Continuous-Discrete Filtering and Smoothing on Submanifolds of Euclidean Space

Filip Tronarp
Department of Computer Science
University of Tübingen
Tübingen, Germany
filip.tronarp@uni-tuebingen.de

Simo Särkkä
Department of Electrical Engineering and Automation
Aalto University
Espoo, Finland
simo.sarkka@aalto.fi

***Abstract*—In this paper the issue of filtering and smoothing in continuous discrete time is studied when the state variable evolves in some submanifold of Euclidean space, which may not have the usual Lebesgue measure. Formal expressions for prediction and smoothing problems are reviewed, which agree with the classical results except that the formal adjoint of the generator is different in general. These results are used to generalise the projection approach to filtering and smoothing to the case when the state variable evolves in some submanifold that lacks a Lebesgue measure. The approach is used to develop projection filters and smoothers based on the von Mises–Fisher distribution, which are shown to be outperform Gaussian estimators both in terms of estimation accuracy and computational speed in simulation experiments involving the tracking of a gravity vector.**

***Index Terms*—Continuous Discrete Filtering and Smoothing, Directional Statistics, Nonlinear Filtering and Smoothing, Riemann manifolds.**

I. Introduction

In this paper we consider the problem of inference in state-space models of the following form:

$$\mathrm{d}X(t) = a(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t), \tag{1a}$$

$$Y(t_n) \mid X(t_n) \sim m(t_n, y \mid X(t_n)), \tag{1b}$$

where $a\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ is the drift, $\sigma\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times q}$ is the diffusion coefficient, W is a standard Wiener process on $\mathbb{R}^q$, and X is the state variable, which is measured by $\{Y(t_n)\}_{n=1}^N$ with measurement densities $m(t_n, y \mid X(t_n))$. The likelihood at time t_n is denoted by $L(t_n, x)$ and the process noise covariance rate is denoted by

$$Q(t,x) = \sigma(t,x)\sigma^{\mathsf{T}}(t,x). \tag{2}$$

The problem of filtering and smoothing for the model in Eq. (1) has been well studied when the filtering and smoothing distributions on $X(t)$ admit densities with respect to the Lebesgue measure on $\mathbb{R}^d$ [1], [2]. Implementation of the exact filtering and smoothing relations are in general intractable, with the notable exception of affine Gaussian systems [3]–[5]. Consequently approaches to approximate inference have been developed such as *assumed density* [6] and the *projection approach* [7], [8].

However, the continuous-discrete time inference problem is not as well studied for the case when the state X is only supported on some submanifold $\mathbb{X}$ of $\mathbb{R}^d$, the assumed density approach has been taken for matrix Fisher distributions on the special orthogonal group $\mathbb{SO}(3)$ [9] and the von Mises–Fisher distribution on the unit sphere $\mathbb{S}^2$ in $\mathbb{R}^3$ [10]. The discrete time problem has been given attention in, for example, [11]–[16].

The contribution of this paper is to generalise the projection approach to filtering [7] and smoothing [8] to the case in which the state variable evolves in a submanifold $\mathbb{X}$ of Euclidean space and for which the filtering and smoothing distributions admits densities with respect to some measure λ on $\mathbb{X}$. This gives the same approximation formulae in terms of generators as in the $\mathcal{L}_2(\mathbb{R}^d)$ case, demonstrating that the Lebesgue density assumption of previous contributions is unnecessary. Furthermore, von Mises–Fisher-based filters and smoothers are developed for gravity vector tracking, which are shown to both be accurate and computationally fast in comparison to Gaussian estimators.

The rest of this paper is organised as follows, the formal solutions to the filtering and smoothing problems are derived in Section II, in Section III the formal solutions are approximated by the projection approach. The methodology is applied to reference vector tracking in Section IV and conclusions are given in Section V.

II. Formal Solution

In this section the formal solutions to the continuous-discrete filtering and smoothing problems are reviewed, with proofs included for completeness.

In the following, the set of measurements up to time t is denoted by $\mathscr{Y}(t) = \{y(t_n)\colon t_n \leq t\}$, the filtering density is denoted by $p_F(t,x) = p(t, x \mid \mathscr{Y}(t))$, and the smoothing density is denoted by $p_S(t,x) = p(t, x \mid \mathscr{Y}(T))$. Recall that the generator of the Itô process X is given by [17]

$$\mathcal{G}[\phi] = \sum_i a_i \partial_i \phi + \frac{1}{2}\sum_{i,j} Q_{i,j}\partial^2_{i,j}\phi, \tag{3}$$

and its adjoint taken in $\mathcal{L}_2(\mathbb{X}, \lambda)$ is denoted by $\mathcal{G}^a$. The present development, *crucially*, does not require the explicit expression for $\mathcal{G}^a$ but it is used in the formal solution formulae. The Fokker–Planck equation on $\mathbb{X}$ with respect to λ is given in Proposition 1.

Proposition 1. *The probability density for $X(t)$ evolves according to*

$$\partial_t p = \mathcal{G}^a[p], \tag{4}$$

where $\mathcal{G}^a$ is the adjoint of $\mathcal{G}$ taken in $\mathcal{L}_2(\mathbb{X}, \lambda)$.

Proof. Itô's formula implies that for arbitrary $\phi \in \mathcal{C}^2(\mathbb{X})$

$$\begin{aligned}\partial_t \mathbb{E}[\phi(X(t))] &= \int_{\mathbb{X}} \phi(x) \partial_t p(t,x) \,\mathrm{d}\lambda(x) = \mathbb{E}[\mathcal{G}[\phi](X(t))] \\ &= \int_{\mathbb{X}} \mathcal{G}[\phi](x) p(t,x) \,\mathrm{d}\lambda(x) \\ &= \int_{\mathbb{X}} \phi(x) \mathcal{G}^a[p](t,x) \,\mathrm{d}\lambda(x),\end{aligned}$$

but ϕ is arbitrary, therefore $\partial_t p = \mathcal{G}^a[p]$ as claimed. □

From the fact that X is a Markov process and Proposition 1 it follows that the filtering distribution, between measurements, evolves as

$$\partial_t p_F = \mathcal{G}^a[p_F]. \tag{5}$$

The update is given by Bayes' rule

$$p_F(t_n, x) = \frac{L(t_n, x) p_F(t_n^-, x)}{\int_{\mathbb{X}} L(t_n, x) p_F(t_n^-, x) \,\mathrm{d}\lambda(x)}. \tag{6}$$

The following Theorem was proved in [1] under the assumption that the filtering and smoothing distributions of X have densities with respect to the Lebesgue measure.

