On-Manifold Recursive Bayesian Estimation for Directional Domains

Kailai Li

Dissertation

Original version available at
https://publikationen.bibliothek.kit.edu/1000142151

Tag der mündlichen Prüfung:	17.12.2021
Erster Gutachter:	Prof. Dr.-Ing. Uwe D. Hanebeck
Zweiter Gutachter:	Prof. Dr. Fredrik Gustafsson

Acknowledgment

This thesis is built upon my research performed at the Intelligent Sensor-Actuator-Systems (ISAS) laboratory, Institute for Anthropomatics and Robotics (IAR), Karlsruhe Institute of Technology (KIT). My first and deepest thanks go to my advisor, Uwe D. Hanebeck, for his consistent trust, guidance and great support over the years, which encourage me to conduct research with a large degree of freedom. I feel grateful to have Fredrik Gustafsson as my co-advisor. Fredrik has long inspired me through his brilliant contributions to information fusion and beyond, and I really look forward to further collaborations with him.

The presented work was mainly funded by the German Research Foundation (DFG) under grant HA 3789/16-1 and later partially funded by the Helmholtz AI Cooperation Unit within the scope of the project "Ubiquitous Spatio-Temporal Learning for Future Mobility" (ULearn4Mobility).

I have many great colleagues. I would like to thank Susanne Radtke for her kindness, also for the joint memory of sharing office and traveling to conferences (and adventures). I appreciate the casual conversations with Daniel Frisch during evenings in the lab. I always feel extremely grateful for the brotherhood with Florian Pfaff, with whom I also developed the research through our fruitful collaborations. Moreover, I want to thank Gerhard Kurz and Igor Gilitschenski for their encouragement and initial technical guidance in this research field, and their attention to my progress. Further, I owe my gratitude to my colleagues Florian Rosenthal, Michael Fennel, Jana Mayer, Antonio Zea, Benjamin Siebler, Marcel Reith-Braun, and Tim Baur. I also want to thank our technician team, Sascha Faber, Achim Langendörfer, Alexander Riffel, and our secretary, Dagmar Gambichler, for supporting my daily work in the lab.

Over the years, I had the opportunity to work with many wonderful students. In particular, I would like to thank Simon Bultmann, Hannes Möls, Meng Li, and Philipp Kern. The collaborations with them have not only delivered innovative work but also deepened my insight to many engineering problems. A special thank to Simon for the solid friendship beyond our joint work.

I feel very lucky to be a member of the KIT long-distance running group – I enjoyed all the training sessions and races with my teammates and trainers. I want to thank Romana Markovic for caring about me as an inspiring and amazing friend.

Finally, I thank my parents for their unconditional and invaluable support. They brought me into the world of poems and literature, which retain my imagination and shape my mind to a significant degree. I am blessed to have them in my life.

Karlsruhe, December 2021 Kailai Li

Contents

Notation

General Conventions

x	scalar
$\underline{x}$	vector
$\boldsymbol{x}$	random variable
x_k	k-th entry of the vector $\underline{x}$
$\underline{x}_t$	vector $\underline{x}$ at time step t
$\underline{x}^\top$	transpose of $\underline{x}$
$a(\cdot)$	function
$\underline{a}(\cdot)$	function with vector-valued output
$\mathbf{A}$	matrix
$\mathbf{0}_{n\times m}$	$n \times m$ zero matrix
$\underline{0}_n$	n-dimensional zero vector
$\mathbf{I}_{n\times n}$	$n \times n$ identity matrix
$\mathrm{E}(\boldsymbol{x})$	expected value of random variable $\boldsymbol{x}$
$\mathrm{Re}(z)$	real part of a complex number z
$\boldsymbol{\Sigma}^{\underline{\boldsymbol{v}}}$	covariance matrix of random vector $\underline{\boldsymbol{v}}$
$\underline{a} \otimes \underline{b}$	product of quaternions $\underline{a}$ and $\underline{b}$
$\underline{a} \boxtimes \underline{b}$	product of dual quaternions $\underline{a}$ and $\underline{b}$
[KS13]	regular citation
[O1]	citation of own work
[S1]	citation of supervised student thesis

Probabilistic Modeling and Filtering on Directional Manifolds

π	circular constant
$\mathcal{I}_v$	modified Bessel function of the first kind and of order v
exp	exponential function
f	probability density
δ	Dirac delta function
$\mathcal{N}$	Gaussian distribution
$\mathcal{B}$	Bingham distribution
$\mathcal{VM}$	von Mises distribution
$\mathcal{VMF}$	von Mises–Fisher distribution
$\mathcal{PB}$	SE(2)-Bingham distribution
$\mathcal{VMM}$	von Mises mixture distribution
$\mathcal{VMFM}$	von Mises–Fisher mixture distribution
$\mathsf{i}, \mathsf{j}, \mathsf{k}$	basic quaternions

d	dimension of the manifold
$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbb{M}$	manifold
$\mathbb{S}^{d-1}$	$(d-1)$-sphere (a subset of $\mathbb{R}^d$)
$\mathbb{B}^d$	d-ball (a subset of $\mathbb{R}^d$, bounded by $\mathbb{S}^{d-1}$)
$\mathbb{DH}_1$	manifold of unit dual quaternions
$\mathbb{T}_{\underline{v}}\mathbb{M}$	tangent space of manifold $\mathbb{M}$ at $\underline{v} \in \mathbb{M}$
$\mathrm{Log}_{\underline{v}}$	logarithm map operated at $\underline{v}$
$\mathrm{Exp}_{\underline{v}}$	exponential map operated at $\underline{v}$
$\mathbb{X}$	sample set consisting of elements denoted as $\underline{x}$
t	time step
$f_t^{\boldsymbol{w}}$	density of the random variable $\boldsymbol{w}_t$
f_t^{e}	estimated density at time step t incorporating measurements up to time step t
f_{t+1}^{p}	predicted density at time step $t+1$ incorporating measurements up to time step t

Abbreviations

BPB	Bingham principal basis
CDS	configurable deterministic sampling
CvMD	Cramér–von Mises distance
DMA	Dirac mixture approximation
DoF	degrees of freedom
DQPF	dual quaternion particle filter
FDS	flexible deterministic sampling
HCvMD	hyperspherical Cramér–von Mises distance
HDMR	hyperspherical Dirac mixture reapproximation
HEP	hyperspherical equal partitioning
HPT	hyperspherical parallel transport
HRDF	hyperspherical reapproximation discrete filter
IDS	isotropic deterministic sampling
IMU	inertial measurement unit
LATS	locally augmented tangent space
LCD	localized cumulative distribution
LiDAR	light detection and ranging
MCP	mode-centric partitioning
MCvMD	on-manifold Cramér–von Mises distance
MDF	on-manifold discrete filter
MDMR	on-manifold Dirac mixture reapproximation
MLE	maximum likelihood estimation
PDF	probability density function
PF	particle filter
QBF	quaternion Bingham filter
QCB	quaternion concomitant basis
R&R	reapproximation and reconstruction
RMSE	root mean squared error
SCH	spherical convex hull
TS	tangent space

UDQPF	unscented dual quaternion particle filter
UKF	unscented Kalman filter
UPF	unscented particle filer
UQBF	unscented quaternion Bingham filter
UT	unscented transform
vMFF	von Mises–Fisher filter

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Etablierung eines Frameworks für probabilistische Modellierung und Filterung direkt auf der Mannigfaltigkeit für Zufallsvariablen auf direktionalen Domänen. Konventionelle Filter, welche die Topologie berücksichtigen, basieren auf den klassischen Annahmen parametrischer Modelle für zirkuläre oder hypersphärische Domänen. Sie nutzen meist rudimentäre Schätzverfahren, die zu Unzulänglichkeiten in praktischen Anwendungen führen können. Daher werden in dieser Arbeit die Konzepte für zirkuläre und hypersphärische Domänen so generalisiert, dass sie auf cartesische Produkte mit euklidischen Räumen angewandt werden können. Dies erlaubt es auch, sie für praktische Anwendungen, wie dem Modellieren und Schätzen von Starrkörperbewegungen, anzuwenden. Im Rahmen der betrachteten direktionalen Domänen wird das Framework wie im Folgenden beschrieben aufgebaut.

Um theoretisch fundierte Modelle und Filter herleiten zu können, beginnen wir mit einer Untersuchung der Geometrie der betrachteten direktionalen Domänen. Einsichten zu hypersphärischen Geometrien werden erst kombiniert mit Verfahren zur Abbildung von hypersphärischen Räumen auf Tangentialräume und zurück, wie beispielsweise Logarithm/Exponential Maps, Orthogonalprojektionen und gnomonische Projektionen/Retraktionen. Für cartesische Produkte von Hypersphären und euklidischen Räumen oder direktionale Domänen mit besonderen topologischen Merkmalen (wie antipodale Symmetrie) werden mehrere geometrische Werkzeuge betrachtet. Hierbei wird der Fokus auf die Mannigfaltigkeiten der Einheitsquaternionen oder dualen Quaternionen gelegt. Diese sind von enormer Wichtigkeit zur Schätzung von Starrkörperbewegungen. Darüber hinaus werden Methoden für den Aufbau/Transport lokaler Basen, die Tangentialräume auf nichtlinearen Mannigfaltigkeiten aufspannen, vorgestellt, um eine Interpretation der Unsicherheit in den Tangentialraum zu ermöglichen.

Anhand der gewonnenen Erkenntnisse erweitern wir zunächst die parametrischen direktionalen Modelle und Filter. Als Beispiel für eine Domäne, die sich aus einer direktionalen und einer anderen Domäne zusammensetzt, wird die Mannigfaltigkeit der dualen Quaternionen gewählt. Hierfür wird ein Ansatz basierend auf Paralleltransport mit Bingham-Gaussians umgesetzt. Neben der Modellierung der hypersphärischen Komponente mit Binghamverteilungen und der linearen Komponente mit Gaußverteilungen werden Korrelationen zwischen den Komponenten durch hypersphärischen Paralleltransport berücksichtigt. Zur Erhöhung der Güte von parametrischen direktionalen Filtern werden zunächst neue Ansätze zum deterministischen Sampling mit unterschiedlichen Stichprobengrößen vorgestellt. Hierbei werden die ersten beiden Momente präzise berücksichtigt, höhere Momente jedoch nur approximiert. Für praktische Anwendungen wird die konkrete Methodik für Bingham-Verteilungen und von Mises–Fisher-Verteilungen vorgestellt. Des Weiteren werden deterministische progressive Updateschritte vorgestellt, um starke Nichtlinearitäten im Systemmodell oder Messmodelle, die nicht dem Identitätsmodell entsprechen, berücksichtigen zu können. Simulationen zeigen, dass durch Einbeziehung des einstellbaren deterministischen Samplings in den progressiven Filter in nichtlinearen direktionalen Schätzproblemen eine deutlich höhere Genauigkeit erzielt werden kann.

Darüber hinaus werden nichtparametrische Modelle und Filter basierend auf Dirac Mischverteilungen auf systematische Weise hergeleitet. Für Hypersphären, die stets kompakt und beschränkt sind, werden neue Diskretisierungsstrategien vorgestellt, um gleichmäßige oder an den Moden der Verteilung zentrierte hypersphärische Grids zu erhalten. Erstere erlauben eine feste, gleichmäßige Auflösung, wohingegen Letztere eine genauere Approximation in der Nähe der Mode erlauben und sich der Dispersion im Falle von antipodaler Symmetrie anpassen können. Auch wird ein generischer Ansatz für das diskrete Filtern auf der Mannigfaltigkeit vorstellt. Es wird dargelegt, wie das diskrete Modell basierend auf dem modenzentrierten Grid zum nichtlinearen Schätzen von Quaternionen darin integriert werden kann. Um die Anwendbarkeit von sequenziellen Monte-Carlo-Methoden für nichtparametrische direktionale Schätzung zu verbessern, wird ein Unscented Particle Filter eingeführt. Bei diesem wird für jedes Partikel ein on-manifold Unscented Kalman Filter verwendet, um die Proposal Density zu finden, welche neue Observationen berücksichtigt. Im Gegensatz zu einfachen Partikel Filtern, welche die Transitionsdichte als Proposal Density verwenden, können mit dem neuen Verfahren bessere Ergebnisse unter schwierigen Bedingungen, wie starker Nichtlinearität, Nichtstationarität oder konzentrierten Likelihoods, erzielt werden.

Zu guter Letzt wird ein einheitliches Konzept für die Reapproximation von Dirac Mischverteilungen vorgestellt. Für eine gegebene Dirac Mischverteilung basierend auf vielen empirisch gewonnenen Samples wird das Konzept auf die unterliegende Geometrie der Mannigfaltigkeit angepasst, um so eine Dirac Mischdichte mit weniger und anpassbarer Anzahl an Komponenten zu erhalten, die auch die Dispersion berücksichtigt. Basierend auf Beispielen für den Kreis, die Hypersphäre und die Mannigfaltigkeite der dualen Quaternionen, erlaubt das neue Konzept eine effizientere diskrete Darstellung als Monte-Carlo-Methoden oder Methoden basierend auf Grids. Für den hypersphärischen Fall wird eine Reapproximations- und Rekonstruktionsmethode basierend auf von Mises–Fisher Mixturen vorgestellt, die für unbekannte zugrundeliegende kontinuierliche Dichten anwendbar ist. Darüber hinaus wird das vorgestellte Reapproximationsverfahren in den generischen diskreten Filteransatz integriert, um auf direktionalen Domänen ohne Vorliegen konkreter Rauschverteilungen nichtparametrisch filtern zu können. Simulationen zeigen, dass dieser Ansatz parametrischen Filtern und Filtern basierend auf Monte-Carlo-Methoden überlegen ist.

Der Hauptbeitrag dieser Arbeit ist die systematische Entwickelung eines theoretischen Frameworks zur Bayes'schen Schätzung unter Berücksichtigung der Topologie. Sie liefert eine theoretisch fundierte und solide Basis für die Anwendung von geomeriegetriebenen Methoden zur Quantifizierung und Behandlung von Unsicherheiten in vielen technischen Anwendungsfällen.

Abstract

We establish a methodological framework for on-manifold probabilistic modeling and filtering of random variables in directional domains. Conventional topology-aware directional filters still adhere to the classical assumptions of parametric models defined in circular or hyperspherical domains, and are mostly equipped with rudimentary estimation techniques, which may suffer from multiple issues in the face of requirements close to engineering practices. In this thesis, we first extend the traditional scope of directional domains from circle/hyperspheres to a generalized concept that also allows their products with Euclidean spaces. This further allows addressing relevant practical application scenarios, such as modeling and estimating uncertain rigid body motions. Within the scope of the considered directional domains, the proposed framework is built up as follows.

As a preparation for establishing theoretically sound modeling and filtering methods, we start off with providing a geometric investigation on the considered directional manifolds. Insights to hyperspherical geometry is first brought together with approaches to bridging the nonlinear manifold to a hyperspherical tangent space including the logarithm/exponential maps, orthographic and gnomonic projections/retractions. For hyperspheres in conjunction with Euclidean spaces or those with additional topological structures (such as antipodal symmetry), we elaborate several geometric tools with showcases on the unit quaternion and unit dual quaternion manifolds, which are of vital importance for spatial transformation estimation. Moreover, methods are provided for establishing/transporting local bases spanning tangent spaces on nonlinear manifolds for enabling in-tangent-space uncertainty interpretation.

Based on the geometric insights to directional manifolds, we first focus on upgrading the parametric directional modeling and filtering methods. For extending topology-aware parametric models in composite directional domains, we select the unit dual quaternion manifold for demonstration and propose the parallel transport-based Bingham–Gaussian modeling approach. Besides the individual modeling of the hyperspherical and linear components using the Bingham and Gaussian distributions, respectively, probabilistic correlation between the components is interpreted by expressing the linear model w.r.t. the dispersion on the hypersphere via parallel transport. For enhancing the performance of parametric directional filters, we first propose novel sampling schemes to allow configurable sizes of deterministic samples under the scheme of unscented transform, namely, by preserving moments up to the second order of the underlying distribution, meanwhile approximating its higher-order shape information. For practice, the methodology is concretized for the Bingham distribution and the von Mises–Fisher distribution. Further, we introduce the deterministic progressive update step for handling estimation with strong nonlinearities and non-identity measurement models. Simulations show that incorporating the configurable deterministic sampling into the progressive fusion scheme can deliver considerable performance enhancements for nonlinear directional estimation tasks.

Moreover, Dirac mixture-based nonparametric modeling and filtering methods are established in a systematic fashion. For unit hyperspheres, which are compact and bounded, we propose discretization strategies for generating quasi-equidistant/mode-centric hyperspherical grid points.

The former enables a uniform and fixed resolution, while the latter provides a more detailed approximation around the mode and its grid topology enables inherent coherence to dispersions of antipodal symmetry. We further provide a generic design of on-manifold discrete filtering and showcase how to integrate the mode-centric grid-based discrete model into it for nonlinear quaternion estimation. In order to improve the utility of sequential Monte Carlo methods for nonparametric directional estimation, we establish the unscented particle filtering scheme with each particle running an individual on-manifold unscented Kalman filter to obtain a proposal incorporating the recent observation. Compared with the plain particle filter with transition density being the proposal, the new scheme shows superior performance even under challenging conditions such as strong nonlinearities, non-stationary systems or peaky likelihoods.

Last but not least, for enabling efficient discrete modeling, we propose a unified paradigm for on-manifold Dirac mixture reapproximation. Given a Dirac mixture with many components located at directional samples that are collected from empirical data, the novel paradigm is then customized according to the underlying manifold geometry to produce a target Dirac mixture with fewer and configurable components of dispersion-adaptive layout. Based on showcases with customization on the unit circle, unit hyperspheres and unit dual quaternion manifolds, the proposed paradigm enables more efficient discrete modeling than grid-based or Monte Carlo-based methods. Built upon its hyperspherical variant, a reapproximation & reconstruction procedure is proposed for continuous probabilistic modeling of unknown underlying distribution in the form of von Mises–Fisher mixtures. Further, we integrate the proposed reapproximation paradigm into the generic discrete filtering scheme for nonparametric directional estimation with unknown form of noise distribution. Evaluations on nonlinear spherical estimation manifest its leading performance in comparison with Monte Carlo-based or parametric filters.

The major technical content of this thesis systematically drafts out a theoretical framework for topology-aware recursive Bayesian estimation. It provides a theoretically sound and solid basis for applying geometry-driven methodologies to quantifying and tackling uncertainties in extensive engineering practices.

CHAPTER 1

Introduction

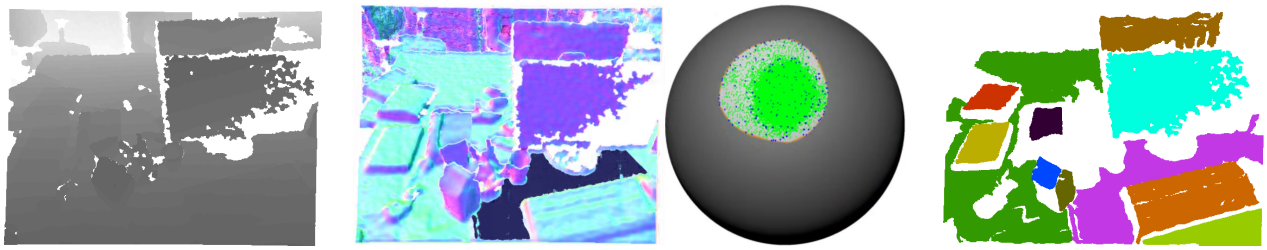
Uncertainty quantification and state estimation of random variables in directional domains play important roles in ubiquitous application scenarios, in particular, robotic perception, navigation and control, computer vision and graphics, and data mining, among others [1–7]. Unlike Euclidean spaces, directional manifolds, e.g., the unit hyperspheres[1], exhibit nonlinear and periodic topological structure, to which dispersion of random variables are thereon confined. Conventional schemes of recursive estimation for directional domains often rely on local linearization at a given state vector, e.g., via Lie algebra or Riemannian geometry, inducing an Euclidean space (namely, a tangent plane), where the uncertainty underlying on the manifold is modeled by a Gaussian distribution. Based thereon, classical filtering approaches, e.g., the Kalman filter and its variants such as the unscented Kalman filter (UKF) [8], can be deployed to the linearized space with proper modifications [9–11]. However, in theory, such a paradigm produces a warped dispersion upon interpreting the on-tangent-plane Gaussian distribution back on the manifold (via the exponential map). To alleviate this artifact in practice, the assumption of small perturbation is often imposed, albeit with the risk of being violated by fast system transition or large uncertainty.

Recently, enormous effort has been devoted to developing theoretically sound approaches to recursive estimation for directional domains. A major advance, inter alia, has been made in applying directional statistics [12] to probabilistic modeling and recursive estimation of random variables with directional components [13]. Defined on nonlinear manifolds, distributions from directional statistics inherently model the uncertainty in coherence with the underlying manifold geometry, thereby eliminating the approximation artifacts induced by the linearization.

Within the scope of manifold-adaptive directional estimation, promising work has been pioneered in various disciplines ranging from robotics to bioinformatics [14–17]. However, existing techniques are typically application- and domain-specific, and one still faces various challenges with current rudimentary designs. From the generic viewpoint of data fusion involving directional quantities, there still remains a considerable theoretical gap between the previous state of the art[2] and a systematic architecture of topology-aware probabilistic modeling and filtering.

In this thesis, we aim to establish a methodological framework for on-manifold recursive Bayesian estimation. Though highly robotics-related as later shown by the applications in Sec. 1.1, the proposed approaches provide new and generically applicable "geometric" insights to data fusion problems involving directional quantities, inducing a theoretically sound basis for extensive engineering practices.

[1] For conciseness, the term "hypersphere" denotes a sphere of any dimension when specifying the term is not needed.
[2] As of the time before the contributions in this thesis were made in own publications.



(A) input depth image (B) region growing in normal map transferred to $\mathbb{S}^2$ (C) segmented planes

Figure 1.1.: Plane extraction of real-world scene [20] transferred to clustering normal vectors via SCH-based region growing (SCH-RG) [19].

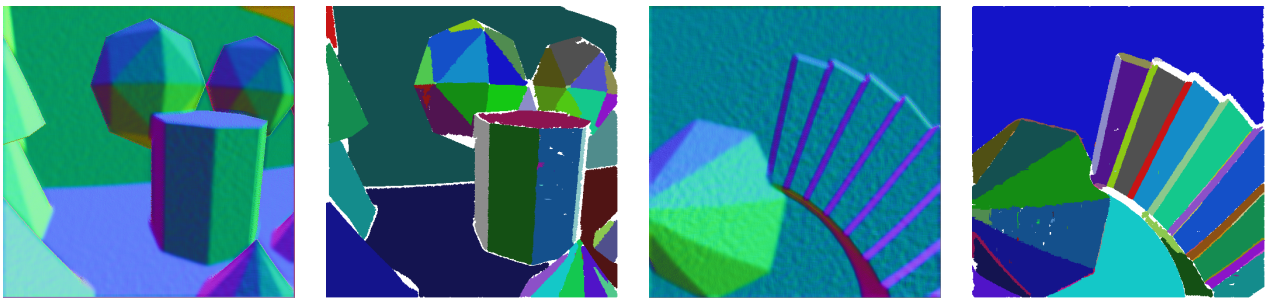
In the remainder of this chapter, a few application scenarios incorporating uncertain directional quantities are first introduced in Sec. 1.1 to elaborate representative topological spaces which are of high interest to engineering. Afterward in Sec. 1.2, we scrutinize the major limitations of past state-of-the-art directional filtering theories, on which the presented thesis focuses, thereby delivering our major contributions. Further, a brief overview of the related work is provided in Sec. 1.3. Finally, a sketch of the thesis structure is outlined in Sec. 1.4.

1.1 Random Variables Composing Directional Quantities

In the sense of statistics, directional manifolds refer to the topological spaces defined by unit vectors in the Euclidean spaces (with extensions to rotation matrices) [12, 18]. We show the following application scenarios from the own publications of the author to concretize the fundamental roles played by tackling (uncertain) directional variables in reality.

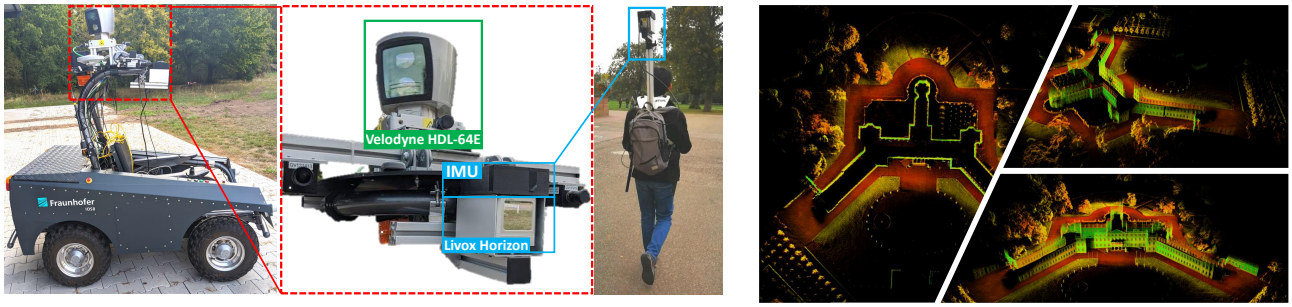
Transferred primitive recognition on the unit sphere: In man-made environments, extracting geometric primitives such as planes can largely raise the performance of scene representation and understanding for intelligent mobile agents. Modern visual sensory modalities often provide depth information in the format of organized point cloud (e.g., using RGB-D cameras), from which the pixelwise surface normals can be easily obtained. Most existing approaches to plane extractions exploit explicit plane parameterization for optimizing a certain point-to-plane metric, which is hardly feasible for parallel processing on graphics hardware and requires downsampling for real-time performance. A normal map can be transferred onto the unit sphere $\mathbb{S}^2$, with each spherical point representing the pixelwise surface normal vector. In [19], we introduced a new concept, the spherical convex hull (SCH), for implicit and nonparametric plane representation. As shown in Fig. 1.1, normals are clustered on the unit sphere via expanding SCHs to find the maximal pairwise orientation differences among spherical points along with region growing in the normal map. The whole plane extraction scheme inherently realizes parallelism at both algorithmic and implementationary levels, which is mostly enabled by the mechanism of comparing orientations w.r.t. the arc length on $\mathbb{S}^2$. Fig. 1.2 further illustrates examples of extracting planes given two synthesized normal maps. Deployed on common embedded GPUs, the approach guarantees state-of-the-art segmentation accuracy and real-time processing in full resolution for typical RGB-D images. This specific application underlines the benefits of tackling direction-related problems on directional manifolds respecting the underlying topological structure.

Rigid body motion estimation and robotic perception: Accurate and robust state estimation, in particular of the rigid body motions, is crucial for realizing high-performance autonomy of



normals → planes normals → planes

Figure 1.2.: Plane extraction results of two synthetic scenes from [21] given by SCH-RG [19].

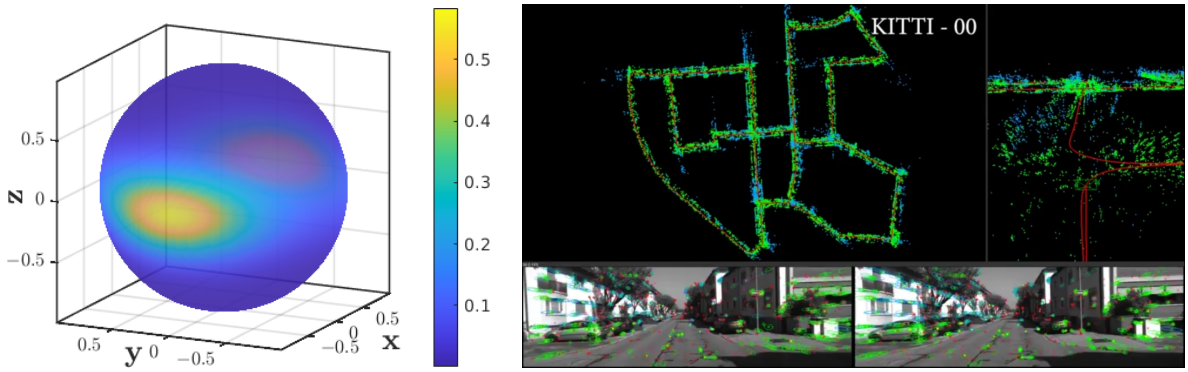


(A) LiDAR-inertial suites onboard different mobile platforms (B) reconstructed 3D map

Figure 1.3.: LiDAR-inertial fusion for rigid body motion estimation and scene reconstruction [25].

mobile robots [1]. Mathematically speaking, the six-DoF rigid body motions belong to the special Euclidean group SE(3), which inherently comprises nonlinear topological structure due to the rotation component belonging to the special orthogonal group SO(3). Minimal representations such as Euler angles plus translations suffer from issues of discontinuities or singularities (e.g., the issue of "gimbal lock"). Such issues can be eliminated through overparameterization, for instance, by using the well-known 4×4 transformation matrices. But they introduce a large degree of redundancy, inducing numerical instabilities and memory inefficiencies. Quaternions of unit norm can be formulated into four-dimensional vectors confined to the unit hypersphere $\mathbb{S}^3 \subset \mathbb{R}^4$. With only one degree of redundancy, they parameterize spatial rotations without ambiguities and enable concise kinematic formulations incorporating measurements close to the sensory hardware such as magnetometers and inertial measurement units (IMUs). Therefore, they are widely used for spatial transformation representation (together with translation vectors) in sensor fusion and control tasks [22–24]. For instance, we developed a tightly-coupled LiDAR-inertial odometry and mapping system in [25] for robotic mobile perception. There, a keyframe-based sensor fusion scheme estimates rigid body motions (parameterized by unit quaternions and translation vectors) by directly fusing the extracted LiDAR features and preintegrated IMU measurements via sliding window optimization. As shown in Fig. 1.3-(A), the system can be deployed to different mobile platforms and is compatible for both conventional (HDL-64E) and solid-state (Livox Horizon) LiDARs. It delivers real-time efficiency and superior tracking accuracy over existing competitive systems (as of the date of the publication). A showcase of using the system for high-quality 3D reconstruction of Schloss Karlsruhe is given in Fig. 1.3-(B).

Based on the quaternion representation and the dual number theory, dual quaternions are further introduced for parameterizing six-DoF rigid body motions [26]. Formulated into eight-dimensional



(A) a Bingham distribution on $\mathbb{S}^2$

(B) stereo visual odometry and mapping on [3]

Figure 1.4.: Dual quaternion-based recursive SE(3) estimation using Bingham distributions on $\mathbb{S}^3$ [31].

vectors (essentially composed of two quaternions, the real part and the dual part, respectively), they represent the spatial rotation and translation simultaneously without ambiguities, inducing only two degrees of redundancy and enabling similar kinematic formulations as those of quaternions [27]. As alternatives to the transformation matrices, they are commonly deployed to applications in computer graphics, navigation, and robotics [28–30]. In [31], we presented a stereo visual odometry system using an unscented transform-based filtering method [8], with dual quaternions representing the six-DoF egomotion. In order to model the uncertainty of the rotation components (i.e., quaternions) of the dual quaternion states, the Bingham distribution on the unit hypersphere $\mathbb{S}^3$ is exploited. For the sake of visualization, we plot the distribution on $\mathbb{S}^2$ in Fig. 1.4-(A). Since an arbitrary quaternion $\underline{x} \in \mathbb{S}^3$ and its antipode $-\underline{x} \in \mathbb{S}^3$ represent the same rotation, the Bingham distribution is a natural fit for probabilistic modeling given its dispersion of antipodal symmetry [32]. Fig. 1.4-(B) provides a snapshot of running the system on the 00 sequence of KITTI data set [3]. Given consecutive stereo images, the proposed dual quaternion-based recursive filter is effective and delivers state-of-the-art tracking performance. Moreover, the use of Bingham distributions provides a theoretically sound measure to quantify the uncertainty of the rotation estimate. A detailed introduction of unit quaternions and unit dual quaternions will be given in Sec. 3.2 and Sec. 3.3, respectively.

Short summary: As introduced in the aforementioned examples, directional quantities occur ubiquitously in a variety of applications. Choosing the proper representation for a directional quantity may drastically reformulate the problem, and thus provides better perspectives and solutions for tackling complex tasks. Furthermore, directional variables are often coupled with other directional or non-directional components in a covariant manner, inducing products of manifolds of more complex geometric structure (e.g., the manifold of dual quaternions).

In this thesis, we extend the scope of directional domains to a more generic sense by also considering their Cartesian products with Euclidean spaces. Consequently, recursive Bayesian estimations in the *generalized* directional domains refer not only to the classical settings w.r.t. parametric directional statistics [13] but also to more challenging scenarios of on-manifold probabilistic modeling and filtering in the face of extensive engineering practices.

1.2 Beyond Limitations: Considered Problems and Contributions

The major focus of this thesis is to establish a methodological framework for theoretically sound probabilistic modeling and filtering in the generalized directional domains[3]. For showcasing and evaluating the proposed techniques, we exploit unit hyperspheres and their products with Euclidean spaces (e.g., the unit dual quaternion manifold). More specifically, we elaborate the major considered problems and underline the corresponding contributions in the following sections.

1.2.1 Parametric Probabilistic Modeling for Composite Directional Domains

Limitation: Classical directional statistics and estimation methods have no consideration on the manifolds composing both circle/hyperspheres and Euclidean spaces. As a representative example, there existed in the past no parametric distribution for modeling uncertain dual quaternions adaptively to the underlying topological structure. Particularly, probabilistic modeling on composite directional manifolds requires not only handling the uncertainty in each individual component but also the correlations between the components. As for dual quaternions representing six-DoF rigid body motions, such a probabilistic correlation established between the component domains is essential as it reflects how uncertain rotation and translation terms covary with each other.

Contribution: We upgrade parametric probabilistic modeling from hyperspheres to composite directional manifolds incorporating both hyperspherical and Euclidean spaces with a showcase on the manifold of unit dual quaternions representing SE(3) states [26]. For that, we first provide an in-depth and systematic investigation on topological structures of the unit quaternion and unit dual quaternion manifolds. Based thereon, a novel on-manifold parametric method is proposed for modeling uncertain unit dual quaternions in coherence with the underlying manifold geometry. A Bingham distribution and a Gaussian distribution are exploited to model uncertainties on the hyperspherical and linear component domains, respectively, meanwhile the probabilistic correlations between them are inherently considered w.r.t. parallel-transported Bingham principal directions. Afterward, corresponding deterministic sampling and parameter fitting schemes are also provided and validated for further deployment to recursive estimation.

1.2.2 Enhancing Parametric Directional Filters for Nonlinear Estimation

Limitation: Applying parametric models from directional statistics to state estimation requires the facilitation of samples. As the random sampling scheme cannot reproduce the results and lacks efficiency, deterministic sampling approaches following the idea of unscented transform (UT) have been proposed [32, 33]. However, the basic UT-based sampling scheme only produces a limited number of samples that preserve moments of the underlying distribution up to the second order, deteriorating the estimation performance under strong nonlinearities or peaky likelihoods. Furthermore, past state-of-the-art parametric directional filters are mostly designed for identity measurement models – measurements are confined to the state domain and to following the form of the state distribution. Given arbitrary sensory modalities, adaptations to the identity setup induce additional approximation error and effort for preprocessing [32, 34]. In particular, such issues are highlighted in high-dimensional cases (e.g., hyperspheres and their products with Euclidean

[3] For conciseness in the remainder of this thesis, we use the term "directional domains" to denote both unit circle/hyperspheres and their products with Euclidean spaces for unspecified cases.

spaces), leading to theoretical restrictions of their applicability in practice (e.g., for estimating spatial transformations).

Contribution: We introduce novel deterministic sampling approaches that enable configurable sample sizes for two commonly used hyperspherical distributions, namely, the Bingham distribution [35, 36] and the von Mises–Fisher distribution [37, 38]. Based on hyperspherical geometry, samples are drawn to guarantee the unscented transform (preserving moments up to the second order) while approximating higher-order shape information of the underlying distribution. For nonlinear directional filtering using non-identity measurement models, we further introduce the deterministic progressive update scheme for stepwise measurement fusion adapted to the difference among the sample likelihoods [38–40]. Applying the two enhancements in conjunction considerably improves the applicability and performance of parametric directional filters for nonlinear estimation tasks.

1.2.3 Topology-Aware Nonparametric Probabilistic Modeling and Filtering

Limitation: Parametric models impose assumptions on the arbitrary nature of uncertainties, yet the past state of art has hardly given systematic consideration on nonparametric modeling and filtering for directional domains. One preliminary trial was briefly made on discrete recursive estimation based on discretized density in the circular domain [41]. The sequential Monte Carlo methods are in general applicable for modeling arbitrary directional uncertainties, but a large number of particles are typically desired due to degeneracy issues caused by high-dimensional state spaces and strong nonlinearities (which is common for estimation on hyperspheres or composite directional manifolds). Furthermore, a particle filter typically employs the transition prior as the proposal, and thus disregards the newly observed evidence from the measurement [42]. This induces deteriorated performance in challenging estimation scenarios such as those with non-stationary models, heavy-tailed distributions and peaky likelihoods.

Contribution: We provide novel discrete modeling approaches based on hyperspherical grid points of quasi-equidistant or mode-centric layouts [43]. The former has a uniform discretization resolution and the latter enables a more detailed approximation around the mode of the underlying distribution with inherent consideration on its potential antipodal symmetry. We further introduce a generic discrete filtering scheme using Dirac mixtures of deterministic supports and customize it w.r.t. the mode-centric grid for quaternion filtering. This pioneers a series of followup work for grid-based spherical, toroidal and rigid body motion estimation [44–48]. Furthermore, we establish the on-manifold unscented particle filtering scheme with showcases on unit quaternion and unit dual quaternion manifolds [49, 50]. By running a novel on-manifold UKF particlewise, the scheme incorporates recently observed evidence into estimating the proposals. Evaluations under challenging conditions of directional estimation show its evidently superior performance over the parametric and basic Monte Carlo-based directional filters.

1.2.4 Efficient Discrete Probabilistic Modeling on Directional Manifolds

Limitation: Exploiting random particles is an intuitive solution to nonparametric modeling, and the proposed particlewise on-manifold UKF leads to a considerable performance boost for filtering. However, the innate characteristics of randomness inherently prohibits efficient approximations of arbitrary densities. For propagating samples through the system dynamics, a given transition

density is required, which is often derived based on the system noise represented by a parametric model. Given numerous on-manifold samples of empirical data collected from a stochastic process, approximating them using a parametric model may lead to a loss of useful information. A Dirac mixture located on raw samples is information-preserving, but it lacks efficiency for discrete probabilistic modeling and filtering.

Contribution: We propose a novel on-manifold Dirac mixture reapproximation paradigm for efficient discrete modeling in directional domains [51, 52]. Given a source Dirac mixture with components located at many raw samples from empirical data, it produces a target Dirac mixture with a configurable number of components in a layout that is adapted to the shape of the underlying unknown dispersion. For demonstration, we customize the paradigm to the unit circle, unit hyperspheres, and manifolds of dual quaternions. Based on the customized variant for unit hyperspheres, a two-stage reapproximation & reconstruction procedure is introduced for a followup continuous modeling of the underlying unknown distribution in the form of von Mises–Fisher mixtures. Furthermore, we integrate the proposed reapproximation paradigm into the generic discrete filtering scheme and provide the on-manifold reapproximation discrete filter for nonparametric directional estimation without given form of system noise distribution. Evaluations for nonlinear spherical estimation show its leading performance in comparison with Monte Carlo-based or parametric directional filters.

1.3 Related Work

In this section, we provide a brief retrospect of related work given the past state of the art within the scope of the considered problems in this thesis.

1.3.1 Parametric Directional Estimation

Conventional directional estimation approaches rely on parametric models. Commonly used circular distributions, among others, are the wrapped normal distribution, the von Mises distribution, and the Bingham distribution. Popular ones for hyperspherical filtering include the von Mises–Fisher distribution, the Bingham distribution, and the Watson distribution. Dispersion of the Bingham distribution and the Watson distribution exhibits antipodal symmetry, with the one of the latter being isotropic. Due to its versatility, the Bingham distribution is more of interest for filtering applications, and the one on the unit hypersphere $\mathbb{S}^3$ is inherently appropriate for modeling uncertain quaternions representing SO(3) states. For composite directional manifolds, we only consider the products of circle/hyperspheres with Euclidean spaces such as the manifold of unit dual quaternions. Distributions for other common composite directional domains, such as the tori/hypertori and cylinders, are out of the scope of this thesis – related work can be found in [12, 13, 53, 53, 54]. As briefly mentioned earlier in Sec. 1.2.1 and Sec. 1.2.2, we elaborate related parametric modeling and filtering schemes for the considered directional domains as follows.

Parametric modeling for composite directional domains: In [55], a novel distribution was proposed to model uncertain planar motions represented by dual quaternions in the form of four-dimensional vectors. Parameterized by a 4×4 matrix, the distribution inherently considers the antipodal symmetry of dual quaternions representing SE(2) states. Essentially, the distribution models the real part with a circular Bingham distribution, conditioned on which the dual part is modeled by a two-dimensional Gaussian distribution (a detailed introduction follows in Sec. 2.2.3).

Based thereon, an unscented transform-based SE(2) estimator was proposed in [56] with an identity setting of the measurement model. For modeling uncertain dual quaternions representing SE(3) states, a heuristic method was given in [31], where the incorporated rotation quaternion and translation term are modeled separately by a Bingham distribution on $\mathbb{S}^3$ and a Gaussian distribution in $\mathbb{R}^3$, respectively, without tackling the probabilistic correlation between them.

Deterministic sampling: The deterministic sampling scheme has become a default tool for facilitating system propagation (prediction step) or likelihood reweighting (update step) of parametric directional filtering. Samples are drawn adaptively to the shape of the underlying distribution by at least preserving its first and second (trigonometric) moments, inducing a series of "unscented" directional filters (following the idea of unscented transform used in the UKF). The past existing deterministic sampling methods for the von Mises–Fisher and Bingham distributions generate only $2 \cdot d - 1$ samples on $\mathbb{S}^{d-1}$ as given in [32,34], which do not include shape information of orders higher than two. On the unit circle $\mathbb{S}^1$, a basic version was proposed for drawing five deterministic samples preserving the first two trigonometric moments of typical circular distributions in closed form [57]. Meanwhile, effort has also been dedicated to drawing deterministic samples of larger sizes. By combining scaled sample sets given by the basic UT-based version via superposition, shapes of the circular densities can be better approximated while maintaining the first two trigonometric moments [58]. In [59], arbitrary numbers of samples are drawn by minimizing a squared integral distance to the underlying distribution under the constraint of circular moments. In [60], deterministic samples of configurable sizes are drawn via Voronoi quantization on circular distributions. As of the time before our own publications, there has been no published methods of deterministic sampling that allow freely configurable sample sizes in the sense of unscented transform for commonly used hyperspherical distributions. Corresponding parametric filters thus suffered from deteriorated performance in nonlinear estimation tasks due to the sample degeneracy issue.

Progressive filtering: As presented in Sec. 1.2.2, the past existing unscented hyperspherical filtering methods (e.g., the unscented von Mises–Fisher or Bingham filters) only allow for measurement models of identity form with the system noise following the same distribution as the state. Given a non-identity measurement model, simply reweighting the deterministic samples using the likelihood in the update step is prone to degeneracy. The progressive measurement update scheme aims to confine the ratio of sample weights to a pregiven threshold by splitting the likelihood into a product of likelihoods with adaptively determined exponents. The scheme has been originally proposed and well established for nonlinear filtering in Euclidean spaces [61, 62]. In the past, extensions to directional domains were only proposed for the unit circle [33] to tackle nonlinear angular estimation with non-identity measurement models [33]. For filtering on hyperspheres or composite directional domains with non-identity models, previous considerations focused on reformulating the measurement model into the common identity setting, leading to additional effort and approximation errors [55, 56, 63].

1.3.2 Nonparametric Directional Estimation

For the considered directional domains, there was no systematic investigation on nonparametric probabilistic modeling and estimation in the past state of the art. Within the scope of this thesis, we mainly focus on developing Dirac mixture-based discrete models and corresponding filtering techniques. In this regard, a few relevant methods are listed as follows.

Sequential Monte Carlo methods: Often exploited as the baseline, the sequential Monte Carlo methods, i.e., the particle filters, are a straightforward solution to achieving nonparametric recursive estimation on directional manifolds. Theoretically, it enables complete modeling of arbitrary uncertainties in stochastic processes, provided that sufficiently high numbers of particles are deployed. In practice, however, large quantities of particles are desired for nonlinear estimation on directional manifolds that are typically high-dimensional (e.g., for quaternion or dual quaternion estimation). To mitigate the limitation of the plain particle filtering scheme as mentioned in Sec. 1.2.3, the unscented particle filter (UPF) [64] has been proposed for state estimation in Euclidean spaces. Though considerable improvements were achieved for recursive estimation in challenging scenarios, the method cannot be trivially applied to the considered directional domains due to the underlying nonlinear topology.

Grid-based discrete Bayesian estimation: For directional domains that are compact and bounded (e.g., circle and hyperspheres), it is intuitive to discretize the state space with grid points of a certain layout that allows for approximating distributions of arbitrary shape. Exploiting grids for recursive estimation has been pioneered by the well-known Wonham filter [65] for Euclidean spaces. In the context of angular filtering, work has been done in [66], where circular distributions are discretized by either piecewise constant densities or Dirac mixtures w.r.t. the grid points of uniform interval on $\mathbb{S}^1$. An extension to recursive SE(2) estimation was further built thereupon via Rao–Blackwellization in [41]. Though the fundamental design of grid-based discrete filtering was established previously, adapting it to hyperspheres is not straightforward. An efficient grid generation method on unit hyperspheres has been introduced in [67], but the (quasi-)equidistant grid layout does not lead to memory efficiency when approximating densities of high concentration or with additional topological structures such as the antipodal symmetry. In [68], a novel recursive tesseract subdivision scheme was proposed for discretizing the SO(3) state space based on the quaternion representation. The approach resembles the former equal partitioning method [67] in the sense of uniform subdivision and additionally considers the antipodal symmetry underlying on the rotation group. For both of the discretization methods proposed for directional state space, there existed in the past no corresponding filtering algorithm.

Dirac mixture approximation: For certain directional manifolds that have unbounded components, a complete grid-covering of the entire space is infeasible to achieve. Also, grid-based discrete modeling may suffer from inefficiency as grid points of a fixed resolution and layout cannot adapt to the shape of the underlying dispersion. Monte Carlo-based schemes may enable complete and dispersion-adaptive discrete modeling, but they often contain redundant information due to the randomness in particles. Thus, deterministic and dispersion-aware discrete models are appealing for efficient nonparametric modeling and filtering. In order to achieve discrete modeling with components of deterministic locations, effort has been first dedicated to Dirac mixture approximation of parametric distributions in Euclidean spaces. In [69,70], the Cramér–von Mises distance (CvMD) was generalized w.r.t. multivariate distributions by extending the classical concept of cumulative distribution to the so-called localized cumulative distribution (LCD). A Dirac mixture is obtained by matching an underlying Gaussian (mixture) distribution in the sense of least CvMD, with its layout adapted to the shape of the underlying continuous density. In [71], the same methodology was modified for optimal sample reduction. Though only proposed for Euclidean spaces, the approach can potentially prune redundant information due to the randomness in empirical data and be extended to the considered directional domains.

1.4 Outline

The remainder of this thesis is structured as follows.

In Chapter 2, we first provide preliminaries about recursive Bayesian estimation in Sec. 2.1. Afterward, typical parametric directional distributions (the von Mises–Fisher distribution, the Bingham distribution and the SE(2)-Bingham distribution) are introduced in Sec. 2.2.1 to Sec. 2.2.3, based on which a generic description of parametric directional filters is given in Sec. 2.2.4.

In Chapter 3, we provide a geometric investigation on the considered directional domains. As one of the most fundamental directional manifolds, the unit hypersphere is first studied with several important geometric tools introduced in Sec. 3.1. Based thereon, the manifolds of unit quaternions and unit dual quaternions are introduced in Sec. 3.2 and Sec. 3.3, respectively, with a focus on their geometric structures and engineering backgrounds (e.g., for representing rigid body motions).

In Chapter 4, methodologies of upgrading parametric modeling and filtering on directional manifolds are introduced. To underline the contribution mentioned in Sec. 1.2.1, we showcase a parametric modeling method for composite directional domains on the manifold of unit dual quaternions in Sec. 4.1. Afterward, two methodologies are introduced for enhancing parametric directional filtering corresponding to the contribution claimed in Sec. 1.2.2. First, we propose a novel scheme for configurable deterministic sampling in Sec. 4.2 with practices on the Bingham and von Mises–Fisher distributions in Sec. 4.2.1 and Sec. 4.2.2, respectively. Second, a progressive measurement update scheme is introduced in Sec. 4.3 for handling nonlinear parametric directional filtering with non-identity measurement models. Further, the two enhancements are validated by a case study for nonlinear SO(3) estimation in Sec. 4.4.

In Chapter 5, nonparametric probabilistic modeling and filtering schemes for directional domains are systematically established to address the considered limitation mentioned in Sec. 1.2.3. Dirac mixtures are employed for discrete modeling of on-manifold uncertainties in a nonparametric manner. To discretize the space of unit hyperspheres, approaches for generating hyperspherical grids are introduced in Sec. 5.1. Afterward, an on-manifold discrete filtering scheme using Dirac mixtures of deterministic supports is described in Sec. 5.2 including a generic design given in Sec. 5.2.1 and a showcase on the unit quaternion manifold in Sec. 5.2.2. Furthermore, an on-manifold unscented particle filtering scheme is established in Sec. 5.3 using Dirac mixtures of random supports following sequential Monte Carlo methods, for which a case study for nonlinear SE(3) estimation is given.

In Chapter 6, we propose a novel on-manifold Dirac mixture reapproximation paradigm for efficient discrete modeling of directional uncertainties considering the limitation mentioned in Sec. 1.2.4. We first introduce the concept of on-manifold reapproximation in Sec. 6.1 and its generic design in Sec. 6.2. Afterward, customization of the paradigm to unit hyperspheres is elaborated in Sec. 6.3 followed by further showcases on other considered directional manifolds in Sec. 6.4. Using the reapproximation-based discrete models on hyperspheres, a followup method is provided in Sec. 6.5 for continuous reconstruction of the underlying distribution in the form of von Mises–Fisher mixtures. Moreover, we propose an on-manifold reapproximation discrete filtering scheme in Sec. 6.6 for nonparametric directional estimation with unknown form of system noise, which we evaluate in a case study for nonlinear spherical estimation.

The thesis is concluded in Chapter 7.

CHAPTER 2

Preliminaries

In this chapter, we first give the general formulation of on-manifold recursive Bayesian estimation in Sec. 2.1. Then, we introduce several popular parametric probabilistic models for the considered directional domains in Sec. 2.2.1 to Sec. 2.2.3. Based thereon, a generic description of nonlinear directional filtering algorithms is provided in Sec. 2.2.4.

2.1 Recursive Bayesian Estimation

In this thesis, we consider discrete-time stochastic systems in the following form

$$\underline{\boldsymbol{x}}_{t+1} = \underline{a}(\underline{\boldsymbol{x}}_t, \underline{\boldsymbol{w}}_t)\,, \tag{2.1}$$

with system states $\underline{\boldsymbol{x}}_t$, $\underline{\boldsymbol{x}}_{t+1} \in \mathbb{M}$ being the random variables on a certain manifold $\mathbb{M}$. $\underline{\boldsymbol{w}}_t \in \mathbb{W}$ denotes system noise that belongs to an arbitrary domain $\mathbb{W}$. Function $\underline{a}\colon \mathbb{M} \times \mathbb{W} \to \mathbb{M}$ denotes the system model. Meanwhile, measurements are given by

$$\underline{\boldsymbol{z}}_t = \underline{h}(\underline{\boldsymbol{x}}_t, \underline{\boldsymbol{v}}_t)\,, \tag{2.2}$$

where $\underline{\boldsymbol{z}}_t \in \mathbb{Z}$ and $\underline{\boldsymbol{v}}_t \in \mathbb{V}$ are the measurement and the measurement noise belonging to their individual domains, respectively. Function $\underline{h}\colon \mathbb{M}\times\mathbb{V} \to \mathbb{Z}$ denotes the measurement model. Further, we assume that the system noise term $\underline{\boldsymbol{w}}_t$ in (2.1) and the measurement noise term $\underline{\boldsymbol{v}}_t$ in (2.2) have the probability densities $f_t^{\underline{\boldsymbol{w}}}$ and $f_t^{\underline{\boldsymbol{v}}}$, respectively. Note that the two models above are given in general form and are potentially nonlinear.

In general, recursive Bayesian estimation aims to obtain a posterior distribution f_t^{e} of the state $\underline{\boldsymbol{x}}_t$ at a given time step t. The estimate can be approximated by a distribution of parametric form or a nonparametric model (e.g., in the form of Dirac mixtures). In the upcoming paragraphs, we provide the theoretical derivations of the update and prediction steps for the recursive Bayesian estimation initialized with an initial prior estimate $f_0^{\mathrm{p}}(\underline{x}_0)$ at time step zero. Alongside the narrative in this thesis, implementation of the two steps will be frequently specified according to different settings of individual probabilistic modeling and filtering approaches.

Update step: Given a measurement $\hat{\underline{z}}_t$ from (2.2), the posterior density $f_t^{\mathrm{e}}(\underline{x}_t|\hat{\underline{z}}_{1:t})$ up to time step t can be obtained via updating the predicted or initialized prior density $f_t^{\mathrm{p}}(\underline{x}_t|\hat{\underline{z}}_{1:t-1})$ according to the Bayes' rules [72, Sec. 10.3] in the following form

$$f_t^{\mathrm{e}}(\underline{x}_t|\hat{\underline{z}}_{1:t}) = \frac{f_t^{\mathrm{L}}(\hat{\underline{z}}_t|\underline{x}_t) f_t^{\mathrm{p}}(\underline{x}_t|\hat{\underline{z}}_{1:t-1})}{\int_{\mathbb{M}} f_t^{\mathrm{L}}(\hat{\underline{z}}_t|\underline{x}_t) f_t^{\mathrm{p}}(\underline{x}_t|\hat{\underline{z}}_{1:t-1})\,\mathrm{d}\underline{x}_t} \propto f_t^{\mathrm{L}}(\hat{\underline{z}}_t|\underline{x}_t) f_t^{\mathrm{P}}(\underline{x}_t|\hat{\underline{z}}_{1:t-1})\,. \tag{2.3}$$

Here, $f_t^{\mathrm{L}}(\hat{\underline{z}}_t|\underline{x}_t)$ denotes the likelihood function determined by the measurement model (2.2) and the measurement noise density $f_t^{\underline{\boldsymbol{v}}}$. The integral of the density product over the state $\underline{x}_t$ on manifold

$\mathbb{M}$ is essentially a normalization and is independent of the state. Thus, the posterior distribution can be quantified by scaling the numerator as derived in the second stage of (2.3).

