Sliced Optimal Transport on the Sphere

Michael Quellmalz[1,2] Robert Beinert[1,2] Gabriele Steidl[1]

August 2, 2023

Abstract

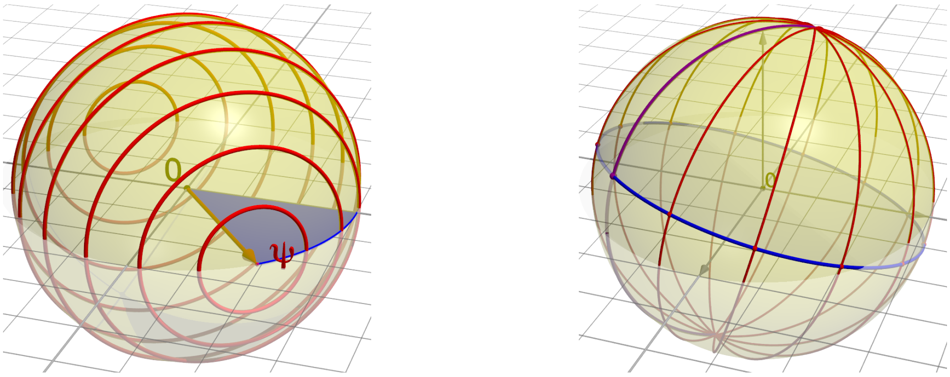
Sliced optimal transport reduces optimal transport on multi-dimensional domains to transport on the line. More precisely, sliced optimal transport is the concatenation of the well-known Radon transform and the cumulative density transform, which analytically yields the solutions of the reduced transport problems. Inspired by this concept, we propose two adaptions for optimal transport on the 2-sphere. Firstly, as counterpart to the Radon transform, we introduce the vertical slice transform, which integrates along all circles orthogonal to a given direction. Secondly, we introduce a semicircle transform, which integrates along all half great circles with an appropriate weight function. Both transforms are generalized to arbitrary measures on the sphere. While the vertical slice transform can be combined with optimal transport on the interval and leads to a sliced Wasserstein distance restricted to even probability measures, the semicircle transform is related to optimal transport on the circle and results in a different sliced Wasserstein distance for arbitrary probability measures. The applicability of both novel sliced optimal transport concepts on the sphere is demonstrated by proof-of-concept examples dealing with the interpolation and classification of spherical probability measures. The numerical implementation relies on the singular value decompositions of both transforms and fast Fourier techniques. For the inversion with respect to probability measures, we propose the minimization of an entropy-regularized Kullback–Leibler divergence, which can be numerically realized using a primal-dual proximal splitting algorithm.

1. Introduction

Optimal transport and in particular Wasserstein distances between measures have received much attention from a theoretical and practical point of view [62,79,87] and recently became of interest in neural gradient flows [4,26,49]. While Wasserstein distances are in general hard to compute, there exist analytic formulas for optimal transport on the line. Therefore sliced Wasserstein distances, which basically combine the Radon transform in Euclidean spaces with optimal transport on the line, have become quite popular [58, 72, 79]. In particular, the related Radon cumulative distribution transform has been applied for interpolation and classification as well as for model reduction [17,37,48,73,81]. The idea behind sliced optimal transport has been generalized and transferred to many related problems. There exists sliced

[1]Institute of Mathematics, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany, {quellmalz, beinert, steidl}@math.tu-berlin.de http://tu.berlin/imageanalysis

[2]These authors have contributed equally to the work.



(a) Vertical slices for a fixed direction. (b) Semicircles (red) starting in a fixed point.

Figure 1: Areas of integration of the spherical transforms.

variants [8, 16] of partial optimal transport [19, 27], where only a fraction of mass is transported, and a sliced version [20] of multi-marginal optimal transport [10, 12, 30], considering the transport between several measures instead of only two. For optimal transport on Riemannian manifolds, sliced Wasserstein distances based on the push-forward of the eigenfunctions of the Laplacian have been proposed in [77]. Especially for shape and graph analysis, sliced optimal transport has been transferred to the Gromov–Wasserstein setting [86], which more generally defines a metric between metric measure spaces [11, 55, 83]. Differently from the Wasserstein formulation with its analytic solution, the Gromov–Wasserstein transport on the line is more involved [13, 24].

In this paper, we transfer the slicing approach to optimal transport on the two-dimensional sphere. Spherical optimal transport has been intensely studied in recent years. For instance, the problem can be solved using a Monge–Ampère type equation [38, 54, 88] or a variational framework [22]. The regularity of optimal maps has been investigated in [51]. Spherical Wasserstein barycenters have been computed using a stochastic projected subgradient method [82] and have been estimated on random graphs [84].

To introduce slicing frameworks on the sphere, we do not follow the Laplacian approach in [77], but focus on spherical counterparts of the Radon transform. A well-known one is the Funk–Radon transform [29, 39, 52, 70], which takes integrals along all great circles. Integration along all circles of a fixed radius were studied in [74, 80]. Further Radon-type transforms were considered based on intersections with planes containing a fixed point inside the sphere [60, 66, 68, 78], on the sphere [1, 76], and outside the sphere [2]. Moreover, transforms including derivatives were proposed in [53, 69]. However, in the context of sliced optimal transport, we require that probability density functions on the sphere are mapped to a family of probability density functions on one-dimensional domains. For this purpose, we consider two specific spherical transforms, namely the vertical slice transform and the normalized semicircle transform.

The vertical slice transform was first considered in [32] and applied in [43, 89] for photoacoustic tomography. The generalization to higher dimensions is due to [75]. The basic idea is to take means along parallel circles, see Figure 1a, which gives a probability density function on an interval. The process is then repeated for further directions. Geometrically, the areas of integration for a fixed direction can be imagined like an “egg cutter” applied to the sphere. We generalize the vertical slice transform to probability measures and use it to define a vertical sliced Wasserstein distance. Radon transforms of measures have been considered in the context of a dual fibration, cf. [31, 59] and [39, Chap. 2, § 2], where they are defined via duality. We will see that our definition via the push-forward of measures can be also derived from that point of view.

An (unnormalized) semicircle transform was examined in [36, 41]. It takes integrals along semicircles starting in a fixed point, see Figure 1b, and yields a function defined on the one-dimensional unit circle. The process is then repeated for further starting points. This transform has been combined with optimal transport on the circle to obtain a sliced Wasserstein distance [15]. However, the crucial point is here that the unnormalized semicircle transform does not map probability density functions to probability density functions, meaning that optimal transport techniques on the circle cannot be applied. In the numeric part of [15], the authors restrict themselves to point measures, which then are projected onto great circles. This approach corresponds to an appropriately *normalized* semicircle transform instead, where the integrand is multiplied with a certain weight function. In this paper, we introduce and study this normalized semicircle transform in a rigours manner to obtain a semicircular sliced Wasserstein distance.

Main contributions

- We give rigorous definitions of the vertical slice and the normalized semicircle transform, which are originally considered only for functions, and generalize them to measures using an appropriate push-forward. For absolutely continuous measures, the generalized and initial definitions coincide in the sense that merely the density function has to be transformed. Furthermore, probability measures are transformed to probability measures.

- We prove a singular value decomposition of the normalized semicircle transform, which provides an approach for numerical computations. Moreover, the singular value decompositions of the vertical slice and the normalized semicircle transform allow the inversion via their Moore–Penrose pseudoinverses.

- We define sliced Wasserstein distances on the sphere based on both transform. We show that the normalized semicircle transform is injective for all measures, and hence the sliced Wasserstein distance indeed fulfills the properties of a metric. Furthermore, the vertical sliced Wasserstein distance is a metric for even measures on the sphere.

- We propose a Tikhonov-type regularization which minimizes a variational model consisting of the entropy-regularized Kullback–Leibler divergence. This ensures that the inverse is a probability measure and in particular non-negative. Further, this allows to compute a sliced CDT interpolation between spherical probability measures to approximate Wasserstein barycenters.

Outline of the Paper We start in Section 2 with the necessary preliminaries on optimal transport, the unit sphere and the rotation group on $\mathbb{R}^3$. Then, we introduce the two counterparts of the Radon transform on $\mathbb{S}^2$, namely the vertical slice transform in Section 3 and the normalized semicircle transform in Section 4. First, we define the transforms for functions and derive their adjoint operators and singular value decompositions on $L^2(\mathbb{S}^2)$. In order to combine these transforms with optimal transport on the interval and the circle respectively, we have to enlarge their definitions to measure spaces which we have not found in a mathematically rigorous form in the literature. Section 5 connects the above transforms with optimal transport to introduce spherical sliced Wasserstein distances for measures on the sphere. Section 6.1 deals with the discretization of the spherical transforms and their inversion, which is an ill-posed problem. For an approximate inversion, we can use the truncated Moore–Penrose pseudoinverse. However, when dealing with probability density functions, this inversion does not guarantee the non-negativity of the reconstructed function. Therefore, we suggest another reconstruction which minimizes a variational model consisting of an entropy-regularized Kullback–Leibler divergence, see Section 6.2. The actual minimization can be done by a primal-dual splitting. Numerical proof-of-concept results are reported in Section 7, where we provide two kinds of experiments. First, we show in Section 7.1 that Wasserstein barycenters on the sphere can be approximated using sliced Wasserstein transforms and Wasserstein interpolation on the interval and the circle respectively. These results require in particular the inversion of the sliced spherical transforms. Second, we demonstrate by a synthetic example that the binary classification of different measures is in principle possible in Section 7.2.

2. Preliminaries

In this section, we first provide the notation and necessary preliminaries on optimal transport, in particular on the interval and the circle. Then, we recall basic facts about the unit sphere and the rotation group on $\mathbb{R}^3$.

2.1. Measures and Optimal Transport

Let $\mathbb{X}$ be a compact metric space with metric $d\colon \mathbb{X}\times\mathbb{X}\to\mathbb{R}$, and let $\mathcal{B}(\mathbb{X})$ be the Borel σ-algebra induced by d. By $\mathcal{M}(\mathbb{X})$, we denote the Banach space of signed, finite measures, and by $\mathcal{P}(\mathbb{X})$ the subset of probability measures on $\mathbb{X}$. The pre-dual space of $\mathcal{M}(\mathbb{X})$ is $C(\mathbb{X})$. Let $\mathbb{Y}$ be another compact metric space and $T\colon \mathbb{X}\to\mathbb{Y}$ be measurable. For $\mu\in\mathcal{M}(\mathbb{X})$, we define the *push-forward measure* $T_\#\mu \coloneqq \mu\circ T^{-1}\in\mathcal{M}(\mathbb{Y})$. For any measure $\pi\in\mathcal{M}(\mathbb{X}\times\mathbb{Y})$ with first marginal $\mu\in\mathcal{M}(\mathbb{X})$, i.e., $\pi(B\times\mathbb{Y})=\mu(B)$ for all $B\in\mathcal{B}(\mathbb{X})$, we call a collection of measures $\pi_x\in\mathcal{M}(\mathbb{Y})$, $x\in\mathbb{X}$, a *disintegration family* if

$$\int_{\mathbb{X}\times\mathbb{Y}} f(x,y)\,\mathrm{d}\pi(x,y) = \int_{\mathbb{X}}\int_{\mathbb{Y}} f(x,y)\,\mathrm{d}\pi_x(y)\,\mathrm{d}\mu(x)$$

for all measurable functions f on $\mathbb{X}\times\mathbb{Y}$.

The *p-Wasserstein distance*, $p\in[1,\infty)$, of $\mu,\nu\in\mathcal{P}(\mathbb{X})$ is given by

$$\mathrm{W}_p^p(\mu,\nu) \coloneqq \min_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{X}^2} d^p(x,y)\,\mathrm{d}\pi(x,y), \tag{1}$$

with $\Pi(\mu,\nu)\coloneqq\{\pi\in\mathcal{M}(\mathbb{X}\times\mathbb{X}) : \pi(B\times\mathbb{X})=\mu(B), \pi(\mathbb{X}\times B)=\nu(B) \text{ for all } B\in\mathcal{B}(\mathbb{X})\}$. It defines a metric on $\mathcal{P}(\mathbb{X})$. The metric space $\mathcal{P}^p(\mathbb{X})\coloneqq(\mathcal{P}(\mathbb{X}),W_p)$ is called p-Wasserstein space and, in case $p=2$, just Wasserstein space. The above Wasserstein distance is just a special case of the more general optimal transport problem, where $d^p(x,y)$ can be replaced by a more general cost function $c(x,y)$. For $\delta \in [0,1]$, the *p-Wasserstein barycenter* between $\mu, \nu \in \mathcal{P}^p(\mathbb{X})$ is the minimizer of

$$\min_{\omega \in \mathcal{P}(\mathbb{X})} (1-\delta)\ \mathrm{W}_p^p(\mu,\omega) + \delta\ \mathrm{W}_p^p(\nu,\omega), \tag{2}$$

see [3]. Note that the Wasserstein barycenter between absolutely continuous measures is unique, cf. [45].

Optimal Transport on the Interval If $\mathbb{X}$ is the *unit interval* $\mathbb{I} \coloneqq [-1,1]$ with the distance $d(x,y) = |x-y|$, the optimal transport between two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{I})$ can be computed easily [62,79,87] using the *cumulative distribution function* $F_\mu(x) \coloneqq \mu([-1,x])$, $x \in \mathbb{I}$, which is non-decreasing and right-continuous. Its pseudoinverse, the *quantile function* $F_\mu^{-1}(r) \coloneqq \min\{x \in \mathbb{I} : F_\mu(x) \geq r\}$, $r \in [0,1]$, is non-decreasing and left-continuous. The measure μ can be recovered by $\mu = (F_\mu^{-1})_\# \sigma_{[0,1]}$, where $\sigma_{[0,1]}$ denotes the Lebesque measure on $[0,1]$. The p-Wasserstein distance (1) between $\mu, \nu \in \mathcal{P}^p(\mathbb{I})$ now equals $\mathrm{W}_p(\mu,\nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L^p([0,1])}$. Moreover, if $\mu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{I})$, where $\mathcal{P}_{\mathrm{ac}}(\mathbb{I})$ denotes the probability measures that are *absolutely continuous* with respect to the Lebesgue measure, then the optimal transport plan π in (1) is uniquely given by

$$\pi = (\mathrm{Id}, T^{\mu,\nu})_\# \mu \quad \text{with} \quad T^{\mu,\nu}(x) \coloneqq F_\nu^{-1}(F_\mu(x)), \quad x \in \mathbb{I}.$$