Theorem 1. *The smoothing density satisfies*

$$\partial_t p_S = \frac{p_S}{p_F} \mathcal{G}^a[p_F] - p_F \mathcal{G}\Big[\frac{p_S}{p_F}\Big]. \tag{7}$$

The statement still holds when $X(t)$ evolves in some submanifold $\mathbb{X}$ of $\mathbb{R}^d$, with time marginal distributions absolutely continuous with respect to some measure λ on $\mathbb{X}$. The argument is the same as in [1], *mutatis mutandis*, and it is given in the following proof.

Proof. By the Markov property we have

$$\begin{aligned}p_S(t,x) &= \int_{\mathbb{X}} p_S(t, x \mid X(t+\mathrm{d}t) = z) p_S(t+\mathrm{d}t, z) \,\mathrm{d}\lambda(z) \\ &= \int_{\mathbb{X}} p_F(t, x \mid X(t+\mathrm{d}t) = z, \mathrm{d}\mathscr{Y}(t)) p_S(t+\mathrm{d}t, z) \,\mathrm{d}\lambda(z) \\ &= \int_{\mathbb{X}} p_F(t, x \mid X(t+\mathrm{d}t) = z, \mathrm{d}\mathscr{Y}(t)) \\ &\quad \times \big(p_S(t,z) + \partial_t p_S(t,z) \,\mathrm{d}t\big) \,\mathrm{d}\lambda(z) + o(\mathrm{d}t).\end{aligned}$$

The first term in this integral can be evaluated using Bayes' rule

$$\begin{aligned}&p_F(t, x \mid X(t+\mathrm{d}t) = z, \mathrm{d}\mathscr{Y}(t)) \\ &= \frac{p(t+\mathrm{d}t, z \mid X(t) = x, \mathrm{d}\mathscr{Y}(t)) p_F(t, x \mid \mathrm{d}\mathscr{Y}(t))}{p_F(t+\mathrm{d}t, z)} \\ &= \frac{p(t+\mathrm{d}t, z \mid X(t) = x) p_F(t,x)}{p_F(t+\mathrm{d}t, z)} \\ &= \frac{\big(\delta(z-x) + \mathcal{G}^a[\delta(z-x)]\,\mathrm{d}t\big) p_F(t,x)}{p_F(t+\mathrm{d}t, z)} + o(\mathrm{d}t),\end{aligned} \tag{8}$$

where the second equality follows from the fact that $\mathrm{d}\mathscr{Y}(t) = \{y(t_n) \colon t \le t_n \le t + \mathrm{d}t\}$ is empty for sufficiently small $\mathrm{d}t$. The last equality follows from Eq. (4). Inserting this expression for the filtering density into the former expression for the smoothing density yields

$$\begin{aligned}p_S(t,x) &= \frac{p_F(t,x)}{p_F(t+\mathrm{d}t, x)} \big(p_S(t,x) + \partial_t p_S(t,x) \,\mathrm{d}t\big) \\ &+ p_F(t,x) \int_{\mathbb{X}} \mathcal{G}^a[\delta(z-x)] \frac{p_S(t,z)}{p_F(t+\mathrm{d}t, z)} \,\mathrm{d}t \,\mathrm{d}\lambda(z) + o(\mathrm{d}t) \\ &= p_S(t,x) + \partial_t p_S(t,x) \,\mathrm{d}t - \frac{p_S(t,x)}{p_F(t,x)} \mathcal{G}^a[p_F](t,x) \,\mathrm{d}t \\ &+ \mathcal{G}\Big[\frac{p_S}{p_F}\Big](t,x) \,\mathrm{d}t + o(\mathrm{d}t),\end{aligned}$$

and the conclusion follows. □

III. The Projection Method

In the following the differential geometric setup of [7] is reviewed. Consider the metric space of square root densities $\mathcal{P}^{1/2}(\mathbb{X}) \subset \mathcal{L}_2(\mathbb{X}, \lambda)$, for which the Hellinger metric is induced by the $\mathcal{L}_2(\mathbb{X}, \lambda)$ norm. Furthermore, consider an m-dimensional manifold in $\mathcal{P}^{1/2}(\mathbb{X})$ with one global and smoothing coordinate chart

$$\mathcal{P}_\Theta^{1/2} = \{\sqrt{p_\theta}, \quad \theta \in \Theta \subset \mathbb{R}^m\}. \tag{9}$$

The tangent space at $\sqrt{p_\theta}$ is a closed subspace of $\mathcal{L}_2(\mathbb{X}, \lambda)$, which is given by

$$\mathcal{T}_{\sqrt{p_\theta}} \mathcal{P}_\Theta^{1/2} = \operatorname{span}\{\partial_1^\theta \sqrt{p_\theta}, \dots \partial_m^\theta \sqrt{p_\theta}\} \tag{10}$$

and the inner product between basis elements of the tangent space is given by

$$\langle \partial_i^\theta \sqrt{p_\theta}, \partial_j^\theta \sqrt{p_\theta} \rangle = g(\theta)/4, \tag{11}$$

where $g(\theta)$ is the Fisher information matrix. The projection $\Pi_\theta \colon \mathcal{L}_2(\mathbb{X}, \lambda) \mapsto \mathcal{T}_{\sqrt{p_\theta}} \mathcal{P}_\Theta^{1/2}$ is given by

$$\Pi_\theta v = 4 \sum_{i,j} g_{i,j}^{-1}(\theta) \langle v, \partial_j^\theta \sqrt{p_\theta} \rangle \partial_i^\theta \sqrt{p_\theta}. \tag{12}$$

For particular forms of v the projection formula simplifies according to the following Lemma [7].

Lemma 1. *Let $u \in \mathcal{L}_2(\mathbb{X}, \lambda)$ satisfy $\mathbb{E}_\theta[\|u|] < \infty$. Then the projection of $v = \frac{1}{2}\sqrt{p_\theta} u$ onto $\mathcal{T}_{\sqrt{p_\theta}}$ is given by*

$$\Pi_\theta v = \mathbb{E}_\theta[u \nabla_\theta \log p_\theta]^\mathsf{T} g^{-1}(\theta) \nabla_\theta \sqrt{p_\theta}. \tag{13}$$

A. Projection Filtering

It follows from Eq. (5) that $\sqrt{p_F}$ is governed by

$$\begin{aligned}\mathcal{F}^{1/2}[\phi] &= \frac{\phi}{2\phi^2} \mathcal{G}^a[\phi^2], \\ \partial_t \sqrt{p_F} &= \mathcal{F}^{1/2}[\sqrt{p_F}].\end{aligned}$$

Let $p_F(t_{n-1}, x) \in \mathcal{P}_\Theta$, then the prediction formula from t_{n-1} due to the projection approach is given by [7]

$$\partial_t \sqrt{p_{\theta_F}} = \Pi_{\theta_F} \circ \mathcal{F}^{1/2}[\sqrt{p_{\theta_F}}]. \tag{15}$$

Proposition 2. *The curve in Θ defined by Eq.* (15) *is given by*

$$\dot{\theta}_F = g^{-1}(\theta_F)\mathbb{E}_{\theta_F}\big[\mathcal{G}[\nabla_{\theta_F}\log p_{\theta_F}]\big]. \tag{16}$$

Proposition 2 was proved in [7] for the case in which p_F is a density with respect to the Lebesgue measure. The proof method is essentially the same but is given below for completeness.