Prediction step: For predicting the prior density $f^{\mathrm{p}}_{t+1}(\underline{x}_{t+1}|\hat{\underline{z}}_{1:t})$ at time step $t+1$ from the previous estimate $f^{\mathrm{e}}_t(\underline{x}_t|\hat{\underline{z}}_{1:t})$ given by (2.3), we apply the Chapman–Kolmogorov equation and obtain

$$f^{\mathrm{p}}_{t+1}(\underline{x}_{t+1}|\hat{\underline{z}}_{1:t}) = \int_{\mathbb{M}} f^{\mathrm{T}}_t(\underline{x}_{t+1}|\underline{x}_t) f^{\mathrm{e}}_t(\underline{x}_t|\hat{\underline{z}}_{1:t}) \,\mathrm{d}\underline{x}_t \,. \tag{2.4}$$

$f^{\mathrm{T}}_t(\underline{x}_{t+1}|\underline{x}_t)$ denotes the transition density function determined by the system model (2.1) under system noise $f^{\boldsymbol{w}}_t$. For identity system models with a manifold-additive noise term, the transition density can often be derived in a symbolic form, and the integral in (2.4) essentially denotes an "on-manifold convolution" $(f^{\boldsymbol{w}}_t \circledast f^{\mathrm{e}}_t)(\underline{x}_{t+1})$ of the system noise and the previous posterior densities ($f^{\boldsymbol{w}}_t$ and f^{e}_t, respectively), with $\circledast$ being a convolution operator adapted to specific manifolds [32,34].

2.2 Parametric Filtering Based on Directional Statistics

In this section, we first provide an overview of commonly used continuous distributions from directional statistics [12] for parametric modeling of directional random variables for the considered domains. Based thereon, a generic description of on-manifold parametric filtering scheme is provided for nonlinear directional estimation.

2.2.1 The von Mises–Fisher Distribution

Parameterized by the mode $\underline{\alpha} \in \mathbb{S}^{d-1}$ and the concentration parameter $\kappa \geq 0$, the von Mises–Fisher distribution $\underline{\boldsymbol{x}} \sim \mathcal{VMF}(\underline{\alpha}, \kappa)$ is defined on the unit hypersphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ ($d \geq 3$) with the probability density function (PDF) [12, 73]

$$f_{\mathcal{VMF}}(\underline{x}; \underline{\alpha}, \kappa) = N_d(\kappa) \cdot \exp(\kappa\, \underline{\alpha}^{\mathsf{T}} \underline{x}) \,,\ \ \underline{x} \in \mathbb{S}^{d-1} \,. \tag{2.5}$$

The normalization constant is only dependent on the concentration and follows

$$N_d(\kappa) = \Big(\int_{\mathbb{S}^{d-1}} \exp(\kappa\, \underline{\alpha}^{\mathsf{T}} \underline{x}) \,\mathrm{d}\underline{x} \Big)^{-1} = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} \boldsymbol{\mathcal{I}}_{d/2-1}(\kappa)} \,, \tag{2.6}$$

with $\boldsymbol{\mathcal{I}}_{d/2-1}$ denoting the modified Bessel function of the first kind and of order $d/2-1$. As shown in (2.5), the distribution models hyperspherical uncertainties w.r.t. the arc length and exhibits isotropic dispersion. For quantifying the expectation of von Mises–Fisher-distributed random variables, the *mean resultant vector* is introduced by generalizing the trigonometric moment from circles to hyperspheres. It is given by the formula

$$\underline{v} = \mathrm{E}(\underline{\boldsymbol{x}}) = \int_{\mathbb{S}^{d-1}} \underline{x} \cdot f_{\mathcal{VMF}}(\underline{x}; \underline{\alpha}, \kappa) \,\mathrm{d}\underline{x} = \boldsymbol{\mathcal{A}}_d(\kappa)\, \underline{\alpha} \,, \quad \text{with} \quad \boldsymbol{\mathcal{A}}_d = \boldsymbol{\mathcal{I}}_{d/2}(\kappa) / \boldsymbol{\mathcal{I}}_{d/2-1}(\kappa) \tag{2.7}$$

being a Bessel function ratio. Thus, the hyperspherical mean $\underline{v}$ points in the direction given by $\underline{\alpha}$ and is of length $\boldsymbol{\mathcal{A}}_d(\kappa)$ determined by the concentration κ. Consequently, fitting a von Mises–Fisher distribution given weighted hyperspherical samples $\{(\underline{x}_i, \omega_i)\}_{i=1}^n$ can be performed via

$$\hat{\underline{\alpha}} = \hat{\underline{v}} / \,||\hat{\underline{v}}|| \quad \text{and} \quad \hat{\kappa} = \boldsymbol{\mathcal{A}}_d^{-1}(\hat{\underline{v}}) \,, \quad \text{with} \quad \hat{\underline{v}} = \sum_{i=1}^{n} \omega_i\, \underline{x}_i \tag{2.8}$$

being the mean of the samples. Computing the inverse of the Bessel function ratio $\mathcal{A}_d$ to obtain the concentration $\hat{\kappa}$ in (2.8) can be done efficiently using the Newton's method given by [74]. As justified in [Sec. A.1] [7], matching the mean resultant vector of a von Mises–Fisher distribution produces results equivalent to those obtained via maximum likelihood estimation (MLE). For fitting an arbitrary hyperspherical distribution to the von Mises–Fisher form, this procedure also guarantees minimum information loss w.r.t. the Kullback–Leibler divergence [75, 76].

The von Mises–Fisher distribution is a popular statistical tool for modeling uncertain hyperspherical data [77, 77, 78]. However, limitations may arise when hyperspherical uncertainties are inherently of antipodal symmetry or improper to be modeled with isotropic dispersion, e.g., for unit quaternions representing SO(3) states. To eliminate such issues, the Bingham distribution comes into play.

2.2.2 The Bingham Distribution

The Bingham distribution $\underline{\boldsymbol{x}} \sim \mathcal{B}(\mathbf{M}, \mathbf{Z})$ can be obtained by intersecting a zero-mean multivariate Gaussian distribution in $\mathbb{R}^d$ with the unit hypersphere $\mathbb{S}^{d-1}$, on which a renormalization is followed afterward. It has the following probability density function

$$f_{\mathcal{B}}(\underline{x}) = \frac{1}{N(\mathbf{Z})} \exp(\underline{x}^\top \mathbf{M}\,\mathbf{Z}\,\mathbf{M}^\top \underline{x})\,,\ \underline{x} \in \mathbb{S}^{d-1} \subset \mathbb{R}^d\,. \tag{2.9}$$

The parameter matrix $\mathbf{Z} = \operatorname{diag}(z_1, \dots, z_{d-1}, z_d)$ controls the concentration of dispersion, with $z_1 \leq \cdots \leq z_{d-1} \leq z_d = 0$ being the diagonal elements[1]. $\mathbf{M} \in \mathbb{R}^{d\times d}$ is a real orthogonal matrix (thus $\mathbf{M}\,\mathbf{M}^\top = \mathbf{M}^\top\mathbf{M} = \mathbf{I}_{d\times d}$), which determines the orientation of dispersion on the hypersphere. In practice, the parameters $\mathbf{M}$ and $\mathbf{Z}$ can be obtained via performing eigenvalue decomposition to a negative semidefinite matrix $\mathbf{C}_{\mathcal{B}} \in \mathbb{R}^{d\times d}$, with $\mathbf{C}_{\mathcal{B}} = \mathbf{M}\,\mathbf{Z}\,\mathbf{M}^\top$ (the distribution can thus be denoted as $\underline{\boldsymbol{x}} \sim \mathcal{B}(\mathbf{C}_{\mathcal{B}})$ for brevity as well). Afterward, eigenvalues in the diagonal matrix $\mathbf{Z}$ are rearranged in ascending order with eigenvectors in $\mathbf{M}$ reordered correspondingly. And by convention, we subtract the diagonal elements of $\mathbf{Z}$ with the largest eigenvalue to fit the definition in (2.9) (there, $z_d = 0$). Consequently, the last column of matrix $\mathbf{M}$ denotes the mode of the Bingham distribution as it corresponds to the largest diagonal element in the concentration parameter matrix $\mathbf{Z}$.

As shown in (2.9), the normalization constant of the Bingham distribution is purely determined by the concentration matrix $\mathbf{Z}$. Computing the normalization constant follows

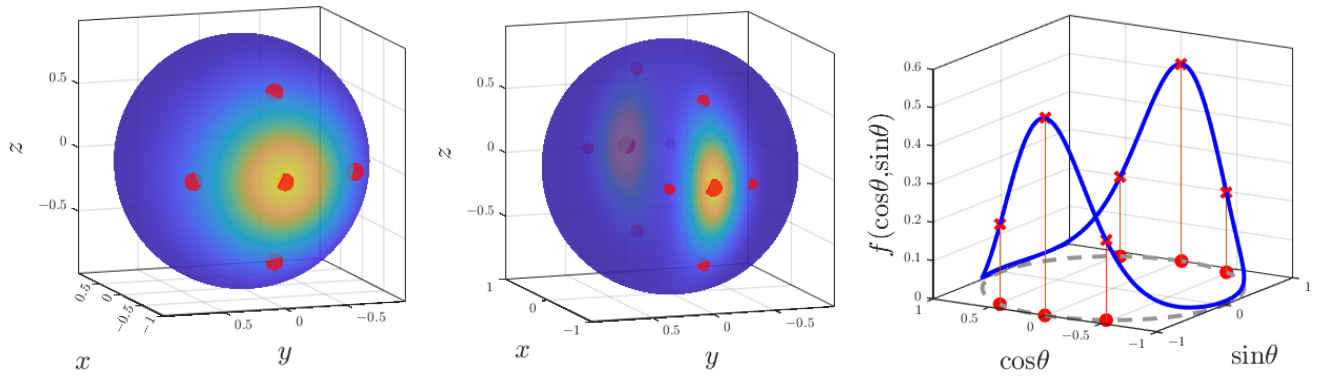
$$N(\mathbf{Z}) = \int_{\mathbb{S}^{d-1}} \exp\left(\underline{x}^\top \mathbf{M}\,\mathbf{Z}\,\mathbf{M}^\top \underline{x}\right) \mathrm{d}\underline{x} = \left|\mathbb{S}^{d-1}\right| \cdot {}_1\mathcal{F}_1\big(1/2, d/2, \mathbf{Z}\big)\,, \tag{2.10}$$

with $\left|\mathbb{S}^{d-1}\right|$ being the surface of the unit hypersphere $\mathbb{S}^{d-1}$ and ${}_1\mathcal{F}_1$ the hypergeometric function of a matrix argument, for which no closed-form formula exists [79]. For practical applications such as quaternion filtering, efficient approaches have been proposed based on lookup tables or saddle point approximation [17, 79, 80]. Further, the distribution has a mean at $\underline{0} \in \mathbb{R}^d$ and second-order moment in the form

$$\operatorname{cov}_{\mathcal{B}} = \mathbf{M} \cdot \operatorname{diag}(c_1, \dots, c_d) \cdot \mathbf{M}^\top\,, \quad \text{with} \quad c_i = \frac{\partial N(\mathbf{Z})}{\partial z_i} / N(\mathbf{Z})\,,\ i = 1, \dots, d \tag{2.11}$$

denoting the diagonal entries of the matrix in the middle and fulfilling $\sum_{i=1}^{d} c_i = 1$. For any $\underline{x} \in \mathbb{S}^{d-1}$, it is trivial to validate that $f_{\mathcal{B}}(\underline{x}) = f_{\mathcal{B}}(-\underline{x})$ holds. Therefore, the dispersion of the Bingham distribution is inherently antipodally symmetric on unit hyperspheres and is directly applicable to modeling uncertain unit quaternions belonging to $\mathbb{S}^3$ as elaborated in Sec. 3.2.

[1] The function diag maps each input to each diagonal entry of a diagonal matrix.



(A) $\mathcal{VMF}$ on $\mathbb{S}^2$ in Example 1.1 (B) $\mathcal{B}$ on $\mathbb{S}^2$ in Example 1.2 (C) $\mathcal{B}$ on $\mathbb{S}^1$ in Example 1.2

Figure 2.1.: Basic UT-based deterministic sampling for common parametric directional models in Example 1.

2.2.3 The SE(2)-Bingham Distribution

As a good example of topology-aware parametric models on composite directional manifolds, the SE(2)-Bingham distribution is defined on $\mathbb{S}^1 \times \mathbb{R}^2 \subset \mathbb{R}^4$ (i.e., the Cartesian product of the unit circle and the two-dimensional Euclidean space) for modeling uncertain planar motions parameterized by dual quaternions [55]. Formulated as a four-dimensional vector $\underline{x} = [\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top]^\top \in \mathbb{R}^4$, the so-called planar dual quaternion represents the planar rotation with the real part $\underline{x}_\mathrm{r} \in \mathbb{S}^1$ and encodes the planar translation with the dual part $\underline{x}_\mathrm{s} \in \mathbb{R}^2$ (a detailed introduction is provided in Sec. 3.3.3). Correspondingly, the SE(2)-Bingham distribution $\mathcal{PB}(\mathbf{C})$ has the PDF

$$f_{\mathcal{PB}}(\underline{x}) = \frac{1}{N(\mathbf{C})} \exp(\underline{x}^\top \mathbf{C}\, \underline{x})\,, \quad \text{with} \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2^\top \\ \mathbf{C}_2 & \mathbf{C}_3 \end{bmatrix} \tag{2.12}$$

being the parameter matrix, where $\mathbf{C}_1 \in \mathbb{R}^{2\times 2}$ is symmetric, $\mathbf{C}_2 \in \mathbb{R}^{2\times 2}$ arbitrary and $\mathbf{C}_3 \in \mathbb{R}^{2\times 2}$ symmetric negative definite. Based on the Schur complement of $\mathbf{C}_3$, the PDF in (2.12) can be decomposed into

$$f_{\mathcal{PB}}(\underline{x}) = \frac{1}{N(\mathbf{C})} \exp\left(\underline{x}_\mathrm{r}^\top \mathbf{T}_1\, \underline{x}_\mathrm{r} + (\underline{x}_\mathrm{s} - \mathbf{T}_2\, \underline{x}_\mathrm{r})^\top \mathbf{C}_3\, (\underline{x}_\mathrm{s} - \mathbf{T}_2\, \underline{x}_\mathrm{r})\right), \tag{2.13}$$

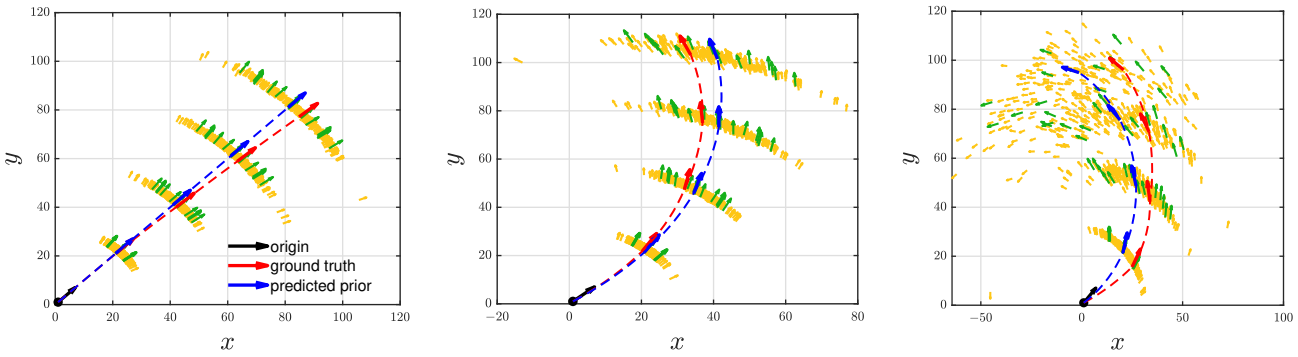
with $\mathbf{T}_1 = \mathbf{C}_1 - \mathbf{C}_2^\top \mathbf{C}_3^{-1} \mathbf{C}_2$ and $\mathbf{T}_2 = \mathbf{C}_3^{-1} \mathbf{C}_2$. Thus, the rotation term $\underline{x}_\mathrm{r}$ follows a Bingham distribution parameterized by $\mathbf{T}_1$ and the translation term $\underline{x}_\mathrm{s}$ a conditional Gaussian distribution $\mathcal{N}(\mathbf{T}_2\, \underline{x}_\mathrm{r}, -0.5\, \mathbf{C}_3^{-1})$ [40, 81]. Consequently, the normalization constant of the PDF in (2.12) follows

$$N(\mathbf{C}) = \frac{2\pi \sqrt{\det(-0.5\, \mathbf{C}_3^{-1})}}{N_{\mathcal{B}}(\mathbf{T}_1)}\,,$$

with $N_{\mathcal{B}}(\mathbf{T}_1)$ denoting the normalization constant (2.10) of the Bingham component parameterized by matrix $\mathbf{T}_1$ in (2.13). While considering the antipodal symmetry of the uncertainty underlying on the manifold, the SE(2)-Bingham distribution also properly interprets the probabilistic correlation between the two component domains (namely, $\mathbb{S}^1$ and $\mathbb{R}^2$).

2.2.4 Nonlinear Directional Filtering Using Parametric Models

Deterministic sampling for unscented transform: For parametric directional distributions, sampling is almost always essential for topology-aware recursive estimation, in particular in



(A) $\mathcal{PB}$ with $\lambda = 1000$, $\theta = 0$ (B) $\mathcal{PB}$ with $\lambda = 500$, $\theta = \pi/12$ (C) $\mathcal{PB}$ with $\lambda = 50$, $\theta = \pi/6$

Figure 2.2.: Consecutive UT-based predictions with SE(2)-Bingham-distributed system noises in Example 1.3. Yellow arrows denote samples randomly drawn from the predicted prior density whereas the green ones are drawn deterministically [55].

scenarios with nonlinearities or composite directional domains. As random samples cannot guarantee reproducible results and are inefficient for representing the underlying distributions, deterministic sampling techniques have been proposed in accordance with the concept of statistical linearization [82, 83]. For typical parametric directional models, samples are drawn in the fashion of the unscented transform [8] to match the first- and second-order moments of the underlying distribution. The following example is provided for showcasing this basic sampling scheme practiced on the previously introduced distributions.

Example 1: *We parameterize the three parametric models in Sec. 2.2.1 to Sec. 2.2.3 as follows:*

1. *A von Mises–Fisher distribution* $\mathcal{VMF}(\underline{\alpha}, \kappa)$ *is defined on the unit sphere* $\mathbb{S}^2$ *with mode* $\underline{\alpha} = [\,-1, 0, 0\,]^\top \in \mathbb{S}^2$ *and concentration* $\kappa = 5$.
2. *Bingham distributions* $\mathcal{B}(\mathbf{C}_\mathcal{B})$ *are parameterized on the unit sphere* $\mathbb{S}^2$ *and unit circle* $\mathbb{S}^1$ *by matrices* $\mathbf{C}_\mathcal{B} = -\mathrm{diag}(1, 20, 5)$ *and* $\mathbf{C}_\mathcal{B} = -\mathrm{diag}(1, 5)$*, respectively.*
3. *SE(2)-Bingham distributions* $\mathcal{PB}(\mathbf{C})$ *are configured with matrices* $\mathbf{C} = -\mathrm{diag}([\,1, \lambda, \lambda, \lambda\,])$*, where* $\lambda \in \{1000, 500, 50\}$ *control concentrations of dispersion. For the configured distributions, we additionally set their modes at planar motions composing rotation angles* $\theta \in \{\,0, \pi/12, \pi/6\,\}$ *and the same translation term* $\underline{t} = [\,20, 20\,]^\top$.

The basic UT-based sampling schemes for circular and hyperspherical distributions always draw $2 \cdot d - 1$ *samples on* $\mathbb{S}^{d-1} \subset \mathbb{R}^d$*. As shown in Fig. 2.1-(A), samples are drawn from the von Mises–Fisher distribution in an isotropic layout on the unit sphere* $\mathbb{S}^2$ *coherently to the underlying isotropic dispersion via the method given by [34]. The Bingham distribution has non-isotropic dispersion of antipodal symmetry. Samples are drawn according to [32] in a topology-adaptive manner on* $\mathbb{S}^2$ *and* $\mathbb{S}^1$ *as depicted in Fig. 2.1-(B) and (C), respectively.*

Using the sampling approach mentioned in [55, 56], planar dual quaternion samples are obtained from the SE(2)-Bingham distribution by combining deterministic samples separately drawn from the Bingham and Gaussian components in (2.13) *via Cartesian product. As plotted in Fig. 2.2, we deploy the SE(2)-Bingham distributions configured in the third scenario above to an identity planar motion model for simulating the noise term. After each time step, a predicted prior is fitted according to the propagated deterministic samples via moment matching. The SE(2)-Bingham*

Algorithm 1: Generic Nonlinear Parametric Directional Filter

Input: prior density f_t^{p}, measurement $\hat{\underline{z}}_t$, system noise density $f_t^{\underline{\boldsymbol{w}}}$
Output: predicted prior density f_{t+1}^{p}

/* update step */
1 $\{(\underline{\sigma}_{t,i}^{\mathrm{p}}, \nu_{t,i}^{\mathrm{p}})\}_{i=1}^{n} \leftarrow$ `sampleDeterministic` (f_t^{p});
2 **for** $i \leftarrow 1$ **to** n **do**
3 $\quad \nu_{t,i}^{\mathrm{p}} \leftarrow \nu_{t,i}^{\mathrm{p}} \cdot f_t^{\mathrm{L}}(\hat{\underline{z}}_t | \underline{\sigma}_{t,i}^{\mathrm{p}})$;
4 $f_t^{\mathrm{e}} \leftarrow$ `matchMoment` $(\{(\underline{\sigma}_{t,i}^{\mathrm{p}}, \nu_{t,i}^{\mathrm{p}})\}_{i=1}^{n})$;
/* prediction step */
5 $\{(\underline{\sigma}_{t,i}^{\mathrm{e}}, \nu_{t,i}^{\mathrm{e}})\}_{i=1}^{n} \leftarrow$ `sampleDeterministic` (f_t^{e});
6 $\{(\underline{\sigma}_{t,k}^{\mathrm{w}}, \nu_{t,k}^{\mathrm{w}})\}_{k=1}^{n_{\mathrm{w}}} \leftarrow$ `sampleDeterministic` $(f_t^{\underline{\boldsymbol{w}}})$;
7 **for** $i \leftarrow 1$ **to** n **do**
8 $\quad$ **for** $k \leftarrow 1$ **to** n_{w} **do**
9 $\quad\quad \underline{\sigma}_{t+1,r}^{\mathrm{p}} \leftarrow \underline{a}(\underline{\sigma}_{t,i}^{\mathrm{e}}, \underline{\sigma}_{t,k}^{\mathrm{w}})$;
10 $\quad\quad \nu_{t+1,r}^{\mathrm{p}} \leftarrow \nu_{t,i}^{\mathrm{e}} \cdot \nu_{t,k}^{\mathrm{w}}$;
11 $f_{t+1}^{\mathrm{p}} \leftarrow$ `matchMoment` $(\{(\underline{\sigma}_{t+1,r}^{\mathrm{p}}, \nu_{t+1,r}^{\mathrm{p}})\}_{r=1}^{n \cdot n_{\mathrm{w}}})$;
12 **return** f_{t+1}^{p};

distribution is able to model the uncertain rotation and translation individually, meanwhile the probabilistic correlation between the two components is interpreted.

Parametric directional filtering scheme: As showcased in Example 1.3 by the procedure of obtaining predicted priors, parametric directional filters are typically operated in a sampling–matching pattern for nonlinear estimation tasks. To concretize this concept, we show a generic description for directional estimation using parametric models according to the derivation in Sec. 2.1.

As shown in Algorithm 1, given the prior density f_t^{p} from the initialization or previous step, we first draw (weighted) deterministic samples $\{(\underline{\sigma}_{t,i}^{\mathrm{p}}, \nu_{t,i}^{\mathrm{p}})\}_{i=1}^{n}$ according to the specific form of the parametric model as shown in Example 1 (Algorithm 1, line 1). We assume that a likelihood function can be obtained from the measurement model in (2.2), which we exploit to reweight the prior samples given the current measurement $\hat{\underline{z}}_t$ (Algorithm 1, lines 2–3). Afterward, the posterior density f_t^{e} is obtained via moment-matching to the reweighted prior samples (Algorithm 1, line 4). In the upcoming prediction step, we first perform deterministic sampling on the posterior density given by the update step and the noise density $f_t^{\underline{\boldsymbol{w}}}$. (Algorithm 1, lines 5–6). The two sample sets are then combined via Cartesian product and propagated through the system model in (2.1) (Algorithm 1, lines 7–10). The resulting prior sample sets are then moment-matched according to the parametric form of the state distribution, and the predicted prior density f_{t+1}^{p} is then obtained (Algorithm 1, line 11).

In the past state of the art, parametric directional filters are usually tailored w.r.t. the specific forms of system or measurement models. For instance, the measurement model is typically assumed to be identity, with measurements restricted to the manifold of the state. Therefore, measurement models and the measurement noise distributions need to be adapted to the identity setting, leading to faster measurement update (often in closed form without sampling) but also additional approximation error. Directly reweighting the basic UT-based deterministic samples with their likelihoods derived

from the original measurement model induces degeneracy even under mild nonlinearities due to limited sample sizes. To address such issues, we will introduce several enhancements in Chapter 4. Before that, we first give a systematic geometric investigation in the upcoming chapter on several important directional manifolds, which are highly relevant to engineering practices of directional estimation.

CHAPTER 3

Geometric Investigation on Directional Manifolds

In this chapter, we start off by investigating the hyperspherical geometry in Sec. 3.1, which plays a fundamental role in constructing composite directional manifolds. Then, we specify unit hyperspheres to the unit quaternion manifold in Sec. 3.2, based on which the unit dual quaternion manifold is further introduced in Sec. 3.3 to illustrate approaches to inspecting directional manifolds composing Euclidean spaces (including the degenerate case of planar dual quaternions). Note both of the two directional variables are well-established approaches to representing rigid body motions, which we also exploit recurrently to construct nonlinear estimation scenarios for showcasing the filtering schemes proposed later. Knowledge about other directional domains such as tori and hypertori is not detailed in this thesis, for which we refer to related literature in [12, 47, 53].

3.1 Hyperspherical Geometry

A unit hypersphere is a compact and bounded Riemannian manifold with constant curvature [84]. In this section, we perform a geometric investigation on unit hyperspheres using common analysis tools from Riemannian geometry [85,86]. At the end of this section, we showcase the introduced geometric tools on the circular Bingham distribution and discuss their potential usage for probabilistic modeling and filtering of directional random variables.

3.1.1 Exponential and Logarithm Maps on Unit Hyperspheres

Given an arbitrary point $\underline{v} \in \mathbb{S}^{d-1} \subset \mathbb{R}^d$ on the unit hypersphere, a hyperspherical tangent space $\mathbb{T}_{\underline{v}}\mathbb{S}^{d-1} \subset \mathbb{R}^d$ can be located on it [84]. Any point $\underline{x} \in \mathbb{S}^{d-1}$ can be mapped to $\mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ via the logarithm map

$$\tilde{\underline{x}} = \mathrm{Log}_{\underline{v}}(\underline{x}) = \big(\underline{x} - \cos(\gamma)\,\underline{v}\big)\frac{\gamma}{\sin(\gamma)} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}\,, \quad \text{with} \quad \gamma = \arccos(\underline{v}^\top \underline{x})\,, \tag{3.1}$$

while preserving its geodesic curve length to $\underline{v}$, namely, $\|\mathrm{Log}_{\underline{v}}(\underline{x})\| = \gamma$. Inversely, any point $\tilde{\underline{x}} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ can be retracted back to the hyperspherical manifold via the exponential map

$$\underline{x} = \mathrm{Exp}_{\underline{v}}(\tilde{\underline{x}}) = \cos(\|\tilde{\underline{x}}\|)\underline{v} + \frac{\sin(\|\tilde{\underline{x}}\|)}{\|\tilde{\underline{x}}\|}\tilde{\underline{x}} \in \mathbb{S}^{d-1}\,. \tag{3.2}$$

Points in the tangent space $\mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ can also be expressed locally w.r.t. an orthonormal basis $\mathcal{E}_{\underline{v}}$ via $\tilde{\underline{x}}^\iota = \mathbf{E}_{\underline{v}}^\top \tilde{\underline{x}} \in \mathbb{R}^{d-1}$, where columns of matrix $\mathbf{E}_{\underline{v}}$ are basis vectors of $\mathcal{E}_{\underline{v}}$. Expressed w.r.t. $\mathcal{E}_{\underline{v}}$, hyperspherical points undergoing the logarithm map in (3.1) essentially form a $(d-1)$-ball of radius π, namely, $\tilde{\underline{x}}^\iota \in \mathbb{B}_\pi^{d-1} \subset \mathbb{R}^{d-1}$, which is bounded by the hypersphere $\mathbb{S}_\pi^{d-2}$ of radius π.

3.1.2 Projections on Unit Hyperspheres

Mapping points from a unit hypersphere $\mathbb{S}^{d-1}$ to its tangent space $\mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ at an arbitrary point $\underline{v} \in \mathbb{S}^{d-1}$ can also be performed via projections [87]. Among others, we introduce two types of mapping approaches, the orthographic and gnomonic projections, and their inverse operations, the orthographic and gnomonic retractions, respectively.

Orthographic projection and retraction: The orthographic projection projects a point $\underline{x} \in \mathbb{S}^{d-1}$ to the tangent space $\mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ orthogonally [35]. Mathematically, the operation is expressed as

$$\underline{\tilde{x}} = \boldsymbol{\mathcal{P}}^{\mathrm{o}}_{\underline{v}}(\underline{x}) = \underline{x} - (\underline{v}^{\top}\underline{x})\,\underline{v} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}\,,\ \forall \underline{x} \in \mathbb{S}^{d-1}\,.$$

Inversely, any point $\underline{\tilde{x}} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ can be retracted back to the underlying hypersphere via

$$\underline{x} = \boldsymbol{\mathcal{R}}^{\mathrm{o}}_{\underline{v}}(\underline{\tilde{x}}) = \underline{\tilde{x}} + \sqrt{1 - \|\underline{\tilde{x}}\|^2}\,\underline{v} \in \mathbb{S}^{d-1}\,,\ \forall \underline{\tilde{x}} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}\,.$$

Gnomonic projection and retraction: Intuitively, the gnomonic projection shoots a line from the origin through the given hyperspherical point $\underline{x} \in \mathbb{S}^{d-1}$ and takes the line intersection with the tangent space $\mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}$ as the output [35]. This procedure can be formulated as

$$\underline{\tilde{x}} = \boldsymbol{\mathcal{P}}^{\mathrm{g}}_{\underline{v}}(\underline{x}) = \frac{\underline{x}}{\underline{v}^{\top}\underline{x}} - \underline{v} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}\,,\ \forall \underline{x} \in \mathbb{S}^{d-1}\,. \tag{3.3}$$

Correspondingly, the gnomonic retraction follows

$$\underline{x} = \boldsymbol{\mathcal{R}}^{\mathrm{g}}_{\underline{v}}(\underline{\tilde{x}}) = \frac{\underline{v} + \underline{\tilde{x}}}{\|\underline{v} + \underline{\tilde{x}}\|} \in \mathbb{S}^{d-1}\,,\ \forall \underline{\tilde{x}} \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}\,, \tag{3.4}$$

which essentially normalizes the coordinates of point $\underline{v} + \underline{\tilde{x}}$. The following example provides an illustration of bridging a directional manifold with its tangent space for a circular Bingham distribution using the aforementioned logarithm/exponential maps and the projection-based mapping approaches.

Example 2: *Similar to Example 1.2, we configure a Bingham distribution $\mathcal{B}(\mathbf{C})$ on the unit circle with $\mathbf{C} = -\mathrm{diag}(1, 2)$. As shown in Fig. 3.1, we first draw samples using the basic UT-based deterministic sampling method in [32] (shown as black dots on $\mathbb{S}^1$). Afterward, we establish the tangent plane $\mathbb{T}_{\underline{m}}\mathbb{S}^1$ at the Bingham mode $\underline{m} \in \mathbb{S}^1$ (drawn as a gray line, the one at its antipode is discarded) and map the samples drawn on the same side to it using the three mapping approaches introduced previously. As shown in the plot, the logarithm map preserves the arc length from the samples to the mode in the tangent space. Expressed w.r.t. the local basis of the tangent plane $\mathbb{T}_{\underline{m}}\mathbb{S}^1$, the orthographic projection outputs a bounded range of $[-1, 1] \subset \mathbb{R}^1$, whereas an unbounded output domain is theoretically guaranteed by the gnomonic projection given circular inputs.*

As shown in Example 2, the aforementioned mapping approaches can all in theory facilitate in-tangent-space interpretation of on-manifold uncertainties. Suppose a Gaussian distribution is defined in the tangent space at a certain point on $\mathbb{S}^{d-1}$, and it is retracted back to the unit hypersphere via one of the three maps above to quantify the underlying uncertainty. It is then evident that the commonly used exponential map [11] induces a warped distribution on $\mathbb{S}^{d-1}$, and the orthographic retraction truncates the in-tangent-space dispersion orthogonally to the hypersphere due to the bounded input range. Therefore, both phenomena lead to problematic

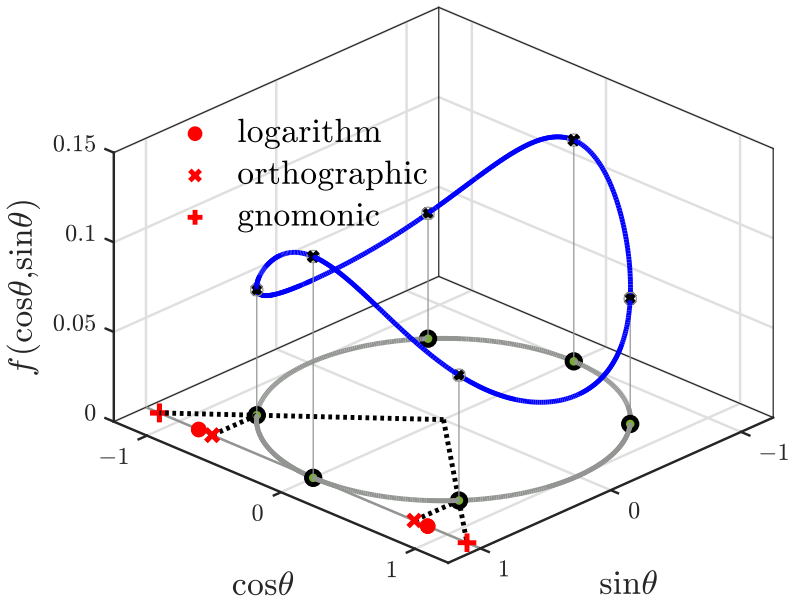


Figure 3.1.: Hyperspherical geometry illustrated on the unit circle $\mathbb{S}^1$ with a Bingham distribution.

interpretation of on-manifold uncertainties and can only be practically employed under certain assumptions (e.g., local perturbation). The gnomonic projection/retraction, however, provide (so far) a theoretically sound solution to bridging a directional manifold and its tangent spaces for probabilistic modeling.

3.2 Unit Quaternion Manifold

By convention, a quaternion $\underline{x}$ is defined as $\underline{x} = x_0 + x_1\,\mathsf{i} + x_2\,\mathsf{j} + x_3\,\mathsf{k}$ with i, j, k being the basic quaternions that fulfill $\mathsf{i}^2 = \mathsf{j}^2 = \mathsf{k}^2 = \mathsf{ijk} = -1$. For the sake of conciseness, we write quaternions in vector form $\underline{x} = [\,x_0, x_1, x_2, x_3\,]^\top \in \mathbb{R}^4$ and exploit matrix-vector operations for expressing quaternion arithmetic in this thesis.

3.2.1 Quaternion Arithmetic and SO(3) State Representation

The product of two quaternions is computed via the Hamilton product $\otimes$, in which the basis elements are multiplied under the distributive law [88]. In accordance with the vector formulation, the Hamilton product can also be expressed via ordinary matrix-vector multiplications. For instance, given two arbitrary quaternions $\underline{r} = [\,r_0, r_1, r_2, r_3\,]^\top$ and $\underline{s} = [\,s_0, s_1, s_2, s_3\,]^\top$, their Hamilton product can be expressed as $\underline{r} \otimes \underline{s} = \mathbf{Q}^{\llcorner}_{\underline{r}}\,\underline{s} = \mathbf{Q}^{\lrcorner}_{\underline{s}}\,\underline{r}$, with

$$\mathbf{Q}^{\llcorner}_{\underline{r}} = \begin{bmatrix} r_0 & -r_1 & -r_2 & -r_3 \\ r_1 & r_0 & -r_3 & r_2 \\ r_2 & r_3 & r_0 & -r_1 \\ r_3 & -r_2 & r_1 & r_0 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}^{\lrcorner}_{\underline{s}} = \begin{bmatrix} s_0 & -s_1 & -s_2 & -s_3 \\ s_1 & s_0 & s_3 & -s_2 \\ s_2 & -s_3 & s_0 & s_1 \\ s_3 & s_2 & -s_1 & s_0 \end{bmatrix} \tag{3.5}$$

being the left and right representations of quaternions $\underline{r}$ and $\underline{s}$, respectively [50]. Moreover, the norm of a quaternion $\underline{r}$ is given as $\sqrt{\underline{r} \otimes \underline{r}^*}$, where $\underline{r}^* = [\,r_0, -r_1, -r_2, -r_3\,]^\top$ denotes the conjugate of $\underline{r}$. Consequently, quaternions of unit norm are called unit quaternions.

As a reparameterization of the axis–angle representation, unit quaternions have become a popular tool for representing spatial rotations belonging to the special orthogonal group SO(3). Given a spatial rotation around an axis $\underline{n}$ through an angle θ, it can be represented by a unit quaternion defined as

$$\underline{r} = [\,\cos(\theta/2)\,, \underline{n}^\top \sin(\theta/2)\,]^\top\,, \tag{3.6}$$

where the rotation axis $\underline{n}$ is of unit length, i.e., $\underline{n} \in \mathbb{S}^2$. Any point $\underline{v} \in \mathbb{R}^3$ can be rotated to $\underline{v}'$ by the rotation parameterized in $\underline{r}$ according to

$$\underline{v}' = \left(\underline{r} \otimes [\,0, \underline{v}^\top\,]^\top \otimes \underline{r}^*\right)_{2:4}\,. \tag{3.7}$$

We augment the vector $\underline{v}$ with a scalar term of zero to its quaternion form for compatibility with the Hamilton product and use the operator $(\cdot)_{2:4}$ to extract the last three elements of a quaternion to recover the rotated vector. By reformulating the quaternion rotation law in (3.7) using the matrix expressions in (3.5), we obtain

$$\underline{r} \otimes [\,0, \underline{v}^\top\,]^\top \otimes \underline{r}^* = (\mathbf{Q}_{\underline{r}}^{\lrcorner})^\top \mathbf{Q}_{\underline{r}}^{\llcorner}\, [\,0, \underline{v}^\top]^\top\,, \quad \text{with} \quad (\mathbf{Q}_{\underline{r}}^{\lrcorner})^\top \mathbf{Q}_{\underline{r}}^{\llcorner} = \begin{bmatrix} 1 & \mathbf{0}_{3\times1} \\ \mathbf{0}_{1\times3} & \mathbf{R}_\mathrm{r} \end{bmatrix}\,, \tag{3.8}$$

where $\mathbf{R}_\mathrm{r} \in \mathrm{SO}(3)$ denotes the SO(3) matrix representing an identical rotation. Therefore, we have $\underline{v}' = \mathbf{R}_\mathrm{r}\, \underline{v}$.

3.2.2 Geometric Structure of Unit Quaternion Manifold

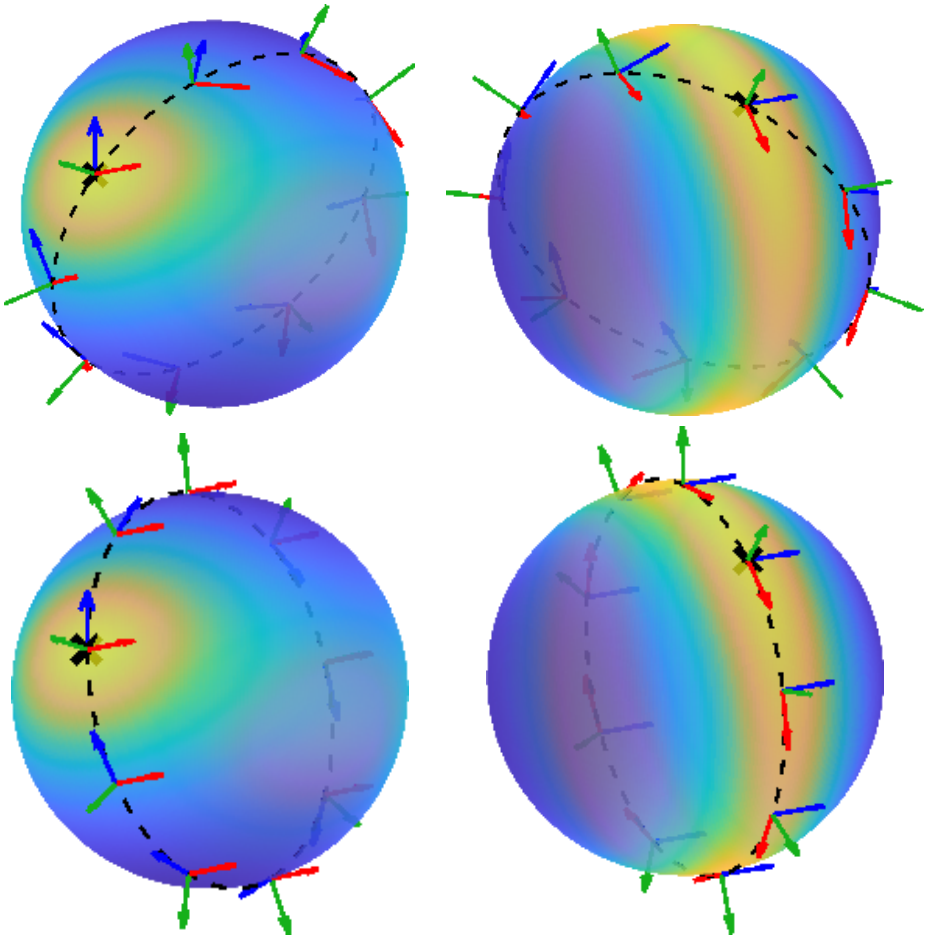
According to the definition of quaternion norm in Sec. 3.2.1, unit quaternions are also of unit length when expressed as vectors in the four-dimensional Euclidean space. Therefore, they are confined to the unit hypersphere $\mathbb{S}^3 \subset \mathbb{R}^4$. As shown in (3.7), two antipodally symmetric unit quaternions $\underline{x}$ and $-\underline{x}$ represent the same spatial rotation. Thus, the unit hypersphere $\mathbb{S}^3$ is a double covering of the SO(3) group. Therefore, the geometric tools introduced in Sec. 3.1 for unit hyperspheres are also applicable to the manifold of unit quaternions. For instance, the logarithm map in (3.1) operated at the identity quaternion $\mathbb{1} = [\,1, 0, 0, 0\,]^\top \in \mathbb{S}^3$ boils down to

$$\mathrm{Log}_{\mathbb{1}}(\underline{x}) = [\,0\,, \frac{\theta}{2}\, \underline{n}^\top\,]^\top \in \mathbb{T}_{\mathbb{1}}\mathbb{S}^3\,, \tag{3.9}$$

with $\underline{x} = [\,\cos(\theta/2), \underline{n}^\top \sin(\theta/2)]^\top \in \mathbb{S}^3$ being an arbitrary unit quaternion as defined in (3.6).

For any unit quaternion, its matrix expressions respecting the Hamilton product in (3.5) inherently belong to the four-dimensional special orthogonal group SO(4), namely, $\mathbf{Q}_{\underline{x}}^{\circ}(\mathbf{Q}_{\underline{x}}^{\circ})^\top = (\mathbf{Q}_{\underline{x}}^{\circ})^\top \mathbf{Q}_{\underline{x}}^{\circ} = \mathbf{I}_{4\times4}$ and $\det(\mathbf{Q}_{\underline{x}}^{\circ}) = 1$ (the superscript $\circ$ summarizes the left $\llcorner$ and right $\lrcorner$ superscripts) [35, 43]. Thus, the Hamilton product of unit quaternions essentially denotes four-dimensional rotations, under which the manifold $\mathbb{S}^3$ is closed. Furthermore, the inverse of a unit quaternion $\underline{x}$ is identical to its conjugate, i.e., $\underline{x}^{-1} = \underline{x}^*$, with corresponding matrix representations fulfilling $\mathbf{Q}_{\underline{x}^{-1}}^{\circ} = (\mathbf{Q}_{\underline{x}}^{\circ})^{-1} = (\mathbf{Q}_{\underline{x}}^{\circ})^\top$.

As indicated in (3.7) and (3.9), the manifold of unit quaternions representing SO(3) states contains the additional topological structure of antipodal symmetry compared with the ordinary $\mathbb{S}^3$. Therefore, unit quaternions undergoing the logarithm map w.r.t. a local basis of the tangent space form a ball $\mathbb{B}^3_{\pi/2}$ of radius $\pi/2$. This is different from the finding for ordinary unit hyperspheres stated in Sec. 3.1.1, where the ball is equipped with a radius of π.



(A) $\mathcal{B}(\mathbf{C}_1)$ on $\mathbb{S}^2$ (B) $\mathcal{B}(\mathbf{C}_2)$ on $\mathbb{S}^2$

Figure 3.2.: Parallel transport of the Bingham principal basis on $\mathbb{S}^2$ in Example 3. The cross '×' denotes the mode of the underlying distribution. For each configured distribution shown in column (A) and (B), coordinate systems with basis vectors pointing along the principal curves stay parallel w.r.t. the connection on different spherical trajectories.

From the perspective of Riemannian geometry, the four-dimensional rotation enabled by the Hamilton product also corresponds to the concept of parallel transport on $\mathbb{S}^3$ [89]. Based thereon, logarithm and exponential maps at an arbitrary quaternion $\underline{v} \in \mathbb{S}^3$ can be transferred to the identity quaternion $\mathbb{1} \in \mathbb{S}^3$ and operated according to (3.9), namely,

$$\begin{aligned} \underline{\tilde{x}} &= \mathrm{Log}_{\underline{v}}(\underline{x}) = \underline{v} \otimes \mathrm{Log}_{\mathbb{1}}(\underline{v}^{-1} \otimes \underline{x}) \in \mathbb{T}_{\underline{v}}\mathbb{S}^{d-1}\,, \\ \underline{x} &= \mathrm{Exp}_{\underline{v}}(\underline{\tilde{x}}) = \underline{v} \otimes \mathrm{Exp}_{\mathbb{1}}(\underline{v}^{-1} \otimes \underline{\tilde{x}}) \in \mathbb{S}^{d-1}\,. \end{aligned} \tag{3.10}$$

In [26], we proposed a novel parametric approach for modeling uncertain dual quaternions in a topology-aware manner. There, the Hamilton product is exploited to transfer coordinate systems respecting principal curves of the underlying Bingham distribution to arbitrary points on $\mathbb{S}^3$ via the hyperspherical parallel transport. To give an intuitive illustration of this procedure, we provide the following brief example based on spherical Bingham distributions.

Example 3: *We configure two Bingham distributions* $\mathcal{B}(\mathbf{C}_1)$ *and* $\mathcal{B}(\mathbf{C}_2)$ *as in Sec. 2.2.2 with*

$$\mathbf{C}_1 = -\mathrm{diag}(1,2,3) \quad \textit{and} \quad \mathbf{C}_2 = -\begin{bmatrix} 5 & 0.1 & 0.2 \\ 0.1 & 1 & 0.1 \\ 0.2 & 0.1 & 1 \end{bmatrix},$$

respectively. As pointed out in [35, 36], the column vectors of the matrix $\mathbf{M}$ *in* (2.9) *naturally provide an orthonormal basis of* $\mathbb{R}^d$ *in accordance with principal curves of the underlying dispersion (namely, the Bingham principal basis that will be elaborated shortly after in Sec. 3.3.2).*

As shown in Fig. 3.2, we depict this local basis at the mode of the distribution and perform parallel transport to it along an arbitrarily given smooth curve on $\mathbb{S}^2$*. Drawn in the same color, the basis vectors pointing along the same principal curve (in blue and red) stay parallel with each other on* $\mathbb{S}^2$ *w.r.t. the connection on different trajectories. Meanwhile, vectors stay orthogonal to each other throughout parallel transports on the sphere.*

3.3 Unit Dual Quaternion Manifold

A dual quaternion is defined as $\underline{x} = \underline{x}_\mathrm{r} + \epsilon \underline{x}_\mathrm{s}$, with quaternions $\underline{x}_\mathrm{r}$ and $\underline{x}_\mathrm{s}$ being the real and dual parts, respectively [90]. ϵ denotes the dual unit which satisfies $\epsilon^2 = 0$. By concatenating the real and dual parts, dual quaternions can be formulated into the vector form $\underline{x} = [\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top\,]^\top$. The Corresponding arithmetic of dual quaternions is then a combination of the quaternion arithmetic and the dual number theory [26]. In this section, a concise introduction of basic dual quaternion operations is first provided, after which the manifold of unit dual quaternions will be investigated in the context of SE(3) state representation.

3.3.1 Dual Quaternion Arithmetic and SE(3) State Representation

Similar to quaternions, the mutiplication of two arbitrary dual quaternions $\underline{x} = [\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top\,]^\top$ and $\underline{y} = [\,\underline{y}_\mathrm{r}^\top, \underline{y}_\mathrm{s}^\top\,]^\top$ can be expressed as ordinary matrix-vector multiplications, namely, $\underline{x} \boxtimes \underline{y} = \mathbf{Q}_{\underline{x}}^{\ulcorner}\, \underline{y} = \mathbf{Q}_{\underline{y}}^{\urcorner}\, \underline{x}$, with

$$\mathbf{Q}_{\underline{x}}^{\ulcorner} = \begin{bmatrix} \mathbf{Q}_{\underline{x}_\mathrm{r}}^{\llcorner} & \mathbf{0}_{4\times4} \\ \mathbf{Q}_{\underline{x}_\mathrm{s}}^{\llcorner} & \mathbf{Q}_{\underline{x}_\mathrm{r}}^{\llcorner} \end{bmatrix}, \quad \mathbf{Q}_{\underline{y}}^{\urcorner} = \begin{bmatrix} \mathbf{Q}_{\underline{y}_\mathrm{r}}^{\lrcorner} & \mathbf{0}_{4\times4} \\ \mathbf{Q}_{\underline{y}_\mathrm{s}}^{\lrcorner} & \mathbf{Q}_{\underline{y}_\mathrm{r}}^{\lrcorner} \end{bmatrix}. \tag{3.11}$$

Here, $\mathbf{Q}_{\underline{x}}^{\ulcorner}$ and $\mathbf{Q}_{\underline{y}}^{\urcorner}$ denote the matrix representations of the dual quaternions when multiplied from left- and right-hand sides, respectively. The norm of a dual quaternion $\underline{x}$ is defined as $\sqrt{\underline{x} \boxtimes \underline{x}^*}$, with $\underline{x}^* = [\,(\underline{x}_\mathrm{r}^*)^\top, (\underline{x}_\mathrm{s}^*)^\top]^\top$ being the *classical conjugate* of the dual quaternion $\underline{x}$ (obtained by conjugating the quaternions of real and dual parts individually) [26].

For representing the six-DoF rigid body motion, a dual quaternion $\underline{x} = [\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top]^\top$ can be defined, with its real part being a unit quaternion parameterizing a rotation of angle θ around axis $\underline{n} \in \mathbb{S}^2$ and its dual part encoding a spatial translation $\underline{t} \in \mathbb{R}^3$ in the form

$$\underline{x}_\mathrm{r} = [\,\cos(\theta/2)\,, \underline{n}^\top \sin(\theta/2)\,]^\top \quad \text{and} \quad \underline{x}_\mathrm{s} = 0.5\,[\,0, \underline{t}^\top]^\top \otimes \underline{x}_\mathrm{r}\,, \tag{3.12}$$

respectively. In the dual part $\underline{x}_\mathrm{s}$, the translation term $\underline{t}$ is combined with the real part $\underline{x}_\mathrm{r}$ via the Hamilton product. Given any point $\underline{v} \in \mathbb{R}^3$, it can be transformed by a dual quaternion defined in (3.12) according to

$$\underline{v}' = \left(\underline{x} \boxtimes [\,1,0,0,0,0,\underline{v}^\top]^\top \boxtimes \underline{x}^{\odot}\right)_{6:8}, \tag{3.13}$$

with $\underline{x}^{\odot} = [\,(\underline{x}_{\mathrm{r}}^{*})^{\top}, -(\underline{x}_{\mathrm{s}}^{*})^{\top}\,]^{\top}$ denoting the *full conjugate* (conjugating both the dual unit ϵ and the two component quaternions) of $\underline{x}$. The operator $()_{6:8}$ is hereby applied to recover the transformed points by extracting the last three entries of its coordinates in the form of dual quaternion. A proof of (3.13) has been given in [26, Appendix D]. The operation has a similar formulation to performing a rotation using a unit quaternion in (3.7). With both the rotation and translation terms parameterized together in an eight-dimensional vector, dual quaternions exhibit no ambiguity (compared with Euler angle-based representations) and much less redundancy than the well-known 4×4 transformation matrices. Also, similar to the antipodal symmetry of unit quaternions, a dual quaternion $\underline{x}$ and its antipode $-\underline{x}$ represent the same rigid body motion as indicated by (3.13).

3.3.2 Geometric Structure of Unit Dual Quaternion Manifold

Dual quaternions of unit norm, i.e., the unit dual quaternions, are confined to the following manifold embedded in the eight-dimensional Euclidean space

$$\mathbb{DH}_1 = \left\{ [\,\underline{x}_{\mathrm{r}}^{\top}, \underline{x}_{\mathrm{s}}^{\top}\,]^{\top} \,\middle|\, \underline{x}_{\mathrm{r}} \in \mathbb{S}^3, \underline{x}_{\mathrm{r}}^{\top}\underline{x}_{\mathrm{s}} = 0 \right\} \subset \mathbb{R}^8 \,. \tag{3.14}$$

A proof of the manifold definition above is provided in Appendix A.1. Moreover, dual quaternions representing SE(3) states naturally fulfill the constraints of unit norm and are therefore located on the manifold $\mathbb{DH}_1$ (a proof is provided in Appendix A.2). The two additional constraints imposed on the eight-dimensional vector in (3.14) also cohere with the fact that the rigid body motions have six DoF. Geometrically speaking, the real part of any unit dual quaternion is located on the unit hypersphere $\mathbb{S}^3$, whereas the dual part is perpendicular to the real part. In other words, given any unit dual quaternion $\underline{x} = [\,\underline{x}_{\mathrm{r}}^{\top}, \underline{x}_{\mathrm{s}}^{\top}\,]^{\top}$, the dual part belongs to the tangent space at the real part on the unit hypersphere, namely, $\underline{x}_{\mathrm{s}} \in \mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3$.

Composing the unit hypersphere with the Euclidean space in a covariant manner, the unit dual quaternion manifold will be later exploited to showcase probabilistic modeling and filtering for composite directional domains. Investigating its geometric structure is nontrivial since the dual part is coupled with the real part located on a nonlinear manifold. Therefore, we aim to decouple this nonlinear relationship between the real and dual parts, and more precisely, to express the dual part in a local basis of the tangent space at the real part, such that modeling the dual part can be executed in a linear fashion. Selection of an orthonormal basis spanning the tangent space $\mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3$ is theoretically arbitrary. For the operational convenience of on-manifold probabilistic modeling and filtering, we introduce two methods for establishing local bases of $\mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3$ in accordance with the manifold structure of $\mathbb{DH}_1$ [26, 35, 37, 50].

