Tronarp, Filip; Hostettler, Roland; Särkkä, Simo

Continuous-Discrete von Mises-Fisher Filtering on S^2 for Reference Vector Tracking

Published in:
Proceedings of the 21st International Conference on Information Fusion, FUSION 2018

DOI:
10.23919/ICIF.2018.8455299

Published: 05/09/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Tronarp, F., Hostettler, R., & Särkkä, S. (2018). Continuous-Discrete von Mises-Fisher Filtering on S^2 for Reference Vector Tracking. In *Proceedings of the 21st International Conference on Information Fusion, FUSION 2018* (pp. 1345-1352). Article 8455299 IEEE. https://doi.org/10.23919/ICIF.2018.8455299

Continuous-Discrete von Mises–Fisher Filtering on S^2 for Reference Vector Tracking

Filip Tronarp, Roland Hostettler, Simo Särkkä

This is a post-print of a paper published in *21th International Conference on Information Fusion (FUSION)*. When citing this work, you must always cite the original article:

F. Tronarp, R. Hostettler, and S. Särkkä, "Continuous-discrete von Mises–Fisher filtering on S^2 for reference vector tracking," in *21th International Conference on Information Fusion (FUSION)*, Cambridge, UK, July 2018

DOI:

10.23919/ICIF.2018.8455299

Continuous-Discrete von Mises–Fisher Filtering on S^2 For Reference Vector Tracking

Filip Tronarp, Roland Hostettler, and Simo Särkkä
Department of Electrical Engineering and Automation
Aalto University, Finland
E-Mail: { filip.tronarp, roland.hostettler, simo.sarkka }@aalto.fi

***Abstract*—This paper is concerned with tracking of reference vectors in the continuous-discrete-time setting. For this end, an Itô stochastic differential equation, using the gyroscope as input, is formulated that explicitly accounts for the geometry of the problem. The filtering problem is solved by restricting the prediction and filtering distributions to the von Mises–Fisher class, resulting in ordinary differential equations for the parameters. A strategy for approximating Bayesian updates and marginal likelihoods is developed for the class of conditionally spherical measurement distributions, which is realistic for sensors such as accelerometers and magnetometers, and includes robust likelihoods. Furthermore, computationally efficient and numerically robust implementations are presented. The method is compared to other state-of-the-art filters in simulation experiments involving tracking of the local gravity vector. Additionally, the methodology is demonstrated in the calibration of a smartphone's accelerometer and magnetometer. Lastly, the method is compared to state-of-the-art in gravity vector tracking for smartphones in two use cases, where it is shown to be more robust to unmodeled accelerations.**

***Index Terms*—Directional statistics, von Mises–Fisher distribution, robust filtering, sensor calibration.**

I. Introduction

Tracking of directional quantities, such as reference vectors, is an important problem in signal processing. It is, for example, used for tracking of the local gravity vector in pedestrian dead reckoning systems [1], magnetic field based positioning [2], screen orientation tracking for smartphones [3], target tracking using omnidirectional cameras [4], and speaker tracking using a microphone array [5]. An important application is orientation tracking, which has previously been tackled by a variety of approaches, such as quaternion methods [6]–[10], where either a non-linear Kalman filter [7]–[10] or gradient descent [6] is used. These approaches have the drawback of not accounting for the geometry of the problem or are doing so in an ad-hoc manner.

On the other hand, reference vectors can be explicitly modeled on the unit sphere, S^2, using directional statistics [11], [12]. Approaches to the Bayesian tracking of reference vectors have recently been developed on this principle. Discrete-time filters have been developed by using the von Mises–Fisher distribution, based on moment-matching [4], [5], [13], [14]. While another approach, based on score-matching [15], [16], was proposed in [17].

In this paper, a continuous-discrete von Mises–Fisher filter for reference vector tracking is developed. In contrast to previous reference vector tracking methods such as [3], the model is specified as a stochastic differential equation that obeys the geometrical restrictions of the problem. Based on this, ordinary differential equations (ODEs) for the von Mises–Fisher parameters are derived and strategies for incorporating data from sensors such as accelerometers and magnetometers are developed using spherical likelihoods, which includes robust likelihoods. Furthermore, methods are provided for computing approximate marginal likelihoods that can be used for parameter estimation [7], [18], [19]. The methods are compared against the state-of-the-art both in simulated and real data scenarios. The experiments include local gravity tracking using simulated and real data, as well as sensor calibration using real data.

The rest of the paper is organised as follows, notation, problem formulation, and contributions are presented in Section II, the basics of the von Mises–Fisher distribution are outlined in Section III, the dynamic model is developed in Section IV, the filter is derived in Section V, experimental results are presented in Section VI, and lastly, the conclusions are given in Section VII.

II. Problem Formulation and Notation

A. Notation

Here some notation is established, $\mathbb{R}_+$ is the positive real half-line, $\mathrm{S}^2 \subset \mathbb{R}^p$ is the unit sphere, and $\mathrm{SO}(3)$ is the special orthogonal group acting on $\mathbb{R}^3$. For vectors $u, v \in \mathbb{R}^3$ the matrix $[u]_\times$ corresponds to the linear transform defined by cross-product from the left, $u \times v =$

$[u]_\times v$, for a set A, χ_A is its indicator function, for a function of a scalar variable, $\mathscr{V}\colon \mathbb{R} \to \mathbb{R}^p$, $\mathscr{V}'$ is its derivative, and ∂_t is used for time derivative. Furthermore, let $\{t_k\}_{k=1}^K,\ t_k < t_{k+1}$ be a subset of $\mathbb{R}_+$ and define the following sets for a stochastic process, $\{Y(t_k)\}_{k=1}^K$,

$$\mathscr{Y}(t) = \{y(t_k) \mid t_k \leq t\}, \tag{1a}$$

$$\mathscr{Y}(t^-) = \{y(t_k) \mid t_k < t\}. \tag{1b}$$

For random variables R and Y, $\mathbb{E}[R]$ is the expectation of R and $\mathbb{E}[R \mid Y]$ is the expectation of R conditioned on Y.