Proof. It follows from Lemma 1 that

$$\begin{aligned}\partial_t\sqrt{p_{\theta_F}} &= \Pi_{\theta_F}\circ\mathcal{F}^{1/2}[\sqrt{p_{\theta_F}}]\\ &= \mathbb{E}_{\theta_F}\Big[\frac{\mathcal{G}^a[p_{\theta_F}]}{p_{\theta_F}}\nabla_{\theta_F}\log p_{\theta_F}\Big]g^{-1}(\theta_F)\nabla_{\theta_F}\sqrt{p_{\theta_F}}\\ &= \mathbb{E}_{\theta_F}\big[\mathcal{G}[\nabla_{\theta_F}\log p_{\theta_F}]\big]^{\mathsf{T}}g^{-1}(\theta_F)\nabla_{\theta_F}\sqrt{p_{\theta_F}}.\end{aligned}$$

On the other hand, by the chain rule $\partial_t\sqrt{p}_{\theta_F} = \dot{\theta}_F^{\mathsf{T}}\nabla_{\theta_F}\sqrt{p_{\theta_F}}$ and matching terms gives the result. □

Corollary 1. *Let $\mathcal{P}_\Theta$ be an exponential family, $p_\theta \propto \exp(\theta^{\mathsf{T}}s(x)-\psi(\theta))$. Then Eq.* (16) *reduces to*

$$\dot{\theta}_F = g^{-1}(\theta_F)\mathbb{E}_{\theta_F}\big[\mathcal{G}[s]\big]. \tag{17}$$

Proof. It follows from Proposition 2 and direct calculation. □

For the filter update, it is herein assumed that $\mathcal{P}_\Theta$ is a conjugate family for the likelihoods. For a projection approach to the filter update, see [18].

B. Projection Smoothing

It follows from Theorem 1 that $\sqrt{p_S}$ is governed by

$$\begin{aligned}\mathcal{B}^{1/2}[\phi] &= \frac{\phi}{2\phi^2}\Big(\frac{\phi^2}{p_F}\mathcal{G}^a[p_F] - p_F\mathcal{G}\Big[\frac{\phi^2}{p_F}\Big]\Big),\\ \partial_t\sqrt{p_S} &= \mathcal{B}^{1/2}[\sqrt{p_S}].\end{aligned}$$

In order to arrive at a tractable algorithm, similarly to [8], the operator $\mathcal{B}^{1/2}$ is approximated by

$$\widehat{\mathcal{B}}^{1/2}[\phi] = \frac{\phi}{2\phi^2}\Big(\frac{\phi^2}{p_{\theta_F}}\mathcal{G}^a[p_{\theta_F}] - p_{\theta_F}\mathcal{G}\Big[\frac{\phi^2}{p_{\theta_F}}\Big]\Big) \tag{19}$$

and the projection smoother is given by

$$\partial_t\sqrt{p_{\theta_S}} = \Pi_{\theta_S}\circ\widehat{\mathcal{B}}^{1/2}[\sqrt{p_{\theta_S}}]. \tag{20}$$

Remark 1. *The approximation of $\mathcal{B}^{1/2}$ by $\widehat{\mathcal{B}}^{1/2}$ adds an additional source of approximation error. At present we have no method to analyse this error but it is certainly an interesting topic for future research.*

Proposition 3 was proved in [8] for the case in which p_F and p_S are densities with respect to the Lebesgue measure. The proof method is essentially the same but is given below for completeness.

Proposition 3. *The curve in Θ defined by Eq.* (20) *is given by*

$$\dot{\theta}_S = g^{-1}(\theta_S)\mathbb{E}_{\theta_S}\Big[\nabla_{\theta_S}\frac{p_{\theta_F}}{p_{\theta_S}}\mathcal{G}\Big[\frac{p_{\theta_S}}{p_{\theta_F}}\Big]\Big]. \tag{21}$$

Proof. It follows from Lemma 1 that

$$\begin{aligned}\partial_t\sqrt{p_{\theta_S}} &= \Pi_{\theta_S}\circ\widehat{\mathcal{B}}^{1/2}[\sqrt{p_{\theta_S}}]\\ &= \mathbb{E}_{\theta_S}\left[\frac{\mathcal{G}^a[p_{\theta_F}]}{p_{\theta_F}}\nabla_{\theta_S}\log p_{\theta_S}\right]^{\mathsf{T}}g^{-1}(\theta_S)\nabla_{\theta_S}\sqrt{p_{\theta_S}}\\ &\quad - \mathbb{E}_{\theta_S}\left[\frac{p_{\theta_F}}{p_{\theta_S}}\mathcal{G}\Big[\frac{p_{\theta_S}}{p_{\theta_F}}\Big]\nabla_{\theta_S}\log p_{\theta_S}\right]^{\mathsf{T}}g^{-1}(\theta_S)\nabla_{\theta_S}\sqrt{p_{\theta_S}}.\end{aligned}$$

The first expectation can be simplified according to

$$\begin{aligned}&\mathbb{E}_{\theta_S}\left[\frac{\mathcal{G}^a[p_{\theta_F}]}{p_{\theta_F}}\nabla_{\theta_S}\log p_{\theta_S}\right] = \nabla_{\theta_S}\mathbb{E}_{\theta_S}\left[\frac{\mathcal{G}^a[p_{\theta_F}]}{p_{\theta_F}}\right]\\ &= \nabla_{\theta_S}\mathbb{E}_{\theta_F}\left[\mathcal{G}\Big[\frac{p_{\theta_S}}{p_{\theta_F}}\Big]\right] = \nabla_{\theta_S}\mathbb{E}_{\theta_S}\left[\frac{p_{\theta_F}}{p_{\theta_S}}\mathcal{G}\Big[\frac{p_{\theta_S}}{p_{\theta_F}}\Big]\right]\\ &= \mathbb{E}_{\theta_S}\left[\nabla_{\theta_S}\frac{p_{\theta_F}}{p_{\theta_S}}\mathcal{G}\Big[\frac{p_{\theta_S}}{p_{\theta_F}}\Big]\right]\\ &\quad + \mathbb{E}_{\theta_S}\left[\frac{p_{\theta_F}}{p_{\theta_S}}\mathcal{G}\Big[\frac{p_{\theta_S}}{p_{\theta_F}}\Big]\nabla_{\theta_S}\log p_{\theta_S}\right].\end{aligned}$$

