to the Fokker-Planck equation associated with Equation (1a) we get an approximation to the prediction step of continuous-discrete Bayesian filter [17]. The update is problematic in this formulation, because it corresponds to an instantaneous jump in the posterior distribution. For that reason [16, 17] formulated the discrete-time updates only for conjugate models. In this paper we present a more general procedure to perform the updates in the projection formalism.

## 2. BAYESIAN PROJECTION UPDATE

In this section, we present the proposed projection filter and its specialisation to Gaussian manifolds.

### 2.1. A Smooth Mapping From Prior To Posterior

As discussed above, the prediction step of a Bayesian filter can be approximated by Equation (4) with  $\mathcal{A}$  selected to be the Fokker-Planck operator [17]. Hence, here we only need to consider the update step of the Bayesian filter.

Let  $\pi_{\theta_0} \in \mathcal{P}_\Theta$ ,  $\theta_0 \in \Theta$  be a predictive density and  $f(y | x)$  the likelihood. The filtering density is then given by:

$$\pi(x | y) = \frac{f(y | x)\pi_{\theta_0}(x)}{\int_{\mathcal{X}} f(y | x)\pi_{\theta_0}(x) dx}. \quad (5)$$

The goal is to find an element  $\theta_1 \in \Theta$  such that  $\pi_{\theta_1}$  approximates the exact posterior, Equation (5), well. In order to make use of Equation (4), we introduce a smooth mapping from prior to posterior:

$$p_\tau(x | y) : [0, 1] \rightarrow \mathcal{P} \supseteq \mathcal{P}_\Theta,$$

with  $p_0(x | y) = \pi_{\theta_0}(x)$  and  $p_1(x | y) = \pi(x | y)$ . An obvious candidate is given by

$$p_\tau(x | y) = \frac{[f(y | x)]^\tau \pi_{\theta_0}(x)}{\int_{\mathcal{X}} [f(y | x)]^\tau \pi_{\theta_0}(x) dx}. \quad (6)$$

Differentiating  $p_\tau$  with respect to  $\tau$  gives

$$\partial_\tau p_\tau(x | y) = (\ell(x) - \mathbb{E}_\tau[\ell(X)])p_\tau(y | x),$$

where  $\ell(x) \triangleq \log f(y | x)$  and  $\mathbb{E}_\tau$  is the expectation operator associated with  $p_\tau$ . Or for the square root:

$$\mathcal{A}_{X|Y}(u) = \frac{1}{2}(\ell(x) - \mathbb{E}_{u^2}[\ell(X)])u, \quad u \in \mathcal{L}^2, \quad (7a)$$

$$\partial_\tau p_\tau^{1/2}(x | y) = \mathcal{A}_{X|Y}(p_\tau^{1/2}(x | y)), \quad (7b)$$

where  $\mathbb{E}_{u^2}$  is the expectation with respect to  $u^2 \in \mathcal{P}$ .

### 2.2. Projecting The Posterior Onto A Manifold

Applying the projection operator, as defined in Equation (2), to Equation (7b) gives:

$$\begin{aligned} & \Pi_{\theta(\tau)} \circ \mathcal{A}_{X|Y}(\widehat{p}_{\theta(\tau)}^{1/2}) \\ &= \frac{1}{2} \widehat{\mathbb{E}} \left[ \ell(X) \nabla_\theta^\top \log \widehat{p}_{\theta(\tau)} \right] g^{-1}(\theta(\tau)) (\nabla_\theta \log \widehat{p}_{\theta(\tau)}) \widehat{p}_{\theta(\tau)}^{1/2}, \end{aligned}$$

where  $\widehat{\mathbb{E}}$ ,  $\widehat{\mathbb{C}}$ , and  $\widehat{\mathbb{V}}$  are expectation, cross-covariance, and covariance operators with respect to  $\widehat{p}_{\theta(\tau)}$ , respectively. On the other hand, we have

$$\partial_\tau \widehat{p}_{\theta(\tau)}^{1/2}(x) = \frac{1}{2} \widehat{p}_{\theta(\tau)}^{1/2} [\partial_\tau \theta(\tau)]^\top g(\theta(\tau)) g^{-1}(\theta(\tau)) \nabla_\theta \log \widehat{p}_{\theta(\tau)}.$$

