Article

Progressive von Mises–Fisher Filtering Using Isotropic Sample Sets for Nonlinear Hyperspherical Estimation †

Kailai Li *, Florian Pfaff and Uwe D. Hanebeck

Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany; florian.pfaff@kit.edu (F.P.); uwe.hanebeck@kit.edu (U.D.H.)

* Correspondence: kailai.li@kit.edu; Tel.: +49-721-6084-8973

† This paper is an extended version of our paper published in K. Li, F. Pfaff, and U. D. Hanebeck, "Nonlinear von Mises–Fisher Filtering Based on Isotropic Deterministic Sampling", in Proceedings of the 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2020), Virtual, 14–16 September 2020.

Abstract: In this work, we present a novel scheme for nonlinear hyperspherical estimation using the von Mises–Fisher distribution. Deterministic sample sets with an isotropic layout are exploited for the efficient and informative representation of the underlying distribution in a geometrically adaptive manner. The proposed deterministic sampling approach allows manually configurable sample sizes, considerably enhancing the filtering performance under strong nonlinearity. Furthermore, the progressive paradigm is applied to the fusing of measurements of non-identity models in conjunction with the isotropic sample sets. We evaluate the proposed filtering scheme in a nonlinear spherical tracking scenario based on simulations. Numerical results show the evidently superior performance of the proposed scheme over state-of-the-art von Mises–Fisher filters and the particle filter.

Keywords: sensor fusion; recursive Bayesian estimation; directional statistics; unscented transform; nonlinear hyperspherical filtering

1. Introduction

The use of inferences on (hyper-)spherical states is ubiquitous in a large variety of application scenarios, such as protein structure prediction [1], rigid-body motion estimation [2,3], remote sensing [4], omnidirectional robotic perception [5,6] and scene segmentation and understanding [7,8]. In most of these tasks, quantifying uncertainties in hyperspherical domains is crucial (for brevity, the word "hypersphere" is used to denote spheres of any dimension in unspecified cases throughout the paper) . Therefore, the von Mises–Fisher distribution [9,10] has become a popular probabilistic model defined on the unit hypersphere $\mathbb{S}^{d-1} = \left\{ \underline{x} \in \mathbb{R}^d : \|\underline{x}\| = 1 \right\}$.

The recursive estimation of hyperspherical random variables using the von Mises–Fisher distribution is nontrivial due to its nonlinear and periodic nature on directional manifolds [10]. Samples generated from the underlying distribution are typically employed to propagate estimates through system dynamics or to evaluate the likelihoods given certain measurements. In [11,12], rejection sampling-based approaches were proposed to generate random samples for von Mises–Fisher distributions in arbitrary dimensions with an unbounded runtime. A deterministic runtime without resorting to rejection schemes is achievable, although only for specific numbers of dimensions; e.g., using the methods proposed in [13] on the unit sphere $\mathbb{S}^2$ or that given in [14] for odd numbers of dimensions.

Although a random sampling-based von Mises–Fisher filter is effective for nonlinear hyperspherical estimation, it cannot deliver reproducible results and may lack runtime efficiency (especially under strong nonlinearities or in high-dimensional state spaces). Therefore, deterministic sample sets are desired for an efficient and accurate representation

of the underlying distribution. Reminiscent of the unscented Kalman filter (UKF) for linear domains, the so-called unscented von Mises–Fisher filter (UvMFF) was proposed in [15] on unit hyperspheres $\mathbb{S}^{d-1} \in \mathbb{R}^d$. Following the idea of the unscented transform (UT), $2\,d - 1$ deterministic samples are drawn in a way that preserves the mean resultant vector of the underlying von Mises–Fisher distribution. Compared with confining a UKF to the manifold structure, this approach delivers superior performance for nonlinear hyperspherical tracking.

There remains considerable space for improvement for state-of-the-art von Mises–Fisher filtering schemes in the area of high-performance hyperspherical estimation. The deterministic sampling method used in the current UvMFF [15] only allows fixed numbers of hyperspherical samples (i.e., $2\,d - 1$ samples on $\mathbb{S}^{d-1}$) in accordance with the unscented transform. Moreover, the current UvMFF only allows identity measurement models with the measurement still confined to hyperspheres, and the measurement noise must be von Mises–Fisher-distributed. Thus, its practical deployment to arbitrary sensor modalities requires reapproximating the measurement model and the noise term [16], leading to additional preprocessing and errors. For arbitrary measurement models, directly reweighting the samples and fitting a posterior von Mises–Fisher to them is theoretically feasible. However, a limited number of deterministic samples is prone to degeneration, which is particularly risky under strong nonlinearities or with peaky likelihood functions. Therefore, it is important to enable deterministic sample sets of flexible sizes to better represent the underlying distribution while satisfying the condition of the unscented transform.

Generating deterministic sample sets of configurable sizes for continuous distributions was originally investigated in Euclidean spaces. In [17], deterministic samples were generated from a multivariate Gaussian distribution by minimizing the statistical divergence between its supporting Dirac mixture and the underlying continuous densities. For this, the Cramér–von Mises distance was generalized to the multivariate case based on the so-called localized cumulative distribution (LCD) to quantify the statistical divergence. This Dirac mixture approximation (DMA)-based method was further improved in [18] for better efficiency and extended in [19] for Gaussian mixtures.

In the context of recursive estimation based on distributions from directional statistics [10], deterministic samples are typically drawn by preserving moments up to a certain order. In [20], five deterministic samples were generated for distributions on circular domains; e.g., for the wrapped normal or the von Mises distribution [21]. For this, a sample set was scaled to match the first and the second trigonometric moments of the distribution. Sample sets for different scaling factors can then be merged to a larger set via superposition. In [22], a DMA-like sampling scheme was proposed to generate arbitrary numbers of deterministic samples while preserving the circular moments via optimization. In [23], deterministic samples were drawn from typical circular distributions via optimal quadratic quantification based on the Voronoi cells. For unit hyperspheres, major efforts have been dedicated to the Bingham distribution, where the basic UT-based sampling scheme in [24] ($2\,d - 1$ samples as for $\mathbb{S}^{d-1} \in \mathbb{R}^d$) was extended for arbitrary sample sizes, first in the principal directions [25] and then for the entire hyperspherical domain [26]. The sampling paradigm was DMA-based, with an on-manifold optimizer minimizing the statistical divergence of the samples to the underlying distribution under the moment constraints up to the second order. Based on this, improved filtering performance has been shown for quaternion-based orientation estimation.

Although non-identity measurement models can be handled by enlarging the sizes of deterministic samples in the update step, considerably large sample sizes are still desired in the face of degeneration issues (e.g., due to strong nonlinearities or peaky likelihoods). Thus, there remains the necessity to improve sample utilization. In [27], a novel progressive update scheme was proposed for nonlinear Gaussian filtering by decomposing the likelihood into a product of likelihoods with exponents adaptively determined by confining sample weight ratios within a pre-given threshold. Consequently, deterministic samples of small sizes are less likely to degenerate and become more deployable for nonlinear

estimation. Similar schemes have also been proposed for estimating angular systems [28] and Bingham-based hyperspherical filtering [29] with non-identity measurement models.

To date, there exists no flexible deterministic sampling scheme for von Mises–Fisher distributions. Existing optimization-based paradigms may have undesirable properties, such as local minima or a deteriorated runtime, for large sample sizes, which prohibit their deployment to online estimation tasks. Furthermore, no sample-efficient method is available for von Mises–Fisher filtering with non-identity measurement models.

In consideration of the state-of-the-art approaches above, we propose a novel algorithm for von Mises–Fisher distributions in arbitrary dimensions to obtain deterministic sample sets with manually configurable sizes. Based on hyperspherical geometries, samples are drawn coherently to the isotropic dispersion of the underlying distribution without optimization while satisfying the requirement of the unscented transform. Moreover, a novel progressive update scheme is developed in conjunction with the proposed sampling approach for nonlinear von Mises–Fisher filtering. Furthermore, an extensive evaluation of nonlinear spherical estimation is provided. Compared with existing von Mises–Fisher filtering schemes and the particle filter, the proposed progressive von Mises–Fisher filter using isotropic sample sets delivers superior performance with regard to tracking accuracy, runtime and memory efficiency.

The remainder of the paper is structured as follows. Preliminaries for the von Mises–Fisher distributions and the corresponding hyperspherical geometry are given in Section 2. Based on this, the novel isotropic deterministic sampling scheme is introduced in Section 3. In Section 4, the proposed progressive deterministic update for von Mises–Fisher filtering is provided, followed by a simulation-based benchmark of nonlinear spherical tracking in Section 5. The work is concluded in Section 6.

2. Preliminaries

2.1. General Conventions of Notations

We use underlined lowercase variables $\underline{x} \in \mathbb{R}^d$ to denote vectors. Random variables are denoted by lowercase boldface letters $\underline{\mathbf{x}}$. Uppercase boldface letters $\mathbf{B}$ are used to denote matrices. $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ denotes the unit $(d-1)$-sphere embedded in the d-dimensional Euclidean space. In the context of recursive Bayesian estimation, we denote the posterior density of the state at time step t, which relates to all measurements up to time step t, by f_t^{e}. f_{t+1}^{p} is used for the predicted density of the state at time step $t+1$ with regard to all measurements up to t. The rest of the symbols used are explained in the course of the following pages.