Bingham principal basis: As introduced in Sec. 2.2.2, the Bingham distribution is commonly used for parametric modeling of uncertain unit quaternions [32, 36] and can thus be exploited for modeling the real part of unit dual quaternions [26]. Given a quaternion Bingham distribution $\mathcal{B}(\mathbf{M}, \mathbf{Z})$ on $\mathbb{S}^3$, the parameter matrix $\mathbf{M} = [\,\underline{m}_1, \underline{m}_2, \underline{m}_3, \underline{m}\,] \in \mathbb{R}^{4\times4}$ that controls the orientation of dispersion is real orthogonal[1], with each column pointing along a principal curve of the distribution. Therefore, the first three columns of $\mathbf{M}$ naturally provide an orthonormal basis of the tangent space at the mode of the distribution on $\mathbb{S}^3$, namely,

$$\mathbb{T}_{\underline{m}}\mathbb{S}^3 = \mathrm{Span}(\mathcal{E}_{\mathrm{m}}^{\mathcal{B}}) \,, \quad \text{with} \quad \mathcal{E}_{\mathrm{m}}^{\mathcal{B}} = \{\underline{m}_1, \underline{m}_2, \underline{m}_3\} \tag{3.15}$$

[1] As introduced in Sec. 2.2.2, columns of $\mathbf{M}$ is adjusted according to their corresponding concentrations in ascending order with the last one indicating the Bingham mode.

denoting the so-called *Bingham principal basis* (BPB)[2]. We concatenate the basis vectors column wise into a matrix $\mathbf{E}_{\mathrm{m}}^{\mathcal{B}} = [\,\underline{m}_1, \underline{m}_2, \underline{m}_3\,] \in \mathbb{R}^{4\times 3}$ such that any point $\tilde{\underline{x}} \in \mathbb{T}_{\underline{m}}\mathbb{S}^3$ can be expressed w.r.t. $\mathcal{E}_{\mathrm{m}}^{\mathcal{B}}$ in its local coordinates via $\tilde{\underline{x}}^{\iota} = (\mathbf{E}_{\mathrm{m}}^{\mathcal{B}})^{\top}\tilde{\underline{x}}$. Moreover, we have $(\mathbf{E}_{\mathrm{m}}^{\mathcal{B}})^{\top}\mathbf{E}_{\mathrm{m}}^{\mathcal{B}} = \mathbf{I}_{3\times 3}$, given that matrix $\mathbf{M}$ is real orthogonal.

As introduced in Example 3 and Fig. 3.2, the BPB can be transferred from the Bingham mode $\underline{m} \in \mathbb{S}^3$ to any given quaternion $\underline{x}_{\mathrm{r}} \in \mathbb{S}^3$ via hyperspherical parallel transport. On $\mathbb{S}^3$, this procedure can be conveniently implemented via the SO(4) rotation induced by the Hamilton product, namely,

$$\mathbf{E}_{\mathrm{x_r}}^{\mathcal{B}} = \mathbf{Q}^{\llcorner}_{\underline{x}_{\mathrm{r}}\otimes\underline{m}^{-1}}\mathbf{E}_{\mathrm{m}}^{\mathcal{B}} = \mathbf{Q}^{\llcorner}_{\underline{x}_{\mathrm{r}}}(\mathbf{Q}^{\llcorner}_{\underline{m}})^{\top}\mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\,, \tag{3.16}$$

as proposed in [26]. Thus, (3.16) provides a realization to characterizing the Bingham dispersion at any location on $\mathbb{S}^3$ while preserving its principal curve directions in the sense of parallel transport. Exploiting the transported BPB to span a tangent space on $\mathbb{S}^{d-1}$ essentially endows the in-tangent-space geometric structure with interpretation of the on-manifold uncertainty. For probabilistic modeling on composite directional manifolds, this provides a solution to establishing the probabilistic correlation between the hyperspherical and linear component domains (an introduction is detailed in Sec. 4.1).

Based on any orthonormal basis of the tangent space at $\underline{x}_{\mathrm{r}} \in \mathbb{S}^3$, the orthographic projection to $\mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3$ in Sec. 3.1.2 can be transformed to multiplication with the corresponding *projection matrix*. Without loss of generality, the projection matrix w.r.t. $\mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3$ can be obtained using the transported BPB in (3.16) according to

$$\mathbf{P}_{\mathrm{x_r}} = \mathbf{E}_{\mathrm{x_r}}^{\mathcal{B}}(\mathbf{E}_{\mathrm{x_r}}^{\mathcal{B}})^{\top} \in \mathbb{R}^{4\times 4}\,. \tag{3.17}$$

Given any $\underline{v} \in \mathbb{R}^4$, its projected coordinates on $\mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3$ follow $\tilde{\underline{v}} = \mathbf{P}_{\mathrm{x_r}}\underline{v} \in \mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3$. Furthermore, it is trivial to verify that $(\mathbf{E}_{\mathrm{x_r}}^{\mathcal{B}})^{\top}\mathbf{E}_{\mathrm{x_r}}^{\mathcal{B}} = \mathbf{I}_{3\times 3}$ and $\underline{x}_{\mathrm{r}}\underline{x}_{\mathrm{r}}^{\top} + \mathbf{P}_{\mathrm{x_r}} = \mathbf{I}_{4\times 4}$ hold.

Quaternion concomitant basis: As introduced in Sec. 3.2, the matrix representations of unit quaternions shown in (3.5) belong to the SO(4) group. Taking the last three columns from matrix $\mathbf{Q}^{\circ}_{\underline{x}_r}$ (either left- or right-hand expressions) given a unit quaternion $\underline{x}_{\mathrm{r}} \in \mathbb{S}^3$, we obtain one pair of *quaternion concomitant bases* (QCBs)[3], namely,

$$\mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3 = \mathrm{Span}(\mathcal{A}^{\circ}_{\mathrm{x_r}})\,, \quad \text{with} \quad \mathcal{A}^{\circ}_{\mathrm{x_r}} = \{\underline{e}_1, \underline{e}_2, \underline{e}_3\} \quad \text{and} \quad \mathbf{Q}^{\circ}_{\underline{x}_{\mathrm{r}}} =: [\,\underline{x}_{\mathrm{r}}, \underline{e}_1, \underline{e}_2, \underline{e}_3\,]\,. \tag{3.18}$$

We also concatenate basis vectors in $\mathcal{A}_{\mathrm{x_r}}$ column wise into a matrix $\mathbf{A}^{\circ}_{\mathrm{x_r}} \in \mathbb{R}^{4\times 3}$ and obtain $\mathbf{Q}^{\circ}_{\underline{x}_{\mathrm{r}}} = [\,\underline{x}_{\mathrm{r}}, \mathbf{A}^{\circ}_{\mathrm{x_r}}]$, with $(\mathbf{A}^{\circ}_{\mathrm{x_r}})^{\top}\mathbf{A}^{\circ}_{\mathrm{x_r}} = \mathbf{I}_{3\times 3}$, because $\mathbf{Q}^{\circ}_{\underline{x}_r} \in \mathrm{SO}(4)$. When expressing the dual part of a unit dual quaternion representing a SE(3) state as defined in (3.12) w.r.t. the QCB $\mathcal{A}^{\lrcorner}_{\mathrm{x_r}}$, we obtain a scaled translation term $\underline{t}$ according to the following formula [50]

$$\underline{t} = (2\,\underline{x}_{\mathrm{s}} \otimes \underline{x}_{\mathrm{r}}^{-1})_{2:4} = 2\,(\mathbf{A}^{\lrcorner}_{\mathrm{x_r}})^{\top}\underline{x}_{\mathrm{s}}\,.$$

3.3.3 Planar Dual Quaternions Representing SE(2) States

For parameterizing planar rigid body motions belonging to the SE(2) group, unit dual quaternions can be employed in a degenerate form. In this thesis, we name dual quaternions representing SE(2) states as planar dual quaternions.

[2] For brevity, we write an orthonormal basis $\mathcal{E}_{\underline{m}}$ at $\underline{m}$ as $\mathcal{E}_{\mathrm{m}}$ (same for $\mathbf{E}_{\underline{m}}$ and $\mathbf{E}_{\mathrm{m}}$).

[3] The name indicates that the basis inherently comes along with a quaternion and does not rely on external settings.

More specifically, a planar motion is conventionally composed of a rotation around the z-axis and a translation $\underline{t} = t_x\,\mathsf{i} + t_y\,\mathsf{j}$ on the x–y coordinate plane. For parameterizations using dual quaternions, the real part is defined as the quaternion $\underline{x}_\mathrm{r} = \cos(\theta/2) + \mathsf{k}\sin(\theta/2)$ and is expressed in the vector form $\underline{x}_\mathrm{r} = [\,\cos(\theta/2), \sin(\theta/2)\,]^\top \in \mathbb{S}^1 \subset \mathbb{R}^2$. The dual part is defined in a specified form for the planar case according to (3.12), namely,

$$\underline{x}_\mathrm{s} = \frac{1}{2}\,\underline{t} \otimes \underline{x}_\mathrm{r} = \frac{1}{2}\mathbf{Q}^{\lrcorner}_{\underline{x}_\mathrm{r}}\,\underline{t}\,, \quad \text{with} \quad \mathbf{Q}^{\lrcorner}_{\underline{x}_\mathrm{r}} = \begin{bmatrix} x_{r,1} & x_{r,2} \\ -x_{r,2} & x_{r,1} \end{bmatrix} \tag{3.19}$$

being the right-matrix representation of the real part by customizing (3.5) w.r.t. planar rotations [40, 81]. Therefore, planar dual quaternions in the form $\underline{x} = [\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top\,]^\top$ belong to the Cartesian product of the unit circle and the two-dimensional Euclidean space, i.e., $\underline{x} \in \mathbb{S}^1 \times \mathbb{R}^2 \subset \mathbb{R}^4$ [40, 81]. For two arbitrary planar dual quaternions $\underline{x} = [\,x_0, x_1, x_2, x_3\,]^\top$ and $\underline{y} = [\,y_0, y_1, y_2, y_3\,]^\top$, their product can be derived by modifying (3.11) as $\underline{x} \boxtimes \underline{y} = \mathbf{Q}^{\ulcorner}_{\underline{x}}\,\underline{y} = \mathbf{Q}^{\urcorner}_{\underline{y}}\,\underline{x}$, with

$$\mathbf{Q}^{\ulcorner}_{\underline{x}} = \begin{bmatrix} x_0 & -x_1 & 0 & 0 \\ x_1 & x_0 & 0 & 0 \\ x_2 & x_3 & x_0 & -x_1 \\ x_3 & -x_2 & x_1 & x_0 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}^{\urcorner}_{\underline{y}} = \begin{bmatrix} y_0 & -y_1 & 0 & 0 \\ y_1 & y_0 & 0 & 0 \\ y_2 & -y_3 & y_0 & y_1 \\ y_3 & y_2 & -y_1 & y_0 \end{bmatrix} \tag{3.20}$$

being the left- and right-matrix representations of $\underline{x}$ and $\underline{y}$, respectively. For any point $\underline{v} \in \mathbb{R}^2$, it can be transformed by a given planar dual quaternion $\underline{x} = [\,x_0, x_1, x_2, x_3\,]^\top$ according to

$$\underline{v}' = \left(\underline{x} \boxtimes [\,1, 0, \underline{v}^\top]^\top \boxtimes \underline{x}^{\odot}\right)_{3:4}, \tag{3.21}$$

with $\underline{x}^{\odot}$ being the full conjugate of $\underline{x}$ for the planar case. Here, $\underline{v}$ is first converted into the form of a planar dual quaternion, and the operator $()_{3:4}$ extracts the last two entries of a vector to recover the transformed coordinates $\underline{v}' \in \mathbb{R}^2$.

Exploiting dual quaternions to represent planar motions enables modeling rotations and translations simultaneously in a joint vector space. Based on this concise form, a PDF was proposed in [55] for modeling uncertain planar dual quaternions while inherently considering the probabilistic correlation between the real and dual parts. The distribution has been introduced in Sec. 2.2.3 and depicted in Fig. 2.2.

CHAPTER 4

On-Manifold Upgrade for Parametric Directional Filtering

Considering the limitations mentioned in Sec. 1.2.1 and Sec. 1.2.2, we aim to upgrade parametric probabilistic modeling and filtering on directional manifolds in this chapter. More specifically, three generic methodologies are introduced: (1) topology-aware modeling in composite directional domains based on hyperspherical parallel transport in Sec. 4.1, (2) on-manifold configurable deterministic sampling for parametric directional models (such as the von Mises–Fisher distribution and the Bingham distribution) in Sec. 4.2, and (3) deterministic progressive update for nonlinear directional estimation with non-identity measurement models in Sec. 4.3. We showcase the first upgrade on the manifold of unit dual quaternions. Furthermore, the latter two enhancements are validated in a simulated scenario of quaternion-based nonlinear SO(3) estimation in Sec. 4.4.

4.1 Upgrade I: Parametric Modeling for Composite Directional Domains

One of the key problems of parametric modeling on composite directional manifolds is to consider the correlation across individual component domains in a topology-aware manner. The methods introduced here are tailored to unit hyperspheres combined with Euclidean spaces through a Cartesian product. Other types of composite directional manifolds, e.g., tori and hypertori, are out of the scope of this thesis. For demonstration, we explicitly consider the unit dual quaternion manifold introduced in Sec. 3.3.2 throughout this section.

4.1.1 Bingham–Gaussian Modeling Scheme Based on Parallel Transport

Given an arbitrary unit dual quaternion $\underline{x} = [\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top]^\top \in \mathbb{DH}_1$ as defined in (3.14), we decompose its underlying probability density as

$$f(\underline{x}_\mathrm{r}, \underline{x}_\mathrm{s}) = f(\underline{x}_\mathrm{r}) f(\underline{x}_\mathrm{s} | \underline{x}_\mathrm{r}) \,, \tag{4.1}$$

where the real part is assumed to be Bingham-distributed, namely, $\underline{x}_\mathrm{r} \in \mathcal{B}(\mathbf{M}_\mathrm{r}, \mathbf{Z}_\mathrm{r})$, with $\mathbf{M}_\mathrm{r}$ indicating its principal curve directions as shown by (3.15). For modeling the conditional part $\underline{x}_\mathrm{s} \in \mathbb{T}_{\underline{x}_\mathrm{r}} \mathbb{S}^3$, we first express it w.r.t. a local basis[1] $\mathcal{E}_\mathrm{r}$ of $\mathbb{T}_{\underline{x}_\mathrm{r}} \mathbb{S}^3$ for decoupling from the nonlinear real part and obtain its local coordinates $\underline{x}_\mathrm{s}^\iota = \mathbf{E}_\mathrm{r}^\top \underline{x}_\mathrm{s} \in \mathbb{R}^3$. The choice of the local basis is in theory arbitrary. We thus take this chance and design a basis $\mathcal{E}_\mathrm{r}$ with consideration of the hyperspherical dispersion such that its probabilistic correlation with the uncertainty in the coupled Euclidean

[1] We denote $\mathcal{E}_{\mathrm{x_r}}$ in Sec. 3.3.2 as $\mathcal{E}_\mathrm{r}$ in this chapter for brevity.

space can be established. As proposed in [26], we set up the Bingham principal basis $\mathcal{E}_{\mathrm{r}}^{\mathcal{B}}$ at $\underline{x}_{\mathrm{r}}$ using the hyperspherical parallel transport (HPT) enabled by the Hamilton product as introduced in (3.16). Based thereon, the conditional density in (4.1) follows

$$\begin{aligned} f(\underline{x}_{\mathrm{s}}|\underline{x}_{\mathrm{r}}) &= \int_{\mathbb{R}^3} f(\underline{x}_{\mathrm{s}}, \underline{x}_{\mathrm{s}}^{\iota}|\underline{x}_{\mathrm{r}}) \,\mathrm{d}\underline{x}_{\mathrm{s}}^{\iota} \\ &= \int_{\mathbb{R}^3} f(\underline{x}_{\mathrm{s}}|\underline{x}_{\mathrm{s}}^{\iota}, \underline{x}_{\mathrm{r}}) f(\underline{x}_{\mathrm{s}}^{\iota}|\underline{x}_{\mathrm{r}}) \,\mathrm{d}\underline{x}_{\mathrm{s}}^{\iota} \\ &= \int_{\mathbb{R}^3} \delta(\underline{x}_{\mathrm{s}} - \mathbf{E}_{\mathrm{r}}^{\mathcal{B}} \underline{x}_{\mathrm{s}}^{\iota}) f(\underline{x}_{\mathrm{s}}^{\iota}|\underline{x}_{\mathrm{r}}) \,\mathrm{d}\underline{x}_{\mathrm{s}}^{\iota}. \end{aligned} \tag{4.2}$$

Since the dual part $\underline{x}_{\mathrm{s}}$ is in the tangent space at $\underline{x}_{\mathrm{r}} \in \mathbb{S}^3$, a multiplication with the projection matrix in (3.17) does not change its coordinates, i.e., $\underline{x}_{\mathrm{s}} = \mathbf{P}_{\mathrm{r}} \underline{x}_{\mathrm{s}}$. Following the multivariate extension of the Dirac delta function [91], we have

$$\begin{aligned} \delta(\underline{x}_{\mathrm{s}} - \mathbf{E}_{\mathrm{r}}^{\mathcal{B}} \underline{x}_{\mathrm{s}}^{\iota}) &= \delta\big(\mathbf{P}_{\mathrm{r}} \underline{x}_{\mathrm{s}} - \mathbf{E}_{\mathrm{r}}^{\mathcal{B}} \underline{x}_{\mathrm{s}}^{\iota}\big) \\ &= \delta\big(\mathbf{E}_{\mathrm{r}}^{\mathcal{B}} (\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \underline{x}_{\mathrm{s}} - \mathbf{E}_{\mathrm{r}}^{\mathcal{B}} \underline{x}_{\mathrm{s}}^{\iota}\big) \\ &= \det\big((\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \mathbf{E}_{\mathrm{r}}^{\mathcal{B}}\big)^{-0.5} \cdot \delta(\underline{x}_{\mathrm{s}}^{\iota} - (\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \underline{x}_{\mathrm{s}}) \,. \end{aligned} \tag{4.3}$$

For the basis matrix $\mathbf{E}_{\mathrm{r}}^{\mathcal{B}} \in \mathbb{R}^{4\times 3}$, $(\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \mathbf{E}_{\mathrm{r}}^{\mathcal{B}} = \mathbf{I}_{3\times 3}$ holds as introduced in Sec. 3.3.2. We thus obtain $\delta(\underline{x}_{\mathrm{s}} - \mathbf{E}_{\mathrm{r}}^{\mathcal{B}} \underline{x}_{\mathrm{s}}^{\iota}) = \delta\big(\underline{x}_{\mathrm{s}}^{\iota} - (\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \underline{x}_{\mathrm{s}}\big)$ for (4.3). As a result, (4.2) is reduced to

$$f(\underline{x}_{\mathrm{s}}|\underline{x}_{\mathrm{r}}) = \int_{\mathbb{R}^3} \delta\big(\underline{x}_{\mathrm{s}}^{\iota} - (\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \underline{x}_{\mathrm{s}}\big) f(\underline{x}_{\mathrm{s}}^{\iota}|\underline{x}_{\mathrm{r}}) \,\mathrm{d}\underline{x}_{\mathrm{s}}^{\iota} = f\big((\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \underline{x}_{\mathrm{s}}|\underline{x}_{\mathrm{r}}\big).$$

Consequently, the density in (4.1) can be rewritten as

$$f(\underline{x}_{\mathrm{r}}, \underline{x}_{\mathrm{s}}) = f(\underline{x}_{\mathrm{r}}) f\big((\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \underline{x}_{\mathrm{s}}|\underline{x}_{\mathrm{r}}\big) \,,$$

with $\mathbf{E}_{\mathrm{r}}^{\mathcal{B}}$ representing the BPB transferred from the Bingham mode $\underline{m} \in \mathbb{S}^3$ to $\underline{x}_{\mathrm{r}} \in \mathbb{S}^3$ via HPT. By further assuming that the conditioned dual part expressed w.r.t. basis $\mathcal{E}_{\mathrm{r}}^{\mathcal{B}}$ follows a Gaussian distribution in $\mathbb{R}^3$, namely, $(\mathbf{E}_{\mathrm{r}}^{\mathcal{B}})^{\mathsf{T}} \underline{x}_{\mathrm{s}} \sim \mathcal{N}(\underline{\tau}_{\mathrm{s}}, \mathbf{\Sigma}_{\mathrm{s}})$, the probabilistic correlation between the real and dual parts is interpreted. Consequently, the parallel transport-based Bingham–Gaussian (PTBG) modeling scheme is established on the unit dual quaternion manifold $\mathbb{DH}_1$ in a topology-aware manner.

In principle, the PTBG scheme proposed for the unit dual quaternion manifold is applicable to any directional domain constructed in a similar fashion. In this regard, certain modifications might be necessary, such as choosing a variable-specific hyperspherical PDF and a domain-specific implementation of the hyperspherical parallel transport (here handily done by the quaternion product). Another approach to the considered problem is to marginalize out the linear state variable conditioned on the hyperspherical state variable via the Rao–Blackwellization, which is well-established for marginalized particle filters [92]. Though this methodology has more popularity for nonparametric filtering, it may also have good potential for adaption to parametric modeling for the considered composite directional domains. Based on the theoretical setup of the PTBG scheme above, we further propose the following preliminary techniques to pave its way to recursive estimation on the unit dual quaternion manifold.

Parameter matching: Given a set of weighted unit dual quaternion samples $\{(\underline{x}_i, \nu_i)\}_{i=1}^n$ with $\underline{x}_i = [\,\underline{x}_{i,\mathrm{r}}^{\mathsf{T}}, \underline{x}_{i,\mathrm{s}}^{\mathsf{T}}\,]^{\mathsf{T}} \in \mathbb{DH}_1$ and weights fulfilling $\sum_{i=1}^n \nu_i = 1$, parameters of the Bingham and Gaussian components of the PTBG scheme can be matched as detailed in Algorithm 2 [26]. We

Algorithm 2: Parameter Matching for PTBG Scheme

Input: weighted sample set $\{(\underline{x}_i, \nu_i)\}_{i=1}^{n}$
Output: Bingham component density $f_{\mathcal{B}}(\mathbf{M}_{\mathrm{r}}, \mathbf{Z}_{\mathrm{r}})$, Gaussian component density $f_{\mathcal{N}}(\underline{\tau}_{\mathrm{s}}, \mathbf{\Sigma}_{\mathrm{s}})$

1 $f_{\mathcal{B}} \leftarrow \texttt{matchBingham}\,(\{(\underline{x}_{i,\mathrm{r}}, \nu_i)\}_{i=1}^{n})$;
2 $[\,\mathbf{E}_{\mathrm{m}}^{\mathcal{B}}, \underline{m}\,] \leftarrow \mathbf{M}_{\mathrm{r}}$;
3 **for** $i \leftarrow 1$ **to** n **do**
4 $\quad \mathbf{E}_{\mathrm{r},i}^{\mathcal{B}} \leftarrow \mathbf{Q}_{\underline{x}_{i,\mathrm{r}}}^{\llcorner}(\mathbf{Q}_{\underline{m}}^{\llcorner})^{\top}\mathbf{E}_{\mathrm{m}}^{\mathcal{B}}$; // see (3.16)
5 $\quad \underline{x}_{i,\mathrm{s}}^{\iota} \leftarrow (\mathbf{E}_{\mathrm{r},i}^{\mathcal{B}})^{\top}\underline{x}_{i,\mathrm{s}}$;
6 $\underline{\tau}_{\mathrm{s}} \leftarrow \sum_{i=1}^{n} \nu_i\, \underline{x}_{i,\mathrm{s}}^{\iota}$;
7 $\mathbf{\Sigma}_{\mathrm{s}} \leftarrow \sum_{i=1}^{n} \nu_i\, (\underline{x}_{i,\mathrm{s}}^{\iota} - \underline{\tau}_{\mathrm{s}})\,(\underline{x}_{i,\mathrm{s}}^{\iota} - \underline{\tau}_{\mathrm{s}})^{\top}$;
8 **return** $f_{\mathcal{B}}(\mathbf{M}_{\mathrm{r}}, \mathbf{Z}_{\mathrm{r}}), f_{\mathcal{N}}(\underline{\tau}_{\mathrm{s}}, \mathbf{\Sigma}_{\mathrm{s}})$;

Algorithm 3: Deterministic Sampling for PTBG Scheme

Input: Bingham component density $f_{\mathcal{B}}(\mathbf{M}_{\mathrm{r}}, \mathbf{Z}_{\mathrm{r}})$, Gaussian component density $f_{\mathcal{N}}(\underline{\tau}_{\mathrm{s}}, \mathbf{\Sigma}_{\mathrm{s}})$
Output: weighted sample set $\{(\underline{x}_i, \nu_i)\}_{i=1}^{n}$

1 $\{(\underline{x}_{j,\mathrm{r}}, \nu_{j,\mathrm{r}})\}_{j=1}^{n_{\mathrm{r}}} \leftarrow \texttt{sampleDeterministic}\,(f_{\mathcal{B}})$;
2 $\{(\underline{x}_{k,\mathrm{s}}^{\iota}, \nu_{k,\mathrm{s}})\}_{k=1}^{n_{\mathrm{s}}} \leftarrow \texttt{sampleDeterministic}\,(f_{\mathcal{N}})$;
3 $[\,\mathbf{E}_{\mathrm{m}}^{\mathcal{B}}, \underline{m}\,] \leftarrow \mathbf{M}_{\mathrm{r}}$;
4 **for** $j \leftarrow 1$ **to** n_r **do**
5 $\quad \mathbf{E}_{\mathrm{r},j}^{\mathcal{B}} \leftarrow \mathbf{Q}_{\underline{x}_{j,\mathrm{r}}}^{\llcorner}(\mathbf{Q}_{\underline{m}}^{\llcorner})^{\top}\mathbf{E}_{\mathrm{m}}^{\mathcal{B}}$; // see (3.16)
6 $\quad$ **for** $k \leftarrow 1$ **to** n_s **do**
7 $\quad\quad \underline{x}_{k,\mathrm{s}} \leftarrow \mathbf{E}_{\mathrm{r},j}^{\mathcal{B}}\, \underline{x}_{k,\mathrm{s}}^{\iota}$;
8 $\quad\quad \underline{x}_i \leftarrow [\,\underline{x}_{j,\mathrm{r}}^{\top}, \underline{x}_{k,\mathrm{s}}^{\top}\,]^{\top}$;
9 $\quad\quad \nu_i \leftarrow \nu_{j,\mathrm{r}} \cdot \nu_{k,\mathrm{s}}$;
10 **return** $\{(\underline{x}_i, \nu_i)\}_{i=1}^{n_{\mathrm{r}} \cdot n_{\mathrm{s}}}$;

first approximate the Bingham density using all weighted real-part samples as proposed in [32], after which the BPB is extracted from the obtained parameter matrix $\mathbf{M}_{\mathrm{r}}$ (Algorithm 2, lines 1–2). For each sample $\underline{x}_i$, we span the tangent space at its real part $\mathbb{T}_{\underline{x}_{i,\mathrm{r}}}\mathbb{S}^3$ with the local BPB basis $\mathcal{E}_{\mathrm{r},i}^{\mathcal{B}}$ as given in (3.16), w.r.t. which the dual part $\underline{x}_{i,\mathrm{s}}$ is expressed in its local coordinates $\underline{x}_{i,\mathrm{s}}^{\iota}$ (Algorithm 2, lines 3–5). We then collect all the rotation-invariant dual parts $\{\underline{x}_{i,\mathrm{s}}^{\iota}\}_{i=1}^{n}$, based on which the parameters of the Gaussian component are obtained (Algorithm 2, lines 6–7).

Deterministic sampling: As discussed in Algorithm 1, deterministic samples are typically required in system propagation or measurement update by parametric directional filters for most nonlinear estimation tasks. Given a set of PTBG parameters, we provide the deterministic sampling scheme formulated in Algorithm 3 [26]. As shown in lines 1–2, weighted deterministic samples are first drawn individually from the Bingham and Gaussian components with cardinalities n_{r} and n_{s}, respectively. Afterward, we establish the local BPB for the tangent space at each real part sample $\underline{x}_{j,\mathrm{r}} \in \mathbb{S}^3$, where all the Gaussian samples $\{\underline{x}_{k,\mathrm{s}}^{\iota}\}_{k=1}^{n_{\mathrm{s}}}$ w.r.t. the local basis $\mathcal{E}_{\mathrm{r},j}^{\mathcal{B}}$ are transformed to their global coordinates (Algorithm 3, lines 3–9). The combination of the real and dual parts is done in the fashion of Cartesian product, thereby resulting in overall $n_{\mathrm{r}} \cdot n_{\mathrm{s}}$ samples on the

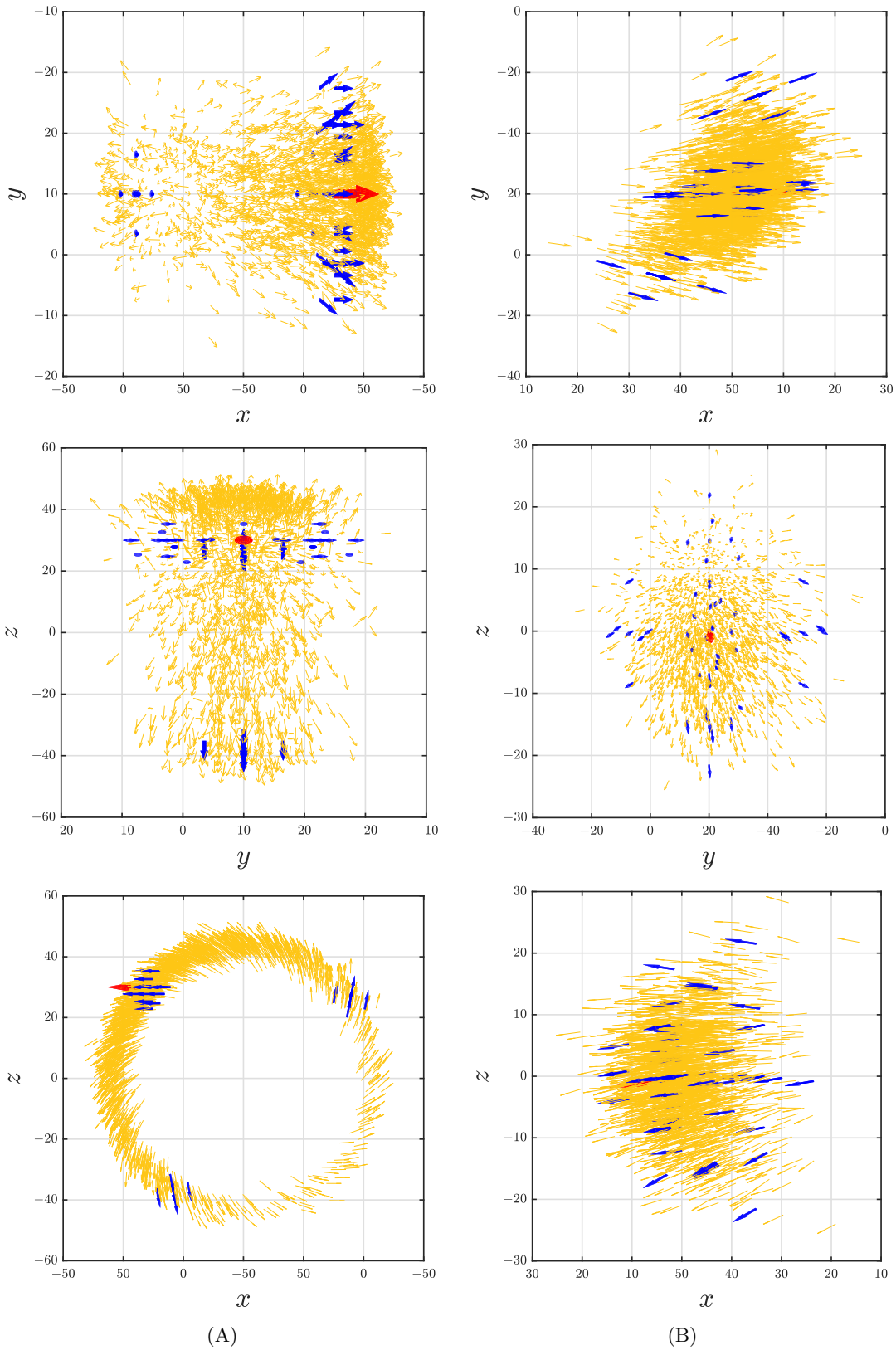


Figure 4.1.: Random and deterministic unit dual quaternion samples drawn according to the two PTBG configurations in Example 4 (shown from different view angles in column (A) and (B)). Yellow and blue arrows depict random and deterministic samples, respectively, whereas the red one denotes the mode.

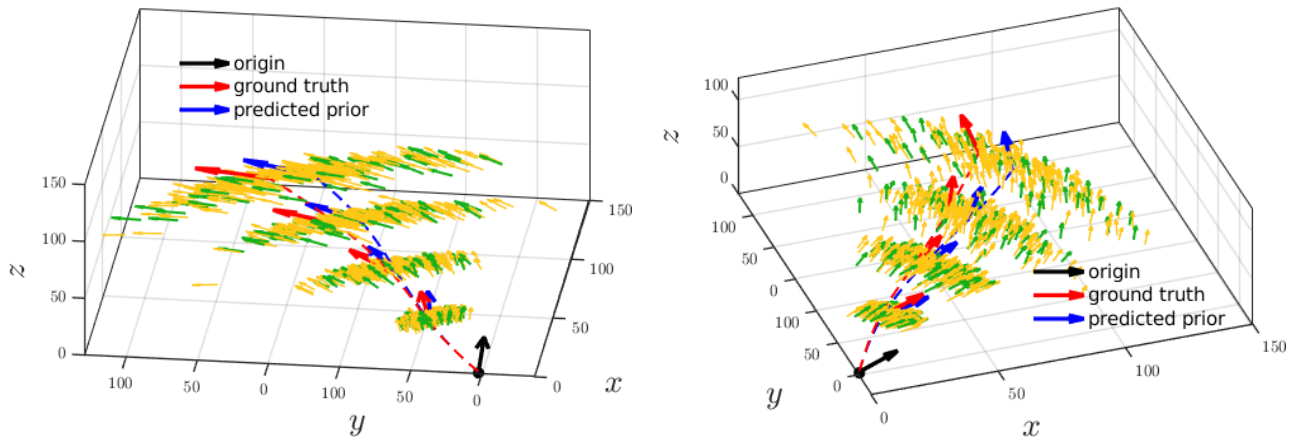


Figure 4.2.: Consecutive PTBG-based predictions on $\mathbb{DH}_1$ (viewed from two different angles). Random and deterministic samples are depicted by yellow and green arrows, respectively.

manifold of unit dual quaternions. We showcase the PTBG modeling scheme on $\mathbb{DH}_1$ with different configurations in the following example.

Example 4: *Two PTBG configurations are provided. The Bingham components* $\mathcal{B}(\mathbf{C})$ *are parameterized in the concise form given in Sec. 2.2.2 with* $\mathbf{C} \in \{-\mathrm{diag}(5,100,8,100), -\mathrm{diag}(1,50,50,50)\}$, *thus having both modes at the identity quaternion* $\mathbb{1} = [1,0,0,0]^\top \in \mathbb{S}^3$. *Additionally, we manipulate the mode of the second distribution to be a unit quaternion parameterizing a spatial rotation of angle* $\pi/12$ *around axis* $[1,15,1]^\top$ *(after normalization) according to* (3.6). *For the Gaussian components* $\mathcal{N}(\underline{\tau}_s, \mathbf{\Sigma}_s)$, *we set*

$$\underline{\tau}_s \in \left\{ [15,15,0]^\top, [-15,5,0]^\top \right\} \quad \textit{and} \quad \mathbf{\Sigma}_s \in \left\{ \mathrm{diag}(1,2,3)\,, \begin{bmatrix} 2 & 0.2 & 0.3 \\ 0.2 & 1 & 0.1 \\ 0.3 & 0.1 & 5 \end{bmatrix} \right\}.$$

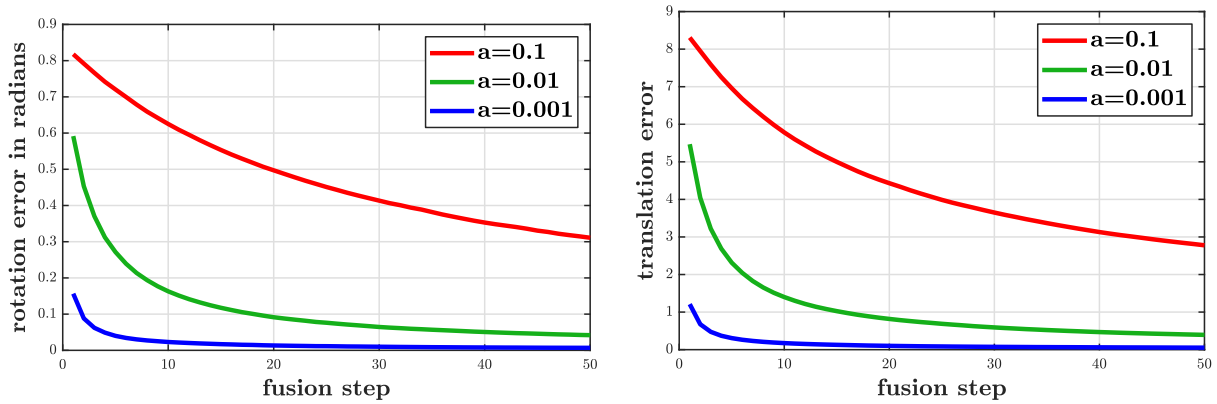
As shown in Fig. 4.1, random (yellow) and deterministic (blue) samples are drawn according to the two PTBG configurations in column (A) and (B) (each viewed from different angles). Unit dual quaternion samples are visualized by transforming an initial orientation vector $[1,0,0]^\top$ *at the origin according to* (3.13). *We use the basic UT-based methods for deterministic sampling of the Bingham and Gaussian components [8, 32, 93], thereby leading to a cardinality of* 49 *for the dual quaternion sample sets. We also combine the Bingham and Gaussian modes for showing modes of the sample sets (depicted in red). It is evident that the PTBG scheme is able to model uncertain translations and rotations simultaneously with consideration of their correlations.*

Based on the proposed deterministic sampling and parameter matching approaches, a PTBG-based prediction step can be established for recursive estimation on $\mathbb{DH}_1$ by following the basic sampling–matching scheme given in Sec. 2.2.4. We briefly demonstrate this procedure in Fig. 4.2 using an identity system model $\underline{\boldsymbol{x}}_{t+1} = \underline{\boldsymbol{x}}_t \boxtimes \underline{u} \boxtimes \underline{\boldsymbol{w}}_t$, where we assume a constant step-wise input $\underline{u} \in \mathbb{DH}_1$ parameterizing a rotation of angle $\pi/4$ around the axis $1/\sqrt{3} \cdot [1,1,1]^\top$ and a translation $[15,15,15]^\top$. The system noise is simulated by the proposed PTBG scheme with a configuration of the Bingham component $\mathcal{B}(-\mathrm{diag}(1,400,400,400))$ and the Gaussian component $\mathcal{N}(\underline{0}_3, 0.1 \cdot \mathbf{I}_{3\times 3})$. Four consecutive prediction steps are simulated. With reference to the ground truth, the sampling–matching-based system propagation functions effectively.

Algorithm 4: Measurement Fusion for PTBG Scheme

Input: prior density $f^{\mathrm{p}}_{\mathcal{B}}$ and $f^{\mathrm{p}}_{\mathcal{N}}$, noise density $f^{\underline{\boldsymbol{v}}}_{\mathcal{B}}$ and $f^{\underline{\boldsymbol{v}}}_{\mathcal{N}}$, measurement $\hat{\underline{z}}$
Output: posterior density $f^{\mathrm{e}}_{\mathcal{B}}$ and $f^{\mathrm{e}}_{\mathcal{N}}$

1 $f^{\mathrm{e}}_{\mathcal{B}}(\mathbf{M}^{\mathrm{e}}_{\mathrm{r}}, \mathbf{Z}^{\mathrm{e}}_{\mathrm{r}}) \leftarrow \texttt{fuseBingham}\,(f^{\mathrm{p}}_{\mathcal{B}}, f^{\underline{\boldsymbol{v}}}_{\mathcal{B}}, \hat{\underline{z}}_{\mathrm{r}})$; // see [32]
2 $\{\underline{x}_{i,\mathrm{r}}\}_{i=1}^{n_{\mathrm{r}}} \leftarrow \texttt{sampleDeterministic}\,(f^{\mathrm{e}}_{\mathcal{B}})$; // equally weighted
3 $f^{\mathrm{e}}_{\mathcal{N}} \leftarrow f^{\mathrm{p}}_{\mathcal{N}}$;
4 $[\,\mathbf{E}^{\mathcal{B}}_{\mathrm{m}}, \underline{m}\,] \leftarrow \mathbf{M}^{\mathrm{e}}_{\mathrm{r}}$;
5 **for** $i \leftarrow 1$ **to** n_r **do**
6 $\quad \mathbf{E}^{\mathcal{B}}_{\mathrm{r},i} \leftarrow \mathbf{Q}^{\llcorner}_{\underline{x}_{i,\mathrm{r}}} (\mathbf{Q}^{\llcorner}_{\underline{m}})^{\top} \mathbf{E}^{\mathcal{B}}_{\mathrm{m}}$; // see (3.16)
7 $\quad \hat{\underline{z}}^{\iota}_{\mathrm{s}} \leftarrow (\mathbf{E}^{\mathcal{B}}_{\mathrm{r},i})^{\top} \hat{\underline{z}}_{\mathrm{s}}$;
8 $\quad f^{\mathrm{e}}_{\mathcal{N}} \leftarrow \texttt{updateUKF}\,(f^{\mathrm{e}}_{\mathcal{N}}, f^{\underline{\boldsymbol{v}}}_{\mathcal{N}}, \hat{\underline{z}}^{\iota}_{\mathrm{s}})$;
9 **return** $f^{\mathrm{e}}_{\mathcal{B}}, f^{\mathrm{e}}_{\mathcal{N}}$;



(A) rotation error (B) translation error

Figure 4.3.: Consecutive measurement fusion using a PTBG-based identity model under different noise levels.

Measurement fusion with identity measurement model: As the next building block for PTBG-based recursive estimation, we propose a basic method for measurement fusion as formulated in Algorithm 4. Currently, it only considers an identity measurement model in the form $\underline{\boldsymbol{z}}_t = \underline{\boldsymbol{x}}_t \boxtimes \underline{\boldsymbol{v}}_t$. As shown in line 1, the real part $\hat{\underline{z}}_{\mathrm{r}}$ of the measurement $\hat{\underline{z}} = [\,\hat{\underline{z}}^{\top}_{\mathrm{r}}, \hat{\underline{z}}^{\top}_{\mathrm{s}}\,]^{\top}$ is first fused into the prior Bingham density $f^{\mathrm{p}}_{\mathcal{B}}$ to obtain the its posterior $f^{\mathrm{e}}_{\mathcal{B}}$ as given in [32]. Afterward, we draw deterministic samples $\{\underline{x}_{i,\mathrm{r}}\}_{i=1}^{n_{\mathrm{r}}}$ from the posterior Bingham density $f^{\mathrm{e}}_{\mathcal{B}}$ (Algorithm 4, line 2). At each quaternion sample $\underline{x}_{i,\mathrm{r}} \in \mathbb{S}^3$, a tangent space is spanned by its local BPB $\mathcal{E}^{\mathcal{B}}_{\mathrm{r},i}$ obtained via HPT, w.r.t. which the dual part $\hat{\underline{z}}_{\mathrm{s}}$ of the measurement $\hat{\underline{z}}$ is fused into the Gaussian component via an ordinary UKF update [8] (Algorithm 4, lines 3–8). We validate the proposed measurement fusion method in a simulated scenario in the following example.

Example 5: *We simulate the hidden state of the unit dual quaternion signal according to* $\underline{\boldsymbol{x}} = \underline{x}_0 \boxtimes \underline{\boldsymbol{w}}$. *For that, we synthesize the noise term* $\underline{\boldsymbol{w}}$ *according to the PTBG scheme with the Bingham and Gaussian components parameterized as* $f^{\underline{\boldsymbol{w}}}_{\mathcal{B}}(-\mathrm{diag}(1, 400, 400, 400))$ *and* $f^{\underline{\boldsymbol{w}}}_{\mathcal{N}}(\underline{0}_3, 0.002 \cdot \mathbf{I}_{3\times 3})$, *respectively. An additional offset* $\underline{x}_0$ *is added, which represents a rotation of angle* $\pi/3$ *around the axis* $1/\sqrt{3} \cdot [\,1, 1, 1\,]^{\top}$ *followed by a translation* $[\,5, 4, 6\,]^{\top}$. *Given an identity measurement model* $\underline{\boldsymbol{z}} = \underline{\boldsymbol{x}} \boxtimes \underline{\boldsymbol{v}}$, *we model the noise term* $\underline{\boldsymbol{v}}$ *under the scheme of PTBG with the Bingham and Gaussian components configured as* $f^{\underline{\boldsymbol{v}}}_{\mathcal{B}}(-\mathrm{diag}(1, 1/a, 2/a, 3/a))$ *and* $f^{\underline{\boldsymbol{v}}}_{\mathcal{N}}(\underline{0}_3, a \cdot \mathbf{I}_{3\times 3})$, *respectively. Here,* $a \in$

$\{0.1, 0.01, 0.001\}$ *control the measurement noise levels (a larger* a *gives a larger uncertainty). For initializing the prior estimate, we exploit a PTBG configuration with the Bingham and Gaussian components parameterized as* $f_{\mathcal{B}}^{p}(-\mathrm{diag}(1, 500, 500, 500))$ *and* $f_{\mathcal{N}}^{p}(\underline{0}_3, 0.001 \cdot \mathbf{I})$*, respectively. We perform measurement fusion for* 50 *consecutive steps with* 200 *Monte Carlo runs and record estimation errors for rotation (in radians) and translation after each step w.r.t. the ground truth.*

As depicted in Fig. 4.3, the averaged estimation errors converge over fusion steps for both rotation and translation. Also, a lower measurement noise level induces faster convergence. This validates the efficacy of the proposed measurement fusion method under the identity setting.

4.2 Upgrade II: Configurable Deterministic Sampling

As introduced in Sec. 2.2.4, deterministic samples are drawn in a reproducible manner and carry information of the underlying distribution more efficiently for nonlinear estimation compared with random samples. In the past state of the art, deterministic sampling for parametric models on hyperspheres $\mathbb{S}^{d-1}$ could only be performed with a fixed sample size fully determined by the dimension d while preserving the first- and second-order moments of the underlying distribution. To improve the performance of parametric directional filtering in nonlinear estimation tasks, we propose the configurable deterministic sampling (CDS) methodology to generate deterministic samples of manually-configurable sizes. The obtained samples fulfill the moment constraints, meanwhile they approximate higher-order shape information of the underlying dispersion. As shown in related own publications [35–37], we concretize this methodology for two typical hyperspherical distributions, namely, the von Mises–Fisher distribution of isotropic dispersion and the Bingham distribution that has antipodally symmetric and non-isotropic dispersion.

4.2.1 Flexible Deterministic Sampling for Bingham Distributions

In [69, 70, 94], approaches have been proposed for deterministic Dirac mixture approximation (DMA) of Gaussian distributions in the sense of least Cramér–von Mises distance based on the concept of the localized cumulative distribution (LCD)[2]. Deterministic samples are drawn to optimally approximate the shape of a Gaussian distribution while preserving its moments up to the second order for UT-based filtering. As it can be executed efficiently [95] thanks to its closed-form derivations, this methodology has brought profound improvement for nonlinear estimation in Euclidean spaces [96].

The key idea of establishing the flexible deterministic sampling (FDS) for a Bingham distribution (introduced in Sec. 2.2.2) is: (1) approximating the hyperspherical dispersion with a Gaussian distribution in the tangent space $\mathbb{T}_{\underline{m}}\mathbb{S}^{d-1}$ at the mode $\underline{m}$ w.r.t. the BPB, (2) performing LCD-based deterministic sampling on the approximate model and retracting the obtained samples back to the manifold, and (3) on-manifold layout correction to the freshly retracted samples for fulfilling the moment constraints. As introduced in own publications [35] and [36], this idea has been tried out progressively – first along the principal curves, then on the whole hypersphere. We hereby only elaborate the procedure of the latter since the former is a degenerate version following the same methodology.

Given a Bingham distribution $\mathcal{B}(\mathbf{M}, \mathbf{Z})$ as defined in (2.9) on the unit hypersphere $\mathbb{S}^{d-1}$, its parameter matrix $\mathbf{M}$ can be expressed column wise as $\mathbf{M} \coloneqq [\,\underline{m}_1, \underline{m}_2, \ldots, \underline{m}_{d-1}, \underline{m}\,]$, with the last

[2] An on-manifold version of the Cramér–von Mises distance will be elaborated in Chapter 6.

column being the mode and the rest composing the BPB $\mathcal{E}_{\mathrm{m}}^{\mathcal{B}} = \{\underline{m}_i\}_{i=1}^{d-1}$ spanning the tangent space $\mathbb{T}_{\underline{m}}\mathbb{S}^{d-1}$ as introduced in Sec. 3.3.2. Since samples are first to be drawn from a Gaussian distribution by means of the LCD-based DMA unbounded in $\mathbb{T}_{\underline{m}}\mathbb{S}^{d-1}$ w.r.t. the BPB, we exploit the gnomonic projection in (3.3) to interpret the on-manifold density in the tangent space.

Expressed w.r.t. the BPB, points in the tangent space $\mathbb{T}_{\underline{m}}\mathbb{S}^{d-1}$ essentially form the Euclidean space $\mathbb{R}^{d-1}$. Given any point $\tilde{\underline{x}}^{\iota} \in \mathbb{R}^{d-1}$ expressed w.r.t. the BPB of $\mathbb{T}_{\underline{m}}\mathbb{S}^{d-1}$, we transform it to its global coordinates and pick the on-manifold density value via gnomonic retraction. Mathematically, this procedure can be formulated as

$$f_{\mathcal{B}}(\tilde{\underline{x}}^{\iota}) = \frac{1}{N_{\mathcal{B}}} \exp\left(\left(\mathcal{R}_{\underline{m}}^{\mathrm{g}}(\mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{x}}^{\iota}) \right)^{\top} \mathbf{M}\, \mathbf{Z}\, \mathbf{M}^{\top} \mathcal{R}_{\underline{m}}^{\mathrm{g}}(\mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{x}}^{\iota}) \right), \quad \text{with} \quad \tilde{\underline{x}}^{\iota} \in \mathbb{R}^{d-1}\,.$$

Taking the definition of the gnomonic retraction $\mathcal{R}_{\underline{m}}^{\mathrm{g}}$ in (3.4), we then obtain

$$f_{\mathcal{B}}(\tilde{\underline{x}}^{\iota}) = \frac{1}{N_{\mathcal{B}}} \exp\left\{ \frac{(\underline{m} + \mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{x}}^{\iota})^{\top} \mathbf{M}\, \mathbf{Z}\, \mathbf{M}^{\top} (\underline{m} + \mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{x}}^{\iota})}{\| \underline{m} + \mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{x}}^{\iota} \|^2} \right\}. \tag{4.4}$$

As introduced in Sec. 3.3.2, the transformation matrix $\mathbf{E}_{\mathrm{m}}^{\mathcal{B}} \in \mathbb{R}^{d\times(d-1)}$ denoting the BPB $\mathcal{E}_{\mathrm{m}}^{\mathcal{B}}$ is given by $\mathbf{M} = [\, \mathbf{E}_{\mathrm{m}}^{\mathcal{B}}, \underline{m}\,]$ and fulfills $(\mathbf{E}_{\mathrm{m}}^{\mathcal{B}})^{\top}\mathbf{E}_{\mathrm{m}}^{\mathcal{B}} = \mathbf{I}_{(d-1)\times(d-1)}$. Also, the Bingham concentration matrix is given in the form $\mathbf{Z} = \mathrm{diag}(z_1, \dots, z_{d-1}, 0)$ by definition. Thus, we have $(\mathbf{E}_{\mathrm{m}}^{\mathcal{B}})^{\top}\mathbf{M} = (\mathbf{E}_{\mathrm{m}}^{\mathcal{B}})^{\top}[\, \mathbf{E}_{\mathrm{m}}^{\mathcal{B}}, \underline{m}\,] = [\, \mathbf{I}_{(d-1)\times(d-1)}, \underline{0}_{d-1}\,] \in \mathbb{R}^{(d-1)\times d}$ and $\underline{m}^{\top}\mathbf{M} = [\, \underline{0}_{d-1}^{\top}, 1\,] \in \mathbb{R}^{1\times d}$ for the numerator in the exponent of (4.4). Meanwhile, the denominator in the exponent of (4.4) follows $\| \underline{m} + \mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{x}}^{\iota} \|^2 = (\underline{m} + \mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{x}}^{\iota})^{\top}(\underline{m} + \mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{x}}^{\iota}) = 1 + \| \tilde{\underline{x}}^{\iota} \|^2$. Therefore, (4.4) can be reduced to

$$f_{\mathcal{B}}(\tilde{\underline{x}}^{\iota}) = \frac{1}{N_{\mathcal{B}}} \exp\left(\frac{(\tilde{\underline{x}}^{\iota})^{\top}\, \mathrm{diag}(z_1, \dots, z_{d-1})\, \tilde{\underline{x}}^{\iota}}{1 + \| \tilde{\underline{x}}^{\iota} \|^2} \right).$$

The density expressed above is not a PDF in $\mathbb{R}^{d-1}$. However, as shown in [36], its shape can be approximated by a scaled Gaussian distribution in $\mathbb{R}^{d-1}$, from which n arbitrary deterministic samples $\{\tilde{\underline{\sigma}}_i^{\iota}\}_{i=1}^{n}$ can be drawn via the LCD-based DMA. Afterward, the samples are transformed to global coordinates and mapped back to the manifold via the gnomonic retraction $\underline{\sigma}_i = \mathcal{R}_{\underline{m}}^{\mathrm{g}}(\mathbf{E}_{\mathrm{m}}^{\mathcal{B}}\, \tilde{\underline{\sigma}}_i^{\iota}) \in \mathbb{S}^{d-1}$. For typical scenarios of hyperspherical estimation, we set samples to be equally weighted.

The next step is to correct the freshly retracted samples according to the requirement of the unscented transform, namely, to yield a sample covariance that is identical to the $\mathrm{cov}_{\mathcal{B}}$ in (2.11) of the underlying Bingham distribution. For this, we formulate an optimization problem based on the Frobenius norm to minimize the difference of the two covariance matrices. It follows

$$\mathbf{S}^{*} = \underset{\mathbf{S}\in\mathbb{OB}(d,n)}{\arg\min} \left\| \mathbf{S}\, \mathbf{S}^{\top}/n - \mathrm{cov}_{\mathcal{B}} \right\|_{\mathrm{F}}^{2}, \quad \text{with} \quad \mathbf{S} = [\, \underline{\sigma}_1, \dots, \underline{\sigma}_n\,]$$

concatenating the retracted samples $\{\underline{\sigma}_i\}_{i=1}^{n}$ column wise. Since all samples are confined to the unit hypersphere $\mathbb{S}^{d-1} \subset \mathbb{R}^{d}$, matrix $\mathbf{S}$ belongs to the oblique manifold $\mathbb{OB}(d,n) \subset \mathbb{R}^{d\times n}$ [85]. As recommended in [36], the optimization problem can be efficiently solved by the Riemannian steepest descent method implemented in [86]. To demonstrate the efficacy of the proposed FDS approach for Bingham distributions (including the one in [35] proposed for sampling on principal curves), we now provide an example.