Matching the expressions gives  $\partial_\tau \theta$  as

$$\partial_\tau \theta(\tau) = g^{-1}(\theta(\tau)) \widehat{\mathbb{E}} \left[ \nabla_\theta \log \widehat{p}_{\theta(\tau)} \ell(X) \right]. \quad (8)$$

If  $\widehat{p}_{\theta(\tau)}$  is in an exponential family of densities:

$$\widehat{p}_{\theta(\tau)}(x) = \exp(\theta^\top(\tau)T(x) - \kappa(\theta(\tau)))h(x), \quad (9)$$

then by standard results on exponential families [21],

$$\nabla_\theta \log \widehat{p}_{\theta(\tau)}(x) = T(x) - \nabla_\theta \kappa(\theta(\tau)) \quad (10a)$$

$$g(\theta(\tau)) = \nabla_\theta^2 \kappa(\theta(\tau)). \quad (10b)$$

Therefore,  $\theta(\tau)$  is given by

$$\partial_\tau \theta(\tau) = [\nabla_\theta^2 \kappa(\theta(\tau))]^{-1} \widehat{\mathbb{C}}[T(X), \ell(X)]. \quad (11)$$

Note that if  $\ell(X)$  is a polynomial in  $T(X)$ , then the right hand side of Equation (11) can be expressed in terms of derivatives of  $\kappa(\theta(\tau))$  [21]. When  $\ell(X)$  is linear in  $T(X)$ , exact posterior inference is retrieved, this is Theorem 1.

**Theorem 1.** *Let  $\ell(x) = \eta^\top(y)T(x) + c$ , for some constant  $c$  that doesn't depend on  $x$ . Then the projection update Equation (11) gives the exact posterior for a prior in the exponential family Equation (9).*

*Proof.* Plugging in the log-likelihood into Equation (11) gives

$$\begin{aligned} \partial_\tau \theta(\tau) &= [\nabla_\theta^2 \kappa(\theta(\tau))]^{-1} \widehat{\mathbb{C}}[T(X), \eta^\top(y)T(X)] \\ &= [\nabla_\theta^2 \kappa(\theta(\tau))]^{-1} \widehat{\mathbb{V}}[T(X)] \eta(y) = \eta(y), \end{aligned} \quad (12)$$

where we used  $\widehat{\mathbb{V}}[T(X)] = \nabla_\theta^2 \kappa(\theta(\tau))$  [21]. Therefore, solving Equation (12) gives

$$\theta(1) = \theta_0 + \eta(y), \quad (13)$$

which is the required result since Equation (9) is a conjugate prior and Equation (13) is the posterior parameter [21].  $\square$

### 2.2.1. The Gaussian Manifold

A particularly interesting and useful example of an exponential family is the Gaussian distributions:

$$\widehat{p}_{\theta(\tau)}(x) = \mathcal{N}(x; \mu(\tau), \Sigma(\tau)). \quad (14)$$

We use the parametrisation  $\theta^\top = [\mu^\top, (\text{vec } \Sigma)^\top]$ . The Fisher matrix and score vector ( $\nabla_{\theta} \log \widehat{p}_{\theta(\tau)}$ ) are then given by

$$g(\theta) = \text{blkdiag} \left[ \Sigma^{-1}, \frac{1}{2} \Sigma^{-1} \otimes \Sigma^{-1} \right],$$

$$\nabla_{\theta} \log \widehat{p}_{\theta(\tau)}(x) = \begin{bmatrix} \Sigma^{-1}(x - \mu) \\ \frac{1}{2} \text{vec}[\Sigma^{-1}(x - \mu)(x - \mu)^\top \Sigma^{-1}] \\ 0 \\ -\frac{1}{2} \text{vec}[\Sigma^{-1}] \end{bmatrix},$$

