Joint Calibration of Inertial Sensors and Magnetometers using von Mises–Fisher Filtering and Expectation Maximization

Roland Hostettler, Ángel García-Fernández, Filip Tronarp, and Simo Särkkä

Joint Calibration of Inertial Sensors and Magnetometers using von Mises–Fisher Filtering and Expectation Maximization

Roland Hostettler*, Ángel F. García-Fernandez†, Filip Tronarp*, and Simo Särkkä*
*Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland
E-mail: { roland.hostettler, filip.tronarp, simo.sarkka }@aalto.fi
†Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K.
E-mail: angel.garcia-fernandez@liverpool.ac.uk

***Abstract*—Microelectromechanical-systems-based inertial sensors and magnetometers are low-cost, off-the-shelf sensors that are widely used in both consumer and industrial applications. However, these sensors suffer from biases and effects such as axis misalignment or scale errors, which require careful system design and periodic sensor calibration. In this paper, we propose a fast calibration method for jointly calibrating inertial sensors and magnetometers based on discrete-time von Mises–Fisher filtering and expectation maximization. We evaluate the method on built-in smartphone sensors and show that the proposed method efficiently estimates the sensors' parameters and reduces the overall error.**

***Index Terms*—Magnetometers, inertial sensors, expectation-maximization, calibration, von Mises–Fisher filtering**

I. Introduction

Recent advances in low-cost manufacturing methods have enabled the use of inertial measurement units (IMUs) and magnetometers in a wide range of novel motion sensing applications. Examples of such applications include pedestrian navigation [1], [2], rehabilitation [3], [4], or cardiac signal reconstruction in electrocardiography [5], [6]. An important challenge when using IMUs for motion sensing tasks is, however, their proper calibration [7]–[9]. Accelerometers, gyroscopes, and magnetometers typically suffer from sensor biases or drifts, scale errors, or axis misalignment. If the sensors are not calibrated properly, motion estimation may degrade severely and the overall system performance can be greatly reduced. This problem is particularly pronounced in low-cost sensors that are typically employed in smartphones or wearable devices.

Calibration is most important for magnetometers, since the sensing elements are easily affected by surrounding ferromagnetic materials and external magnetic fields other than the earth's magnetic field. Hence, many calibration algorithms have been developed for magnetometers. The first calibration algorithms were based on estimating the biases by comparing the measured heading to the true heading [10], which requires knowledge of the latter that might be difficult to obtain. However, since the measurements of a properly calibrated sensor that is rotated around its own axes should lie on a sphere, the magnetometer calibration problem can be considered a sphere-fitting problem, which can be implemented based on batch processing [11]–[14] or Kalman filtering [14], [15]. These algorithms do not require an external reference, that is, the magnetometer can be calibrated without knowledge of the true heading and inclination. Furthermore, while these methods were derived for calibration of magnetometers, the same principle can be used for the calibration of accelerometers. However, these approaches apply nonlinear transformations to the noisy measurement data which leads to a biased parameter estimate [16].

To overcome this problem, the calibration problem can also be solved using state augmentation in Kalman filtering [17], [18] or by maximum likelihood estimation [19]. The latter approach guarantees that the calibration parameters are asymptotically unbiased. It can be further improved by formulating the problem using a state-space model that uses gyroscope measurements to more accurately estimate the rotation. This yields a comprehensive calibration method that is able to exploit and jointly calibrate the complete IMU and magnetometer measurement system [20]–[22]. Solving the direct maximum likelihood problem, however, requires solving a nonlinear and non-convex optimization problem, which is not guaranteed to converge. Furthermore, Kalman-filtering-based approaches normally represent the sensor attitude using quaternions, which can not be handled properly using standard Kalman filtering due to the unit sphere constraints [23].

This paper addresses some of these drawbacks by proposing a new calibration method for inertial sensors and magnetometers based on von Mises–Fisher [24] assumed density filtering and expectation maximization (EM), which iteratively approximates the asymptotically unbiased maximum likelihood solution. In particular, we formulate the calibration problem as a reference vector tracking problem on the unit sphere using a state-space model parametrized by the unknown sensor parameters. We develop a von Mises–Fisher assumed density filter for inference based on the resulting model and use EM to estimate the parameters.

The remainder of this paper is organized as follows. Section II presents the dynamic and measurement models as well as the problem formulation. The von Mises–Fisher filter and EM

update are derived in Section III and Section IV, respectively. The proposed method is evaluated on real data examples in Section V and some concluding remarks follow in Section VI.

II. Model and Problem Formulation

In this section, the systems' dynamic and sensor models are derived and the estimation problem is formulated. The basic underlying assumptions are that there is no significant translational motion present during the calibration sequence and that the sensors do not suffer from axis misalignment. Furthermore, for the magnetometer, we also assume that no external magnetic disturbances are present.

A. Dynamic Rotation Model for Reference Vectors

Accelerometers and magnetometers can be used to track the earth's gravity and magnetic field vectors. In the sensors' local coordinate frame, these vectors can be expressed as a scale factor times a unit reference vector, that is, as

$$\begin{aligned} \boldsymbol{g}^{\mathsf{L}}(t) &= g\,\boldsymbol{r}_a(t), \\ \boldsymbol{B}^{\mathsf{L}}(t) &= B\,\boldsymbol{r}_m(t), \end{aligned}$$

where $\boldsymbol{g}^{\mathsf{L}}(t)$ and $\boldsymbol{B}^{\mathsf{L}}(t)$ are the local gravity and magnetic field vectors, respectively, $\boldsymbol{r}_a(t) \in \mathbb{R}^3$ with $\|\boldsymbol{r}_a(t)\| = 1$ is the unit vector for the accelerometer, g is the accelerometer scale factor, and $\boldsymbol{r}_m(t) \in \mathbb{R}^3$ ($\|\boldsymbol{r}_m(t)\| = 1$) and B are the corresponding unit vector and scale factor for the magnetic field, respectively.

Furthermore, the dynamic model for rotations of a unit vector $\boldsymbol{r}_j(t)$ (for $j \in \{a, m\}$) with the local rotational velocity vector $\boldsymbol{\omega}(t)$ is given by [9]

$$\dot{\boldsymbol{r}}_j(t) = -\boldsymbol{\omega}(t) \times \boldsymbol{r}_j(t),$$

where $\dot{\boldsymbol{r}}(t)$ denotes the time-derivative of $\boldsymbol{r}(t)$, and $\boldsymbol{a} \times \boldsymbol{b}$ is the cross-product of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

Assuming that the rotational velocity is constant on the interval $(t_{n-1}, t_n]$ (zero-order hold) such that $\boldsymbol{\omega}(t) = \boldsymbol{\omega}_n$ on the interval $(t_{n-1}, t_n]$, the continuous-time dynamic model has the discrete-time equivalent [25]

$$\begin{aligned} \boldsymbol{r}_{j,n} &= \Bigg(\boldsymbol{I}_3 + \sin(\|\boldsymbol{\omega}_n\|\Delta t) \frac{[-\boldsymbol{\omega}_n]_\times}{\|\boldsymbol{\omega}_n\|} \\ &\quad + (1 - \cos(\|\boldsymbol{\omega}_n\|\Delta t)) \frac{[-\boldsymbol{\omega}_n]_\times^2}{\|\boldsymbol{\omega}_n\|^2} \Bigg) \boldsymbol{r}_{j,n-1} \\ &= \boldsymbol{F}(\boldsymbol{\omega}_n)\,\boldsymbol{r}_{j,n-1}, \end{aligned} \tag{1}$$

where $\boldsymbol{r}_{j,n} \triangleq \boldsymbol{r}_j(t_n)$ with

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{\omega}_n) \triangleq \boldsymbol{I}_3 &+ \sin(\|\boldsymbol{\omega}_n\|\Delta t) \frac{[-\boldsymbol{\omega}_n]_\times}{\|\boldsymbol{\omega}_n\|} \\ &+ (1 - \cos(\|\boldsymbol{\omega}_n\|\Delta t)) \frac{[-\boldsymbol{\omega}_n]_\times^2}{\|\boldsymbol{\omega}_n\|^2}, \end{aligned} \tag{2}$$

and $[\boldsymbol{a}]_\times$ denotes the left cross product matrix of the vector $\boldsymbol{a}$ such that $\boldsymbol{a} \times \boldsymbol{b} = [\boldsymbol{a}]_\times \boldsymbol{b}$, which is given by

$$[\boldsymbol{a}]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}.$$

Note that the dynamic model (1) is norm-preserving, that is, it is ensured that $\|\boldsymbol{r}_{j,n}\| = 1$.