Example 6: *We perform the proposed FDS method for Bingham distributions* $\mathcal{B}(\mathbf{C})$ *(defined in the concise form given in Sec. 2.2.2) of different parameterizations. Due to the antipodal symmetry of dispersion, the configured sample sizes refer only to the semicircle or hemisphere.*

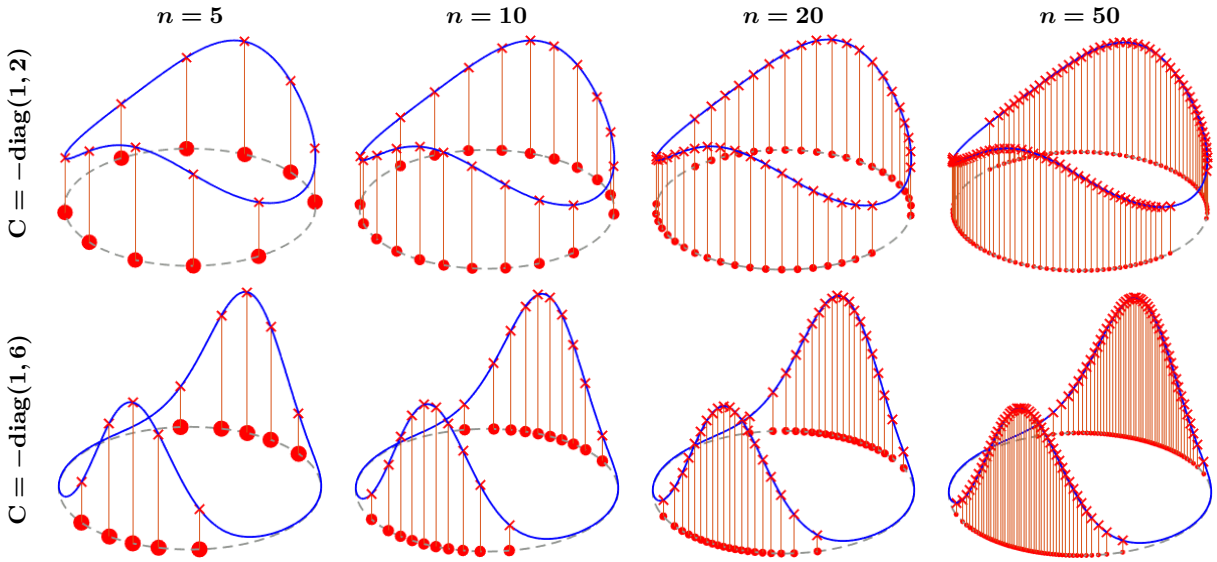


Figure 4.4.: FDS for Bingham distributions on $\mathbb{S}^1$ in Example 6.1.

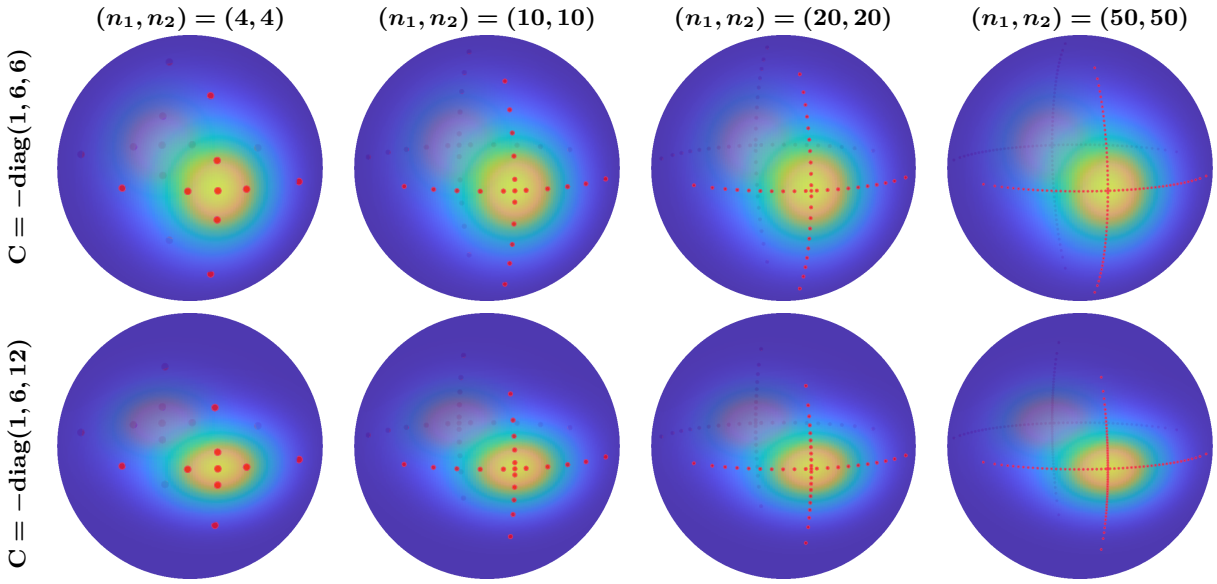


Figure 4.5.: FDS on principal curves for Bingham distributions on $\mathbb{S}^2$ in Example 6.2.

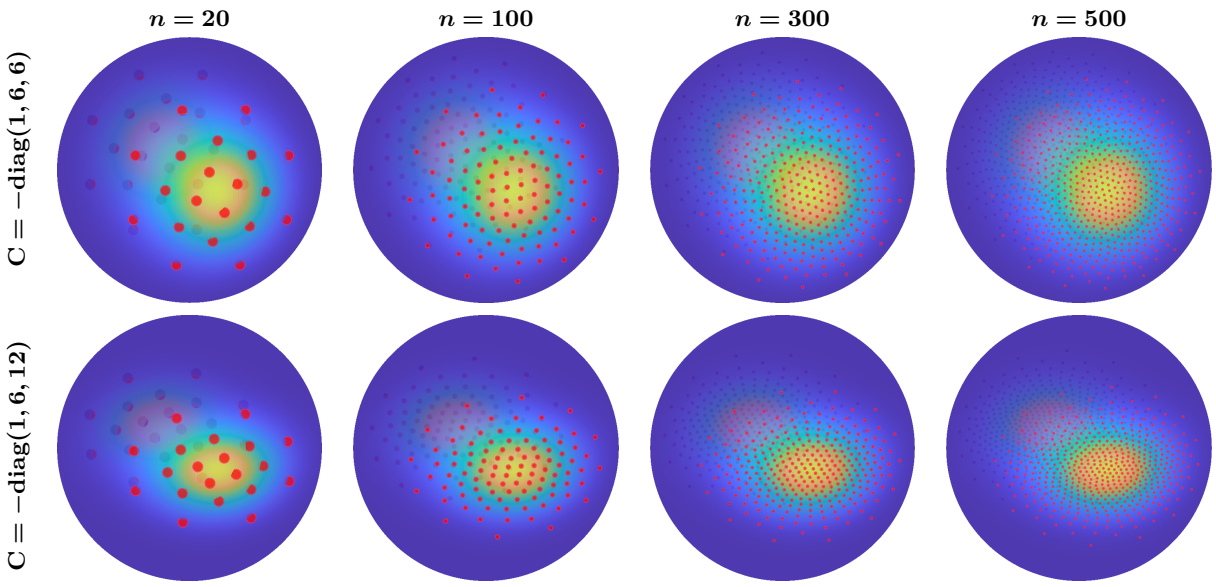


Figure 4.6.: FDS on the whole sphere for Bingham distributions on $\mathbb{S}^2$ in Example 6.3.

1. *On the unit circle* $\mathbb{S}^1$*, we parameterize Bingham distributions with parameter matrices* $\mathbf{C} \in \{-\mathrm{diag}(1, 2), -\mathrm{diag}(1, 6)\}$*. Sample sizes are given as* $n \in \{5, 10, 20, 50\}$*.*
2. *Another two Bingham distributions are further parameterized on the unit sphere* $\mathbb{S}^2$ *with* $\mathbf{C} \in \{-\mathrm{diag}(1, 6, 6), -\mathrm{diag}(1, 6, 12)\}$*. We first perform FDS of its original version proposed in [35], with* $(n_1, n_2) \in \{(4, 4), (10, 10), (20, 20), (50, 50)\}$ *denoting the sample sizes along the two principal curves of Bingham distributions.*
3. *We exploit the previous setting of Bingham distributions and perform FDS on the whole sphere using the variant elaborated in this section according to [36], with* $n \in \{20, 100, 300, 500\}$ *being the sample sizes.*

As shown in Fig. 4.4, Fig. 4.5 and Fig. 4.6, samples are drawn from different Bingham distributions given various configurations and mirrored to their antipodes. The proposed FDS scheme is able to generate deterministic samples of layouts adaptive to the shape of dispersion underlying on the manifold while fulfilling the moment constraints. Moreover, a comparison between the sampling results in Fig. 4.5 and Fig. 4.6 clearly hints that the FDS-variant on the whole hypersphere delivers a better sample covering of Bingham distributions on $\mathbb{S}^{d-1}$ *(*$d \geq 3$*), especially for the area between principal curves that interprets the correlation. Therefore, this variant is recommended for deployment to nonlinear hyperspherical estimation.*

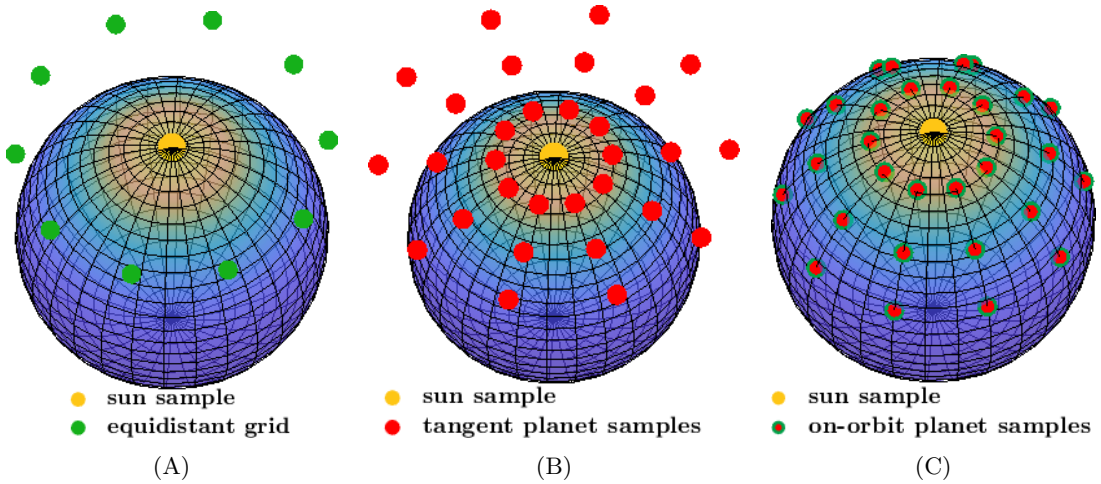
4.2.2 Isotropic Deterministic Sampling for von Mises–Fisher Distributions

Another important parametric distribution modeling directional random variables is the von Mises–Fisher distribution as introduced in Sec. 2.2.1 [6, 12]. Defined on the unit hypersphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ as illustrated in Fig. 2.1-(A), the distribution has isotropic dispersion. The conventional deterministic sampling in the past [34] only produces $2 \cdot d - 1$ samples that preserve the mean resultant vector in (2.7) of the underlying distribution. In this section, we introduce a new sampling method proposed in [37, 38] to obtain configurable numbers of deterministic samples that preserve

Algorithm 5: Isotropic Deterministic Sampling (IDS) for von Mises–Fisher Distributions

Input: $\mathcal{VMF}(\underline{\alpha}, \kappa)$, number of orbits n_1, on-orbit resolution n_2
Output: sample set $\mathbb{X}$

1 $\mathbb{X} \leftarrow \underline{\alpha}$;
2 $\mathcal{E}_\alpha \leftarrow$ `getBasisTangentSpace` $(\underline{\alpha})$;
3 $\{\tilde{\underline{x}}_j^\iota\}_{j=1}^{n_2} \leftarrow$ `HEP` $(\mathbb{S}^{d-2}, n_2)$;
4 $\lambda \leftarrow$ `computeInterval` (n_1, n_2, κ);
5 **for** $k \leftarrow 1$ **to** n_1 **do**
6 **for** $j \leftarrow 1$ **to** n_2 **do**
7 $\mathbb{X} \leftarrow \mathbb{X} \cup \mathrm{Exp}_{\underline{\alpha}}(k\lambda\, \mathbf{E}_\alpha\, \tilde{\underline{x}}_j^\iota)$;
8 **return** $\mathbb{X}$;



(A) (B) (C)

Figure 4.7.: Illustration of isotropic deterministic sampling procedure for von Mises–Fisher distributions.

the mean resultant vector and approximate higher-order shape information of the underlying von Mises–Fisher distribution.

Considering that a von Mises–Fisher distribution $\mathcal{VMF}(\underline{\alpha}, \kappa)$ has isotropic dispersion, we design the layout of the sample set to also be isotropic. As shown in Fig. 4.7, we put one sample at the location of its mode $\underline{\alpha} \in \mathbb{S}^{d-1}$ and call it the *sun sample*. It is surrounded by n_1 hyperspherical orbits of uniform interval λ, with n_2 *planet* samples placed (quasi-)equidistantly on each of them. Thus, a sample set $\mathbb{X} \subset \mathbb{S}^{d-1}$ of cardinality $n_1 \cdot n_2 + 1$ is produced. Similar to the case of the FDS on Bingham distributions, we set all samples to be equally weighted. The uniform interval λ for arranging the hyperspherical orbits is determined in a way to preserve the mean resultant vector of $\mathcal{VMF}(\underline{\alpha}, \kappa)$. Before deriving it, we first give an overview of the proposed isotropic deterministic sampling (IDS) method in Algorithm 5 as follows.

As shown in Fig. 4.7-(A), the mode $\underline{\alpha}$ of the distribution is first set as the sun sample (Algorithm 5, line 1). The tangent space $\mathbb{T}_{\underline{\alpha}}\mathbb{S}^{d-1}$ is established at the mode with $\mathcal{E}_\alpha$ denoting a local basis of it[3] (Algorithm 5, line 2). As introduced in Sec. 3.1.1, points on the unit hypersphere $\mathbb{S}^{d-1}$ undergoing the logarithm map to a tangent space form the ball $\mathbb{B}_\pi^{d-1}$ bounded by the hypersphere $\mathbb{S}_\pi^{d-2} \subset \mathbb{R}^{d-1}$ of radius π w.r.t. its local basis. As depicted in Fig. 4.7-(A), we first apply the hyperspherical equal

[3] A matrix $\mathbf{E}_\alpha \in \mathbb{R}^{d\times(d-1)}$ concatenates basis vectors of $\mathcal{E}_\alpha$ column wise and fulfills $\mathbf{E}_\alpha^\top \mathbf{E}_\alpha = \mathbf{I}_{(d-1)\times(d-1)}$.

partitioning (HEP) algorithm from [67] to $\mathbb{S}^{d-2}$ for obtaining n_2 on-orbit (quasi-)equidistant planet samples $\{\tilde{\underline{x}}_j^{\iota}\}_{j=1}^{n_2} \subset \mathbb{S}^{d-2}$ (Algorithm 5, line 3). Suppose the interval λ is obtained (Algorithm 5, line 4), the equidistant grid points are scaled on each orbit of $k = 1, \dots, n_1$ according to λ and afterward transformed to global coordinates by $\mathbf{E}_{\alpha}$ (Fig. 4.7-(B)). Then, they are retracted back to $\mathbb{S}^{d-1}$ via the exponential map in (3.2) (Fig. 4.7-(C) and Algorithm 5, lines 5–7). Mathematically, this procedure follows

$$\underline{x}_{k,j} = \operatorname{Exp}_{\underline{\alpha}}(k\,\lambda\,\mathbf{E}_{\alpha}\,\tilde{\underline{x}}_j^{\iota}) = \cos(k\lambda)\,\underline{\alpha} + \sin(k\lambda)\,\mathbf{E}_{\alpha}\,\tilde{\underline{x}}_j^{\iota} \in \mathbb{S}^{d-1}\,,$$

with $\underline{x}_{k,j}$ being the j-th sample on the k-th hyperspherical orbit on $\mathbb{S}^{d-1}$. Taking both planet and sun samples, we derive their hyperspherical mean to be the form

$$\hat{\underline{v}} = \frac{1}{n_1 n_2 + 1}\Big(\underline{\alpha} + \sum_{k=1}^{n_1}\sum_{j=1}^{n_2}\underline{x}_{k,j}\Big) = \frac{1}{n_1 n_2 + 1}\Big(\underline{\alpha} + \sum_{k=1}^{n_1}\sum_{j=1}^{n_2}\big(\cos(k\lambda)\,\underline{\alpha} + \sin(k\lambda)\,\mathbf{E}_{\alpha}\tilde{\underline{x}}_j^{\iota}\big)\Big).$$

Given that the grid point set $\{\tilde{\underline{x}}_j^{\iota}\}_{j=1}^{n_2}$ drawn by the HEP proposed in [67] is (numerically) zero-centered for typical IDS configurations, we thus have

$$\hat{\underline{v}} = \frac{1}{n_1 n_2 + 1}\Big(1 + n_2\sum_{k=1}^{n_1}\cos(k\lambda)\Big)\underline{\alpha}\,.$$

By imposing the constraint of the mean resultant vector, namely, $\hat{\underline{v}} \overset{!}{=} \mathcal{A}_d(\kappa)\,\underline{\alpha}$, the condition of the unscented transform [34] is fulfilled, and we further obtain

$$\sum_{k=1}^{n_1}\cos(k\lambda) = \frac{(n_1 n_2 + 1)\mathcal{A}_d(\kappa) - 1}{n_2}\,.$$

The finite series on the left-hand side of the equation above refers to Lagrange's trigonometric identity [97, Sec. 2.4.1.6]. Considering this fact, we have

$$\frac{\sin\big((n_1 + 0.5)\lambda\big)}{2\sin(0.5\,\lambda)} = \frac{(n_1 n_2 + 1)\mathcal{A}_d(\kappa) - 1}{n_2} + \frac{1}{2}\,,$$

with the left-hand side taking the form of a scaled Dirichlet kernel [98], which we denote as $\mathcal{J}_{n_1}(\lambda)$. Computing λ then boils down to solving the equation

$$\mathcal{J}_{n_1}(\lambda) = \frac{(n_1 n_2 + 1)\mathcal{A}_d(\kappa) - 1}{n_2} + \frac{1}{2}\,, \quad \text{with} \quad \mathcal{J}_{n_1}(\lambda) = \frac{\sin\big((n_1 + 0.5)\lambda\big)}{2\sin(0.5\,\lambda)}\,,\ \lambda \in [\,0, \pi/n_1\,]\,. \tag{4.5}$$

The scaled Dirichlet kernel $\mathcal{J}_{n_1}(\lambda)$ obtains its maximum at $\lambda = 0$ with $\mathcal{J}_{n_1}(0) = n_1 + 0.5$. Since the Bessel function ratio $\mathcal{A}_d(\kappa)$ is in the range of $(0, 1)$ for concentration $\kappa > 0$, the value on the right-hand side of (4.5) is less than $\mathcal{J}_{n_1}(0)$. Thus, (4.5) is solvable in non-analytic form. For that, we provide a numerical solver by tailoring Newton's method to the problem at hand in Appendix B. The following example further showcases the proposed IDS method applied to von Mises–Fisher distributions on $\mathbb{S}^2$ of different configurations.

Example 7: *We set up von Mises–Fisher distributions $\mathcal{VMF}([\,0, 0, 1\,]^{\top}, \kappa)$ on the unit sphere $\mathbb{S}^2$ with different concentrations $\kappa \in \{0.4, 3, 5\}$. For each configuration, we perform IDS with resolutions of $(n_1, n_2) \in \{(3, 12), (6, 12), (12, 12), (12, 24)\}$.*

As shown in Fig. 4.8, deterministic samples are drawn in dispersion-adaptive layouts, meanwhile mean resultant vectors of the underlying von Mises–Fisher distributions are preserved.

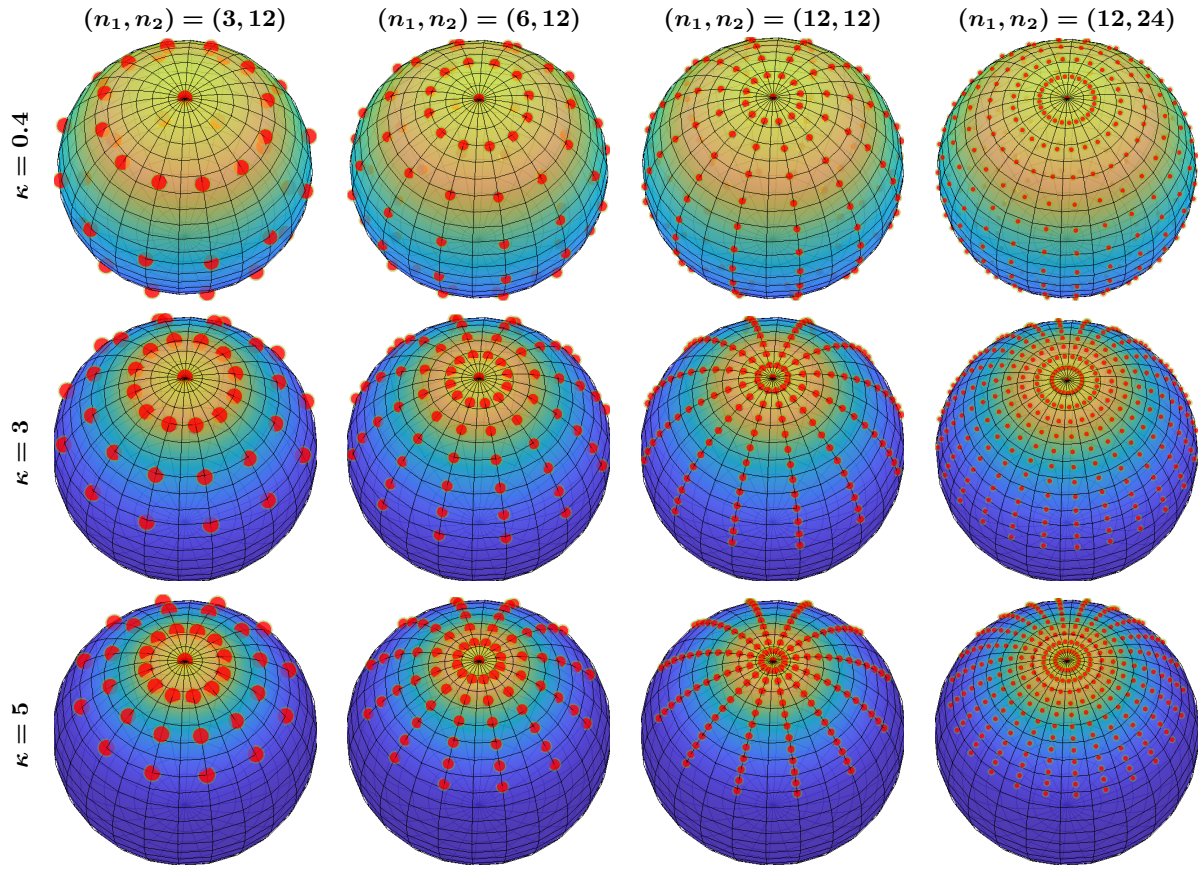


Figure 4.8.: IDS for von Mises–Fisher distributions on $\mathbb{S}^2$ in Example 7.

4.3 Upgrade III: Deterministic Progressive Update

In most directional estimation tasks, measurement models as shown in (2.2) are of non-identity form, namely, the measurement $\underline{\boldsymbol{z}}_t$ and the state $\underline{\boldsymbol{x}}_t$ belong to different domains, with each following its own respective distribution [31,39]. For the basic nonlinear filtering scheme in Algorithm 1, deterministic samples $\{\underline{x}^{\mathrm{p}}_{t,i}\}_{i=1}^{n}$ drawn from the prior density (obtained via initialization or prediction) are reweighted by their likelihoods $\{f^{\mathrm{L}}_t(\hat{\underline{z}}_t \,|\, \underline{x}^{\mathrm{p}}_{t,i})\}_{i=1}^{n}$ given the measurement $\hat{\underline{z}}_t$. Though the UT-based deterministic sampling schemes have been largely improved by enabling configurable sample sizes as introduced in Sec. 4.2, such a rudimentary update step may still suffer from sample degeneracy issue under strong nonlinearities or peaky likelihoods. As a straightforward demonstration of this issue, Fig. 4.9 is plotted with a von Mises–Fisher prior density (of concentration $\kappa = 5$) updated by a relatively peaky likelihood density function[4]. Though overall 21 deterministic samples are drawn for the measurement update via the IDS proposed in Sec. 4.2.2 (resolution configured as $(n_1, n_2) = (2, 10)$), the prior samples are still prone to degenerate after a single-step reweighting with a huge ratio of the largest to the smallest weight.

As mentioned in Sec. 1.3.1, this issue can be mitigated by exploiting the progressive update scheme that was originally proposed for linear state spaces [61,62] and previously evaluated on the circular domain [33]. Based on own publications [38–40], we hereby extend this scheme to generic directional

[4] Another von Mises–Fisher distribution of $\kappa = 50$ is deployed as measurement noise for demonstration purposes.

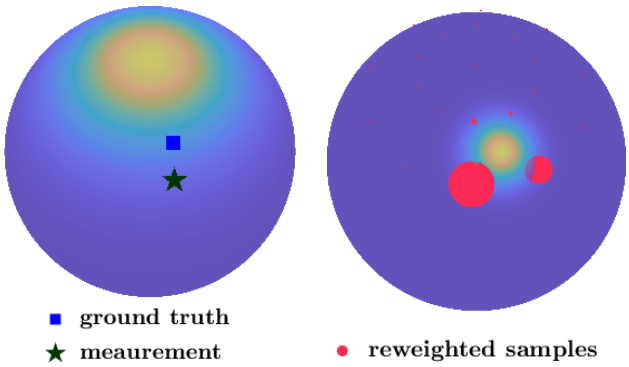


Figure 4.9.: Illustration of a single-step update for a von Mises–Fisher filter using isotropic deterministic samples of resolution $(n_1, n_2) = (2, 10)$. Samples are plotted with sizes proportional to their weights after being reweighted by their likelihoods.

Algorithm 6: On-Manifold Deterministic Progressive Update

Input: prior density f_t^{p}, measurement $\hat{\underline{z}}_t$, threshold ϵ
Output: posterior density f_t^{e}

$f_t^{\mathrm{e}} \leftarrow f_t^{\mathrm{p}}$;
$\Delta \leftarrow 1$, $k \leftarrow 1$;
while $\Delta > 0$ **do**
  $\{\underline{x}_{k,i}\}_{i=1}^{n} \leftarrow \mathtt{CDS}\,(f_t^{\mathrm{e}}, n)$; // distribution-specific
  $\{\omega_{k,i}\}_{i=1}^{n} \leftarrow \{f_t^{\mathrm{L}}(\hat{\underline{z}}_t|\underline{x}_{k,i})\}_{i=1}^{n}$; // element-wise assignment
  $\omega_k^{\min} \leftarrow \min\left(\{\omega_{k,i}\}_{i=1}^{n}\right)$;
  $\omega_k^{\max} \leftarrow \max\left(\{\omega_{k,i}\}_{i=1}^{n}\right)$;
  $\Delta_k \leftarrow \min\left(\Delta,\, \log(\epsilon)/\log(\omega_k^{\min}/\omega_k^{\max})\right)$;
  $\{\omega_{k,i}\}_{i=1}^{n} \leftarrow \{(\omega_{k,i})^{\Delta_k}\}_{i=1}^{n}$; // element-wise assignment
  $f_t^{\mathrm{e}} \leftarrow \mathtt{matchMoment}\left(\{\underline{x}_{k,i}\}_{i=1}^{n}, \{\omega_{k,i}\}_{i=1}^{n}\right)$; // distribution-specific
  $\Delta \leftarrow \Delta - \Delta_k$;
  $k \leftarrow k + 1$;
return f_t^{e} ;

domains in conjunction with the configurable deterministic sampling scheme introduced in Sec. 4.2. The resulting deterministic progressive update step is derived as follows.

We substitute $f_t^{\mathrm{p}}(\underline{x}_t)$ in (2.3) with a Dirac mixture supported by deterministic samples $\{\underline{x}_{t,i}^{\mathrm{p}}\}_{i=1}^{n}$ of size n drawn from the prior density, namely, $f_t^{\mathrm{p}}(\underline{x}_t) = \sum_{i=1}^{n} \nu_{t,i}^{\mathrm{p}}\, \delta(\underline{x}_t - \underline{x}_{t,i}^{\mathrm{p}})$ and obtain

$$f_t^{\mathrm{e}}(\underline{x}_t|\hat{\underline{z}}_t) \propto f_t^{\mathrm{L}}(\hat{\underline{z}}_t|\underline{x}_t)\, f_t^{\mathrm{p}}(\underline{x}_t) = \sum_{i=1}^{n} \nu_{t,i}^{\mathrm{p}}\, f_t^{\mathrm{L}}(\hat{\underline{z}}_t|\underline{x}_{t,i}^{\mathrm{p}})\, \delta(\underline{x}_t - \underline{x}_{t,i}^{\mathrm{p}})\,, \tag{4.6}$$

with $\nu_{t,i}^{\mathrm{p}}$ being the prior sample weights that fulfill $\sum_{i=1}^{n} \nu_{t,i}^{\mathrm{p}} = 1$ (uniform if given by the CDS). The key idea of the progressive measurement update is to decompose the single-step likelihood reweighting in (4.6) into l progressions as

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

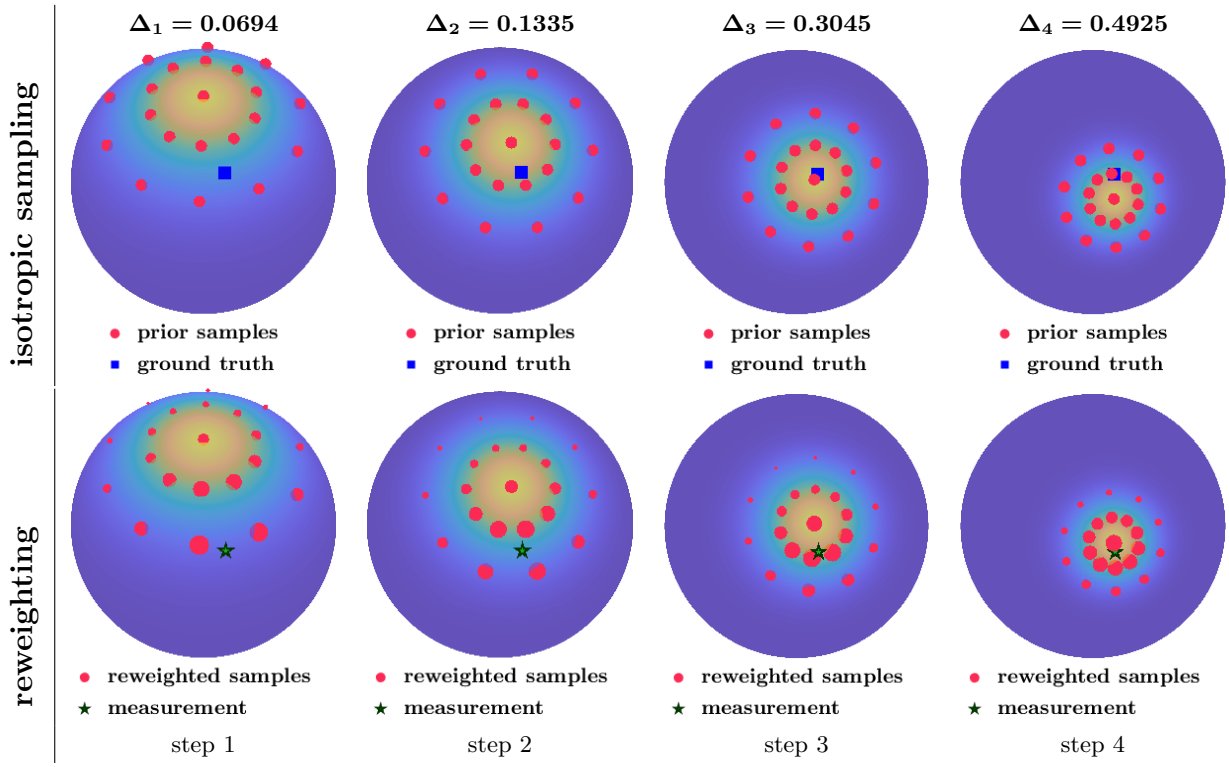


Figure 4.10.: Demonstration of an IDS-based progressive update step for von Mises–Fisher filtering.

with $\sum_{k=1}^{l} \Delta_k = 1$, where Δ_k denotes the stride length at each step. It is determined by imposing a pregiven threshold ϵ on the minimal ratio among the likelihoods of current posterior samples $\{\underline{x}_{k,i}\}_{i=1}^{n}$ via

$$\left(\frac{\min\left(\{\omega_{k,i}\}_{i=1}^{n}\right)}{\max\left(\{\omega_{k,i}\}_{i=1}^{n}\right)} \right)^{\Delta_k} \geq \epsilon\,,$$

with $\omega_{k,i} = f_t^{\mathrm{L}}(\hat{\underline{z}}_t \,|\, \underline{x}_{k,i})$ denoting the likelihood of sample $\underline{x}_{k,i}$ given the measurement $\hat{\underline{z}}_t$. The maximal stride at the k-th progression then follows

$$\Delta_k \leq \frac{\log(\epsilon)}{\log(\omega_k^{\min}) - \log(\omega_k^{\max})}\,, \quad \text{with} \quad \omega_k^{\min} = \min(\{\omega_{k,i}\}_{i=1}^{n})\,,\ \omega_k^{\max} = \max(\{\omega_{k,i}\}_{i=1}^{n})\,. \tag{4.7}$$

Each stride length Δ_k is adapted to the current likelihoods ratio given the threshold ϵ, and the posterior density f_t^{e} is updated via moment matching. The progression cycle is repeated until the information from observation is fully fused into the prior via likelihood reweighting regulated by ϵ.

We detail the proposed progressive procedure in Algorithm 6 along with a graphic illustration in Fig. 4.10 showing its application to an IDS-based von Mises–Fisher filter. The progression horizon is initialized as $\Delta = 1$ with the posterior set as the given prior density (Algorithm 6, lines 1–2). Within each progression step, we draw a deterministic sample set $\{\underline{x}_{k,i}\}_{i=1}^{n}$ from the current posterior f_t^{e} with a configurable size n and compute their likelihoods given the measurement $\hat{\underline{z}}_t$ (Algorithm 6, lines 3–5). The current fusion stride length Δ_k is computed according to (4.7) (Algorithm 6, lines 6–8). The obtained Δ_k then scales the sample likelihoods, and the posterior density is updated via moment matching. The cycle ends when the initialized progression horizon Δ is used up (Algorithm 6, lines 9–12). Note that performing the configurable deterministic sampling or the moment matching (line 4 and line 10 in Algorithm 6, respectively) depends on the specific form of the distribution representing the estimate.

As demonstrated in Fig. 4.10, we set up an identical scenario to the one given in Fig. 4.9 for comparison with the single measurement update mentioned before. The proposed progressive scheme with IDS-based samples adaptively decomposes the entire update step into four consecutive substeps given $\epsilon = 0.02$. Consequently, the prior samples have considerably less tendency to degenerate, leading to superior fusion results over the ones obtained in a single-step fashion.

4.4 Case Study: Nonlinear Quaternion Estimation

The proposed configurable deterministic sampling and progressive update methods are applicable to common tasks of nonlinear directional estimation using parametric models. In [38], we have provided a detailed evaluation of the IDS-based von Mises–Fisher filtering with progressive measurement fusion for nonlinear spherical estimation. In comparison with other state-of-the-art directional filters, superior performance of the unscented von Mises–Fisher filter is achieved by combining the two introduced enhancements (IDS + progressive update) w.r.t. tracking accuracy and runtime efficiency. We hereby provide another scenario of quaternion-based nonlinear SO(3) estimation to further evaluate the proposed enhancements for the unscented Bingham filter [32].

Based on the logarithm/exponential maps in (3.1) and (3.2), respectively, we specify the system model in (2.1) based on quaternion interpolation [99] to be

$$\underline{\boldsymbol{x}}_{t+1} = \underline{a}(\underline{\boldsymbol{x}}_t) \otimes \underline{\boldsymbol{w}}_t\,, \quad \text{with} \quad \underline{a}(\underline{\boldsymbol{x}}_t) = \operatorname{Exp}_{\underline{\boldsymbol{x}}_t}\big(\zeta \cdot \operatorname{Log}_{\underline{\boldsymbol{x}}_t}(\hat{\underline{x}}_0)\big)$$

denoting the motion model interpolating between the current state $\underline{\boldsymbol{x}}_t$ and a target quaternion $\hat{\underline{x}}_0 = [\,1,1,1,1\,]^\top/2 \in \mathbb{S}^3$. We set the interpolation ratio to be constant with $\zeta = 0.1$. The system noise is assumed to follow a Bingham distribution with concentration matrix $\mathbf{Z} = -\operatorname{diag}(1, 500, 500, 500)$ and mode at a quaternion representing a rotation of $\theta = \pi/12$ around the axis $\underline{n} = 1/\sqrt{3} \cdot [\,1,1,1\,]^\top$. The measurement model in (2.2) is set to be non-identity as follows

$$\underline{\boldsymbol{z}}_t = \big(\underline{\boldsymbol{x}}_t \otimes [\,0, \hat{\underline{z}}_0^\top\,]^\top \otimes \underline{\boldsymbol{x}}_t\big)_{2:4} + \underline{\boldsymbol{v}}_t\,.$$

It rotates a point initialized at $\hat{\underline{z}}_0 = [\,2,3,4\,]^\top \in \mathbb{R}^3$ via the current quaternion state $\underline{\boldsymbol{x}}_t$ according to the formula in (3.7) and imposes an additive noise term $\underline{\boldsymbol{v}}_t$. We assume that the measurement noise follows a zero-mean Gaussian distribution, i.e., $\underline{\boldsymbol{v}}_t \sim \mathcal{N}(\underline{0}_3, \boldsymbol{\Sigma}^{\underline{v}})$, with $\boldsymbol{\Sigma}^{\underline{v}} = 0.01 \cdot \mathbf{I}_{3\times 3}$.

We apply the FDS scheme to the unscented quaternion Bingham filter (UQBF) with sample sizes ranging from 10 to 5000. The basic version of UQBF proposed in [32] with $n = 7$ deterministic samples is also included. The proposed progressive update scheme is further integrated with the FDS and the basic UT-based sampling method, leading to the variant called progressive unscented quaternion Bingham filter (Prog-UQBF). Another random sampling-based quaternion Bingham filter (QBF) [80] is deployed with sample sizes ranging from 10 to 10000. For all the filters, the prediction step is done in a similar fashion as in [32, 38] given an on-manifold additive system noise – samples (either deterministically or randomly) drawn from the previous posterior are propagated through the motion model $\underline{a}\colon \mathbb{S}^3 \to \mathbb{S}^3$ and moment-matched to a Bingham density, which is then composed with the noise distribution (also Bingham) to deliver the prior estimate. The tracking accuracy is quantified according to $\arccos\left(|\hat{\underline{x}}_t^\top \underline{x}_t|\right)$, with $\hat{\underline{x}}_t$ and $\underline{x}_t$ denoting the estimate and the ground truth, respectively. 1000 Monte Carlo runs are performed with 30 time steps in each run, based on which deviations in the form of the root mean squared error (RMSE) are computed w.r.t. the last estimate of each run.

For the proposed FDS method, the LCD-based DMA approach (provided by [95]) caches precomputed samples every time a newly configured sample size is used, inducing different interpretation

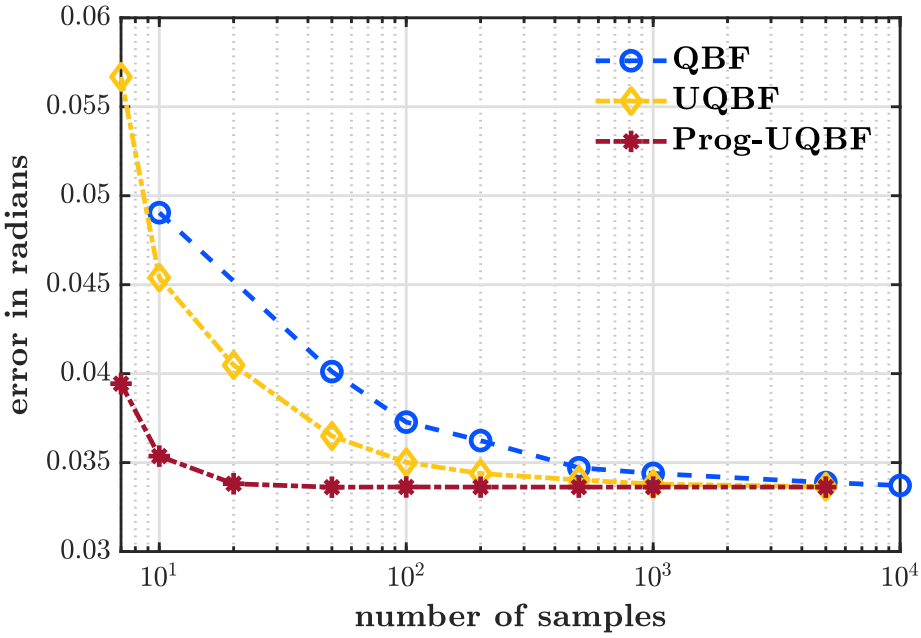


Figure 4.11.: Error over sample sizes given by the evaluated filters for nonlinear SO(3) estimation.

of runtime results compared with other online estimators (e.g., QBF). Therefore, a quantitative evaluation of runtime efficiency is excluded. In practice, execution of the FDS including the Riemannian optimization-based moment correction is typically fast. W.r.t. tracking accuracy plotted in Fig. 4.11, the proposed Bingham FDS method delivers an evident performance boost compared with the random sampling-based variant and the basic version using seven samples [32] (leftmost markers). Moreover, the Prog-UQBF that exploits the proposed progressive fusion scheme converges considerably faster than the ones using the rudimentary single-step update. Here, the results in the Bingham-based quaternion estimation scenario coheres to our findings in the evaluation of von Mises–Fisher-based hyperspherical estimation in [38]. They further validate the efficacy of the proposed enhancements in Sec. 4.2 and Sec. 4.3 for upgrading the parametric directional filtering scheme.

4.5 Short Summary

The three methodologies for upgrading the parametric probabilistic modeling and filtering of directional random variables in this chapter, though proposed individually, can be integrated together according to requirements of a specific application scenario [35, 37–40]. For instance, predicting/updating the Bingham component in the PTBG modeling scheme of the first upgrade in Sec. 4.1 can be potentially facilitated by the FDS method given in Sec. 4.2. In [40], we have also tested the progressive fusion scheme for SE(2)-Bingham-based planar dual quaternion filtering (mentioned in Sec. 2.2.3 and Sec. 3.3.3). There, the basic UT-based filter also exhibited considerable improvement w.r.t. tracking accuracy for nonlinear planar motion estimation.

For real-world applications, we have deployed the deterministic progressive update scheme to the stereo visual odometry system in [31] with quaternion-based Bingham filtering. There, the measurement model is set up on the camera image plane, with pixel coordinates of feature points being the measurements. Though the employed unscented Bingham filter is only equipped with the basic version of deterministic sampling (seven samples), the progressive update scheme still

facilitates the system to achieve an accuracy comparable to the state of the art for egomotion estimation.

CHAPTER 5

Topology-Aware Nonparametric Directional Filtering

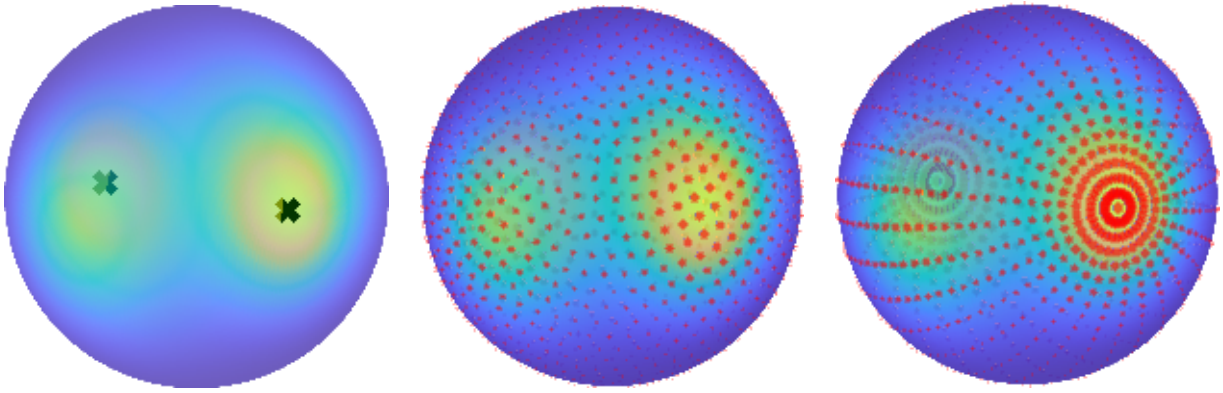
While the parametric models allow for a concise representation framework and have so far delivered promising results, several drawbacks may still arise for nonlinear directional estimation in practice. First, deploying parametric models to directional estimation requires recursive executions of the (deterministic) sampling–approximation cycle, where the normalization constant of a parametric model (see Sec. 2.2.1 and Sec. 2.2.2 for PDFs of common distributions) needs to be computed numerically during moment matching. Attaining the enhancements of configurable deterministic samples as suggested in Sec. 4.2 also relies on on-manifold optimizers or numerical solvers. Though runtime performance is not a bottleneck in executing parametric directional filters thanks to solutions such as precomputation [17, 80, 96] and customized numerical solvers [38, 79] as previously mentioned, issues may still occur concerning approximation accuracy or numerical stability. More importantly, parametric distributions impose their own assumptions on the random nature of uncertainties in the physical world. For instance, a random quaternion variable may exhibit multiple modes, thereby violating the unimodal dispersion (considering antipodal symmetry) assumed by the Bingham distribution. Or the assumption of isotropic dispersion introduced by utilizing von Mises–Fisher distributions can be easily violated by the uncertain hyperspherical states in practice.

As pointed out in Sec. 1.2.3, no systematic establishment of topology-aware non-parametric filtering methods existed previously for directional domains. In this chapter, we give a dedicated investigation in this regard by proposing two categories of methodologies – one based on hyperspherical grids, another one based on the sequential Monte Carlo methods (i.e., particle filtering) [42]. The two resemble each other in the sense of Dirac mixture-based discrete modeling of on-manifold distributions. However, the former relies on grid points at fixed locations as elaborated in Sec. 5.1, whereas the latter (shown in Sec. 5.3) contains randomness in point layouts, for which a point-wise on-manifold UKF is proposed for enhancing the performance of the plain particle filtering method [64].

Moreover, we provide a generic description of the on-manifold discrete filter (MDF) in Sec. 5.2.1 to show how to incorporate Dirac mixtures of deterministic supports (e.g., based on grids) as discrete probabilistic models for nonparametric directional estimation. Besides its showcase for grid-based discrete quaternion estimation in Sec. 5.2.2, the MDF will be further customized to a reapproximation-based variant introduced in Sec. 6.6 of the upcoming chapter.

5.1 Grid-Based Discrete Modeling on Directional Manifolds

As the most basic topological component of directional domains, the unit hypersphere is a compact and bounded Riemannian manifold as previously introduced in Sec. 3.1. This allows us to generate



(A) ground truth (B) HEP grid (C) MCP grid

Figure 5.1.: Generating HEP and MCP grids on a synthesized distribution of antipodal symmetry on the unit sphere $\mathbb{S}^2$ in Example 8.1. The two antipodally symmetric modes of the underlying distribution are marked in (A). In (B) and (C), grid points are depicted by red dots with sizes proportional to the underlying densities.

a grid of points, using which underlying distributions of arbitrary forms can be approximated via Dirac mixtures or piecewise constant densities.

Given an arbitrary hyperspherical distribution characterized by its PDF $g\colon \mathbb{S}^{d-1} \to \mathbb{R}^+$, we perform discrete modeling using a Dirac mixture with its components located at grid points $\{\underline{x}_i\}_{i=1}^n \subset \mathbb{S}^{d-1}$. The resulting approximate distribution is expressed as

$$\hat{g} = \sum_{i=1}^{n} \nu_i \, \delta(\underline{x} - \underline{x}_i) \,, \quad \text{with} \quad \nu_i \propto g(\underline{x}_i) \tag{5.1}$$

being the grid weights that are the normalized function values of the underlying density $g(\underline{x}_i)$ (thus $\sum_{i=1}^n \nu_i = 1$) and δ the Dirac delta function.

5.1.1 Generating Hyperspherical Grids

Using grid points for state estimation can be traced back to the Wonham filter [65]. A first trial on directional manifolds can be found in [66] on the circular domain, and an extension for SE(2) estimation was made in [41] via Rao–Blackwellization. There, the circular domain is discretized using a set of equidistantly-spaced grid points $\{\underline{x}_i\}_{i=1}^n \subset \mathbb{S}^1$. For discretizing the unit hyperspheres $\mathbb{S}^{d-1}$ ($d \geq 3$), we provide the following two approaches for generating grid points of different layouts.

Hyperspherical equal partitioning: As pioneered in [41, 66], one straightforward strategy for discretizing unit hyperspheres is to use equidistant grid points. In [67], an algorithm was proposed for partitioning the unit hypersphere into patches of equal area, inducing a grid of approximately equidistant points $\{\underline{x}_i\}_{i=1}^n \subset \mathbb{S}^{d-1}$. The algorithm has been employed for IDS in Sec. 4.2.2 to obtain the planet samples in the tangent space at the mode of the von Mises–Fisher distribution. Due to its recursive algorithmic structure, the method can be executed very efficiently. This motivates us to directly apply this hyperspherical equal partitioning (HEP) method for generating grid points of uniform resolution for discrete probabilistic modeling on unit hyperspheres.

Mode-centric partitioning: The plain HEP discretizes the entire hypersphere uniformly without explicit consideration on the shape of underlying dispersion. For certain hyperspherical random

Algorithm 7: Mode-Centric Partitioning (MCP) (customized to unit quaternion manifold)

Input: number of orbits n_1, on-orbit resolution n_2, mode $\underline{\alpha}$
Output: set of grid points $\mathbb{X} \subset \mathbb{S}^3$

1 $\mathbb{X} \leftarrow \underline{\alpha}$;
2 $\{\tilde{\underline{x}}_j^{\iota}\}_{j=1}^{n_2} \leftarrow \mathtt{HEP}\,(\mathbb{S}^2, n_2)$;
3 **for** $k \leftarrow 1$ **to** n_1 **do**
4 **for** $j \leftarrow 1$ **to** n_2 **do**
5 $\underline{x}_{k,j} \leftarrow \underline{\alpha} \otimes \mathrm{Exp}_{\mathbb{1}}(\frac{\pi k}{2n_1}[\,0, \tilde{\underline{x}}_j^{\iota}\,]^{\top})$; // see (3.9) and (3.10)
6 $\mathbb{X} \leftarrow \mathbb{X} \cup \underline{x}_{k,j}$;
7 **return** $\mathbb{X}$;

variables, their uncertainties are inherently equipped with additional topological structure (e.g., unit quaternions follow antipodally symmetric distributions on $\mathbb{S}^3$). Also, the approximation quality of HEP grids depends purely on one number that determines the resolution, resulting in inefficient discrete models for representing densities of high concentration or in high-dimensional spaces. Dispersion of a hyperspherical random variable is almost never distributed uniformly on the hypersphere, and a high approximation resolution is often desired in the vicinity of the mode to alleviate potential grid point degeneracy for nonlinear filtering. Therefore, we propose the mode-centric partitioning (MCP) method. For clearness, we present the method in accordance with the geometric structure of the unit quaternion manifold introduced in Sec. 3.2.

As detailed in Algorithm 7, the method resembles the IDS procedure in Algorithm 5 (illustrated in Fig. 4.7) proposed for von Mises–Fisher distributions. However, since no parametric assumption is imposed on the discrete model, the resulting grid layout is purely determined by the configured resolution, with n_1 and n_2 denoting the number of orbits and number of points per orbit, respectively. We place the grid center at the given mode location $\underline{\alpha} \in \mathbb{S}^3$ (Algorithm 7, line 1). As introduced in Sec. 3.2.2, the tangent space on the unit quaternion manifold is bounded by a sphere $\mathbb{S}^2_{\pi/2}$ of radius $\pi/2$ under logarithm/exponential maps w.r.t. its local basis[1]. Therefore, the sample set $\{\tilde{\underline{x}}_j^{\iota}\}_{j=1}^{n_2}$ given by the basic HEP is scaled with uniform interval up to $\pi/2$ given n_1 and retracted back to $\mathbb{S}^3$ via the quaternion exponential map in (3.10) (Algorithm 7, lines 2–6). The resulting grid sample set has a cardinality of $|\mathbb{X}| = n_1 \cdot n_2 + 1$. We provide the following example to illustrate grid-based Dirac mixture modeling given by (5.1) on unit hyperspheres.

Example 8: *We consider two scenarios for grid-based hyperspherical discrete modeling.*

1. *An antipodally symmetric distribution is synthesized on $\mathbb{S}^2$ in the form of a von Mises–Fisher mixture $f_{\mathcal{VMFM}}(\underline{x}) = \sum_{i=1}^{3} 1/6 \cdot (f_{\mathcal{VMF}}(\underline{x}; \underline{\alpha}_i, \kappa_i) + f_{\mathcal{VMF}}(\underline{x}; -\underline{\alpha}_i, \kappa_i))$ with means $(\underline{\alpha}_1, \underline{\alpha}_2, \underline{\alpha}_3)$ and concentrations $(\kappa_1, \kappa_2, \kappa_3)$ configured as*

$$\left(\frac{1}{5\sqrt{5}}[\,0, 5, 10\,]^{\top}, \frac{1}{\sqrt{501}}[\,20, 1, 10\,]^{\top}, \frac{1}{\sqrt{326}}[\,1, 15, 10\,]^{\top}\right) \quad \textit{and} \quad (6, 8, 10)\,,$$

respectively. For discrete approximations of the distribution using Dirac mixtures, we perform MCP and HEP to generate spherical grids of resolutions $(n_1, n_2) = (20, 30)$ and $n = 20 \times 30 + 1 = 601$, respectively.

[1] Due to the additional antipodal symmetry, this is different from the case of IDS where the tangent space is bounded by $\mathbb{S}^2_{\pi}$.