2.2. The von Mises–Fisher Distribution

Defined on the unit hypersphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$, the von Mises–Fisher distribution $\underline{\mathbf{x}} \sim \mathcal{VMF}(\underline{v}, \kappa)$ is parameterized by the mode location $\underline{v} \in \mathbb{S}^{d-1} \subset \mathbb{R}^d$ and the concentration parameter $\kappa \geq 0$. Its probability density function is given in the form

$$f_{\mathrm{vMF}}(\underline{x}) = N_d(\kappa) \cdot \exp(\kappa \underline{v}^\top \underline{x}), \ \underline{x} \in \mathbb{S}^{d-1}, \tag{1}$$

with the normalization constant

$$N_d(\kappa) = \left(\int_{\mathbb{S}^{d-1}} \exp(\kappa \underline{v}^\top \underline{x}) \, \mathrm{d}\underline{x} \right)^{-1} = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} \, \mathcal{I}_{d/2-1}(\kappa)} \tag{2}$$

depending on the concentration κ and the dimension d. $\mathcal{I}_{d/2-1}(\kappa)$ denotes the modified Bessel function of the first kind and order $d/2 - 1$. Note that the von Mises–Fisher distribution quantifies uncertainties using an arc length-based metric, which is coherent to the hyperspherical manifold structure. The distribution is unimodal and exhibits an isotropic dispersion. By generalizing the trigonometric moment from the circular to the hyper-

spherical domain, we obtain the *mean resultant vector* of the von Mises–Fisher distribution as follows

$$\underline{\alpha} = \mathbb{E}(\underline{x}) = \int_{\mathbb{S}^{d-1}} \underline{x} f_{\text{vMF}}(\underline{x})\, d\underline{x} = \mathcal{A}_d(\kappa)\, \underline{v}, \text{ with } \mathcal{A}_d(\kappa) = \frac{\mathcal{I}_{d/2}(\kappa)}{\mathcal{I}_{d/2-1}(\kappa)}$$

denoting the ratio of two Bessel functions. Therefore, the mean resultant vector is essentially a re-scaled hyperspherical mean $\underline{v}$, with a length of $\|\underline{\alpha}\| = \mathcal{A}_d(\kappa)$. Given a set of weighted samples $\{(\underline{x}_i, \omega_i)\}_{i=1}^n \subset \mathbb{S}^{d-1}$ with the weights $\sum_{i=1}^n \omega_i = 1$, a von Mises–Fisher distribution can be fitted to the sample mean $\underline{\hat{\alpha}} = \sum_{i=1}^n \omega_i \underline{x}_i$ via

$$\hat{\underline{v}} = \underline{\hat{\alpha}} / \|\underline{\hat{\alpha}}\| \quad \text{and} \quad \hat{\kappa} = \mathcal{A}_d^{-1}(\underline{\hat{\alpha}}). \tag{3}$$

To obtain the concentration $\hat{\kappa}$, one needs to solve the inverse of the Bessel function ratio in Equation (2), which can be efficiently obtained using the algorithm introduced in [30]. Moment matching to the mean resultant vector has been proven to be equivalent to maximum likelihood estimation (MLE) [31] (Section A.1) for the von Mises–Fisher distribution. Moreover, this also guarantees minimal information loss (quantified by the Kullback–Leibler divergence) when fitting a von Mises–Fisher to an arbitrary distribution [32] for stochastic filtering.

2.3. Geometric Structure of Hyperspherical Manifolds

The von Mises–Fisher distribution quantifies hyperspherical uncertainty in relation to the geodesic curve length on the manifold to the mode. To establish the proposed deterministic sampling scheme, we first investigate the hyperspherical domain from the perspective of Riemannian geometry [33]. Any point $\underline{x} \in \mathbb{S}^{d-1}$ can be mapped to the tangent space at $\underline{v} \in \mathbb{S}^{d-1}$ via the logarithm map

$$\underline{\tilde{x}} = \text{Log}_{\underline{v}}(\underline{x}) = (\underline{x} - \cos(\gamma)\,\underline{v}) \frac{\gamma}{\sin(\gamma)} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}, \text{ with } \gamma = \arccos(\underline{v}^\top \underline{x}),$$

while preserving its geodesic distance to $\underline{v}$; i.e., $|\gamma| = \|\text{Log}_{\underline{v}}(\underline{x})\|$. Inversely, any point $\underline{\tilde{x}} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ can be retracted to the unit hypersphere via the exponential map

$$\underline{x} = \text{Exp}_{\underline{v}}(\underline{\tilde{x}}) = \cos\left(\|\underline{\tilde{x}}\|\right)\underline{v} + \frac{\sin\left(\|\underline{\tilde{x}}\|\right)}{\|\underline{\tilde{x}}\|}\,\underline{\tilde{x}} \in \mathbb{S}^{d-1}.$$

When expressing logarithm-mapped points $\underline{\tilde{x}} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ with regard to an orthonormal basis of the tangent space, their local coordinates $\underline{\tilde{x}}^\ell$ essentially form a $(d-1)$-ball of radius π—i.e., $\underline{\tilde{x}}^\ell \in \mathbb{B}_\pi^{d-1} \subset \mathbb{R}^{d-1}$—which is bounded by the hypersphere $\mathbb{S}_\pi^{d-2}$ of radius π. To avoid ambiguities, we denote the logarithm and exponential maps defined for hyperspherical geometry above with capitalized first letters to distinguish them from the common logarithmic and exponential functions used in algebra [33].

3. Isotropic Deterministic Sampling

Considering the isotropic dispersion of the von Mises–Fisher distribution, we design a sample set layout with one sun sample at the mode surrounded by λ hyperspherical orbits of interval ζ. On each orbit, τ planet samples are placed (quasi-)equidistantly, thereby inducing a sample set $\mathbb{X} \subset \mathbb{S}^{d-1}$ of cardinality $\lambda\tau + 1$. All samples are equally weighted. One has to determine the interval value ζ that ensures that the samples are confined to the mean resultant vector of the underlying distribution, thereby preserving the unscented von Mises–Fisher filtering paradigm. Based on the introduction in Section 2, details about deriving the orbit interval follow.

The proposed isotropic deterministic sampling scheme is detailed in Algorithm 1 and illustrated in Figure 1. Given a von Mises–Fisher distribution on $\mathbb{S}^{d-1}$, we first place the sun sample at its mode (Algorithm 1, line 1). The tangent space at the mode, $\mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$, is bounded

by the hypersphere $\mathbb{S}^{d-2}_\pi$ with regard to its local basis $\mathcal{B}_{\underline{v}}$ (Algorithm 1, line 2). To obtain τ planet samples on each hyperspherical orbit, we first generate equidistant grid points on the unit hypersphere $\mathbb{S}^{d-2}$ using the equal area partitioning algorithm from [34] (Algorithm 1, line 3 and Figure 1A). Given the sampling configuration, the hyperspherical orbit interval ζ is then computed in accordance with the requirement of the unscented transform (Algorithm 1, line 4). Afterward, the obtained sample set $\{\underline{\tilde{x}}^\ell_s\}^\tau_{s=1} \subset \mathbb{S}^{d-2}$ with regard to $\mathcal{B}_{\underline{v}}$ is transformed into global coordinates scaled by each orbit radius $r\zeta$ $(r = 1, \ldots, \lambda)$ and undergoes the exponential map to land on the r-th orbit on $\mathbb{S}^{d-1}$ (Algorithm 1, line 5–9, Figure 1B,C). In order to determine the orbit interval ζ that guarantees the unscented transform for filtering, we provide the following derivations.