Based on the *optimal transport map* $T^{\mu,\nu}$, the Wasserstein space $\mathcal{P}^p(\mathbb{I})$ can be isometrically embedded into $L^p_\omega(\mathbb{I})$ with $\omega \in \mathcal{P}_{\mathrm{ac}}(\mathbb{I})$ [11,48,61], where $L^p_\omega(\mathbb{I})$ consists of all p-integrable functions with respect to ω. More precisely, for the reference measure $\omega \in \mathcal{P}_{\mathrm{ac}}(\mathbb{I})$, the *cumulative distribution transform* (CDT) is defined by $\mathrm{CDT}_\omega \colon \mathcal{P}^p(\mathbb{I}) \to L^p_\omega(\mathbb{I})$ with

$$\mathrm{CDT}_\omega[\mu](x) \coloneqq (T^{\omega,\mu} - \mathrm{Id})(x) = \left(F_\mu^{-1} \circ F_\omega\right)(x) - x, \quad x \in \mathbb{I}, \tag{3}$$

and we especially have $\mathrm{W}_p(\mu,\nu) = \|\mathrm{CDT}_\omega[\mu] - \mathrm{CDT}_\omega[\nu]\|_{L^p_\omega(\mathbb{I})}$. The CDT is in fact a mapping from $\mathcal{P}^p(\mathbb{I})$ into the tangent space of $\mathcal{P}^p(\mathbb{I})$ at ω, see [5, § 8.5]. Due to the relation to the optimal transport map, the CDT can be inverted by $\mu = \mathrm{CDT}_\omega^{-1}[h] \coloneqq (h + \mathrm{Id})_\# \omega$ for $h = \mathrm{CDT}_\omega[\mu]$. If $\mu, \omega \in \mathcal{P}_{\mathrm{ac}}(\mathbb{I})$ possess the density functions $f_\mu > 0$, $f_\omega > 0$, then, by the transformation formula for push-forward measures, f_μ can be recovered by

$$f_\mu(x) = \left(g^{-1}\right)'(x)\, f_\omega(g^{-1}(x)) \quad \text{with} \quad g(x) = \mathrm{CDT}_\omega[\mu](x) + x, \quad x \in \mathbb{I}. \tag{4}$$

For $\mu, \nu \in \mathcal{P}(\mathbb{I})$, and an arbitrary reference measure $\omega \in \mathcal{P}_{\mathrm{ac}}(\mathbb{I})$, the 2-Wasserstein barycenter (2) has the form

$$\mathrm{CDT}_\omega^{-1}\left(\delta\, \mathrm{CDT}_\omega[\nu] + (1-\delta)\, \mathrm{CDT}_\omega[\mu]\right), \tag{5}$$

see [48]. In particular for $\omega = \mu$, we have by (3) that $\mathrm{CDT}_\mu[\mu](x) = F_\mu^{-1}(F_\mu(x)) - x = 0$ and therefore the barycenter (5) becomes

$$\mathrm{CDT}_\mu^{-1}\left(\delta\, \mathrm{CDT}_\mu[\nu]\right).$$

Optimal Transport on the Circle On the *circle* $\mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$ equipped with the metric $d(x, y) := \min_{k\in\mathbb{Z}} |x - y + 2\pi k|$, the optimal transport can be computed in a similar manner by incorporating the periodicity. Following [23, 71], we define the *(extended) cumulative distribution function* by $\tilde{F}_\mu(x) := \mu([0, x])$ for $x \in [0, 2\pi]$ and extend it to $\mathbb{R}$ by the convention $\tilde{F}_\mu(x + 2\pi) := \tilde{F}_\mu(x) + 1$. Its pseudoinverse, the *(extended) quantile function*, is defined as $\tilde{F}_\mu^{-1}(r) := \min\{x \in \mathbb{R} : \tilde{F}_\mu(x) \geq r\}$ for $r \in \mathbb{R}$. Note that $\tilde{F}$ and $\tilde{F}^{-1}$ are mappings defined on entire $\mathbb{R}$. The p-Wasserstein distance between $\mu, \nu \in \mathcal{P}(\mathbb{T})$ is given by

$$W_p^p(\mu, \nu) = \min_{\theta\in\mathbb{R}} \int_0^1 |\tilde{F}_\mu^{-1}(r) - (\tilde{F}_\nu - \theta)^{-1}(r)|^p \,\mathrm{d}r, \tag{6}$$

where $(\tilde{F}_\nu - \theta)^{-1}$ is the pseudoinverse of the shifted cumulative distribution function [71]. For $\mu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{T})$, each minimizer θ of (6) yields an optimal transport plan

$$\pi = (\mathrm{Id}, \iota(\tilde{T}^{\mu,\nu}))_\# \mu \quad \text{with} \quad \tilde{T}^{\mu,\nu}(x) := (\tilde{F}_\nu - \theta)^{-1}(\tilde{F}_\mu(x)), \quad x \in [0, 2\pi),$$

where $\iota \colon \mathbb{R} \to \mathbb{T}$ denotes the canonical projection from the line to the circle. Note that $\tilde{T}^{\mu,\nu}(x) \in \mathbb{R}$ is the representative of $\iota(\tilde{T}^{\mu,\nu}(x)) \in \mathbb{T}$ with

$$d(x, \iota(\tilde{T}^{\mu,\nu}(x))) = |x - \tilde{T}^{\mu,\nu}(x)|.$$

If $p > 1$ and $\mu, \nu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{T})$, the minimizer θ of (6) is unique. This follows by the proof of [23, Lem. 5.2], where it is shown that the objective of (6) is convex in θ, but the argument even implies strict convexity. In analogy to (3), we define for $p \in (1, \infty)$, the *circular CDT* (cCDT) of $\mu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{T})$ with reference measure $\omega \in \mathcal{P}_{\mathrm{ac}}(\mathbb{T})$ by $\mathrm{cCDT}_\omega \colon \mathcal{P}^p(\mathbb{T}) \to L^p_\omega(\mathbb{T})$ with

$$\mathrm{cCDT}_\omega[\mu](x) := (\tilde{T}^{\omega,\mu} - \mathrm{Id})(x) = \big((\tilde{F}_\mu - \theta_{\omega,\mu})^{-1} \circ \tilde{F}_\omega\big)(x) - x, \quad x \in [0, 2\pi),$$

where $\tilde{T}^{\omega,\mu}$ is the optimal transport plan and $\theta_{\omega,\mu}$ the minimizer of (6). Note that the cCDT is no longer an isometric embedding. The cCDT can be inverted by $\mu = \mathrm{cCDT}_\omega^{-1}[h] := (\iota \circ (h + \mathrm{Id}))_\# \omega$ for $h = \mathrm{cCDT}_\omega[\mu]$. If $\mu, \omega \in \mathcal{P}_{\mathrm{ac}}(\mathbb{T})$ have densities $f_\mu > 0$, $f_\omega > 0$, the density f_μ can be recovered similarly to (4) via

$$f_\mu(x) = \big(g^{-1}\big)'(x)\, f_\omega(g^{-1}(x)) \quad \text{with} \quad g(x) = \iota\left(\mathrm{cCDT}_\omega[\mu](x) + x\right), \quad x \in \mathbb{T}.$$

In analogy to (5) with $\omega = \mu$, we interpolate between the measures $\mu, \nu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{T})$ by

$$\mathrm{cCDT}_\mu^{-1}[\delta\, \mathrm{cCDT}_\mu[\nu]].$$

2.2. Sphere and Rotation Group

Unit Sphere The two-dimensional unit sphere is defined as $\mathbb{S}^2 := \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\| = 1\}$. The canonical unit vectors are henceforth denoted by $\boldsymbol{e}^j$, $j = 1, 2, 3$. Points $\boldsymbol{\xi} \in \mathbb{S}^2$ can be parameterized in *spherical coordinates*

$$\boldsymbol{\xi} = \Phi(\varphi, \vartheta) := (\cos\varphi\, \sin\vartheta,\, \sin\varphi\, \sin\vartheta,\, \cos\vartheta) \in \mathbb{S}^2, \quad \varphi \in \mathbb{T},\ \vartheta \in [0, \pi].$$

The restriction $\Phi \colon (\mathbb{T} \times (0, \pi)) \cup (\{0\} \times \{0, \pi\}) \to \mathbb{S}^2$ is a bijective mapping. We denote the first and second component of this restriction as *azimuth angle* $\mathrm{azi}(\boldsymbol{\xi})$ and *zenith angle* $\mathrm{zen}(\boldsymbol{\xi})$, respectively, which are uniquely given by

$$\mathrm{azi}(\Phi(\varphi, \vartheta)) = \varphi \quad \text{and} \quad \mathrm{zen}(\Phi(\varphi, \vartheta)) = \vartheta$$

for all $(\varphi, \vartheta) \in (\mathbb{T} \times (0,\pi)) \cup (\{0\} \times \{0,\pi\})$. The *surface measure* $\sigma_{\mathbb{S}^2}$ on the sphere is given by

$$\int_{\mathbb{S}^2} f(\boldsymbol{\xi}) \,\mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}) = \int_0^\pi \int_{\mathbb{T}} f(\Phi(\varphi,\vartheta)) \sin\vartheta \,\mathrm{d}\varphi \,\mathrm{d}\vartheta.$$

Normalizing $\sigma_{\mathbb{S}^2}$ yields the *uniform measure* $u_{\mathbb{S}^2} \coloneqq (4\pi)^{-1}\sigma_{\mathbb{S}^2}$. We denote by $L^p(\mathbb{S}^2)$, $p \in [1,\infty]$, the Banach space of all (equivalence classes of) p-integrable functions on $\mathbb{S}^2$, where we use the above surface measure.

We define the *spherical harmonics* of degree $n \in \mathbb{N}_0$ and order $k = -n, \dots, n$ by

$$Y_n^k(\Phi(\varphi,t)) \coloneqq \sqrt{\frac{2n+1}{4\pi}\frac{(n-k)!}{(n+k)!}}\, P_n^k(\cos\vartheta)\,\mathrm{e}^{\mathrm{i}k\varphi}, \tag{7}$$

where $P_n^k \colon [-1,1] \to \mathbb{R}$ denotes the *associated Legendre functions* defined by

$$P_n^k(t) \coloneqq \frac{(-1)^k}{2^n n!}(1-t^2)^{\frac{k}{2}} \frac{\mathrm{d}^{n+k}(t^2-1)^n}{\mathrm{d}t^{n+k}}, \qquad n \in \mathbb{N}_0,\ k \in \{0,\dots,n\} \tag{8}$$

and

$$P_n^{-k} \coloneqq (-1)^k \frac{(n-k)!}{(n+k)!} P_n^k. \tag{9}$$

The spherical harmonics $\{Y_n^k : n \in \mathbb{N}_0,\ k = -n,\dots,n\}$ form an orthonormal basis of $L^2(\mathbb{S}^2)$. Finally, the *Sobolev space* $H^s(\mathbb{S}^2)$ with $s \geq 0$, is defined as the completion of $C^\infty(\mathbb{S}^2)$ with respect to the norm

$$\|f\|_{H^s(\mathbb{S}^2)} \coloneqq \sum_{n=0}^\infty \left(n+\tfrac{1}{2}\right)^{2s} \sum_{k=-n}^n \left|\langle f, Y_n^k\rangle_{L^2(\mathbb{S}^2)}\right|^2. \tag{10}$$

Rotation Group Next, we are interested in the *rotation group*

$$\mathrm{SO}(3) \coloneqq \{\boldsymbol{Q} \in \mathbb{R}^{3\times 3} : \boldsymbol{Q}^\top \boldsymbol{Q} = I, \det(\boldsymbol{Q}) = 1\}.$$

Any matrix in SO(3) has an *Euler angle* parameterization

$$\Psi(\alpha,\beta,\gamma) \coloneqq \boldsymbol{R}_3(\alpha)\boldsymbol{R}_2(\beta)\boldsymbol{R}_3(\gamma) \in \mathrm{SO}(3), \qquad \alpha,\gamma \in \mathbb{T},\ \beta \in [0,\pi],$$

where

$$\boldsymbol{R}_3(\alpha) \coloneqq \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ \boldsymbol{R}_2(\beta) \coloneqq \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}. \tag{11}$$

The rotation group SO(3) can be identified with the product $\mathbb{S}^2 \times \mathbb{T}$ via the bijection

$$\mathbb{S}^2 \times \mathbb{T} \ni (\boldsymbol{\xi},\gamma) \mapsto \Psi(\mathrm{azi}(\boldsymbol{\xi}), \mathrm{zen}(\boldsymbol{\xi}), \gamma) \in \mathrm{SO}(3),$$

cf. [34]. In Euler angles, the rotationally invariant measure $\sigma_{\mathrm{SO}(3)}$ on SO(3) is given by

$$\begin{aligned} \int_{\mathrm{SO}(3)} f(\boldsymbol{Q}) \,\mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{Q}) &= \int_0^{2\pi}\int_0^\pi\int_0^{2\pi} f(\Psi(\alpha,\beta,\gamma)) \sin(\beta) \,\mathrm{d}\alpha \,\mathrm{d}\beta \,\mathrm{d}\gamma \\ &= \int_{\mathbb{T}}\int_{\mathbb{S}^2} f(\Psi(\alpha,\beta,\gamma)) \,\mathrm{d}\sigma_{\mathbb{S}^2}(\Phi(\alpha,\beta)) \,\mathrm{d}\gamma. \end{aligned} \tag{12}$$

The uniform measure on SO(3) is $u_{\mathrm{SO}(3)} \coloneqq (8\pi^2)^{-1}\sigma_{\mathrm{SO}(3)}$.

The *rotational harmonics* or *Wigner D-functions* $D_n^{k,j}$ of degree $n \in \mathbb{N}_0$ and orders $k, j \in \{-n, \dots, n\}$ are defined by

$$D_n^{k,j}(\Psi(\alpha,\beta,\gamma)) \coloneqq \mathrm{e}^{-\mathrm{i}k\alpha}\, d_n^{k,j}(\cos\beta)\, \mathrm{e}^{-\mathrm{i}j\gamma}, \tag{13}$$

where the *Wigner d-functions* are given for $t \in [-1,1]$ by

$$d_n^{k,j}(t) \coloneqq \frac{(-1)^{n-j}}{2^n}\sqrt{\frac{(n+k)!(1-t)^{j-k}}{(n-j)!(n+j)!(n-k)!(1+t)^{j+k}}}\frac{\mathrm{d}^{n-k}}{\mathrm{d}t^{n-k}}\frac{(1+t)^{n+j}}{(1-t)^{-n+j}},$$

see [85, chap. 4]. The rotational harmonics are the matrix entries of the left angular representations of SO(3), i.e.,

$$Y_n^k(\boldsymbol{Q}^\top \boldsymbol{\xi}) = \sum_{j=-n}^{n} D_n^{j,k}(\boldsymbol{Q}) Y_n^j(\boldsymbol{\xi}), \quad \boldsymbol{Q} \in \mathrm{SO}(3),\ \boldsymbol{\xi} \in \mathbb{S}^2. \tag{14}$$

They satisfy the orthogonality relation

$$\int_{\mathrm{SO}(3)} D_n^{j,k}(\boldsymbol{Q}) D_{n'}^{j',k'}(\boldsymbol{Q})\,\mathrm{d}\boldsymbol{Q} = \frac{8\pi^2}{2n+1}\delta_{n,n'}\delta_{k,k'}\delta_{j,j'}, \tag{15}$$

for all $n, n' \in \mathbb{N}_0$, $j, k = -n, \dots, n$, and $j', k' = -n', \dots, n'$, where δ denotes the Kronecker symbol. Then $\{(\frac{2n+1}{8\pi^2})^{\frac{1}{2}} D_n^{j,k} : n \in \mathbb{N}_0, j, k = -n, \dots, n\}$ form an orthonormal basis of $L^2(\mathrm{SO}(3))$.