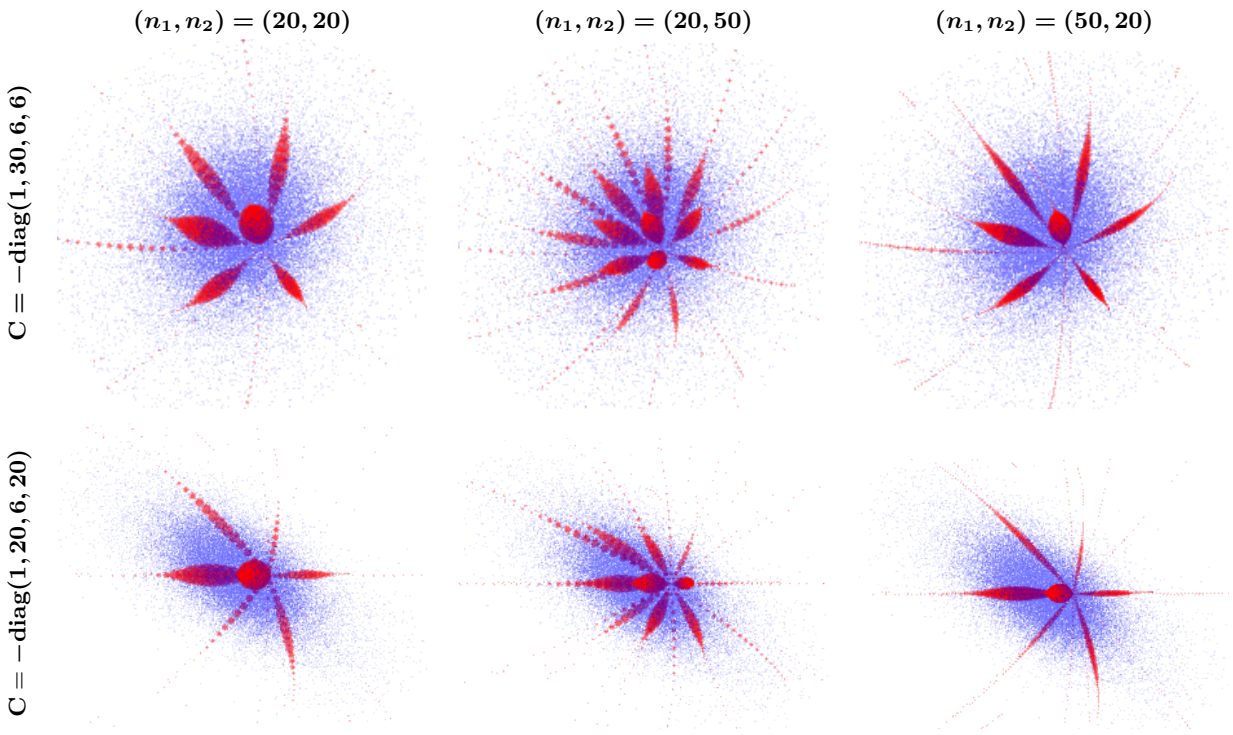


Figure 5.2.: MCP-based discrete modeling of Bingham distributions on $\mathbb{S}^3$ visualized in the tangent space at the modes in Example 8.2. Blue dots denote random samples drawn from the ground truth for characterizing the uncertainties. Grid points are depicted by red dots with sizes proportional to the underlying densities.

2. *We first set up two Bingham distributions defined in the concise form* $\mathcal{B}(\mathbf{C})$ *on the unit hypersphere* $\mathbb{S}^3$ *with parameter matrices* $\mathbf{C} \in \{-\mathrm{diag}(1, 30, 6, 6), -\mathrm{diag}(1, 20, 6, 20)\}$. *Afterward, we set both of their modes to a unit quaternion encoding a rotation of* $\theta = \pi/3$ *around axis* $[1, 0, 0]^\top$. *For both settings, we perform MCP with resolutions configured as* $(n_1, n_2) \in \{(20, 20), (20, 50), (50, 20)\}$.

As shown in Fig. 5.1, the synthesized distribution in the first scenario exhibits dispersion of antipodal symmetry on $\mathbb{S}^2$. *We depict the grid points using red dots with their sizes proportional to the Dirac component weights as defined in* (5.1). *Though configured with identical grid resolutions, the mode-centric variant induces evidently less information loss than the HEP grid for Dirac mixture modeling. In consideration of the additional topological structure of antipodal symmetry, the MCP grid only discretizes the hemisphere, thereby also enabling a more efficient discrete representation compared with the HEP method. Fig. 5.2 further shows the MCP-based Dirac mixture modeling on* $\mathbb{S}^3$. *To plot hyperspherical uncertainties, we draw* 5×10^4 *random samples (in blue) from each ground truth distribution and map them to the tangent space at the mode (thereby enabling visualizations in* $\mathbb{R}^3$*) together with the weighted grid points (in red). While all configurations of the MCP grids deliver effective Dirac mixture models with consideration of the antipodal symmetry of unit quaternion states, the approximation quality differs over different grid resolutions.*

5.2 On-Manifold Discrete Filtering

As a first trial of nonparametric on-manifold directional filtering, we generalize the Dirac mixture expression in (5.1) from the hypersphere to generic directional manifolds $\mathbb{M}$ using deterministic

supports (e.g., at hyperspherical grids) for discrete representation of the posterior density. As discrete modeling approaches may differ given different state space topology, we categorize this class of filtering methods under the name of the on-manifold discrete filter (MDF) [41, 51, 100].

5.2.1 Generic Design

We represent the posterior density at time step t using a Dirac mixture-based discrete model

$$f_t^{\mathrm{e}}(\underline{x}_t) = \sum_{i=1}^{n} \nu_{t,i}^{\mathrm{e}} \, \delta(\underline{x}_t - \underline{x}_{t,i}^{\mathrm{e}}) \,, \tag{5.2}$$

with $\{\underline{x}_{t,i}^{\mathrm{e}}\}_{i=1}^{n} \subset \mathbb{M}$ and $\{\nu_{t,i}^{\mathrm{e}}\}_{i=1}^{n}$ ($\sum_i^n \nu_{t,i}^{\mathrm{e}} = 1$ holds) being locations and weights of the Dirac components, respectively. Given the generic setup of nonlinear recursive Bayesian estimation in Sec. 2.1, we introduce the generic design of the on-manifold discrete filter as follows [41, 43].

Prediction step: By specifying (2.4) with the proposed discrete model in (5.2), we obtain the prediction step of MDF in the form[2]

$$\begin{aligned} f_{t+1}^{\mathrm{p}}(\underline{x}_t) &= \int_{\mathbb{M}} f_t^{\mathrm{T}}(\underline{x}_{t+1}|\underline{x}_t) f_t^{\mathrm{e}}(\underline{x}_t) \, \mathrm{d}\underline{x}_t \\ &= \int_{\mathbb{M}} f_t^{\mathrm{e}}(\underline{x}_t) \int_{\mathbb{W}} f(\underline{x}_{t+1}|\underline{w}_t, \underline{x}_t) \, f_t^{\boldsymbol{w}}(\underline{w}_t) \, \mathrm{d}\underline{w}_t \, \mathrm{d}\underline{x}_t \\ &= \sum_{i=1}^{n} \nu_{t,i}^{\mathrm{e}} \int_{\mathbb{W}} f(\underline{x}_{t+1}|\underline{w}_t, \underline{x}_{t,i}^{\mathrm{e}}) \, f_t^{\boldsymbol{w}}(\underline{w}_t) \, \mathrm{d}\underline{w}_t \,. \end{aligned}$$

We further represent the noise distribution using a Dirac mixture $f_t^{\boldsymbol{w}}(\underline{w}_t) = \sum_{k=1}^{n_{\mathrm{w}}} \nu_{t,k}^{\mathrm{w}} \, \delta(\underline{w}_t - \underline{w}_{t,k})$, where $\sum_{k=1}^{n_{\mathrm{w}}} \nu_{t,k}^{\mathrm{w}} = 1$ holds. Taking the generic form of the system model in (2.1), we obtain

$$f_{t+1}^{\mathrm{p}}(\underline{x}_{t+1}) = \sum_{i=1}^{n} \sum_{k=1}^{n_{\mathrm{w}}} \nu_{t,i}^{\mathrm{e}} \, \nu_{t,k}^{\mathrm{w}} \, \delta\big(\underline{x}_{t+1} - \underline{a}(\underline{x}_{t,i}^{\mathrm{e}}, \underline{w}_{t,k})\big) \,. \tag{5.3}$$

Here, the on-manifold state sample set $\{\underline{x}_{t,i}^{\mathrm{e}}\}_{i=1}^{n} \subset \mathbb{M}$ is combined with the noise sample set $\{\underline{w}_{t,k}\}_{k=1}^{n_{\mathrm{w}}} \subset \mathbb{W}$ via Cartesian product and propagated together through the system dynamics $\underline{a} \colon \mathbb{M} \times \mathbb{W} \to \mathbb{M}$. Further, (5.3) can be rewritten as

$$f_{t+1}^{\mathrm{p}}(\underline{x}_{t+1}) = \sum_{r=1}^{n_{\mathrm{p}}} \nu_{t+1,r}^{\mathrm{p}} \, \delta(\underline{x}_{t+1} - \underline{x}_{t+1,r}^{\mathrm{p}}) \,, \tag{5.4}$$

with $\underline{x}_{t+1,r}^{\mathrm{p}} = a(\underline{x}_{t,i}^{\mathrm{e}}, \underline{w}_{t,k})$ and $\nu_{t+1,r}^{\mathrm{p}} = \nu_{t,i}^{\mathrm{e}} \, \nu_{t,k}^{\mathrm{w}}$ (thus $n_{\mathrm{p}} = n \cdot n_{\mathrm{w}}$ and $r = (i-1) \cdot n_{\mathrm{w}} + k$) denoting the location and the weight of each propagated Dirac component, respectively. If a prior estimate is required at this stage (e.g., for consecutive predictions), the raw prior sample set $\{\underline{x}_{t+1,r}^{\mathrm{p}}\}_{r=1}^{n_{\mathrm{p}}}$ is to be approximated from cardinality $n \cdot n_{\mathrm{w}}$ to n given the setting of a specific discrete model.

Update step: Suppose a specific discrete filter directly fuses the measurement $\hat{\underline{z}}_{t+1}$ to the prior Dirac mixture in (5.4), the update step in (2.3) is then given by

$$\begin{aligned} f_{t+1}^{\mathrm{e}}(\underline{x}_{t+1}|\hat{\underline{z}}_{t+1}) &\propto f_{t+1}^{\mathrm{L}}(\hat{\underline{z}}_{t+1}|\underline{x}_{t+1}) f_{t+1}^{\mathrm{p}}(\underline{x}_{t+1}) \\ &= \sum_{r=1}^{n_{\mathrm{p}}} f_{t+1}^{\mathrm{L}}(\hat{\underline{z}}_{t+1}|\underline{x}_{t+1,r}^{\mathrm{p}}) \, \nu_{t+1,r}^{\mathrm{p}} \, \delta(\underline{x}_{t+1} - \underline{x}_{t+1,r}^{\mathrm{p}}) \,. \end{aligned} \tag{5.5}$$

[2] For brevity, we omit the dependency on measurements $\hat{\underline{z}}_{1:t}$ in (2.4) for the upcoming derivations.

Algorithm 8: MCP-based Discrete Quaternion Filter

Input: posterior set $\{(\underline{x}^{\mathrm{e}}_{t,i}, \nu^{\mathrm{e}}_{t,i})\}_{i=1}^{n}$, noise set $\{(\underline{w}_{t,k}, \nu^{\mathrm{w}}_{t,k})\}_{k=1}^{n_{\mathrm{w}}}$, measurement $\hat{\underline{z}}_{t+1}$
Output: posterior set $\{(\underline{x}^{\mathrm{e}}_{t+1,i}, \nu^{\mathrm{e}}_{t+1,i})\}_{i=1}^{n}$

/* prediction step */
1 $\{(\underline{x}^{\mathrm{p}}_{t+1,r}, \nu^{\mathrm{p}}_{t+1,r})\}_{r=1}^{n_{\mathrm{P}}} \leftarrow$ `propagate` $(\{(\underline{x}^{\mathrm{e}}_{t,i}, \nu^{\mathrm{e}}_{t,i})\}_{i=1}^{n}, \{(\underline{w}_{t,k}, \nu^{\mathrm{w}}_{t,k})\}_{k=1}^{n_{\mathrm{w}}})$; // see (5.3)
2 $\underline{\alpha}^{\mathrm{e}}_{t+1} \leftarrow$ `getMode` $(\{(\underline{x}^{\mathrm{p}}_{t+1,r}, \nu^{\mathrm{p}}_{t+1,r})\}_{r=1}^{n_{\mathrm{P}}})$;
3 $\{\underline{x}^{\mathrm{e}}_{t+1,i}\}_{i=1}^{n} \leftarrow$ `transportGrid` $(\{\underline{x}^{\mathrm{e}}_{t,i}\}_{i=1}^{n}, \underline{\alpha}^{\mathrm{e}}_{t}, \underline{\alpha}^{\mathrm{e}}_{t+1})$; // see (5.6)
4 $\{\nu^{\mathrm{p}}_{t+1,i}\}_{i=1}^{n} \leftarrow$ `reallocateWeights` $(\{(\underline{x}^{\mathrm{p}}_{t+1,r}, \nu^{\mathrm{p}}_{t+1,r})\}_{r=1}^{n_{\mathrm{P}}}, \{\underline{x}^{\mathrm{e}}_{t+1,i}\}_{i=1}^{n})$; // see (5.7)
/* update step */
5 **for** $i \leftarrow 1$ **to** n **do**
6 $\quad \nu^{\mathrm{e}}_{t+1,i} \leftarrow \nu^{\mathrm{p}}_{t+1,i} \cdot f^{\mathrm{L}}_{t+1}(\hat{\underline{z}}_{t+1} | \underline{x}^{\mathrm{e}}_{t+1,i})$;
7 **return** $\{(\underline{x}^{\mathrm{e}}_{t+1,i}, \nu^{\mathrm{e}}_{t+1,i})\}_{i=1}^{n}$;

The reweighted Dirac mixture then needs to be approximated in accordance with the form of the employed discrete model.

Implementing this procedure depends on the specific discrete modeling and filtering approach. In principle, both grid-based discrete probabilistic models in Sec. 5.1 can be incorporated into the generic design of on-manifold discrete filtering. In fact, studies on both variants are undergoing actively. For showcasing the introduced discrete filtering scheme, we first provide one variant based on the MCP grid in the upcoming section for recursive quaternion estimation [43]. For detailed elaboration on the HEP-based discrete filtering, the author provides a reading list of own publications including [45, 46, 48]. Besides the grid-based discrete filtering, another variant using Dirac mixture reapproximation will be provided in Sec. 6.6.

5.2.2 Grid-Based Discrete Quaternion Filter

One concern arises when customizing the general discrete filtering scheme to the unit quaternion manifold using the MCP-based grid representation – the propagated weighted grid points $\{(\underline{x}^{\mathrm{p}}_{t+1,r}, \nu^{\mathrm{p}}_{t+1,r})\}_{r=1}^{n_{\mathrm{P}}}$ in (5.4) have an arbitrary layout and a size of $n \cdot n_{\mathrm{w}}$ instead of n (see Algorithm 8, line 1). Therefore, an additional step is desired for reallocating them to a mode-centric grid (of fixed layout) placed at the new mode. As introduced in Sec. 3.2.2, the Hamilton product indicates an SO(4) rotation via the matrix representation in (3.5), under which the unit hypersphere $\mathbb{S}^3$ is closed. Thus, we transport the MCP grid $\{\underline{x}^{\mathrm{e}}_{t,i}\}_{i=1}^{n} \subset \mathbb{S}^3$ centered at the previous mode $\underline{\alpha}^{\mathrm{e}}_{t}$ to the new one $\underline{\alpha}^{\mathrm{e}}_{t+1}$ given by the system propagation via

$$\underline{x}^{\mathrm{e}}_{t+1,i} = \underline{x}^{\mathrm{e}}_{t,i} \otimes (\underline{\alpha}^{\mathrm{e}}_{t})^{-1} \otimes \underline{\alpha}^{\mathrm{e}}_{t+1} = \mathbf{Q}^{\lrcorner}_{\underline{\alpha}^{\mathrm{e}}_{t+1}} (\mathbf{Q}^{\lrcorner}_{\underline{\alpha}^{\mathrm{e}}_{t}})^{\top} \underline{x}^{\mathrm{e}}_{t,i} =: \mathbf{R}^{t+1}_{t} \underline{x}^{\mathrm{e}}_{t,i} \,, \tag{5.6}$$

without changing the grid topology (since $\mathbf{R}^{t+1}_{t} \in \mathrm{SO}(4)$ denotes a rigid body rotation) (Algorithm 8, lines 2–3). Each propagated grid point $\underline{x}^{\mathrm{p}}_{t+1,r}$ is then absorbed into the transported grid by reallocating its weight among the nearest m grid points $\{\underline{x}^{\mathrm{e}}_{t+1,j}\}_{j=1}^{m}$ proportionally to the inverse geodesic distance $\mathcal{D}(\underline{x}^{\mathrm{e}}_{t+1,j}, \underline{x}^{\mathrm{p}}_{t+1,r})$ (Algorithm 8, line 4) [43]. Executed in a cumulative manner through all propagated grid points, this procedure can be expressed as

$$\nu^{\mathrm{p}}_{t+1,j} \leftarrow \nu^{\mathrm{p}}_{t+1,j} + \frac{\mathcal{D}(\underline{x}^{\mathrm{e}}_{t+1,j}, \underline{x}^{\mathrm{p}}_{t+1,r})}{\sum_{j=1}^{m} \mathcal{D}(\underline{x}^{\mathrm{e}}_{t+1,j}, \underline{x}^{\mathrm{p}}_{t+1,r})} \nu^{\mathrm{p}}_{t+1,r} \,, \tag{5.7}$$

with $\mathcal{D}(\underline{x}^{\mathrm{e}}_{t+1,j}, \underline{x}^{\mathrm{p}}_{t+1,r}) = 1/\arccos(|(\underline{x}^{\mathrm{e}}_{t+1,j})^{\top}\underline{x}^{\mathrm{p}}_{t+1,r}|)$ being the inverse geodesic distance w.r.t. arc length. Afterward, we fuse the measurement $\hat{\underline{z}}_{t+1}$ into the grid-based representation by reweighting the grid points using their likelihoods (Algorithm 8, lines 5–6).

As a variant of the presented grid-based quaternion filter, we also proposed a method in [46] that involves modifying the original HEP algorithm [67] for generating equidistant hyperhemispherical grid for quaternion estimation using the hyperhemisphere. Due to the uniform topology, the approach enables precomputation of a system-transition matrix given a time-invariant transition density $f^{\mathrm{T}}(\underline{x}_{t+1}|\underline{x}_t)$ for efficiently predicting grid weights without reallocation. Similar filter designs were also introduced for the general hyperspherical [45] and hypertoridal manifolds [47], and extensions to the composite directional domains were made based on Rao–Blackwellization for planar motion estimation [41, 48]. As studies on grid-based directional filtering are still ongoing with many variants emerging, we only focus on its basic design in this thesis and do not elaborate on evaluating its performance[3]. A detailed evaluation of the on-manifold discrete filtering scheme is provided in combination with the Dirac mixture reapproximation technique in Sec. 6.6.1 of the upcoming chapter.

5.3 On-Manifold Unscented Particle Filtering

Sequential Monte Carlo methods are a straightforward strategy for nonparametric filtering and are directly applicable to directional domains. In theory, they allow exact modeling of arbitrary uncertainties provided that adequately many particles are available. In practice, recursive Bayesian estimation for directional domains typically has to deal with strong nonlinearities and high-dimensional spaces (e.g., on hyperspheres or the unit dual quaternion manifold), inducing particles to quickly degenerate. A plain particle filter deploys the transition density as the proposal distribution and thus disregards the recent observation, leading to deteriorated performance specifically in the face of heavy-tailed distributions, non-stationary models, or peaky likelihoods.

The aforementioned issues may occur in a combined manner, making deploying PFs much less appealing for directional estimation compared with Euclidean cases. In order to mitigate such difficulties, we propose a topology-aware unscented particle filtering scheme based on its original linear counterpart [64], making the sequential Monte Carlo methods deployable for nonparametric directional estimation.

5.3.1 Sequential Monte Carlo Methods and the Unscented Particle Filter

The PF [42] deployed to directional manifolds exploits a Dirac mixture distribution for representing the posterior density in the form $f(\underline{x}_{0:t}|\hat{\underline{z}}_{1:t}) = \sum_{i=1}^{n} \nu_{t,i}\, \delta(\underline{x}_{0:t} - \underline{x}_{0:t,i})$ with $\{\underline{x}_{0:t,i}\}_{i=1}^{n} \subset \mathbb{M}$ being the particles and $\{\nu_{t,i}\}_{i=1}^{n}$ their weights summing to one, i.e., $\sum_{i=1}^{n} \nu_{t,i} = 1$. As sampling from the true density is infeasible in theory, the common strategy is to exploit a proposal distribution $g(\underline{x}_{t,i}|\underline{x}_{0:t-1,i}, \hat{\underline{z}}_{1:t})$ that is easy to sample, updating the importance weight of each particle in a recursive fashion according to

$$\nu_{t,i} = \nu_{t-1,i}\, \frac{f(\hat{\underline{z}}_t|\underline{x}_{t,i}) f(\underline{x}_{t,i}|\underline{x}_{t-1,i})}{g(\underline{x}_{t,i}|\underline{x}_{0:t-1,i}, \hat{\underline{z}}_{1:t})}\,. \tag{5.8}$$

[3] Evaluations of specific variants of grid-based discrete filters are elaborated in own publications [43–48].

The most popular choice for the proposal distribution is the transition prior $f(\underline{x}_{t,i}|\underline{x}_{t-1,i})$ [42]. While fulfilling the recursive weight update (5.8) in a concise manner, this strategy disregards the most recently observed information from $\hat{\underline{z}}_t$. As previously mentioned, this leads to deteriorated performance for state estimation with strong nonlinearities, heavy-tailed distributions, non-stationary models or peaky likelihoods, and the circumstance may even exacerbate for directional estimation.

In Euclidean spaces, the unscented particle filter (UPF) was proposed by exploiting particlewise UKFs to obtain proposal distributions that incorporate the recently obtained measurement, leading to evidently superior tracking performance over a plain PF. Moreover, it has been theoretically justified that the UPF delivers a convergence rate that is independent of the dimensionality, which also motivates us to adopt it for directional domains. As introduced in [49, 50], the key component for such extensions is to establish the particlewise UKF on directional manifolds. In the upcoming section, we demonstrate such techniques in a unified manner with customization to hyperspheres and the manifold of unit dual quaternions (as a showcase for composite directional domains).

5.3.2 Particlewise On-Manifold UKF

As shown in Sec. 3.1, the gnomonic projection enables an unbounded probabilistic interpretation in the tangent space at points on hyperspheres without issues such as the warping and truncation effects (induced by the logarithm/exponential maps and orthographic projection/retraction, respectively). Therefore, we set up a Gaussian distribution in the (augmented) tangent space of each on-manifold particle $\underline{x}_i \in \mathbb{M}$ w.r.t. its local basis and propose the topology-aware on-manifold UKF (MUKF) based upon the gnomonic projection/retraction. We specify these operations for the considered directional domains as follows.

Unit hyperspheres: At any $\underline{v} \in \mathbb{S}^{d-1}$, a tangent space can be established w.r.t. a local basis $\mathcal{E}_\mathrm{v}$. We concatenate the basis vectors column wise to form a matrix $\mathbf{E}_\mathrm{v} \in \mathbb{R}^{d\times(d-1)}$ for transforming points in local coordinates to global coordinates. Based thereon, the hyperspherical gnomonic projection/retraction (given in (3.3) and (3.4), respectively) can be derived w.r.t. $\mathcal{E}_\mathrm{v}$ as

$$\begin{aligned}\tilde{\underline{x}}^\iota &= \texttt{projectSph}(\underline{v}, \underline{x}) = \mathbf{E}_\mathrm{v}^\top \big(\underline{x}/(\underline{v}^\top \underline{x}) - \underline{v}\big) \in \mathbb{R}^{d-1}\,, \quad \text{and}\\ \underline{x} &= \texttt{retractSph}(\underline{v}, \tilde{\underline{x}}^\iota) = \big(\underline{v} + \mathbf{E}_\mathrm{v}\, \tilde{\underline{x}}^\iota\big)\big/\sqrt{1+\|\tilde{\underline{x}}^\iota\|^2} \in \mathbb{S}^{d-1} \subset \mathbb{R}^d\,,\end{aligned}$$

respectively, with $\underline{x}$ being an arbitrary point on the unit hypersphere. Setting a local basis $\mathcal{E}_\mathrm{v}$ is flexible. For states on the $\mathbb{S}^3$, a convenient choice of $\mathcal{E}_\mathrm{v}$ is to use the quaternion concomitant bases (QCBs) $\mathcal{A}^\circ_\mathrm{x}$ given in (3.18). We therefore exploit them in the upcoming case of the unit dual quaternion manifold.

Unit dual quaternion manifold: We establish the so-called locally augmented tangent space (LATS) based on the SE(3) representation in (3.12) of unit dual quaternions. Given an arbitrary unit dual quaternion $\underline{v} = [\,\underline{v}_\mathrm{r}^\top, \underline{v}_\mathrm{s}^\top\,]^\top \in \mathbb{DH}_1$, we augment the tangent space $\mathbb{T}_{\underline{v}_\mathrm{r}}\mathbb{S}^3$ at the real part $\underline{v}_\mathrm{r}$ with the dual part expressed w.r.t. the local basis $\mathcal{A}^\lrcorner_{\mathrm{x}_\mathrm{r}}$ taken from the QCBs in (3.18), inducing a six-dimensional Euclidean space for particlewise quantification of the on-manifold uncertainty. As proposed in [50], any point $\underline{x} = [\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top\,]^\top \in \mathbb{DH}_1$ can be mapped to the LATS at $\underline{v} \in \mathbb{DH}_1$ via the augmented gnomonic projection[4]

$$\tilde{\underline{x}}^\iota = \texttt{projectDH1}(\underline{v}, \underline{x}) = [\,(\tilde{\underline{x}}_\mathrm{r}^\iota)^\top, (\tilde{\underline{x}}_\mathrm{s}^\iota)^\top\,]^\top \in \mathbb{R}^6\,,$$

[4] The antipodal symmetry underlying in the unit dual quaternion topology needs to be first addressed.

Algorithm 9: Particlewise On-Manifold UKF (MUKF)

Input: $(\underline{x}_t^{\mathrm{e}}, \mathbf{C}_t^{\mathrm{e}})$, measurement $\hat{\underline{z}}_{t+1}$
Output: $(\underline{x}_{t+1}^{\mathrm{e}}, \mathbf{C}_{t+1}^{\mathrm{e}})$

/* prediction step */
1 $\{(\tilde{\underline{\sigma}}_{t,k}^{\iota}, \omega_k)\}_{k=1}^{m} \leftarrow \texttt{sampleDeterministic}\,(f_{\mathcal{N}}(\underline{0}, \mathbf{C}_t^{\mathrm{e}}))$;
2 $\{\underline{\sigma}_{t,k}\}_{k=1}^{m} \leftarrow \texttt{retract}\,(\underline{x}_t^{\mathrm{e}}, \{\tilde{\underline{\sigma}}_{t,k}^{\iota}\}_{k=1}^{m})$; // domain-specific
3 $\{\underline{\sigma}_{t+1,k}\}_{k=1}^{m} \leftarrow \texttt{propagate}\,(\{\underline{\sigma}_{t,k}\}_{k=1}^{m})$;
4 $\underline{x}_{t+1}^{\mathrm{p}} \leftarrow \texttt{average}\,(\{(\underline{\sigma}_{t+1,k}, \omega_k)\}_{k=1}^{m})$;
5 $\{\tilde{\underline{\sigma}}_{t+1,k}^{\iota}\}_{k=1}^{m} \leftarrow \texttt{project}\,(\underline{x}_{t+1}^{\mathrm{p}}, \{\underline{\sigma}_{t+1,k}\}_{k=1}^{m})$; // domain-specific
6 $\mathbf{C}_{t+1}^{\mathrm{p}} \leftarrow \sum_k \omega_k\, \tilde{\underline{\sigma}}_{t+1,k}^{\iota} (\tilde{\underline{\sigma}}_{t+1,k}^{\iota})^{\top}$;
7 $\{\underline{z}_{t+1,k}\}_{k=1}^{m} \leftarrow \texttt{predictMeasurement}\,(\underline{x}_{t+1}^{\mathrm{p}}, \{\tilde{\underline{\sigma}}_{t+1,k}^{\iota}\}_{k=1}^{m})$;
/* update step */
8 $\underline{z}_{t+1}^{\mathrm{p}} \leftarrow \sum_k \omega_k\, \underline{z}_{t+1,k}$;
9 $\mathbf{C}_1 \leftarrow \sum_k \omega_k (\underline{z}_{t+1,k} - \underline{z}_{t+1}^{\mathrm{p}})(\underline{z}_{t+1,k} - \underline{z}_{t+1}^{\mathrm{p}})^{\top}$;
10 $\mathbf{C}_2 \leftarrow \sum_k \omega_k\, \tilde{\underline{\sigma}}_{t+1,k}^{\iota} (\underline{z}_{t+1,k} - \underline{z}_{t+1}^{\mathrm{p}})^{\top}$;
11 $\mathbf{K} \leftarrow \mathbf{C}_2 (\mathbf{C}_1)^{-1}$;
12 $\tilde{\underline{x}}_{t+1}^{\iota} \leftarrow \mathbf{K}(\hat{\underline{z}}_{t+1} - \underline{z}_{t+1}^{\mathrm{p}})$;
13 $\mathbf{C}_{t+1}^{\mathrm{e}} \leftarrow \mathbf{C}_{t+1}^{\mathrm{p}} - \mathbf{K}\,\mathbf{C}_1\,\mathbf{K}^{\top}$;
14 $\underline{x}_{t+1}^{\mathrm{e}} \leftarrow \texttt{retract}\,(\underline{x}_{t+1}^{\mathrm{p}}, \tilde{\underline{x}}_{t+1}^{\iota})$; // domain-specific
15 **return** $(\underline{x}_{t+1}^{\mathrm{e}}, \mathbf{C}_{t+1}^{\mathrm{e}})$;

with $\tilde{\underline{x}}_{\mathrm{r}}^{\iota} = \texttt{projectSph}(\underline{v}_{\mathrm{r}}, \underline{x}_{\mathrm{r}}) \in \mathbb{R}^3$ and $\tilde{\underline{x}}_{\mathrm{s}}^{\iota} = \mathbf{R}_{\mathrm{v_r}}^{\top}(\underline{t}_{\mathrm{x}} - \underline{t}_{\mathrm{v}}) \in \mathbb{R}^3$. $\mathbf{R}_{\mathrm{v_r}}$ denotes the SO(3) rotation matrix given by the real part $\underline{v}_{\mathrm{r}}$ of $\underline{v}$ as shown in (3.8). $\underline{t}_{\mathrm{x}}$ and $\underline{t}_{\mathrm{v}}$ are the translation terms represented by the dual quaternions $\underline{x}$ and $\underline{v}$, respectively. The derivation for mapping the dual part is provided in Appendix A.3. Inversely, any point $\tilde{\underline{x}}^{\iota} = [\,(\tilde{\underline{x}}_{\mathrm{r}}^{\iota})^{\top}, (\tilde{\underline{x}}_{\mathrm{s}}^{\iota})^{\top}\,]^{\top}$ in the LATS at $\underline{v} \in \mathbb{DH}_1$ can be mapped back to the manifold $\mathbb{DH}_1$ via the augmented gnomonic retraction

$$\underline{x} = \texttt{retractDH1}(\underline{v}, \tilde{\underline{x}}^{\iota}) = [\,\underline{x}_{\mathrm{r}}^{\top}, \underline{x}_{\mathrm{s}}^{\top}\,]^{\top} \in \mathbb{DH}_1\,,$$

with $\underline{x}_{\mathrm{r}} = \texttt{retractSph}(\underline{v}_{\mathrm{r}}, \tilde{\underline{x}}_{\mathrm{r}}^{\iota}) \in \mathbb{S}^3$ and $\underline{x}_{\mathrm{s}} = 0.5\mathbf{A}_{\mathrm{x_r}}^{\lrcorner}(\mathbf{R}_{\mathrm{v_r}}\tilde{\underline{x}}_{\mathrm{s}}^{\iota} + \underline{t}_{\mathrm{v}}) \in \mathbb{T}_{\underline{x}_{\mathrm{r}}}\mathbb{S}^3$.

Based on the particlewise (locally augmented) tangent space (TS) enabled by the (augmented) gnomonic projection/retraction, we adapt the UKF to the topology of the considered directional domains and propose a unified on-manifold design shown in Algorithm 9. As separately elaborated in [49] and [50], each particle is assigned with a covariance matrix to characterize the on-manifold uncertainty in its (LA)TS in the form of a zero-mean Gaussian distribution w.r.t. a local basis. We first draw UT-based deterministic samples in the (LA)TS at each particle $\underline{x}_t^{\mathrm{e}}$ from the previous posterior density[5], retract them to the underlying manifold, and propagate them through the system dynamics in (2.1) (Algorithm 9, lines 1–3). The propagated samples are then averaged to obtain the prior state $\underline{x}_{t+1}^{\mathrm{p}}$ to locate a new (LA)TS, to which the propagated samples are mapped via (augmented) gnomonic projection, and the in-tangent-space covariance is computed to characterize the prior proposal (Algorithm 9, lines 4–6). After the measurement is predicted for each deterministic sample using the measurement model in (2.2) (gnomonic retractions are here exploited), we perform an ordinary UKF update step in the (LA)TS at $\underline{x}_{t+1}^{\mathrm{p}}$ to obtain the posterior estimate w.r.t. the local basis (Algorithm 9, lines 7–13). The mean of the Gaussian posterior is

[5] Since the MUKF is executed for every particle following the identical fashion, the particle index is omitted in Algorithm 9.

Algorithm 10: On-Manifold Unscented Particle Filter (MUPF)

Input: $\{(\underline{x}^{\mathrm{e}}_{t,i}, \nu^{\mathrm{e}}_{t,i}, \mathbf{C}^{\mathrm{e}}_{t,i})\}_{i=1}^{n}$, measurement $\hat{\underline{z}}_{t+1}$
Output: $\{(\underline{x}^{\mathrm{e}}_{t+1,i}, \nu^{\mathrm{e}}_{t+1,i}, \mathbf{C}^{\mathrm{e}}_{t+1,i})\}_{i=1}^{n}$

1 **for** $i \leftarrow 1$ **to** n **do**
  /* particlewise MUKF */
2 $(\underline{x}^{\mathrm{e}}_{t+1,i}, \mathbf{C}^{\mathrm{e}}_{t+1,i}) \leftarrow \texttt{MUKF}(\underline{x}^{\mathrm{e}}_{t,i}, \mathbf{C}^{\mathrm{e}}_{t,i}, \hat{\underline{z}}_{t+1})$;
  /* importance weighting */
3 $\tilde{\underline{x}}^{\iota}_{t+1,i} \leftarrow \texttt{sampleRnd}\left(f_{\mathcal{N}}(\underline{0}, \mathbf{C}^{\mathrm{e}}_{t+1,i})\right)$;
4 $\hat{\underline{x}}_{t+1,i} \leftarrow \texttt{retract}\left(\underline{x}^{\mathrm{e}}_{t+1,i}, \tilde{\underline{x}}^{\iota}_{t+1,i}\right)$; // domain-specific
5 $\nu^{\mathrm{e}}_{t+1,i} \leftarrow \nu^{\mathrm{e}}_{t,i} \frac{f(\hat{\underline{z}}_{t+1}|\hat{\underline{x}}_{t+1,i}) f(\hat{\underline{x}}_{t+1,i}|\underline{x}_{t,i})}{g(\hat{\underline{x}}_{t+1,i}|\underline{x}_{0:t,i}, \hat{\underline{z}}_{1:t+1})}$;
6 $\{\nu^{\mathrm{e}}_{t+1,i}\}_{i=1}^{n} \leftarrow \texttt{normalize}(\{\nu^{\mathrm{e}}_{t+1,i}\}_{i=1}^{n})$;
7 $\{(\underline{x}^{\mathrm{e}}_{t+1,i}, \nu^{\mathrm{e}}_{t+1,i}, \mathbf{C}^{\mathrm{e}}_{t+1,i})\}_{i=1}^{n} \leftarrow \texttt{resample}(\{(\hat{\underline{x}}_{t+1,i}, \nu^{\mathrm{e}}_{t+1,i}, \mathbf{C}^{\mathrm{e}}_{t+1,i})\}_{i=1}^{n})$;
8 **return** $\{(\underline{x}^{\mathrm{e}}_{t+1,i}, \nu^{\mathrm{e}}_{t+1,i}, \mathbf{C}^{\mathrm{e}}_{t+1,i})\}_{i=1}^{n}$;

then mapped back to the manifold via the (augmented) gnomonic retraction, and we obtain the new location of the particle (Algorithm 9, line 14).

5.3.3 The On-Manifold Unscented Particle Filter

As introduced in [49, 50], the proposed on-manifold UKF naturally enables the unscented particle filtering scheme on the considered directional manifolds. As shown in Algorithm 10, each particle from the posterior Dirac mixture is associated with a 3-tuple including its location $\underline{x}^{\mathrm{e}}_{t,i}$, weight $\nu^{\mathrm{e}}_{t,i}$, and the covariance matrix $\mathbf{C}^{\mathrm{e}}_{t,i}$ in its (LA)TS for quantifying the on-manifold proposal density. By running the proposed MUKF for each particle and taking the resulting UKF posterior as the particlewise proposal $g(\underline{x}_{t+1,i}|\underline{x}_{0:t,i}, \hat{\underline{z}}_{1:t+1})$, the recently observed evidence $\hat{\underline{z}}_{t+1}$ is inherently considered for updating the importance weight following (5.8) (Algorithm 10, lines 1–5). Afterward, we normalize the updated weights and perform a typical sampling-importance resampling procedure to obtain equally weighted particles (Algorithm 10, lines 6–7).

The presented MUPF scheme was first specified on unit hyperspheres and deployed to quaternion-based orientation estimation in [49]. Further, it was applied to the manifold of unit dual quaternions in [50] for rigid body motion estimation. Both variants showed the superiority of the proposed methodology over the plain particle filter and parametric directional filters w.r.t. accuracy and robustness, especially under conditions of strong nonlinearities and peaky likelihoods. In these cases, the parametric filters mostly lose tracking, and a plain PF requires a large amount of particles for normal functionality. In the following case study, we set up a similarly difficult scenario compared with [50] for nonlinear SE(3) estimation using the variant on $\mathbb{DH}_1$ for showcasing its convergence behavior over numbers of particles and runtime.

5.3.4 Case Study: Nonlinear SE(3) Estimation

We specify the system model in (2.1) on $\mathbb{DH}_1$ into the following form

$$\underline{\boldsymbol{x}}_{t+1} = \underline{\boldsymbol{x}}_t \boxtimes \underline{\boldsymbol{w}}_t \,,$$

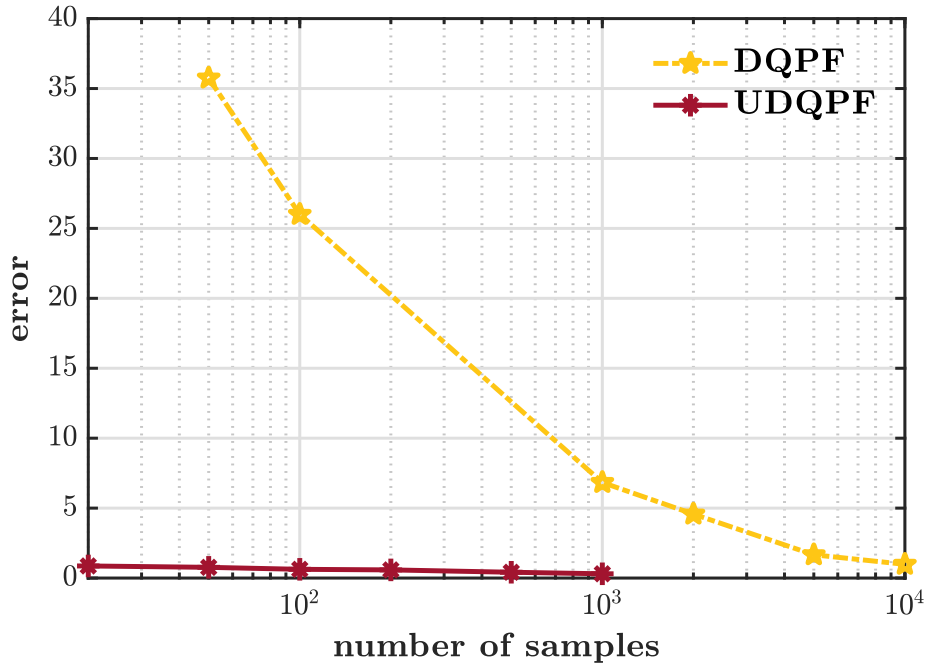


Figure 5.3.: Error over numbers of particles given by the evaluated filters.

with $\underline{\boldsymbol{x}}_t \in \mathbb{DH}_1$ and $\underline{\boldsymbol{w}}_t \in \mathbb{DH}_1$ being the unit dual quaternion state and the noise term as defined in (3.12), respectively. Similar to the evaluation scenario in [50], we synthesize the noise term $\underline{\boldsymbol{w}}_t = [\underline{\boldsymbol{w}}_{\mathrm{r},t}^\top, \underline{\boldsymbol{w}}_{\mathrm{s},t}^\top]^\top$ to be state-dependent. Its real part represents a rotation of angle $\boldsymbol{\theta}_t^{\mathrm{w}}$ around axis $\underline{\boldsymbol{u}}_t^{\mathrm{w}}$, with the former following a von Mises distribution and the latter a von Mises–Fisher distribution parameterized as $\mathcal{VM}(\pi/12, 200)$ and $\mathcal{VMF}(1/\sqrt{3} \cdot [1,1,1]^\top, 200)$, respectively. The dual part $\underline{\boldsymbol{w}}_{\mathrm{s},t}$ of the noise term encodes a translation given by $\underline{\boldsymbol{t}}_t^{\mathrm{w}} = (\boldsymbol{x}_{t,1} + \boldsymbol{w}_{t,1})\, \breve{\underline{\boldsymbol{t}}}_t^{\mathrm{w}}$. Here, $\breve{\underline{\boldsymbol{t}}}_t^{\mathrm{w}}$ denotes a state-independent component that is assumed to follow a Gaussian distribution $\mathcal{N}([5,2,10]^\top, 0.1 \cdot \mathbf{I}_{3\times3})$. $\boldsymbol{x}_{t,1}$ and $\boldsymbol{w}_{t,1}$ are the first elements in the state and noise vectors, respectively.

The measurement model in (2.2) is specified to be

$$\underline{\boldsymbol{z}}_t = \left(\underline{\boldsymbol{x}}_t \boxtimes [1,0,0,0,0,\hat{\underline{z}}_0^\top]^\top \boxtimes \underline{\boldsymbol{x}}_t\right)_{6:8} + \underline{\boldsymbol{v}}_t\,,$$

which transforms a point initialized at $\hat{\underline{z}}_0 = [2,3,4]^\top$ according to the rigid body motion represented by the state $\underline{\boldsymbol{x}}_t$ following (3.13). The additive noise $\underline{\boldsymbol{v}}_t$ is assumed to follow a zero-mean Gaussian distribution $\mathcal{N}(\underline{0}_3, \boldsymbol{\Sigma}^{\underline{\boldsymbol{v}}})$ with $\boldsymbol{\Sigma}^{\underline{\boldsymbol{v}}} = 0.1 \cdot \mathbf{I}_{3\times3}$.

We compare the proposed MUPF-variant, the unscented dual quaternion particle filter (UDQPF), with a plain (dual quaternion) particle filter (DQPF) using the same resampling method. The numbers of particles in the proposed UDQPF are $\{20, 50, 100, 200, 500, 1000\}$, and the DQPF is equipped with $\{50, 100, 1000, 2000, 5000, 10000\}$ particles. The error metric for unit dual quaternion states is the geodesic distance (expressed without unit) defined on $\mathbb{DH}_1$ as in [9,89], where rotation and translation errors are quantified simultaneously[6]. Overall 1000 Monte Carlo runs of simulations are executed with 30 steps in each run. Based thereon, deviations are summarized into RMSE of the last estimate in each run as in Sec. 4.4.

[6] A degenerate case of the same metric on planar dual quaternion manifold was employed as the residual term for pose graph optimization in [101].

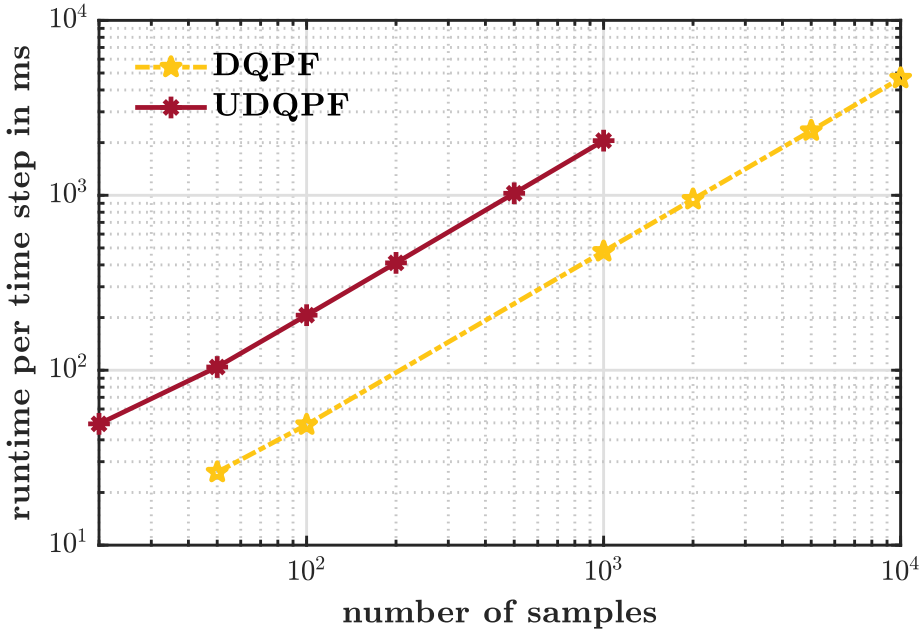


Figure 5.4.: Runtime over numbers of particles given by the evaluated filters.

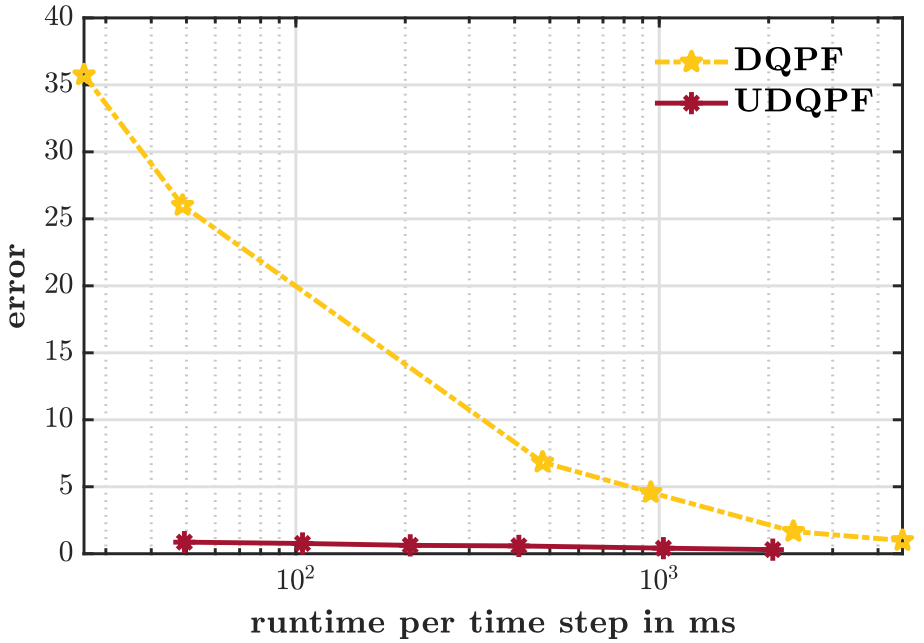


Figure 5.5.: Error over runtime given by the evaluated filters.

As shown in Fig. 5.3, the UDQPF delivers a better tracking accuracy even using 20 particles compared with the ordinary DQPF exploiting 5000 particles. Further shown in Fig. 5.4, the UDQPF takes more runtime than the DQPF given same numbers of particles due to the additional particlewise MUKF. However, it exhibits a much better tracking accuracy and convergence over runtime as shown in Fig. 5.5. Combined with our findings in [49, 50], it can be validated that the proposed on-manifold unscented particle filtering scheme exhibits superior performance in challenging directional estimation scenarios (e.g., involving strong nonlinearities, state-dependency in noise or peaky likelihoods) over the plain PF and competitive parametric filters.

5.4 Short Summary

In this chapter, we introduce two novel on-manifold methodologies for nonparametric filtering in directional domains – one based on discrete modeling using grids of different topology, another one based on sequential Monte Carlo methods with topology-aware enhancements for difficult estimation scenarios. While both of them exploit Dirac mixtures for nonparametric modeling of arbitrary distributions on directional manifolds, the former has fixed locations of Dirac components due to the underlying grid structure, whereas the latter contains randomness in particles that support Dirac components.

As discussed in Example 8 together with the illustrations in Fig. 5.1 and Fig. 5.2, the grid-based discrete modeling scheme has the drawback of fixed grid structure that is nonadaptive to the shape of the underlying distribution. The Monte Carlo-based method models the on-manifold uncertainty adaptively to its shape but in theory contains redundant information due to the randomness in particles. Though both of them are successfully deployed to nonparametric directional estimation and deliver promising results, we may still search for an alternative Dirac mixture-based probabilistic model somewhere in between – that has the adaptiveness of the Monte Carlo-based method for shape approximation and deterministic supports on the manifold (similar to a grid). This interest motivates us to propose the on-manifold Dirac mixture reapproximation paradigm in the upcoming chapter.

CHAPTER 6

On-Manifold Dirac Mixture Reapproximation

As introduced in Chapter 4, the parametric probabilistic models impose strong assumptions on the uncertain nature of directional random variables. Grid-based or Monte Carlo-based schemes as discussed in Chapter 5 are possible solutions to nonparametric modeling and filtering. However, they may lack the representation efficiency due to the nonadaptive grid topology or the randomness in particles. In this chapter, we propose a unified paradigm for enabling efficient Dirac mixture-based nonparametric modeling on directional manifolds.

We first introduce the concept of reapproximation-based discrete modeling in Sec. 6.1. Given a *source* Dirac mixture of many components representing an unknown directional distribution, the on-manifold Dirac mixture reapproximation (MDMR) paradigm is established in Sec. 6.2 to yield a *target* Dirac mixture with configurable numbers of components while respecting the underlying domain topology. As further showcased in Sec. 6.3 and Sec. 6.4 with MDMR-variants on various directional manifolds, the target Dirac components exhibit a deterministic layout that is adaptive to the shape of underlying dispersion, thereby considerably improving the representation efficiency of discrete probabilistic models. Built upon the hyperspherical variant of the proposed MDMR, we further introduce a two-stage reapproximation & reconstruction (R&R) procedure in Sec. 6.5 for continuous modeling of unknown distributions on $\mathbb{S}^{d-1}$ ($d \geq 3$) in the form of von Mises–Fisher mixtures. Furthermore, we integrate the proposed MDMR paradigm into the generic discrete filtering scheme given in Sec. 5.2 and introduce the on-manifold reapproximation discrete filter (MRDF) in Sec. 6.6 for nonparametric directional estimation with unknown form of system noise. We showcase the proposed MRDF with the customization to unit hyperspheres, and underline its superior performance in nonlinear spherical estimation over parametric and Monte Carlo-based directional filters.

6.1 Efficient Discrete Modeling via Reapproximation

As previously shown in Chapter 5, Dirac mixtures provide an intuitive solution to representing arbitrary directional distributions in a nonparametric manner. Given raw samples of empirical data collected from a stochastic process, a Dirac mixture model can be established by placing each component on each sample point and associating it with the corresponding weight. However, such a naive modeling solution is inefficient due to the inherent randomness of raw samples. In [71], an optimization-based scheme was proposed for reducing Dirac mixture components in Euclidean spaces. There, the Cramér–von Mises distance (CvMD) was generalized for measuring the statistical divergence between two multivariate Dirac mixtures based on the concept of the localized cumulative distribution with an isotropic Gaussian kernel. Consequently, a large set of Euclidean samples can be reduced to a sample set of smaller size in the sense of least CvMD. The approach has

the merit of establishing smooth characterization of discrete probabilistic models, further enabling the optimization-based scheme for reduction. However, computing the proposed metric introduces approximations at several stages and relies on a fixed weighting scheme. Moreover, the source and target samples are required to be centralized at the same spot. As an extension to directional domains, a heuristic approach was introduced in [102] for the unit sphere $\mathbb{S}^2$. Since this extension is essentially a transformation from the linear version w.r.t. the spherical coordinate system, the aforementioned issues remain unresolved, and further generalization to generic directional domains is hardly feasible.

In the upcoming section of this chapter, we propose a paradigm of on-manifold Dirac mixture reapproximation to enable efficient discrete modeling of unknown and arbitrary directional distributions given a source Dirac mixture of many components. The paradigm is developed based on theoretically sound derivations and is presented in a unified fashion. By introducing the new concept of the on-manifold localized cumulative distribution (MLCD), the on-manifold Cramér–von Mises distance (MCvMD) is established for quantifying the statistical divergence between Dirac mixtures adaptively to the confining topological structure. Optimally preserving the probability mass of the source Dirac mixture in the least-MCvMD sense, the target Dirac mixture reapproximates the underlying unknown distribution with configurable number of components in a dispersion-adaptive layout. Customizing the generic MDMR procedure for a certain directional manifold only requires selecting a proper kernel with a geometry-aware distance metric and a corresponding weighting function. For showcasing the general applicability of the MDMR paradigm, we provide several variants, including ones for the unit circle, unit hyperspheres of arbitrary dimensions, the planar dual quaternion manifold, and the unit dual quaternion manifold. Optimization of the MCvMD is confined to the underlying manifold structure. Instead of implicitly handling the geometric constraints (e.g., via Lagrange multipliers), we employ the Riemannian optimization scheme [103] with inherent consideration of the confining geometry for better convergence properties. For that, we provide closed-form gradients and Hessians in symbolic form to guarantee runtime performance.

6.2 On-Manifold Dirac Mixture Reapproximation: A Unified Paradigm

Suppose we are provided with a source Dirac mixture of many components placed at raw samples from empirical directional data. We aim to reduce the influence of the randomness in raw data without introducing strong assumptions imposed by parametric models. For that, we choose to reapproximate the unknown underlying distribution on the manifold $\mathbb{M}$ using another Dirac mixture of fewer components at more representative locations.