In practice, the exact rotational velocity $\boldsymbol{\omega}_n$ is unknown. Instead, gyroscope measurements $\boldsymbol{y}_{\omega,n}$ are available and these measurements can be used in the dynamic model (1). However, the measurements are noisy and subject to a gyroscope bias, that is, the gyroscope measurement model is given by

$$\boldsymbol{y}_{\omega,n} = \boldsymbol{\omega}_n + \boldsymbol{b}_\omega + \boldsymbol{e}_{\omega,n}, \tag{3}$$

where $\boldsymbol{b}_\omega$ and $\boldsymbol{e}_{\omega,n}$ are the gyroscope bias and measurement noise, respectively.

Hence, these two factors have to be taken into account. Assuming a known bias and zero-mean noise, the dynamic model can be written as

$$\boldsymbol{r}_{j,n} \approx \boldsymbol{F}(\boldsymbol{y}_{\omega,n} - \boldsymbol{b}_\omega)\,\boldsymbol{r}_{j,n-1} \tag{4}$$

where the approximation is due to the gyroscope measurement noise. Naturally, the latter needs to be taken into account as well. Traditionally, this has been done by adding a process noise term to the dynamic model (4). However, if not done carefully, this violates the constraint of the dynamic model being norm-preserving.

Instead, here we propose to use a probabilistic dynamic model based on the von Mises–Fisher probability density function (pdf). In particular, we model the dynamics as an uncertain rotation on the sphere with mean direction given by the transition equation (4), which yields

$$p(\boldsymbol{r}_{j,n} \mid \boldsymbol{r}_{j,n-1}) = \mathcal{VMF}(\boldsymbol{r}_{j,n}; \boldsymbol{F}(\boldsymbol{y}_{\omega,n} - \boldsymbol{b}_\omega)\,\boldsymbol{r}_{j,n-1}, \kappa_{j,n})$$

where $p(\boldsymbol{r}_{j,n} \mid \boldsymbol{r}_{j,n-1})$ is the pdf of the dynamic model and $\mathcal{VMF}(\boldsymbol{r}; \boldsymbol{\mu}, \eta)$ denotes the von Mises–Fisher pdf of the random variable $\boldsymbol{r} \in \mathbb{R}^p$ such that $\|\boldsymbol{r}\| = 1$ with mean direction $\boldsymbol{\mu}$, concentration η, and normalization constant $C_p(\eta)$ given by (see Appendix A and [24] for more information about the von Mises–Fisher pdf)

$$\mathcal{VMF}(\boldsymbol{x}; \boldsymbol{\mu}, \eta) = C_p(\eta)^{-1} \exp(\eta \boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{x}).$$

Next, defining the state as

$$\boldsymbol{x}_n = \begin{bmatrix} \boldsymbol{r}_{a,n} \\ \boldsymbol{r}_{m,n} \end{bmatrix}$$

and noting that the dynamic models for both vectors $\boldsymbol{r}_{a,n}$ and $\boldsymbol{r}_{m,n}$ are conditionally independent, the dynamic model becomes

$$\begin{aligned} p(\boldsymbol{x}_n \mid \boldsymbol{x}_{n-1}) &= \prod_j p(\boldsymbol{r}_{j,n} \mid \boldsymbol{r}_{j,n-1}) \\ &= \prod_j \mathcal{VMF}(\boldsymbol{r}_{j,n}; \boldsymbol{F}(\boldsymbol{y}_{\omega,n} - \boldsymbol{b}_\omega)\,\boldsymbol{r}_{j,n-1}, \kappa_{j,n}), \end{aligned} \tag{5}$$

for $j \in \{a, m\}$.

Finally, we also assume that the initial state is distributed according to a von Mises–Fisher distribution with pdf

$$p(\boldsymbol{x}_0) = \prod_j \mathcal{VMF}(\boldsymbol{r}_{j,0}; \boldsymbol{\mu}_{j,0}, \eta_{j,0}). \tag{6}$$

B. Accelerometer and Magnetometer Measurement Models

Assuming negligible translational motion, the measurements of accelerometers and magnetometers consist of scaled gravity and magnetic field vectors, sensor biases, and measurement noise. Additionally, the sensors may also suffer from axis misalignment (i.e., imperfections in the orthogonality between the sensors' x-, y-, and z-axes) [21]. However, the dominating effects are the sensor biases, and to some extent imperfect scale factors. Hence, in this paper, we focus on these two parameters, which yields the measurement models

$$\boldsymbol{y}_{a,n} = g\,\boldsymbol{r}_{a,n} + \boldsymbol{b}_a + \boldsymbol{e}_{a,n}, \tag{7a}$$
$$\boldsymbol{y}_{m,n} = B\,\boldsymbol{r}_{m,n} + \boldsymbol{b}_m + \boldsymbol{e}_{m,n}, \tag{7b}$$

where g and B are the magnitudes of gravity and the magnetic field, respectively, $\boldsymbol{b}_j$ (for $j \in \{a, m\}$) are the sensor biases, and $\boldsymbol{e}_{j,n}$ are the measurement noises, which are assumed to be zero-mean, independent, and identically distributed Gaussian noises with $\boldsymbol{e}_{j,n} \sim \mathcal{N}(0, \sigma_j^2 \boldsymbol{I}_3)$. Letting

$$\boldsymbol{y}_n = \begin{bmatrix} \boldsymbol{y}_{a,n} \\ \boldsymbol{y}_{m,n} \end{bmatrix},$$

the models in (7) thus give rise to the likelihood

$$\begin{aligned} p(\boldsymbol{y}_n \mid \boldsymbol{x}_n) &= \prod_j p(\boldsymbol{y}_{j,n} \mid \boldsymbol{r}_{j,n}) \\ &= \prod_j \mathcal{N}(\boldsymbol{y}_{j,n}; \alpha_j\,\boldsymbol{r}_{j,n} + \boldsymbol{b}_j, \sigma_j^2 \boldsymbol{I}_3), \end{aligned} \tag{8}$$

where $\alpha_a \triangleq g$ and $\alpha_m \triangleq B$.