Finally, the *Sobolev space* $H^s(\mathrm{SO}(3))$ with $s \ge 0$ is defined as the completion of $C^\infty(\mathrm{SO}(3))$ with respect to the Sobolev norm

$$\|g\|^2_{H^s(\mathrm{SO}(3))} \coloneqq \sum_{n=0}^{\infty} \left(n + \tfrac{1}{2}\right)^{2s} \sum_{j,k=-n}^{n} \frac{8\pi^2}{2n+1} |\langle g, D_n^{j,k}\rangle|^2.$$

3. Vertical Slice Transform

3.1. Vertical Slice Transform of Functions

In analogy to the Radon transform, the main idea behind the vertical slice transform is to integrate a given function $f \colon \mathbb{S}^2 \to \mathbb{R}$ along parallel vertical slices. To describe these slices mathematically, we define the *slicing operator* $\mathcal{S}_\psi \colon \mathbb{S}^2 \to \mathbb{I}$ for any fixed $\psi \in \mathbb{T}$ by

$$\mathcal{S}_\psi(\boldsymbol{\xi}) \coloneqq \langle \boldsymbol{\xi}, (\cos\psi, \sin\psi, 0)^\top \rangle = \cos(\psi)\,\xi_1 + \sin(\psi)\,\xi_2,$$

and the corresponding *slice/circle* by

$$C_\psi^t \coloneqq \mathcal{S}_\psi^{-1}(t) = \{\boldsymbol{\xi} \in \mathbb{S}^2 : \mathcal{S}_\psi(\boldsymbol{\xi}) = t\}, \qquad t \in \mathbb{I}.$$

The slice C_ψ^t is the intersection of $\mathbb{S}^2$ and the plane with normal $(\cos\psi, \sin\psi, 0)^\top$ and distance t from the origin, An illustration of the slices C_ψ^t for fixed ψ is given in Figure 1a. The *vertical slice transform* $\mathcal{V}$ is defined by

$$\mathcal{V}f(\psi, t) \coloneqq \frac{1}{2\pi\sqrt{1-t^2}} \int_{C_\psi^t} f(\boldsymbol{\xi})\,\mathrm{d}s(\boldsymbol{\xi}), \qquad \psi \in \mathbb{T},\ t \in (-1,1), \tag{16}$$

where $\mathrm{d}s$ denotes the arc-length on C_ψ^t. For $t = \pm 1$, the *vertical slice transform* is

$$\mathcal{V} f(\psi, 1) \coloneqq f(\cos\psi, \sin\psi, 0) \quad \text{and} \quad \mathcal{V} f(\psi, -1) \coloneqq f(-\cos\psi, -\sin\psi, 0).$$

For fixed $\psi \in \mathbb{T}$, we define the (normalized) restrictions

$$\mathcal{V}_\psi \coloneqq 2\pi\, \mathcal{V}(\psi, \cdot). \tag{17}$$

This corresponds to projecting the mean values of f along C_ψ^t to $t \in \mathbb{I}$. For an illustration see again Figure 1a. The different normalizations of $\mathcal{V}$ and $\mathcal{V}_\psi$ are chosen with respect to the later generalization to measures and ensure that density functions are transformed to density functions by $\mathcal{V}$ and $\mathcal{V}_\psi$. By the following proposition, both operators are well defined almost everywhere.

Proposition 3.1. *Let $1 \le p \le \infty$. For every $f \in L^p(\mathbb{S}^2)$, it holds*

$$\int_{\mathbb{I}} \mathcal{V}_\psi f(t) \,\mathrm{d}t = \int_{\mathbb{S}^2} f(\boldsymbol{\xi}) \,\mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}) \quad \textit{and} \quad \int_{\mathbb{T}} \int_{\mathbb{I}} \mathcal{V} f(\psi, t) \,\mathrm{d}t \,\mathrm{d}\psi = \int_{\mathbb{S}^2} f(\boldsymbol{\xi}) \,\mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}). \tag{18}$$

Let $\psi \in \mathbb{T}$. The operators $\mathcal{V}_\psi \colon L^p(\mathbb{S}^2) \to L^p(\mathbb{I})$ and $\mathcal{V} \colon L^p(\mathbb{S}^2) \to L^p(\mathbb{T} \times \mathbb{I})$ are bounded with

$$\|\mathcal{V}_\psi\|_{L^p \to L^p} = (2\pi)^{1-1/p} \quad \textit{and} \quad \|\mathcal{V}\|_{L^p \to L^p} = 1.$$

Moreover, it holds $\mathcal{V}_\psi \colon C(\mathbb{S}^2) \to C(\mathbb{I})$ and $\mathcal{V} \colon C(\mathbb{S}^2) \to C(\mathbb{T} \times \mathbb{I})$.

Proof. We parameterize the *upper* and *lower hemispheres* by

$$H_\psi^\pm(s, t) \coloneqq \begin{pmatrix} t\cos(\psi) - s\sin(\psi) \\ t\sin(\psi) + s\cos(\psi) \\ \pm\sqrt{1 - t^2 - s^2} \end{pmatrix}, \quad s \in \sqrt{1 - t^2}\,\mathbb{I},\ t \in \mathbb{I}.$$

Then the upper and lower semicircle of C_ψ^t can be parameterized via $H_\psi^\pm(\cdot, t)$. Thus we obtain

$$\begin{aligned} \int_{\mathbb{S}^2} f(\boldsymbol{\xi}) \,\mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}) &= \int_{\mathbb{I}} \int_{\sqrt{1-t^2}\,\mathbb{I}} \Big(f(H_\psi^+(s, t) + f(H_\psi^-(s, t)) \Big) \frac{1}{\sqrt{1 - t^2 - s^2}} \,\mathrm{d}s \,\mathrm{d}t \\ &= \int_{\mathbb{I}} \frac{1}{\sqrt{1 - t^2}} \int_{C_\psi^t} f(\boldsymbol{\xi}) \,\mathrm{d}s(\boldsymbol{\xi}) \,\mathrm{d}t = \int_{\mathbb{I}} \mathcal{V}_\psi f(t) \,\mathrm{d}t. \end{aligned}$$

Using (17) and integrating over ψ immediately yields the second identity in (18). By Fubini's theorem, $\mathcal{V}_\psi$ and $\mathcal{V}$ are well defined.

Following the above computation for the absolute value of f, we obtain with the triangle inequality $\|\mathcal{V}_\psi\|_{L^1 \to L^1} = \|\mathcal{V}\|_{L^1 \to L^1} = 1$. Since the vertical slice transform is essentially bounded by

$$|\mathcal{V}_\psi f(t)| \le \frac{1}{\sqrt{1 - t^2}} \int_{C_\psi^t} |f(\boldsymbol{\xi})| \mathrm{d}s(\boldsymbol{\xi}) \le 2\pi \operatorname*{ess\,sup}_{\boldsymbol{\xi} \in \mathbb{S}^2} |f(\boldsymbol{\xi})|,$$

we further have $\|\mathcal{V}_\psi\|_{L^\infty \to L^\infty} = 2\pi$ and $\|\mathcal{V}\|_{L^\infty \to L^\infty} = 1$. Now the second assertion follows from the Riesz–Thorin interpolation theorem.

The last assertion is an immediate consequence of Lebesgue's dominated convergence theorem. $\square$

Since all circles C_ψ^t are symmetric with respect to the ξ_1-ξ_2 plane, $\mathcal{V}f$ vanishes for functions f which are odd in the third coordinate, i.e., $f(\xi_1, \xi_2, \xi_3) = -f(\xi_1, \xi_2, -\xi_3)$. For brevity, we call these functions *odd*. In [32], an explicit inversion formula for *even* functions, i.e., $f(\xi_1, \xi_2, \xi_3) = f(\xi_1, \xi_2, -\xi_3)$, is derived. However, as for the Radon inversion formula, this formula leads to instable practical computations if we leave the range of $\mathcal{V}$. For numerical simulation, we will invert $\mathcal{V}$ using its singular value decomposition. For this purpose, notice that the spherical harmonics Y_n^k with even $k+n$ are even functions, while those with odd $k+n$ are odd functions.

Theorem 3.2 ([43, Thm. 3.3]). *The vertical slice transform* (16) *fulfills*

$$\mathcal{V}Y_n^k(\psi, t) = \mathrm{v}_n^k \sqrt{\tfrac{2n+1}{4\pi}}\, \mathrm{e}^{\mathrm{i}k\psi}\, P_n(t), \qquad n \in \mathbb{N}_0,\ k \in \{-n, \dots, n\},\ n+k \textit{ even},$$

where

$$\mathrm{v}_n^k \coloneqq (-1)^{\frac{n+k}{2}} \sqrt{\frac{(n-k)!}{(n+k)!} \frac{(n+k-1)!!}{(n-k)!!}}.$$

There exist constants $C_1, C_2 > 0$ such that for all $n \in \mathbb{N}_0$, $k \in \{-n, \dots, n\}$ with $n+k$ even,

$$C_1(n+1/2)^{-1/2} \le |\mathrm{v}_n^k| \le C_2(n+1/2)^{-1/4}. \tag{19}$$

Noting that the functions

$$B_n^k(\psi, t) \coloneqq \sqrt{\frac{2n+1}{4\pi}}\, P_n(t)\, \mathrm{e}^{\mathrm{i}k\psi}, \qquad \forall (\psi, t) \in \mathbb{T} \times \mathbb{I}, \tag{20}$$

form an orthonormal basis of $L^2(\mathbb{T} \times \mathbb{I})$ and that $\mathrm{v}_n^k \to 0$ as $n \to \infty$, we deduce that $\mathcal{V}\colon L^2(\mathbb{S}^2) \to L^2(\mathbb{T} \times \mathbb{I})$ is a compact operator with singular value decomposition

$$\mathcal{V}f(\psi, t) = \sum_{n \in \mathbb{N}_0} \sum_{\substack{k=-n \\ n+k\,\text{even}}}^{n} \mathrm{v}_n^k\, B_n^k(\psi, t). \tag{21}$$

Restricting $\mathcal{V}$ to even functions $L^2_{\text{sym}}(\mathbb{S}^2)$, where $L^p_{\text{sym}}(\mathbb{S}^2)$ with $1 \le p \le \infty$ is defined as

$$L^p_{\text{sym}}(\mathbb{S}^2) \coloneqq \left\{ f \in L^p(\mathbb{S}^2) : f(\xi) = \check{f}(\xi) \text{ a.e. on } \mathbb{S}^2 \right\}$$

and $\check{f}(\boldsymbol{\xi}) \coloneqq f(\xi_1, \xi_2, -\xi_3)$ is the reflection at the ξ_1-ξ_2 plane, the operator $\mathcal{V}\colon L^2_{\text{sym}}(\mathbb{S}^2) \to L^2(\mathbb{T} \times \mathbb{I})$ is injective. Its *Moore–Penrose pseudoinverse*, cf. [25], is given by

$$\mathcal{V}^\dagger\colon \mathcal{R}(\mathcal{V}) \oplus \mathcal{R}(\mathcal{V})^\perp \to L^2_{\text{sym}}(\mathbb{S}^2), \qquad \mathcal{V}^\dagger g = \sum_{n \in \mathbb{N}_0} \sum_{\substack{k=-n \\ n+k\,\text{even}}}^{n} (\mathrm{v}_n^k)^{-1} \langle g, B_n^k \rangle\, Y_n^k, \tag{22}$$

where $\mathcal{R}(\mathcal{V})$ denotes the range of $\mathcal{V}$. We will further need the adjoint operator of $\mathcal{V}$.

Proposition 3.3. *Let $1 \le p, q \le \infty$ with $1/p + 1/q = 1$. For $1 \le p < \infty$, the adjoint $\mathcal{V}^*\colon L^q(\mathbb{T} \times \mathbb{I}) \to L^q(\mathbb{S}^2)$ of $\mathcal{V}\colon L^p(\mathbb{S}^2) \to L^p(\mathbb{T} \times \mathbb{I})$ is given by*

$$\mathcal{V}^* g(\boldsymbol{\xi}) = \frac{1}{2\pi} \int_{\mathbb{T}} g(\psi, \xi_1 \cos\psi + \xi_2 \sin\psi)\, \mathrm{d}\psi, \tag{23}$$

and the adjoint $\mathcal{V}_\psi^*\colon L^q(\mathbb{I}) \to L^q(\mathbb{S}^2)$ *of* $\mathcal{V}_\psi\colon L^p(\mathbb{S}^2) \to L^p(\mathbb{I})$ *by*

$$\mathcal{V}_\psi^* g(\boldsymbol{\xi}) = g(\xi_1 \cos\psi + \xi_2 \sin\psi). \tag{24}$$

Moreover, it holds $\mathcal{V}^*\colon C(\mathbb{T}\times\mathbb{I}) \to C(\mathbb{S}^2)$ *and* $\mathcal{V}_\psi^*\colon C(\mathbb{I}) \to C(\mathbb{S}^2)$.

Proof. The assertion follows from Proposition 3.1, which yields

$$\begin{aligned}\langle \mathcal{V}f, g\rangle &= \int_{\mathbb{T}}\int_{\mathbb{I}} \mathcal{V}f(\psi,t)\, g(\psi,t)\;\mathrm{d}t\;\mathrm{d}\psi = \int_{\mathbb{T}}\int_{\mathbb{I}} \frac{1}{2\pi\sqrt{1-t^2}} \int_{C_\psi^t} f(\boldsymbol{\xi})\, g(\psi,t)\,\mathrm{ds}(\boldsymbol{\xi})\,\mathrm{d}t\,\mathrm{d}\psi \\ &= \frac{1}{2\pi}\int_{\mathbb{T}}\int_{\mathbb{S}^2} f(\boldsymbol{\xi})\, g(\psi, \xi_1\cos\psi + \xi_2\sin\psi)\,\mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi})\,\mathrm{d}\psi = \langle f, \mathcal{V}^* g\rangle\end{aligned}$$

for all $f \in L^p(\mathbb{S}^2)$, $g \in L^q(\mathbb{T}\times\mathbb{I})$. The adjoint of $\mathcal{V}_\psi$ can be established analogously—without the integral over $\mathbb{T}$ and the factor $(2\pi)^{-1}$. The last assertion again follows from Lebesgue's dominated convergence theorem and by the definition of the adjoint. □

3.2. Vertical Slice Transform of Measures

For functions $f\colon \mathbb{S}^2 \to \mathbb{R}$, the vertical slice transform $\mathcal{V}f(\psi,t)$ in (16) and its restriction $\mathcal{V}_\psi f(t)$ in (17) are integrals of f along the slices $\mathcal{S}_\psi^{-1}(t)$. Heuristically, the related concept for measures $\mu \in \mathcal{M}(\mathbb{S}^2)$ would be to consider $\mu(\mathcal{S}_\psi^{-1}(t))$. In this manner, for a fixed angle $\psi \in \mathbb{T}$, we generalize the (restricted) *vertical slice transform* $\mathcal{V}_\psi$ by

$$\mathcal{V}_\psi\colon \mathcal{M}(\mathbb{S}^2) \to \mathcal{M}(\mathbb{I}), \quad \mu \mapsto (\mathcal{S}_\psi)_\# \mu = \mu \circ \mathcal{S}_\psi^{-1}. \tag{25}$$

In the function setting, we figuratively obtain $\mathcal{V}f$ by gluing the (rescaled) functions $\frac{1}{2\pi}\mathcal{V}_\psi f$ together along the angle ψ. In the measure setting, the corresponding concept is to consider $\mathcal{V}_\psi$ as disintegration family. We define the *vertical slice transform* $\mathcal{V}\colon \mathcal{M}(\mathbb{S}^2) \to \mathcal{M}(\mathbb{T}\times\mathbb{I})$ by

$$\mathcal{V}\mu \coloneqq (T_{\mathcal{V}})_\#(u_{\mathbb{T}} \times \mu) \quad \text{with} \quad T_{\mathcal{V}}(\psi, \boldsymbol{\xi}) \coloneqq (\psi, \mathcal{S}_\psi(\boldsymbol{\xi})). \tag{26}$$

The disintegration aspect becomes clear in the following proposition.