where  $\otimes$  is Kronecker's product. Straightforward calculations then give

$$\Pi_{\theta(\tau)} \circ \mathcal{A}_{X|Y}(\widehat{p}_{\theta(\tau)}^{1/2}) = \widehat{\mathbb{C}}[\ell(X), (X - \mu)] \Sigma^{-1} (x - \mu) \frac{1}{2} \widehat{p}_{\theta(\tau)}^{1/2} \\ + \frac{1}{2} (x - \mu)^\top \Sigma^{-1} \widehat{\mathbb{C}}[(X - \mu)(X - \mu)^\top, \ell(X)] \Sigma^{-1} (x - \mu) \\ \times \frac{1}{2} \widehat{p}_{\theta(\tau)}^{1/2} - \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \widehat{\mathbb{C}}[(X - \mu)(X - \mu)^\top, \ell(X)] \right] \frac{1}{2} \widehat{p}_{\theta(\tau)}^{1/2}.$$

On the other hand, differentiating the square root of Equation (14) gives

$$\partial_{\tau} \widehat{p}_{\theta(\tau)}^{1/2}(x) = \left( [\partial_{\tau} \mu]^\top \Sigma^{-1} (x - \mu) - \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \partial_{\tau} \Sigma \right] \right) \\ \times \frac{1}{2} \widehat{p}_{\theta(\tau)}^{1/2}(x) + \frac{1}{2} (x - \mu)^\top \Sigma^{-1} [\partial_{\tau} \Sigma] \Sigma^{-1} (x - \mu) \frac{1}{2} \widehat{p}_{\theta(\tau)}^{1/2}(x).$$

Matching terms gives:

$$\partial_{\tau} \mu = \widehat{\mathbb{E}}[(X - \mu) \ell(X)], \quad (16a)$$

$$\partial_{\tau} \Sigma = \widehat{\mathbb{E}}[(X - \mu)(X - \mu)^\top (\ell(X) - \mathbb{E}[\ell(X)])]. \quad (16b)$$

Using Stein's lemma [22], Equation (16) can be written as

$$\partial_{\tau} \mu = \Sigma \widehat{\mathbb{E}}[\nabla_X \ell(X)], \quad (17a)$$

$$\partial_{\tau} \Sigma = \Sigma \widehat{\mathbb{E}}[\nabla_X^2 \ell(X)] \Sigma. \quad (17b)$$

Another form of Equation (16) is given by [3, Exercise 5.3]:

$$\partial_{\tau} \mu = \Sigma \nabla_{\mu} \widehat{\mathbb{E}}[\ell(X)], \quad (18a)$$

$$\partial_{\tau} \Sigma = \Sigma \nabla_{\mu}^2 \widehat{\mathbb{E}}[\ell(X)] \Sigma. \quad (18b)$$

Which of the formulations Equations (16) to (18) is most appropriate depends on the context. Equation (18) is convenient when  $\ell(x)$  is non-differentiable, while  $\widehat{\mathbb{E}}[\ell(X)]$  is both tractable and differentiable in  $\mu$ . This is the case, for example, when  $\ell(X)$  is a Laplace log-likelihood. On the other

hand, Equation (17) can be used when derivatives of the log-likelihood are easy and cheap to compute and can then be paired with numerical integration [10] if their expectations are intractable. In all other cases, Equation (16) together with numerical integration might be preferable. Lastly, if  $\ell(X)$  is quadratic in  $X$  we get exact updates, see Theorem 2.