B. Problem formulation

In this paper, the problem of tracking a reference vector, r, using a three-axis rate gyro is considered. Without loss of generality let $r \in \mathrm{S}^2$, then the deterministic kinematics for r are given by [20]

$$\partial_t r(t) = -[\Omega(t)]_\times r(t) \tag{2}$$

where $\Omega(t)$ and $r(t)$ are the angular rate and the reference vector in the local frame, respectively. Furthermore, it is assumed that noisy measurements, $\breve{\Omega}$, of Ω are taken at a relatively high frequency to warrant the interpretation of $\breve{\Omega}$ as a continuous-time signal. The direction r is assumed to be measured at a set of discrete time instants, $\{t_k\}_{k=1}^K,\ t_k < t_{k+1}$,

$$f(y(t_k) \mid r(t_k)) = \exp\big(-\mathscr{V}(\rho_k^2)/2\big) \tag{3}$$

where

$$\rho_k^2 = ||y(t_k) - gQr(t_k) - b||^2/\sigma_Y^2, \tag{4}$$

$g \in \mathbb{R}_+$ is a gain (e.g., magnitude of the gravity vector), $Q \in \mathrm{SO}(3)$, and $\mathscr{V}\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a differentiable potential function, with derivative $\mathscr{V}'$. Note that Eq. (3) belongs to the class of spherical densities [21], in particular normal scale mixture densities (e.g Student's t distributions) are of this class [22], hence robust likelihoods are considered.

C. Contribution

The contributions of this paper are as follows:

- Using a gyroscope in dynamic replacement mode [20], a continuous-time model for the reference direction is developed, guaranteeing that the reference vector stays on S^2 with probability 1.
- The geometry of the problem is explicitly accounted for by using the von Mises–Fisher distribution, in contrast to Kalman based solutions [3].
- The kinematics are accounted for by formulating a continuous-time model, in difference to other von Mises–Fisher approaches [4], [13], [17].
- Approximate and exact updates and marginal likelihoods for spherical measurement densities are developed (including robust likelihoods), hence sensor calibration is possible.

III. The von Mises–Fisher Distribution

A random variable $R \in \mathrm{S}^2$ is said to be von Mises–Fisher (VMF) distributed, $R \sim \mathcal{VMF}(\mu, \eta)$, if its probability density function is given by [12]

$$f(r) = \mathcal{C}_3^{-1}(\eta)\exp\Big(\eta\mu^\mathsf{T} r\Big)\chi_{\mathrm{S}^2}(r), \tag{5}$$

where $\eta > 0$ a concentration parameter, $\mu \in \mathrm{S}^2$ determines the mode of the distribution, and $\mathcal{C}_3(\eta)$ is the normalization constant, given by

$$\mathcal{C}_3^{-1}(\eta) = \frac{\eta}{(4\pi\sinh\eta)}, \tag{6}$$

Furthermore, the derivative of $\log\mathcal{C}_3(\eta)$ is

$$A_3(\eta) := \partial_\eta \log\mathcal{C}_p(\eta) = \coth\eta - 1/\eta. \tag{7}$$

The expected value of a von Mises–Fisher distributed variable is given by [12], [23]

$$\mathbb{E}[R] = A_p(\eta)\mu. \tag{8}$$

IV. A Dynamic Model For Reference Vectors

In this section, a dynamic model suitable for tracking a reference vector is developed, in essence it is a modification of the model used in [3], with the added feature that the stochastic differential equation is norm preserving in Itô sense. While the kinematics for a local reference vector, $r(t)$, are given by Eq. (2), the angular rate, $\Omega(t)$, is rarely available, but rather a noisy version $\breve{\Omega}(t)$. This problem has previously been solved by adding a Wiener differential to the dynamics according to [3]

$$\mathrm{d}R(t) = -[\breve{\Omega}(t)]_\times R(t)\,\mathrm{d}t + \gamma\,\mathrm{d}W(t), \tag{9}$$

where $\breve{\Omega}(t)$ is the measured angular rate, $\gamma \in \mathbb{R}_+$, and $W(t)$ is a vector of independent standard Wiener processes. While Eq. (9) is a pragmatic model that allows for tracking using a Kalman filter [3], it does not properly account for the geometry, that is to say the Itô differential, $\mathrm{d}||R(t)||^2/2$, does not vanish. With this in mind, the following model for the reference direction is proposed:

$$\mathrm{d}R(t) = -\Big([\breve{\Omega}(t)]_\times + \gamma^2\mathrm{I}\Big)R(t)\,\mathrm{d}t + \gamma[R(t)]_\times\,\mathrm{d}W(t). \tag{10}$$

This model does indeed preserve the norm of $R(t)$, as asserted by Lemma 1 below.

Lemma 1. *Assume R(t) is governed by*

$$\mathrm{d}R(t) = -\Big([\breve{\Omega}(t)]_\times + \gamma^2\mathrm{I}\Big)R(t)\,\mathrm{d}t + \gamma[R(t)]_\times\,\mathrm{d}W(t),$$

with initial condition, $R(0)$*, such that* $R(0) \in \mathrm{S}^2$ *with probability 1, then* $R(t) \in \mathrm{S}^2, t \geq 0$ *with probability 1.*