Proposition 3.4. *Let* $\mu \in \mathcal{M}(\mathbb{S}^2)$. *Then* $\mathcal{V}\mu$ *can be disintegrated into the family* $\mathcal{V}_\psi\mu$ *with respect to the uniform measure* $u_{\mathbb{T}}$, *i.e., for all* $g \in C(\mathbb{T}\times\mathbb{I})$, *it holds*

$$\int_{\mathbb{T}\times\mathbb{I}} g(\psi,t)\,\mathrm{d}\mathcal{V}\mu(\psi,t) = \int_{\mathbb{T}}\int_{\mathbb{I}} g(\psi,t)\,\mathrm{d}\mathcal{V}_\psi\mu(t)\,\mathrm{d}u_{\mathbb{T}}(\psi).$$

Proof. Incorporating (26), and using Fubini's theorem, we obtain

$$\langle \mathcal{V}\mu, g\rangle = \int_{\mathbb{T}}\int_{\mathbb{S}^2} g(\psi, \mathcal{S}_\psi(\boldsymbol{\xi}))\,\mathrm{d}\mu(\boldsymbol{\xi})\,\mathrm{d}u_{\mathbb{T}}(\psi) = \int_{\mathbb{T}}\int_{\mathbb{I}} g(\psi,t)\,\mathrm{d}((\mathcal{S}_\psi)_\#\mu)(t)\,\mathrm{d}u_{\mathbb{T}}(\psi)$$

for every $g \in C(\mathbb{T}\times\mathbb{I})$. By (25) this implies the assertion. □

The defined measure-valued versions of $\mathcal{V}$ and $\mathcal{V}_\psi$ are in fact the adjoints of $\mathcal{V}^*\colon C(\mathbb{T}\times\mathbb{I}) \to C(\mathbb{S}^2)$ in (23) and $\mathcal{V}_\psi^*\colon C(\mathbb{I}) \to C(\mathbb{S}^2)$ in (24), which explains the generalizations from the duality point of view.

Proposition 3.5. *The vertical slice transforms* (26) *and* (25) *satisfy*

$$\begin{aligned} \langle \mathcal{V}\mu, g\rangle &= \langle \mu, \mathcal{V}^* g\rangle \quad \text{for all } g \in C(\mathbb{T}\times\mathbb{I}) \quad \text{and} \\ \langle \mathcal{V}_\psi\mu, g\rangle &= \langle \mu, \mathcal{V}_\psi^* g\rangle \quad \text{for all } g \in C(\mathbb{I}),\ \psi\in\mathbb{T} \end{aligned} \tag{27}$$

with the adjoint operators from (23) *and* (24).

Proof. For $\mu \in \mathcal{M}(\mathbb{S}^2)$ and $g \in C(\mathbb{T}\times\mathbb{I})$, the conjecture can be established by

$$\langle \mathcal{V}\mu, g\rangle = \int_{\mathbb{T}\times\mathbb{I}} g(\psi,t)\, \mathrm{d}(T_{\mathcal{V}})_\#(u_{\mathbb{T}}\times\mu)(\psi,t) = \int_{\mathbb{S}^2}\int_{\mathbb{T}} g(\psi,\mathcal{S}_\psi(\boldsymbol{\xi}))\,\mathrm{d}u_{\mathbb{T}}(\psi)\,\mathrm{d}\mu(\boldsymbol{\xi}) = \langle \mu, \mathcal{V}^* g\rangle$$

and, for $\mu\in\mathcal{M}(\mathbb{S}^2)$, $g \in C(\mathbb{I})$, and fixed $\psi\in\mathbb{T}$, by

$$\langle \mathcal{V}_\psi\mu, g\rangle = \int_{\mathbb{I}} g(t)\,\mathrm{d}(\mathcal{S}_\psi)_\#\mu(t) = \int_{\mathbb{S}^2} g(\mathcal{S}_\psi(\boldsymbol{\xi}))\,\mathrm{d}\mu(\boldsymbol{\xi}) = \langle \mu, \mathcal{V}_\psi^* g\rangle. \qquad \square$$

One could equivalently use the identity (27) to define the vertical slice transform of a measure, analogously as it was done for the Radon transform in [39, Chap. 2, § 2]. For absolutely continuous measures with respect to $\sigma_{\mathbb{S}^2}$, the measure- and function-valued vertical slice transforms coincide, which now justify the different scalings in (16) and (17).

Proposition 3.6. *For $f \in L^1(\mathbb{S}^2)$, the vertical slice transforms satisfy*

$$\mathcal{V}[f\sigma_{\mathbb{S}^2}] = (\mathcal{V}f)\,\sigma_{\mathbb{T}\times\mathbb{I}} \quad \text{and} \quad \mathcal{V}_\psi[f\sigma_{\mathbb{S}^2}] = (\mathcal{V}_\psi f)\,\sigma_{\mathbb{I}}.$$

In particular, the transformed measures are again absolutely continuous.

Proof. Let $\langle\cdot,\cdot\rangle_{\mathcal{M}}$ denotes the dual pairing for measures and continuous function and $\langle\cdot,\cdot\rangle_L$ the dual pairing between L^1 and L^∞ functions. Then the identity follows directly from Proposition 3.5 by

$$\langle \mathcal{V}[f\sigma_{\mathbb{S}^2}], g\rangle_{\mathcal{M}} = \langle f\sigma_{\mathbb{S}^2}, \mathcal{V}^* g\rangle_{\mathcal{M}} = \langle f, \mathcal{V}^* g\rangle_L = \langle \mathcal{V}f, g\rangle_L = \langle (\mathcal{V}f)\,\sigma_{\mathbb{T}\times\mathbb{I}}, g\rangle_{\mathcal{M}}$$

for all $g \in C(\mathbb{T}\times\mathbb{I})$. For $\mathcal{V}_\psi$, the identity follows analogously. $\square$

By the following theorem, we see that similarly to the function setting, the vertical slice transform is injective when restricted to the even measures (with respect to the ξ_1-ξ_2 plane) given by

$$\mathcal{M}_{\mathrm{sym}}(\mathbb{S}^2) \coloneqq \{\mu\in\mathcal{M}(\mathbb{S}^2) : \langle\mu, f\rangle = \langle \mu, \check{f}\rangle \text{ for all } f\in C(\mathbb{S}^2)\}. \tag{28}$$

Theorem 3.7. *The vertical slice transform $\mathcal{V}\colon \mathcal{M}_{\mathrm{sym}}(\mathbb{S}^2) \to \mathcal{M}(\mathbb{T}\times\mathbb{I})$ is injective.*

The proof is given in Appendix A.

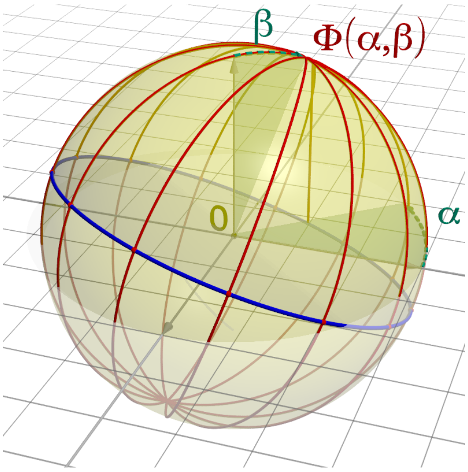


Figure 2: Semicircles $M^\gamma_{\alpha,\beta}$ (red) starting at a fixed point $\Phi(\alpha,\beta)$ and with varying $\gamma \in \mathbb{T}$. Here β is the angle of $\Phi(\alpha,\beta)$ to the north pole and α the angle of its projection in the ξ_1-ξ_2 plane to the ξ_1 axis. The blue circle is orthogonal to the semicircles.

4. Normalized Semicircle Transform

4.1. Normalized Semicircle Transform of Functions

Instead of integrating over parallel slices, the semicircle transform integrates a function along all meridians with respect to a fixed zenith on the sphere. For any *zenith* $\Phi(\alpha,\beta) \in \mathbb{S}^2$ with $\alpha \in \mathbb{T}$ and $\beta \in [0,\pi]$, we define the *azimuth operator* $\mathcal{A}_{\alpha,\beta} \colon \mathbb{S}^2 \to \mathbb{T}$ and the *zenith operator* $\mathcal{Z}_{\alpha,\beta} \colon \mathbb{S}^2 \to [0,\pi]$ as

$$\mathcal{A}_{\alpha,\beta}(\boldsymbol{\xi}) \coloneqq \operatorname{azi}(\Psi(\alpha,\beta,0)^\top \boldsymbol{\xi}),$$
$$\mathcal{Z}_{\alpha,\beta}(\boldsymbol{\xi}) \coloneqq \operatorname{zen}(\Psi(\alpha,\beta,0)^\top \boldsymbol{\xi}),$$

i.e., we rotate the zenith back to the north pole and take the azimuth and zenith angle, see Figure 2. For the zenith $\Phi(\alpha,\beta)$ and fixed $\gamma \in \mathbb{T}$, we consider the *semicircles/meridians*

$$M^\gamma_{\alpha,\beta} \coloneqq \mathcal{A}^{-1}_{\alpha,\beta}(\gamma) = \{\boldsymbol{\xi} \in \mathbb{S}^2 : \mathcal{A}_{\alpha,\beta}(\boldsymbol{\xi}) = \gamma\}.$$

If $\gamma \neq 0$, we have

$$M^\gamma_{\alpha,\beta} = \{\boldsymbol{\xi} = \Psi(\alpha,\beta,0)\,\Phi(\gamma,\vartheta) = \Psi(\alpha,\beta,\gamma)\,\Phi(0,\vartheta) : \vartheta \in (0,\pi)\}.$$

Otherwise, if $\gamma = 0$, we need to replace the open interval by a closed one, i.e., $\vartheta \in [0,\pi]$. Figuratively, $M^\gamma_{\alpha,\beta}$ is a rotation of the meridian $\{\Phi(\gamma,\vartheta) : \vartheta \in (0,\pi)\}$ with azimuth γ by $\Psi(\alpha,\beta,0)$. The *normalized semicircle transform* $\mathcal{W}$ of $f \colon \mathbb{S}^2 \to \mathbb{R}$ is defined by

$$\mathcal{W}f(\alpha,\beta,\gamma) \coloneqq \frac{1}{4\pi} \int_{M^\gamma_{\alpha,\beta}} f(\boldsymbol{\xi}) \sin\left(\mathcal{Z}_{\alpha,\beta}(\boldsymbol{\xi})\right) \mathrm{d}s(\boldsymbol{\xi})$$
$$= \frac{1}{4\pi} \int_0^\pi f(\Psi(\alpha,\beta,0)\,\Phi(\gamma,\vartheta)) \sin(\vartheta) \mathrm{d}\vartheta.$$

We may interpret $M_{\alpha,\beta}^{\gamma}$ as rotation of the prime median by $\Psi(\alpha, \beta, \gamma)$. Based on the substitution $\boldsymbol{Q} = \Psi(\alpha, \beta, \gamma)$, the normalized semicircle transform defines a function on SO(3) via

$$\mathcal{W} f(\boldsymbol{Q}) \coloneqq \frac{1}{4\pi} \int_0^{\pi} f\left(\boldsymbol{Q}\, \Phi(0, \vartheta)\right) \sin(\vartheta)\, \mathrm{d}\vartheta. \tag{29}$$

Henceforth, we will not distinguish between $\mathcal{W} f(\alpha, \beta, \gamma)$ and $\mathcal{W} f(\boldsymbol{Q})$. Especially for the inversion formula by the singular value decomposition, we will make use of the latter definition. The multiplication with $(4\pi)^{-1} \sin(\vartheta)$ in the latitude ensures that density functions are mapped to density functions allowing the later generalization to measures. For the zenith $\Phi(\alpha, \beta)$, we define the (normalized) restriction

$$\mathcal{W}_{\alpha,\beta} f \coloneqq 4\pi\, \mathcal{W} f(\alpha, \beta, \cdot).$$

Remark 4.1. The *(unnormalized) semicircle transform* $\widetilde{\mathcal{W}} \colon C(\mathbb{S}^2) \to C(\mathrm{SO}(3))$ is defined by

$$\widetilde{\mathcal{W}} f(\boldsymbol{Q}) \coloneqq \int_{-\pi/2}^{\pi/2} f\left(\boldsymbol{Q}^{\top} (\Phi(\varphi, \tfrac{\pi}{2}))\right) \mathrm{d}\varphi, \qquad \boldsymbol{Q} \in \mathrm{SO}(3),$$

see [41]. It computes the mean values of f along all half great circles of the sphere, i.e., without the weight $\sin(\vartheta)$ of (29). The injectivity of $\widetilde{\mathcal{W}}$ was shown in [36]. A singular value decomposition and inversion algorithms were provided in [41]. The authors of [15] reinvented this transform with another parameterization using the plane through $\Psi(\alpha, \beta, 0)\boldsymbol{e}^1$ and $\Psi(\alpha, \beta, 0)\boldsymbol{e}^2$. More precisely, their notation was not clear to us since it seems that they have applied the normalized transform in the numerical examples, but certain parts in their analysis rely on the unnormalized transform.

The semicircle transforms $\mathcal{W}$ and $\mathcal{W}_{\alpha,\beta}$ are well defined for continuous functions as well as for p-integrable functions. Moreover, both transforms are continuous operators.

Proposition 4.2. *Let $1 \le p \le \infty$, and let $\Phi(\alpha, \beta) \in \mathbb{S}^2$. For every $f \in L^p(\mathbb{S}^2)$, it holds*

$$\int_{\mathbb{T}} \mathcal{W}_{\alpha,\beta} f(\gamma)\, \mathrm{d}\gamma = \int_{\mathbb{S}^2} f(\boldsymbol{\xi})\, \mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}) \quad \text{and} \quad \int_{\mathrm{SO}(3)} \mathcal{W} f(\boldsymbol{Q})\, \mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{Q}) = \int_{\mathbb{S}^2} f(\boldsymbol{\xi})\, \mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}).$$

The operators $\mathcal{W}_{\alpha,\beta} \colon L^p(\mathbb{S}^2) \to L^p(\mathbb{T})$ and $\mathcal{W} \colon L^p(\mathbb{S}^2) \to L^p(\mathrm{SO}(3))$ are bounded with

$$\|\mathcal{W}_{\alpha,\beta}\|_{L^p \to L^p} \le 2^{1-1/p} \quad \text{and} \quad \|\mathcal{W}\|_{L^p \to L^p} \le (2\pi)^{1/p-1}.$$

Moreover, it holds $\mathcal{W}_{\alpha,\beta} \colon C(\mathbb{S}^2) \to C(\mathbb{T})$ and $\mathcal{W} \colon C(\mathbb{S}^2) \to C(\mathrm{SO}(3))$.