Furthermore, note that the measurement models (7) can also be written in the form

$$\begin{aligned} \boldsymbol{y}_{j,n} &= \boldsymbol{G}(\boldsymbol{r}_{j,n})\,\boldsymbol{\lambda}_j + \boldsymbol{e}_{j,n} \\ &= \boldsymbol{G}_{j,n}\boldsymbol{\lambda}_j + \boldsymbol{e}_{j,n}, \end{aligned}$$

where the matrix $\boldsymbol{G}_{j,n}$ is

$$\boldsymbol{G}_{j,n} \triangleq \boldsymbol{G}(\boldsymbol{r}_{j,n}) = \begin{bmatrix} \boldsymbol{r}_{j,n} & \boldsymbol{I}_3 \end{bmatrix}, \tag{9}$$

$\boldsymbol{\lambda}_a = \begin{bmatrix} g & \boldsymbol{b}_a^\mathsf{T} \end{bmatrix}^\mathsf{T}$, and $\boldsymbol{\lambda}_m = \begin{bmatrix} B & \boldsymbol{b}_m^\mathsf{T} \end{bmatrix}^\mathsf{T}$. This, in turn, also yields the alternative form of the likelihood

$$p(\boldsymbol{y}_n \mid \boldsymbol{x}_n) = \prod_j \mathcal{N}(\boldsymbol{y}_{j,n}; \boldsymbol{G}_{j,n}\boldsymbol{\lambda}_j, \sigma_j^2 \boldsymbol{I}_3). \tag{10}$$

C. Problem Formulation

Given the dynamic model (5), the measurement model (8) (together with the alternative formulation (10)), and the calibration data $\boldsymbol{y}_{1:N} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N\}$, the objective is to estimate the IMU and magnetometer model parameters. The latter are the scale factors g and B, as well as the accelerometer and magnetometer biases $\boldsymbol{b}_j$, the measurement noise variances σ_j^2, and additionally, the gyroscope bias $\boldsymbol{b}_\omega$. This yields the full parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix} g & \boldsymbol{b}_a^\mathsf{T} & \sigma_a^2 & B & \boldsymbol{b}_m^\mathsf{T} & \sigma_m^2 & \boldsymbol{b}_\omega^\mathsf{T} \end{bmatrix}^\mathsf{T}.$$

In order to estimate the unknown parameters $\boldsymbol{\theta}$, we consider a solution based on the EM algorithm, which approximates the maximum likelihood estimate of $\boldsymbol{\theta}$ by iteratively maximizing the expected log-likelihood. Specifically, in the kth iteration, we maximize the expected log-likelihood given the $k-1$th iteration's parameter estimate $\hat{\boldsymbol{\theta}}^{(k-1)}$ to find an improved parameter estimate $\hat{\boldsymbol{\theta}}^{(k)}$ by maximizing

$$\hat{\boldsymbol{\theta}}^{(k)} = \operatorname*{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$$

with

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) \triangleq \mathrm{E}\{\log p(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N}; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:N}\}.$$

Here, $p(\cdot; \boldsymbol{\theta})$ indicates the parametrization of the corresponding pdf by $\boldsymbol{\theta}$ and $\mathrm{E}\{\cdot\}$ denotes the expectation. The latter is with respect to the smoothing density $p(\boldsymbol{x}_{0:N} \mid \boldsymbol{y}_{1:N})$, which was calculated using $\hat{\boldsymbol{\theta}}^{(k-1)}$.

It can be shown that $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$ can be written as [26]

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) = Q_1(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) + Q_2(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) + Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}),$$

with

$$Q_1(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) = \mathrm{E}\{\log p(\boldsymbol{x}_0; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:N}\}, \tag{11a}$$
$$Q_2(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) = \sum_{n=1}^{N} \mathrm{E}\{\log p(\boldsymbol{x}_n \mid \boldsymbol{x}_{n-1}; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:N}\}, \tag{11b}$$
$$Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) = \sum_{n=1}^{N} \mathrm{E}\{\log p(\boldsymbol{y}_n \mid \boldsymbol{x}_n; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:N}\}. \tag{11c}$$

These expressions make use of the initial state (6), the dynamic model (5), and the likelihood (8) (or (10)).

Hence, the objective is to calculate the smoothing distribution $p(\boldsymbol{x}_{1:N} \mid \boldsymbol{y}_{1:N})$ given the $k-1$th iterations parameters (expectation step) and then maximize the expectations (11) with respect to the parameters $\boldsymbol{\theta}$ (maximization step). The implementation of these two steps is discussed in the following sections.

III. von Mises–Fisher Filtering

Since the state is composed of the gravity and magnetic field reference vectors that are constrained on the unit sphere, a filtering algorithm that takes these constraints into account should be used. One approach is to use constrained Kalman filtering algorithms that use normalization of the state to ensure the constraint is met and project the reference vector onto the unit sphere [23]. An alternative is to instead consider inference on the unit sphere itself, which requires a suitable filtering algorithm [27], [28].

In this paper, we chose this latter option, motivated by the fact that it systematically accounts for the geometry of the problem. In particular, we employ a discrete-time von Mises–Fisher assumed density filtering algorithm that approximates the filtering posterior of the gravity and magnetic field vectors according to

$$\begin{aligned} p(\boldsymbol{x}_n \mid \boldsymbol{y}_{1:n}) &= \prod_j p(\boldsymbol{r}_{j,n} \mid \boldsymbol{y}_{1:n}) \\ &\approx \prod_j \mathcal{VMF}(\boldsymbol{r}_{j,n}; \boldsymbol{\mu}_{j,n|n}, \eta_{j,n|n}). \end{aligned} \tag{12}$$

The remainder of this section introduces the prediction and measurement update steps used to obtain the posterior (12).

A. Prediction

Assume that we are given the assumed filtering density of the form (12) with parameters $\boldsymbol{\mu}_{j,n-1|n-1}$ and $\eta_{j,n-1|n-1}$ at time step $n-1$. Then, the predictive density of $\boldsymbol{x}_n$ given the data $\boldsymbol{y}_{1:n-1}$ is given by

$$\begin{aligned} p(\boldsymbol{x}_n \mid \boldsymbol{y}_{1:n-1}) &= \int p(\boldsymbol{x}_n \mid \boldsymbol{x}_{n-1}) p(\boldsymbol{x}_{n-1} \mid \boldsymbol{y}_{1:n-1}) \mathrm{d}\boldsymbol{x}_{n-1} \\ &= \prod_j \int \mathcal{VMF}(\boldsymbol{r}_{j,n}; \boldsymbol{F}(\boldsymbol{y}_{\omega,n} - \boldsymbol{b}_\omega)\, \boldsymbol{r}_{j,n-1}, \kappa_{j,n}) \\ &\quad \mathcal{VMF}(\boldsymbol{r}_{j,n-1}; \boldsymbol{\mu}_{j,n-1|n-1}, \eta_{j,n-1|n-1}) \mathrm{d}\boldsymbol{r}_{j,n-1}, \end{aligned} \tag{13}$$

where we have made used of the dynamic model (5).