Inserting this into the previous equation gives

$$\partial_t\sqrt{p_{\theta_S}} = \mathbb{E}_{\theta_S}\Big[\nabla_{\theta_S}\frac{p_{\theta_F}}{p_{\theta_S}}\mathcal{G}\Big[\frac{p_{\theta_S}}{p_{\theta_F}}\Big]\Big]^{\mathsf{T}}g^{-1}(\theta_S)\nabla_{\theta_S}\sqrt{p_{\theta_S}}$$

and the conclusion follows by the same argument as in the proof of Proposition 2. □

Corollary 2. *Let $\mathcal{P}_\Theta$ be an exponential family, $p_\theta \propto \exp(\theta^{\mathsf{T}}s(x)-\psi(\theta))$. Then Eq.* (21) *reduces to*

$$\dot{\theta}_S = g^{-1}(\theta_S)\Big(\mathbb{E}_{\theta_S}\big[\mathcal{G}[s]\big] + \mathbb{E}_{\theta_S}[J_s Q J_s^{\mathsf{T}}](\theta_S-\theta_F)\Big), \tag{22}$$

where J_s is the Jacobian of the sufficient statistic s.

Proof. It follows from Proposition 3 and direct calculation. □

IV. Application: Tracking on $\mathbb{S}^2$

Consider the following state space model

$$\mathrm{d}X = -\breve{\Omega}(t)\times X\,\mathrm{d}t - \gamma^2 X\,\mathrm{d}t + \gamma X\times\mathrm{d}W, \tag{23a}$$

$$Y(t_n)\mid X(t_n) \sim \mathcal{N}(gX(t_n),\alpha^2\mathrm{I}), \tag{23b}$$

where $\times$ denotes the vector cross-product and $X(t)\in\mathbb{S}^2$ for $t>0$ if $X(0)\in\mathbb{S}^2$ [10]. The model in Eq. (23) can be used to track the local gravity direction by using gyroscope measurements $\breve{\Omega}$, accelerometer measurements Y, and setting g to the local gravity constant. For this model Q and $\mathcal{G}$ are

$$Q(x) = \gamma^2(\mathrm{I}\|x\|^2 - xx^{\mathsf{T}}), \tag{24a}$$

$$\mathcal{G}[\phi] = -(\breve{\Omega}(t)\times x + \gamma^2 x)^{\mathsf{T}}\nabla_x\phi + \frac{1}{2}\operatorname{tr}[Q\nabla_x^2\phi]. \tag{24b}$$

To develop projection filters and smoothers for this inference problem, a class of densities on the unit sphere needs to be selected. An obvious choice is the von Mises–Fisher family, which is an exponential family with respect to the uniform

measure on $\mathbb{S}^2$ [19]. The von Mises–Fisher densities on $\mathbb{S}^2$ are

$$p_\theta(x) = \exp(\theta^\mathsf{T} x - \kappa(\|\theta\|)), \tag{25}$$

where $\kappa(r) = -\log r + \log(4\pi) + \log\sinh r$ and the sufficient statistic is $s(x) = x$. Furthermore, for the von Mises–Fisher distribution the following holds [19], [20]

$$\mathbb{E}_\theta[s(X)] = \mathbb{E}_\theta[X] = \kappa'(\|\theta\|)\frac{\theta}{\|\theta\|}, \tag{26a}$$

$$g(\theta) = \frac{\kappa'(\|\theta\|)}{\|\theta\|}P_\perp(\theta) + \kappa''(\|\theta\|)P(\theta), \tag{26b}$$

$$g^{-1}(\theta) = \frac{\|\theta\|}{\kappa_3'(\|\theta\|)}P_\perp(\theta) + \frac{1}{\kappa''(\|\theta\|)}P(\theta), \tag{26c}$$

where $P(\theta) = \theta\theta^\mathsf{T}/\|\theta\|^2$ and $P_\perp(\theta) = \mathrm{I} - P(\theta)$.

A. von Mises–Fisher Filtering

The projection based prediction equation is retrieved by Corollary 1, Eq. (24), and Eq. (26). It is given by

$$\dot\theta_F = -\breve\Omega(t) \times \theta_F - \frac{\gamma^2\kappa'(\|\theta_F\|)}{\|\theta_F\|\,\kappa''(\|\theta_F\|)}\theta_F. \tag{27}$$

The von Mises–Fisher family is conjugate to the likelihood (Proposition 1 in [10]), yielding the following update

$$\theta_F(t_n) = \theta_F(t_n^-) + gy(t_n)/\alpha^2. \tag{28}$$

Another parametrisation of the von Mises–Fisher distribution is given by the mapping

$$\theta_F \mapsto (\|\theta_F\|, \theta_F/\|\theta_F\|) = (\beta_F, \mu_F). \tag{29}$$

By the chain-rule and Eq. (27), the predictions in this parametrisation are

$$\dot\mu_F = -\breve\Omega(t) \times \mu_F, \quad \dot\beta_F = -\gamma^2\kappa'(\beta_F)[\kappa''(\beta_F)]^{-1}. \tag{30}$$

These are the prediction equations derived in [10] by the assumed density method. An advantage of the current development is that a smoother is readily obtained as well.

B. von Mises–Fisher Smoothing

Using Corollary 2, Eq. (24), and Eq. (26) gives the following formula for the von Mises–Fisher smoother

$$\begin{aligned} G(\theta) &= \frac{\gamma^2\|\theta\|}{\kappa'(\|\theta\|)}P_\perp(\theta) + \frac{\gamma^2[1-(\kappa'(\|\theta\|))^2]}{\kappa''(\|\theta\|)}P(\theta) - \gamma^2\mathrm{I}, \\ \dot\theta_S &= -\breve\Omega(t) \times \theta_S - \frac{\gamma^2\kappa'(\|\theta_S\|)\theta_S}{\|\theta_S\|\,\kappa''(\|\theta_S\|)} + G(\theta_S)\big(\theta_S - \theta_F\big), \end{aligned} \tag{31}$$

where θ_F is the output of the filter.

C. Experimental Results

The von Mises–Fisher filter and smoother (VMFF/VMFS) are evaluated on the model in Eq. (23) with $g = 9.82$, corresponding to tracking the local gravity vector. The system is simulated one hundred times for all combinations of $\gamma \in \{10^{-3}, 10^{-2}\}$ and $\alpha^2 \in \{10^{-3}, 10^{-2}\}$ and for sample rates of Y and $\breve\Omega$ at $50, 100$, and 200Hz. The coordinates of $\breve\Omega$ are governed by zero mean Ornstein–Uhlenbeck processes with mean reversion rate -5 and diffusion constant 2.5.