Proof. Since the surface measure on $\mathbb{S}^2$ is invariant under rotations, we have

$$\begin{aligned} \int_{\mathbb{T}} \mathcal{W}_{\alpha,\beta} f(\gamma)\, \mathrm{d}\gamma &= \int_{\mathbb{T}} \int_0^{\pi} f(\Psi(\alpha, \beta, 0)\ \Phi(\gamma, \vartheta)) \sin(\vartheta)\, \mathrm{d}\vartheta\, \mathrm{d}\gamma \\ &= \int_{\mathbb{S}^2} f(\Psi(\alpha, \beta, 0)\, \boldsymbol{\xi})\, \mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}) = \int_{\mathbb{S}^2} f(\boldsymbol{\xi})\, \mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}). \end{aligned}$$

The definition of $\mathcal{W}_{\alpha,\beta}$ and integration over α and β gives the second identity. Thus $\mathcal{W}_{\alpha,\beta}$ and $\mathcal{W}$ are well defined almost everywhere by Fubini's theorem. Using absolute values and the triangle inequality in the above computation yields $\|\mathcal{W}_{\alpha,\beta}\|_{L^1\to L^1} = \|\mathcal{W}\|_{L^1\to L^1} = 1$, where we deduce a lower bound of the norm by inserting the constant $f = 1$. The semicircle transform is further essentially bounded by

$$|\mathcal{W}_{\alpha,\beta} f(\gamma)| \leq \int_0^\pi |f(\Psi(\alpha,\beta,0)\,\Phi(\gamma,\vartheta))| \sin(\vartheta)\mathrm{d}\vartheta \leq 2 \operatorname*{ess\,sup}_{\boldsymbol{\xi}\in\mathbb{S}^2} |f(\boldsymbol{\xi})|;$$

so $\|\mathcal{W}_{\alpha,\beta}\|_{L^\infty\to L^\infty} = 2$ and $\|\mathcal{W}\|_{L^\infty\to L^\infty} = (2\pi)^{-1}$. The second assertion now follows from the Riesz–Thorin interpolation theorem. The last assertion is an immediate consequence of Lebesgue's dominated convergence theorem. □

Considering the semicircle transform in the Hilbert space setting, i.e., $\mathcal{W}\colon L^2(\mathbb{S}^2) \to L^2(\mathrm{SO}(3))$, we are interested in its singular value decomposition.

Theorem 4.3. *The normalized semicircle transform fulfills*

$$\mathcal{W} Y_n^k = \mathrm{w}_n Z_n^k, \quad n \in \mathbb{N}_0,\ k \in \{-n,\dots,n\}, \tag{30}$$

with the singular values $\mathrm{w}_n \coloneqq \|\mathcal{W} Y_n^k\|_{L^2(\mathbb{S}^2)}$ *and the orthonormal functions*

$$Z_n^k \coloneqq \mathrm{w}_n^{-1} \sum_{j=-n}^{n} \lambda_n^j\, \overline{D_n^{k,j}} \in L^2(\mathrm{SO}(3)), \tag{31}$$

where $\lambda_0^0 \coloneqq 2(4\pi)^{-3/2}$ *and, for* $n \in \mathbb{N}$ *and* $j \in \{1,\dots,n\}$ *with* $n+j$ *even,*

$$\lambda_n^j \coloneqq \frac{(-1)^j}{4\pi} \sqrt{\frac{2n+1}{4\pi} \frac{(n-j)!}{(n+j)!}}\, \frac{j\,(n-2)!!\,(n+j-1)!!}{(n-j)!!\,(n+1)!!} \begin{cases} 2: & n \text{ even}, \\ \pi: & n \text{ odd}, \end{cases} \tag{32}$$

$\lambda_n^{-j} \coloneqq (-1)^j \lambda_n^j$, *and* $\lambda_n^j = 0$ *otherwise. Here* Y_n^k *denote the spherical harmonics* (7) *and* $D_n^{k,j}$ *the rotational harmonics* (13)*. Moreover, there are constants* $C_1, C_2 > 0$ *such that*

$$C_1\ (n+1)^{-1/2} \leq \mathrm{w}_n \leq C_2\ (n+1)^{-1/2} \quad \text{for all } n \in \mathbb{N}_0. \tag{33}$$

The proof is given in Appendix B. Analogously to [67, Thm. 3.13], we see that the semicircle transform is a smoothing operator.

Corollary 4.4. *For* $s \geq 0$*, the operator* $\mathcal{W}\colon H^s(\mathbb{S}^2) \to H^{s+1/2}(\mathrm{SO}(3))$ *is continuous.*

Theorem 4.3 implies that $\mathcal{W}\colon L^2(\mathbb{S}^2) \to L^2(\mathrm{SO}(3))$ is an injective, compact operator with

$$\mathcal{W} f = \sum_{n\in\mathbb{N}_0} \mathrm{w}_n \sum_{k=-n}^{n} \langle f, Y_n^k\rangle Z_n^k. \tag{34}$$

Its *Moore–Penrose pseudoinverse* is given by

$$\mathcal{W}^\dagger g = \sum_{n=0}^{\infty} \frac{1}{\mathrm{w}_n} \sum_{k=-n}^{n} \left\langle g, Z_n^k \right\rangle Y_n^k = \sum_{n=0}^{\infty} \frac{1}{(\mathrm{w}_n)^2} \sum_{k=-n}^{n} \sum_{j=-n}^{n} \lambda_n^j \left\langle g, \overline{D_n^{k,j}} \right\rangle Y_n^k. \tag{35}$$

We will also need the adjoint operator.

Proposition 4.5. *Let $1 \le p, q \le \infty$ with $1/p + 1/q = 1$. For $1 \le p < \infty$, the adjoint $\mathcal{W}^* \colon L^q(\mathrm{SO}(3)) \to L^q(\mathbb{S}^2)$ of $\mathcal{W} \colon L^p(\mathbb{S}^2) \to L^p(\mathrm{SO}(3))$ is given by*

$$\mathcal{W}^* g(\boldsymbol{\xi}) = \frac{1}{4\pi} \int_{\mathbb{T}} \int_0^\pi g(\Psi(\alpha, \beta, \mathcal{A}_{\alpha,\beta}\, \boldsymbol{\xi}))\, \sin(\beta)\, \mathrm{d}\beta\, \mathrm{d}\alpha, \tag{36}$$

and the adjoint $\mathcal{W}^*_{\alpha,\beta} \colon L^q(\mathbb{T}) \to L^q(\mathbb{S}^2)$ of $\mathcal{W}_{\alpha,\beta} \colon L^p(\mathbb{S}^2) \to L^p(\mathbb{T})$ by

$$\mathcal{W}^*_{\alpha,\beta} g(\boldsymbol{\xi}) = g(\mathcal{A}_{\alpha,\beta}\, \boldsymbol{\xi}). \tag{37}$$

Moreover, it holds $\mathcal{W}^* \colon C(\mathrm{SO}(3)) \to C(\mathbb{S}^2)$, but $\mathcal{W}^*_{\alpha,\beta} \colon C(\mathbb{T}) \not\to C(\mathbb{S}^2)$.

Proof. Let $f \in L^p(\mathbb{S}^2)$ and $g \in L^q(\mathrm{SO}(3))$. Based on (12) and (29), and using the substitution $\boldsymbol{\eta} \coloneqq \Phi(\gamma, \vartheta)$ with $\gamma = \mathrm{azi}(\boldsymbol{\eta})$, we compute the adjoint by

$$\begin{aligned}
\langle \mathcal{W} f, g \rangle &= \frac{1}{4\pi} \int_{\mathbb{T}} \int_0^\pi \int_{\mathbb{T}} \int_0^\pi f(\Psi(\alpha, \beta, 0)\, \Phi(\gamma, \vartheta))\, g(\Psi(\alpha, \beta, \gamma))\, \sin(\vartheta) \sin(\beta)\, \mathrm{d}\vartheta\, \mathrm{d}\alpha\, \mathrm{d}\beta\, \mathrm{d}\gamma \\
&= \frac{1}{4\pi} \int_{\mathbb{T}} \int_0^\pi \int_{\mathbb{S}^2} f(\Psi(\alpha, \beta, 0)\, \boldsymbol{\eta})\, g(\Psi(\alpha, \beta, \mathrm{azi}(\boldsymbol{\eta}))\sin(\beta)\, \mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\eta})\, \mathrm{d}\beta\, \mathrm{d}\alpha \\
&= \frac{1}{4\pi} \int_{\mathbb{S}^2} \int_{\mathbb{T}} \int_0^\pi f(\boldsymbol{\xi})\, g(\Psi(\alpha, \beta, \mathcal{A}_{\alpha,\beta}(\boldsymbol{\xi}))\sin(\beta)\, \mathrm{d}\beta\, \mathrm{d}\alpha\, \mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}) = \langle f, \mathcal{W}^* g \rangle,
\end{aligned}$$

where we used $\boldsymbol{\xi} \coloneqq \Psi(\alpha, \beta, 0)\, \boldsymbol{\eta}$ in the last line. The adjoint of $\mathcal{W}_{\alpha,\beta}$ follows analogously. The continuity of $\mathcal{W}^* g$ for $g \in C(\mathrm{SO}(3))$ can be established by Lebesgue's dominated convergence theorem. For non-constant $g \in C(\mathbb{T})$, the adjoint $\mathcal{W}^*_{\alpha,\beta}$ is discontinuous at $\Phi(\alpha, \beta)$. □

4.2. Normalized Semicircle Transform of Measures

The generalization from functions to measures can be done analogously to Section 3.2. For the zenith $\Phi(\alpha, \beta)$, we generalize the (restricted) *semicircle transform* $\mathcal{W}_{\alpha,\beta}$ by

$$\mathcal{W}_{\alpha,\beta} \colon \mathcal{M}(\mathbb{S}^2) \to \mathcal{M}(\mathbb{T}), \quad \mu \mapsto (\mathcal{A}_{\alpha,\beta})_\# \mu = \mu \circ \mathcal{A}^{-1}_{\alpha,\beta}. \tag{38}$$

Considering the measures $\mathcal{W}_{\alpha,\beta}\, \mu$ as disintegration family, we define the *(normalized) semicircle transform* $\mathcal{W} \colon \mathcal{M}(\mathbb{S}^2) \to \mathcal{M}(\mathrm{SO}(3))$ by

$$\mathcal{W} \mu \coloneqq (T_{\mathcal{W}})_\# (u_{\mathbb{S}^2} \times \mu) \quad \text{with} \quad T_{\mathcal{W}}(\Phi(\alpha, \beta), \boldsymbol{\xi}) \coloneqq \Psi(\alpha, \beta, \mathcal{A}_{\alpha,\beta}(\boldsymbol{\xi})). \tag{39}$$

Proposition 4.6. *Let $\mu \in \mathcal{M}(\mathbb{S}^2)$. Then $\mathcal{W}\mu$ can be disintegrated into the family $\mathcal{W}_{\alpha,\beta}\, \mu$ with respect to the uniform measure $u_{\mathbb{S}^2}$, i.e., for all $g \in C(\mathrm{SO}(3))$, it holds*

$$\int_{\mathrm{SO}(3)} g(\boldsymbol{Q})\, \mathrm{d}(\mathcal{W}\mu)(\boldsymbol{Q}) = \int_{\mathbb{S}^2} \int_{\mathbb{T}} g(\Psi(\alpha, \beta, \gamma))\, \mathrm{d}(\mathcal{W}_{\alpha,\beta}\mu)(\gamma)\, \mathrm{d}\sigma_{\mathbb{S}^2}(\Phi(\alpha, \beta)).$$

Proof. Inserting (39) and using Fubini's theorem, we obtain

$$\begin{aligned}
\langle \mathcal{W}\mu, g \rangle &= \int_{\mathbb{S}^2} \int_{\mathbb{S}^2} g(\Psi(\alpha, \beta, \mathcal{A}_{\alpha,\beta}\, \boldsymbol{\xi}))\, \mathrm{d}\mu(\boldsymbol{\xi})\, \mathrm{d}u_{\mathbb{S}^2}(\Phi(\alpha, \beta)) \\
&= \int_{\mathbb{S}^2} \int_{\mathbb{T}} g(\Psi(\alpha, \beta, \gamma))\, \mathrm{d}(\mathcal{A}_{\alpha,\beta})_\# \mu(\gamma)\, \mathrm{d}u_{\mathbb{S}^2}(\Phi(\alpha, \beta)).
\end{aligned}$$

for every $g \in C(\mathrm{SO}(3))$ establishing the assertion. □

While $\mathcal{W}$ can be interpreted as the adjoint of $\mathcal{W}^*\colon C(\mathrm{SO}(3)) \to C(\mathbb{S}^2)$ in (36), the same reasoning does not hold for $\mathcal{W}_{\alpha,\beta}$ by the lack of continuity of $\mathcal{W}^*_{\alpha,\beta} g$ for continuous g.

Proposition 4.7. *The semicircle transforms* (39) *and* (38) *satisfy*

$$\begin{aligned} \langle \mathcal{W}\mu, g\rangle &= \langle \mu, \mathcal{W}^* g\rangle \quad \text{for all } g \in C(\mathrm{SO}(3)) \text{ and} \\ \langle \mathcal{W}_{\alpha,\beta}\mu, g\rangle &= \int_{\mathbb{S}^2} \mathcal{W}^*_{\alpha,\beta}\, g(\boldsymbol{\xi})\,\mathrm{d}\mu(\boldsymbol{\xi}) \quad \text{for all } g \in C(\mathbb{T}) \end{aligned} \tag{40}$$

with the adjoint operator from (36) *and* (37).

Proof. Plugging in the push-forward definition (39), we obtain

$$\langle \mathcal{W}\mu, g\rangle = \int_{\mathbb{S}^2}\int_{\mathbb{S}^2} g(\Psi(\alpha,\beta,\mathcal{A}_{\alpha,\beta}\,\boldsymbol{\xi}))\,\mathrm{d}\mu(\boldsymbol{\xi})\,\mathrm{d}u_{\mathbb{S}^2}(\Phi(\alpha,\beta)) = \langle \mu, \mathcal{W}^* g\rangle\,.$$

The second identity follows analogously. □

Remark 4.8. The restricted semicircle transform $\mathcal{W}_{\alpha,\beta} : \mathcal{M}(\mathbb{S}^2) \to \mathcal{M}(\mathbb{T})$ is indeed related to the L^p adjoint $W^*_{\alpha,\beta}$ in (37). Although the integral on the right-hand side of (40) is always well defined, i.e., $W^*_{\alpha,\beta}\, g \in L^\infty(\mathbb{S}^2)$, the integral is no dual pairing in the measure/continuous function sense since $W^*_{\alpha,\beta}\, g$ is discontinuous at the zenith $\Phi(\alpha,\beta)$ in general. Therefore, unlike in Proposition 3.5 for the vertical slice transform, equation (40) does not constitute a proper definition of $W_{\alpha,\beta}$ for measures via the dual pairing.

The definitions of the semicircle transform in the function and measure setting are consistent in the sense that both coincide for absolutely continuous measures.

Proposition 4.9. *For $f \in L^1(\mathbb{S}^2)$, the semicircle transforms satisfy*

$$\mathcal{W}[f\sigma_{\mathbb{S}^2}] = (\mathcal{W}f)\,\sigma_{\mathrm{SO}(3)} \quad \text{and} \quad \mathcal{W}_{\alpha,\beta}[f\sigma_{\mathbb{S}^2}] = (\mathcal{W}_{\alpha,\beta} f)\,\sigma_{\mathbb{T}}.$$

In particular, the transformed measures are again absolutely continuous.

Proof. Both identities directly follow from Proposition 4.7 in analogy to the proof of Proposition 3.6. For $\mathcal{W}_{\alpha,\beta}$ the second dual pairing has to be replaced by an integral. □

By the following theorem, the injectivity of $\mathcal{W}\colon L^2(\mathbb{S}^2) \to L^2(\mathrm{SO}(3))$ generalizes to $\mathcal{W}\colon \mathcal{M}(\mathbb{S}^2) \to \mathcal{M}(\mathrm{SO}(3))$, which also implies the injectivity of $\mathcal{W}\colon L^p(\mathbb{S}^2) \to L^p(\mathrm{SO}(3))$.

Theorem 4.10. *The semicircle transform $\mathcal{W}\colon \mathcal{M}(\mathbb{S}^2) \to \mathcal{M}(\mathrm{SO}(3))$ defined by* (39) *is injective.*

The proof is given in Appendix C.