**Theorem 2.** *Let  $Y | X \sim \mathcal{N}(CX, R)$  then the Gaussian projection update Equations (16) to (18) give the exact posterior.*

*Proof.* Note that  $\nabla_X \ell(X) = C^\top R^{-1}(y - CX)$  and  $\nabla_X^2 \ell(X) = -C^\top R^{-1}C$  and change parametrisation to information form  $\xi = \Sigma^{-1}\mu$  and  $\Lambda = \Sigma^{-1}$ . Then Equation (17) gives:

$$\partial_{\tau} \xi = C^\top R^{-1}y, \quad (19a)$$

$$\partial_{\tau} \Lambda = C^\top R^{-1}C. \quad (19b)$$

Therefore we have

$$\xi(1) = \xi_0 + C^\top R^{-1}y, \quad (20a)$$

$$\Lambda(1) = \Lambda_0 + C^\top R^{-1}C, \quad (20b)$$

which is the information form of the Kalman update [23].  $\square$

## 3. EXPERIMENTS

### 3.1. Linear $\ell_1$ -Filtering

Consider a Wiener velocity model:

$$dX(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \otimes I_2 X(t) dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes I_2 dB(t), \quad (21a)$$

$$Y(t_n) = [0, I_2] X(t_n) + R_n^{1/2} V_n. \quad (21b)$$

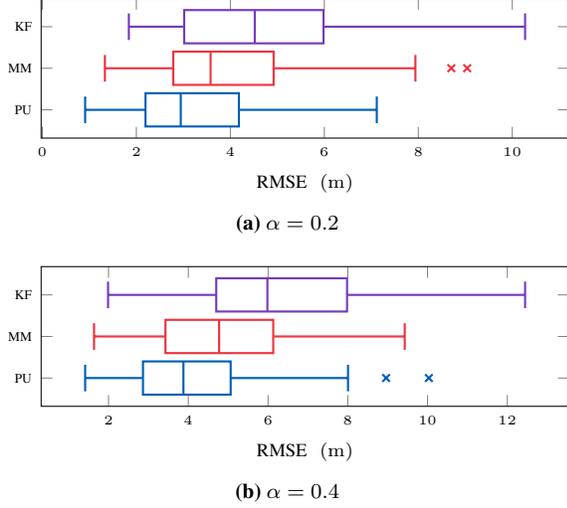
The initial condition is Gaussian with mean  $\mathbb{E}[X(0)] = [0, 0, 10, 10]^\top$  and covariance  $\mathbb{V}[X(0)] = I$ . We use a similar simulation model as in [24], where  $R_n = R_0$  with probability  $1 - \alpha$  and  $R_n = 20R_0$  otherwise,  $V_n$  is standard Gaussian, the system is measured at intervals  $\delta t = 0.1$ , and  $N = 1000$  measurements per trajectory are simulated with  $M = 100$  trajectories in total. Lastly, we set  $R_0 = I_2$  and conduct two experiments for  $\alpha = 0.2$  and  $\alpha = 0.4$ , respectively.

For the projection update (PU) we take  $R_n V_n$  to be Laplace distributed,  $\mathcal{L}(0, R_0)$  and we compare to the method by [24] (MM) and the standard Kalman filter (KF). PU uses 5 integration steps and MM uses 5 iterations. Furthermore, note that the statistical linear regression methods [13, 14, 15] reduce to the Kalman filter for this model.

The projection update is implemented as follows: The expectation of the log-likelihood is just a sum of means of folded Normal distributions [25], which can be differentiated with respect to  $\mu$ . Therefore, Equation (18) gives

$$\partial_{\tau} \mu(\tau) = \Sigma(\tau) C^\top R_0^{-1/2} w(\tau), \quad (22a)$$

$$\partial_{\tau} \Sigma(\tau) = -\Sigma(\tau) C^\top R_0^{-1/2} W(\tau) R_0^{-1/2} C \Sigma(\tau), \quad (22b)$$



**Fig. 1:** Boxplots of the RMSES in the tracking experiments for KF (Purple), MM [24] (Red), and PU (Blue).

where the entries of the vector  $w$  and the diagonal matrix  $W$  are given by

$$w_i(\tau) = \sqrt{2} \operatorname{Erf} \left( \frac{e_i^\top R_0^{-1/2} (y(t_n) - C\mu(\tau))}{\sqrt{2[R_0^{-1/2} C \Sigma(\tau) C^\top R_0^{-1/2}]_{ii}}} \right), \quad (23a)$$