The von Mises–Fisher filter and smoother estimate the state by the mode $\theta/\|\theta\|$ and are compared against a Gaussian filter and a Gaussian smoother (GF/GS) (Type II, see [21]), which use norm-constrained minimum mean-square estimators [22]. The von Mises–Fisher filter and smoother are initialised by the uniform distribution on $\mathbb{S}^2$ and the Gaussian estimators are initialised by moment matching the uniform distribution on $\mathbb{S}^2$. The von Mises–Fisher based estimators are implemented with an exponential Rosenbrock–Euler integrator and the Kalman filter integrates the covariance by a zeroth-order hold approximation to $\mathbb{E}[Q(X(t))]$, the mean is integrated exactly.

The methods are evaluated in terms of the mean angular error. The angular error when estimating $r \in \mathbb{S}^2$ by $\hat r \in \mathbb{S}^2$ is given by

$$\varepsilon(\hat r, r) = \frac{180}{\pi}\cos^{-1}\hat r^\mathsf{T} r, \tag{32}$$

and the mean angular error for a series $r(t_n)$ estimated by $\hat r(t_n)$, $n = 1, 2, \ldots, N$ is given by

$$\bar\varepsilon_N = \frac{1}{N}\sum_{n=1}^{N}\varepsilon(\hat r(t_n), r(t_n)). \tag{33}$$

Additionally, for the von Mises–Fisher distribution, the parameter $|\theta|$ determines how concentrated the distribution is around the mode $\theta/|\theta|$ [19]. The difference between the von Mises–Fisher filter and the smoother shall be examined in this respect.

The mean angular error for the sample rate at 200Hz is listed in Table I, where the mean is calculated over all time stamps and Monte Carlo samples. While the improvement of the von Mises–Fisher filter over its Gaussian coutnerpart is marginal, it can be seen that the situation for smoothing is different. Indeed, the Gaussian smoother performs worse than the filter for higher noise levels while the von Mises–Fisher smoother consistently improves upon the filter estimate. The angular error for one trajectory at the various noise levels is shown in Figure 1, and the corresponding concentration parameters are shown in Figure 2. The performance degradation of the Gaussian smoother is particularly visible. For the concentration parameter, it can be seen that the von Mises–Fisher filter starts with very low concentration (high uncertainty) but rapdily converges to a very high steady-state concentration. The von Mises–Fisher smoother also converges to a steady-state concentration, which is higher than the filter, as expected.

The mean angular error for the sample rate at 100Hz is listed in Table II. The situation is similar as for the case of 200Hz in that the von Mises–Fisher filter is slightly better than the Gaussian filter, while again the Gaussian smoother

TABLE I: Mean angular error (degrees) at sampling frequency of 200Hz.

(α^2, γ)	VMFF	VMFS	GF	GS
$(10^{-3}, 10^{-3})$	1.0805	**0.7853**	1.0835	0.9393
$(10^{-2}, 10^{-3})$	1.9072	**1.3745**	1.9144	1.4999
$(10^{-3}, 10^{-2})$	3.0830	**2.4117**	3.1182	3.6434
$(10^{-2}, 10^{-2})$	5.8900	**4.2922**	6.0534	6.4925

TABLE II: Mean angular error (degrees) at sampling frequency of 100Hz.

(α^2, γ)	VMFF	VMFS	GF	GS
$(10^{-3}, 10^{-3})$	1.3013	**0.9719**	1.3053	1.1028
$(10^{-2}, 10^{-3})$	2.3238	**1.6702**	2.3345	1.7851
$(10^{-3}, 10^{-2})$	3.5199	**2.9060**	3.5523	4.3347
$(10^{-2}, 10^{-2})$	6.8638	**5.0812**	7.0968	7.6803

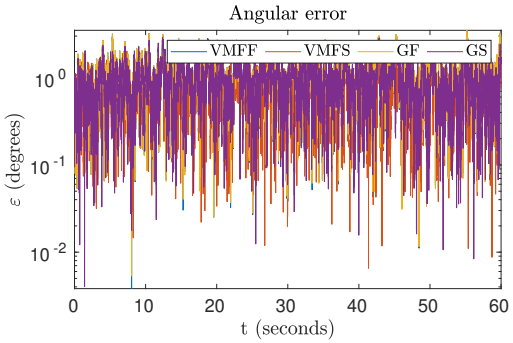
TABLE III: Mean angular error (degrees) at sampling frequency of 50Hz.

(α^2, γ)	VMFF	VMFS	GF	GS
$(10^{-3}, 10^{-3})$	1.6476	**1.3014**	1.6528	1.3781
$(10^{-2}, 10^{-3})$	2.8490	**2.1143**	2.8632	2.1944
$(10^{-3}, 10^{-2})$	4.0003	**3.5788**	3.9928	5.1751
$(10^{-2}, 10^{-2})$	8.0107	**6.0837**	8.3466	9.1550

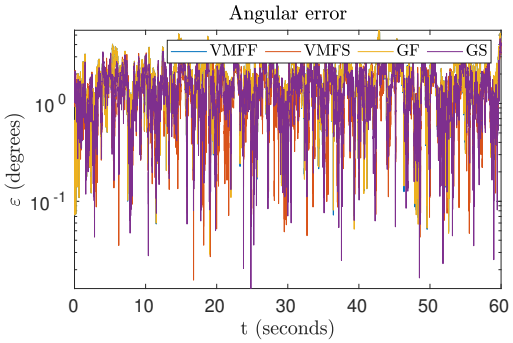
degrades as the noise level increases. However, the von Mises–Fisher smoother consistently improves upon the filter estimate. The angular error for one trajectory at the various noise levels is shown in Figure 3, and the corresponding concentration parameters are shown in Figure 4. Most notable is gain the performance degradation of the Gaussian smoother. For the concentration parameter, the results are the same as in the 200Hz case, except the stead-state levels perhaps being slightly lower.

The experiemntal results for 50Hz are summarised in Table III. The situation is similar as for the case of 200 and 100Hz in that the von Mises–Fisher filter is slightly better than the Gaussian filter. Now it can also be seen that the perfomance of all estimators degrades as the sample rate decreases. However, the performance degradation is most notable for the Gaussian smoother at high noise levels, where its error in comparison with the Gaussian filter appears to increase as the sample rate decreases. This problem is not present for the von Mises–Fisher smoother, which consistently improves upon the filter estimate. The angular error for one trajectory at the various noise levels is shown in Figure 5, and the corresponding concentration parameters are shown in Figure 6. What can be discerned from these figures is qualitatively the same as for the 100 and 200Hz case.