Unfortunately, the von Mises–Fisher pdf is not closed under marginalization. Instead, we have to approximate the predictive density as

$$p(\boldsymbol{x}_n \mid \boldsymbol{y}_{n-1}) \approx \prod_j \mathcal{VMF}(\boldsymbol{r}_{j,n}; \boldsymbol{\mu}_{j,n|n-1}, \eta_{j,n|n-1}) \tag{14}$$

and use maximum likelihood estimation of the VMF parameters to calculate the predicted mean direction $\boldsymbol{\mu}_{j,n|n-1}$ and concentration $\eta_{j,n|n-1}$ in (13). These are found through the predictive mean (see Appendix A and [24])

$$\begin{aligned} \boldsymbol{m}_{j,n|n-1} &\triangleq \mathrm{E}\{\mathrm{E}\{\boldsymbol{r}_{j,n} \mid \boldsymbol{r}_{j,n-1}\} \mid \boldsymbol{y}_{1:n-1}\} \\ &= A_3(\kappa_{j,n}) A_3(\eta_{j,n-1|n-1}) \boldsymbol{F}(\boldsymbol{y}_{\omega,n} - \boldsymbol{b}_\omega)\, \boldsymbol{\mu}_{j,n-1|n-1} \end{aligned} \tag{15}$$

and are given by [24]

$$\boldsymbol{\mu}_{j,n|n-1} = \frac{\boldsymbol{m}_{j,n|n-1}}{\|\boldsymbol{m}_{j,n|n-1}\|}, \tag{16a}$$

$$\eta_{j,n|n-1} = A_3^{-1}(\|\boldsymbol{m}_{j,n|n-1}\|). \tag{16b}$$

Here, $A_p^{-1}(x)$ denotes the solution of the nonlinear equation

$$A_p(\eta) = x$$

for η where $A_p(\kappa) = \partial / \partial \eta \log C_p(\kappa)$ (see Appendix A). This nonlinear equation can only be solved numerically, and here, we use the solution method introduced in [29].

B. Measurement Update

Given the von Mises–Fisher prediction (14) together with the likelihood (8), a closed form measurement update is obtained. This is given by [27]

$$\begin{aligned} p(\boldsymbol{x}_n \mid \boldsymbol{y}_{1:n}) &\propto \prod_j p(\boldsymbol{y}_{j,n} \mid \boldsymbol{r}_{j,n}) p(\boldsymbol{r}_{j,n} \mid \boldsymbol{y}_{j,1:n-1}) \\ &\propto \prod_j \mathcal{VMF}(\boldsymbol{r}_{j,n}; \boldsymbol{\mu}_{j,n|n}, \eta_{j,n|n}), \end{aligned} \tag{17}$$

where the parameters of the updated filtering posterior are

$$\eta_{j,n|n} = \left\| \frac{\alpha_j}{\sigma_j^2} (\boldsymbol{y}_{j,n} - \boldsymbol{b}_j) + \eta_{j,n|n-1} \boldsymbol{\mu}_{j,n|n-1} \right\|, \tag{18a}$$

$$\boldsymbol{\mu}_{j,n|n} = \frac{\frac{\alpha_j}{\sigma_j^2} (\boldsymbol{y}_{j,n} - \boldsymbol{b}_j) + \eta_{j,n|n-1} \boldsymbol{\mu}_{j,n|n-1}}{\eta_{j,n|n-1}}. \tag{18b}$$

Algorithm 1 von Mises–Fisher Filter Update Step

Input: Previous filtering parameters $\boldsymbol{\mu}_{j,n-1|n-1}$, $\eta_{j,n-1|n-1}$ and measurements $\boldsymbol{y}_{\omega,n}$, $\boldsymbol{y}_{j,n}$

Output: Updated filtering parameters $\boldsymbol{\mu}_{j,n|n}$, $\eta_{j,n|n}$

1: Calculate the predictive means $\boldsymbol{m}_{j,n|n-1}$ using (15)
2: Calculate the predictive mean direction $\boldsymbol{\mu}_{j,n|n-1}$ and concentrations $\eta_{j,n|n-1}$ using (16)
3: Calculate the filtered mean directions $\boldsymbol{\mu}_{j,n|n}$ and concentrations $\eta_{j,n|n}$ using (18)
4: Calculate the marginal likelihood increments $p(\boldsymbol{y}_n \mid \boldsymbol{y}_{1:n-1})$ using (19)

Furthermore, during the measurement update, the marginal likelihood $p(\boldsymbol{y}_{1:n})$ may also be computed. First, note that the marginal likelihood can be factorized as

$$p(\boldsymbol{y}_{1:n}) = p(\boldsymbol{y}_n \mid \boldsymbol{y}_{1:n-1}) p(\boldsymbol{y}_{1:n-1}),$$

where $p(\boldsymbol{y}_{1:n-1})$ is the marginal likelihood at the previous time step and $p(\boldsymbol{y}_n \mid \boldsymbol{y}_{1:n-1})$ the increment of the current time step. The latter is calculated using (14) and (8) as (see [27])

$$\begin{aligned} p(\boldsymbol{y}_n \mid \boldsymbol{y}_{1:n-1}) &= \int p(\boldsymbol{y}_n \mid \boldsymbol{x}_n) p(\boldsymbol{x}_n \mid \boldsymbol{y}_{1:n-1}) \mathrm{d}\boldsymbol{x}_n \\ &= \prod_j \frac{C_3(\eta_{j,n|n})}{C_3(\eta_{j,n|n-1})(2\pi\sigma_j^2)^{3/2}} \exp\left(-\frac{\|\boldsymbol{y}_{j,n} - \boldsymbol{b}_j\|^2 + \alpha_j^2}{2\sigma_j^2} \right). \end{aligned} \tag{19}$$

This concludes the discrete-time von Mises–Fisher assumed density filtering algorithm and one complete update step is shown in Algorithm 1.

As discussed in Section II, in EM, the parameter update is estimated by maximizing the expected log-likelihood with respect to the smoothing density $p(\boldsymbol{x}_{0:N} \mid \boldsymbol{y}_{1:N})$. In principle, this would require to run a smoother. However, in this application, we propose to approximate the smoothing pdf using the filtering pdf as

$$\begin{aligned} p(\boldsymbol{x}_n \mid \boldsymbol{y}_{1:N}) &\approx p(\boldsymbol{x}_n \mid \boldsymbol{y}_{1:n}) \\ &= \prod_j \mathcal{VMF}(\boldsymbol{r}_{j,n}; \boldsymbol{\mu}_{j,n|n}, \eta_{j,n|n}). \end{aligned} \tag{20}$$

This is based on the fact that accelerometers and magnetometers measure the reference vectors with large scale factors, which leads to very high concentration parameters $\eta_{j,n|n}$ during filtering, see (18). This implies that the von Mises–Fisher filtering posterior is highly concentrated and the smoothing density is practically identical to the filtering density and the latter can safely be used to approximate $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$.

IV. Parameter Estimation

As discussed in the problem formulation, the parameters to estimate are the scale factors g and B as well as the sensor biases $\boldsymbol{b}_a$, $\boldsymbol{b}_m$, and $\boldsymbol{b}_\omega$. Additionally, we also estimate the noise magnitudes σ_a^2 and σ_m^2. In this section, we derive the necessary estimators that maximize the expected log-likelihood $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$. However, first note that due to the

approximation (20) of the smoothing density, the approximation of the expected log-likelihood becomes

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) \approx \mathrm{E}\{\log p(\boldsymbol{y}_{1:N}, \boldsymbol{x}_{0:N}; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:n}\},$$

which will be used throughout the remainder of this section.

A. Gyroscope Bias

The gyroscope parameters only consist of the bias $\boldsymbol{b}_\omega$, which enters the dynamic models for both the gravity and magnetic field reference vectors, see (5). Hence, in order to maximize $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$ with respect to $\boldsymbol{b}_\omega$, it is enough to maximize $Q_2(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$.