Proof. As $R \in \mathrm{S}^2$ if and only if $||R(t)||^2 = 1$ it is sufficient to show that the Itô differential of $||R(t)||^2/2$ vanishes. From Itô's lemma it follows that

$$\begin{aligned}
\mathrm{d}\frac{||R(t)||^2}{2} &= -R^{\mathsf{T}}(t)\Big([\breve{\Omega}(t)]_\times + \gamma^2 \mathrm{I}\Big)R(t)\,\mathrm{d}t \\
&\quad + \frac{\gamma^2}{2}\mathrm{tr}\{[R(t)]_\times^{\mathsf{T}}[R(t)]_\times\}\,\mathrm{d}t \\
&\quad + R^{\mathsf{T}}(t)\gamma[R(t)]_\times\,\mathrm{d}W(t) \\
&= 0,
\end{aligned}$$

where it was used that $[\breve{\Omega}(t)]_\times$ is skew symmetric, $R(t) \times R(t) = 0$, and

$$\mathrm{tr}\{[R(t)]_\times^{\mathsf{T}}[R(t)]_\times\} = 2||R(t)||^2.$$

□

V. Continuous-Discrete von Mises–Fisher Filtering

The purpose of this section is to develop a von Mises–Fisher based assumed density filter, hence dynamics for the parameters need to be derived, as well as strategies for approximating Bayesian updates [24]. That is, approximations to the family of filtering densities

$$f(r, t \mid \mathscr{Y}(t)),\ t \in [t_1, t_K],$$

are sought, such that the approximation remains in the von Mises–Fisher class at all times. The resulting algorithm, using an explicit ODE solver together with the assumption that $\breve{\Omega}(t)$ and $Y(t)$ are synchronously sampled, is given in Alg. 1[1]. The remainder of this section is dedicated to the derivations and strategies for implementation.

A. Prediction

In order to derive a predictive distribution based on the von Mises–Fisher distribution an ODE for $\mathbb{E}[R \mid \mathscr{Y}(t)]$ can be derived that is valid in between adjacent measurement instants, $[t_k, t_{k+1})$. Taking the expectation of Eq. (10) and exploiting the martingale property of the Itô integral (see [25]) gives

$$\partial_t \mathbb{E}[R \mid \mathscr{Y}(t)] = -\Big([\breve{\Omega})]_\times + \gamma^2 \mathrm{I}\Big)\mathbb{E}[R \mid \mathscr{Y}(t)]. \quad (13)$$

Assume the filtering density is von Mises–Fisher,

$$f(r, t \mid \mathscr{Y}(t)) \sim \mathcal{VMF}(r; \bar{\mu}(t), \bar{\eta}(t)),$$

[1]This assumption is not necessary but makes for clearer presentation.

Algorithm 1 Continuous-Discrete von Mises–Fisher Filter (Explicit)

Input: Initial parameters $\bar{\mu}(t_0^-)$, $\bar{\eta}(t_0^-)$, and sampling intervals $\{\delta t_k\}_{k=0}^{K-1}$
Output: Filtering parameters $\{\bar{\mu}(t_k)\}_{k=0}^{K}$ and $\{\bar{\eta}(t_k)\}_{k=0}^{K}$
for $k = 0$ **to** $K-1$
{Predict}

$$\begin{aligned}
\bar{\eta}(t_{k+1}^-) &\leftarrow \exp\left(-\frac{\gamma^2 A_3(\bar{\eta}(t_k))\delta t_k}{A_3'(\bar{\eta}(t_k))\bar{\eta}(t_k)}\right)\bar{\eta}(t_k) \\
\breve{\theta}(t_k) &\leftarrow ||\breve{\Omega}(t_k)||\delta t_k \\
\bar{\mu}(t_{k+1}^-) &\leftarrow \bar{\mu}(t_k) - \frac{\sin\breve{\theta}(t_k)\delta t_k}{\breve{\theta}(t_k)}\breve{\Omega}(t_k) \times \bar{\mu}(t_k) \\
&\quad + \frac{(1-\cos\breve{\theta}(t_k))(\delta t_k)^2}{\breve{\theta}(t_k)^2}\breve{\Omega}(t_k) \times \breve{\Omega}(t_k) \times \bar{\mu}(t_k)
\end{aligned}$$

{Update}

$$\begin{aligned}
\hat{\rho}^2 &\leftarrow ||y(t_{k+1}) - gQ\mu(t_{k+1}^-) - b||^2/\sigma_Y^2 \\
\xi(t_{k+1}) &\leftarrow g/\sigma_Y^2 \mathscr{V}'(\hat{\rho}^2)Q^{\mathsf{T}}(y(t_{k+1}) - b) \\
\bar{\eta}(t_{k+1}) &\leftarrow ||\bar{\eta}(t_{k+1}^-)\bar{\mu}(t_{k+1}^-) + \xi(t_{k+1})|| \\
\bar{\mu}(t_{k+1}) &\leftarrow (\bar{\eta}(t_{k+1}^-)\bar{\mu}(t_{k+1}^-) + \xi(t_{k+1}))/\bar{\eta}(t_{k+1})
\end{aligned}$$

end

then it follows from Eq. (8) that

$$\mathbb{E}[R \mid \mathscr{Y}(t)] = A_3(\bar{\eta}(t))\bar{\mu}(t) \quad (14a)$$
$$||\mathbb{E}[R \mid \mathscr{Y}(t)]|| = A_3(\bar{\eta}(t)). \quad (14b)$$

Taking the time derivative of $A_3(\bar{\eta}(t))$ and using the norm identity in Eq. (14) gives

$$\begin{aligned}
A_3'(\bar{\eta}(t))\partial_t\bar{\eta}(t) &= \frac{\langle \partial_t \mathbb{E}[R \mid \mathscr{Y}(t)], \mathbb{E}[R \mid \mathscr{Y}(t)]\rangle}{||\mathbb{E}[R \mid \mathscr{Y}(t)]||} \\
&= -\gamma^2 A_3(\bar{\eta}(t)),
\end{aligned}$$

where it was used that $\langle u, [v]_\times u\rangle = 0$ for all vectors u and v, hence the predictive ODE for $\bar{\eta}$ is given by

$$\partial_t \bar{\eta}(t) = -\gamma^2 A_3(\bar{\eta}(t))/A_3'(\bar{\eta}(t)). \quad (15)$$