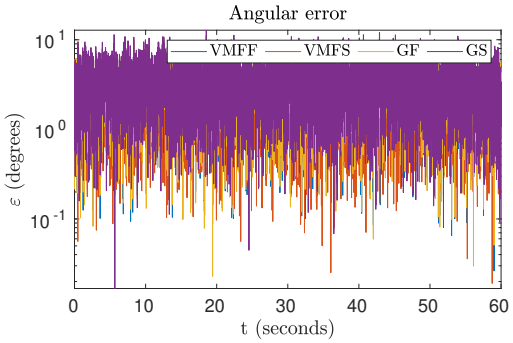
Additionally, the average processing time per sample is calculated for one of the experiments at a sample rate of 200Hz, see Table IV. It can be seen that another benefit of the von Mises–Fisher based estimators is their computational speed. Both the von Mises–Fisher and Gaussian filter can employ for Rodriguez formula for predictions, but the Gaussian filter also has to operate on a 3×3 covariance matrix slowing it down. However, there appears to be no corresponding trick that can be applied for the von Mises–Fisher smoother, which is why it suffers a major increase in computation time, whiole still being cheaper than its Gaussian counterpart.



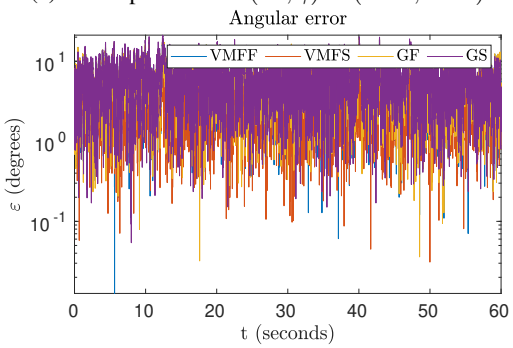
(a) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-3})$



(b) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-3})$

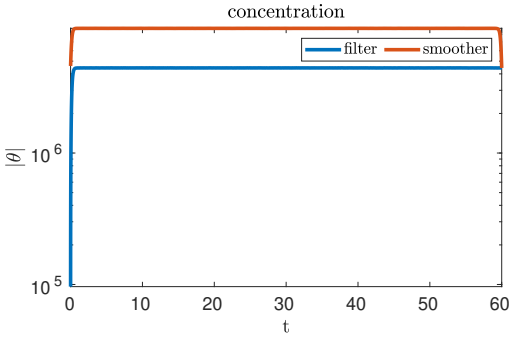


(c) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-2})$

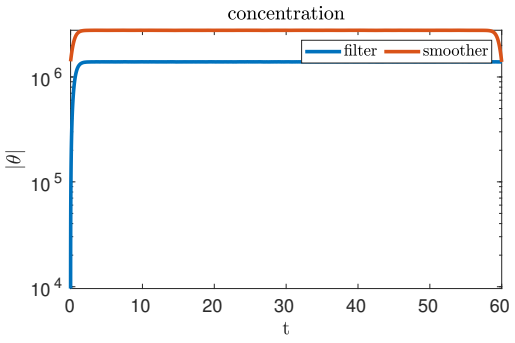


(d) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-2})$

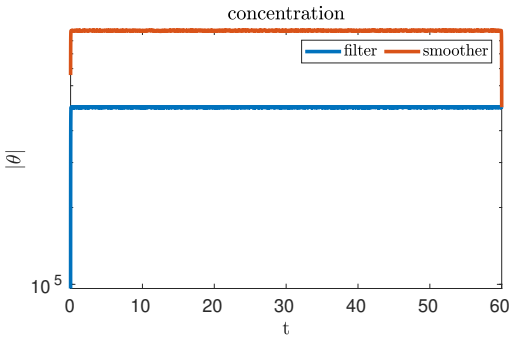
Fig. 1: The angular errors for one simulated trajectory at a sampling rate of 200Hz with varying noise parameters.



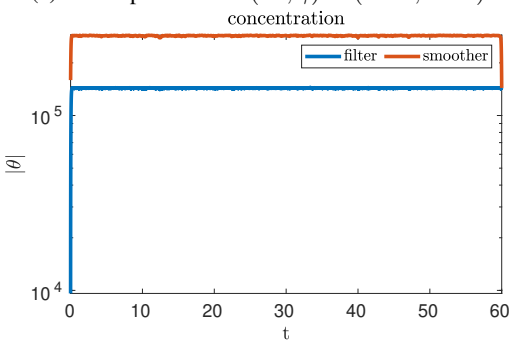
(a) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-3})$



(b) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-3})$

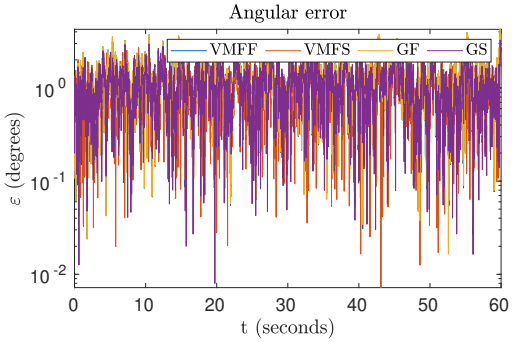


(c) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-2})$

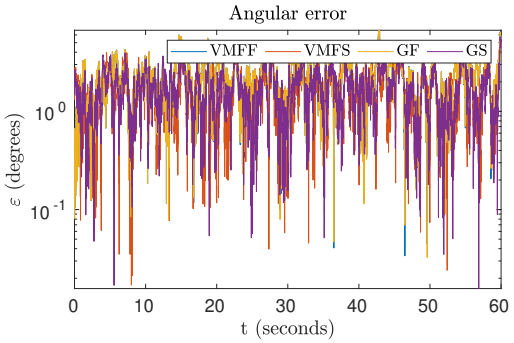


(d) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-2})$

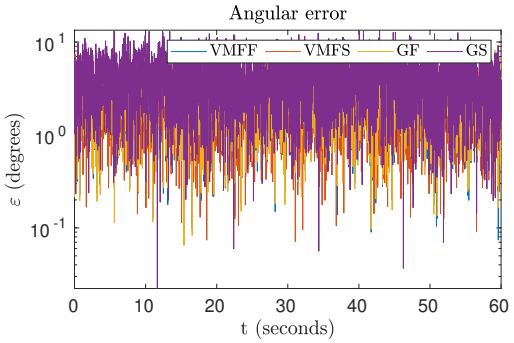
Fig. 2: The concentration parameter, $|\theta|$, of the von Mises–Fisher filter/smoother for one simulated trajectory at a sampling rate of 200Hz with varying noise parameters.