This term is given by

$$\begin{aligned} Q_2(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) &\approx \sum_{n=1}^{N} \mathrm{E}\{\log p(\boldsymbol{x}_n \mid \boldsymbol{x}_{n-1}; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:n}\} \\ &= \sum_{n=1}^{N} \sum_{j} -\log C_3(\kappa_{j,n}) \\ &\quad + \mathrm{E}\{\kappa_{j,n} (\boldsymbol{F}(\boldsymbol{y}_{\omega,n} - \boldsymbol{b}_\omega)\, \boldsymbol{r}_{j,n-1})^\mathsf{T} \boldsymbol{r}_{j,n} \mid \boldsymbol{y}_{1:n}\}, \end{aligned}$$

which would require the joint density $p(\boldsymbol{x}_n, \boldsymbol{x}_{n-1} \mid \boldsymbol{y}_{1:n})$. This is the joint density of the one-lag smoothing density, and due to the high concentration of the von Mises–Fisher filtering posterior discussed in Section III, we approximate this as

$$p(\boldsymbol{x}_n, \boldsymbol{x}_{n-1} \mid \boldsymbol{y}_{1:n}) \approx p(\boldsymbol{x}_n \mid \boldsymbol{y}_{1:n}) p(\boldsymbol{x}_{n-1} \mid \boldsymbol{y}_{1:n-1}).$$

This approximation then leads to

$$\begin{aligned} Q_2(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) &\approx \sum_{n=1}^{N} \sum_{j} -\log C_3(\kappa_{j,n}) \\ &\quad + \kappa_{j,n} (\boldsymbol{F}(\boldsymbol{y}_{\omega,n} - \boldsymbol{b}_\omega)\, \boldsymbol{m}_{j,n-1|n-1})^\mathsf{T} \boldsymbol{m}_{j,n|n}. \end{aligned} \tag{21}$$

It can be seen from the definition of the matrix $\boldsymbol{F}(\cdot)$ in (2) and (21) that the bias term $\boldsymbol{b}_\omega$ enters the expected log-likelihood in a highly nonlinear way. Hence, a closed-form maximization step for this parameter cannot be found and numerical optimization methods have to be used instead. Here, we use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton method [30].

B. Accelerometer and Magnetometer Parameters

From the likelihood (10) it can be seen that the parameters of the accelerometer and magnetometer (scale α_j, bias $\boldsymbol{b}_j$, and noise variance σ_j^2) enter the model in the same way. Furthermore, since the parameters only affect the likelihood, maximizing $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$ is equivalent to maximizing

$$\begin{aligned} Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) &\approx \sum_{n=1}^{N} \mathrm{E}\{\log p(\boldsymbol{y}_n \mid \boldsymbol{x}_n; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:n}\} \\ &= \sum_{n=1}^{N} \sum_{j} -\frac{3}{2} \log 2\pi\sigma_j^2 - \frac{1}{2\sigma_j^2} \mathrm{E}\{\|\boldsymbol{y}_{j,n} - \boldsymbol{G}_{j,n} \boldsymbol{\lambda}_j\|^2 \mid \boldsymbol{y}_{1:n}\}, \end{aligned} \tag{22}$$

only, see (10)–(11).

Algorithm 2 Joint IMU and Magnetometer Calibration

Input: Measurements $\boldsymbol{y}_{\omega,1:N}$, $\boldsymbol{y}_{a,1:N}$, $\boldsymbol{y}_{m,1:N}$, and initial parameter guess $\hat{\boldsymbol{\theta}}^{(0)}$
Output: Calibration parameter estimate $\hat{\boldsymbol{\theta}}$

```
while Not converged do
    for n = 1, 2, ..., N do
        Calculate μ_{j,n|n} and η_{j,n|n} using Algorithm 1
        Calculate z_{j,n} and H_{j,n} using (25)
        Calculate
            z_{j,1:n} = z_{j,1:n-1} + z_{j,n}
            H_{j,1:n} = H_{j,1:n-1} + H_{j,n}
    end for
    Estimate λ̂_j^{(k)} and σ̂_j^{2,(k)} using (23)
    Estimate b̂_ω^{(k)} by numerically maximizing (21)
end while
Set θ̂ ← θ̂^{(k)}
```

Taking the gradient of (22) with respect to the parameters $\boldsymbol{\lambda}_j = \begin{bmatrix} \alpha_j & \boldsymbol{b}_j^\mathsf{T} \end{bmatrix}^\mathsf{T}$ and σ_j^2, and then solving the resulting equation system yields the closed form parameter updates (see Appendix B)

$$\hat{\boldsymbol{\lambda}}_j^{(k)} = \boldsymbol{H}_{j,1:N}^{-1} \boldsymbol{z}_{j,1:N}, \tag{23a}$$

$$\hat{\sigma}_j^{2,(k)} = \frac{1}{3N} \left(\sum_{n=1}^{N} \boldsymbol{y}_{j,n}^\mathsf{T} \boldsymbol{y}_{j,n} - \boldsymbol{z}_{j,1:N}^\mathsf{T} \boldsymbol{H}_{j,1:N}^{-1} \boldsymbol{z}_{j,1:N} \right), \tag{23b}$$

where

$$\boldsymbol{H}_{j,1:N} = \sum_{n=1}^{N} \boldsymbol{H}_{j,n}, \tag{24a}$$

$$\boldsymbol{z}_{j,1:N} = \sum_{n=1}^{N} \boldsymbol{z}_{j,n}, \tag{24b}$$

and

$$\boldsymbol{z}_{j,n} = \begin{bmatrix} A_3(\eta_{j,n|n}) \boldsymbol{\mu}_{j,n|n}^\mathsf{T} \\ \boldsymbol{I}_3 \end{bmatrix} \boldsymbol{y}_{j,n}, \tag{25a}$$

$$\boldsymbol{H}_{j,n} = \begin{bmatrix} 1 & A_3(\eta_{j,n|n}) \boldsymbol{\mu}_{j,n|n}^\mathsf{T} \\ A_3(\eta_{j,n|n}) \boldsymbol{\mu}_{j,n|n} & \boldsymbol{I}_3 \end{bmatrix}. \tag{25b}$$

Note that the quantities (24) can be calculated during the filtering step (expectation step) and thus, the parameter update (23) can be calculated efficiently once the filtering step is completed. This yields the full calibration algorithm summarized in Algorithm 2.

V. Results

In this section, the proposed method is evaluated on real data for calibration of smartphone IMUs.

A. Experimental Setup

The experimental setup used to evaluate the proposed method is as follows. We use a Sony Xperia XZ1 smartphone, which integrates a single-chip IMU (combined accelerometer

and gyroscope) and a magnetometer. We collect data for three different motion sequences to assess the impact on the calibration. First, the phone is attached to a block, which is rotated around all three axes by 90° sequentially, placing the block on all of its faces for $5\,\mathrm{s}$ each. Second, starting from a level position, the phone is tilted $\pm 90^\circ$ around each of its axes. Third, the phone is rotated freely around all of its axes. The three datasets are termed *block*, *tilt*, and *free*, respectively. Each experiment is performed twice, the raw data is recorded at $100\,\mathrm{Hz}$ with the internal compensation algorithms disabled, and the data is processed offline.

The initial guesses for the biases and scale factors for the accelerometer and magnetometer are obtained using sphere-fitting [21] and the initial gyroscope bias is set to zero. The initial guess for the sensor noise variance σ_j^2 is set based on the sample variance of the first ten samples of $\boldsymbol{y}_{j,n}$ and the parameters of the initial state's pdf are set as $\boldsymbol{\mu}_{j,0} = \boldsymbol{y}_{j,1}/\|\boldsymbol{y}_{j,1}\|$ and $\eta_{j,0} = 1 \times 10^{-3}$. The concentration parameter $\kappa_{j,n}$ of the dynamic model is set to $\kappa_{a,n} = \kappa_{m,n} = 100$. The EM algorithm is terminated whenever the relative change in parameter estimate drops below the threshold $\epsilon = 1 \times 10^{-6}$ or after a maximum of $K = 100$ iterations.