Algorithm 1: Isotropic Deterministic Sampling

    Input: $\mathcal{VMF}(\underline{v}, \kappa)$, number of orbits λ, per-orbit resolution τ
    Output: deterministic sample set $\mathbb{X}$
1 $\mathbb{X} \leftarrow \underline{v}$;
2 $\mathbf{B}_{\underline{v}} \leftarrow$ getBasis$(\underline{v})$;
3 $\{\underline{\tilde{x}}^\ell_s\}^\tau_{s=1} \leftarrow$ equalPartition$(\mathbb{S}^{d-2}, \tau)$;
4 $\zeta \leftarrow$ computeInterval(λ, τ);
5 **for** $r \leftarrow 1$ *to* λ **do**
6   **for** $s \leftarrow 1$ *to* τ **do**
7         $\underline{x}_{r,s} \leftarrow \mathrm{Exp}_{\underline{v}}(r\,\zeta\,\mathbf{B}_{\underline{v}}\,\underline{\tilde{x}}^\ell_s)$;
8         $\mathbb{X} \leftarrow \mathbb{X} \cup \underline{x}_{r,s}$;
9   **end**
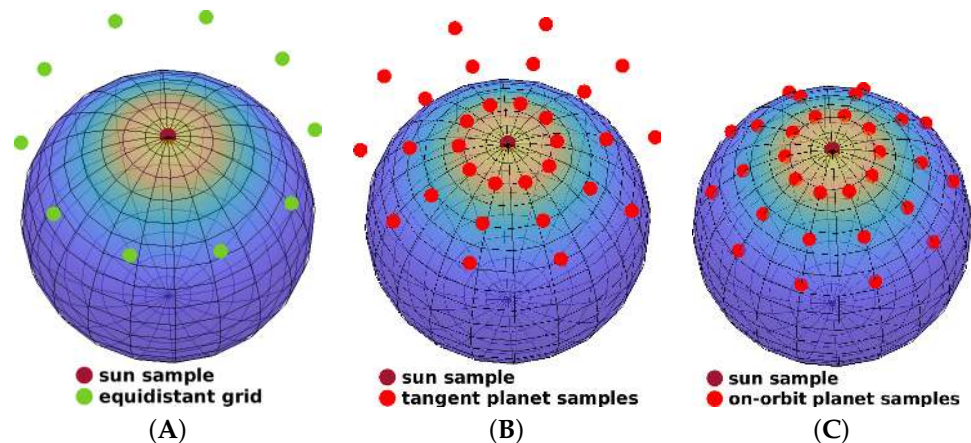10 **end**
11 **return** $\mathbb{X}$;

Figure 1. Illustration of isotropic deterministic sampling with $(\lambda, \tau) = (3, 10)$ for a von Mises–Fisher distribution ($\kappa = 4$) on $\mathbb{S}^2$. (**A**) Equal partitioning in $\mathbb{T}_{\underline{v}}\mathbb{S}^2$ with regard to its local basis. (**B**) Scaling with the UT-preserving interval in $\mathbb{T}_{\underline{v}}\mathbb{S}^2$. (**C**) Exponential map from $\mathbb{T}_{\underline{v}}\mathbb{S}^2$ to $\mathbb{S}^2$ for placing planet samples on hyperspherical orbits.

We map each point $\underline{\tilde{x}}^\ell_s$ generated by the equal area partitioning algorithm [34] with regard to $\mathcal{B}_{\underline{v}}$ to $\mathbb{S}^{d-1}$ according to Algorithm 1, line 7, and obtain

$$\underline{x}_{r,s} = \mathrm{Exp}_{\underline{v}}(r\,\zeta\,\mathbf{B}_{\underline{v}}\,\underline{\tilde{x}}^\ell_s) = \cos(r\,\zeta)\,\underline{v} + \sin(r\,\zeta)\,\mathbf{B}_{\underline{v}}\,\underline{\tilde{x}}^\ell_s \in \mathbb{S}^{d-1}\,,$$

with $\underline{x}_{r,s}$ denoting the s-th planet sample on the r-th hyperspherical orbit. The bold letter $\mathbf{B}_{\underline{v}} \in \mathbb{R}^{d\times(d-1)}$ denotes the matrix transforming coordinates from the local basis $\mathcal{B}_{\underline{v}}$ to the global one. Then, the hyperspherical mean of the whole sample set (including the sun sample) is

$$
\begin{aligned}
\underline{\alpha} &= \frac{1}{\lambda\tau+1}\left(\underline{\nu} + \sum_{r=1}^{\lambda}\sum_{s=1}^{\tau} \underline{x}_{r,s}\right) \\
&= \frac{1}{\lambda\tau+1}\left(\underline{\nu} + \tau\sum_{r=1}^{\lambda}\cos(r\,\zeta)\,\underline{\nu} + \sum_{r=1}^{\lambda}\sum_{s=1}^{\tau}\sin(r\,\zeta)\,\mathbf{B}_{\underline{\nu}}\tilde{\underline{x}}_s^{\ell}\right).
\end{aligned}
\tag{4}
$$

For typical configurations of the equal area partitioning algorithm for unit hyperspheres [34], the sample set $\{\tilde{\underline{x}}_s^{\ell}\}_{s=1}^{\tau}$ is zero-centered. Therefore, the formula in Equation (4) can be further simplified as

$$
\underline{\alpha} = \frac{1}{\lambda\tau+1}\left(1 + \tau\sum_{r=1}^{\lambda}\cos(r\,\zeta)\right)\underline{\nu}.
$$

By constraining the sample set mean to be identical to the mean resultant vector of the underlying distribution—i.e., $\underline{\alpha} = \mathcal{A}_d(\kappa)\,\underline{\nu}$—the hyperspherical moment in Equation (3) is maintained, thereby satisfying the requirement of the unscented transform. Consequently, we have

$$
\sum_{r=1}^{\lambda}\cos(r\,\zeta) = \frac{(\lambda\tau+1)\mathcal{A}_d(\kappa)-1}{\tau}.
$$

By exploiting Lagrange's trigonometric identity [35] (Section 2.4.1.6), the finite series in the equation above can be further simplified, and we obtain

$$
\frac{\sin\left((\lambda+0.5)\zeta\right)}{2\sin(0.5\,\zeta)} = \frac{(\lambda\tau+1)\mathcal{A}_d(\kappa)-1}{\tau} + \frac{1}{2}.
$$

The left-hand side fits the form of the (scaled) Dirichlet kernel [36] $\mathcal{D}_\lambda(\zeta)$ and the desired orbit interval ζ is obtained by solving the equation

$$
\mathcal{D}_\lambda(\zeta) = \frac{(\lambda\tau+1)\mathcal{A}_d(\kappa)-1}{\tau} + \frac{1}{2}, \text{ with } \mathcal{D}_\lambda(\zeta) = \frac{\sin\left((\lambda+0.5)\zeta\right)}{2\sin(0.5\,\zeta)},\ \zeta\in\left[0,\frac{\pi}{\tau}\right].
\tag{5}
$$

Note that Equation (5) does not have a closed-form solution. It is trivial to prove that the maximum of the Dirichlet kernel is obtained at $\zeta=0$ with $\mathcal{D}_\lambda^{\max}(\zeta) = \mathcal{D}_\lambda(0) = \lambda + 0.5$. Meanwhile, the constant on the right-hand side of Equation (5) is smaller than $\mathcal{D}_\lambda^{\max}(\zeta)$ given the Bessel function ratio $\mathcal{A}_d(\kappa)\in(0,1)$ for $\kappa>0$. Therefore, Equation (5) is solvable for $\zeta\in[0,\pi/\tau]$.

3.1. Numerical Solution for Equation (5)

Instead of deploying a universal numerical solver (e.g., the function `solve` in Matlab) to solve Equation (5) as in our preceding work [37], we now provide a tailored Newton's method with iterative steps of a closed form. For that, the first derivative of the Dirichlet kernel is provided as follows:

$$
\begin{aligned}
\mathcal{D}_\lambda'(\zeta) &= \frac{(\lambda+0.5)\cos\left((\lambda+0.5)\zeta\right)\sin(0.5\,\zeta) - 0.5\sin\left((\lambda+0.5)\zeta\right)\cos(0.5\,\zeta)}{2\left(\sin(0.5\,\zeta)\right)^2} \\
&= 0.5\,(\lambda+0.5)\cos\left((\lambda+0.5)\zeta\right)\csc(0.5\,\zeta) - 0.5\,\mathcal{D}_\lambda(\zeta)\cot(0.5\,\zeta).
\end{aligned}
$$

Then, the $(k+1)$-th Newton step for updating ζ_k is given as

$$
\zeta_{k+1} = \zeta_k - \frac{\mathcal{D}_\lambda(\zeta_k) - \left((\lambda\tau+1)\mathcal{A}_d(\kappa)-1\right)/\tau - 0.5}{0.5\,(\lambda+0.5)\cos\left((\lambda+0.5)\zeta_k\right)\csc(0.5\,\zeta_k) - 0.5\,\mathcal{D}_\lambda(\zeta_k)\cot(0.5\,\zeta_k)}.
$$

To initialize ζ of the Newton's method, we perform a linear interpolation between 0 and the first non-negative zero of the Dirichlet kernel, $\pi/(\lambda+0.5)$, with regard to their

function values $\mathcal{D}_\lambda(0) = \lambda + 0.5$ and $\mathcal{D}_\lambda(\pi/(\lambda+0.5)) = 0$, respectively. We substitute the right-hand side of Equation (5) with $c = (\lambda\tau + 1)\mathcal{A}_d(\kappa)/\tau - 1/\tau + 0.5$ and obtain

$$\zeta_0 = \frac{\pi\,(\lambda + 0.5 - c)}{(\lambda + 0.5)^2} = \frac{\pi(\lambda + 1/\tau)(1 - \mathcal{A}_d(\kappa))}{(\lambda + 0.5)^2}\,.$$

In practice, the Newton's method specified above with the proposed initialization results in convergence below the error threshold 10^{-7} within five steps, which is faster than our implementation in [37] by two orders of magnitude, thereby guaranteeing efficient sampling performance for online estimation. We now consider the following example to illustrate the efficacy of the proposed isotropic sampling scheme on von Mises–Fisher distributions of various configurations.