$$W_{ii}(\tau) = \frac{2}{\sqrt{\pi}} \frac{\exp \left( -\frac{[e_i^\top R_0^{-1/2} (y(t_n) - C\mu(\tau))]^2}{2[R_0^{-1/2} C \Sigma(\tau) C^\top R_0^{-1/2}]_{ii}} \right)}{[R_0^{-1/2} C \Sigma(\tau) C^\top R_0^{-1/2}]_{ii}}, \quad (23b)$$

where Erf is the error function and  $e_i$  is a canonical basis vector. The results are presented by boxplots of the root mean square errors (RMSE) in Figure 1. As can be seen, MM is better than KF, while PU is better than MM. As MM and PU use the same modelling assumptions it is feasible that PU is better at approximating the filtering distribution.

### 3.2. Stochastic Volatility

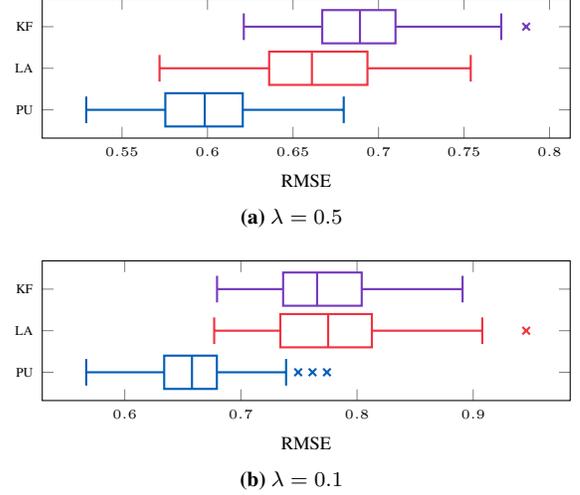
Consider the stochastic volatility model:

$$dX(t) = -\lambda(X(t) - m) dt + \sigma B(t), \quad (24a)$$

$$Y(t_n) = \exp(X(t_n)/2) V_n. \quad (24b)$$

The initial condition is Gaussian with moments  $\mathbb{E}[X(0)] = \mathbb{V}[X(0)] = 1$  and  $\sigma = m = 1$ . The system is simulated with a measurement interval of  $\delta t = 0.1$  and  $N = 1000$  measurements per trajectory, with  $M = 100$  trajectories simulated in total. Lastly, we make two experiments for  $\lambda = 0.5$  and  $\lambda = 0.1$ , respectively.

The log-likelihood derivatives  $\nabla_X \ell(X)$  and  $\nabla_X^2 \ell(X)$  are easy to compute and have tractable expectations. Therefore,



**Fig. 2:** Boxplots of the RMSES in the stochastic volatility model for KF (Purple), LA (Red), and PU (Blue).

Equation (17) gives the following projection update (PU):

$$\partial_\tau \mu = \Sigma \frac{1}{2} \left( y^2 \exp \left( -\mu + \frac{\Sigma}{2} \right) - 1 \right), \quad (25a)$$

$$\partial_\tau \Sigma = -\Sigma^2 \frac{1}{2} y^2 \exp \left( -\mu + \frac{\Sigma}{2} \right). \quad (25b)$$

We compare this to Laplace approximation (LA), which also uses  $\nabla_X \ell(X)$  and  $\nabla_X^2 \ell(X)$ . Furthermore, the measurement in Equation (24) can be transformed according to:

$$\log Y^2(t_n) = X(t_n) + \log V_n^2, \quad (26)$$

where  $\mathbb{E}[\log V_n^2] = \psi(1) - \log 2$ ,  $\mathbb{V}[\log V_n^2] = \pi^2/2$ , and  $\psi$  is the digamma function [26]. This enables the implementation of a Kalman filter (KF), which we also compare to. Lastly, note that the statistical linear regression methods are not applicable here as the conditional mean is zero, (see [15]), and they reduce to the Kalman filter for the transformation in Equation (26).

The results are presented by boxplots of the RMSES in Figure 2. It can be seen that LA offers a moderate improvement in performance over KF while PU in turn offers a significant improvement in performance over LA. It can also be seen that the RMSE for all filters go down for slower mean reversion ( $\lambda = 0.1$ ), while the difference in performance increases.