5. Spherical Sliced Wasserstein Distances

The computation of the Wasserstein distance on the sphere consists in determining a transport plan between the considered probability measures. To avoid the occurring optimization problem, the general idea behind so-called sliced Wasserstein distances [15, 47] is to transform the measures first to one-dimensional domains, and to exploit the explicit solution formula of the one-dimensional transport. Based on the vertical slice and the normalized semicircle transform, we can define two kinds of spherical sliced distances. For $p \in [1, \infty)$ and $\mu, \nu \in \mathcal{M}(\mathbb{S}^2)$, we define the *vertical sliced Wasserstein distance*

$$\mathrm{VSW}_p^p(\mu, \nu) \coloneqq \int_{\mathbb{T}} \mathrm{W}_p^p(\mathcal{V}_\psi \mu, \mathcal{V}_\psi \nu) \,\mathrm{d}\psi,$$

and the *semicircular sliced Wasserstein distance*

$$\mathrm{SSW}_p^p(\mu, \nu) \coloneqq \int_{\mathbb{S}^2} \mathrm{W}_p^p(\mathcal{W}_{\alpha,\beta} \mu, \mathcal{W}_{\alpha,\beta} \nu) \,\mathrm{d}\sigma_{\mathbb{S}^2}(\Phi(\alpha, \beta)),$$

which are integrals over Wasserstein distances on $\mathbb{I}$ and $\mathbb{T}$, respectively.

Theorem 5.1. *For every $1 \le p < \infty$, the vertical sliced Wasserstein distance VSW_p is a metric on $\mathcal{M}_{\mathrm{sym}}(\mathbb{S}^2)$, which was defined in (28), and the semicircular Wasserstein distance SSW_p is a metric on $\mathcal{M}(\mathbb{S}^2)$.*

Proof. The symmetry and the triangle inequality follow from the corresponding properties of the Wasserstein distance and the p-norm on $\mathbb{T}$ and $\mathbb{S}^2$. The positive definiteness follows from the injectivity of $\mathcal{V}$ and $\mathcal{W}$ in Theorem 3.7 and Theorem 4.10. □

Since the geodesic distance $d(\boldsymbol{\xi}, \boldsymbol{\eta}) = \arccos(\boldsymbol{\xi}^\top \boldsymbol{\eta})$ on the sphere $\mathbb{S}^2$ is rotationally invariant, i.e., $d(\boldsymbol{Q}\boldsymbol{\xi}, \boldsymbol{Q}\boldsymbol{\eta}) = d(\boldsymbol{\xi}, \boldsymbol{\eta})$ for all $\boldsymbol{Q} \in \mathrm{SO}(3)$, the Wasserstein distance (1) on $\mathbb{S}^2$ inherits this property, i.e., $\mathrm{W}_p(\mu, \nu) = \mathrm{W}_p(\mu \circ \boldsymbol{Q}, \nu \circ \boldsymbol{Q})$ for all $\boldsymbol{Q} \in \mathrm{SO}(3)$. The vertical sliced Wasserstein distance is only partially rotation invariant.

Proposition 5.2. *For any $p \in [1, \infty)$, the vertical sliced Wasserstein distance VSW_p is invariant with respect to rotations (11) around the vertical axis, i.e., for all $\mu, \nu \in \mathcal{M}(\mathbb{S}^2)$ and $\alpha \in \mathbb{T}$, it holds*

$$\mathrm{VSW}_p(\mu, \nu) = \mathrm{VSW}_p(\mu \circ \boldsymbol{R}_3(\alpha), \mu \circ \boldsymbol{R}_3(\alpha)).$$

Proof. We have

$$\begin{aligned}\mathrm{VSW}_p^p(\mu \circ \boldsymbol{R}_3(\alpha), \mu \circ \boldsymbol{R}_3(\alpha)) &= \int_{\mathbb{T}} \mathrm{W}_p^p(\mathcal{V}_\psi[\mu \circ \boldsymbol{R}_3(\alpha)], \mathcal{V}_\psi[\nu \circ \boldsymbol{R}_3(\alpha)]) \,\mathrm{d}\psi \\ &= \int_{\mathbb{T}} \mathrm{W}_p^p([\mathcal{S}_\psi \circ \boldsymbol{R}_3(\alpha)^\top]_\# \mu, [\mathcal{S}_\psi \circ \boldsymbol{R}_3(\alpha)^\top]_\# \nu) \,\mathrm{d}\psi.\end{aligned}$$

Since $[\mathcal{S}_\psi \circ \boldsymbol{R}_3(\alpha)^\top](\boldsymbol{\xi}) = \langle \boldsymbol{\xi}, \boldsymbol{R}_3(\alpha)\,(\cos\psi, \sin\psi, 0)^\top \rangle = \mathcal{S}_{\psi+\alpha}(\boldsymbol{\xi})$, we further obtain

$$\mathrm{VSW}_p^p(\mu \circ \boldsymbol{R}_3(\alpha), \mu \circ \boldsymbol{R}_3(\alpha)) = \int_{\mathbb{T}} \mathrm{W}_p^p(\mathcal{V}_{\psi+\alpha} \mu, \mathcal{V}_{\psi+\alpha} \nu) \,\mathrm{d}\psi = \mathrm{VSW}_p^p(\mu, \nu).$$ □

In contrast to the vertical sliced Wasserstein distance, the semicircular sliced Wasserstein distance is invariant to general rotations.

Proposition 5.3. *For any $p \in [1,\infty)$, the semicircular sliced Wasserstein distance* SSW_p *is rotationally invariant, i.e., for every $\mu,\nu \in \mathcal{M}(\mathbb{S}^2)$ and $\boldsymbol{Q} \in \mathrm{SO}(3)$, it holds*

$$\mathrm{SSW}_p(\mu,\nu) = \mathrm{SSW}_p(\mu \circ \boldsymbol{Q}, \mu \circ \boldsymbol{Q}).$$

Proof. For $\gamma \in \mathbb{T}$, let $\mathcal{T}_\gamma \colon \mathbb{T} \to \mathbb{T}$ be the shift operator given by $\mathcal{T}_\gamma(\psi) \coloneqq \psi - \gamma$. The key observation to show the statement is the identity

$$\mathcal{T}_\gamma \circ \mathcal{A}_{\alpha,\beta}(\boldsymbol{\xi}) = \operatorname{azi}(\Psi(\alpha,\beta,0)^\top \boldsymbol{\xi}) - \gamma = \operatorname{azi}(\Psi(\alpha,\beta,\gamma)^\top \boldsymbol{\xi}). \tag{41}$$

Exploiting the shift invariance of the Wasserstein distance on $\mathbb{T}$, the identity (41), and the rotation invariance of the surface measure on $\mathrm{SO}(3)$, we have

$$\begin{aligned}
&\mathrm{SSW}_p^p(\mu \circ \boldsymbol{Q}, \nu \circ \boldsymbol{Q}) = \int_{\mathbb{S}^2} \mathrm{W}_p^p(\mathcal{W}_{\alpha,\beta}[\mu \circ \boldsymbol{Q}], \mathcal{W}_{\alpha,\beta}[\nu \circ \boldsymbol{Q}])\, \mathrm{d}\sigma_{\mathbb{S}^2}(\Phi(\alpha,\beta)) \\
&\quad= \frac{1}{2\pi} \int_{\mathbb{T}} \int_{\mathbb{S}^2} \mathrm{W}_p^p([\mathcal{T}_\gamma \circ \mathcal{A}_{\alpha,\beta}]_\#(\mu \circ \boldsymbol{Q}), [\mathcal{T}_\gamma \circ \mathcal{A}_{\alpha,\beta}]_\#(\nu \circ \boldsymbol{Q}))\, \mathrm{d}\sigma_{\mathbb{S}^2}(\Phi(\alpha,\beta))\, \mathrm{d}\gamma \\
&\quad= \frac{1}{2\pi} \int_{\mathrm{SO}(3)} \mathrm{W}_p^p([\operatorname{azi}(\Psi(\alpha,\beta,\gamma)^\top \boldsymbol{Q}^\top]_\#\mu, [\operatorname{azi}(\Psi(\alpha,\beta,\gamma)^\top \boldsymbol{Q}^\top \cdot)]_\#\nu)\, \mathrm{d}\sigma_{\mathrm{SO}(3)}(\Psi(\alpha,\beta,\gamma)) \\
&\quad= \frac{1}{2\pi} \int_{\mathrm{SO}(3)} \mathrm{W}_p^p([\operatorname{azi}(\Psi(\tilde\alpha,\tilde\beta,\tilde\gamma)^\top \cdot)]_\#\mu, [\operatorname{azi}(\Psi(\tilde\alpha,\tilde\beta,\tilde\gamma)^\top \cdot)]_\#\nu)\, \mathrm{d}\sigma_{\mathrm{SO}(3)}(\Psi(\tilde\alpha,\tilde\beta,\tilde\gamma)) \\
&\quad= \frac{1}{2\pi} \int_{\mathbb{T}} \int_{\mathbb{S}^2} \mathrm{W}_p^p([\mathcal{T}_{\tilde\gamma} \circ \mathcal{A}_{\tilde\alpha,\tilde\beta}]_\#\mu, [\mathcal{T}_{\tilde\gamma} \circ \mathcal{A}_{\tilde\alpha,\tilde\beta}]_\#\nu)\, \mathrm{d}\sigma_{\mathbb{S}^2}(\Phi(\tilde\alpha,\tilde\beta))\, \mathrm{d}\tilde\gamma = \mathrm{SSW}_p^p(\mu,\nu).
\end{aligned}$$

□

6. Discrete Spherical Transforms and Inversion

6.1. Discretization and Inversion via Moore–Penrose Pseudoinverse

We will compute the sliced spherical transforms of (probability density) functions numerically based on the singular value decomposition in a similar way as in [41]. To this end, we need an appropriate discretization. In particular we need quadrature formulas on $\mathbb{S}^2$ as well as on the image domains $\mathbb{T} \times \mathbb{I}$ and $\mathrm{SO}(3)$ of $\mathcal{V}$ and $\mathcal{W}$, respectively.

Let $N \in \mathbb{N}$. We choose quadrature nodes $\boldsymbol{\xi}^m \in \mathbb{S}^2$ and respective weights $w_m > 0$, $m \in [M] \coloneqq \{1,\dots,M\}$ such that all spherical harmonics of degree $\le 2N$ are exactly integrated by the corresponding quadrature rule, see [35,42]. To be more precise, we use the equispaced nodes $\varphi_i = i\pi/(N+1)$, $i \in [2N+2]$, and the Gauss–Legendre nodes $t_j \in [-1,1]$, $j \in [N+1]$, given by the roots of the $(N+1)$st Legendre polynomial. We denote the corresponding Gauss–Legendre weights by r_j. Now we obtain the quadrature

$$\boldsymbol{\xi}^{m(i,j)} \coloneqq \Phi(\varphi_i, \arccos t_j) \quad \text{and} \quad w_{m(i,j)} \coloneqq 2\pi r_j/(2N+2), \qquad i \in [2N+2],\ j \in [N+1],$$

where $m(i,j) \in [M]$ denotes the index related to the pair (i,j) and $M = 2(N+1)^2$. Setting

$$\boldsymbol{f} \coloneqq (f_m)_{m=1}^M \coloneqq (f(\boldsymbol{\xi}^m))_{m=1}^M, \quad \boldsymbol{Y}_n^k \coloneqq (Y_n^k(\boldsymbol{\xi}^m))_{m=1}^M, \quad \text{and} \quad \boldsymbol{w} \coloneqq (w_m)_{m=1}^M, \tag{42}$$

we approximate the spherical harmonics coefficients $\langle f, Y_n^k \rangle$ by

$$\left\langle \boldsymbol{f}, \boldsymbol{Y}_n^k \right\rangle_{\boldsymbol{w}} \coloneqq \sum_{m=1}^{M} f(\boldsymbol{\xi}^m)\, \overline{Y_n^k(\boldsymbol{\xi}^m)}\, w_m, \qquad n = 0, \ldots, N,\ k = -n, \ldots, n.$$

In particular, we have that $\langle f, Y_n^k \rangle = \langle \boldsymbol{f}, \boldsymbol{Y}_n^k \rangle_{\boldsymbol{w}}$ if f is a spherical polynomial of degree $\le N$. All discrete Fourier coefficients can be computed efficiently in $\mathcal{O}(N^2 \log^2 N)$ arithmetic operations utilizing the *nonuniform fast spherical Fourier transform* (NFSFT) [50, 63].

Discrete Vertical Slice Transform For discretizing $\mathbb{T} \times \mathbb{I}$, we use again equispaced nodes $\psi_i = i\pi/(N+1)$, $i \in [2N+2]$ and Gauss–Legendre nodes t_j and weights r_j, $j \in [N+1]$. We denote the respective quadrature weights on $\mathbb{T} \times \mathbb{I}$ by $\tilde{\boldsymbol{w}}$, where $\tilde{w}_{\ell(i,j)} = \pi r_j/(N+1)$ and $\ell(i,j) \in [L]$ with $L = 2(N+1)^2$ denotes the index related to the pair (i,j). The quadrature is exact of degree $2N$, i.e., for all linear combinations of basis functions B_n^k with $0 \le n, |k| \le 2N$. Using the singular value decomposition (21), we discretize $\mathcal{V}$ by

$$\mathcal{V}_D \colon \mathbb{R}^M \to \mathbb{R}^L, \quad \mathcal{V}_D \boldsymbol{f} \coloneqq \sum_{n=0}^{N} \sum_{\substack{k=-n \\ n+k \text{ even}}}^{n} \mathrm{v}_n^k \left\langle \boldsymbol{f}, \boldsymbol{Y}_n^k \right\rangle_{\boldsymbol{w}} \boldsymbol{B}_n^k, \tag{43}$$

where $\boldsymbol{B}_n^k \coloneqq (B_n^k(\psi_i, t_j))_{\ell(i,j)=1}^L$ and B_n^k is given in (20). Then $\mathcal{V}_D \boldsymbol{f}$ can be computed using the fast Fourier transform (FFT) in ψ and a fast polynomial transform [65] in t in $\mathcal{O}(N^2 \log^3 N)$ arithmetic operations. Based on the quadrature for $\mathbb{T} \times \mathbb{I}$, we analogously discretize the (truncated) Moore–Penrose pseudoinverse (22) by

$$\mathcal{V}_D^\dagger \colon \mathbb{R}^L \to \mathbb{R}^M, \quad \mathcal{V}_D^\dagger \boldsymbol{g} \coloneqq \sum_{n=0}^{N} \sum_{\substack{k=-n \\ n+k \text{ even}}}^{n} \frac{1}{\mathrm{v}_n^k} \left\langle \boldsymbol{g}, \boldsymbol{B}_n^k \right\rangle_{\tilde{\boldsymbol{w}}} \boldsymbol{Y}_n^k,$$

where $\boldsymbol{g} \coloneqq (g(\psi_i, t_j))_{\ell(i,j)=1}^L$ consists of samples of $g \colon \mathbb{T} \times \mathbb{I} \to \mathbb{R}$. For a spherical polynomial f of degree N, the chosen quadratures ensure $\mathcal{V}_D^\dagger \mathcal{V}_D \boldsymbol{f} = \boldsymbol{f}$.