B. Results and Discussion

Table I shows the parameters estimated using the three different calibration data sets. It can be seen that the estimated biases and scale factors are of the same magnitude for all three datasets. The scale factors are both close to their nominal values of $g = 9.819\,\mathrm{m/s^2}$ and $B = 52.17\,\mu\mathrm{T}$ for Espoo, Finland [31]. The biases for the accelerometer and gyroscope are close to zero, whereas the bias for the magnetometer is significant as expected. These results are in line with the identifiability analysis in [21], where it was concluded that successful calibration does not require full rotations around all axes. In particular, the *tilt* calibration sequence is significantly easier to perform and reproduce reliably compared to full rotations but still yields good parameter estimates. The random rotation sequence, however, may suffer from excessive translational motion, which decreases the quality of the calibration. This is also reflected in the estimated sensor noise covariances σ_a^2 and σ_m^2, which are significantly higher for this calibration sequence, see also [32].

Furthermore, Table II shows the dataset sizes (number of samples N) together with the calibration time (for a non-optimized implementation in Matlab running on an Intel i5-8350U $1.7\,\mathrm{GHz}$ CPU with $16\,\mathrm{GB}$ RAM) and the number of iterations required for convergence. Naturally, the calibration time depends on the number of samples and hence varies greatly between the *block* and other datasets. However, despite being much smaller in length, the *tilt* dataset estimates the parameters equally well but in considerably less time. The number of EM iterations on the other hand is similar for all three datasets. Finally, note that the main computational bottleneck is the numerical optimization for updating the gyroscope bias $\boldsymbol{b}_\omega$, which requires extra iterations compared to the other parameters.

TABLE I
ESTIMATED PARAMETERS FOR THE THREE CALIBRATION DATASETS.

Parameter	Block	Tilt	Free
g / $\mathrm{m/s^2}$	9.88	9.97	10.1
$\boldsymbol{b}_a$ / $\mathrm{m/s^2}$	$\begin{bmatrix}0.91\\0.20\\0.12\end{bmatrix}$	$\begin{bmatrix}0.02\\0.09\\-0.01\end{bmatrix}$	$\begin{bmatrix}0.1\\0.08\\-0.65\end{bmatrix}$
σ_a^2 / $(\mathrm{m/s^2})^2$	0.33	0.27	2.17
B / μT	50.16	49.5	49.7
$\boldsymbol{b}_m$ / μT	$\begin{bmatrix}-43.4\\-0.98\\30.5\end{bmatrix}$	$\begin{bmatrix}-43.2\\-1.1\\30\end{bmatrix}$	$\begin{bmatrix}-43.5\\-1.49\\30.3\end{bmatrix}$
σ_m^2 / $(\mu\mathrm{T})^2$	0.43	0.47	1.11
$\boldsymbol{b}_\omega$ / $\mathrm{s^{-1}}$	$\begin{bmatrix}-0.01\\0\\-0.02\end{bmatrix}$	$\begin{bmatrix}0\\0\\-0.03\end{bmatrix}$	$\begin{bmatrix}0\\0.02\\-0.05\end{bmatrix}$

TABLE II
COMPARISON OF THE TRAINING DATASET LENGTHS, COMPUTATIONAL TIME, AND NUMBER OF EM ITERATIONS.

Parameter	Block	Tilt	Free
N	6776	1405	1614
Time	61.9 s	12.6 s	13.5 s
Iterations	32	30	27

Finally, Fig. 1 shows the cumulative distribution function of the root mean squared error (RMSE)

$$e_{\text{RMSE,n}} = \sqrt{\frac{1}{6}\sum_j \|\boldsymbol{y}_{j,n} - \hat{\boldsymbol{y}}_{j,n|n-1}\|^2}$$

of the uncalibrated and calibrated models (calibrated using the *block* calibration dataset) for all three validation datasets, where $\hat{\boldsymbol{y}}_{j,n|n-1}$ is the one step ahead prediction given by

$$\begin{aligned}\hat{\boldsymbol{y}}_{j,n|n-1} &\triangleq \mathrm{E}\{\boldsymbol{y}_{j,n} \mid \boldsymbol{y}_{j,1:n-1}\} \\ &= \alpha_j A_p(\eta_{j,n|n-1})\,\boldsymbol{\mu}_{j,n|n-1}.\end{aligned}$$

As expected, the calibration significantly improves the predictive power, lowers the RMSE, and generalizes well from the calibration to the validation datasets. Also, note that the RMSE is lower and of the same magnitude for the *block* (Fig. 1, left) and *tilt* (Fig. 1, middle) validation datasets compared to the *free* (Fig. 1, right) dataset. This can again be attributed to the increased translational sensor motion.

VI. CONCLUSIONS

In this paper, we presented a von Mises–Fisher-filtering- and EM-based method for joint calibration of inertial sensors and magnetometers. The proposed method exploits the model structure by considering the problem as a scaled and biased reference vector tracking problem on the unit sphere. The resulting algorithm allows for efficient estimation of the unknown parameters. The main drawback is that in its present form, the proposed calibration algorithm can not yet make use of the most complete sensor models (e.g., axis misalignment or rotations between sensor coordinate frames are neglected).

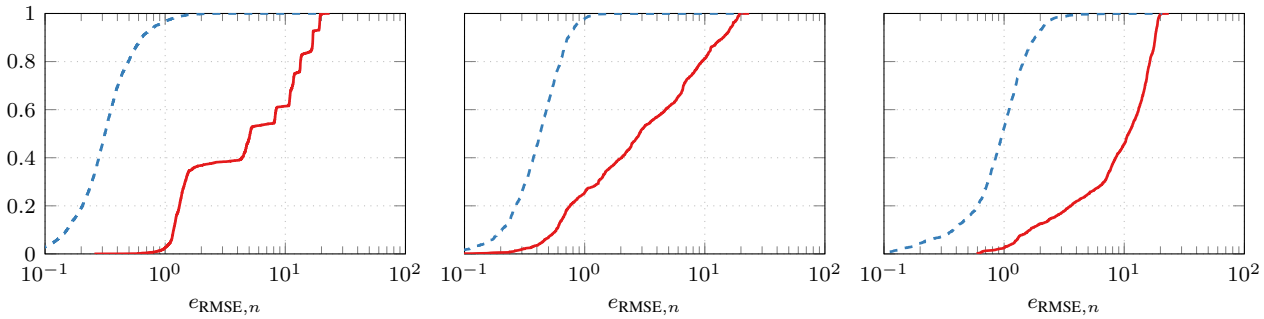

Fig. 1. Cumulative distribution functions of the one-step ahead predictions' RMSEs for the uncalibrated (——) and calibrated (- - -) sensor models for the *block* (left), *tilt* (middle), and *free* validation datasets.

Thus, future work will focus on improving this shortcoming to relax some of the model assumptions.

APPENDIX

A. Von Mises–Fisher pdf

The von Mises–Fisher pdf with mean direction $\boldsymbol{\mu} \in \mathbb{R}^p : \|\boldsymbol{\mu}\| = 1$ and concentration parameter η for a random variable $\boldsymbol{r} \in \mathbb{R}^p : \|\boldsymbol{r}\| = 1$ is defined as [24]

$$\mathcal{VMF}(\boldsymbol{r}; \boldsymbol{\mu}, \eta) = C_p(\eta)^{-1} \exp(\eta \boldsymbol{\mu}^\mathsf{T} \boldsymbol{r}), \tag{26}$$

where

$$C_p(\eta)^{-1} = \frac{\eta^{p/2-1}}{(2\pi)^{p/2} \mathcal{I}_{p/2-1}(\eta)} \tag{27}$$

is the normalization constant with $\mathcal{I}_k(x)$ the modified Bessel function of the first kind of order k.