3.2. Example

We parameterize the von Mises–Fisher distribution on the unit sphere $\mathbb{S}^2$ with three concentration values $\kappa = \{0.5, 2, 4\}$. Without loss of generality, the three distributions are given the same mode $\underline{v} = [0, 0, 1]^\top$. Five configurations are used for the proposed sampling method; i.e., $(\lambda, \tau) = \{(3, 10), (5, 10), (5, 20), (10, 10), (10, 20)\}$. As shown in Figure 2, the isotropic sample sets are adapted to the dispersions for various parameterizations and configurations while preserving the mean resultant vector of the underlying distributions.
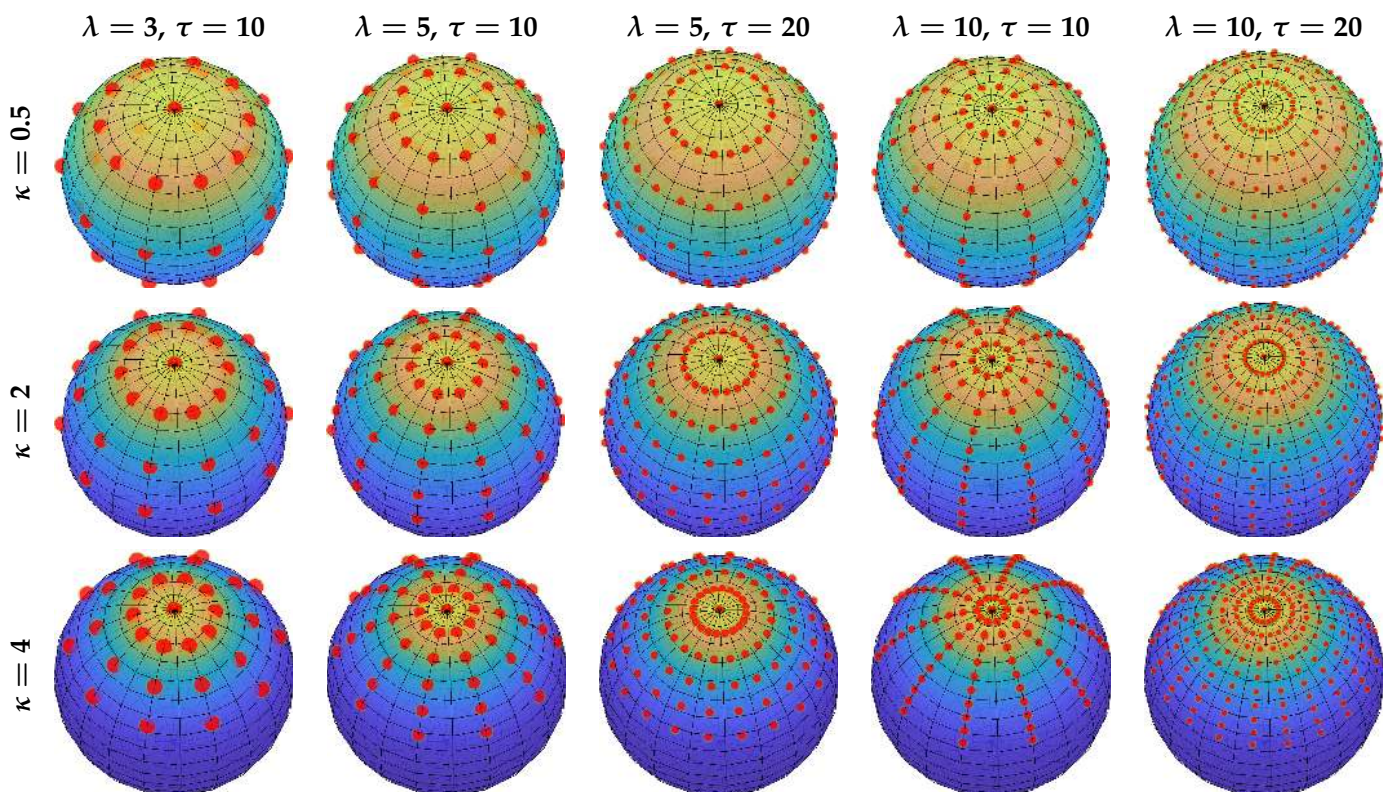

Figure 2. Illustration of the proposed isotropic deterministic sampling schemes with von Mises–Fisher distributions on $\mathbb{S}^2$ of different parameterizations in Section 3.2. Samples (red dots) are uniformly weighted and dotted with sizes proportional to weights.

4. Progressive Unscented von Mises–Fisher Filtering

The proposed sampling method yields isotropic deterministic sample sets of arbitrary sizes that represent the underlying uncertainty more comprehensively for the unscented transform. As shown in our preceding work [37], the current unscented von Mises–Fisher filtering scheme is thus considerably enhanced for nonlinear estimation. However, for nonlinear and non-identity measurement models, the current paradigm simply reweights

prior samples based on the likelihoods for the moment-matching of the posterior estimates. Although superior efficiency was shown over the approach using random samples, large sizes of deterministic samples are still desirable under strong nonlinearity [37] or with peaky likelihoods to avoid degeneration. To alleviate this issue, we propose the progressive unscented von Mises–Fisher filter using isotropic sample sets.

4.1. Task Formulation

We consider the following hyperspherical estimation scenario. The system model is assumed to be given as an equation of random variables:

$$\underline{x}_{t+1} = \underline{a}(\underline{x}_t, \underline{w}_t),$$

with $\underline{x}_t, \underline{x}_{t+1} \in \mathbb{S}^{d-1} \subset \mathbb{R}^d$ representing the hyperspherical states and $\underline{w}_t \in \mathbb{W}$ the system noise. The transition function $\underline{a} \colon \mathbb{S}^{d-1} \times \mathbb{W} \to \mathbb{S}^{d-1}$ maps the state from time step t to $t+1$ under consideration of the noise term. The measurement model is given as

$$\underline{z}_t = \underline{h}(\underline{x}_t, \underline{v}_t),$$

where $\underline{z}_t \in \mathbb{Z}$, $\underline{v}_t \in \mathbb{V}$ are the measurement and the measurement noise, respectively. $\underline{h} \colon \mathbb{S}^{d-1} \times \mathbb{V} \to \mathbb{Z}$ denotes the observation function.

4.2. Prediction Step for Nonlinear von Mises–Fisher Filtering

Given the setup above, one can use the Chapman–Kolmogorov equation to obtain the prior density from the last posterior $f_t^e(\underline{x}_t)$:

$$f_{t+1}^{\mathrm{p}}(\underline{x}_{t+1}) = \int_{\mathbb{S}^{d-1}} f_t^{\mathrm{e}}(\underline{x}_t) \int_{\mathbb{W}} f(\underline{x}_{t+1} \mid \underline{w}_t, \underline{x}_t) f_t^{\underline{w}}(\underline{w}_t) \, \mathrm{d}\underline{w}_t \, \mathrm{d}\underline{x}_t. \tag{6}$$

We follow the generic framework of von Mises–Fisher filtering with samples facilitating the inference procedure. The estimates from the prediction and update steps are thus expressed in the form of von Mises–Fisher distributions. We allow arbitrary motion models. As explained in the following paragraphs, we use two different implementations of the prediction step according to the forms of the transition density.

(1) For a generic transition density, we first represent the posterior density $f_t^e(\underline{x}_t)$ of the previous step using a sample set generated from the von Mises–Fisher distribution; namely,

$$f_t^{\mathrm{e}}(\underline{x}_t) = \sum_{i=1}^{n} \omega_{t,i}^{\mathrm{e}} \, \delta(\underline{x}_t - \underline{x}_{t,i}^{\mathrm{e}}). \tag{7}$$

where $\delta(\cdot)$ denotes the Dirac delta distribution and $\omega_{t,i}^{\mathrm{e}}$ represents the sample weights satisfying $\sum_{i=1}^{n} \omega_{t,i}^{\mathrm{e}} = 1$. The prediction step in Equation (6) now turns into

$$f_{t+1}^{\mathrm{p}}(\underline{x}_{t+1}) = \sum_{i=1}^{n} \omega_{t,i}^{\mathrm{e}} \int_{\mathbb{W}} f(\underline{x}_{t+1} \mid \underline{w}_t, \underline{x}_{t,i}^{\mathrm{e}}) f_t^{\underline{w}}(\underline{w}_t) \, \mathrm{d}\underline{w}_t. \tag{8}$$

Similarly, the noise distribution $f_t^{\underline{w}}(\underline{w}_t)$ of arbitrary form is also represented by a sample set; i.e., $f_t^{\underline{w}}(\underline{w}_t) = \sum_{j=1}^{m} \omega_{t,j}^{\underline{w}} \delta(\underline{w}_t - \underline{w}_{t,j})$, with the sample weights $\sum_{j=1}^{n} \omega_{t,j}^{\underline{w}} = 1$. Equation (6) is then reduced to

$$f_{t+1}^{\mathrm{p}}(\underline{x}_{t+1}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \omega_{t,i}^{\mathrm{e}} \cdot \omega_{t,j}^{\underline{w}} \cdot \delta\big(\underline{x}_{t+1} - \underline{a}(\underline{x}_{t,i}^{\mathrm{e}}, \underline{w}_{t,j})\big),$$

in which all elements of the Cartesian product of the posterior samples and the noise samples are propagated through the system function. As also shown in [37], the predicted von Mises–Fisher is fitted to the samples via moment matching.