The ODE for $\bar{\mu}(t)$ is obtained by taking the time derivative of both sides of the mean identity in Eq. (14)

$$\begin{aligned}
\partial_t \mathbb{E}[R \mid \mathscr{Y}(t)] &= A_3'(\bar{\eta}(t))\partial_t\bar{\eta}(t)\bar{\mu}(t) + A_3(\bar{\eta}(t))\partial_t\bar{\mu}(t) \\
&= -\gamma^2 A_3(\bar{\eta}(t))\bar{\mu}(t) + A_3(\bar{\eta}(t))\partial_t\bar{\mu}(t).
\end{aligned}$$

Re-arranging terms and using Eq. (15) gives the following expression for $\partial_t\bar{\mu}(t)$,

$$\partial_t\bar{\mu}(t) = \Big(\partial_t \mathbb{E}[R \mid \mathscr{Y}(t)] + \gamma^2 A_3(\bar{\eta}(t))\bar{\mu}(t)\Big)/A_3(\bar{\eta}(t)).$$

Continuing by inserting the expression for $\partial_t\mathbb{E}[R \mid \mathscr{Y}(t)]$ given in Eq. (13) and using the mean identity in Eq. (14) gives the end result as follows

$$\partial_t\bar{\mu}(t) = -[\breve{\Omega}(t)]_\times\bar{\mu}(t). \tag{16}$$

In summary, the prediction equations for the von Mises–Fisher parameters are given by

$$\partial_t\bar{\mu}(t) = -[\breve{\Omega}(t)]_\times\bar{\mu}(t) \tag{17a}$$
$$\partial_t\bar{\eta}(t) = -\gamma^2 A_3(\bar{\eta}(t))/A_3'(\bar{\eta}(t)). \tag{17b}$$

The issue of efficiently integrating the ODEs for $\bar{\mu}(t)$ and $\bar{\eta}(t)$ shall be returned to in Section V-C, after examining the problem of measurement updates in Section V-B.

B. Measurement update and marginal likelihood

Here, schemes for approximating the filtering distribution and the marginal likelihood, assuming the predictive distribution is in the von Mises–Fisher class and the measurement is generated by Eq. (3) are developed, such that the filtering distribution remains in the von Mises–Fisher class. For this end, the special case of spherical Gaussian likelihoods is examined first, where exact relations for the filtering distribution and marginal likelihoods is available, this is Proposition 1 below.

Proposition 1. *Let $R \sim \mathcal{VMF}(\mu,\eta)$, $Y \mid R = r \sim \mathcal{N}(y; gQr + b, \sigma_Y^2\mathrm{I})$, with $g \in \mathbb{R}_+$, $b \in \mathbb{R}^3$ and $Q \in \mathrm{SO}(3)$, then*

$$f(r \mid y) = \mathcal{VMF}(r; \mu_{R|Y}, \eta_{R|Y}), \tag{18a}$$
$$f(y) = \frac{\exp\Big(-\frac{||y-b||^2+g^2}{2\sigma_Y^2}\Big)}{(2\pi\sigma_Y^2)^{3/2}}\frac{\mathcal{C}_3(\eta_{R|Y})}{\mathcal{C}_3(\eta)}, \tag{18b}$$

where

$$\eta_{R|Y} = ||g/\sigma_Y^2 Q^\mathsf{T}(y-b) + \eta\mu|| \tag{19a}$$
$$\mu_{R|Y} = (g/\sigma_Y^2 Q^\mathsf{T}(y-b) + \eta\mu)/\eta_{R|Y}. \tag{19b}$$

Proof. According to Baye's theorem, the posterior is given by

$$\begin{aligned} f(r \mid y) &\propto \frac{\exp\Big(-\frac{||y-gQr-b||^2}{2\sigma_Y^2} + \eta\mu^\mathsf{T}r\Big)}{\mathcal{C}_3(\eta)(2\pi\sigma_Y^2)^{3/2}}\chi_{\mathrm{S}^2}(r) \\ &= \frac{\exp\Big(-\frac{||y-b||^2+g^2||r||^2}{2\sigma_Y^2} + \eta_{R|Y}\mu_{R|Y}^\mathsf{T}r\Big)}{\mathcal{C}_3(\eta)(2\pi\sigma_Y^2)^{3/2}}\chi_{\mathrm{S}^2}(r) \end{aligned}$$

However, as $||r||^2$ is constant on the domain of $f(r \mid y)$, due to the indicator function, it follows that

$$f(r \mid y) = \mathcal{VMF}(r; \mu_{R|Y}, \eta_{R|Y}),$$

with the parameters given in Eq. (19). Furthermore, integrating gives the results in Eq. (18). ☐

The results of Proposition 1 can be used to produce an approximate posterior in the von Mises–Fisher class for the measurement model in Eq. (3). Using a similar idea to that of [26], a Taylor series of $\mathscr{V}(\rho^2)$ is given by

$$\mathscr{V}(\rho^2) \approx \mathscr{V}(\hat{\rho}) + \mathscr{V}'(\hat{\rho})(\rho^2 - \hat{\rho}^2) + \mathscr{O}\Big((\rho^2-\hat{\rho}^2)^2\Big).$$

A readily available linearisation point, $\hat{\rho}^2$, would be to evaluate ρ^2 at the prior mode,

$$\hat{\rho}^2 = ||y - gQ\mu - b||^2/\sigma_Y^2.$$

Truncating at the first order, an approximation of the likelihood in Eq. (3) is given by

$$f(y \mid r) \approx \exp\Big(-||y-gQr-b||^2/(2\sigma_Y^2/\mathscr{V}'(\hat{\rho}^2)) + \hat{\psi}\Big),$$

where $\hat{\psi}$ is given by

$$\hat{\psi} = -\frac{1}{2}\Big(\mathscr{V}(\hat{\rho}^2) - \mathscr{V}'(\hat{\rho}^2)\hat{\rho}^2\Big).$$

Repeating the argument in Proposition 1 gives the following approximations.