## 4. CONCLUSIONS

We have shown how differential geometric methods provide an effective and versatile tool for approximating Bayesian filter updates in systems with non-Gaussian likelihoods. The novelty of this paper was to reformulate the update as a continuous flow which can be approximated with projection methods. In addition to providing the general algorithm, we also specialised the algorithm to Gaussian manifolds. Our experimental results show that the algorithm gives good results in comparison to alternative approximation algorithms.

## References

- [1] B. Øksendal, *Stochastic Differential Equations - An Introduction with Applications*, Springer, 2003.
- [2] D. H. Titterton and J. L. Weston, *Strapdown Inertial Navigation Technology*, The Institute of Electrical Engineers, 2004.
- [3] S. Särkkä, *Bayesian filtering and smoothing*, Cambridge University Press, 2013.
- [4] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*, Chapman & Hall/CRC, 2004.
- [5] E. Lindström, H. Madsen, and J. N. Nielsen, *Statistics for Finance*, Chapman and Hall/CRC, 2015.
- [6] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [7] Andrew H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.
- [8] 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*, 1995, pp. 1628–1632.
- [9] 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.
- [10] Y. Wu, D. Hu, M. Wu, and X. Hu, “A numerical-integration perspective on Gaussian filters,” *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2910–2921, Aug 2006.
- [11] S. Särkkä, “On unscented Kalman filtering for state estimation of continuous-time nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 52(9), pp. 1631–1641, 2007.
- [12] S. Särkkä and J. Sarmavuori, “Gaussian filtering and smoothing for continuous-discrete dynamic systems,” *Signal Processing*, vol. 93, pp. 500–510, 2013.
- [13] B. M. Bell and F. W. Cathey, “The iterated Kalman filter update as a Gauss–Newton method,” *IEEE Transaction on Automatic Control*, vol. 38, no. 2, pp. 294–297, 1993.
- [14] 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.
- [15] F. Tronarp, A. F. Garcia-Fernandez, and S. Särkkä, “Iterative filtering and smoothing in non-linear and non-Gaussian systems using conditional moments,” *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 408–412, March 2018.
- [16] D. Brigo, B. Hanzon, and F. LeGland, “A differential geometric approach to nonlinear filtering: the projection filter,” *IEEE Transactions on Automatic Control*, vol. 43, no. 2, pp. 247–252, 1998.
- [17] D. Brigo, B. Hanzon, and F. Le Gland, “Approximate nonlinear filtering by projection on exponential manifolds of densities,” *Bernoulli*, vol. 5, no. 3, pp. 495–534, 1999.
- [18] S. Amari, *Differential-geometrical methods in statistics*, vol. 28, Springer Science & Business Media, 2012.
- [19] S. Koyama, “Projection smoothing for continuous and continuous-discrete stochastic dynamic systems,” *Signal Processing*, vol. 144, pp. 333–340, 2018.
- [20] P. Del Moral, A. Doucet, and A. Jasra, “Sequential Monte Carlo samplers,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.
- [21] F. Nielsen and V. Garcia, “Statistical exponential families: A digest with flash cards,” *arXiv preprint arXiv:0911.4863*, 2009.
- [22] A. Doucet, P. E. Jacob, and S. Rubenthaler, “Derivative-free estimation of the score vector and observed information matrix,” *arXiv preprint 1304.5768*, 2013.
- [23] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Information and System Sciences Series. Prentice-Hall, 1979.
- [24] H. Wang, H. Li, W. Zhang, and H. Wang, “Laplace  $\ell_1$  robust Kalman filter based on majorization minimization,” in *20th International Conference on Information Fusion*. IEEE, 2017.
- [25] M. Tsagris, C. Beneki, and H. Hassani, “On the folded normal distribution,” *Mathematics*, vol. 2, no. 1, pp. 12–28, 2014.
- [26] K. Platanioti, EJ McCoy, and DA Stephens, “A review of stochastic volatility: univariate and multivariate models,” Tech. Rep., 2005.