(2) When the noise term $\underline{w}_t$ is additive and von Mises–Fisher-distributed, we obtain a transition density in the form of a von Mises–Fisher distribution [15]; namely,

$$f_t^{\mathrm{T}}(\underline{x}_{t+1} \,|\, \underline{x}_t) = f_{\mathrm{vMF}}(\underline{x}_{t+1}; \underline{a}_t(\underline{x}_t), \kappa_t^{\underline{w}}),$$

with $\underline{a}(t) \colon \mathbb{S}^{d-1} \to \mathbb{S}^{d-1}$ being a noise-invariant system function of arbitrary form and $\kappa_t^{\underline{w}}$ denoting the concentration of the noise distribution. Then, the predicted density in Equation (6) can be expressed as

$$
\begin{aligned}
f_{t+1}^{\mathrm{P}}(\underline{x}_{t+1}) &= \int_{\mathbb{S}^{d-1}} f_t^{\mathrm{T}}(\underline{x}_{t+1} \,|\, \underline{x}_t)\, f_t^{\mathrm{e}}(\underline{x}_t) \,\mathrm{d}\underline{x}_t \\
&= \int_{\mathbb{S}^{d-1}} f_{\mathrm{vMF}}(\underline{x}_{t+1}; \underline{a}_t(\underline{x}_t), \kappa_t^{\underline{w}})\, f_t^{\mathrm{e}}(\underline{x}_t) \,\mathrm{d}\underline{x}_t .
\end{aligned}
\tag{9}
$$

To obtain the predicted density in Equation (9), we first propagate the posterior sample set $\{\underline{x}_{t,i}^{\mathrm{e}}\}_{i=1}^{n}$ in Equation (7) through the motion model to obtain the propagated sample set $\{\underline{a}_t(\underline{x}_{t,i}^{\mathrm{e}})\}_{i=1}^{n}$. To approximate the density after applying the system function, a von Mises–Fisher distribution $\mathcal{VMF}(\underline{v}_t, \kappa_t)$ is then fitted to the propagated samples $\hat{\alpha} = \sum_{i=1}^{n} \omega_{t,i}\, \underline{a}_t(\underline{x}_{t,i}^{\mathrm{e}})$ via moment matching as introduced in Equation (3). By convolving the fitted von Mises–Fisher distribution with that for the noise term, the predicted estimate $\mathcal{VMF}(\underline{v}_{t+1}^{\mathrm{P}}, \kappa_{t+1}^{\mathrm{P}})$ is obtained via $\underline{v}_{t+1}^{\mathrm{P}} = \underline{v}_t$ and $\kappa_{t+1}^{\mathrm{P}} = \mathcal{A}_d^{-1}\big(\mathcal{A}_d(\kappa_t)\mathcal{A}_d(\kappa_t^{\underline{w}})\big)$. A detailed formulation of the method can be found in [15] (Algorithm 2).

4.3. Deterministic Progressive Update Using Isotropic Sample Sets

For nonlinear and non-identity measurement models, the posterior density can be obtained by reweighting the prior samples $\{\underline{x}_{t,i}^{\mathrm{P}}\}_{i=1}^{n}$ using the likelihood $f_t^{\mathrm{L}}(\underline{\hat{z}}_t \,|\, \underline{x}_t)$ given the measurement $\underline{\hat{z}}_t$ as follows:

$$f_t^{\mathrm{e}}(\underline{x}_t \,|\, \underline{\hat{z}}_t) \propto f_t^{\mathrm{L}}(\underline{\hat{z}}_t \,|\, \underline{x}_t)\, f_t^{\mathrm{P}}(\underline{x}_t) = \sum_{i=1}^{n} \omega_{i,t}^{\mathrm{P}} \cdot f_t^{\mathrm{L}}(\underline{\hat{z}}_t \,|\, \underline{x}_{t,i}^{\mathrm{P}}) \cdot \delta(\underline{x}_t - \underline{x}_{t,i}^{\mathrm{P}}) . \tag{10}$$

The posterior distribution is obtained by fitting a von Mises–Fisher distribution to the reweighted samples via moment matching. Directly applying the likelihood functions to the sample weights can be risky (regardless of whether they are generated randomly or deterministically) for strong nonlinearities or non-identity measurement models with peaky likelihoods due to the sample degeneration.

Therefore, we develop a novel update approach by deploying the proposed isotropic sampling to a progressive measurement update scheme [27]. More specifically, the likelihood in Equation (10) is decomposed into a product of l components:

$$f_t^{\mathrm{e}}(\underline{x}_t \,|\, \underline{\hat{z}}_t) \propto f_t^{\mathrm{L}}(\underline{\hat{z}}_t \,|\, \underline{x}_t) \cdot f_t^{\mathrm{P}}(\underline{x}_t) = \left(\prod_{k=1}^{l} \big(f_t^{\mathrm{L}}(\underline{\hat{z}}_t \,|\, \underline{x}_t)\big)^{\Delta_k} \right) \cdot f_t^{\mathrm{P}}(\underline{x}_t) , \tag{11}$$

with $\sum_{k=1}^{l} \Delta_k = 1$. The exponent Δ_k indicates the progression stride and is determined by a prespecified threshold $\epsilon \in (\,0, 1\,]$ to bound the likelihood ratios among the deterministic samples according to

$$\left(\frac{\min\big(\{v_{k,i}\}_{i=1}^{\lambda\tau+1}\big)}{\max\big(\{v_{k,i}\}_{i=1}^{\lambda\tau+1}\big)} \right)^{\Delta_k} \geq \epsilon , \quad \text{with} \quad v_{k,i} = f_t^{\mathrm{L}}(\underline{\hat{z}}_t \,|\, \underline{x}_{k,i}^{\mathrm{P}}) \tag{12}$$

which is the likelihood of the prior isotropic sample $\underline{x}_{k,i}^{\mathrm{P}}$ at the k-th progression step. Thus, we obtain

$$\Delta_k \leq \frac{\log(\epsilon)}{\log(v_k^{\min}) - \log(v_k^{\max})} ,$$

with $v_k^{\min} = \min \left(\{v_{k,i}\}_{i=1}^{\lambda\tau+1} \right)$ and $v_k^{\max} = \max \left(\{v_{k,i}\}_{i=1}^{\lambda\tau+1} \right)$, respectively. The progression stride is thus adaptively determined based on the variance of the samples' likelihoods at the current progression step. We repeat this sampling–reweighting–fitting cycle based on the density obtained from the previous progression step until the likelihood is fully fused into the result (exponents Δ_k sum to one).

The procedure above is detailed with pseudo-code in Algorithm 2. We first initialize the posterior density with that obtained from the prediction step and set the remaining progression horizon to $\Delta = 1$ (Algorithm 2, line 1–2). At each progression step, an isotropic deterministic sample set is drawn from the current posterior density $f_t^e(\underline{x}_t)$ (Algorithm 2, line 3–4). For each sample, we evaluate the likelihood for the measurement $\underline{z}_t$ and determine the maximal and minimal values of the likelihood values (Algorithm 2, line 5–7). Based on this, the current progression stride Δ_k is then computed according to Equation (12) (Algorithm 2, line 8). The posterior density is then fitted to the samples with the weights re-scaled according to the obtained progression stride Δ_k, as shown in Equation (11) (Algorithm 2, line 9–10). We repeat the progression step until Δ reaches zero, which is when the entire likelihood has been incorporated into the density (Algorithm 2, line 10–11).