Approximation 1. *Let $R \sim \mathcal{VMF}(\mu,\eta)$, $f(y \mid r) = \exp(-\mathscr{V}(\rho^2)/2)$, with $g \in \mathbb{R}_+$, $b \in \mathbb{R}^3$ and $Q \in \mathrm{SO}(3)$, then*

$$f(r \mid y) \approx \mathcal{VMF}(r; \hat{\mu}_{R|Y}, \hat{\eta}_{R|Y}),$$
$$f(y) \approx \frac{\exp\Big(-\frac{||y-b||^2+g^2}{2\sigma_Y^2}\Big)}{\exp(-\hat{\psi})}\frac{\mathcal{C}_3(\hat{\eta}_{R|Y})}{\mathcal{C}_3(\eta)},$$

where

$$\hat{\eta}_{R|Y} = ||g/\sigma_Y^2\mathscr{V}'(\hat{\rho}^2)Q^\mathsf{T}(y-b) + \eta\mu||$$
$$\hat{\mu}_{R|Y} = (g/\sigma_Y^2\mathscr{V}'(\hat{\rho}^2)Q^\mathsf{T}(y-b) + \eta\mu)/\hat{\eta}_{R|Y}.$$

For example, if $f(y \mid r)$ is a Student's t density with ν degrees of freedom, then

$$\mathscr{V}'(\hat{\rho}^2) = (\nu+3)/(\nu+\hat{\rho}^2),$$

and the posterior reverts to the prior for large $||y||$.

C. Implementation considerations

In order to implement the filter, a zeroth order hold method is recommended for the direction parameter $\bar{\mu}(t)$, that is assume $\breve{\Omega}(t)$ is constant on the intervals $[t_k, t_{k+1})$ and define $\delta t_k = t_{k+1} - t_k$, then

$$\bar{\mu}(t_{k+1}^-) = \exp\Big(-[\breve{\Omega}(t_k)]_\times\delta t_k\Big)\bar{\mu}(t_k).$$

Using Rodriguez' formula, this simplifies to the expression given in Alg. 1 (see e.g [20]). Hence the evolution of $\bar{\mu}(t)$ can be implemented with just 2 cross-products, 2 evaluations of elementary functions, and 2 vector additions.

On the issue of numerically adequate integrators for $\bar{\eta}(t)$, assume $\bar{\eta}(t) \neq 0$ on the intervals $[t_k, t_{k+1})$. The differential equation for $\bar{\eta}(t)$ can then be written as

$$\partial_t \bar{\eta}(t) = -\gamma^2 A_3(\bar{\eta}(t))/[A_3'(\bar{\eta}(t))\bar{\eta}(t)]\bar{\eta}(t). \tag{22}$$

The solution at t_{k+1} can then be written as

$$\bar{\eta}(t_{k+1}^-) = \exp\left(- \int_{t_k}^{t_{k+1}} \frac{\gamma^2 A_3(\bar{\eta}(\tau))}{A_3'(\bar{\eta}(\tau))\bar{\eta}(\tau)} \, \mathrm{d}\tau \right) \bar{\eta}(t_k).$$

The integral in the exponent can be approximated by the rectangle rule yielding the explicit scheme given in Alg. 1. On the other hand, if the trapezoidal rule is used, an implicit scheme is obtained, which can be solved by fixed-point iteration.

Approximation 2. *An implicit scheme, based on the trapezoidal rule, for integrating Eq.* (22) *is given by*

$$\bar{\eta}^{(j+1)}(t_{k+1}^-) \approx \exp\left(- \frac{\gamma^2 A_3(\bar{\eta}^{(j)}(t_{k+1}^-))\delta t}{2A_3'(\bar{\eta}^{(j)}(t_{k+1}^-))\bar{\eta}^{(j)}(t_{k+1}^-)} \right) \times \exp\left(- \frac{\gamma^2 A_3(\bar{\eta}(t_k))\delta t}{2A_3'(\bar{\eta}(t_k))\bar{\eta}(t_k)} \right) \bar{\eta}(t_k).$$

Note that the ratio $A_3(\eta)/A_3'(\eta)/\eta$ can be written as

$$A_3(\eta)/A_3'(\eta)/\eta = \frac{1 - \eta \coth \eta}{\eta^2 \operatorname{csch}^2 \eta - 1},$$

where $\coth$ and csch are the hyperbolic cotangent and hyperbolic cosecant functions, respectively.

VI. Experimental Results

The proposed method shall be validated in simulation studies as well as a real data experiment using a smartphone, both pertaining to the estimation of the local gravity vector. Hence, in the simulations, the performance is assessed in terms of reconstruction accuracy of the gravity vector as given by the norm error,

$$\varepsilon(t_k) = ||gr(t_k) - \mathbb{E}[gR(t_k) \mid \mathscr{Y}(t_k)]||,$$

where $r(t)$ is the normalised local gravity vector and g its magnitude ($g \approx 9.82\,\mathrm{m/s^2}$).

A. Simulations: gravity vector tracking

In this experiment the following system is considered

$$\mathrm{d}\Omega(t) = -1.5(s(t) - \Omega(t))\,\mathrm{d}t + \mathrm{d}B(t)/100, \tag{23a}$$

$$\mathrm{d}R(t) = -[\Omega(t)]_\times R(t)\,\mathrm{d}t, \tag{23b}$$

$$\breve{\Omega}(t_k) = \Omega(t_k) + \breve{V}(t_k),\ \breve{V}(t_k) \sim \mathcal{N}(0, \sigma_\omega^2 \mathrm{I}), \tag{23c}$$

$$Y(t_k) \mid R(t_k) = r \sim \exp\big(- \mathscr{V}(\rho_k^2)/2 \big), \tag{23d}$$

where $\rho_k^2 = ||y(t_k) - gr||/\sigma_a^2$ and

$$s(t) = -1/10\mathrm{e}_3\chi_{[0,15]}(t) + 1/10\mathrm{e}_1\chi_{(15,30]}(t) + 1/10\mathrm{e}_2\chi_{(30,45]}(t) + 1/10\mathrm{e}_3\chi_{(45,60]}(t).$$