6.2.1 Problem Formulation

For a source Dirac mixture of components located at a sample set $\widehat{\mathbb{X}} = \{\underline{\widehat{x}}_i\}_{i=1}^{\widehat{n}} \subset \mathbb{M}$ on the manifold $\mathbb{M}$ with weights $\{\widehat{\nu}_i\}_{i=1}^{\widehat{n}}$, its density function is expressed as

$$f_{\widehat{\mathbb{X}}}(\underline{x}) = \sum_{i=1}^{\widehat{n}} \widehat{\nu}_i \, \delta(\underline{x} - \underline{\widehat{x}}_i) \,, \quad \text{with} \quad \sum_{i=1}^{\widehat{n}} \widehat{\nu}_i = 1 \tag{6.1}$$

normalizing the sample weights. As introduced in (5.1), $\delta(\underline{x} - \underline{\widehat{x}}_i)$ denotes the Dirac delta function that characterizes the mass cluster around each sample $\underline{\widehat{x}}_i$. The desired target Dirac mixture is equipped with components of a much smaller size $n \ll \widehat{n}$ placed at the sample set $\mathbb{X} = \{\underline{x}_i\}_{i=1}^{n} \subset \mathbb{M}$,

reapproximating the underlying distribution in the form

$$f_{\mathbb{X}}(\underline{x}) = \sum_{i=1}^{n} \nu_i \, \delta(\underline{x} - \underline{x}_i) \,, \quad \text{with} \quad \sum_{i=1}^{n} \nu_i = 1 \,. \tag{6.2}$$

As previously mentioned, the reapproximation scheme is built upon the minimization of a certain metric quantifying the statistical divergence $\mathcal{D}$ between the two Dirac mixtures located at the source and target sets ($\widehat{\mathbb{X}}$ and $\mathbb{X}$, respectively). In general, this procedure can be formulated as[1]

$$\mathbb{X}^* = \underset{\mathbb{X} \subset \mathbb{M}}{\arg\min} \, \mathcal{D}\big(\mathbb{X}, \widehat{\mathbb{X}}\big) \,. \tag{6.3}$$

Designing a meaningful expression $\mathcal{D}\big(\mathbb{X}, \widehat{\mathbb{X}}\big)$ for measuring the statistical divergence between two Dirac mixtures is not as trivial as it is for two continuous probabilistic models. The major theoretical obstacle lies in the fact that the two discrete models do not share any support, thereby invalidating the applicability of typical choices such as the Kullback–Leibler divergence or the Hellinger distance. For real scalar random variables, the empirical distribution function can be exploited for establishing such measures based on the cumulative distribution. But a straightforward extension to the multivariate case leads to a nonunique expression. As a viable modification, the localized cumulative distribution in [71] provides an extension for the concept of cumulative distribution from scalar to multivariate Euclidean spaces by incorporating an isotropic Gaussian kernel into the integration. This further induces a multivariate generalization of the Cramér–von Mises distance for measuring the difference of two discrete models in continuous domains in a unique and symmetric manner. In reference to this basic approach for Euclidean spaces, we introduce a generic design for on-manifold Dirac mixture reapproximation shown in the upcoming sections.

6.2.2 On-Manifold Localized Cumulative Distribution

Definition 1: *Given a directional random variable $\underline{\boldsymbol{x}} \in \mathbb{M}$ following a probability density function $f : \mathbb{M} \to \mathbb{R}^+$, its on-manifold localized cumulative distribution (MLCD) is defined as*

$$\mathcal{F}(\underline{\alpha}, \tau) = \int_{\mathbb{M}} f(\underline{x}) \, \boldsymbol{k}(\underline{x}; \underline{\alpha}, \tau) \, d\underline{x} \,.$$

$\boldsymbol{k}(\underline{x}; \underline{\alpha}, \tau)$ denotes a manifold-specific kernel function located at $\underline{\alpha} \in \mathbb{M}$ that is isotropic and smooth, with $\tau > 0$ characterizing the kernel concentration.

Given the definition above, the MLCD of the source Dirac mixture in (6.1) follows

$$\widehat{\mathcal{F}}(\underline{\alpha}, \tau) = \int_{\mathbb{M}} \sum_{i=1}^{\hat{n}} \widehat{\nu}_i \, \delta(\underline{x} - \widehat{\underline{x}}_i) \, \boldsymbol{k}(\underline{x}; \underline{\alpha}, \tau) \, \mathrm{d}\underline{x} = \sum_{i=1}^{\hat{n}} \widehat{\nu}_i \, \boldsymbol{k}(\widehat{\underline{x}}_i; \underline{\alpha}, \tau) \,. \tag{6.4}$$

Similarly, the target MLCD is expressed as

$$\mathcal{F}(\underline{\alpha}, \tau) = \int_{\mathbb{M}} \sum_{i=1}^{n} \nu_i \, \delta(\underline{x} - \underline{x}_i) \, \boldsymbol{k}(\underline{x}; \underline{\alpha}, \tau) \, \mathrm{d}\underline{x} = \sum_{i=1}^{n} \nu_i \, \boldsymbol{k}\left(\underline{x}_i; \underline{\alpha}, \tau\right) \,. \tag{6.5}$$

[1] For brevity, the weights of the Dirac components in the objective of (6.3) are not written as inputs, but they are considered in the actual computation as shown later.

We provide the choice of topology-aware kernel functions respecting specific directional manifolds at a later stage in Sec. 6.3 and Sec. 6.4 [51, 52]. Based thereon, the proposed MLCD yields a local characterization of the underlying unknown distribution w.r.t. the kernel location $\underline{\alpha}$ and the concentration parameter τ.

6.2.3 On-Manifold Cramér–von Mises Distance

Given the target and source Dirac mixtures $f_{\mathbb{X}}$ and $f_{\widehat{\mathbb{X}}}$, respectively, the on-manifold Cramér–von Mises distance (MCvMD) quantifies their statistical divergence based on their MLCDs $\mathcal{F}(\underline{\alpha}, \tau)$ and $\widehat{\mathcal{F}}(\underline{\alpha}, \tau)$ in the form

$$\mathcal{D}(\mathbb{X}, \widehat{\mathbb{X}}) = \int_{\mathbb{R}^+} \mathcal{W}(\tau) \int_{\mathbb{M}} \left(\mathcal{F}(\underline{\alpha}, \tau) - \widehat{\mathcal{F}}(\underline{\alpha}, \tau)\right)^2 \mathrm{d}\underline{\alpha}\, \mathrm{d}\tau \,. \tag{6.6}$$

$\mathcal{W}(\tau)$ is a continuous weighting function that regulates the impact of the concentration τ in the kernel function. The MCvMD above integrates the weighted squared difference between the two MLCDs over possible kernel locations $\underline{\alpha} \in \mathbb{M}$ and concentration parameters $\tau > 0$. Further, we expand the expression in (6.6) and obtain the objective function in (6.3) following

$$\mathcal{D}(\mathbb{X}, \widehat{\mathbb{X}}) =: \mathcal{D}_1(\mathbb{X}) - 2\mathcal{D}_2(\mathbb{X}, \widehat{\mathbb{X}}) + \mathcal{D}_3(\widehat{\mathbb{X}}) \,,$$

with the three components expressed as

$$\begin{aligned}
\mathcal{D}_1(\mathbb{X}) &= \int_{\mathbb{R}_+} \mathcal{W}(\tau) \int_{\mathbb{M}} \mathcal{F}^2(\underline{\alpha}, \tau)\, \mathrm{d}\underline{\alpha}\, \mathrm{d}\tau \,, \\
\mathcal{D}_2(\mathbb{X}, \widehat{\mathbb{X}}) &= \int_{\mathbb{R}_+} \mathcal{W}(\tau) \int_{\mathbb{M}} \mathcal{F}(\underline{\alpha}, \tau)\, \widehat{\mathcal{F}}(\underline{\alpha}, \tau)\, \mathrm{d}\underline{\alpha}\, \mathrm{d}\tau \,, \\
\mathcal{D}_3(\widehat{\mathbb{X}}) &= \int_{\mathbb{R}_+} \mathcal{W}(\tau) \int_{\mathbb{M}} \widehat{\mathcal{F}}^2(\underline{\alpha}, \tau)\, \mathrm{d}\underline{\alpha}\, \mathrm{d}\tau \,.
\end{aligned}$$

By substituting the source and target MLCDs, $\widehat{\mathcal{F}}(\underline{\alpha}, \tau)$ and $\mathcal{F}(\underline{\alpha}, \tau)$, with their formulae in (6.4) and (6.5), respectively, the three components above then follow

$$\begin{aligned}
\mathcal{D}_1(\mathbb{X}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \nu_i\, \nu_j \int_{\mathbb{R}_+} \mathcal{W}(\tau) \int_{\mathbb{M}} k(\underline{x}_i; \underline{\alpha}, \tau)\, k(\underline{x}_j; \underline{\alpha}, \tau)\, \mathrm{d}\underline{\alpha}\, \mathrm{d}\tau \,, \\
\mathcal{D}_2(\mathbb{X}, \widehat{\mathbb{X}}) &= \sum_{i=1}^{n} \sum_{k=1}^{\widehat{n}} \nu_i\, \widehat{\nu}_k \int_{\mathbb{R}_+} \mathcal{W}(\tau) \int_{\mathbb{M}} k(\underline{x}_i; \underline{\alpha}, \tau)\, k(\widehat{\underline{x}}_k; \underline{\alpha}, \tau)\, \mathrm{d}\underline{\alpha}\, \mathrm{d}\tau \,, \\
\mathcal{D}_3(\widehat{\mathbb{X}}) &= \sum_{k=1}^{\widehat{n}} \sum_{l=1}^{\widehat{n}} \widehat{\nu}_k\, \widehat{\nu}_l \int_{\mathbb{R}_+} \mathcal{W}(\tau) \int_{\mathbb{M}} k(\widehat{\underline{x}}_k; \underline{\alpha}, \tau)\, k(\widehat{\underline{x}}_l; \underline{\alpha}, \tau)\, \mathrm{d}\underline{\alpha}\, \mathrm{d}\tau \,.
\end{aligned} \tag{6.7}$$

The three components in (6.7) have the same form of summands, which involves integrating the product of kernel function values at sample pairs over the on-manifold kernel locations and concentrations. We extract this general form and express it as the *MCvMD unit*

$$\mathcal{Q}(\underline{u}, \underline{v}) = \int_{\mathbb{R}_+} \mathcal{W}(\tau)\, \mathcal{P}(\underline{u}, \underline{v}, \tau)\, \mathrm{d}\tau \,, \tag{6.8}$$

with the integral over kernel locations formulated as

$$\mathcal{P}(\underline{u}, \underline{v}, \tau) = \int_{\mathbb{M}} k(\underline{u}; \underline{\alpha}, \tau)\, k(\underline{v}; \underline{\alpha}, \tau)\, \mathrm{d}\underline{\alpha} \,. \tag{6.9}$$

Based on the MCvMD unit expressed above, the components of the objective function $\mathcal{D}(\mathbb{X}, \widehat{\mathbb{X}}) = \mathcal{D}_1(\mathbb{X}) - 2\mathcal{D}_2(\mathbb{X}, \widehat{\mathbb{X}}) + \mathcal{D}_3(\widehat{\mathbb{X}})$ in (6.3) then boil down to conciser forms as follows

$$\begin{aligned} \mathcal{D}_1(\mathbb{X}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \nu_i \, \nu_j \, \boldsymbol{Q}(\underline{x}_i, \underline{x}_j) \,, \\ \mathcal{D}_2(\mathbb{X}, \widehat{\mathbb{X}}) &= \sum_{i=1}^{n} \sum_{k=1}^{\widehat{n}} \nu_i \, \widehat{\nu}_k \, \boldsymbol{Q}(\underline{x}_i, \widehat{\underline{x}}_k) \,, \\ \mathcal{D}_3(\widehat{\mathbb{X}}) &= \sum_{k=1}^{\widehat{n}} \sum_{l=1}^{\widehat{n}} \widehat{\nu}_k \, \widehat{\nu}_l \, \boldsymbol{Q}(\widehat{\underline{x}}_k, \widehat{\underline{x}}_l) \,. \end{aligned} \tag{6.10}$$

In the resulting expression of the proposed distance measure, both the kernel function in (6.9) and the weighting function in (6.8) are yet to be specified. In principle, the distance metric in the kernel function needs to cohere with the underlying manifold geometry. Moreover, it needs to have isotropic dispersion, such that the induced integral is symmetric and unique. For instance, a von Mises–Fisher-like kernel $k(\underline{x}; \underline{\alpha}, \tau) = \exp(\tau \underline{\alpha}^\top \underline{x})$ can be chosen for hyperspherical Dirac mixture reapproximation as given in [51]. As a result, the resulting MCvMD only depends on the relative geodesic distance w.r.t. $\underline{u}^\top \underline{v}$ between the two sample points $\underline{u}$ and $\underline{v}$ in (6.9). The weighting function $\mathcal{W}(\tau)$ in (6.8) can be chosen freely as long as the resulting product is integrable. However, in consideration of a practical implementation and runtime efficiency, we propose to choose it in a specific way such that the integral (6.8) yields a closed-form solution.

In the upcoming Sec. 6.3, we elaborate customization of the proposed MDMR paradigm to unit hyperspheres $\mathbb{S}^{d-1}$ of arbitrary dimensions $d \geq 3$ [51]. Afterward, we briefly show several other MDMR-variants for other common directional domains with necessary kernel and weighting functions in Sec. 6.4. Based on the hyperspherical variant of MDMR, a reapproximation & reconstruction procedure is introduced in Sec. 6.5 for continuous modeling of unknown distributions on $\mathbb{S}^{d-1}$ in the form of von Mises–Fisher mixtures.

6.3 Hyperspherical Dirac Mixture Reapproximation

Random variables on unit hyperspheres $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ $(d \geq 3)$ appear in various applications such as quaternion-based rigid body motion estimation [1], orientation tracking [78], scene understanding [6, 19, 84], and text document representation using vector space models [7, 104].

As previously mentioned, we choose a von Mises–Fisher-like kernel $k(\underline{x}; \underline{\alpha}, \tau) = \exp(\tau \underline{\alpha}^\top \underline{x})$ for the hyperspherical MDMR-variant, and the integral in (6.9) is then specified as

$$\mathcal{P}(\underline{u}, \underline{v}, \tau) = \int_{\mathbb{S}^{d-1}} \exp\left(\tau \, (\underline{u} + \underline{v})^\top \underline{\alpha}\right) \mathrm{d}\underline{\alpha} \,,$$

with kernel locations $\underline{\alpha} \in \mathbb{S}^{d-1}$ $(d \geq 3)$. We normalize the sum of unit vectors $\underline{u}$ and $\underline{v}$ to be $\widehat{\underline{u} + \underline{v}} = (\underline{u} + \underline{v})/\|\underline{u} + \underline{v}\|$, with $\|\underline{u} + \underline{v}\| = (2 + 2\,\underline{u}^\top \underline{v})^{1/2}$ being the norm. The integral then follows

$$\mathcal{P}(\underline{u}, \underline{v}, \tau) = \int_{\mathbb{S}^{d-1}} \exp\left(\tau \|\underline{u} + \underline{v}\| \, \widehat{\underline{u} + \underline{v}}^\top \alpha\right) \mathrm{d}\underline{\alpha} \,,$$

As evident from the definition in (2.5), the formula above induces the normalization constant of the von Mises–Fisher distribution $\mathcal{VMF}\big(\underline{\alpha}; \widehat{\underline{u}+\underline{v}}, \tau(2+2\,\underline{u}^\top\underline{v})^{1/2}\big)$. According to (2.6), we obtain

$$\mathcal{P}(\delta,\tau) = \frac{(2\pi)^{d/2}\,\mathcal{I}_{d/2-1}\big(\tau\sqrt{2+2\delta}\,\big)}{\big(\tau\sqrt{2+2\delta}\,\big)^{d/2-1}}\,, \quad \text{with} \quad \delta = \underline{u}^\top\underline{v}$$

quantifying the hyperspherical distance coherently to the manifold geometry. $\mathcal{I}_{d/2-1}$ denotes the modified Bessel function of the first kind and of order $d/2-1$. Further, we design the weighting function in (6.8) to be

$$\mathcal{W}(\tau) = \exp(-\epsilon\tau)\,\tau^{d/2-2}\,, \tag{6.11}$$

with $\epsilon > 2$ being precomputed given the configuration. The integral in (6.8) is then derived as

$$\mathcal{Q}(\delta) = \frac{(2\pi)^{d/2}}{(\sqrt{2+2\delta}\,)^{d/2-1}}\int_0^\infty e^{-\epsilon\tau}\tau^{-1}\mathcal{I}_{d/2-1}\big(\tau\sqrt{2+2\delta}\,\big)\,\mathrm{d}\tau\,. \tag{6.12}$$

The right-hand side of the formula above refers to the Laplace transform of functions in the form $\tau^{-1}\mathcal{I}_\nu(a\tau)$, which is given by [105, 17.13.112] according to the formula

$$\mathcal{L}\big\{\tau^{-1}\mathcal{I}_\nu(a\tau)\big\}(\epsilon) = \int_0^\infty e^{-\epsilon\tau}\tau^{-1}\mathcal{I}_\nu(a\tau)\,\mathrm{d}\tau = \nu^{-1}a^\nu\big(\sqrt{\epsilon^2-a^2}+\epsilon\big)^{-\nu} \tag{6.13}$$

under the condition $\mathrm{Re}(\epsilon) > \mathrm{Re}(a)$. As we have $\epsilon > 2$ and the hyperspherical distance metric δ is in $[-1,1]$ for $a = \sqrt{2+2\delta}$, this condition is fulfilled, and we obtain the following closed form of the integral (6.12) over possible concentration parameters

$$\begin{aligned}\mathcal{Q}(\delta) &= \frac{(2\pi)^{d/2}}{(\sqrt{2+2\delta}\,)^{d/2-1}}\mathcal{L}\Big\{\tau^{-1}\mathcal{I}_{d/2-1}\big(\tau\sqrt{2+2\delta}\,\big)\Big\}(\epsilon)\\ &= \frac{(2\pi)^{d/2}}{d/2-1}\Big(\sqrt{\epsilon^2-2\delta-2}+\epsilon\Big)^{1-d/2}\,.\end{aligned} \tag{6.14}$$

Given the MCvMD unit in (6.8) customized to hyperspheres above, the hyperspherical Cramér–von Mises distance (HCvMD) according to (6.10) then consists of the following components

$$\begin{aligned}\mathcal{D}_1(\mathbb{X}) &= \sum_{i=1}^n\sum_{j=1}^n \nu_i\,\nu_j\,\mathcal{Q}(\underline{x}_i,\underline{x}_j) = \frac{(2\pi)^{d/2}}{d/2-1}\sum_{i=1}^n\sum_{j=1}^n \nu_i\,\nu_j\Big(\sqrt{\epsilon^2-2\,\underline{x}_i^\top\underline{x}_j-2}+\epsilon\Big)^{1-d/2}\,,\\
\mathcal{D}_2(\mathbb{X},\widehat{\mathbb{X}}) &= \sum_{i=1}^n\sum_{k=1}^{\widehat{n}} \nu_i\,\widehat{\nu}_k\,\mathcal{Q}(\underline{x}_i,\underline{\widehat{x}}_k) = \frac{(2\pi)^{d/2}}{d/2-1}\sum_{i=1}^n\sum_{k=1}^{\widehat{n}} \nu_i\,\widehat{\nu}_k\Big(\sqrt{\epsilon^2-2\,\underline{x}_i^\top\underline{\widehat{x}}_k-2}+\epsilon\Big)^{1-d/2}\,,\\
\mathcal{D}_3(\widehat{\mathbb{X}}) &= \sum_{k=1}^{\widehat{n}}\sum_{l=1}^{\widehat{n}} \widehat{\nu}_k\,\widehat{\nu}_l\,\mathcal{Q}(\underline{\widehat{x}}_k,\underline{\widehat{x}}_l) = \frac{(2\pi)^{d/2}}{d/2-1}\sum_{k=1}^{\widehat{n}}\sum_{l=1}^{\widehat{n}} \widehat{\nu}_k\,\widehat{\nu}_l\Big(\sqrt{\epsilon^2-2\,\underline{\widehat{x}}_k^\top\underline{\widehat{x}}_l-2}+\epsilon\Big)^{1-d/2}\,.\end{aligned} \tag{6.15}$$

As shown above, the proposed MCvMD specified with the von Mises–Fisher-like kernel and the weighing function in (6.11) quantifies the statistical divergence between two Dirac mixtures analytically on arbitrary-dimensional unit hyperspheres $\mathbb{S}^{d-1}$ of $d \geq 3$. Due to the isotropic dispersion of the kernel function, the metric only depends on the relative distance between pairs of samples respecting the manifold geometry (here, according to the curve length). This leads to a symmetric and unique measure for comparing the probability mass characterized by two hyperspherical discrete models. Furthermore, the closed-form expressions given in (6.15) are continuous functions w.r.t. the location on the unit hypersphere, leading to computational advantage for the optimization-based reapproximation scheme.

6.3.1 Reapproximation via Riemannian Optimization

Based on the HCvMD components formulated in (6.15), we now concretize the optimization problem in (6.3) for hyperspherical Dirac mixture reapproximation (HDMR). Considering the implementation convenience and runtime efficiency, elements in the source and target sets ($\widehat{\mathbb{X}}$ and $\mathbb{X}$, respectively) are concatenated into matrices $\widehat{\mathbf{X}}$ and $\mathbf{X}$, respectively. The optimization problem for reapproximation is then reformulated into the form

$$\mathbf{X}^* = \underset{\mathbf{X}\in\mathbb{OB}(d,n)}{\arg\min}\ \mathcal{D}(\mathbf{X},\widehat{\mathbf{X}})\,, \tag{6.16}$$

with the matrix variable $\mathbf{X}$ belonging to the oblique manifold $\mathbb{OB}(d,n) \subset \mathbb{R}^{d\times n}$ as each column vector in $\mathbf{X}$ is confined to the unit hypersphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ [85]. From the general perspective of Dirac mixture reapproximation for directional distributions, the desired target Dirac mixture supports are always obtained by solving a constrained optimization problem similar to (6.16) with the constraint imposed by the underlying manifold geometry. For this, conventional solutions handle the constraint implicitly by integrating it into the objective function, e.g., via a Lagrange multiplier. In the proposed MDMR paradigm, however, we choose the Riemannian optimization framework for explicit handling of the manifold constraints instead of reformulating the objective to achieve better numerical and convergence properties [85, 106].

6.3.2 Implementation

As shown in the weighting function (6.11), the parameter ϵ can in theory be chosen freely as long as $\epsilon > 2$ holds, which is required by the Laplace transform-based closed form in (6.14). Considering the numerical stability of the optimization in practice, ϵ is given by a piecewise function of continuous form as follows

$$\epsilon = \begin{cases} 2 + n^{-d}\,, & \text{for } n \leq n_\epsilon \\ 2 + (d\cdot n)^{-1}\,, & \text{otherwise} \end{cases}\,. \tag{6.17}$$

By guaranteeing the continuity of the function (6.17), the threshold ϵ is then determined according to the dimension d and the number n of target Dirac components.

Given the closed-form HCvMD components in (6.15), the objective in (6.16) is a real-valued and smooth function over the oblique manifold $\mathbb{OB}(d,n) \subset \mathbb{R}^{d\times n}$. Therefore, we choose to exploit the Riemannian trust-region (RTR) approach based on the implementation in [86] for reapproximation in the sense of least HCvMD. Exploiting the RTR method for solving optimization problems on directional manifolds requires the gradients and Hessians of the objective w.r.t. the variable in its ambient space. Based thereon, the Riemannian gradients and Hessians are derived with consideration of the manifold geometry. Though methods such as automatic differentiation or finite difference are applicable for such purposes, symbolic forms are almost always preferable given their advantage in numerical stability and runtime efficiency. For the case of hyperspherical Dirac mixture reapproximation, we provide symbolic forms of the required gradients and Hessians in Appendix C.1.2 based on the generic expressions given in Appendix C.1.1. Moreover, we set the target Dirac components to be of equal weights, whereas the source Dirac components are allowed to have uniform or non-uniform weights. For demonstrating the efficacy of the MDMR paradigm specified on $\mathbb{S}^{d-1}$, namely, the HDMR method, we show the following examples in various scenarios of hyperspherical discrete modeling [51].

Example 9: *We first perform the proposed hyperspherical Dirac mixture reapproximation on the unit sphere* $\mathbb{S}^2$ *using* $\widehat{n} = 20000$ *samples drawn randomly from the following continuous distributions (therefore, the source Dirac mixture components are equally weighted).*

1. *We exploit von Mises–Fisher distributions* $\mathcal{VMF}([\,0,0,1\,]^\top, \kappa)$ *as introduced in Sec. 2.2.1 with concentrations* $\kappa \in \{6, 12, 18\}$*. The target Dirac mixtures are set to have* $n \in \{5, 20, 30\}$ *components.*
2. *We configure Bingham distributions* $\mathcal{B}(\mathbf{C})$ *as introduced in Sec. 2.2.2 with parameter matrices* $\mathbf{C} \in \{-\mathrm{diag}(20,4,0), -\mathrm{diag}(10,10,0), -\mathrm{diag}(10,4,0)\}$*. Reapproximations to target Dirac mixtures of* $n \in \{20, 42, 82\}$ *components are performed.*

Reapproximation results for the two scenarios above are depicted in Fig. 6.1 and Fig. 6.2. Given source Dirac mixtures of supports at random samples (green dots), the proposed HDMR method delivers convincing results on the unit sphere – discrete models given by the target Dirac mixtures (red dots) reapproximate underlying continuous distributions in a dispersion-adaptive manner with layouts attaining deterministic patterns. Furthermore, additional topological properties of spherical uncertainties, e.g., the antipodal symmetry of Bingham distributions, are also preserved via reapproximation. Compared with a random sampling-based discrete model, the one given by the HDMR represents the underlying distribution in a much more efficient way.

Example 10: *We synthesize von Mises–Fisher mixtures on the unit sphere* $\mathbb{S}^2$ *of densities* $f_{\mathcal{VMFM}}(\underline{x}) = \sum_{i=1}^{3} 1/3 \cdot f_{\mathcal{VMF}}(\underline{x}; \underline{\alpha}_i, \kappa_i)$*, with means* $(\underline{\alpha}_1, \underline{\alpha}_2, \underline{\alpha}_3) = (1/\sqrt{5}\cdot[\,0,2,1\,]^\top, [\,0,1,0\,]^\top, 1/\sqrt{5}\cdot [\,0,2,-1\,]^\top)$ *and concentrations* $(\lambda_1, \lambda_2, \lambda_3) \in \{(8,5,2), (20,15,12), (24,24,24)\}$*. Instead of drawing random samples for obtaining the source Dirac mixture, the HEP algorithm introduced in Sec. 5.1.1 is employed to generate* $\widehat{n} = 5000$ *grid points for placing the Dirac components, with weights being the normalized PDF values at the grid points. The target discrete models are configured with* $n \in \{7, 50, 80\}$ *components.*

As shown in Fig. 6.3, the resulting target Dirac mixtures with equally weighted components (red dots) approximate the underlying distributions in a dispersion-aware manner. Compared with the grid-based Dirac mixtures (green dots), the HDMR-based discrete models exhibit evidently better representation efficiency.

Example 11: *The MDMR paradigm is further applied to the unit hypersphere* $\mathbb{S}^3$ *in accordance with the SO(3) representation of unit quaternions. Taking the definition in* (3.6)*, we synthesize a random unit quaternion variable* $\underline{\boldsymbol{x}} = [\,\cos(\boldsymbol{\theta}/2), \sin(\boldsymbol{\theta}/2)\,\underline{\boldsymbol{u}}^\top]^\top \in \mathbb{S}^3$ *with the rotation angle following a von Mises distribution*[2] $\mathcal{VM}(\pi/10, 50)$ *and the axis a von Mises–Fisher mixture given by* $f_{\mathcal{VMFM}}(\underline{u}) = \sum_{i=1}^{3} 1/3 \cdot f_{\mathcal{VMF}}(\underline{u}; \underline{\alpha}_i, 200)$*, with means* $(\underline{\alpha}_1, \underline{\alpha}_2, \underline{\alpha}_3) = (1/\sqrt{2}\cdot[\,0,1,1\,]^\top, [\,0,1,0\,]^\top, 1/\sqrt{2}\cdot[\,0,1,-1\,]^\top)$*.* $\widehat{n} = 20000$ *random samples are drawn from the synthesized underlying distribution to provide a source Dirac mixture, based on which HDMRs to target Dirac mixtures of* $n \in \{3, 7, 15\}$ *components are performed.*

As depicted in Fig. 6.4, we visualize unit quaternion quantities by rotating an initial point at $[\,1,0,0\,]^\top$ *according to* (3.7)*. The proposed HDMR method also shows evident efficacy on the unit hypersphere* $\mathbb{S}^3$*, yielding target Dirac mixtures of layouts adaptive to the irregular shape of the underlying uncertainty.*

[2] Similar to the von Mises–Fisher distribution in Sec. 2.2.1, the first parameter of a von Mises distribution is the mean, and its second parameter denotes the concentration.

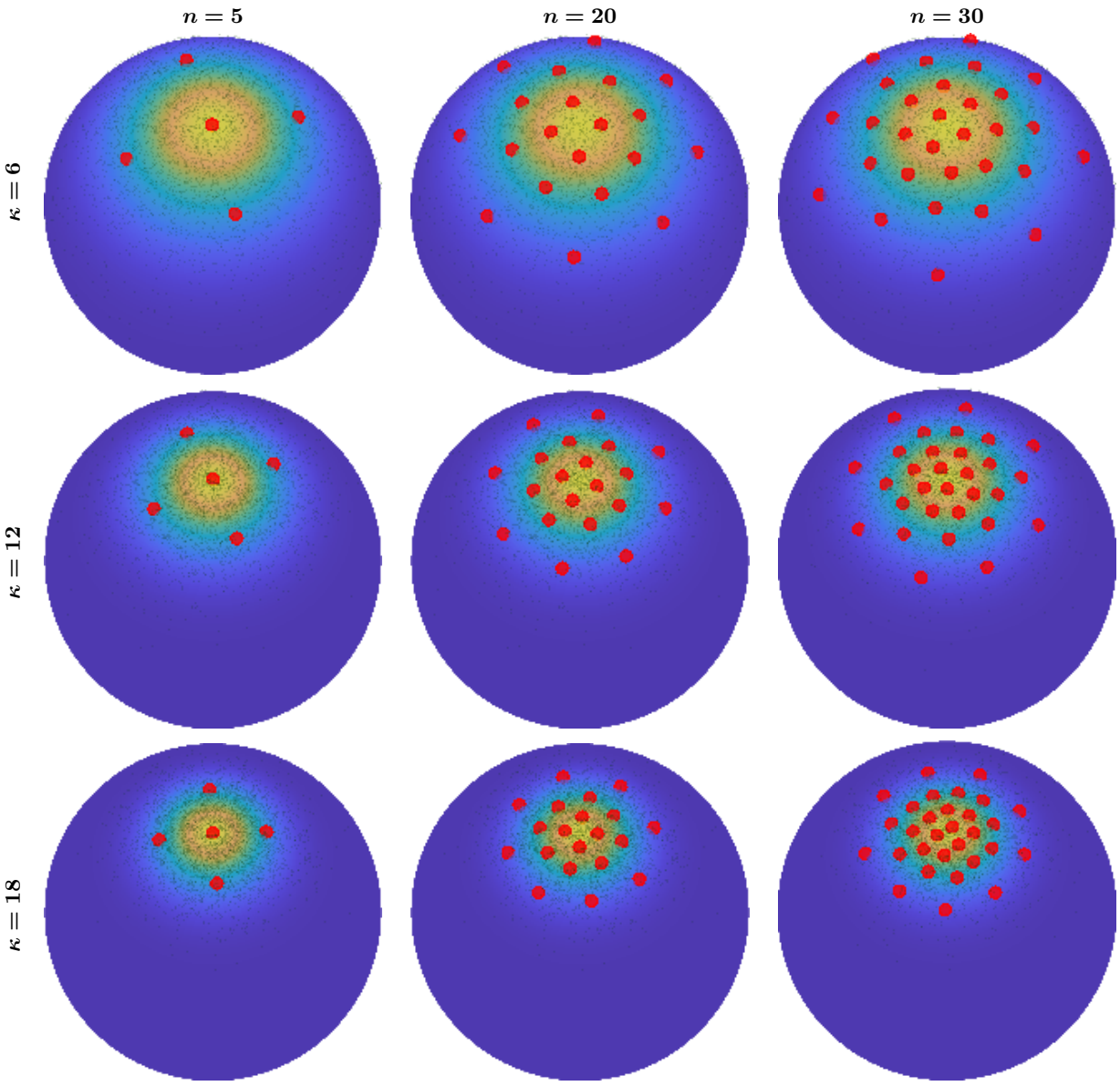


Figure 6.1.: Reapproximations (depicted by red dots) of von Mises–Fisher distributions on $\mathbb{S}^2$ using source Dirac mixtures at random samples (small green dots) in Example 9.1.

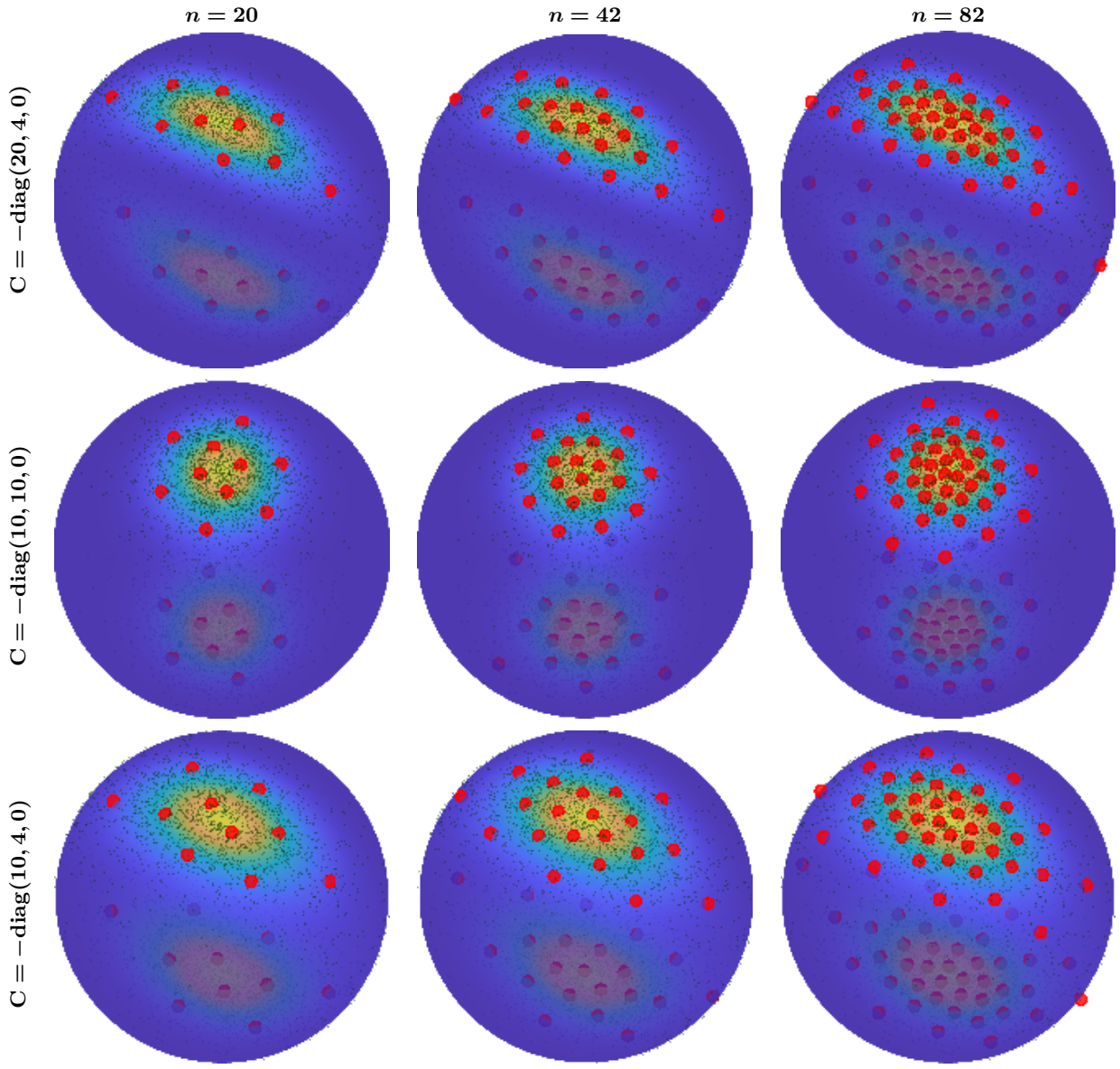


Figure 6.2.: Reapproximations (depicted by red dots) of Bingham distributions on $\mathbb{S}^2$ using source Dirac mixtures at random samples (small green dots) in Example 9.2.

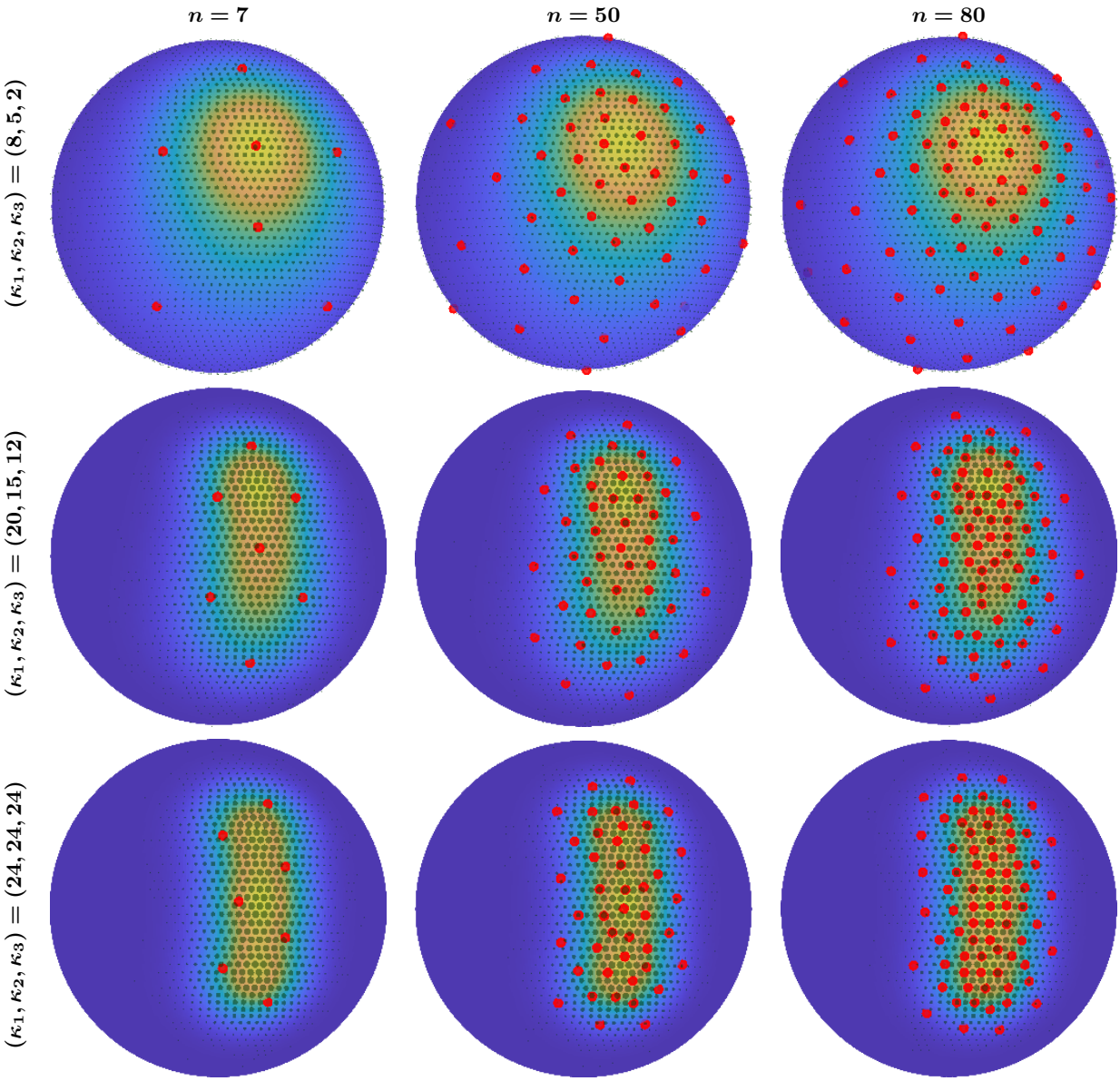


Figure 6.3.: Reapproximations (depicted by red dots) of von Mises–Fisher mixtures on $\mathbb{S}^2$ using source Dirac mixtures on HEP grids in Example 10. Grid points are plotted in green with sizes proportional to the underlying densities (i.e., weights of source Dirac components).

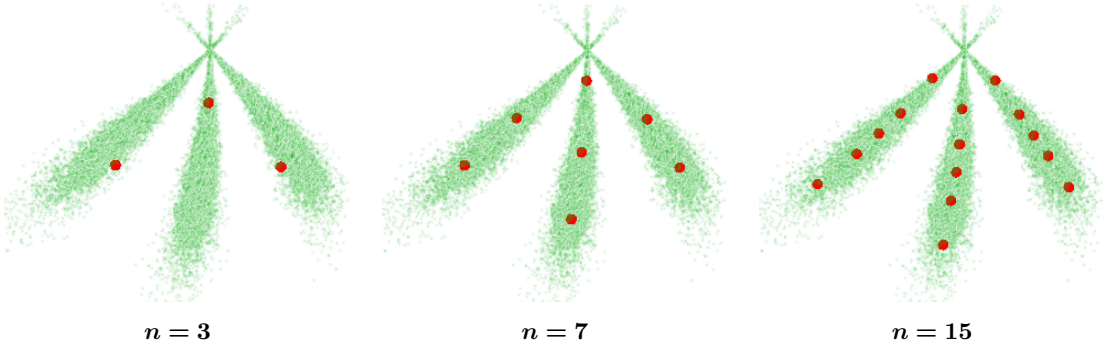


Figure 6.4.: Reapproximations (depicted by red dots) of a synthesized distribution on $\mathbb{S}^3$ using source Dirac mixtures at random samples (small green dots) in Example 11.

6.4 MDMR-Variants for Other Directional Domains

For further demonstrating the applicability of the proposed MDMR paradigm for non-spherical directional domains, we present its customization to the unit circle and the manifolds of planar/unit dual quaternions. For the sake of conciseness, in this section, we only provide necessary adaptations up to the stage of producing functional MDMR units in (6.8) and its corresponding gradient w.r.t. single point location (similar to (C.5)). Complete MCvMD-related terms regarding (6.10) (including gradients and Hessians) can be derived with reference to the generic expressions in Sec. C.1.1 and the customization procedure detailed for hyperspheres in Sec. 6.3.

6.4.1 Unit Circle

The unit circle resembles the geometric structure of the unit hyperspheres presented in Sec. 6.3. We thus set up a similar kernel function based on the von Mises distribution [12] in the form

$$k(\underline{x};\underline{\alpha},\tau) = \exp(\tau\,\underline{\alpha}^\top\underline{x})\,,\ \forall \underline{x}\in\mathbb{S}^1\,,$$

with $\underline{\alpha}\in\mathbb{S}^1$ being the location and $\tau>0$ the concentration of the kernel. Similar to the hyperspherical case, the integral over kernel locations in (6.9) then follows the normalization constant of the von Mises distribution $\mathcal{VM}\big(\widehat{\underline{u}+\underline{v}},\tau(2+2\,\underline{u}^\top\underline{v})^{1/2}\big)$, leading to

$$\boldsymbol{\mathcal{P}}(\underline{u},\underline{v},\tau) = 2\pi\boldsymbol{\mathcal{I}}_0\big(\tau\,(2+2\,\underline{u}^\top\underline{v})^{1/2}\big) =: 2\pi\boldsymbol{\mathcal{I}}_0\big(\tau\,(2+2\,\delta(\underline{u},\underline{v}))^{1/2}\big)\,, \tag{6.18}$$

with $\delta(\underline{u},\underline{v})=\underline{u}^\top\underline{v}$ being the distance metric in accordance with the circular geometry. $\boldsymbol{\mathcal{I}}_0$ denotes the modified Bessel function of the first kind and zeroth order. By specifying the weighting function $\boldsymbol{\mathcal{W}}(\tau)$ in (6.8) as

$$\boldsymbol{\mathcal{W}}(\tau)=\exp(-\epsilon\tau)$$

and following a procedure similar to (6.13) given the Laplace transform formula from [105, 17.13.109], we obtain the following closed-form MCvMD unit customized to the unit circle

$$\boldsymbol{\mathcal{Q}}(\delta)=\int_{\mathbb{R}^+}\boldsymbol{\mathcal{W}}(\tau)\boldsymbol{\mathcal{P}}(\underline{u},\underline{v},\tau)\,\mathrm{d}\tau = 2\pi\big(\epsilon^2-2(1+\delta)\big)^{-1/2} \tag{6.19}$$

under the condition $\epsilon>2$ (valid since $\delta\in[-1,1]$). We provide a brief derivation of the gradient and the Hessian of the circular CvMD unit w.r.t. sample location in Appendix C.1.3. The following example demonstrates the performance of the MDMR paradigm customized for the circular domain.

Example 12: *We synthesize a von Mises mixture on the unit circle of density* $f_{\mathcal{VMM}}(\theta)=\sum_{i=1}^{3}1/3\cdot f_{\mathcal{VM}}(\theta;\alpha_i,\kappa_i)$*, with means*[3] *at* $(\alpha_1,\alpha_2,\alpha_3)=(-\pi/2,\pi,0)$ *and concentrations* $(\kappa_1,\kappa_2,\kappa_3)=(20,50,20)$*. We use* $\widehat{n}=10000$ *equidistantly-spaced grid points to place the source Dirac components, which are then weighted by the underlying density function (with normalization) as also shown in Example 10. We deploy* $n\in\{15,25,42\}$ *target Dirac components to reapproximations.*

As shown in Fig. 6.5, the circular variant of the MDMR paradigm functions well, yielding Dirac mixtures of fewer components (red dots) adaptively to the shape of the continuous densities (blue curves) on $\mathbb{S}^1$*. Thanks to the von Mises-based kernel function, the periodicity in circular data is automatically addressed by the reapproximation method.*

[3] For brevity, here we denote the mean of each von Mises component as angles instead of vectors on $\mathbb{S}^1$.

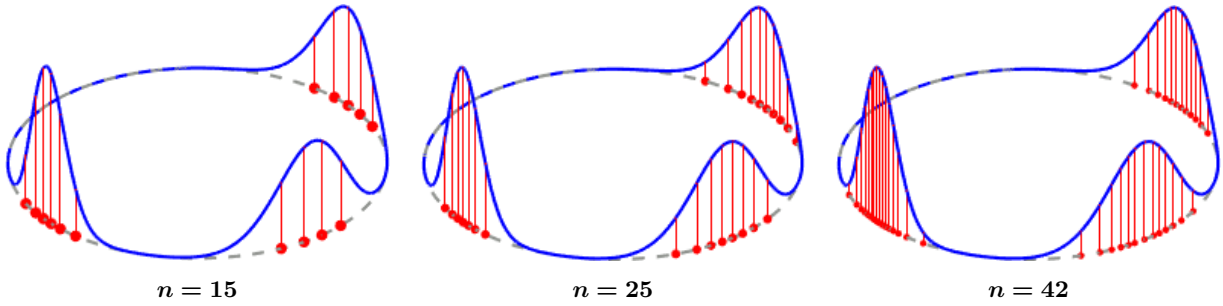


Figure 6.5.: Reapproximations (depicted by red dots) of a synthesized circular distribution (blue curves) using source Dirac mixtures at uniformly-spaced grid points in Example 12.

6.4.2 Manifolds of Dual Quaternions

As for composite directional domains, we customize the MDMR paradigm to the manifolds of planar dual quaternions and unit dual quaternions that are introduced in Sec. 3.3.3 and Sec. 3.3.1, respectively. As the two variants resemble each other to a large degree, we only demonstrate the customization procedure for the planar dual quaternion manifold [52]. However, we show examples for both variants of on-manifold Dirac mixture reapproximation.

Defining the kernel function in the MDMR paradigm for composite directional domains requires consideration on not only the topological structure of each individual component domain but also the probabilistic correlation between them. Meanwhile, a separable kernel function is always preferable for convenience of derivation and computation. According to the manifold structure of planar dual quaternions $\underline{x} = [\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top\,]^\top \in \mathbb{S}^1 \times \mathbb{R}^2 \subset \mathbb{R}^4$, we tailor the kernel function to be separable and isotropic [52] as $k(\underline{x}; \underline{\alpha}, \tau) = k_\mathrm{r}(\underline{x}_\mathrm{r}; \underline{\alpha}_\mathrm{r}, \tau)\, k_\mathrm{s}(\underline{x}_\mathrm{s}; \underline{\alpha}_\mathrm{s}, \tau)$, with

$$\begin{aligned} k_\mathrm{r}(\underline{x}_\mathrm{r}; \underline{\alpha}_\mathrm{r}, \tau) &= \exp(\tau\, \underline{\alpha}_\mathrm{r}^\top \underline{x}_\mathrm{r}) \quad \text{and} \\ k_\mathrm{s}(\underline{x}_\mathrm{s}; \underline{\alpha}_\mathrm{s}, \tau) &= \exp\left(-\tau\, (\underline{x}_\mathrm{s} - \underline{\alpha}_\mathrm{s})^\top (\underline{x}_\mathrm{s} - \underline{\alpha}_\mathrm{s}) \right), \end{aligned}$$

being the component kernels for the real and dual parts, respectively. The two component kernels share one concentration parameter $\tau > 0$ for characterizing the probabilistic correlation between $\mathbb{S}^1$ and $\mathbb{R}^2$. Consequently, the integral over kernel locations in (6.9) can be decomposed into $\mathcal{P}(\underline{u}, \underline{v}, \tau) = \mathcal{P}_\mathrm{r}(\underline{u}_\mathrm{r}, \underline{v}_\mathrm{r}, \tau)\, \mathcal{P}_\mathrm{s}(\underline{u}_\mathrm{s}, \underline{v}_\mathrm{s}, \tau)$. Here, the integral over real part kernel locations $\underline{\alpha}_\mathrm{r} \in \mathbb{S}^1$ is identical to the circular case given by (6.18), namely,

$$\mathcal{P}_\mathrm{r}(\underline{u}_\mathrm{r}, \underline{v}_\mathrm{r}, \tau) = \int_{\mathbb{S}^1} k_\mathrm{r}(\underline{u}_\mathrm{r}; \underline{\alpha}_\mathrm{r}, \tau)\, k_\mathrm{r}(\underline{v}_\mathrm{r}; \underline{\alpha}_\mathrm{r}, \tau)\, \mathrm{d}\underline{\alpha}_\mathrm{r} = 2\pi \mathcal{I}_0\big(\tau(2 + 2\, \underline{u}_\mathrm{r}^\top \underline{v}_\mathrm{r})^{1/2}\big)\,.$$

The integral $\mathcal{P}_\mathrm{s}(\underline{u}_\mathrm{s}, \underline{v}_\mathrm{s}, \tau)$ over dual part kernel locations $\underline{\alpha}_\mathrm{s} \in \mathbb{R}^2$ can be obtained based on the Gaussian identity given by [107, Appendix D] as follows [52]

$$\begin{aligned} \mathcal{P}_\mathrm{s}(\underline{u}_\mathrm{s}, \underline{v}_\mathrm{s}, \tau) &= \int_{\mathbb{R}^2} k_\mathrm{s}(\underline{u}_\mathrm{s}; \underline{\alpha}_\mathrm{s}, \tau)\, k_\mathrm{s}(\underline{v}_\mathrm{s}; \underline{\alpha}_\mathrm{s}, \tau)\, \mathrm{d}\underline{\alpha}_\mathrm{s} \\ &= \left(\frac{\pi}{\tau}\right)^2 \int_{\mathbb{R}^2} f_\mathcal{N}\Big(\underline{u}_\mathrm{s}; \underline{\alpha}_\mathrm{s}, \frac{1}{2\tau}\mathbf{I}_{2\times 2}\Big)\, f_\mathcal{N}\Big(\underline{v}_\mathrm{s}; \underline{\alpha}_\mathrm{s}, \frac{1}{2\tau}\mathbf{I}_{2\times 2}\Big)\, \mathrm{d}\underline{\alpha}_\mathrm{s} \\ &= \left(\frac{\pi}{\tau}\right)^2 f_\mathcal{N}\Big(\underline{v}_\mathrm{s}; \underline{u}_\mathrm{s}, \frac{1}{\tau}\mathbf{I}_{2\times 2}\Big) \int_{\mathbb{R}^2} f_\mathcal{N}(\underline{\alpha}_\mathrm{s})\, \mathrm{d}\underline{\alpha}_\mathrm{s} \\ &= \frac{\pi}{2\tau} \exp\left(-\frac{\tau}{2} (\underline{u}_\mathrm{s} - \underline{v}_\mathrm{s})^\top (\underline{u}_\mathrm{s} - \underline{v}_\mathrm{s}) \right). \end{aligned}$$

Consequently, we have

$$\begin{aligned}\boldsymbol{\mathcal{P}}(\underline{u},\underline{v},\tau) &= \frac{\pi^2}{\tau}\boldsymbol{\mathcal{I}}_0\big(\tau(2+2\,\underline{u}_\mathrm{r}^\top\underline{v}_\mathrm{r})^{1/2}\big)\exp\big(-\frac{\tau}{2}(\underline{u}_\mathrm{s}-\underline{v}_\mathrm{s})^\top(\underline{u}_\mathrm{s}-\underline{v}_\mathrm{s})\big)\\ &=:\frac{\pi^2}{\tau}\boldsymbol{\mathcal{I}}_0\big(\tau(2+2\,\delta_\mathrm{r})^{1/2}\big)\exp\big(-\frac{\tau}{2}\delta_\mathrm{s}\big)\,,\end{aligned}$$

with $\delta_\mathrm{r}=\underline{u}_\mathrm{r}^\top\underline{v}_\mathrm{r}$ and $\delta_\mathrm{s}=\|\underline{u}_\mathrm{s}-\underline{v}_\mathrm{s}\|^2$ indicating distance metrics on the component domains of the real and dual parts, respectively. We further design the weighting function as

$$\boldsymbol{\mathcal{W}}=\tau\exp(-\epsilon\tau)\,,$$

and the MCvMD unit in (6.8) is then specified as

$$\begin{aligned}\boldsymbol{\mathcal{Q}}(\delta_\mathrm{r},\delta_\mathrm{s}) &= \int_{\mathbb{R}^+}\boldsymbol{\mathcal{W}}(\tau)\boldsymbol{\mathcal{P}}(\delta,\tau)\,\mathrm{d}\tau\\ &= \int_{\mathbb{R}^+}\pi^2\exp\big(-(\epsilon+\delta_\mathrm{s}/2)\tau\big)\,\boldsymbol{\mathcal{I}}_0\big(\tau(2+2\delta_\mathrm{r})^{1/2}\big)\,\mathrm{d}\tau\,.\end{aligned}\tag{6.20}$$

According to [105, Sec. 6.611.4], the expression

$$\int_{\mathbb{R}^+}\exp(-a\tau)\,\boldsymbol{\mathcal{I}}_0(b\tau)\,\mathrm{d}\tau=1/\sqrt{a^2-b^2}$$

holds under the conditions $a,b>0$ and $a>b$. Since $\epsilon>2$ and $\delta_\mathrm{r}\in[-1,1]$, this prerequisite is fulfilled for (6.20), leading to the closed-form expression

$$\boldsymbol{\mathcal{Q}}(\delta_\mathrm{r},\delta_\mathrm{s})=\frac{2\pi^2}{\big(\delta_\mathrm{s}^2+4\,\epsilon\,\delta_\mathrm{s}-8\,\delta_\mathrm{r}+4\,\epsilon^2-8\big)^{1/2}}\,.\tag{6.21}$$

We provide the gradient of the customized MCvMD unit above in Appendix C.1.4. For avoiding redundant derivations, the procedure of computing its Hessians is not provided (but included in the implementation). Following a similar procedure, the MDMR can also be customized to the unit dual quaternion manifold $\mathbb{DH}_1$. The upcoming examples further demonstrate the efficacy of these two variants in the context of dual quaternion-based discrete modeling of uncertain rigid body motions.