Algorithm 2: Isotropic Progressive Update

Input: prior density $f_t^P(\underline{x}_t)$, measurement $\underline{\hat{z}}_t$, likelihood $f_t^L(\underline{z}_t \mid \underline{x}_t)$, threshold ϵ
Output: posterior density $f_t^e(\underline{x}_t)$

1　$f_t^e(\underline{x}_t) \leftarrow f_t^P(\underline{x}_t)$;
2　$\Delta \leftarrow 1, k \leftarrow 1$;
3　**while** $\Delta > 0$ **do**
4　　$\{\underline{x}_{k,i}^P\}_{i=1}^{\lambda\tau+1} \leftarrow \texttt{sampleIsoDeterministic}\left(f_t^e(\underline{x}_t), \lambda, \tau \right)$;　　　// see Algorithm 1
5　　$\{v_{k,i}\}_{i=1}^{\lambda\tau+1} \leftarrow \{ f_t^L(\underline{\hat{z}}_t \mid \underline{x}_{k,i}^P) \}_{i=1}^{\lambda\tau+1}$;　　　　// element-wise assignment
6　　$v_k^{\min} \leftarrow \min \left(\{v_{k,i}\}_{i=1}^{\lambda\tau+1} \right)$;
7　　$v_k^{\max} \leftarrow \max \left(\{v_{k,i}\}_{i=1}^{\lambda\tau+1} \right)$;
8　　$\Delta_k \leftarrow \min \left(\Delta, \log(\epsilon) / \log(v_k^{\min}/v_k^{\max}) \right)$;
9　　$\{v_{k,i}\}_{i=1}^{\lambda\tau+1} \leftarrow \{ (v_{k,i})^{\Delta_k} \}^{\lambda\tau+1}$;　　　　　// element-wise assignment
10　　$f_t^e(\underline{x}_t) \leftarrow \text{fitVMF} \left(\{\underline{x}_{k,i}^P\}_{i=1}^{\lambda\tau+1}, \{v_{k,i}\}_{i=1}^{\lambda\tau+1} \right)$;
11　　$\Delta \leftarrow \Delta - \Delta_k$;
12　　$k \leftarrow k + 1$;
13　**end**
14　**return** $f_t^e(\underline{x}_t)$;

A full series of progression steps (for $\epsilon = 0.02$) using isotropic sample sets is illustrated in Figure 3 and compared with a conventional single-step update. For both approaches, 21 samples are used in the configuration $(\lambda, \tau) = (2, 10)$. As shown in Figure 3A, given a prior von Mises–Fisher distribution on $\mathbb{S}^2$ and a relatively peaky likelihood function, the single-step update deteriorates evidently due to sample degeneration. In contrast, the progressive approach (Figure 3B–1 to 4) performs four progression steps and achieves a superior fusion result.
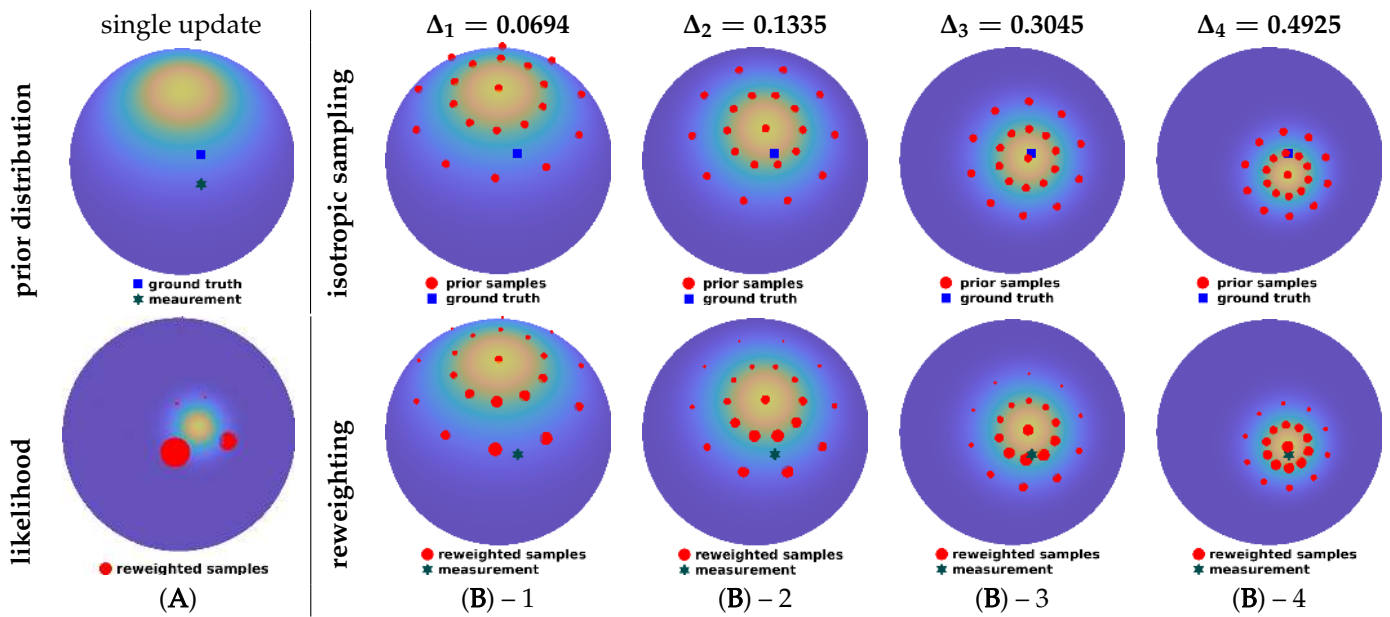
Figure 3. Illustration of the deterministic progressive update using isotropic sample sets. Sizes of red dots are proportional to their weights. The same isotropic sampling configuration, $(\lambda, \tau) = (2, 10)$, is deployed for both the single-step and the progressive updates.

5. Evaluation

We evaluate the proposed Prog-UvMFF using isotropic sample sets for nonlinear spherical estimation with a non-identity measurement model. To underline the merit of isotropic sampling and its integration into the proposed progressive deterministic update, we consider case 2 in Section 4.2 for the transition density with $f_t^{\mathrm{T}}(\underline{x}_{t+1} \mid \underline{x}_t) = f_{\mathrm{vMF}}(\underline{x}_{t+1}; \underline{a}_t(\underline{x}_t), \kappa^{\underline{w}})$, where $\underline{x}_t, \underline{x}_{t+1} \in \mathbb{S}^2$. The system dynamics is given as

$$\underline{a}_t(\underline{x}_t) = \frac{\sin(t/10) \cdot \underline{x}_t + (1 - \sin(t/10)) \cdot \underline{\sigma}}{\left\| \sin(t/10) \cdot \underline{x}_t + (1 - \sin(t/10)) \cdot \underline{\sigma} \right\|}, \quad \text{with} \quad \underline{\sigma} = [1, 1, 1]^\top / \sqrt{3},$$

which corresponds to the normalized linear interpolation [38] with a time-invariant interpolation ratio. We set the concentration κ of the von Mises–Fisher-distributed transition density to 50. Unlike the evaluation scenario in [37], the posterior of the previous step is propagated using samples and the predicted von Mises–Fisher prior is obtained by convolving the fitted density with the system noise as introduced in the second case of Section 4.2.

The nonlinear and non-identity measurement model yields the spherical coordinates (azimuth and elevation) of the state $\underline{x}_t = [\underline{x}_{t,1}, \underline{x}_{t,2}, \underline{x}_{t,3}]^\top$, i.e.,

$$\underline{z}_t = \underline{h}(\underline{x}_t) + \underline{v}_t, \quad \text{with} \quad \underline{h}(\underline{x}_t) = \left[\arctan\left(\frac{\underline{x}_{t,2}}{\underline{x}_{t,1}} \right), \arctan\left(\frac{\underline{x}_{t,3}}{\sqrt{\underline{x}_{t,1}^2 + \underline{x}_{t,2}^2}} \right) \right]^\top.$$

The additive measurement noise is zero-mean Gaussian-distributed; namely, $\underline{v}_t \sim \mathcal{N}(\underline{0}, \Sigma^{\underline{v}})$, with $\underline{0} \in \mathbb{R}^2$ and covariance $\Sigma^{\underline{v}} \in \mathbb{R}^{2 \times 2}$. Thus, the likelihood function is

$$f_t^{\mathrm{L}}(\underline{z}_t \mid \underline{x}_t) = f_{\mathrm{N}}(\underline{z}_t - \underline{h}(\underline{x}_t)). \tag{13}$$

We set the covariance $\Sigma^{\underline{v}} = 0.002 \cdot \mathbf{I}_{2 \times 2}$ to induce a peaky likelihood function.

Three variants from the von Mises–Fisher filtering framework are considered for the evaluation: the plain von Mises–Fisher filter (vMFF) based on random sampling, the unscented von Mises–Fisher filter (UvMFF) using deterministic sample sets and the progressive UvMFF (Prog-UvMFF), which fuses the measurements via progressions. The threshold ϵ controlling the progression stride in Equation (12) is set to 0.02. The random

samples in the vMFF are drawn using the approach in [13]. To generate deterministic samples in the UvMFF and the Prog-UvMFF, we involve both of the UT-based methods with a fixed sample size (as for $\mathbb{S}^2$, $n = 5$) [15] and the proposed isotropic sampling with configurable sizes. Furthermore, we run the particle filter (PF) with a typical sampling–importance resampling approach as a baseline. All the filters are initialized using the same prior von Mises–Fisher distribution $\mathcal{VMF}(\underline{v}_0, \kappa_0)$, where $\underline{v}_0 = [0, 0, 1]^\top$ and $\kappa_0 = 50$. The error between the ground truth $\underline{x}$ and the estimated state $\hat{\underline{x}}$ is quantified by the arc length on $\mathbb{S}^2$ in radians; i.e.,

$$\mathcal{E}(\underline{x}, \hat{\underline{x}}) = \mathrm{acos}(\underline{x}^\top \hat{\underline{x}}).$$

The scenario is simulated for 30 time steps in each run. A total of 1000 Monte Carlo runs is used for the evaluation. A broad scale of sample sizes (from 5 to 10^4) is considered. Deviations are summarized in the form of the root mean squared error (RMSE).