Note that the stochastic differential equations in Eq. (23a) are of state-independent diffusion type, hence the Stratonovich and the Itô interpretations coincide. Therefore, the system can be simulated using the fully implicit midpoint rule [27], that preserves quadratic invariants (i.e the norm of R). The system is simulated with a sampling interval of $\delta_s t = 2 \times 10^{-2}$ and the signals are then downsampled by a factor $M = 3$. The system parameters are $\theta = (\sigma_\omega^2, \sigma_a^2)$ and the filter will be assessed for the parameter settings, $\theta_1 = (1 \times 10^{-2}, 1 \times 10^{-1})$, $\theta_2 = (1 \times 10^{-2}, 1 \times 10^{-2})$, $\theta_3 = (1 \times 10^{-4}, 1 \times 10^{-1})$, and $\theta_4 = (1 \times 10^{-4}, 1 \times 10^{-2})$. For all parameter settings, $\mathscr{V}$ is set to correspond to Student's t distribution with $\nu = 3$ degrees of freedom.

The system is simulated 100 times on the interval $t \in [0, 60]$s for each parameter setting and two implementations of the present filter are considered, one using the explicit rule for integrating $\bar{\eta}(t)$ (CT1) and the other using the implicit rule (CT2). These implementations are compared to two versions of the discrete-time von Mises–Fisher filter, one using moment-matching (DT1) (see e.g. [4]) and the other using score-matching (DT2) [17], both these implementations use an Euler–Maruyama discretisation of Eq. (10) to compute the predictive moments. Five iterations are used for the fixed-point and Newton iterations of CT2 and DT1, respectively. All the von Mises–Fisher based filters use the update scheme presented in Section V-B and they are initialised with $\mu(t_0)$ sampled uniformly on S^2 and $\eta(t_0) = 1 \times 10^{-3}$. The last competitor is a Kalman filter (KF) using the implementation in [3], with the addition of using 5 variational Bayesian iterations for updating using Student's t distributed measurements (see [28], [29]). The Kalman filter mean was initialised as $\mu(t_0)$ with an initial covariance of $1 \times 10^3\mathrm{I}$. The parameter γ was set to $\gamma = 5 \times 10^{-2}$ for all filters and all parameter settings.

The error, $\varepsilon(t_k)$, averaged over time points and Monte Carlo trials is shown in Table I for the different filter candidates. As can be seen, CT2 performs the best, followed by CT1, and then DT1. On the other hand, sometimes DT2 outperforms KF, though it can also perform significantly worse. This is in contrast to findings in [17] where DT1 and DT2 performed similarly, however the simulation setting there is also significantly different. Additionally, the logarithm of the Monte Carlo averaged trajectories of $\varepsilon(t)$ is plotted in Fig. 1. Furthermore, to gain insight into how $\bar{\eta}$ behaves for the different von Mises–Fisher based filters, the logarithm of Monte Carlo averaged trajectories of the aforementioned is shown in Fig. 2. It appears that DT2 underestimates $\bar{\eta}$ at the prediction step, while DT1 overestimates. CT1 and CT2 have very similar trajectories for $\bar{\eta}(t)$ with the latter always being slightly larger. Note that the experiments use a fairly small interval between measurements ($\delta t \approx 6 \times 10^{-3}$ s). It is expected that the contrast between the continuous and discrete time filters grows for larger intervals between measurements.

The computational speed of the filters is shown in Table II. As can be seen, DT1 is the fastest, followed by CT1 and CT2. DT2 is slow due to a matrix inversion in the prediction. KF is the slowest due to matrix computations in both prediction and update step, particularly the latter due to variational Bayesian iterations.

Table I
MEAN NORM ERROR OVER MONTE-CARLO TRIALS AND TIME FOR ALL THE COMPETING FILTERS.

$\theta\#$	CT1	CT2	DT1	DT2	KF
θ_1	0.1176	**0.1174**	0.1189	0.2381	0.1533
θ_2	0.0607	**0.0589**	0.0610	0.0778	0.0859
θ_3	0.1167	**0.1164**	0.1183	0.2381	0.1527
θ_4	0.0600	**0.0581**	0.0606	0.0777	0.0855

Table II
NUMBER OF SAMPLES PROCESSED PER SECOND FOR THE FILTERS.

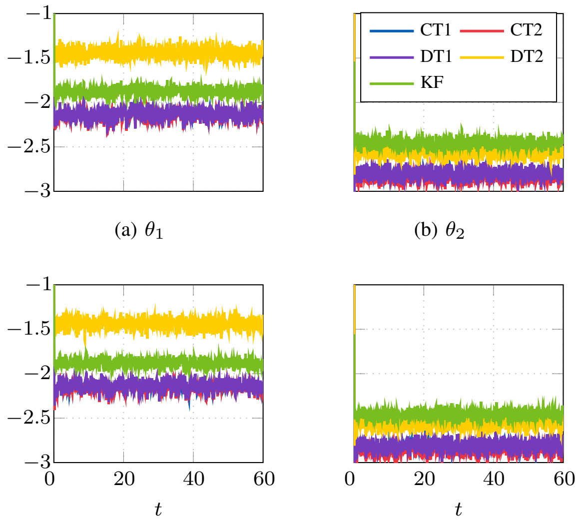
	CT1	CT2	DT1	DT2	KF
1×10^4/s	7.6344	7.0929	**9.2602**	2.881	0.7305

B. Accelerometer and magnetometer calibration

In the second example, the proposed method is used for calibrating the accelerometer and magnetometer of a smartphone [18], [19]. In this case, the objective is to estimate the scale factors (i.e. the magnitude of the gravity g and magnetic field m, which may not correspond to their nominal values due to sensor inaccuracies) as well as the biases (b_a and b_m). The considered measurement models are