The mean of a von Mises–Fisher distributed random variable is given by [24]

$$\mathrm{E}\{\boldsymbol{r}\} = A_p(\eta)\boldsymbol{\mu},$$

where

$$A_p(\eta) = \frac{\partial}{\partial \eta} \log C_p(\eta)$$

is the derivative of the logarithm of the normalization constant with respect to the concentration parameter η. Furthermore, given the mean $\mathrm{E}\{\boldsymbol{x}\} = \boldsymbol{m}$ of a random variable $\boldsymbol{x}$, the maximum likelihood estimates of the mean direction and concentration parameter are [24]

$$\begin{aligned} \boldsymbol{\mu} &= \frac{\boldsymbol{m}}{\|\boldsymbol{m}\|}, \\ \eta &= A_p^{-1}(\|\boldsymbol{m}\|), \end{aligned}$$

where $A_p^{-1}(\|\boldsymbol{m}\|)$ denotes the solution of the nonlinear equation

$$\|\boldsymbol{m}\| = A_p(\eta) \tag{29}$$

with respect to η.

B. Derivation of the Accelerometer and Magnetometer Parameter Updates

In order to maximize $Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$ with respect to the parameters $\boldsymbol{\lambda}_j$ and σ_j^2, we find the zeros of the gradient of $Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$ and solve for the parameters. From (22) it follows that the gradient with respect to $\boldsymbol{\lambda}_j$ is given by

$$\begin{aligned} \nabla_{\boldsymbol{\lambda}_j} Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) &\approx \nabla_{\boldsymbol{\lambda}_j} \sum_{n=1}^{N} \mathrm{E}\{\log p(\boldsymbol{y}_n \mid \boldsymbol{x}_n; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:n}\} \\ &= \sum_{n=1}^{N} \mathrm{E}\{\nabla_{\boldsymbol{\lambda}_j} \log p(\boldsymbol{y}_{j,n} \mid \boldsymbol{r}_{j,n}; \boldsymbol{\theta}) \mid \boldsymbol{y}_{1:n}\}. \end{aligned}$$

Using (10) for the log-likelihood, the gradient of $\log p(\boldsymbol{y}_{j,n} \mid \boldsymbol{r}_{j,n}; \boldsymbol{\theta})$ with respect to $\boldsymbol{\lambda}_j$ becomes

$$\nabla_{\boldsymbol{\lambda}_j} \log p(\boldsymbol{y}_{j,n} \mid \boldsymbol{r}_{j,n}; \boldsymbol{\theta}) = -\frac{1}{\sigma_j^2} \boldsymbol{G}_{j,n}^\mathsf{T} \boldsymbol{y}_{j,n} + \frac{1}{\sigma_j^2} \boldsymbol{G}_{j,n}^\mathsf{T} \boldsymbol{G}_{j,n} \boldsymbol{\lambda}_j$$

and from (9) it follows that

$$\begin{aligned} \boldsymbol{G}_{j,n} &= \begin{bmatrix} \boldsymbol{r}_{j,n} & \boldsymbol{I}_3 \end{bmatrix}, \\ \boldsymbol{G}_{j,n}^\mathsf{T} \boldsymbol{G}_{j,n} &= \begin{bmatrix} \boldsymbol{r}_{j,n}^\mathsf{T} \boldsymbol{r}_{j,n} & \boldsymbol{r}_{j,n}^\mathsf{T} \\ \boldsymbol{r}_{j,n} & \boldsymbol{I}_3 \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{r}_{j,n}^\mathsf{T} \\ \boldsymbol{r}_{j,n} & \boldsymbol{I}_3 \end{bmatrix}. \end{aligned}$$

Furthermore, using (17) and the properties of the von Mises–Fisher pdf (see Appendix A) yields

$$\mathrm{E}\{\boldsymbol{r}_{j,n} \mid \boldsymbol{y}_{1:n}\} = A_3(\eta_{j,n|n})\, \boldsymbol{\mu}_{j,n|n}.$$

Thus, defining $\boldsymbol{z}_{j,n}$ and $\boldsymbol{H}_{j,n}$ as in (25), the gradient of $Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$ with respect to $\boldsymbol{\lambda}_j$ becomes

$$\begin{aligned} \nabla_{\boldsymbol{\lambda}_j} Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) &= \sum_{n=1}^{N} \left(\frac{1}{\sigma_j^2} \boldsymbol{H}_{j,n} \boldsymbol{\lambda}_j - \frac{1}{\sigma_j^2} \boldsymbol{z}_{j,n} \right) \\ &= \frac{1}{\sigma_j^2} \boldsymbol{H}_{j,1:N} \boldsymbol{\lambda}_j + \frac{1}{\sigma_j^2} \boldsymbol{z}_{j,1:N} \end{aligned}$$

with $\boldsymbol{H}_{j,1:N}$ and $\boldsymbol{z}_{j,1:N}$ as in (24). Finally, setting the gradient to zero and solving for $\boldsymbol{\lambda}_j$ yields (23a).

To find the expression (23b) for the measurement noise variance σ_j^2, the derivative of the log-likelihood with respect to σ_j^2 is required. This is given by

$$\nabla_{\sigma_j^2} \log(\boldsymbol{y}_{j,n} \mid \boldsymbol{r}_{j,n}; \boldsymbol{\theta}) = -\frac{3}{2\sigma_j^2} + \frac{1}{2(\sigma_j^2)^2} \|\boldsymbol{y}_{j,n} - \boldsymbol{G}_{j,n} \boldsymbol{\lambda}_j\|^2$$

and it follows that

$$\nabla_{\sigma_j^2} Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) = \sum_{n=1}^{N} \mathrm{E}\{\nabla_{\sigma_j^2} \log p(\boldsymbol{y}_{j,n} \mid \boldsymbol{r}_{j,n}) \mid \boldsymbol{y}_{1:n}\}$$
$$= \sum_{n=1}^{N} -\frac{3}{2\sigma_j^2} + \frac{1}{2(\sigma_j^2)^2} \mathrm{E}\{\|\boldsymbol{y}_{j,n} - \boldsymbol{G}_{j,n}\boldsymbol{\lambda}_j\|^2 \mid \boldsymbol{y}_{1:n}\}.$$