The evaluation results are plotted in Figures 4–6. As shown by the blue curve in Figure 4, the proposed isotropic sampling method allows the ordinary UvMFF [15] to deploy configurable sizes (any number larger than five) of deterministic samples, thereby achieving a superior performance over the random sampling-based filters (vMFF and PF). Due to the peaky likelihood function in Equation (13), however, its progressive variant (Prog-vMFF) delivers much better tracking accuracy (with the same sample size) as well as convergence. Figure 5 shows the runtime efficiency of the evaluated filters. As indicated by the green and blue curves, the runtime of the proposed isotropic sampling method is similar to that of the random one (as the two filters are based on the same filtering procedure) and the two filters are faster than the PF with the same numbers of samples. For the proposed Prog-UvMFF, the progressive measurement fusion induces slightly more runtime than the fusion with a conventional single-step update (while still being faster than the PF). The cost-efficiency (in terms of runtime) of different filters is displayed in Figure 6. Given the same amount of processing time, the proposed isotropic sampling method facilitates the UvMFF in delivering less error than the random counterpart. Furthermore, it enables the Prog-UvMFF to achieve the best tracking accuracy in conjunction with the progressive update.
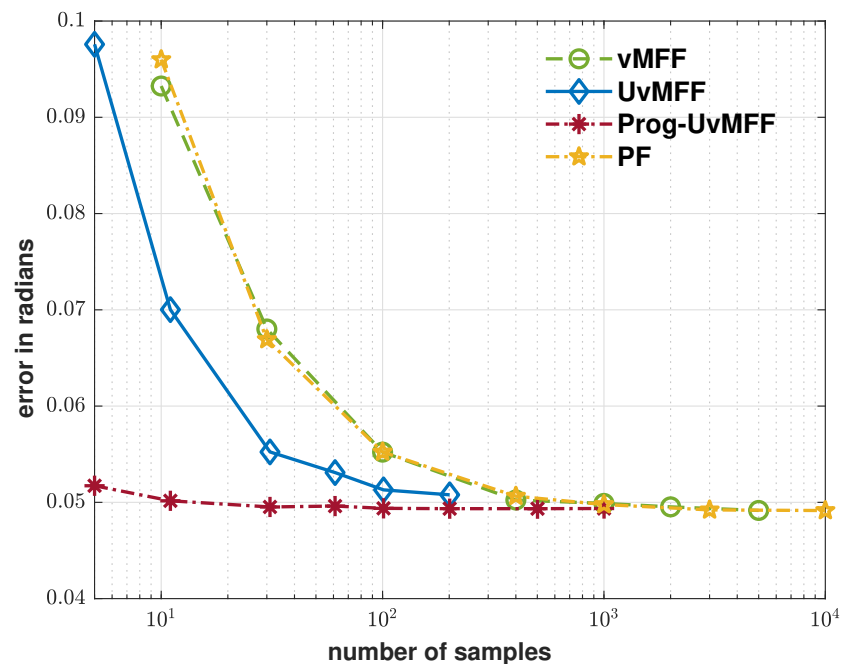

Figure 4. Error over sample numbers (log scale) for the evaluated filters. The configurations with five samples for UvMFF and Prog-UvMFF are based on the original UT-based sampling method in [15].
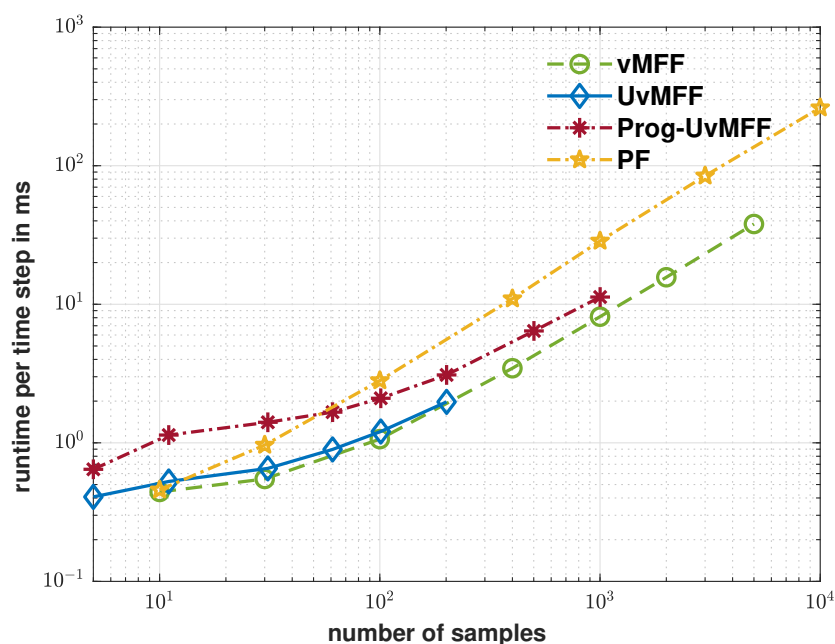
Figure 5. Runtime for each time step in ms over sample size for the evaluated filters.
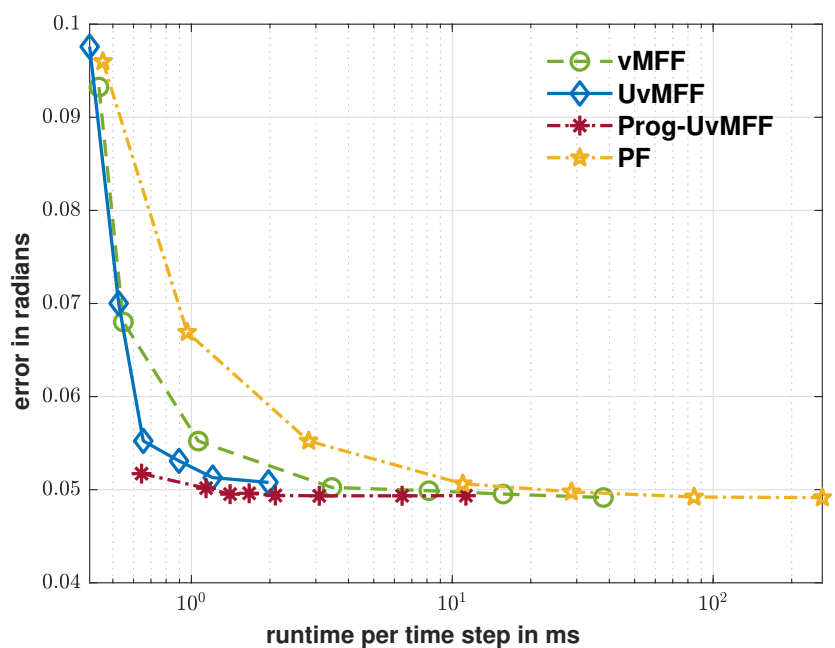

Figure 6. Error over runtime for the evaluated filters.

6. Conclusions

In this work, we propose a new deterministic sampling method for generating equally weighted sample sets of configurable sizes from von Mises–Fisher distributions in arbitrary dimensions. Based on hyperspherical geometries, the sample sets are placed in isotropic layouts adapted to the dispersion of the underlying distribution while satisfying the requirement of the unscented transform. To further enhance nonlinear von Mises–Fisher filtering techniques, we propose a deterministic progressive update step to handle non-identity measurement models. The final product, the Prog-UvMFF, is built upon the progressive filtering scheme with isotropic sample sets and delivers evidently superior performance over state-of-the-art von Mises–Fisher filters and the PF for nonlinear hyperspherical estimation.

Besides the theoretical contribution to recursive estimation for directional manifolds, the presented progressive unscented von Mises–Fisher filter supports generic measurement

models that are directly derived from the true sensor modalities. Thus, it is also of interest to evaluate the filter's performance in real-world tasks. Potential application scenarios include orientation estimation using omnidirectional vision [5], visual tracking on unit hyperspheres [39], bearing-only localization in sensor networks [40], wavefront orientation estimation in the surveillance field [29] and sound source localization [41].

There are multiple directions for further research. In addition to only matching the mean resultant vector, the higher-order shape information of a von Mises–Fisher distribution can be considered, which may lead to further enhancements in the filter performance. For this, deterministic samples can be non-uniformly weighted. Since hyperspherical uncertainties can be of an arbitrary shape in practice, parametric filtering can be error-prone in certain cases (e.g., in the presence of multimodality). Mixtures of von Mises–Fisher distributions can be exploited for more exact modeling, and corresponding recursive estimators are promising.