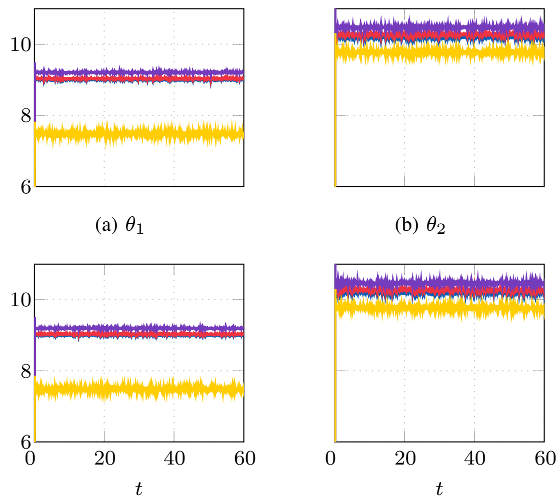
$$\begin{aligned} Y_a(t_k) &= gR_a(t_k) + b_a + V_a(t_k), \\ Y_m(t_k) &= mR_m(t_k) + b_m + V_m(t_k), \end{aligned}$$

where $V_a(t_k)$ and $V_m(t_k)$ are mutually uncorrelated, Gaussian, white noise sequences with covariances $\mathrm{I}\sigma_a^2$ and $\mathrm{I}\sigma_m^2$, respectively. Note that possible axis misalignment is neglected (see, e.g. [19] for details).



(a) θ_1 (b) θ_2 (c) θ_3 (d) θ_4

Figure 1. The logarithm of the error trajectories averaged over Monte Carlo trials for the different parameter settings. Note that CT1 is concealed behind CT2 and DT1.



(a) θ_1 (b) θ_2 (c) θ_3 (d) θ_4

Figure 2. The logarithm of the average concentration parameter, $\bar{\eta}(t)$, over Monte-Carlo trials for the different parameter settings.

For calibration, a Huawei Nexus 6P was strapped to a cardboard box, which was gradually rotated around all its faces, keeping it stationary for about 5 s on each face. The initial bias and reference vector magnitudes was then estimated using traditional sphere fitting [18], [19], followed by maximizing the marginal log-likelihood as

Table III
NEGATIVE MARGINAL LOG-LIKELIHOOD OF THE VALIDATION DATA (LOWER IS BETTER).

	Accelerometer	Magnetometer
Uncalibrated	1.88×10^4	1.35×10^6
Sphere fit	1.87×10^4	2.97×10^4
Maximizing (18)	1.80×10^4	1.61×10^4

described in Section V. In the latter stage, the diffusion coefficient and the measurement noise variances are also estimated. The performance is evaluated by applying the proposed algorithm on an independent validation dataset and comparing the marginal log-likelihood to the uncalibrated (i.e. zero bias and nominal values for gravity and magnetic field strength) and calibrated using sphere fitting cases. The validation dataset was collected by arbitrarily rotating the phone around its own axes without significant displacement.

Table III shows the negative marginal log-likelihood of the validation data. As it can be seen, the proposed algorithm performs best in both cases. For the accelerometer, the negative log-likelihood only decreases slightly (roughly $4\,\%$ compared to the uncalibrated case). For the magnetometer, however, the increase is much more significant. The improvement from uncalibrated to the sphere fit calibration is around $78\,\%$ and $46\,\%$ from sphere fitting calibration to calibration using the proposed approach. This is not a surprising result, since accelerometers generally suffer from less sensor inaccuracies compared to magnetometers.

C. Gravity tracking in smartphones

Similar to the simulation example above, tracking the gravity vector is considered in the second real data example. Specifically, the proposed method is evaluated on two datsets gathered from a smartphone (Huawei Nexus 6P). The first dataset corresponds to the typical motion pattern when answering the phone: Picking up the phone from a table, bringing it to the ear, and putting it back to the table. In the second dataset, the phone is held approximately constant in front of a person while walking. These two mundane tasks pose challenging problems for gravity tracking due to the significant acceleration components superposed. The proposed method (CT2) is compared to the Kalman filter-based gravity tracking algorithm specifically designed for this purpose in [3].

Fig. 3 shows the measurement data for the first experiment, together with the estimated (filtered) gravity components. It can be seen that despite the covariance adaptation made by the Kalman filter (see [3] for details),

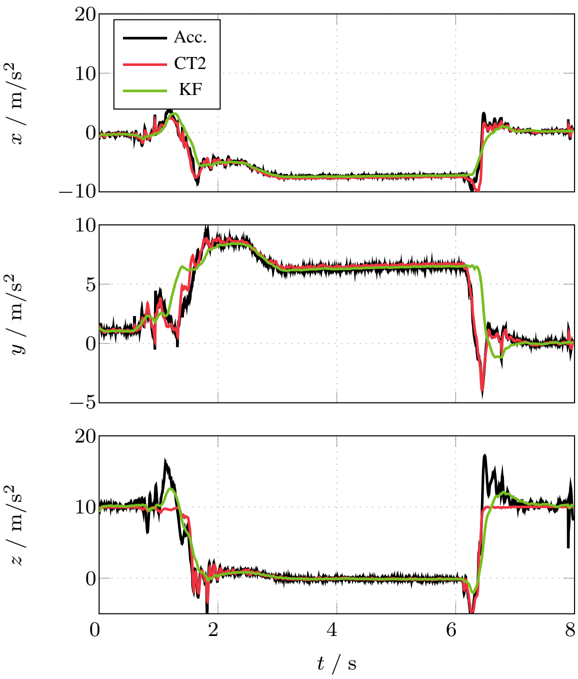


Figure 3. Accelerometer signals and the estimated gravity components for the first data set (answering the phone).

this algorithm has the tendency to absorb the extra acceleration into the gravity estimate, since it is not constrained in magnitude (e.g. around $t \approx 3\,\mathrm{s}$). Furthermore, the covariance adaptation scheme also introduces a certain lag in the tracking (e.g. around $t \approx 6.5\,\mathrm{s}$).