Discrete Semicircle Transform We use quadrature nodes $\boldsymbol{Q}_\ell \in \mathrm{SO}(3)$ and weights $\tilde{\boldsymbol{w}} = (\tilde{w}_\ell)_{\ell=1}^L$ such that all rotational harmonics of degree $\le 2N$ are exactly integrated. Since it becomes clear from the context which weights are addressed, we use again $\tilde{\boldsymbol{w}}$. In particular, we consider a quadrature [34] on $\mathrm{SO}(3) \cong \mathbb{S}^2 \times \mathbb{S}^1$ as product of a Gauss-type quadrature on $\mathbb{S}^2$, see [35], and an equispaced quadrature on $\mathbb{T}$. We use this product structure because we can now discretize $\mathcal{W}_{\alpha,\beta}$ on a uniform grid. Similarly to (43), the singular value decomposition (34) of $\mathcal{W} \colon L^2(\mathbb{S}^2) \to L^2(\mathrm{SO}(3))$ can be truncated as

$$\mathcal{W}_D \colon \mathbb{R}^M \to \mathbb{R}^L, \quad \mathcal{W}_D \boldsymbol{f} \coloneqq \sum_{n=0}^{N} \sum_{j,k=-n}^{n} \lambda_n^j \left\langle \boldsymbol{f}, \boldsymbol{Y}_n^k \right\rangle_{\boldsymbol{w}} \overline{\boldsymbol{D}_n^{k,j}}, \tag{44}$$

where $\boldsymbol{D}_n^{k,j} \coloneqq (D_n^{k,j}(\boldsymbol{Q}_\ell))_{\ell=1}^L$. Then (44) can be computed in $\mathcal{O}(N^3 \log^2 N + L)$ arithmetic operations with the *nonuniform fast SO(3) Fourier transform* (NFSOFT) [64]. Further, we approximate the Moore–Penrose pseudoinverse (35) of $\mathcal{W}$ by

$$\mathcal{W}_D^\dagger\colon \mathbb{R}^L \to \mathbb{R}^M, \quad \mathcal{W}_D^\dagger \boldsymbol{g} \coloneqq \sum_{n=0}^{N} \sum_{j,k=-n}^{n} \frac{1}{(\mathrm{w}_n)^2}\, \lambda_n^j \left\langle \boldsymbol{g}, \overline{\boldsymbol{D}_n^{k,j}} \right\rangle_{\tilde{\boldsymbol{w}}} \boldsymbol{Y}_n^k,$$

where $\boldsymbol{g} \coloneqq (g(\boldsymbol{Q}_\ell))_{\ell=1}^L$ for $g\colon \mathrm{SO}(3) \to \mathbb{R}$. As above, $\mathcal{W}_D^\dagger \boldsymbol{g}$ can be evaluated with NFSFT and NFSOFT algorithms. For a spherical polynomial f of degree N, we have $\mathcal{W}_D^\dagger \mathcal{W}_D \boldsymbol{f} = \boldsymbol{f}$.

6.2. Inversion by Variational Approach

The push-forward definitions of the sliced spherical transforms ensure that probability measures are mapped to probability measures. In the context of optimal transport, we require that the inverse transforms have the same behaviour. Even when restricting to the function setting, we can however construct functions with non-trivial negative part that are transformed into probability densities, e.g. by taking a function that is negative in a sufficiently small spherical cap $\{\boldsymbol{\xi} \in \mathbb{S}^2 : \xi_3 < 1 - \varepsilon\}$ and equals a positive constant otherwise. Thus the Moore–Penrose pseudoinverse applied to a probability density is not necessarily a probability density. To overcome this issue, we consider the inversion of the discretized spherical transforms as inverse problems, which we solve using a variational formulation.

As in (42), let $\boldsymbol{f} \in \mathbb{R}^M$ contain the samples of the probability density function f on $\mathbb{S}^2$, and $\boldsymbol{w} \in \mathbb{R}^M$ the respective quadrature weights. If the quadrature is exact for f, we have $\boldsymbol{f} \ge 0$ and $\int_{\mathbb{S}^2} f \,\mathrm{d}\sigma_{\mathbb{S}^2} = \sum_{m=1}^M w_m f_m = 1$. Thus $\boldsymbol{f}$ can be interpreted as probability density function with respect to the counting measure weighted by $\boldsymbol{w}$. For the numerical inversion, we handle both transforms simultaneously, denoting the discretizations $\mathcal{V}_D$ and $\mathcal{W}_D$ by $\mathcal{T}_D\colon \mathbb{R}^M \to \mathbb{R}^L$. Let $\boldsymbol{g} \in \mathbb{R}^L$ be the samples of the density function g on $\mathbb{T} \times \mathbb{I}$ or $\mathrm{SO}(3)$. We equip $\mathbb{R}^M$ with the weighted Euclidean inner product $\langle \boldsymbol{f}, \tilde{\boldsymbol{f}} \rangle_{\boldsymbol{w}} \coloneqq \sum_{m=1}^M w_m f_m \tilde{f}_m$, and, analogously, $\mathbb{R}^L$ with the inner product $\langle \boldsymbol{g}, \tilde{\boldsymbol{g}} \rangle_{\tilde{\boldsymbol{w}}} \coloneqq \sum_{\ell=1}^L \tilde{w}_\ell g_\ell \tilde{g}_\ell$, where $\tilde{\boldsymbol{w}}$ contains the quadrature weights for $\mathbb{T} \times \mathbb{I}$ or $\mathrm{SO}(3)$. Furthermore, we denote the all-one vector by $\mathbf{1}$.

Now, we aim to find an approximate solution $\boldsymbol{f}$ of the inversion problem $\mathcal{T}_D \boldsymbol{f} = \boldsymbol{g}$ in the weighted probability simplex $\Delta_{\boldsymbol{w}} \coloneqq \{\boldsymbol{f} \in \mathbb{R}^M : \boldsymbol{f} \ge 0, \langle \boldsymbol{f}, \mathbf{1} \rangle_{\boldsymbol{w}} = 1\}$. To this end, we introduce a regularized inverse as the minimizer of the strictly convex optimization problem

$$\operatorname*{arg\,min}_{\boldsymbol{f} \in \Delta_{\boldsymbol{w}}} \mathrm{KL}_{\tilde{\boldsymbol{w}}}(\mathcal{T}_D \boldsymbol{f}, \boldsymbol{g}) + \rho\, \mathrm{KL}_{\boldsymbol{w}}(\boldsymbol{f}, \mathbf{1}), \quad \rho > 0, \tag{45}$$

where $\mathrm{KL}_{\boldsymbol{w}}$ is the discrete *Kullback–Leibler (KL) divergence* on the weighted space $\mathbb{R}^M$ given by

$$\mathrm{KL}_{\boldsymbol{w}}(\boldsymbol{f}, \tilde{\boldsymbol{f}}) \coloneqq \langle \boldsymbol{f}, \log \boldsymbol{f} - \log \tilde{\boldsymbol{f}} \rangle_{\boldsymbol{w}} + \langle \tilde{\boldsymbol{f}} - \boldsymbol{f}, \mathbf{1} \rangle_{\boldsymbol{w}},$$

for $\boldsymbol{f}, \tilde{\boldsymbol{f}} \ge 0$ with $f_m = 0$ whenever $\tilde{f}_m = 0$, and $\mathrm{KL}_{\boldsymbol{w}}(\boldsymbol{f}, \tilde{\boldsymbol{f}}) \coloneqq +\infty$ otherwise. Here the logarithm acts componentwise, and we set $0 \log 0 \coloneqq 0$. Note that $\mathrm{KL}_{\boldsymbol{w}}(\boldsymbol{f}, \mathbf{1})$ is the negative entropy of $\boldsymbol{f}$. The KL divergence on $\mathbb{R}^L$ is defined analoguously.

To find the minimizer of (45), we employ the primal-dual splitting of Chambolle and Pock [18]. To this end, we reformulate (45) as

$$\operatorname*{arg\,min}_{\boldsymbol{f}, \boldsymbol{y}_1, \boldsymbol{y}_2} \mathrm{KL}_{\tilde{\boldsymbol{w}}}(\boldsymbol{y}_1, \boldsymbol{g}) + \rho\, \mathrm{KL}_{\boldsymbol{w}}(\boldsymbol{y}_2, \mathbf{1}) + \chi_{\Delta_{\boldsymbol{w}}}(\boldsymbol{f}) \quad \text{s.t.} \quad \mathcal{T}_D \boldsymbol{f} = \boldsymbol{y}_1,\ \boldsymbol{f} = \boldsymbol{y}_2$$

with the *characteristic function* $\chi_{\Delta_{\boldsymbol{w}}}(\boldsymbol{f}) = 0$ for $\boldsymbol{f} \in \Delta_{\boldsymbol{w}}$ and $\chi_{\Delta_{\boldsymbol{w}}}(\boldsymbol{f}) = +\infty$ else. For $\theta \in (0,1]$ and $\sigma, \tau > 0$ such that $1/\tau\sigma > \|I + \mathcal{T}_D^* \mathcal{T}_D\|$, the algorithm converges and reads as

$$\boldsymbol{f}^{k+1} \coloneqq \operatorname{proj}_{\Delta_{\boldsymbol{w}}}\big(\boldsymbol{f}^k - \tau \mathcal{T}_D^* \boldsymbol{y}_1^k - \tau \boldsymbol{y}_2^k\big), \tag{46a}$$
$$\tilde{\boldsymbol{f}}^{k+1} \coloneqq \boldsymbol{f}^{k+1} + \theta(\boldsymbol{f}^{k+1} - \boldsymbol{f}^k), \tag{46b}$$
$$\boldsymbol{y}_1^{k+1} \coloneqq \operatorname{prox}_{\sigma \operatorname{KL}^*_{\tilde{\boldsymbol{w}}}(\cdot,\boldsymbol{g})}\big(\boldsymbol{y}_1^k + \sigma \mathcal{T}_D \tilde{\boldsymbol{f}}^{k+1}\big) \tag{46c}$$
$$\boldsymbol{y}_2^{k+1} \coloneqq \operatorname{prox}_{\sigma(\rho \operatorname{KL}_{\boldsymbol{w}})^*(\cdot,\mathbf{1})}\big(\boldsymbol{y}_2^k + \sigma\, \tilde{\boldsymbol{f}}^{k+1}\big). \tag{46d}$$

Here $\operatorname{proj}_{\Delta_{\boldsymbol{w}}}$ is the orthogonal projection onto $\Delta_{\boldsymbol{w}}$. Further, for a function $h\colon \mathbb{R}^M \to \mathbb{R}$, the *proximal operator* with respect to the weight $\boldsymbol{w}$ is given by

$$\operatorname{prox}_{\sigma h}(\boldsymbol{x}) \coloneqq \operatorname*{arg\,min}_{\boldsymbol{y} \in \mathbb{R}^M} h(\boldsymbol{y}) + \tfrac{1}{2\sigma} \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{w}}^2 \tag{47}$$

and its *Fenchel conjugate* by $h^*(\boldsymbol{y}) \coloneqq \max_{\boldsymbol{x} \in \mathbb{R}^M} \langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{w}} + h(\boldsymbol{y})$. On $\mathbb{R}^L$ with weight $\tilde{\boldsymbol{w}}$, the proximal operator and conjugate are defined similarly.

Proposition 6.1. *The orthogonal projection onto $\Delta_{\boldsymbol{w}}$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{w}}$ is given by*

$$\operatorname{proj}_{\Delta_{\boldsymbol{w}}}(\boldsymbol{f}) = [\boldsymbol{f} + \lambda \mathbf{1}]_+,$$

where $[\cdot]_+$ denotes the componentwise positive part, and λ is the root of $\langle \mathbf{1}, [\boldsymbol{f} + \lambda \mathbf{1}]_+ \rangle_{\boldsymbol{w}} - 1$.

The statement follows line by line via incorporating the weighted inner product into the argumentation in [9, Thm 6.27]. The function in Proposition 6.1 is monotonically increasing and piecewise linear with finitely many pieces; thus the root can be determined using a bisection method to identify the piece with sign change and solving a linear equation. For the standard probability simplex with $\boldsymbol{w} = \mathbf{1}$, there exist several further numerically efficient approaches [21].

Proposition 6.2. *Let $\sigma, a > 0$, and $\boldsymbol{b} \in \mathbb{R}^M$. On the weighted Euclidean space $\mathbb{R}^M$, the KL divergence satisfies*

$$\operatorname{prox}_{\sigma(a \operatorname{KL})^*(\cdot,\boldsymbol{b})}(\boldsymbol{x}) = \boldsymbol{x} - aW\big(\tfrac{\sigma}{a}\, \boldsymbol{b} \odot \exp\big(\tfrac{1}{a}\, \boldsymbol{x}\big)\big), \qquad \boldsymbol{x} \in \mathbb{R}^M,$$

where W denotes the componentwise applied Lambert's W-function *that maps z to the solution y of $y \exp y = z$ and $\odot$ the componentwise multiplication.*

Proof. Differentiating the objective of the weighted Fenchel conjugate and setting to zero yields

$$(a \operatorname{KL}_{\boldsymbol{w}})^*(\boldsymbol{y}, \boldsymbol{b}) = a \left\langle \boldsymbol{b}, \exp\left(\tfrac{1}{a}\, \boldsymbol{y}\right) \right\rangle_{\boldsymbol{w}} + a \left\langle \boldsymbol{b}, \mathbf{1} \right\rangle_{\boldsymbol{w}}.$$

Inserting the conjugated scaled KL divergence into (47), and setting the derivative again zero, we componentwise obtain

$$\begin{aligned} \boldsymbol{b} \odot \exp\big(\tfrac{1}{a}\, \boldsymbol{y}\big) = \tfrac{1}{\sigma}\,(\boldsymbol{x} - \boldsymbol{y}) \quad &\Leftrightarrow \quad \log(\boldsymbol{b}) + \tfrac{1}{a}\, \boldsymbol{x} = \tfrac{1}{a}\,(\boldsymbol{x} - \boldsymbol{y}) + \log\big(\tfrac{1}{\sigma}\,(\boldsymbol{x} - \boldsymbol{y})\big) \\ &\Leftrightarrow \quad \tfrac{\sigma}{a}\, \boldsymbol{b} \odot \exp\big(\tfrac{1}{a}\, \boldsymbol{x}\big) = \tfrac{1}{a}\,(\boldsymbol{x} - \boldsymbol{y}) \odot \exp\big(\tfrac{1}{a}\,(\boldsymbol{x} - \boldsymbol{y})\big), \end{aligned}$$

which gives the assertion. □

Note that the primal-dual algorithm requires the adjoint $\mathcal{T}_D^*$ in (46a). Based on the discretized spherical transforms in Section 6.1, we obtain their adjoint operators

$$\mathcal{V}_D^* \boldsymbol{g} = \sum_{n=0}^{N} \sum_{\substack{k=-n \\ n+k \,\mathrm{even}}}^{n} \mathrm{v}_n^k \left\langle \boldsymbol{g}, \boldsymbol{B}_n^k \right\rangle_{\tilde{\boldsymbol{w}}} \boldsymbol{Y}_n^k \quad \text{and} \quad \mathcal{W}_D^* \boldsymbol{g} = \sum_{n=0}^{N} \sum_{k,j=-n}^{n} \lambda_n^j \left\langle \boldsymbol{g}, \overline{\boldsymbol{D}_n^{k,j}} \right\rangle_{\tilde{\boldsymbol{w}}} \boldsymbol{Y}_n^k.$$

The primal-dual iteration (46) may be summarized as follows.