Author Contributions: K.L. proposed the deterministic isotropic sampling scheme and the progressive filtering for the von Mises–Fisher filter. F.P. engaged in helpful technical discussions and supported the evaluations. U.D.H. supervised all the modules of the work. K.L. prepared the manuscript, and all the authors contributed to polishing it. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hoff, P.D. Simulation of the Matrix Bingham–von Mises–Fisher Distribution, with Applications to Multivariate and Relational Data. *J. Comput. Graph. Stat.* **2009**, *18*, 438–456. [CrossRef]
2. Bultmann, S.; Li, K.; Hanebeck, U.D. Stereo Visual SLAM Based on Unscented Dual Quaternion Filtering. In Proceedings of the 22nd International Conference on Information Fusion (Fusion 2019), Ottawa, ON, Canada, 2–5 July 2019.
3. Kok, M.; Schön, T.B. A Fast and Robust Algorithm for Orientation Estimation Using Inertial Sensors. *IEEE Signal Process. Lett.* **2019**, *26*, 1673–1677. [CrossRef]
4. Lunga, D.; Ersoy, O. Unsupervised Classification of Hyperspectral Images on Spherical Manifolds. In Proceedings of the Industrial Conference on Data Mining, Vancouver, BC, Canada, 11–14 December 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 134–146.
5. Marković, I.; Chaumette, F.; Petrović, I. Moving Object Detection, Tracking and Following Using an Omnidirectional Camera on a Mobile Robot. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA 2014), Hong Kong, China, 31 May–7 June 2014.
6. Guan, H.; Smith, W.A. Structure-from-Motion in Spherical Video Using the von Mises–Fisher Distribution. *IEEE Trans. Image Process.* **2016**, *26*, 711–723. [CrossRef] [PubMed]
7. Möls, H.; Li, K.; Hanebeck, U.D. Highly Parallelizable Plane Extraction for Organized Point Clouds Using Spherical Convex Hulls. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA 2020), Paris, France, 31 May–31 August 2020.
8. Straub, J.; Chang, J.; Freifeld, O.; Fisher, J., III. A Dirichlet Process Mixture Model for Spherical Data. In Proceedings of the Artificial Intelligence and Statistics, San Diego, CA, USA, 9–12 May 2015; pp. 930–938.
9. Fisher, R.A. Dispersion on a sphere. *Proc. R. Soc. London. Ser. A Math. Phys. Sci.* **1953**, *217*, 295–305. [CrossRef]
10. Mardia, K.V.; Jupp, P.E. *Directional Statistics*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 494.
11. Ulrich, G. Computer Generation of Distributions on the M-Sphere. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1984**, *33*, 158–163. [CrossRef]
12. Wood, A.T. Simulation of the von Mises–Fisher Distribution. *Commun. -Stat.-Simul. Comput.* **1994**, *23*, 157–164. [CrossRef]
13. Jakob, W. *Numerically Stable Sampling of the von Mises–Fisher Distribution on $\mathbb{S}^2$ (and Other Tricks)*; Technical Report; Interactive Geometry Lab, ETH Zürich: Zürich, Switzerland, 2012.
14. Kurz, G.; Hanebeck, U.D. Stochastic Sampling of the Hyperspherical von Mises–Fisher Distribution Without Rejection Methods. In Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2015), Bonn, Germany, 15–17 October 2015.
15. Kurz, G.; Gilitschenski, I.; Hanebeck, U.D. Unscented von Mises–Fisher Filtering. *IEEE Signal Process. Lett. Apr.* **2016**, *23*, 463–467. [CrossRef]
16. Gilitschenski, I.; Kurz, G.; Hanebeck, U.D. Non-Identity Measurement Models for Orientation Estimation Based on Directional Statistics. In Proceedings of the 18th International Conference on Information Fusion (Fusion 2015), Washington, DC, USA, 6–9 July 2015.

17. Hanebeck, U.D.; Huber, M.F.; Klumpp, V. Dirac Mixture Approximation of Multivariate Gaussian Densities. In Proceedings of the 2009 IEEE Conference on Decision and Control (CDC 2009), Shanghai, China, 15–18 December 2009.

18. Gilitschenski, I.; Hanebeck, U.D. Efficient Deterministic Dirac Mixture Approximation. In Proceedings of the 2013 American Control Conference (ACC 2013), Washington, DC, USA, 17–19 June 2013.

19. Gilitschenski, I.; Steinbring, J.; Hanebeck, U.D.; Simandl, M. Deterministic Dirac Mixture Approximation of Gaussian Mixtures. In Proceedings of the 17th International Conference on Information Fusion (Fusion 2014), Salamanca, Spain, 7–10 July 2014.

20. Kurz, G.; Gilitschenski, I.; Siegwart, R.Y.; Hanebeck, U.D. Methods for Deterministic Approximation of Circular Densities. *J. Adv. Inf. Fusion* **2016**, *11*, 138–156.

21. Collett, D.; Lewis, T. Discriminating Between the von Mises and Wrapped Normal Distributions. *Aust. J. Stat.* **1981**, *23*, 73–79. [CrossRef]

22. Hanebeck, U.D.; Lindquist, A. Moment-based Dirac Mixture Approximation of Circular Densities. In Proceedings of the 19th IFAC World Congress (IFAC 2014), Cape Town, South Africa, 24–29 August 2014.

23. Gilitschenski, I.; Kurz, G.; Hanebeck, U.D.; Siegwart, R. Optimal Quantization of Circular Distributions. In Proceedings of the 19th International Conference on Information Fusion (Fusion 2016), Heidelberg, Germany, 5–8 July 2016.

24. Gilitschenski, I.; Kurz, G.; Julier, S.J.; Hanebeck, U.D. Unscented Orientation Estimation Based on the Bingham Distribution. *IEEE Trans. Autom. Control* **2016**, *61*, 172–177. [CrossRef]

25. Li, K.; Frisch, D.; Noack, B.; Hanebeck, U.D. Geometry-Driven Deterministic Sampling for Nonlinear Bingham Filtering. In Proceedings of the 2019 European Control Conference (ECC 2019), Naples, Italy, 25–28 June 2019.

26. Li, K.; Pfaff, F.; Hanebeck, U.D. Hyperspherical Deterministic Sampling Based on Riemannian Geometry for Improved Nonlinear Bingham Filtering. In Proceedings of the 22nd International Conference on Information Fusion (Fusion 2019), Ottawa, ON, Canada, 2–5 July 2019.

27. Steinbring, J.; Hanebeck, U.D. S2KF: The Smart Sampling Kalman Filter. In Proceedings of the 16th International Conference on Information Fusion (Fusion 2013), Istanbul, Turkey, 9–12 July 2013.

28. Kurz, G.; Gilitschenski, I.; Hanebeck, U.D. Nonlinear Measurement Update for Estimation of Angular Systems Based on Circular Distributions. In Proceedings of the 2014 American Control Conference (ACC 2014), Portland, OR, USA, 4–6 June 2014.

29. Li, K.; Frisch, D.; Radtke, S.; Noack, B.; Hanebeck, U.D. Wavefront Orientation Estimation Based on Progressive Bingham Filtering. In Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2018), Bonn, Germany, 9–11 October 2018.

30. Sra, S. A Short Note on Parameter Approximation for von Mises–Fisher Distributions: And a Fast Implementation of Is(x). *Comput. Stat.* **2012**, *27*, 177–190. [CrossRef]

31. Banerjee, A.; Dhillon, I.S.; Ghosh, J.; Sra, S. Clustering on the Unit Hypersphere Using von Mises–Fisher Distributions. *J. Mach. Learn. Res.* **2005**, *6*, 1345–1382.

32. Kurz, G.; Pfaff, F.; Hanebeck, U.D. Kullback–Leibler Divergence and Moment Matching for Hyperspherical Probability Distributions. In Proceedings of the 19th International Conference on Information Fusion (Fusion 2016), Heidelberg, Germany, 5–8 July 2016.

33. Hauberg, S.; Lauze, F.; Pedersen, K.S. Unscented Kalman Filtering on Riemannian Manifolds. *J. Math. Imaging Vis.* **2013**, *46*, 103–120. [CrossRef]

34. Leopardi, P. A Partition of the Unit Sphere Into Regions of Equal Area and Small Diameter. *Electron. Trans. Numer. Anal.* **2006**, *25*, 309–327.

35. Jeffrey, A.; Dai, H.H. *Handbook of Mathematical Formulas and Integrals*; Elsevier: Amsterdam, The Netherlands, 2008.

36. Bruckner, A.M.; Bruckner, J.B.; Thomson, B.S. *Real Analysis*; Prentice Hall: Upper Saddle River, NJ, USA, 1997.

37. Li, K.; Pfaff, F.; Hanebeck, U.D. Nonlinear von Mises–Fisher Filtering Based on Isotropic Deterministic Sampling. In Proceedings of the 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2020), Karlsruhe, Germany, 14–16 September 2020.

38. Lengyel, E. *Mathematics for 3D Game Programming and Computer Graphics*; Nelson Education: Toronto, ON, Canada, 2012.

39. Chiuso, A.; Picci, G. Visual Tracking of Points as Estimation on the Unit Sphere. In *The Confluence of Vision and Control*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 90–105.

40. Radtke, S.; Li, K.; Noack, B.; Hanebeck, U.D. Comparative Study of Track-to-Track Fusion Methods for Cooperative Tracking with Bearings-only Measurements. In Proceedings of the 2019 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2019), Taipei, Taiwan, 6–9 May 2019.

41. Traa, J.; Smaragdis, P. Multiple Speaker Tracking With the Factorial von Mises–Fisher Filter. In Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Reims, France, 21–24 September 2014.