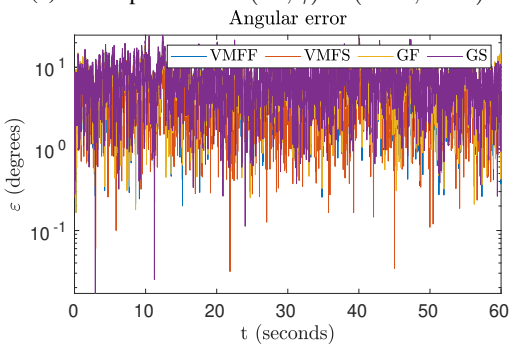
(a) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-3})$



(b) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-3})$



(c) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-2})$

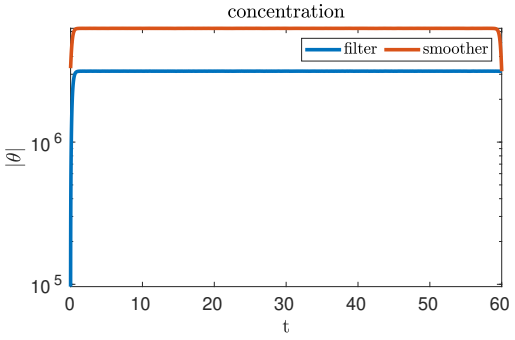


(d) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-2})$

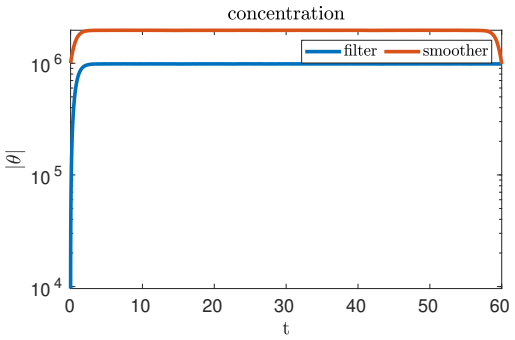
Fig. 3: The angular errors for one simulated trajectory at a sampling rate of 100Hz with varying noise parameters.

V. Conclusion

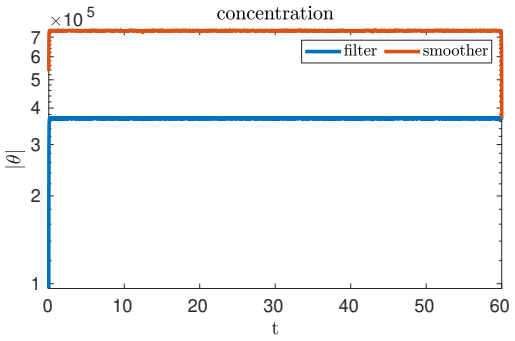
In this paper, the projection approach [7], [8] was generalised to the case in which the state variable evolves in a submanifold of Euclidean space. In particular, von Mises–Fisher based filters and smoothers were developed and shown to offer performance benefits both in terms of accuracy and



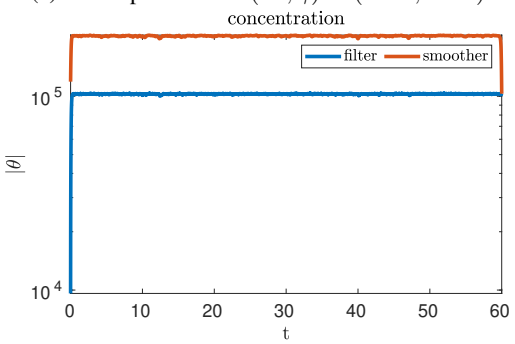
(a) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-3})$



(b) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-3})$

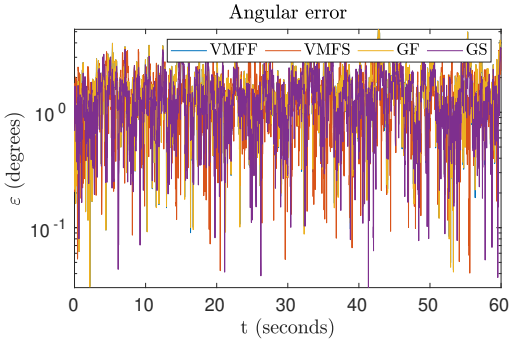


(c) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-2})$

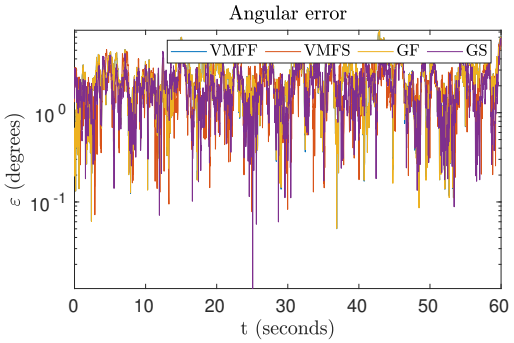


(d) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-2})$

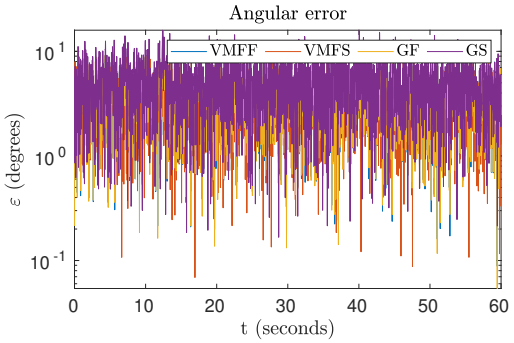
Fig. 4: The concentration parameter, $|\theta|$, of the von Mises–Fisher filter/smoother for one simulated trajectory at a sampling rate of 100Hz with varying noise parameters.