Example 13: *We perform Dirac mixture reapproximation to uncertain planar dual quaternions representing SE(2) states in the following synthetic scenarios.*

1. *An SE(2)-Bingham distribution $\mathcal{PB}(\mathbf{C})$ is parameterized according to* (2.12) *with submatrices $\mathbf{C}_1=-101\cdot\mathbf{I}_{2\times2}$, $\mathbf{C}_2=10\cdot\mathbf{I}_{2\times2}$ and $\mathbf{C}_3=-\mathbf{I}_{2\times2}$. $\widehat{n}=5000$ random samples $\{\widehat{\underline{x}}_i\}_{i=1}^{\widehat{n}}$ are drawn from the continuous distribution for a single-step propagation via $\widehat{\underline{x}}_i\boxtimes\widehat{\underline{x}}_i$ according to the dual quaternion product in* (3.20)*. Afterward, the propagated samples place the source Dirac components of equal weights on $\mathbb{S}^1\times\mathbb{R}^2$ for reapproximations to target Dirac mixtures of $n\in\{20,42,60\}$ components.*

2. *We set up von Mises distributions in the form $\mathcal{VM}(\alpha,50)$ with means $\alpha\in\{\pi/8,\pi/4,3\pi/8\}$ on the unit circle. For each setting, we discretize the circular range $[\,0,\pi\,]$ with $\widehat{n}=2000$ equidistantly-spaced grid points. For each grid point, we draw one random sample from a Gaussian distribution $\mathcal{N}([0.5,0.5]^\top,0.001\cdot\mathbf{I}_{2\times2})$ for incorporating the uncertain planar translation. Based thereon, planar dual quaternion samples $\{\widehat{\underline{x}}_i\}_{i=1}^{\widehat{n}}$ are obtained via* (3.19) *and weighted by the von Mises density of the angular component. A two-step propagation (i.e., $\widehat{\underline{x}}_i\boxtimes\widehat{\underline{x}}_i\boxtimes\widehat{\underline{x}}_i$) is performed to place the source Dirac components with corresponding weights, based on which reapproximations to target Dirac mixtures of $n=6$ components are performed.*

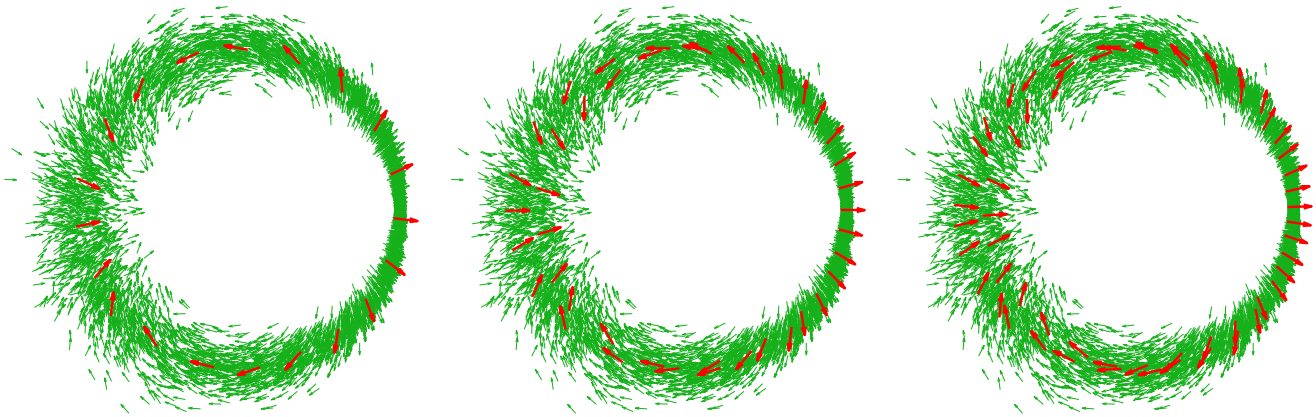


Figure 6.6.: Reapproximations (depicted by red arrows) on the planar dual quaternion manifold using source Dirac mixtures at random samples (small green arrows) in Example 13.1.

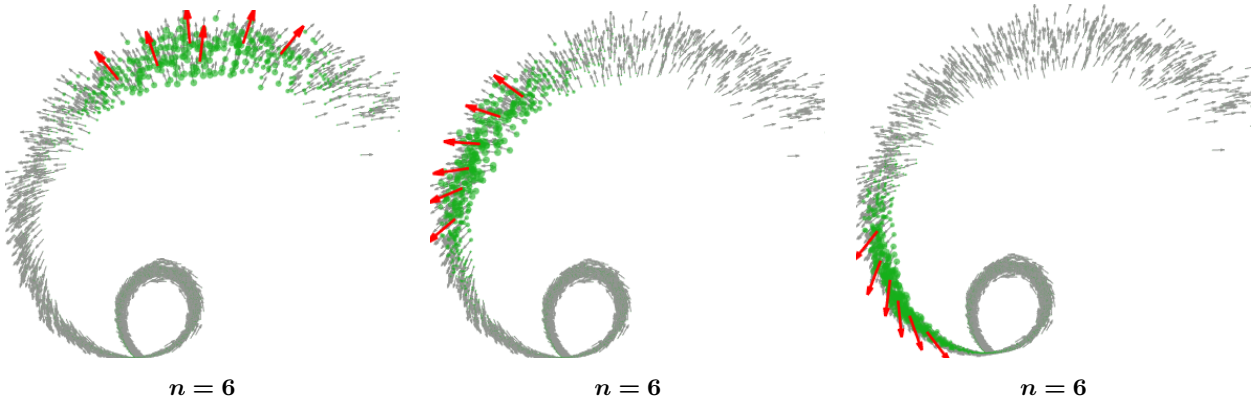
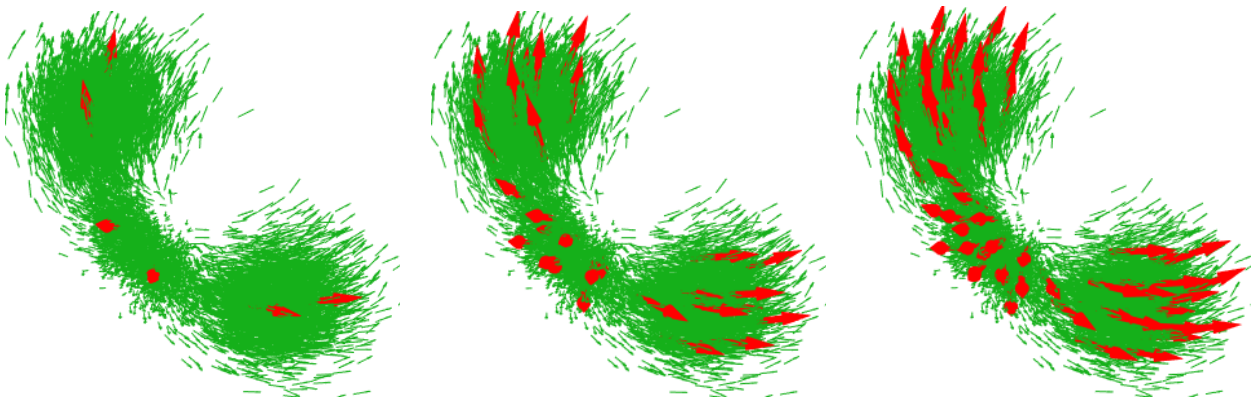


Figure 6.7.: Reapproximations (depicted by red arrows) on the planar dual quaternion manifold in Example 13.2. From left to right, the same source Dirac components are given different weights (proportional to sizes of the green dots, most obtain weights close to zero) for reapproximations using the same number ($n = 6$) of target Dirac components.

For visualization of planar dual quaternion quantities, we apply their SE(2) transformations to a normal vector pointing towards $[1, 0]^\top$ *at the origin according to the formula in* (3.21)*. As depicted in Fig. 6.6, the proposed MDMR scheme customized to* $\mathbb{S}^1 \times \mathbb{R}^2$ *shows viable reapproximations for the first scenario. Symmetries in the underlying dispersion characterized by the source (indicated by green arrows) are also well-preserved in layouts of the target Dirac mixtures (red arrows). Fig. 6.7 further demonstrates reapproximation results for the second scenario. Weights of the source Dirac components are indicated by the sizes of the green dots, and the setting above essentially creates a counterclockwise shifting pattern in the figures (from left to right). Despite the complex dispersion underlying on the planar dual quaternion manifold, the customized MDMR method functions effectively, with the probabilistic correlations between the component domains well-interpreted.*

Example 14: *We further briefly showcase the MDMR-variant on the unit dual quaternion manifold* $\mathbb{DH}_1 \subset \mathbb{R}^8$ *introduced in Sec. 3.3. Based on the SE(3) representation in* (3.12)*, we assume the uncertain angular component to follow a von Mises mixture distribution of density* $f_{\mathcal{VMM}}(\theta) = \sum_{i=1}^{3} 1/3 \cdot f_{\mathcal{VM}}(\theta; \alpha_i, 25)$*, with means* $(\alpha_1, \alpha_2, \alpha_3) = (-\pi/3, \pi/3, 0)$*. Further, the uncertain rotation axis follows a von Mises–Fisher distribution* $\mathcal{VMF}(1/\sqrt{3} \cdot [1, 1, 1]^\top, 300)$*, and the translation*



$n = 6$ $n = 30$ $n = 60$

Figure 6.8.: Reapproximations (depicted by red arrows) on the unit dual quaternion manifold $\mathbb{DH}_1$ using source Dirac mixtures at random samples (small green arrows) in Example 14.

follows a Gaussian distribution $\mathcal{N}([\,10,5,5\,])^\top, 0.005 \cdot \mathbf{I}_{3\times 3})$. $\widehat{n} = 10000$ *equally-weighted samples* $\{\widehat{\underline{x}}_i\}_{i=1}^{\widehat{n}} \subset \mathbb{DH}_1$ *are synthesized based on random sampling from the continuous distributions, after which a single-step propagation is additionally performed via* $\widehat{\underline{x}}_i \boxtimes \widehat{\underline{x}}_i$ *according to* (3.11) *to provide the source Dirac mixture. We perform reapproximations to target Dirac mixtures of* $n \in \{6, 30, 60\}$ *components.*

As depicted in Fig. 6.8, we transform a normal vector pointing toward $[\,1,0,0\,]^\top$ *at the origin using the unit dual quaternion quantities according to* (3.13)*. Reapproximations using the MDMR-variant on* $\mathbb{DH}_1$ *functions effectively under various configurations while preserving the correlation between the real and dual parts. Compared with the source Dirac mixture produced by the Monte Carlo-based scheme (green arrows), the proposed MDMR paradigm delivers more efficient discrete models for representing complex dispersion in high-dimensional composite directional domains.*

6.5 Hyperspherical Reapproximation and Reconstruction

Modeling unknown uncertainties in continuous form is still desired in certain scenarios, e.g., for the purpose of sampling, quantifying probability densities at newly observed data, or applying grid-based filters with continuous transition densities as introduced in Sec. 5.2. Therefore, we further propose a method for continuous probabilistic modeling built upon the discrete model produced by the MDMR paradigm. For clearness of explanation, we present this reconstruction method on unit hyperspheres $\mathbb{S}^{d-1}$ $(d \geq 3)$ based on the HDMR method elaborated in Sec. 6.3, leading to the following two-stage reapproximation and reconstruction (R&R) procedure [51].

We propose to exploit a von Mises–Fisher mixture model as the continuous representation of the unknown hyperspherical distribution that is initially approximated by the source Dirac mixture. Given its reapproximation to the target Dirac mixture of the form (6.2), we place one component of the von Mises–Fisher mixture at each target sample $\underline{x}_i$ (equally weighted) in $\mathbb{X}$ with a joint concentration λ, inducing a density function in the following form

$$\begin{aligned} f_{\mathcal{VMFM}}(\underline{x}; \mathbb{X}, \lambda) &= \sum_{i=1}^{n} \frac{1}{n} f_{\mathcal{VMF}}(\underline{x}; \underline{x}_i, \lambda) \\ &= \sum_{i=1}^{n} \frac{\lambda^{d/2-1}}{n(2\pi)^{d/2}\, \mathcal{I}_{d/2-1}(\lambda)} \exp(\lambda\, \underline{x}^\top \underline{x}_i)\,. \end{aligned}$$

In order to determine the value of the concentration parameter λ, we exploit the source Dirac mixture at $\{\widehat{\underline{x}}_k\}_{k=1}^{\widehat{n}}$ of weights $\{\widehat{\nu}_k\}_{k=1}^{\widehat{n}}$ in (6.1) using maximum likelihood estimation (MLE) based on the log-likelihood-based formulation [108, 109]

$$\lambda^* = \arg\max_{\lambda \geq 0} \left\{ \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \ln \Big(\frac{\lambda^{d/2-1}}{\mathcal{I}_{d/2-1}(\lambda)} \sum_{i=1}^{n} \exp(\lambda\, \widehat{\underline{x}}_k^{\top} \underline{x}_i) \Big) \right\}. \tag{6.22}$$

Using the substitution $\mathcal{C}(\lambda) = \frac{\lambda^{d/2-1}}{\mathcal{I}_{d/2-1}(\lambda)}$, we then obtain

$$\mathcal{L}(\lambda) = \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \ln \Big(\mathcal{C}(\lambda) \sum_{i=1}^{n} \exp(\lambda\, \widehat{\underline{x}}_k^{\top} \underline{x}_i) \Big) = \ln\big(\mathcal{C}(\lambda)\big) + \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \ln \Big(\sum_{i=1}^{n} \exp(\lambda\, \widehat{\underline{x}}_k^{\top} \underline{x}_i) \Big)$$

as the objective function of the optimization for MLE. Further, we derive its first-order derivative w.r.t. λ as

$$\begin{aligned} \mathcal{L}'(\lambda) &= \frac{\mathrm{d}}{\mathrm{d}\,\lambda} \Big\{ \ln\big(\mathcal{C}(\lambda)\big) + \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \ln\big(\sum_{i=1}^{n} \exp(\lambda\, \widehat{\underline{x}}_k^{\top} \underline{x}_i) \big) \Big\} \\ &= \frac{\mathcal{C}'(\lambda)}{\mathcal{C}(\lambda)} + \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \, \frac{\sum_{i=1}^{n} \widehat{\underline{x}}_k^{\top} \underline{x}_i \exp(\lambda\, \widehat{\underline{x}}_k^{\top} \underline{x}_i)}{\sum_{i=1}^{n} \exp(\lambda\, \widehat{\underline{x}}_k^{\top} \underline{x}_i)} \,. \end{aligned} \tag{6.23}$$

Finding the zero of $\mathcal{L}'(\lambda)$ then yields the desired λ^* in the MLE problem of (6.22). For that, we utilize the recurrence relation of differentiating the modified Bessel function (first kind) $\mathcal{I}_{d/2-1}(\lambda)$ in $\mathcal{C}(\lambda)$. As shown in [7], it follows

$$\mathcal{I}'_{d/2-1}(\lambda) = \mathcal{I}_{d/2}(\lambda) + \lambda^{-1}(d/2-1)\mathcal{I}_{d/2-1}(\lambda)\,.$$

Thus, the first term in (6.23) boils down to

$$\frac{\mathcal{C}'(\lambda)}{\mathcal{C}(\lambda)} = \frac{\mathcal{I}_{d/2-1}(\lambda)}{\lambda^{d/2-1}} \cdot \frac{(d/2-1)\lambda^{d/2-2}\mathcal{I}_{d/2-1}(\lambda) - \lambda^{d/2-1}\mathcal{I}'_{d/2-1}(\lambda)}{\big(\mathcal{I}_{d/2-1}(\lambda)\big)^2} = -\frac{\mathcal{I}_{d/2}(\lambda)}{\mathcal{I}_{d/2-1}(\lambda)} = -\mathcal{A}_d(\lambda)\,,$$

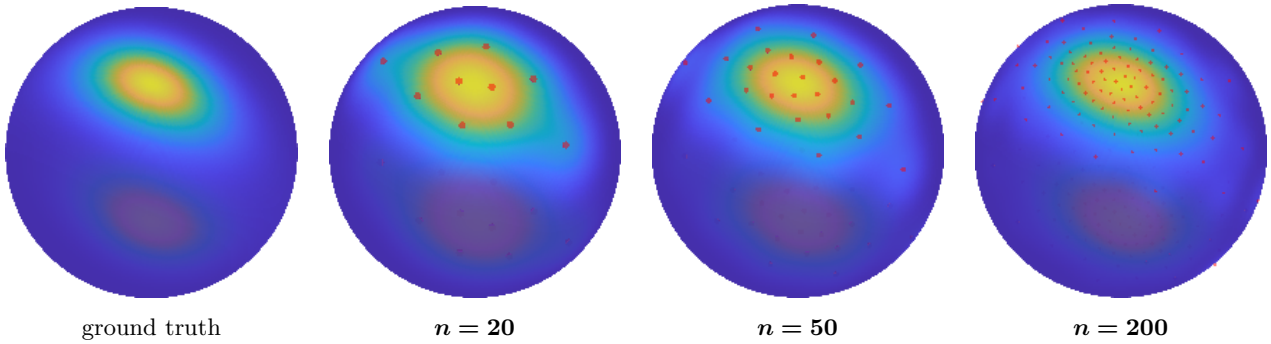
with $\mathcal{A}_d(\lambda)$ being the Bessel function ratio in the mean resultant vector of the von Mises–Fisher distribution shown in (2.7). Then, solving the MLE problem in (6.22) is equal to solving equation $\mathcal{L}'(\lambda) = 0$, with

$$\mathcal{L}'(\lambda) = \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \, \frac{\sum_{i=1}^{n} \underline{x}_i^{\top} \widehat{\underline{x}}_k \exp(\lambda\, \underline{x}_i^{\top} \widehat{\underline{x}}_k)}{\sum_{i=1}^{n} \exp(\lambda\, \underline{x}_i^{\top} \widehat{\underline{x}}_k)} - \mathcal{A}_d(\lambda)\,, \tag{6.24}$$

for which we tailor Newton's method with closed-form Newton steps in Appendix C.2. In practice, we suggest to exploit source Dirac mixtures of equally weighted components for the MLE problem formulated in (6.22) [109]. The following example showcases the performance of the proposed hyperspherical reapproximation & reconstruction scheme on the unit sphere $\mathbb{S}^2$ under various configurations.

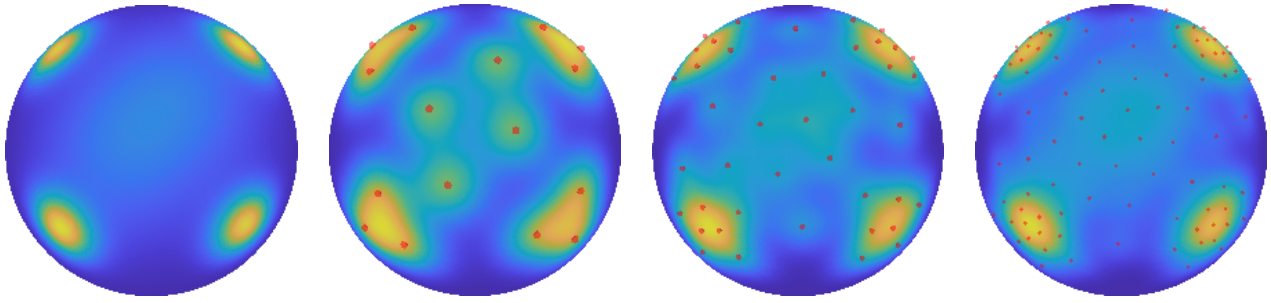
Example 15: *Following two scenarios are considered for testing the proposed R&R scheme on $\mathbb{S}^2$. For evaluating the reconstruction fidelity of a resulting von Mises–Fisher mixture, we determine its Hellinger distance $\mathcal{H}$ [13] to the underlying true distribution.*

1. *We exploit the Bingham distribution $\mathcal{B}(\mathbf{C})$ with $\mathbf{C} = -\mathrm{diag}(10, 4, 0)$ in Example 9.2 (depicted in the last row of Fig. 6.2) as the underlying distribution. We take the same number of random samples ($\widehat{n} = 20000$) for obtaining the source Dirac mixture, based on which the R&R approach is performed using $n \in \{20, 50, 200\}$ components.*



ground truth | $\boldsymbol{n = 20}$ $\kappa^* = 22.2, \mathcal{H} = 0.02$ | $\boldsymbol{n = 50}$ $\kappa^* = 31.5, \mathcal{H} = 0.016$ | $\boldsymbol{n = 200}$ $\kappa^* = 55.3, \mathcal{H} = 0.007$

(A) Continuous probabilistic modeling of a Bingham distribution on $\mathbb{S}^2$ in Example 15.1.



ground truth | $\boldsymbol{n = 16}$ $\lambda^* = 26.7, \mathcal{H} = 0.016$ | $\boldsymbol{n = 42}$ $\lambda^* = 42.3, \mathcal{H} = 0.013$ | $\boldsymbol{n = 90}$ $\lambda^* = 54.8, \mathcal{H} = 0.010$

(B) Continuous probabilistic modeling of a synthesized distribution on $\mathbb{S}^2$ in Example 15.2.

Figure 6.9.: Hyperspherical reapproximation & reconstruction performed on the unit sphere $\mathbb{S}^2$. Red dots depict the component locations of the resulting discrete and continuous models.

2. *A spherical distribution is synthesized in the form of von Mises–Fisher mixture with density* $f_{\mathcal{VMFM}}(\underline{x}) = \sum_{i=1}^{7} 1/7 \cdot f_{\mathcal{VMF}}(\underline{x}; \underline{\alpha}_i, \kappa_i)$. *We set the means of its components with* $(\underline{\alpha}_1, .., \underline{\alpha}_7) =$

$$\left([0,0,1]^\top, \frac{[0,2,1]^\top}{\sqrt{5}}, \frac{[0,-2,1]^\top}{\sqrt{5}}, \frac{[2,0,1]^\top}{\sqrt{5}}, \frac{[-2,0,1]^\top}{\sqrt{5}}, \frac{[0,1,2]^\top}{\sqrt{5}}, \frac{[0,-1,2]^\top}{\sqrt{5}} \right)$$

and their concentrations with $(\kappa_1, \dots, \kappa_7) = (5, 30, 30, 30, 30, 5, 5)$. *We perform R&R to von Mises–Fisher mixtures of* $n \in \{16, 42, 90\}$ *components based on a source Dirac mixture provided by* $\widehat{n} = 20000$ *random samples drawn from the synthesized distribution.*

The reapproximation & reconstruction results of the two scenarios above are illustrated in Fig. 6.9. Based on discrete models given by the HDMR method (component locations depicted by red dots), the followup reconstruction stage produces high-quality continuous models in the form of von Mises–Fisher mixtures compared with the ground truth. As more target components are exploited by the R&R procedure, the induced joint concentration parameter increases, and the modeling quality is improved (i.e., smaller Hellinger distance).

6.6 MDF Revisited: The On-Manifold Reapproximation Discrete Filter

As previously discussed, the proposed MDMR paradigm produces efficient discrete models in the form of Dirac mixtures. Besides the potential limitation of inefficiency in probabilistic modeling,

Algorithm 11: On-Manifold Reapproximation Discrete Filter (MRDF)

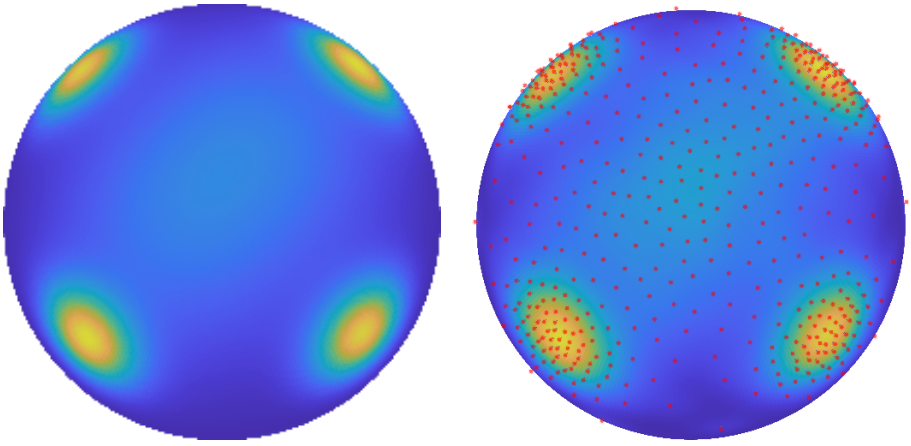
Input: posterior state set $\mathbb{X}_t^{\mathrm{e}} = \{\underline{x}_{t,i}^{\mathrm{e}}\}_{i=1}^{n}$, system noise set $\mathbb{W} = \{\underline{w}_k\}_{k=1}^{n_{\mathrm{w}}}$, measurement $\hat{\underline{z}}_{t+1}$
Output: posterior state set $\mathbb{X}_{t+1}^{\mathrm{e}} = \{\underline{x}_{t+1,i}^{\mathrm{e}}\}_{i=1}^{n}$

1 $\mathbb{X}_{t+1}^{\mathrm{p}} \leftarrow \emptyset,\ \Omega_{t+1}^{\mathrm{p}} \leftarrow \emptyset$;
/* prediction step */
2 **for** $i \leftarrow 1$ **to** n **do**
3 **for** $k \leftarrow 1$ **to** n_{w} **do**
4 $\underline{x}_{t+1,r}^{\mathrm{p}} \leftarrow a\,(\underline{x}_{t,i}^{\mathrm{e}}, \underline{w}_k)$; // see (5.3) and (5.4)
5 $\mathbb{X}_{t+1}^{\mathrm{p}} \leftarrow \{\,\mathbb{X}_{t+1}^{\mathrm{p}},\ \underline{x}_{t+1,r}^{\mathrm{p}}\,\}$;
/* update step */
6 **for** $r \leftarrow 1$ **to** $n \cdot n_{\mathrm{w}}$ **do**
7 $\nu_{t+1,r}^{\mathrm{p}} \leftarrow f_{t+1}^{\mathrm{L}}(\hat{\underline{z}}_{t+1} \,|\, \underline{x}_{t+1,r}^{\mathrm{p}})$;
8 $\Omega_{t+1}^{\mathrm{p}} \leftarrow \{\,\Omega_{t+1}^{\mathrm{p}},\ \nu_{t+1,r}^{\mathrm{p}}\,\}$;
9 $\mathbb{X}_{t+1}^{\mathrm{e}} \leftarrow \mathtt{MDMR}\,(\mathbb{X}_{t+1}^{\mathrm{p}}, \Omega_{t+1}^{\mathrm{p}}, n)$; // domain-specific
10 **return** $\mathbb{X}_{t+1}^{\mathrm{e}}$;

both grid-based and Monte Carlo-based nonparametric filtering schemes in Chapter 5 require the form of the system noise distribution for system propagation. When only raw sample sets (of large sizes) approximating the noise distribution are provided from empirical data, it is then inevitable to conduct moment matching to parametric models, inducing the risk of large approximation error in practice. In consideration of the aforementioned background, we introduce the on-manifold reapproximation discrete filter (MRDF) built upon the MDMR paradigm for nonparametric directional estimation in the absence of a system noise of given form. Following the generic description of on-manifold discrete filtering in Sec. 5.2.1, the new MRDF has a concise design as shown in Algorithm 11 [51, 52].

Suppose that the noise distribution is time-invariant and has an unknown form, for which only a large amount of empirical samples are provided. We first perform the MDMR in a variant that properly respects the noise domain for reapproximation to a Dirac mixture of n_{w} equally weighted components. Since the MDMR scheme typically produces Dirac mixtures of uniform weights, we discard the Dirac mixture expression with weighted components as originally formulated in the generic design of MDF in Sec. 5.2.1. Instead, only locations of Dirac components are taken into account, and the sample set $\mathbb{X}_t^{\mathrm{e}}$ of size n is used as the discrete representation of the posterior estimate at time step t. During the prediction step, we combine the state and noise sample sets via Cartesian product, and propagate them through the system model as derived in (5.3), inducing a prior sample set $\mathbb{X}_{t+1}^{\mathrm{p}}$ of cardinality $n \cdot n_{\mathrm{w}}$ (Algorithm 11, lines 1–5). According to (5.5), we then reweight each prior sample $\underline{x}_{t+1,r}^{\mathrm{p}}$ using its likelihood $f_{t+1}^{\mathrm{L}}(\hat{\underline{z}}_{t+1} \,|\, \underline{x}_{t+1,r}^{\mathrm{p}})$ given the measurement $\hat{\underline{z}}_{t+1}$ (Algorithm 11, lines 6–8). Afterward, the MDMR is performed to the reweighted prior samples to produce a sample set $\mathbb{X}_{t+1}^{\mathrm{e}}$ of size n for discrete modeling of the posterior density at time step $t+1$ (Algorithm 11, line 9).

In the upcoming section, we apply the proposed MRDF scheme on the unit sphere $\mathbb{S}^2$ by specifying the MDMR in line 9 of Algorithm 11 to be its hyperspherical variant given in Sec. 6.3. For evaluating the resulting hyperspherical reapproximation discrete filter (HRDF), a case study on nonlinear spherical estimation is provided as follows.



ground truth

$\boldsymbol{n = 500}$
$\lambda = 98.4, \mathcal{H} = 5.12 \times 10^{-3}$

Figure 6.10.: Hyperspherical reapproximation & reconstruction of the system noise distribution in Sec. 6.6.1.

6.6.1 Case Study: Nonlinear Spherical Estimation with Unknown System Noise

We specify the system model in the generic setting of (2.1) to be a spherical motion according to

$$\underline{\boldsymbol{x}}_{t+1} = \frac{\underline{\boldsymbol{x}}_t + \underline{\boldsymbol{w}}_t}{\| \underline{\boldsymbol{x}}_t + \underline{\boldsymbol{w}}_t \|} ,$$

with $\underline{\boldsymbol{x}}_t$, $\underline{\boldsymbol{x}}_{t+1} \in \mathbb{S}^2$ and $\underline{\boldsymbol{w}}_t \in \mathbb{S}^2$ being the spherical states and the noise term, respectively. The model essentially performs vector addition of the state and noise and then projects the sum to the unit sphere. For underlining the strength of the proposed HRDF, we set the noise distribution to be the one given in Example 15.2, which is visualized in Fig. 6.9-(B). The measurement model takes the spherical coordinates of the state vector $\underline{\boldsymbol{x}}_t = [\, \boldsymbol{x}_{t,1}, \boldsymbol{x}_{t,2}, \boldsymbol{x}_{t,3} \,]^\top$ and applies an additive noise term, namely,

$$\underline{\boldsymbol{z}}_t = \left[\arctan\left(\frac{\boldsymbol{x}_{t,2}}{\boldsymbol{x}_{t,1}}\right), \arctan\left(\frac{\boldsymbol{x}_{t,3}}{(\boldsymbol{x}_{t,1}^2 + \boldsymbol{x}_{t,2}^2)^{1/2}}\right) \right]^\top + \underline{\boldsymbol{v}}_t ,$$

with the measurement noise $\underline{\boldsymbol{v}}_t$ following a zero-mean Gaussian distribution $\mathcal{N}(\underline{0}_2, 0.01 \cdot \mathbf{I}_{2\times 2})$.

Instead of providing the symbolic form of the synthesized noise distribution, we only provide a discrete approximate of $\widehat{n} = 20000$ random samples drawn therefrom. Given the source, we then perform the HDMR in Sec. 6.3 for obtaining Dirac mixtures of $n_{\mathrm{w}} \in \{30, 50, 100, 200, 300, 500, 1000\}$ equally weighted components as discrete noise models for the proposed HRDF. Moreover, the HRDF only deploys a Dirac mixture of $n = 5$ components for discrete modeling of the state distribution. To compare it with state-of-the-art parametric hyperspherical filtering methods, we exploit the same $\widehat{n} = 20000$ random samples for von Mises–Fisher modeling of the noise distribution via moment matching. The fitted noise distribution is then deployed to deriving the transition density for a random sampling-based von Mises–Fisher filter (vMFF) [110] with sample sizes ranging from 100 to 5000. Additionally, we run a hyperspherical particle filter (HPF) as a baseline with a typical sampling-importance resampling scheme exploiting the *true* noise distribution of symbolic form to obtain the transition density (further used as the proposal). Sample sizes of the HPF range from 50 to 5000. The estimation error is quantified by the arc length in radians (thus

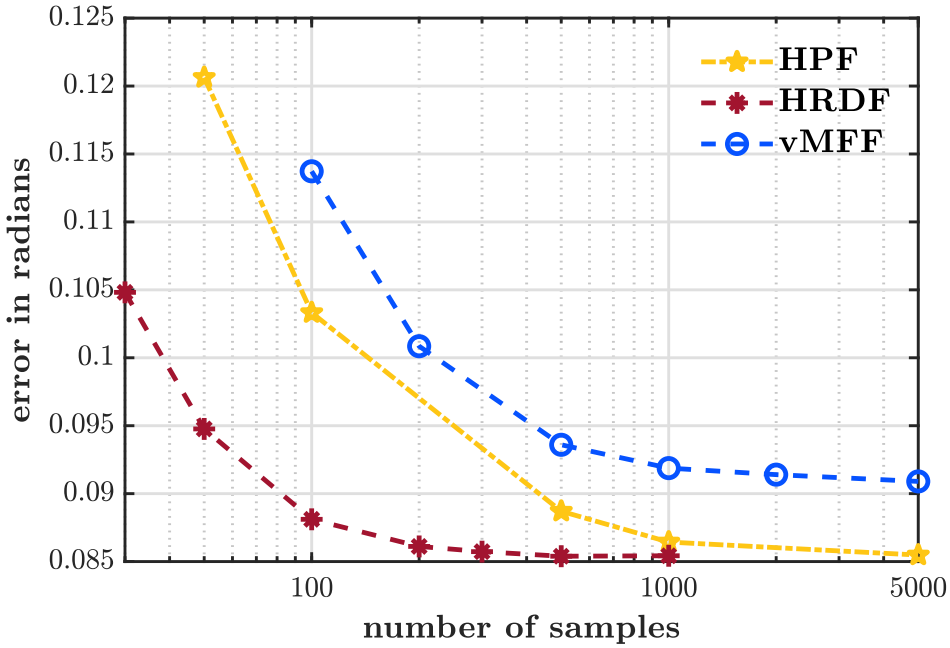


Figure 6.11.: Error over sample sizes given by the evaluated filters.

without unit) on the unit sphere according to $\arccos(\underline{x}_t^\top \hat{\underline{x}}_t)$, with $\underline{x}_t \in \mathbb{S}^2$ and $\hat{\underline{x}}_t \in \mathbb{S}^2$ being the ground truth and the estimate, respectively. Overall 5000 Monte Carlo runs are simulated with 30 consecutive time steps in each run. As in the case studies done previously in Sec. 4.4 and Sec. 5.3.4, we exploit the RMSE of the last estimate in each run to express deviations.

As shown by the error over sample sizes plotted in Fig. 6.11, the proposed HRDF delivers superior tracking accuracy over the other filters with only 5 samples exploited for discrete state modeling. With 300 components deployed to noise Dirac mixture reapproximation, the HRDF induces an error that is comparable to that of the HPF exploiting 5000 random samples and the true noise distribution in symbolic form. This clearly underlines the superior representation efficiency of the discrete models produced by the proposed HDMR method over the ones by the Monte Carlo scheme. In Fig. 6.10, we further plot the von Mises–Fisher mixture modeling of the system noise using 500 components given by the hyperspherical R&R scheme introduced in Sec. 6.5 (based on the same source samples). The proposed scheme produces a high-quality continuous model of the unknown noise distribution, inducing a very small difference compared with the ground truth (indicated by a very small Hellinger distance). Furthermore, the vMFF still exhibits a clear offset to the baseline even until convergence using a large number of samples. This coheres with our understanding about the limitation of imposing parametric settings to approximating complex noise or state distributions.

As further shown in Fig. 6.12, the proposed HRDF has a little worse runtime performance compared with the random sampling-based ones given the same small sample sizes. This results from runtime overhead, caused by, e.g., memory allocation for preparing the reapproximation, and is no longer observed when the sample size increases. More importantly, the HRDF delivers much better tracking accuracy than the others given same runtime as illustrated in Fig. 6.13. In this case, margins of superiority are quickly built up by the proposed HRDF over runtime. In fact, since the system noise is represented by a discrete model of deterministic components, no sampling or resampling is required by the HRDF as in the other Monte Carlo-based filtering approaches. The

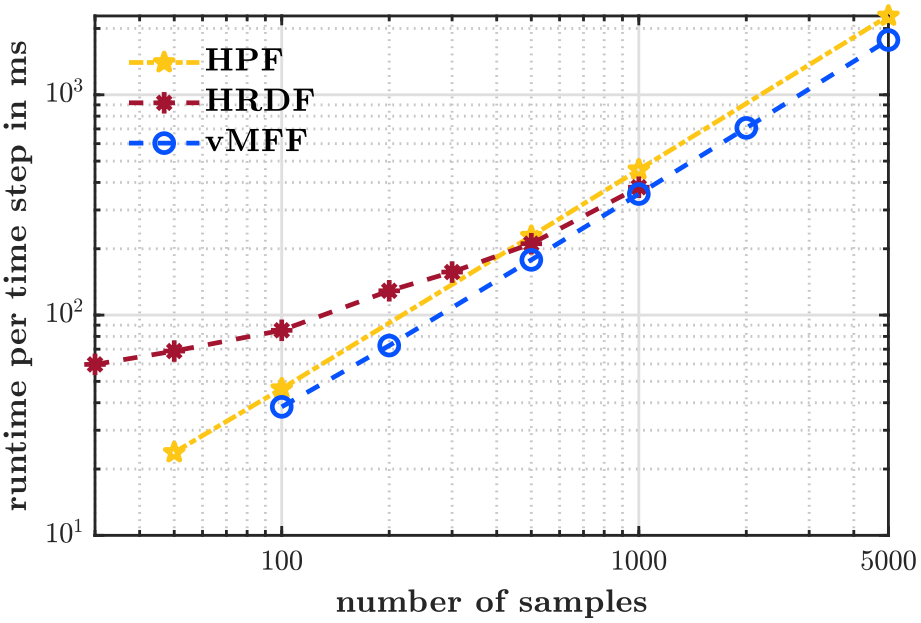


Figure 6.12.: Runtime over sample sizes given by the evaluated filters.

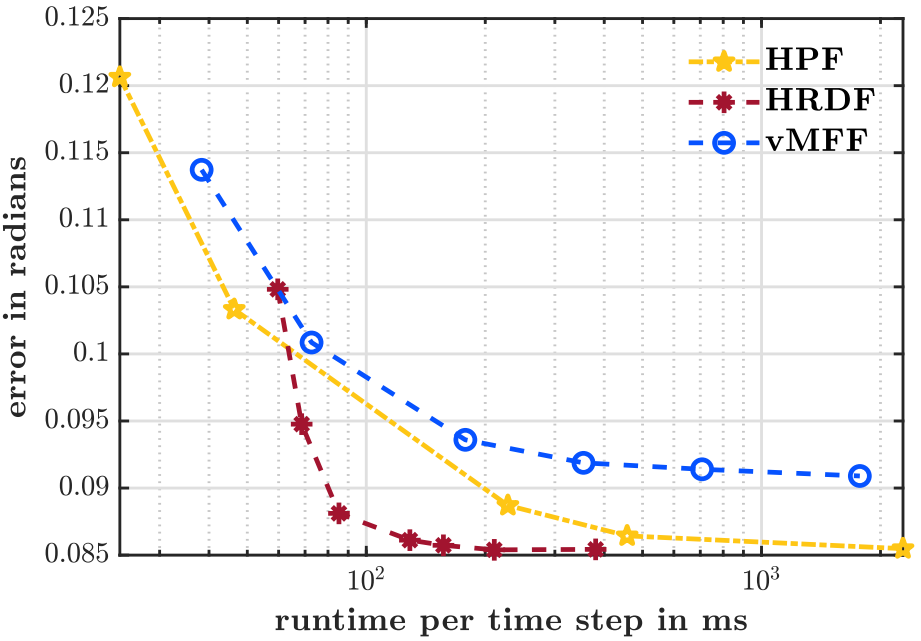


Figure 6.13.: Error over runtime given by the evaluated filters.

most computationally intensive part for the HRDF is the reapproximation for the update step as shown in line 9 of Algorithm 11. Thanks to the Riemannian trust-region method using the symbolic gradients and Hessians of the proposed HCvMD, the reapproximation step exhibits good runtime efficiency.

6.7 Short Summary

In this chapter, we build up a novel and unified paradigm for efficient discrete modeling on typical directional manifolds. Given a source Dirac mixture of equally or unequally weighted components, the proposed on-manifold reapproximation scheme produces target Dirac mixtures of configurable numbers of components with layouts adaptive to the shape of the underlying unknown dispersion in the sense of least MCvMD. Customization to a specific manifold only requires a topology-aware kernel and a weighting function that leads to a closed-form MCvMD unit. For that, we showcase several MDMR-variants for the unit circle, unit hyperspheres of arbitrary dimensions, the manifold of planar dual quaternions, and the manifold of unit dual quaternions. As a followup of the MDMR-based discrete modeling, we further show the two-stage reapproximation & reconstruction procedure on unit hyperspheres to obtain continuous probabilistic models in the form of von Mises–Fisher mixtures. By producing efficient discrete models, the MDMR scheme further contributes to establishing the on-manifold reapproximation discrete filter (nonparametric), which achieves evident superiority over parametric and Monte Carlo-based filters for nonlinear directional estimation with unknown form of system noise.

CHAPTER 7

Conclusions

In this thesis, we have established a methodological framework for systematically investigating recursive Bayesian estimation techniques for directional domains under the spirit of *on-manifold handling of on-manifold uncertainties*. From a geometric perspective, directional random variables are handled in a topology-aware fashion. This leads to theoretically sound interpretations of probabilistic modeling and filtering in comparison with the conventional linearization-based methodology built upon the assumption of local perturbation. As motivated in Chapter 1, recursive Bayesian estimation for directional domains is relevant in ubiquitous application scenarios with various and often complex problem settings. Having emerged in the past few years, however, on-manifold directional estimation techniques still stay at the stage of exploiting basic parametric directional statistical models (e.g., the von Mises, the wrapped normal, the von Mises–Fisher, or the Bingham distributions) and fundamental tools of classical recursive Bayesian estimation (e.g., the unscented transform). While these approaches provide a solid theoretical basis for on-manifold parametric filtering, there exists the urge to upgrade the rudimentary theoretical architecture in the face of requirements from challenging engineering practices.

In consideration of the existing limitations and theoretical demand in real-world applications, we first extend the classical definition of directional domains – the scope of our theoretical architecture is adjusted to topological spaces incorporating both directional manifolds $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ $(d \geq 2)$ and Euclidean spaces (typically coupled via Cartesian product) – leading to the concept of *composite* directional domains. This makes on-manifold directional modeling and filtering theories applicable to a series of important applications, in particular for uncertainty quantification and estimation of rigid body motions (e.g., based on the SE(3) or SE(2) representations of unit dual quaternions). Under the broadened scope of directional domains, we have achieved considerable advances in algorithmic innovations of topology-aware probabilistic modeling and filtering, systematically drafting out a methodological framework for on-manifold recursive Bayesian estimation. More specifically, the key contributions of this thesis are summarized as follows.

7.1 Contributions

We summarize the novel insights and methodologies presented in Chapter 3 to Chapter 6 into the following three categories. Essential geometric tools for investigating directional manifolds are first systematically introduced. Based thereon, we construct our on-manifold methodological framework with upgraded parametric and new nonparametric methods for probabilistic modeling and filtering of uncertain directional variables.

Geometric investigation on directional domains: As the basis of the bottom-up construction of the framework, we first provide a systematic investigation in Chapter 3 on related directional

domains including the unit hyperspheres and those with additional topological structures (such as antipodal symmetry or products with Euclidean spaces) represented by the manifolds of unit quaternions and unit dual quaternions. As a fundamental topological component of directional manifolds, the unit hypersphere is first studied. There, we present several important geometric tools, such as the logarithm/exponential maps, orthographic and gnomonic projection/retraction, to mathematically bridge the hyperspherical manifold to a certain tangent space located on it. The logarithm/exponential maps are geodesic-preserving and are further exploited by the isotropic deterministic sampling (IDS) for von Mises–Fisher distributions (Sec. 4.2.2) and the mode-centric partitioning (MCP) for grid-based discrete modeling of hyperspherical densities (Sec. 5.1.1). We also point out that the logarithm/exponential maps and orthographic projection/retraction have issues of warping and truncation effects, respectively. Therefore, they are not recommended for deployment to in-tangent-space interpretation of on-manifold uncertainties. Instead, the gnomonic projection/retraction technique guarantees unbounded probabilistic modeling in hyperspherical tangent spaces and can also handle dispersion of antipodal symmetry for certain topological spaces (e.g., the unit quaternion manifold). Therefore, this technique is exploited by the flexible deterministic sampling (FDS) for Bingham distributions in Sec. 4.2.1 and further as the geometric basis of establishing the on-manifold unscented particle filtering scheme in Sec. 5.3.

Moreover, we also bring novel insights to the geometric structure of unit quaternion and unit dual quaternion manifolds in the context of probabilistic modeling and filtering. One of the important findings is the SO(4) interpretation of the quaternion Hamilton product that closes the unit hypersphere $\mathbb{S}^3$ (Sec. 3.2). This further facilitates the grid transportation in the MCP-based discrete quaternion filter in Sec. 5.2.2. Probabilistic modeling in hyperspherical tangent spaces needs to be done w.r.t. local bases. For that, two choices are recommended in Sec. 3.3.2, namely, the Bingham principal basis (BPB) and the quaternion concomitant basis (QCB). The former incorporates the underlying probabilistic information geometrically into spanning the tangent space – this is exploited by the FDS on Bingham distributions in Sec. 4.2.1 and enables modeling the correlation between hyperspherical and Euclidean components in the parallel transport-based Bingham–Gaussian (PTBG) modeling scheme for composite directional domains in Sec. 4.1. The latter contributes to establishing the (augmented) tangent space for the on-manifold unscented particle filter in Sec. 5.3 customized to the manifolds of unit quaternions and unit dual quaternions.

Upgraded parametric probabilistic modeling and filtering on directional manifolds: As for parametric modeling on composite directional manifolds, we showcase a novel approach on the unit dual quaternion manifold, which couples the unit hypersphere $\mathbb{S}^3$ and the Euclidean space $\mathbb{R}^3$ via Cartesian product in an orthogonal fashion. With each individual component modeled by a Bingham distribution and a Gaussian distribution, respectively, the method models their correlation w.r.t. the BPB (given in Sec. 3.3.2) under the Hamilton product-based parallel transport (provided in Sec. 3.2.2). Based thereon, we provide several fundamental building blocks for dual quaternion filtering, such as the deterministic sampling method and the update step for identity measurement models in Sec. 4.1, with validations based on simulations.

In the past state of the art of parametric directional filtering, deterministic samples of fixed sample sizes are drawn to preserve moments of the underlying distribution only up to the second order. This basic unscented transform (UT)-based scheme often leads to deteriorated tracking performance of parametric directional filtering under strong nonlinearities due to sample degeneracy issues. Therefore, we enhance it by enabling configurable sizes of deterministic samples that approximate higher-order shape information of the underlying dispersion under the constraint of UT. Based on the geometric investigation on directional manifolds in Chapter 3, we introduce the concept

of configurable deterministic sampling (CDS) in Sec. 4.2 and practice it on the Bingham and von Mises–Fisher distributions, yielding the variants flexible deterministic sampling and isotropic deterministic sampling, respectively. For further improving the sample efficiency in the face of strong nonlinearities and non-identity measurement models, we propose the deterministic progressive update scheme in Sec. 4.3, where a measurement is fused stepwise with stride lengths adapted to the difference of the likelihoods among the deterministic samples. By combining the CDS with the progressive update scheme, performance of parametric filters has gained considerable enhancement for nonlinear directional estimation as shown in Sec. 4.4.

Systematic establishment of nonparametric directional modeling and filtering: We exploit on-manifold Dirac mixtures for nonparametric probabilistic modeling and filtering of directional random variables. Given the bounded and compact topology of unit hyperspheres, we first introduce two grid generation approaches in Sec. 5.1, the hyperspherical equal partitioning (HEP) and mode-centric partitioning (MCP) methods, for grid-based hyperspherical discrete modeling. The former enables quasi-equidistant grid points, allowing for precomputed grid reweighting scheme given the form of a transition density. The latter models the uncertainty in the vicinity of the mode with higher resolution and exhibits better representation efficiency especially when symmetry exists in the underlying topology. For nonparametric directional estimation using Dirac mixtures of deterministic supports, a generic design of the on-manifold discrete filter (MDF) is proposed in Sec. 5.2, with customization to the MCP-based grid showcased for quaternion estimation.

We further practicalize the Monte Carlo-based nonparametric filtering scheme for directional domains. On-manifold directional estimation often requires handling strong nonlinearities in high-dimensional spaces, with peaky likelihoods, heavy-tailed distributions, or non-stationary systems appearing in a joint manner. Each of these may lead to deteriorated performance of a plain particle filter due to particle degeneracy caused by exploiting the transition prior as the proposal. For mitigating this issue, we follow the basic design of the unscented particle filter (UPF) originally proposed for Euclidean spaces and introduce the on-manifold unscented particle filter (MUPF) for directional domains in Sec. 5.3. Based on the geometric investigation on directional manifolds in Chapter 3, the on-manifold UKF (MUKF) is proposed by introducing the (augmented) tangent space established on each particle via the (augmented) gnomonic projection/retraction. By fusing the recent evidence into the proposal particlewise using the MUKF, the MUPF shows evident performance superiority over the plain particle filter and parametric filters for directional estimation under critical conditions.

Last but not least, we introduce the concept of reapproximation-based Dirac mixture modeling in Sec. 6.1 and propose the on-manifold Dirac mixture reapproximation (MDMR) paradigm in Sec. 6.2 for efficient discrete representation of arbitrary directional uncertainties. Dirac mixtures of deterministic and dispersion-adaptive layouts are often preferable for probabilistic modeling over the ones based on grids (of fixed layout) or Monte Carlo schemes (with randomness in particles). Given a source Dirac mixture of many supports provided by empirical data, the MDMR paradigm can be flexibly customized w.r.t. the specific manifold and produces a target Dirac mixture of fewer (also configurable) and more representative components. We then customize the MDMR paradigm to unit hyperspheres in Sec. 6.3 and provide further variants on unit circles and dual quaternion manifolds in Sec. 6.4. Based on the customization to $\mathbb{S}^{d-1}$ ($d \geq 3$), a followup reapproximation & reconstruction procedure is introduced for continuous modeling of unknown hyperspherical distributions in the form of von Mises–Fisher mixtures. Moreover, we integrate the MDMR paradigm into the MDF to establish the on-manifold reapproximation discrete filter (MRDF) in Sec. 6.6 for nonparametric directional filtering with unknown form of system noise distribution.

Evaluations for nonlinear spherical tracking validate the superior performance of the proposed MRDF over parametric and Monte Carlo-based filters.

7.2 Outlook

The work presented in this thesis can be exploited as a basis to flesh out an even more complete theoretical framework for on-manifold directional modeling and filtering. Some results in the presented geometric investigation can be used to establish new probabilistic models. For instance, a new parametric modeling approach for uncertain unit dual quaternions can also be proposed based on the augmented gnomonic projection/retraction given in Sec. 5.3. As also shown in [41, 48], a hybrid scheme can be established for probabilistic modeling on composite directional manifolds – a nonparametric model based on Dirac mixtures (of either random or deterministic supports) can be deployed to the hypersphere, with a parametric model (e.g., a Gaussian distribution) modeling the linear part conditioned on each hyperspherical Dirac component via Rao–Blackwellization. Moreover, combining some of the presented techniques in this thesis may also lead to interesting research results. For example, the proposed MCvMD scheme makes it possible to obtain a hyperspherical grid of adaptive layout. This may further contribute to proposing a new grid-based hyperspherical filter (using either Dirac mixtures or piecewise constant densities) that could achieve better performance than existing grid-based filtering methods.

For paving the way towards broader applicability, some of the proposed approaches may need more extensive evaluations. For instance, the parametric modeling method introduced for composite directional manifolds in Sec. 4.1 is to be tested in common scenarios of SE(3) estimation using the proposed deterministic sampling scheme and measurement update step. Extensions to non-identity measurement models can also be made using the progressive update scheme introduced in Sec. 4.3. Furthermore, the proposed MDMR paradigm should be tested for higher-dimensional directional domains for more extensive use. Some building blocks of its hyperspherical variant, such as the scenario-dependent parameter ϵ in (6.17) and weighting function in (6.11), may need adjustments.

So far, the proposed filtering framework only considers basic directional random variables, thereby disregarding complex dynamics involving first- and second-order derivatives of on-manifold states. When it comes to certain real-world applications, e.g., quaternion or dual quaternion-based egomotion estimation, adaptations to the current filtering methods are essential. For instance, under the proposed MUPF scheme, the augmented tangent space can be further extended to incorporate linear/angular velocities and other essential motion variables for establishing a tightly-coupled state estimation scheme. In this regard, customized kinematic and dynamic models are important for successful deployment of the presented methods. For some typical applications, benchmarking of the proposed filters may also be appealing.

The thesis focuses on the methodological design of on-manifold uncertainty quantification and estimation. While its typical application scenarios lie in the field of data fusion [25, 52], with proper adaptations and extensions, the presented methodologies may also benefit broader research areas such as pattern recognition and machine learning. For instance, the presented MDMR method can be exploited for solving tasks such as on-manifold data pruning and clustering, scene segmentation, or data-driven modeling of complex system dynamics [7, 19, 111]. In these cases, the generic paradigm may require customization to other types of directional domains, and key components such as the kernel and weighting functions are to be designed accordingly.