The remaining expectation is given by

$$\begin{aligned}
&\mathrm{E}\{\|\boldsymbol{y}_{j,n} - \boldsymbol{G}_{j,n}\boldsymbol{\lambda}_j\|^2 \mid \boldsymbol{y}_{1:n}\} \\
&\quad = \mathrm{E}\{\boldsymbol{y}_{j,n}^\mathsf{T}\boldsymbol{y}_{j,n} - 2\boldsymbol{\lambda}_j^\mathsf{T}\boldsymbol{G}_{j,n}^\mathsf{T}\boldsymbol{y}_{j,n} + \boldsymbol{\lambda}_j^\mathsf{T}\boldsymbol{G}_{j,n}^\mathsf{T}\boldsymbol{G}_{j,n}\boldsymbol{\lambda}_j \mid \boldsymbol{y}_{1:n}\} \\
&\quad = \boldsymbol{y}_{j,n}^\mathsf{T}\boldsymbol{y}_{j,n} - 2\boldsymbol{\lambda}_j^\mathsf{T}\boldsymbol{z}_{j,n} + \boldsymbol{\lambda}_j^\mathsf{T}\boldsymbol{H}_{j,n}\boldsymbol{\lambda}_j,
\end{aligned}$$

where $\boldsymbol{z}_{j,n}$ and $\boldsymbol{H}_{j,n}$ are again as defined in (25). Hence, the gradient with respect to σ_j^2 becomes

$$\begin{aligned}
&\nabla_{\sigma_j^2} Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) \\
&\quad = -\frac{3N}{2\sigma_j^2} + \frac{1}{2(\sigma_j^2)^2} \sum_{n=1}^{N} \boldsymbol{y}_{j,n}^\mathsf{T}\boldsymbol{y}_{j,n} - 2\boldsymbol{\lambda}_j^\mathsf{T}\boldsymbol{z}_{j,n} + \boldsymbol{\lambda}_j^\mathsf{T}\boldsymbol{H}_{j,n}\boldsymbol{\lambda}_j \\
&\quad = -\frac{3N}{2\sigma_j^2} + \frac{1}{2(\sigma_j^2)^2} \left(\sum_{n=1}^{N} \boldsymbol{y}_{j,n}^\mathsf{T}\boldsymbol{y}_{j,n} - 2\boldsymbol{\lambda}_j^\mathsf{T}\boldsymbol{z}_{j,1:N} \right. \\
&\qquad \left. + \boldsymbol{\lambda}_j^\mathsf{T}\boldsymbol{H}_{j,1:N}\boldsymbol{\lambda}_j \right).
\end{aligned}$$

Substitution of the solution (23a) for $\boldsymbol{\lambda}_j$ at the stationary point yields

$$\begin{aligned}
&\nabla_{\sigma_j^2} Q_3(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) \\
&\quad = -\frac{3N}{2\sigma_j^2} + \frac{1}{2(\sigma_j^2)^2} \sum_{n=1}^{N} \boldsymbol{y}_{j,n}^\mathsf{T}\boldsymbol{y}_{j,n} - \boldsymbol{z}_{j,1:N}^\mathsf{T}\boldsymbol{H}_{j,1:N}^{-1}\boldsymbol{z}_{j,1:N},
\end{aligned}$$

and finally, solving for solving for σ_j^2 yields the update (23b).

REFERENCES

[1] Z. Xiao, H. Wen, A. Markham, and N. Trigoni, "Robust pedestrian dead reckoning (R-PDR) for arbitrary mobile device placement," in *5th International Conference on Indoor Positioning and Navigation (IPIN)*, October 2014, pp. 187–196.

[2] R. Hostettler and S. Särkkä, "IMU and magnetometer modeling for smartphone-based PDR," in *7th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Alcalá de Henares, Spain, October 2016.

[3] Z. Zhang, Q. Fang, and X. Gu, "Objective assessment of upper-limb mobility for poststroke rehabilitation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 859–868, April 2016.

[4] V. Bonnet, V. Joukov, D. Kulić, P. Fraisse, N. Ramdani, and G. Venture, "Monitoring of hip and knee joint angles using a single inertial measurement unit during lower limb rehabilitation," *IEEE Sensors Journal*, vol. 16, no. 6, pp. 1557–1564, March 2016.

[5] M. A. D. Raya and L. G. Sison, "Adaptive noise cancelling of motion artifact in stress ECG signals using accelerometer," in *24th International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, 2002, pp. 1756–1757.

[6] R. Hostettler, T. Lumikari, L. Palva, T. Nieminen, and S. Särkkä, "Motion artifact reduction in ambulatory electrocardiography using inertial measurement units and Kalman filtering," in *21th International Conference on Information Fusion (FUSION)*, Cambridge, UK, July 2018.

[7] D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology*, 2nd ed. The Institution of Engineering and Technology, 2004.

[8] V. Renaudin, M. H. Afzal, and G. Lachapelle, "Complete triaxis magnetometer calibration in the magnetic domain," *Journal of Sensors*, 2010.

[9] M. Kok, J. D. Hol, and T. B. Schön, "Using inertial sensors for position and orientation estimation," *Foundations and Trends in Signal Processing*, vol. 11, no. 1-2, pp. 1–153, 2017.

[10] G. M. Lerner and M. D. Shuster, "In-flight magnetometer calibration and attitude determination for near-earth spacecraft," *Journal of Guidance, Control, and Dynamics*, vol. 4, pp. 518–522, September 1981.

[11] R. Alonso and M. D. Shuster, "Attitude-independent magnetometer-bias determination: A survey," *Journal of the Astronautical Sciences*, vol. 50, no. 4, pp. 453–476, 2002.

[12] D. Gebre-Egziabher, G. H. Elkaim, J. D. Powell, and B. W. Parkinson, "Calibration of strapdown magnetometers in magnetic field domain," *Journal of Aerospace Engineering*, vol. 19, no. 2, pp. 87–102, 2006.

[13] X. Li, B. Song, W. Wang, J. Niu, and Z. Li, "Calibration and alignment of tri-axial magnetometers for attitude determination," *IEEE Sensors Journal*, vol. 18, no. 18, pp. 7399–7406, September 2018.

[14] H. E. Soken, "A survey of calibration algorithms for small satellite magnetometers," *Measurement*, vol. 122, pp. 417–423, 2018.

[15] 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.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice Hall, 1993.

[17] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *IEEE Computer Graphics and Applications*, vol. 25, no. 6, pp. 38–46, November 2005.

[18] H. E. Soken and C. Hajiyev, "UKF-based reconfigurable attitude parameters estimation and magnetometer calibration," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2614–2627, July 2012.

[19] M. Kok, J. D. Hol, T. B. Schön, F. Gustafsson, and H. Luinge, "Calibration of a magnetometer in combination with inertial sensors," in *15th International Conference on Information Fusion (FUSION)*, July 2012, pp. 787–793.

[20] F. Olsson, M. Kok, K. Halvorsen, and T. B. Schön, "Accelerometer calibration using sensor fusion with a gyroscope," in *IEEE Statistical Signal Processing Workshop (SSP)*, June 2016.

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

[22] A. Kuncar, M. Sysel, and T. Urbanek, "Calibration of low-cost accelerometer and magnetometer with differential evolution," in *2017 International Conference on Military Technologies (ICMT)*, May 2017, pp. 414–418.

[23] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory & Applications*, vol. 4, pp. 1303–1318, August 2010.

[24] K. V. Mardia and P. E. Jupp, *Directional Statistics*. Wiley, 2009.

[25] S. Särkkä, V. Tolvanen, J. Kannala, and E. Rahtu, "Adaptive Kalman filtering and smoothing for gravitation tracking in mobile systems," in *6th International Conference on Indoor Positioning and Indoor Navigation*, October 2015.

[26] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge, UK: Cambridge University Press, 2013.

[27] 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.

[28] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive Bayesian filtering in circular state spaces," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 3, pp. 70–87, March 2016.

[29] S. Sra, "A short note on parameter approximation for von Mises–Fisher distributions: and a fast implementation of $I_s(x)$," *Computational Statistics*, vol. 27, no. 1, pp. 177–190, March 2012.

[30] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 2006.

[31] E. Thébault, C. C. Finlay, C. D. Beggan *et al.*, "International geomagnetic reference field: the 12th generation," *Earth, Planets and Space*, no. 1, pp. 67–79, May 2015.

[32] Y. Wu, D. Zou, P. Liu, and W. Yu, "Dynamic magnetometer calibration and alignment to inertial sensors by Kalman filtering," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 2, pp. 716–723, March 2018.