Similar results are observed in the second experiment (walking and observing the screen) as depicted in Fig. 4. Walking causes very strong accelerations, which significantly affect the Kalman filter and thus affect the estimated gravity vector. This effect could be reduced by increasing the measurement noise covariance, which, however, would increase the lag for tracking the gravity vector in the case when it actually changes. The proposed method (CT2) on the other hand, is not significantly affected by the extra acceleration present due to walking.

VII. CONCLUSION

A continuous-discrete von Mises–Fisher filter was developed for spherical measurements distributions. The method was validated in simulations, in sensor calibration using data collected with a smartphone, as well as gravity vector tracking for two smartphone use cases, showing superior performance to state-of-the-art in all of the experiments. It was found to be particularly robust to unmodeled accelerations in the smartphone experiments.

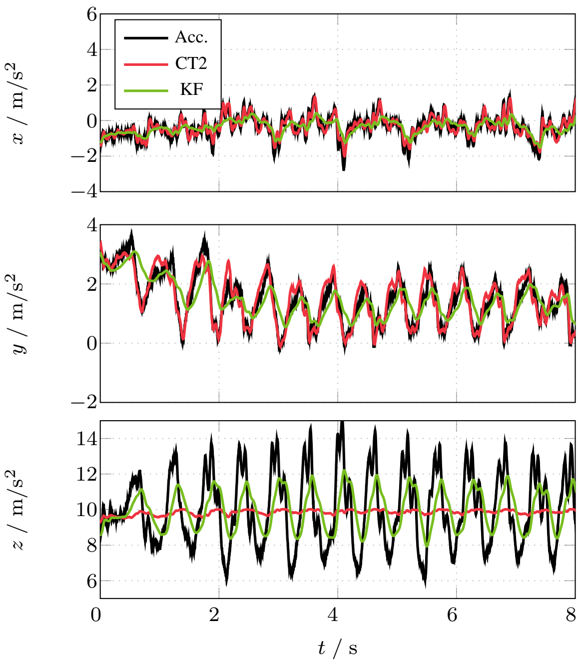


Figure 4. Accelerometer signals and the estimated gravity components for the second data set (walking).

Future work involves online calibration as well as handling elliptic likelihoods.

Acknowledgment

Financial support by the Academy of Finland and Aalto ELEC Doctoral School is acknowledged.

References

[1] Z. Xiao, H. Wen, A. Markham, and N. Trigoni, "Robust pedestrian dead reckoning (R-PDR) for arbitrary mobile device placement," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2014, pp. 187–196.

[2] I. Vallivaara, J. Haverinen, A. Kemppainen, and J. Röning, "Simultaneous localization and mapping using ambient magnetic field," in *2010 IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2010, pp. 14–19.

[3] S. Särkkä, V. Tolvanen, J. Kannala, and E. Rahtu, "Adaptive Kalman filtering and smoothing for gravitation tracking in mobile systems," in *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct 2015, pp. 1–7.

[4] I. Marković, F. Chaumette, and I. Petrović, "Moving object detection, tracking and following using an omnidirectional camera on a mobile robot," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 5630–5635.

[5] 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)*, Sept 2014, pp. 1–6.

[6] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *2011 IEEE International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2011, pp. 1–7.

[7] J. D. Hol, "Sensor fusion and calibration of inertial sensors, vision, ultra-wideband and gps," Ph.D. dissertation, Linköping University Electronic Press, 2011.

[8] J.-O. Nilsson, A. K. Gupta, and P. Händel, "Foot-mounted inertial navigation made easy," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2014, pp. 24–29.

[9] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons, 2007.

[10] D. Titterton and J. L. Weston, *Strapdown inertial navigation technology*. IET, 2004, vol. 17.

[11] S. R. Fisher, "Disperson on a sphere," *Proceedings of the Royal Society A, Mathematical, Physical and Engineering Sciences*, vol. 217, no. 1130, pp. 214–222, 1953.

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

[13] I. Marković, M. Bukal, J. Ćesić, and I. Petrović, "Direction-only tracking of moving objects on the unit sphere via probabilistic data association," in *17th International Conference on Information Fusion (FUSION)*, July 2014, pp. 1–7.

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

[15] A. Hyvärinen, "Estimation of non-normalized statistical models by score matching," *Journal of Machine Learning Research*, vol. 6, no. Apr, pp. 695–709, 2005.

[16] K. V. Mardia, J. T. Kent, and A. K. Laha, "Score matching estimators for directional distributions," *arXiv preprint arXiv:1604.08470*, 2016.

[17] 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 (FUSION)*. IEEE, 2017, pp. 1–6.

[18] J. L. Crassidis, K.-L. Lai, and R. R. Harman, "Real-time attitude-independent three-axis magnetometer calibration," *Journal of Guidance, Control, and Dynamics*, vol. 28, no. 1, pp. 115–120, 2005.

[19] M. Kok and T. B. Schön, "Magnetometer calibration using inertial sensors," *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5679–5689, July 2016.

[20] L. F. Markley and J. L. Crassidis, *Fundamentals of Spacecraft Attitude Determination and Control*. Springer, 2014.

[21] S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *Journal of Multivariate Analysis*, vol. 11, no. 3, pp. 368–385, 1981.

[22] O. Barndorff-Nielsen, J. Kent, and M. Sorensen, "Normal variance-mean mixtures and z distributions," *International Statistical Review*, vol. 50, no. 2, pp. 145–159, 1982.

[23] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra, "Clustering on the unit hypersphere using von Mises–Fisher distributions," *Journal of Machine Learning Research*, vol. 6, pp. 1345–1382, 2005.

[24] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.

[25] B. Øksendal, *Stochastic Differential Equations - An Introduction with Applications*. Springer, 2003.

[26] J. Durbin and S. J. Koopman, "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 1, pp. 3–56, 2000.

[27] A. Abdulle, D. Cohen, G. Vilmart, and K. C. Zygalakis, "High order weak methods for stochastic differential equations," OCCAM

Oxford Centre for Collaborative Applied Mathematics, Tech. Rep. 11/48, 2011.

[28] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, 2009.

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