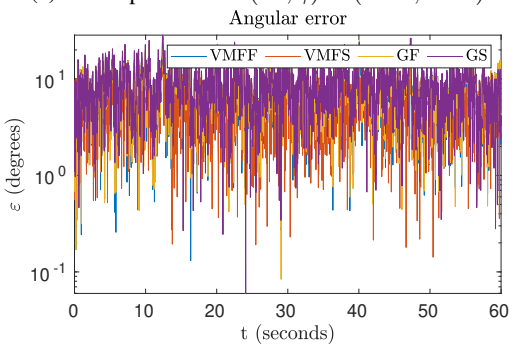
(a) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-3})$



(b) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-3})$



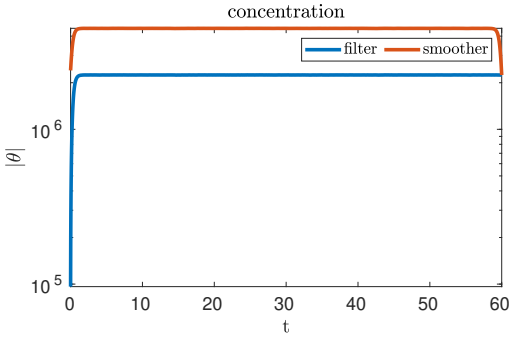
(c) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-2})$



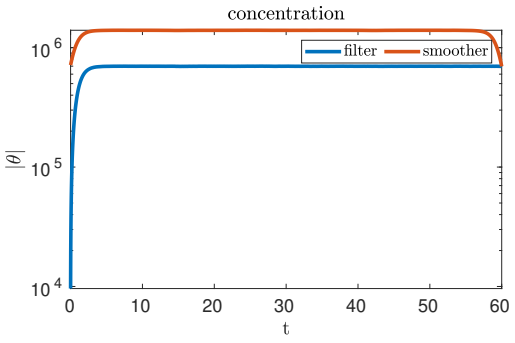
(d) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-2})$

Fig. 5: The angular errors for one simulated trajectory at a sampling rate of 50Hz with varying noise parameters.

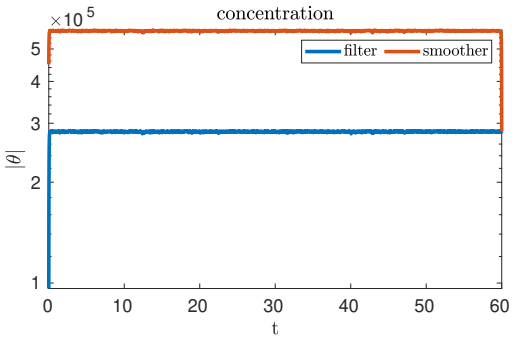
computational speed over Gaussian alternatives. Accuracy benefits are particularly pronounced in noisy systems, where Gaussian smoothers may perform worse than Gaussian filters, owing to the fact that these estimators do not appropriately account for the geometry of the state-space.



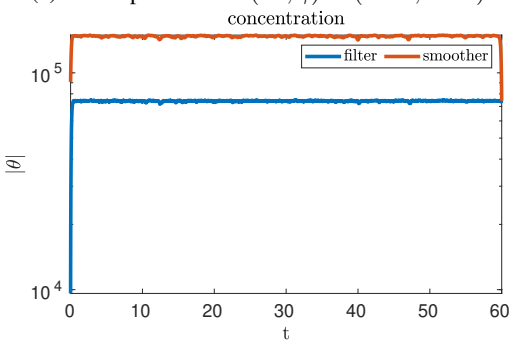
(a) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-3})$



(b) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-3})$



(c) Noise parameters: $(\alpha^2, \gamma) = (10^{-3}, 10^{-2})$



(d) Noise parameters: $(\alpha^2, \gamma) = (10^{-2}, 10^{-2})$

Fig. 6: The concentration parameter, $|\theta|$, of the von Mises–Fisher filter/smoother for one simulated trajectory at a sampling rate of 50Hz with varying noise parameters.

TABLE IV: Average processing time per sample

	VMFF	VMFS	GF	GS
Time (μs)	3.2403	99.31	165.7483	175.7942

REFERENCES

[1] B. D. O. Anderson, "Fixed interval smoothing for nonlinear continuous time systems," *Information and Control*, vol. 20, no. 3, pp. 294–300, 1972.

[2] C. T. Leondes, J. B. Peller, and E. B. Stear, "Nonlinear smoothing theory," *IEEE Transactions on System Science and Cybernetics*, vol. 6, no. 1, pp. 63–71, 1970.

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

[4] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME, Journal of Basic Engineering*, vol. 83, pp. 95–108, 1961.

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

[6] P. S. Maybeck, *Stochastic Models, Estimation and Control*. Academic Press, 1979,1982,1982, vol. 1-3.

[7] 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.

[8] S. Koyama, "Projection smoothing for continuous and continuous-discrete stochastic dynamic systems," *Signal Processing*, vol. 144, pp. 333–340, 2018.

[9] T. Lee, "Bayesian attitude estimation with the matrix Fisher distribution on SO(3)," *IEEE Transactions on Automatic Control*, vol. 63, no. 10, pp. 3377–3392, 2018.

[10] F. Tronarp, R. Hostettler, and S. Särkkä, "Continuous-discrete von Mises–Fisher filtering on S^2 for reference vector tracking," in *21st International Conference on Information Fusion*, ISIF. Cambridge, United Kingdom: IEEE, July 10 - 13 2018, pp. 1345–1352.

[11] M. Bukal, I. Marković, and I. Petrović, "Score matching based assumed density filtering with the von Mises–Fisher distribution," in *20th International Conference on Information Fusion*, ISIF. Xi'an, China: IEEE, July 10 - 13 2017.

[12] J. Traa and P. Smaragdis, "Multiple speaker tracking with the factorial von Mises–Fisher filter," in *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. Reims, France: IEEE, September 21 - 24 2014.

[13] G. Kurz, I. Gilitschenski, S. Julier, and U. D. Hanebeck, "Recursive Bingham filter for directional estimation involving 180 degree symmetry," *Journal on Advances in Information Fusion*, vol. 9, no. 2, pp. 90–105, 2014.

[14] J. Glover and L. P. Kaelbling, "Tracking the spin on a ping pong ball with the quaternion Bingham filter," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. Hong Kong, China: IEEE, May 31 - June 7 2014, pp. 4133–4140.

[15] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Unscented von Mises–Fisher filtering," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 463–467, 2016.

[16] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Unscented orientation estimation based on the Bingham distribution," *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 172–177, 2016.

[17] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*. Cambridge University Press, 2019.

[18] F. Tronarp and S. Särkkä, "Updates in Bayesian filtering by continuous projections on a manifold of densities," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing*. Brighton, United Kingdom: IEEE, May 12 - 17 2019, pp. 5032–5036.

[19] K. V. Mardia and P. E. Jupp, *Directional Statistics*. Wiley, 2000.

[20] Á. F. García-Fernández, F. Tronarp, and S. Särkkä, "Gaussian target tracking with direction-of-arrival von Mises–Fisher measurements," *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 2960–2972, 2019.

[21] S. Särkkä and J. Sarmavuori, "Gaussian filtering and smoothing for continuous-discrete dynamic systems," *Signal Processing*, vol. 93, pp. 500–510, 2013.

[22] R. Zanetti, M. Majji, R. H. Bishop, and D. Mortari, "Norm-constrained Kalman filtering," *Journal of guidance, control, and dynamics*, vol. 32, no. 5, pp. 1458–1465, 2009.