APPENDIX
A

Derivations for Quaternions and Dual Quaternions

A.1 Definition of Unit Dual Quaternion Manifold

For the derivation below, we slightly abuse the vector notation of quaternions through a mixture with the dual unit ϵ and denote a dual quaternion as $\underline{x} = \underline{x}_\mathrm{r} + \epsilon\,\underline{x}_\mathrm{s}$. Further, we rewrite its real and dual parts into

$$\underline{x}_\mathrm{r} = [\,x_{\mathrm{r},0}, \check{\underline{x}}_\mathrm{r}^\top\,]^\top \quad \text{and} \quad \underline{x}_\mathrm{s} = [\,x_{\mathrm{s},0}, \check{\underline{x}}_\mathrm{s}^\top\,]^\top\,,$$

respectively, with $\check{\underline{x}}_\mathrm{r} = [\,x_{\mathrm{r},1}, x_{\mathrm{r},2}, x_{\mathrm{r},3}\,]^\top$ and $\check{\underline{x}}_\mathrm{s} = [\,x_{\mathrm{s},1}, x_{\mathrm{s},2}, x_{\mathrm{s},3}\,]^\top$. By applying the definition of the dual quaternion norm and considering $\epsilon^2 = 0$ [30], we have

$$\underline{x} \boxtimes \underline{x}^* = (\underline{x}_\mathrm{r} + \epsilon\,\underline{x}_\mathrm{s}) \boxtimes (\underline{x}_\mathrm{r}^* + \epsilon\,\underline{x}_\mathrm{s}^*) = \underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{r}^* + \epsilon\,(\underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{s}^* + \underline{x}_\mathrm{s} \otimes \underline{x}_\mathrm{r}^*) = \underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{r}^* + \epsilon\,(\underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{s}^* + (\underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{s}^*)^*)\,.$$

According to the definition of the Hamilton product [99], we further obtain

$$\begin{aligned}\underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{s}^* &= [\,x_{\mathrm{r},0}, \check{\underline{x}}_\mathrm{r}^\top\,]^\top \otimes [\,x_{\mathrm{s},0}, -\check{\underline{x}}_\mathrm{s}^\top\,]^\top \\ &= [\,x_{\mathrm{r},0}\,x_{\mathrm{s},0} + \check{\underline{x}}_\mathrm{r}^\top \check{\underline{x}}_\mathrm{s}\,, (x_{\mathrm{s},0}\,\check{\underline{x}}_\mathrm{r} - x_{\mathrm{r},0}\,\check{\underline{x}}_\mathrm{s} - \check{\underline{x}}_\mathrm{r} \times \check{\underline{x}}_\mathrm{s})^\top]^\top\,.\end{aligned}$$

The dual part of the product then boils down to $\underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{s}^* + \underline{x}_\mathrm{s} \otimes \underline{x}_\mathrm{r}^* = 2(x_{\mathrm{s},0}\,x_{\mathrm{r},0} + \check{\underline{x}}_\mathrm{r}^\top \check{\underline{x}}_\mathrm{s}) = 2\,\underline{x}_\mathrm{r}^\top \underline{x}_\mathrm{s}$. Thus, we have $\underline{x} \boxtimes \underline{x}^* = \underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{r}^* + 2\,\epsilon\,\underline{x}_\mathrm{r}^\top \underline{x}_\mathrm{s}$. By imposing the constraint of unit norm, i.e., $\underline{x} \boxtimes \underline{x}^* = 1$, the real part in the product needs to be one and the dual part zero. Thus, unit dual quaternions are confined to the manifold $\mathbb{DH}_1 = \{[\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top\,]^\top \,|\, \underline{x}_\mathrm{r} \in \mathbb{S}^3, \underline{x}_\mathrm{r}^\top \underline{x}_\mathrm{s} = 0\} \subset \mathbb{R}^8$.

A.2 Unit Dual Quaternions Representing SE(3) States

Considering the derivation in Appendix A.1 and the definition of $\underline{x}_\mathrm{s}$ in (3.12), the squared norm of any dual quaternion $\underline{x} = [\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top\,]^\top$ representing the SE(3) state follows

$$\begin{aligned}\underline{x} \boxtimes \underline{x}^* &= \underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{r}^* + \epsilon\,\underline{x}_\mathrm{r} \otimes ([\,0, 0.5\,\underline{t}^\top\,]^\top \otimes \underline{x}_\mathrm{r})^* + \epsilon\,[\,0, 0.5\,\underline{t}^\top\,]^\top \otimes \underline{x}_\mathrm{r} \otimes \underline{x}_\mathrm{r}^* \\ &= 1 + \epsilon\,[\,0, -0.5\,\underline{t}^\top\,]^\top + \epsilon\,[\,0, 0.5\,\underline{t}^\top\,]^\top = 1\,.\end{aligned}$$

Thus, dual quaternions representing SE(3) states are of unit norm.

A.3 Augmented Gnomonic Projection on Unit Dual Quaternion Manifold

Given the LATS at $\underline{v} = [\,\underline{v}_\mathrm{r}^\top, \underline{v}_\mathrm{s}^\top\,]^\top \in \mathbb{DH}_1$, performing the augmented gnomonic projection to the dual part of $\underline{x} = [\,\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top\,]^\top \in \mathbb{DH}_1$ essentially induces a subtracted translation term. The

subtraction follows

$$\underline{\Delta} = [\,\underline{\Delta}_{\mathrm{r}}^{\top}, \underline{\Delta}_{\mathrm{s}}^{\top}\,]^{\top} \coloneqq \underline{v}^{-1} \boxtimes \underline{x} = \begin{bmatrix} \underline{v}_{\mathrm{r}}^{-1} \otimes \underline{x}_{\mathrm{r}} \\ \underline{v}_{\mathrm{r}}^{-1} \otimes \underline{x}_{\mathrm{s}} + \underline{v}_{\mathrm{s}}^{*} \otimes \underline{x}_{\mathrm{r}} \end{bmatrix} .$$

Given the definition of unit dual quaternions representing SE(3) states in (3.12), the subtracted translation term can be derived according to

$$\begin{aligned} \left[\,0, \underline{t}_{\Delta}^{\top}\,\right]^{\top} &= 2\,(\underline{v}_{\mathrm{r}}^{-1} \otimes \underline{x}_{\mathrm{s}} + \underline{v}_{\mathrm{s}}^{*} \otimes \underline{x}_{\mathrm{r}}) \otimes (\underline{v}_{\mathrm{r}}^{-1} \otimes \underline{x}_{\mathrm{r}})^{-1} \\ &= 2\,(\underline{v}_{\mathrm{r}}^{-1} \otimes \underline{x}_{\mathrm{s}} + \underline{v}_{\mathrm{s}}^{*} \otimes \underline{x}_{\mathrm{r}}) \otimes \underline{x}_{\mathrm{r}}^{-1} \otimes \underline{v}_{\mathrm{r}} \\ &= 2\,\underline{v}_{\mathrm{r}}^{-1} \otimes \underline{x}_{\mathrm{s}} \otimes \underline{x}_{\mathrm{r}}^{-1} \otimes \underline{v}_{\mathrm{r}} + 2\,\underline{v}_{\mathrm{s}}^{*} \otimes \underline{v}_{\mathrm{r}} \\ &= \underline{v}_{\mathrm{r}}^{-1} \otimes [\,0, \underline{t}_{\mathrm{x}}^{\top}\,]^{\top} \otimes \underline{v}_{\mathrm{r}} + \underline{v}_{\mathrm{r}}^{*} \otimes [\,0, -\underline{t}_{\mathrm{v}}^{\top}]^{\top} \otimes \underline{v}_{\mathrm{r}} \\ &= \underline{v}_{\mathrm{r}}^{-1} \otimes [\,0, \underline{t}_{\mathrm{x}}^{\top} - \underline{t}_{\mathrm{v}}^{\top}\,]^{\top} \otimes \underline{v}_{\mathrm{r}} \,. \end{aligned}$$

Here, $\underline{t}_{\mathrm{v}}$ and $\underline{t}_{\mathrm{x}}$ denote the translation vectors encoded in the corresponding dual parts $\underline{v}_{\mathrm{s}}$ and $\underline{x}_{\mathrm{s}}$, respectively. Considering the rotation law of unit quaternions given in (3.7), we then obtain

$$\underline{t}_{\Delta} = \mathbf{R}_{\mathrm{v_r}}^{-1}(\underline{t}_{\mathrm{x}} - \underline{t}_{\mathrm{v}}) = \mathbf{R}_{\mathrm{v_r}}^{\top}(\underline{t}_{\mathrm{x}} - \underline{t}_{\mathrm{v}}) \,.$$

APPENDIX
B

Orbit Interval for Isotropic Deterministic Sampling

Obtaining the orbit interval λ for isotropic deterministic sampling on von Mises–Fisher distributions refers to solving (4.5), which can be done using a generic numerical solver (e.g., `solve` in Matlab). To achieve better runtime efficiency, however, we tailor Newton's Method with iterations performed in closed form. For that, we derive the first derivative of the Dirichlet kernel $\mathcal{J}_{n_1}(\lambda)$ w.r.t. λ as

$$\begin{aligned}\frac{\mathrm{d}\mathcal{J}_{n_1}(\lambda)}{\mathrm{d}\lambda} &= \frac{(n_1+0.5)\cos\big((n_1+0.5)\lambda\big)\sin(0.5\,\lambda) - 0.5\,\sin\big((n_1+0.5)\lambda\big)\cos(0.5\,\lambda)}{2\big(\sin(0.5\,\lambda)\big)^2} \\ &= 0.5\,(n_1+0.5)\cos\big((n_1+0.5)\lambda\big)\csc(0.5\,\lambda) - 0.5\,\mathcal{J}_{n_1}(\lambda)\cot(0.5\,\lambda)\,,\end{aligned}$$

with $\mathcal{J}_{n_1}(\lambda) = 0.5\sin((n_1+0.5)\lambda)/\sin(0.5\,\lambda)$. In the context of this appendix, we use i as the index of the Newton step. Then, λ can be updated from the i-th to the $(i+1)$-th step according to

$$\lambda_{i+1} = \lambda_i - \frac{\mathcal{J}_{n_1}(\lambda_i) - \big((n_1 n_2+1)\mathcal{A}_d(\kappa)-1\big)/n_2 - 0.5}{0.5\,(n_1+0.5)\cos\big((n_1+0.5)\lambda_i\big)\csc(0.5\,\lambda_i) - 0.5\,\mathcal{J}_{n_1}(\lambda_i)\cot(0.5\,\lambda_i)}\,.$$

We initialize the Newton's method via linear interpolation between 0 and the first non-negative zero of $\mathcal{J}_{n_1}(\lambda)$, i.e., $\pi/(n_1+0.5)$, w.r.t. their corresponding Dirichlet kernel values $n_1+0.5$ and 0. By substituting the right-hand side of (4.5) with $c = (n_1 n_2+1)\mathcal{A}_d(\kappa)/n_2 - 1/n_2 + 0.5$, we obtain the initial value

$$\lambda_0 = \frac{\pi\,(n_1+0.5-c)}{(n_1+0.5)^2} = \frac{\pi(n_1+1/n_2)(1-\mathcal{A}_d(\kappa))}{(n_1+0.5)^2}\,.$$

The customized Newton's method above with closed-form derivative delivers fast convergence in practice (within five steps given a typical stopping criterion). This is considerably faster than a universal numerical solver (e.g., two orders of magnitude faster than `solve` in Matlab) and delivers a good runtime performance in nonlinear von Mises–Fisher filtering as shown in [38].

APPENDIX C

Derivations for On-Manifold Dirac Mixture Reapproximations

C.1 Pointwise Derivatives of MCvMD in the Ambient Space

For the sake of conciseness, we first derive the pointwise derivatives in generic expressions in Appendix C.1.1. Afterward, customization to specific manifolds is provided in Appendix C.1.2 to Appendix C.1.4.

C.1.1 Generic Expressions

We express the MCvMD unit in (6.8) as $\boldsymbol{Q}(\underline{u},\underline{v}) = \boldsymbol{Q}(\delta(\underline{u},\underline{v}))$, with $\delta(\underline{u},\underline{v})$ denoting the distance metric in accordance with the underlying manifold geometry. The gradient of the MCvMD unit w.r.t. the point location $\underline{u}$ then follows

$$\frac{\partial \boldsymbol{Q}(\underline{u},\underline{v})}{\partial \underline{u}} = \frac{\mathrm{d}\boldsymbol{Q}(\delta)}{\mathrm{d}\delta}\frac{\partial \delta(\underline{u},\underline{v})}{\partial \underline{u}} =: \chi(\underline{u},\underline{v})\frac{\partial \delta(\underline{u},\underline{v})}{\partial \underline{u}}\,, \tag{C.1}$$

with the scalar coefficient $\chi(\underline{u},\underline{v}) = \mathrm{d}\boldsymbol{Q}(\delta)/\,\mathrm{d}\delta$ being its first derivative w.r.t. to the distance metric. The pointwise gradient of the MCvMD in (6.10) w.r.t. $\underline{x}_i$ is then expressed as

$$\frac{\partial \mathcal{D}(\mathbb{X},\widehat{\mathbb{X}})}{\partial \underline{x}_i} = \frac{\partial \mathcal{D}_1(\mathbb{X})}{\partial \underline{x}_i} - 2\frac{\partial \mathcal{D}_2(\mathbb{X},\widehat{\mathbb{X}})}{\partial \underline{x}_i}\,,$$

with derivative of the last component $\mathcal{D}_3$ in (6.10) being zero (since its value is constant given the source Dirac mixture) and derivatives of the other two components being

$$\begin{aligned}
\frac{\partial \mathcal{D}_1(\mathbb{X})}{\partial \underline{x}_i} &= 2\,\nu_i \sum_{j=1}^{n} \nu_j \frac{\partial \boldsymbol{Q}(\underline{x}_i,\underline{x}_j)}{\partial \underline{x}_i} = 2\,\nu_i \sum_{j=1}^{n} \nu_j\, \chi(\underline{x}_i,\underline{x}_j)\frac{\partial \delta(\underline{x}_i,\underline{x}_j)}{\partial \underline{x}_i} \quad \text{and} \\
\frac{\partial \mathcal{D}_2(\mathbb{X},\widehat{\mathbb{X}})}{\partial \underline{x}_i} &= \nu_i \sum_{k=1}^{\hat{n}} \widehat{\nu}_k \frac{\partial \boldsymbol{Q}(\underline{x}_i,\widehat{\underline{x}}_k)}{\partial \underline{x}_i} = \nu_i \sum_{k=1}^{\hat{n}} \widehat{\nu}_k\, \chi(\underline{x}_i,\widehat{\underline{x}}_k)\frac{\partial \delta(\underline{x}_i,\widehat{\underline{x}}_k)}{\partial \underline{x}_i}\,,
\end{aligned} \tag{C.2}$$

respectively. Further, based on the expression in (C.1), we obtain the Hessians of the MCvMD unit w.r.t. the sample locations in the form

$$\begin{aligned}
\frac{\partial^2 \boldsymbol{Q}(\underline{u},\underline{v})}{\partial \underline{u}\,\partial \underline{u}^\mathsf{T}} &= \frac{\partial \chi(\underline{u},\underline{v})}{\partial \underline{u}}\left(\frac{\partial \delta(\underline{u},\underline{v})}{\partial \underline{u}}\right)^\mathsf{T} + \chi(\underline{u},\underline{v})\frac{\partial^2 \delta(\underline{u},\underline{v})}{\partial \underline{u}\,\partial \underline{u}^\mathsf{T}}\,, \\
\frac{\partial^2 \boldsymbol{Q}(\underline{u},\underline{v})}{\partial \underline{u}\,\partial \underline{v}^\mathsf{T}} &= \frac{\partial \chi(\underline{u},\underline{v})}{\partial \underline{v}}\left(\frac{\partial \delta(\underline{u},\underline{v})}{\partial \underline{u}}\right)^\mathsf{T} + \chi(\underline{u},\underline{v})\frac{\partial^2 \delta(\underline{u},\underline{v})}{\partial \underline{u}\,\partial \underline{v}^\mathsf{T}}\,.
\end{aligned} \tag{C.3}$$

Consequently, the pointwise Hessians of the MCvMD components in (6.10) w.r.t. target Dirac locations can be expressed componentwise according to

$$\begin{aligned}
\frac{\partial^2 \mathcal{D}_1(\mathbb{X})}{\partial \underline{x}_i \partial \underline{x}_i^\top} &= 2\,\nu_i \sum_{j=1}^{n} \nu_j \frac{\partial^2 \boldsymbol{Q}(\underline{x}_i, \underline{x}_j)}{\partial \underline{x}_i \partial \underline{x}_i^\top} \\
&= 2\,\nu_i \sum_{j=1}^{n} \nu_j \left(\frac{\partial \chi(\underline{x}_i, \underline{x}_j)}{\partial \underline{x}_i} \left(\frac{\partial \delta(\underline{x}_i, \underline{x}_j)}{\partial \underline{x}_i} \right)^\top + \chi(\underline{x}_i, \underline{x}_j) \frac{\partial^2 \delta(\underline{x}_i, \underline{x}_j)}{\partial \underline{x}_i \, \partial \underline{x}_i^\top} \right), \\
\frac{\partial^2 \mathcal{D}_1(\mathbb{X})}{\partial \underline{x}_i \partial \underline{x}_j^\top} &= 2\nu_i \sum_{j=1}^{n} \nu_j \frac{\partial^2 \boldsymbol{Q}(\underline{x}_i, \underline{x}_j)}{\partial \underline{x}_i \partial \underline{x}_j^\top} \\
&= 2\,\nu_i \sum_{j=1}^{n} \nu_j \left(\frac{\partial \chi(\underline{x}_i, \underline{x}_j)}{\partial \underline{x}_j} \left(\frac{\partial \delta(\underline{x}_i, \underline{x}_j)}{\partial \underline{x}_i} \right)^\top + \chi(\underline{x}_i, \underline{x}_j) \frac{\partial^2 \delta(\underline{x}_i, \underline{x}_j)}{\partial \underline{x}_i \, \partial \underline{x}_j^\top} \right), \\
\frac{\partial^2 \mathcal{D}_2(\mathbb{X}, \widehat{\mathbb{X}})}{\partial \underline{x}_i \partial \underline{x}_i^\top} &= \nu_i \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \frac{\partial^2 \boldsymbol{Q}(\underline{x}_i, \widehat{\underline{x}}_k)}{\partial \underline{x}_i \partial \underline{x}_i^\top} \\
&= \nu_i \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \left(\frac{\partial \chi(\underline{x}_i, \widehat{\underline{x}}_k)}{\partial \underline{x}_i} \left(\frac{\partial \delta(\underline{x}_i, \widehat{\underline{x}}_k)}{\partial \underline{x}_i} \right)^\top + \chi(\underline{x}_i, \widehat{\underline{x}}_k) \frac{\partial^2 \delta(\underline{x}_i, \widehat{\underline{x}}_k)}{\partial \underline{x}_i \, \partial \underline{x}_i^\top} \right).
\end{aligned} \tag{C.4}$$

Given a specific distance metric $\delta(\underline{u}, \underline{v})$, the expressions above are applicable to non-composite directional domains including the unit circle, unit hyperspheres, or Euclidean spaces.

For composite directional domains[1], deriving the gradients can still follows the same procedure for each component domain provided that the kernel function $k(\underline{u}, \underline{v})$ is separable. Regarding Hessians, besides the derivations for each individual component domain, there also requires the partial derivatives of the gradient w.r.t. variables across the component domains, which we do not elaborate in this thesis to avoid redundant derivations. However, they are incorporated in symbolic form into the implementation of all MDMR-variants introduced in Sec. 6.3 and Sec. 6.4 to guarantee efficient convergence for reapproximation.

C.1.2 Specification on Unit Hyperspheres

For the sake of clarity, we derive the gradients and Hessians w.r.t. each target set location (namely, each column in $\mathbf{X} \in \mathbb{OB}(d, n)$) in the ambient space of $\mathbb{S}^{d-1} \subset \mathbb{R}^d$. In practice, however, their computation is implemented fully matrix-wise for runtime efficiency of the optimization.

Pointwise gradients in the ambient space of $\mathbb{S}^{d-1}$: We first rewrite the HCvMD unit in (6.14) into the form

$$\boldsymbol{Q}(\underline{u}, \underline{v}) = \frac{(2\pi)^{d/2}}{d/2 - 1} \left(\sqrt{\epsilon^2 - 2(1 + \underline{u}^\top \underline{v})} + \epsilon \right)^{1-d/2} =: \frac{(2\pi)^{d/2}}{d/2 - 1} \big(\zeta(\underline{u}, \underline{v}) + \epsilon \big)^{1-d/2},$$

with $\zeta(\underline{u}, \underline{v}) = \sqrt{\epsilon^2 - 2(1 + \delta(\underline{u}, \underline{v}))}$ substituting the square root term, where the inner product $\delta(\underline{u}, \underline{v}) = \underline{u}^\top \underline{v}$ denotes the distance metric on $\mathbb{S}^{d-1}$ in accordance with the hyperspherical geometry.

[1] E.g., the planar dual quaternion manifold or the general unit dual quaternion manifold, where the variable $\underline{x} = [\underline{x}_\mathrm{r}^\top, \underline{x}_\mathrm{s}^\top]^\top \in \mathbb{S}^1 \times \mathbb{R}^2$ or $\mathbb{DH}_1$, respectively.

Further, the first-order derivative of $\boldsymbol{Q}(\underline{u},\underline{v})$ w.r.t. the point location $\underline{u}$ is obtained according to (C.1) in the form

$$\frac{\partial \boldsymbol{Q}(\underline{u},\underline{v})}{\partial \underline{u}} = \frac{(2\pi)^{d/2}}{\zeta(\underline{u},\underline{v})\big(\zeta(\underline{u},\underline{v})+\epsilon\big)^{d/2}} \frac{\partial \delta(\underline{u},\underline{v})}{\partial \underline{u}} =: \chi(\underline{u},\underline{v})\,\underline{v}\,, \tag{C.5}$$

with the substitute of the coefficient given by

$$\chi(\underline{u},\underline{v}) = \frac{(2\pi)^{d/2}}{\zeta(\underline{u},\underline{v})\,\big(\zeta(\underline{u},\underline{v})+\epsilon\big)^{d/2}}\,. \tag{C.6}$$

Based thereon, the pointwise gradients of the HCvMD components in (6.15) can be specified according to (C.2) as

$$\frac{\partial \mathcal{D}_1(\mathbb{X})}{\partial \underline{x}_i} = 2\,\omega_i \sum_{j=1}^{n} \omega_j\,\chi(\underline{x}_i,\underline{x}_j)\,\underline{x}_j \quad \text{and} \quad \frac{\partial \mathcal{D}_2(\mathbb{X},\widehat{\mathbb{X}})}{\partial \underline{x}_i} = \omega_i \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k\,\chi(\underline{x}_i,\widehat{\underline{x}}_k)\,\widehat{\underline{x}}_k\,,$$

with $\chi(\underline{x}_i,\underline{x}_j)$ and $\chi(\underline{x}_i,\widehat{\underline{x}}_k)$ expressed according to (C.6), respectively.

Pointwise Hessians in the ambient space of $\mathbb{S}^{d-1}$: For deriving Hessians of the HCvMD unit in (6.14), we first provide the derivative of the coefficient in (C.6) w.r.t. point locations $\underline{u}$ and $\underline{v}$ in the form

$$\begin{aligned}
\frac{\partial \chi(\underline{u},\underline{v})}{\partial \underline{u}} &= \frac{(2\pi)^{d/2}\big((2+d)\,\zeta(\underline{u},\underline{v})+2\epsilon\big)}{2\,\zeta^3(\underline{u},\underline{v})(\zeta(\underline{u},\underline{v})+\epsilon)^{1+d/2}}\,\underline{v} = \frac{(2+d)\zeta(\underline{u},\underline{v})+2\epsilon}{2(\zeta(\underline{u},\underline{v})+\epsilon)\,\zeta^2(\underline{u},\underline{v})}\chi(\underline{u},\underline{v})\,\underline{v}\,,\\
\frac{\partial \chi(\underline{u},\underline{v})}{\partial \underline{v}} &= \frac{(2\pi)^{d/2}\big((2+d)\,\zeta(\underline{u},\underline{v})+2\epsilon\big)}{2\,\zeta^3(\underline{u},\underline{v})(\zeta(\underline{u},\underline{v})+\epsilon)^{1+d/2}}\,\underline{u} = \frac{(2+d)\zeta(\underline{u},\underline{v})+2\epsilon}{2(\zeta(\underline{u},\underline{v})+\epsilon)\,\zeta^2(\underline{u},\underline{v})}\chi(\underline{u},\underline{v})\,\underline{u}\,,
\end{aligned}$$

respectively. Given that $\delta(\underline{u},\underline{v}) = \underline{u}^\top\underline{v}$, we have the second derivative of $\delta(\underline{u},\underline{v})$ w.r.t. the point location $\underline{u}$ holding

$$\frac{\partial^2 \delta(\underline{u},\underline{v})}{\partial \underline{u}\,\partial \underline{u}^\top} = \mathbf{0}_{d\times d}\,.$$

Following the generic expression in (C.3), we then obtain

$$\begin{aligned}
\frac{\partial^2 \boldsymbol{Q}(\underline{u},\underline{v})}{\partial \underline{u}\,\partial \underline{u}^\top} &= \frac{\partial \chi(\underline{u},\underline{v})}{\partial \underline{u}}\underline{v}^\top = \frac{(2+d)\,\zeta(\underline{u},\underline{v})+2\epsilon}{2\,(\zeta(\underline{u},\underline{v})+\epsilon)\,\zeta^2(\underline{u},\underline{v})}\chi(\underline{u},\underline{v})\,\underline{v}\,\underline{v}^\top\,,\\
\frac{\partial^2 \boldsymbol{Q}(\underline{u},\underline{v})}{\partial \underline{u}\,\partial \underline{v}^\top} &= \frac{\partial \chi(\underline{u},\underline{v})}{\partial \underline{v}}\underline{v}^\top + \chi(\underline{u},\underline{v})\,\mathbf{I}_{d\times d} = \chi(\underline{u},\underline{v})\bigg(\frac{(2+d)\,\zeta(\underline{u},\underline{v})+2\epsilon}{2\,(\zeta(\underline{u},\underline{v})+\epsilon)\,\zeta^2(\underline{u},\underline{v})}\,\underline{u}\,\underline{v}^\top + \mathbf{I}_{d\times d}\bigg)\,.
\end{aligned}$$

The pointwise Hessians of the HCvMD components in (6.15) are then derived according to (C.4) as

$$\begin{aligned}
\frac{\partial^2 \mathcal{D}_1(\mathbb{X})}{\partial \underline{x}_i \partial \underline{x}_i^\top} &= 2\,\nu_i \sum_{j=1}^{n} \nu_j \frac{\partial^2 \boldsymbol{Q}(\underline{x}_i,\underline{x}_j)}{\partial \underline{x}_i \partial \underline{x}_i^\top} = 2\,\nu_i \sum_{j=1}^{n} \nu_j \frac{(2+d)\,\zeta(\underline{x}_i,\underline{x}_j)+2\epsilon}{(\zeta(\underline{x}_i,\underline{x}_j)+\epsilon)\,\zeta^2(\underline{x}_i,\underline{x}_j)}\chi(\underline{x}_i,\underline{x}_j)\,\underline{x}_j\,\underline{x}_j^\top\,,\\
\frac{\partial^2 \mathcal{D}_1(\mathbb{X})}{\partial \underline{x}_i \partial \underline{x}_j^\top} &= 2\,\nu_i \sum_{j=1}^{n} \nu_j \frac{\partial^2 \boldsymbol{Q}(\underline{x}_i,\underline{x}_j)}{\partial \underline{x}_i \partial \underline{x}_j^\top} = 2\,\nu_i \sum_{j=1}^{n} \nu_j \bigg(\frac{(2+d)\,\zeta(\underline{x}_i,\underline{x}_j)+2\epsilon}{2\,(\zeta(\underline{x}_i,\underline{x}_j)+\epsilon)\,\zeta^2(\underline{x}_i,\underline{x}_j)}\,\underline{x}_i\,\underline{x}_j^\top + \mathbf{I}_{d\times d}\bigg)\chi(\underline{x}_i,\underline{x}_j)\,,\\
\frac{\partial^2 \mathcal{D}_2(\mathbb{X},\widehat{\mathbb{X}})}{\partial \underline{x}_i \partial \underline{x}_i^\top} &= \nu_i \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \frac{\partial^2 \boldsymbol{Q}(\underline{x}_i,\widehat{\underline{x}}_k)}{\partial \underline{x}_i \partial \underline{x}_i^\top} = \nu_i \sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \frac{(2+d)\,\zeta(\underline{x}_i,\widehat{\underline{x}}_k)+2\epsilon}{2\,(\zeta(\underline{x}_i,\widehat{\underline{x}}_k)+\epsilon)\,\zeta^2(\underline{x}_i,\widehat{\underline{x}}_k)}\chi(\underline{x}_i,\widehat{\underline{x}}_k)\,\widehat{\underline{x}}_k\,\widehat{\underline{x}}_k^\top\,.
\end{aligned}$$

C.1.3 Specification on the Unit Circle

Given the circular variant of the MDMR unit in (6.19), we obtain its first-order derivative w.r.t. the circular distance metric $\delta(\underline{u},\underline{v}) = \underline{u}^\top\underline{v}$ following (C.1) as follows

$$\frac{\mathrm{d}\boldsymbol{Q}(\delta)}{\mathrm{d}\delta} = \frac{2\pi}{\left(\epsilon^2 - 2(1+\delta)\right)^{3/2}} = \frac{\boldsymbol{Q}(\delta)}{\epsilon^2 - 2(1+\delta)}\,.$$

The gradient of $\boldsymbol{Q}(\underline{u},\underline{v})$ w.r.t. the sample location $\underline{u}$ is then given by

$$\frac{\partial\boldsymbol{Q}(\underline{u},\underline{v})}{\partial\underline{u}} = \frac{\boldsymbol{Q}(\underline{u},\underline{v})}{\epsilon^2 - 2(1+\underline{u}^\top\underline{v})}\underline{v} =: \chi(\underline{u},\underline{v})\,\underline{v}\,, \quad \text{with} \quad \chi(\underline{u},\underline{v}) = \frac{\boldsymbol{Q}(\underline{u},\underline{v})}{\epsilon^2 - 2(1+\underline{u}^\top\underline{v})} \tag{C.7}$$

being the coefficient. Further, the Hessian of $\boldsymbol{Q}(\underline{u},\underline{v})$ w.r.t. sample locations can be derived according to the generic expression in (C.3) as

$$\begin{aligned}
\frac{\partial^2\boldsymbol{Q}(\underline{u},\underline{v})}{\partial\underline{u}\,\partial\underline{u}^\top} &= \frac{3}{\epsilon^2 - 2(1+\underline{u}^\top\underline{v})}\chi(\underline{u},\underline{v})\,\underline{v}\,\underline{v}^\top\,,\\
\frac{\partial^2\boldsymbol{Q}(\underline{u},\underline{v})}{\partial\underline{u}\,\partial\underline{v}^\top} &= \left(\frac{3}{\epsilon^2 - 2(1+\underline{u}^\top\underline{v})}\underline{u}\,\underline{v}^\top + \mathbf{I}_{2\times 2}\right)\chi(\underline{u},\underline{v})\,.
\end{aligned}$$

C.1.4 Specification on the Manifold of Planar Dual Quaternions

We rewrite the customized MCvMD unit in (6.21) into the following form

$$\boldsymbol{Q}(\delta_\mathrm{r},\delta_\mathrm{s}) = \frac{2\pi^2}{\left(\delta_\mathrm{s}^2 + 4\,\epsilon\,\delta_\mathrm{s} - 8\,\delta_\mathrm{r} + 4\,\epsilon^2 - 8\right)^{1/2}} =: \frac{2\pi^2}{\mathcal{J}(\delta_\mathrm{r},\delta_\mathrm{s})}\,,$$

with $\mathcal{J}(\delta_\mathrm{r},\delta_\mathrm{s}) = \left(\delta_\mathrm{s}^2 + 4\,\epsilon\,\delta_\mathrm{s} - 8\delta_\mathrm{r} + 4\epsilon^2 - 8\right)^{1/2}$ substituting the denominator. $\delta_\mathrm{r} = \underline{u}_\mathrm{r}^\top\underline{v}_\mathrm{r}$ and $\delta_\mathrm{s} = \|\underline{u}_\mathrm{s} - \underline{v}_\mathrm{s}\|^2$ denote distance metrics on component domains of the real and dual parts, respectively. Consequently, we obtain the gradients w.r.t. the real part and the dual part at the location $\underline{u} = [\,\underline{u}_\mathrm{r}^\top, \underline{u}_\mathrm{s}^\top\,]^\top$ as

$$\begin{aligned}
\frac{\partial\boldsymbol{Q}(\underline{u},\underline{v})}{\partial\underline{u}_\mathrm{r}} &= \frac{\partial\boldsymbol{Q}(\delta_\mathrm{r},\delta_\mathrm{s})}{\partial\delta_\mathrm{r}}\frac{\partial\delta_\mathrm{r}(\underline{u}_\mathrm{r},\underline{v}_\mathrm{s})}{\partial\underline{u}_\mathrm{r}} = \frac{8\pi^2}{\mathcal{J}^3(\delta_\mathrm{r},\delta_\mathrm{s})}\underline{v}_\mathrm{r}\,,\\
\frac{\partial\boldsymbol{Q}(\underline{u},\underline{v})}{\partial\underline{u}_\mathrm{s}} &= \frac{\partial\boldsymbol{Q}(\delta_\mathrm{r},\delta_\mathrm{s})}{\partial\delta_\mathrm{s}}\frac{\partial\delta_\mathrm{s}(\underline{u}_\mathrm{r},\underline{v}_\mathrm{s})}{\partial\underline{u}_\mathrm{s}} = -4\pi^2\frac{\delta_\mathrm{s} + 2\epsilon}{\mathcal{J}^3(\delta_\mathrm{r},\delta_\mathrm{s})}(\underline{u}_\mathrm{s} - \underline{v}_\mathrm{s})\,,
\end{aligned}$$

respectively. As mentioned in Sec. 6.4.2, the expressions of the Hessians are not provided due to redundant derivations. However, they are implemented for the corresponding showcases.

C.2 Concentration Parameter for Hyperspherical R&R

Since no analytical solution exists to (6.24), we propose to tailor Newton's method for solving it numerically. For that, we derive the derivative of $\mathcal{L}'(\lambda)$ w.r.t. λ as

$$\mathcal{L}''(\lambda) = \sum_{k=1}^{\widehat{n}}\widehat{\nu}_k\frac{\sum_{i=1}^n(\underline{x}_i^\top\widehat{\underline{x}}_k)^2\exp(\lambda\,\underline{x}_i^\top\widehat{\underline{x}}_k)}{\sum_{i=1}^n\exp(\lambda\,\underline{x}_i^\top\widehat{\underline{x}}_k)} - \sum_{k=1}^{\widehat{n}}\widehat{\nu}_k\left(\frac{\sum_{i=1}^n\underline{x}_i^\top\widehat{\underline{x}}_k\exp(\lambda\,\underline{x}_i^\top\widehat{\underline{x}}_k)}{\sum_{i=1}^n\exp(\lambda\,\underline{x}_i^\top\widehat{\underline{x}}_k)}\right)^2 - \mathcal{A}_d'(\lambda)\,, \tag{C.8}$$

which involves computing the first derivative of the Bessel function ratio. Based on its expression given by [12] in the form

$$\mathcal{A}_d'(\lambda) = 1 - \mathcal{A}_d(\lambda)^2 - \frac{d-1}{\lambda}\mathcal{A}_d(\lambda)\,,$$

λ_i at the i-th Newton step can then be updated according to

$$\lambda_{i+1} = \lambda_i - \mathcal{L}'(\lambda_i) \,/\, \mathcal{L}''(\lambda_i)\,,$$

with $\mathcal{L}'(\lambda_i)$ and $\mathcal{L}''(\lambda_i)$ given by (6.24) and (C.8), respectively. We initialize Newton's method by averaging the lower and upper bounds of λ. The initial guess follows $\lambda_0 = (\lambda_{\min} + \lambda_{\max}) \,/\, 2$, with

$$\lambda_{\min} = \mathcal{A}_d^{-1}\Big(\sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \sum_{i=1}^{n} \underline{x}_i^\top \widehat{\underline{x}}_k\Big) \quad \text{and} \quad \lambda_{\max} = \mathcal{A}_d^{-1}\Big(\sum_{k=1}^{\widehat{n}} \widehat{\nu}_k \max_{\underline{x}_i \in \mathbb{X}} \big(\{\underline{x}_i^\top \widehat{\underline{x}}_k\}\big)\Big)\,.$$

Inversion of the Bessel function $\mathcal{A}_d$ above can be solved efficiently using the method given by [7]. In practice, we also implement an additional bisection scheme based on the upper and lower bounds above for the tailored Newton's method to achieve better convergence properties [112].

Bibliography

[1] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, "Iterated Extended Kalman Filter Based Visual-Inertial Odometry Using Direct Photometric Feedback," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, 2017.

[2] I. Gilitschenski, R. Sahoo, W. Schwarting, A. Amini, S. Karaman, and D. Rus, "Deep Orientation Uncertainty Learning Based on a Bingham Loss," in *Proceedings of the 2020 International Conference on Learning Representations (ICLR 2020)*, Virtual, Apr. 2020.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision Meets Robotics: The KITTI Dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[4] J. Engel, J. Sturm, and D. Cremers, "Camera-Based Navigation of a Low-Cost Quadrocopter," in *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, Vilamoura, Portugal, Oct. 2012.

[5] T. Birdal, M. Arbel, U. Simsekli, and L. J. Guibas, "Synchronizing Probability Measures on Rotations via Optimal Transport," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1569–1579.

[6] J. Straub, O. Freifeld, G. Rosman, J. J. Leonard, and J. W. Fisher, "The Manhattan Frame Model–Manhattan World Inference in the Space of Surface Normals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 235–249, 2018.

[7] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra, "Clustering on the Unit Hypersphere Using von Mises–Fisher Distributions," *Journal of Machine Learning Research*, vol. 6, no. 9, pp. 1345–1382, 2005.

[8] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[9] J. Brookshire and S. Teller, "Extrinsic Calibration from Per-Sensor Egomotion," in *Proceedings of Robotics: Science and Systems (RSS 2013)*, Berlin, Germany, June 2013.

[10] M. Brossard, A. Barrau, and S. Bonnabel, "A Code for Unscented Kalman Filtering on Manifolds (UKF-M)," in *Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA 2020)*, Virtual, May 2020.

[11] S. Hauberg, F. Lauze, and K. S. Pedersen, "Unscented Kalman Filtering on Riemannian Manifolds," *Journal of Mathematical Imaging and Vision*, vol. 46, no. 1, pp. 103–120, 2013.

[12] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, 2009, vol. 494.

[13] G. Kurz, I. Gilitschenski, F. Pfaff, L. Drude, U. D. Hanebeck, R. Haeb-Umbach, and R. Y. Siegwart, "Directional Statistics and Filtering Using libDirectional," *Journal of Statistical Software*, May 2019. [Online]. Available: https://dx.doi.org/10.18637/jss.v089.i04

[14] T. Hamelryck, K. Mardia, and J. Ferkinghoff-Borg, *Bayesian Methods in Structural Bioinformatics*. Springer, 2012.

[15] R. A. Srivatsan, M. Xu, N. Zevallos, and H. Choset, "Probabilistic Pose Estimation Using a Bingham Distribution-Based Linear Filter," *The International Journal of Robotics Research*, vol. 37, no. 13-14, pp. 1610–1631, 2018.

[16] R. A. Srivatsan, G. T. Rosen, D. F. N. Mohamed, and H. Choset, "Estimating SE(3) Elements Using a Dual Quaternion Based Linear Kalman Filter," in *Proceedings of Robotics: Science and Systems (RSS 2016)*, Boston, USA, 2016.

[17] J. Glover, G. Bradski, and R. B. Rusu, "Monte Carlo Pose Estimation with Quaternion Kernels and the Bingham Distribution," in *Proceedings of Robotics: Science and Systems (RSS 2012)*, Los Angeles, CA, USA, June 2012.

[18] C. Ley and T. Verdebout, *Modern Directional Statistics*. CRC Press, 2017.

[19] H. Möls, K. Li, and U. D. Hanebeck, "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)*, Virtual, May 2020.

[20] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A Benchmark for the Evaluation of RGB-D SLAM Systems," in *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, Vilamoura, Portugal, Oct. 2012.

[21] A. Schaefer, J. Vertens, D. Büscher, and W. Burgard, "A Maximum Likelihood Approach to Extract Finite Planes from 3-D Laser Scans," in *Proceedings of the 2019 IEEE International Conference on Robotics and Automation (ICRA 2019)*, Montreal, QC, Canada.

[22] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-Based Visual-Inertial Odometry Using Nonlinear Optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[23] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[24] A. Tayebi, "Unit Quaternion-Based Output Feedback for the Attitude Tracking Problem," *IEEE Transactions on Automatic Control*, vol. 53, no. 6, pp. 1516–1520, 2008.

[25] K. Li, M. Li, and U. D. Hanebeck, "Towards High-Performance Solid-State-LiDAR-Inertial Odometry and Mapping," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5167–5174, 2021.

[26] K. Li, F. Pfaff, and U. D. Hanebeck, "Geometry-Driven Stochastic Modeling of SE(3) States Based on Dual Quaternion Representation," in *Proceedings of the 2019 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2019)*, Taipei, Republic of China, May 2019.

[27] A. Sveier and O. Egeland, "Dual Quaternion Particle Filtering for Pose Estimation," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 5, pp. 2012–2025, 2021.

[28] W. Feiten, P. Atwal, R. Eidenberger, and T. Grundmann, "6D Pose Uncertainty in Robotic Perception," in *Advances in Robotics Research*. Heidelberg, Germany: Springer, 2009, pp. 89–98.

[29] L. Kavan, S. Collins, J. Žára, and C. O'Sullivan, "Skinning with Dual Quaternions," in *Proceedings of the 2007 Symposium on Interactive 3D Graphics and Games*, 2007, pp. 39–46.

[30] W. D. Price, "Nonlinear Control for Dual Quaternion Systems," Ph.D. dissertation, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, Nov. 2013.

[31] S. Bultmann, K. Li, and U. D. Hanebeck, "Stereo Visual SLAM Based on Unscented Dual Quaternion Filtering," in *Proceedings of the 22nd International Conference on Information Fusion (Fusion 2019)*, Ottawa, Canada, July 2019.

[32] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Unscented Orientation Estimation Based on the Bingham Distribution," *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 172–177, Jan. 2016.

[33] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D.C., USA, June 2013.

[34] ——, "Unscented von Mises–Fisher Filtering," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 463–467, Apr. 2016.

[35] K. Li, D. Frisch, B. Noack, and U. D. Hanebeck, "Geometry-Driven Deterministic Sampling for Nonlinear Bingham Filtering," in *Proceedings of the 2019 European Control Conference (ECC 2019)*, Naples, Italy, June 2019.

[36] K. Li, F. Pfaff, and U. D. Hanebeck, "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, Canada, July 2019.

[37] ——, "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, Sept. 2020.

[38] ——, "Progressive von Mises–Fisher Filtering Using Isotropic Sample Sets for Nonlinear Hyperspherical Estimation," *Sensors*, vol. 21, no. 9, p. 2991, 2021.

[39] K. Li, D. Frisch, S. Radtke, B. Noack, and U. D. Hanebeck, "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, Oct. 2018.

[40] K. Li, G. Kurz, L. Bernreiter, and U. D. Hanebeck, "Nonlinear Progressive Filtering for SE(2) Estimation," in *Proceedings of the 21st International Conference on Information Fusion (Fusion 2018)*, Cambridge, United Kingdom, July 2018.

[41] G. Kurz, F. Pfaff, and U. D. Hanebeck, "Application of Discrete Recursive Bayesian Estimation on Intervals and the Unit Circle to Filtering on SE(2)," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 1197–1206, Mar. 2018. [Online]. Available: https://doi.org/10.1109/TII.2017.2757011

[42] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[43] K. Li, F. Pfaff, and U. D. Hanebeck, "Grid-Based Quaternion Filter for SO(3) Estimation," in *Proceedings of the 2020 European Control Conference (ECC 2020)*, Virtual, May 2020.

[44] F. Pfaff, K. Li, and U. D. Hanebeck, "Fourier Filters, Grid Filters, and the Fourier-Interpreted Grid Filter," in *Proceedings of the 22nd International Conference on Information Fusion (Fusion 2019)*, Ottawa, Canada, July 2019.

[45] ——, "The Spherical Grid Filter for Nonlinear Estimation on the Unit Sphere," in *Proceedings of the 1st Virtual IFAC World Congress (IFAC-V 2020)*, July 2020.

[46] ——, "A Hyperhemispherical Grid Filter for Orientation Estimation," in *Proceedings of the 23rd International Conference on Information Fusion (Fusion 2020)*, Virtual, July 2020.

[47] ——, "Estimating Correlated Angles Using the Hypertoroidal Grid Filter," in *Proceedings of the 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2020)*, Virtual, Sept. 2020.

[48] ——, "The State Space Subdivision Filter for Estimation on SE(2)," *Sensors*, vol. 21, no. 18, p. 6314, 2021.

[49] K. Li, F. Pfaff, and U. D. Hanebeck, "Hyperspherical Unscented Particle Filter for Nonlinear Orientation Estimation," in *Proceedings of the 1st Virtual IFAC World Congress (IFAC-V 2020)*, July 2020.

[50] ——, "Unscented Dual Quaternion Particle Filter for SE(3) Estimation," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 647–652, 2021.

[51] ——, "Hyperspherical Dirac Mixture Reapproximation," *arXiv preprint arXiv:2110.10411*, Oct. 2021. [Online]. Available: https://arxiv.org/abs/2110.10411

[52] ——, "Dual Quaternion Sample Reduction for SE(2) Estimation," in *Proceedings of the 23rd International Conference on Information Fusion (Fusion 2020)*, Virtual, July 2020.

[53] G. Kurz and U. D. Hanebeck, "Deterministic Sampling on the Torus for Bivariate Circular Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 530–534, Feb. 2017.

[54] F. Pfaff, K. Li, and U. D. Hanebeck, "Deep Likelihood Learning for 2-D Orientation Estimation Using a Fourier Filter," in *Proceedings of the 24th International Conference on Information Fusion (Fusion 2021)*, South Africa, Nov. 2021.

[55] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "A New Probability Distribution for Simultaneous Representation of Uncertain Position and Orientation," in *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, Salamanca, Spain, July 2014.

[56] I. Gilitschenski, G. Kurz, and U. D. Hanebeck, "A Stochastic Filter for Planar Rigid-Body Motions," in *Proceedings of the 2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2015)*, San Diego, California, USA, Sept. 2015.

[57] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Deterministic Approximation of Circular Densities with Symmetric Dirac Mixtures Based on Two Circular Moments," in *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, Salamanca, Spain, July 2014.

[58] G. Kurz, I. Gilitschenski, R. Y. Siegwart, and U. D. Hanebeck, "Methods for Deterministic Approximation of Circular Densities," *Journal of Advances in Information Fusion*, vol. 11, no. 2, pp. 138–156, Dec. 2016. [Online]. Available: https://confcats_isif.s3.amazonaws.com/web-files/journals/entries/JAIF_Vol11_2_3.pdf

[59] U. D. Hanebeck and A. Lindquist, "Moment-Based Dirac Mixture Approximation of Circular Densities," in *Proceedings of the 19th IFAC World Congress (IFAC 2014)*, Cape Town, South Africa, Aug. 2014.

[60] I. Gilitschenski, G. Kurz, U. D. Hanebeck, and R. Siegwart, "Optimal Quantization of Circular Distributions," in *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)*, Heidelberg, Germany, July 2016.

[61] U. D. Hanebeck, K. Briechle, and A. Rauh, "Progressive Bayes: A New Framework for Nonlinear State Estimation," in *Proceedings of SPIE, AeroSense Symposium*, vol. 5099, Orlando, Florida, USA, May 2003, pp. 256–267.

[62] J. Steinbring and U. D. Hanebeck, "S2KF: The Smart Sampling Kalman Filter," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, July 2013.

[63] I. Gilitschenski, G. Kurz, and U. D. Hanebeck, "Non-Identity Measurement Models for Orientation Estimation Based on Directional Statistics," in *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)*, Washington D.C., USA, July 2015.

[64] R. v. d. Merwe, A. Doucet, N. d. Freitas, and E. A. Wan, "The Unscented Particle Filter," in *Proceedings of the 2001 Advances in Neural Information Processing Systems*, 2001, pp. 584–590.

[65] W. M. Wonham, "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, vol. 2, no. 3, pp. 347–369, 1964.

[66] G. Kurz, F. Pfaff, and U. D. Hanebeck, "Discrete Recursive Bayesian Filtering on Intervals and the Unit Circle," in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2016)*, Baden-Baden, Germany, Sept. 2016.

[67] P. Leopardi, "A Partition of the Unit Sphere into Regions of Equal Area and Small Diameter," *Electronic Transactions on Numerical Analysis*, vol. 25, no. 12, pp. 309–327, 2006.

[68] G. Kurz, F. Pfaff, and U. D. Hanebeck, "Discretization of SO(3) Using Recursive Tesseract Subdivision," in *Proceedings of the 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2017)*, Daegu, Republic of Korea, Nov. 2017.

[69] U. D. Hanebeck, M. F. Huber, and V. Klumpp, "Dirac Mixture Approximation of Multivariate Gaussian Densities," in *Proceedings of the 2009 IEEE Conference on Decision and Control (CDC 2009)*, Shanghai, China, Dec. 2009.

[70] I. Gilitschenski and U. D. Hanebeck, "Efficient Deterministic Dirac Mixture Approximation," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D.C., USA, June 2013.

[71] U. D. Hanebeck, "Optimal Reduction of Multivariate Dirac Mixture Densities," *at – Automatisierungstechnik, Oldenbourg Verlag*, vol. 63, no. 4, pp. 265–278, Apr. 2015. [Online]. Available: https://dx.doi.org/10.1515/auto-2015-0005

[72] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., 1993.

[73] R. A. Fisher, "Dispersion on a Sphere," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 217, no. 1130, pp. 295–305, 1953.

[74] 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, 2012.

[75] A. Chiuso and G. Picci, "Visual Tracking of Points as Estimation on the Unit Sphere," in *The Confluence of Vision and Control*. Springer, 1998, pp. 90–105.

[76] G. Kurz and U. D. Hanebeck, "Trigonometric Moment Matching and Minimization of the Kullback–Leibler Divergence," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 3480–3484, Oct. 2015.

[77] J. Straub, T. Campbell, J. P. How, and J. W. Fisher, "Small-Variance Nonparametric Clustering on the Hypersphere," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 334–342.

[78] I. Marković, F. Chaumette, and I. Petrović, "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)*, Hongkong, May 2014.

[79] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Efficient Bingham Filtering Based on Saddlepoint Approximations," in *Proceedings of the 2014 IEEE International Conference on Multisensor Fusion and Information Integration (MFI 2014)*, Beijing, China, Sept. 2014.

[80] J. Glover and L. P. Kaelbling, "Tracking the Spin on a Ping Pong Ball with the Quaternion Bingham Filter," in *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA 2014)*, Hongkong, May 2014.

[81] K. Li, G. Kurz, L. Bernreiter, and U. D. Hanebeck, "Simultaneous Localization and Mapping Using a Novel Dual Quaternion Particle Filter," in *Proceedings of the 21st International Conference on Information Fusion (Fusion 2018)*, Cambridge, United Kingdom, July 2018.

[82] R. C. Booton, "The Analysis of Nonlinear Control System with Random Inputs," in *Proceedings of the MRI Symposium on Nonlinear Circuits*, New York, NY, USA, 1953, pp. 341–344.

[83] A. Gelb, *Applied Optimal Estimation*. MIT press, 1974.

[84] S. Hauberg, "Directional Statistics with the Spherical Normal Distribution," in *Proceedings of the 21st International Conference on Information Fusion (Fusion 2018)*, Cambridge, United Kingdom, July 2018.

[85] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

[86] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab Toolbox for Optimization on Manifolds," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1455–1459, 2014.

[87] P.-A. Absil and J. Malick, "Projection-Like Retractions on Matrix Manifolds," *SIAM Journal on Optimization*, vol. 22, no. 1, pp. 135–158, 2012.

[88] J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace, and Virtual Reality*. Princeton university press, 1999.

[89] B. Busam, T. Birdal, and N. Navab, "Camera Pose Filtering with Local Regression Geodesics on the Riemannian Manifold of Dual Quaternions," in *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2436–2445.

[90] J. M. McCarthy, *Introduction to Theoretical Kinematics*. MIT press, 1990.

[91] L. Zhang, "Dirac Delta Function of Matrix Argument," *International Journal of Theoretical Physics*, pp. 1–28, 2020.

[92] T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized Particle Filters for Mixed Linear/Nonlinear State-Space Models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, 2005.

[93] E. A. Wan and R. v. d. Merwe, "The Unscented Kalman Filter for Nonlinear Estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, Lake Louise, AB, Canada, Oct. 2000, pp. 153–158.

[94] U. D. Hanebeck and V. Klumpp, "Localized Cumulative Distributions and a Multivariate Generalization of the Cramér-von Mises Distance," in *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)*, Seoul, Republic of Korea, Aug. 2008.

[95] J. Steinbring, "Nonlinear Estimation Toolbox," https://bitbucket.org/nonlinearestimation/toolbox, 2017.

[96] J. Steinbring and U. D. Hanebeck, "LRKF Revisited: The Smart Sampling Kalman Filter (S2KF)," *Journal of Advances in Information Fusion*, vol. 9, no. 2, pp. 106–123, Dec. 2014. [Online]. Available: https://confcats_isif.s3.amazonaws.com/web-files/journals/entries/441_1_art_11_17020.pdf

[97] A. Jeffrey and H. H. Dai, *Handbook of Mathematical Formulas and Integrals*. Elsevier, 2008.

[98] A. M. Bruckner, J. B. Bruckner, and B. S. Thomson, *Real Analysis*. Prentice Hall, Upper Saddle River, NJ, 1997.

[99] E. B. Dam, M. Koch, and M. Lillholm, *Quaternions, Interpolation and Animation*. Citeseer, 1998, vol. 2.

[100] J. Duník, M. Soták, M. Veselỳ, O. Straka, and W. Hawkinson, "Design of Rao–Blackwellized Point-Mass Filter with Application in Terrain Aided Navigation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 1, pp. 251–272, 2018.

[101] K. Li, J. Cox, B. Noack, and U. D. Hanebeck, "Improved Pose Graph Optimization for Planar Motions Using Riemannian Geometry on the Manifold of Dual Quaternions," in *Proceedings of the 1st Virtual IFAC World Congress (IFAC-V 2020)*, July 2020.

[102] D. Frisch, K. Li, and U. D. Hanebeck, "Optimal Reduction of Dirac Mixture Densities on the 2-Sphere," in *Proceedings of the 1st Virtual IFAC World Congress (IFAC-V 2020)*, July 2020.

[103] S. Bonnabel, "Stochastic Gradient Descent on Riemannian Manifolds," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2217–2229, 2013.

[104] R. C. Wilson, E. R. Hancock, E. Pekalska, and R. P. Duin, "Spherical and Hyperbolic Embeddings of Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2255–2269, 2014.

[105] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Academic press, 2014.

[106] K. Sato, "Riemannian Optimal Control and Model Matching of Linear Port-Hamiltonian Systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6575–6581, 2017.

[107] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House Norwood, MA, 2007, vol. 685.

[108] K. Hornik and B. Grün, "On Maximum Likelihood Estimation of the Concentration Parameter of von Mises–Fisher Distributions," *Computational Statistics*, vol. 29, no. 5, pp. 945–957, 2014.

[109] C.-A. Deledalle, L. Denis, and F. Tupin, "Iterative Weighted Maximum Likelihood Denoising With Probabilistic Patch-Based Weights," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2661–2672, 2009.

[110] W. Jakob, "Numerically Stable Sampling of the von Mises–Fisher Distribution on $\mathbb{S}^2$ (and Other Tricks)," Interactive Geometry Lab, ETH Zürich, Tech. Rep., 2012.

[111] S. Sra, "Directional Statistics in Machine Learning: a Brief Review," *Applied Directional Statistics: Modern Methods and Case Studies*, p. 225, 2018.

[112] P. Deuflhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*. Springer Science & Business Media, 2005.