Algorithm 6.3 (Primal-Dual for Regularized Inversion).
Input: $\boldsymbol{g} \in \mathbb{R}^L$, $\theta \in (0,1]$, $\sigma, \tau, \rho > 0$.
Initialization: $\boldsymbol{f}^0 := (4\pi)^{-1}\,\mathbf{1}$, $\boldsymbol{y}_1^0 := \mathbf{0}$, $\boldsymbol{y}_2^0 := \mathbf{0}$.
Iteration: For $k = 0, 1, \dots$ until convergence do
(a) $\boldsymbol{f}^{k+1} := \mathrm{proj}_{\Delta_{\boldsymbol{w}}}(\boldsymbol{f}^k - \tau \mathcal{T}_D^* \boldsymbol{y}_1^k - \tau \boldsymbol{y}_2^k)$,
(b) $\tilde{\boldsymbol{f}}^{k+1} := \boldsymbol{f}^{k+1} + \theta(\boldsymbol{f}^{k+1} - \boldsymbol{f}^k)$,
(c) $\tilde{\boldsymbol{y}}_1^{k+1} := \boldsymbol{y}_1^k + \sigma \mathcal{T}_D \tilde{\boldsymbol{f}}^{k+1}$,
(d) $\boldsymbol{y}_1^{k+1} := \tilde{\boldsymbol{y}}_1^{k+1} - W\big(\sigma \boldsymbol{g} \odot \exp(\tilde{\boldsymbol{y}}_1^{k+1})\big)$,
(e) $\tilde{\boldsymbol{y}}_2^{k+1} := \boldsymbol{y}_1^k + \sigma\, \tilde{\boldsymbol{f}}^{k+1}$,
(f) $\boldsymbol{y}_2^{k+1} := \tilde{\boldsymbol{y}}_2^{k+1} - \rho\, W\big(\frac{\sigma}{\rho} \exp(\frac{1}{\rho} \tilde{\boldsymbol{y}}_2^{k+1})\big)$.
Output: $\boldsymbol{f} \in \mathbb{R}^M$ solving (45).

7. Numerical Results

In this section, we provide proof-of-concept examples that the sliced spherical transforms can be combined in a meaningful way with optimal transport on the interval and the circle. First, we deal with the approximation of Wasserstein barycenters on the sphere. In particular, this requires the inversion of the sliced spherical transforms. Second, we show that these transforms combined with optimal transport can be used for classifying classes of measures. All numerical tests are performed in Matlab R2022a on an Intel Core i7-10700 CPU with 16 GB memory.

7.1. Interpolation between Probability Measures

Given two probability measures on the sphere, we generate a measure "between" them, as proposed in [48] for the Radon transform on $\mathbb{R}^2$. In particular, we compute the CDT or cCDT of their spherical transform $\mathcal{V}$ or $\mathcal{W}$, then we interpolate in the CDT space and go back to $\mathbb{S}^2$ via the inverse of the CDT or cCDT and the spherical transforms.

For computing the forward, inverse, and adjoint spherical transforms, we truncate the singular value decomposition at degree $N = 44$ and use the software package [44] for the NFSFT and NFSOFT. We have $M = (2N+2)(N+1) = 4050$ quadrature nodes on the sphere, cf. Section 6.1.

Interpolation between Mises–Fisher Distributions As test function on $C(\mathbb{S}^2)$, we choose the density of the *von Mises–Fisher (vMF) distribution*

$$f_{\kappa,\boldsymbol{\eta}}(\boldsymbol{\xi}) = c_\kappa\, \mathrm{e}^{\kappa \langle \boldsymbol{\eta}, \boldsymbol{\xi} \rangle}, \quad \boldsymbol{\xi} \in \mathbb{S}^2, \tag{48}$$

with the mean direction $\boldsymbol{\eta} \in \mathbb{S}^2$ and the concentration $\kappa > 0$, where c_κ is chosen such that $\int_{\mathbb{S}^2} f_{\kappa,\boldsymbol{\eta}} \, \mathrm{d}\sigma_{\mathbb{S}^2} = 1$. Since $\mathcal{V}$ acts only on even functions, we make our first tests with symmetrized vMF distributions via $(f_{\kappa,\boldsymbol{\eta}}(\xi_1, \xi_2, \xi_3) + f_{\kappa,\boldsymbol{\eta}}(\xi_1, \xi_2, -\xi_3))/2$, see Figure 3.

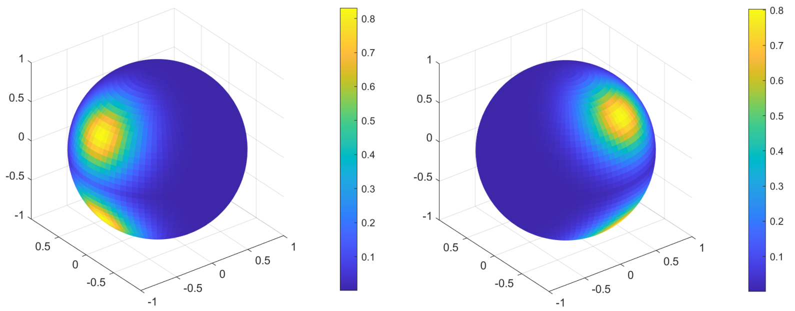


Figure 3: Density functions of two symmetrized vMF distributions (48).

Let $\mu, \nu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{S}^2)$ be two given measures. For some $\delta \in [0, 1]$, we set the *unregularized $\mathcal{V}$-CDT interpolation* between μ and ν as

$$\mathcal{V}^\dagger h, \quad \text{where} \quad h(\psi, t) = \mathrm{CDT}^{-1}_{\mathcal{V}_\psi \mu} \left[\delta \; \mathrm{CDT}_{\mathcal{V}_\psi \mu}[\mathcal{V}_\psi \nu]\right](t), \quad \psi \in \mathbb{T}, \; t \in \mathbb{I}.$$

Here, we discretize $\mathcal{V}_\psi$ via (43) with $2(N+1)^2 = 4050$ quadrature nodes on $\mathbb{T} \times \mathbb{I}$. Our implementation of the CDT and its inverse is based on [48].[1]

Analogously, we define the *unregularized $\mathcal{W}$-CDT interpolation* $\mathcal{W}^\dagger h$, where

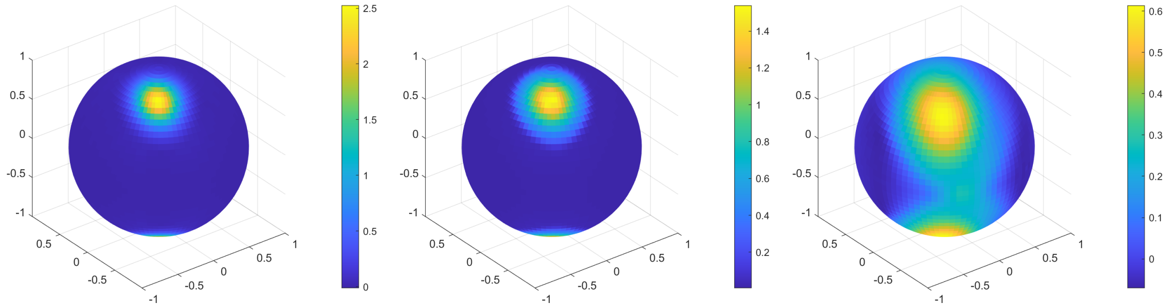
$$h(\Psi(\alpha, \beta, \gamma)) = \mathrm{CDT}^{-1}_{\mathcal{W}_{\alpha,\beta}\mu} \left[\delta \; \mathrm{cCDT}_{\mathcal{W}_{\alpha,\beta}\mu}[\mathcal{W}_{\alpha,\beta}\nu]\right](\gamma), \quad \Psi(\alpha, \beta, \gamma) \in \mathrm{SO}(3).$$

Here, the optimal parameter θ of (6), which is required for the cCDT, is determined by the algorithm [23].[2] Moreover, we compute $\mathcal{W}\mu$ and $\mathcal{W}\nu$ by (44), where we use $L = 118\,944$ quadrature points $\Psi(\alpha, \beta, \gamma)$ on $\mathrm{SO}(3)$, which are obtained as the product of a Gauss-type quadrature[3] in $\Phi(\alpha, \beta) \in \mathbb{S}^2$ and a uniform grid in γ.

Instead of the Moore–Penrose pseudoinverse $\mathcal{V}^\dagger$ or $\mathcal{W}^\dagger$, we also apply the primal-dual Algorithm 6.3 to obtain the regularized inverse (45) of h, which we call the regularized $\mathcal{V}$-CDT or $\mathcal{W}$-CDT interpolation. Here we choose the regularization parameter $\rho = 0.1$ and step sizes $\sigma = 1$ and $\tau = 1/4$, and we terminate the algorithm after 200 iterations. The CDT interpolations for $\delta = 0.5$ are plotted in Figure 4. While the regularization has a comparably small effect on the $\mathcal{V}$-CDT interpolation, we note that the unregularized $\mathcal{W}$-CDT interpolation is severely negative in some areas and therefore not a probability density, which is circumvented by the primal-dual algorithm.

As a reference, we consider the spherical 2-Wasserstein barycenter (2) and its entropy-regularized counterpart [62], whose computation with the Sinkhorn algorithm [46] can be implemented efficiently, cf. [7, 14]. We apply the Python optimal transport library [28] for both, where the Sinkhorn algorithm uses the regularization parameter 0.01 and a maximum number of 1000 iterations. In our example in Figure 4, the regularized 2-Wasserstein barycenter looks similar to the $\mathcal{W}$-CDT interpolation, while the unregularized barycenter is very noisy and takes very long to compute with a linear program solver.

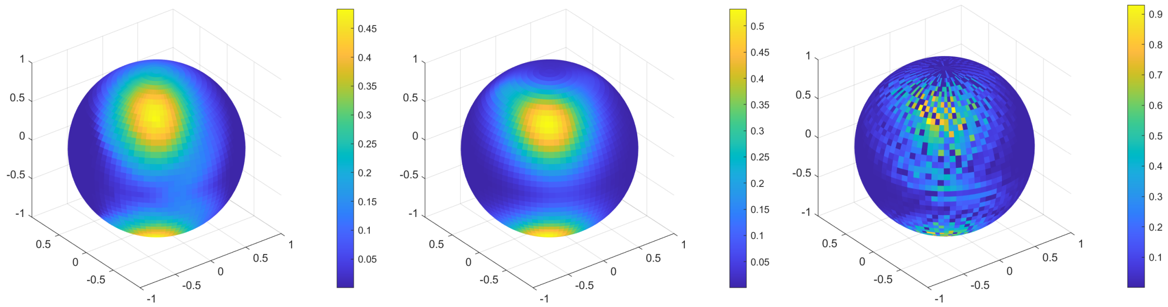
[1] See the Python code https://github.com/skolouri/Radon-Cumulative-Distribution-Transform
[2] See the Matlab code https://users.mccme.ru/ansobol/otarie/software.html
[3] Quadrature rule on $\mathbb{S}^2$ from http://www.tu-chemnitz.de/~potts/workgroup/graef/quadrature.



(a) $\mathcal{V}$-CDT interpolation (0.01 s)

(b) $\mathcal{V}$-CDT interpolation with regularized inverse (0.3 s)

(c) $\mathcal{W}$-CDT interpolation with Moore–Penrose pseudoinverse $\mathcal{W}^\dagger$ (3.2 s)



(d) $\mathcal{W}$-CDT interpolation with regularized inverse (128 s)

(e) Regularized 2-Wasserstein barycenter (19 s)

(f) Unregularized 2-Wasserstein barycenter (33 h)

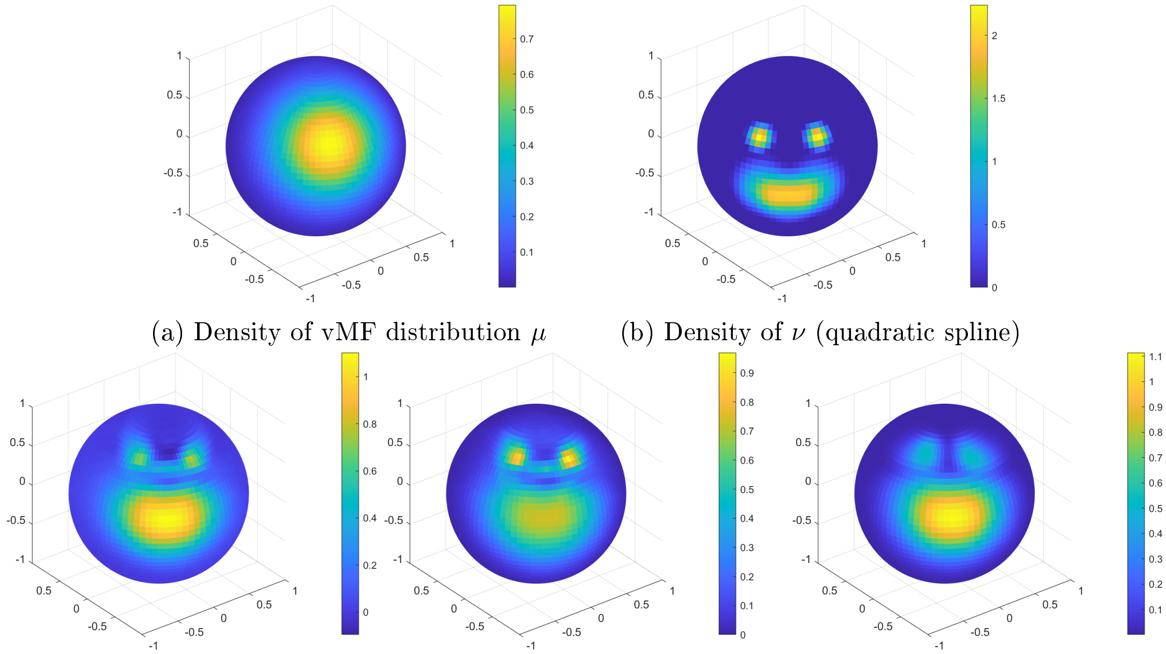
Figure 4: CDT interpolation with $\delta = 0.5$ of the vMF distributions from Figure 3.

Interpolation between vMF Distribution and a Mixture The $\mathcal{W}$-CDT interpolation of more evolved test functions, which are not symmetric, is depicted in Figure 5. We notice that the $\mathcal{W}$-CDT interpolation shows the "eyes" more clearly than the regularized 2-Wasserstein barycenter. This might be caused by a too large regularization parameter of the Sinkhorn algorithm, but when making it smaller the algorithm fails with a division by zero error.

7.2. Classification of Probability Measures

In one dimension, the cumulative distribution transform (3) is known to increase the separability between certain classes of probability measures. If the considered classes are build from prototypes using certain transformations like shifts or scalings, the constructed classes are linearly separable in the CDT space [57, 61]. For probability measures on multi-dimensional domains, the separability of the CDT can be exploited by transforming the considered measures to a series of line measures using the Radon or generalized Radon transform [47, 48]. If we replace the Radon transform by the vertical slice or the normalized semicircle transform, the procedure can be immediately transferred to measures on the sphere.

To show that the vertical slice and semicircle transform can in principle improve the separability between different classes of probability measures, we built five datasets consisting of 100 measures each. Each datum represents a (discretized) single or mixture density function of vMF distributions. The concentration is always chosen as $\kappa = 50$. The means are randomly generated on $\mathbb{S}^2$ satisfying the restrictions in Table 1. All distributions in the



(a) Density of vMF distribution μ (b) Density of ν (quadratic spline)

(c) $\mathcal{W}$-CDT interpolation with Moore–Penrose pseudoinverse $\mathcal{W}^\dagger$ (d) $\mathcal{W}$-CDT interpolation with regularized inverse (e) Regularized 2-Wasserstein barycenter

Figure 5: CDT interpolation of density functions with $\delta = 0.5$.

mixtures are equally weighted. Figure 6 shows some examples of the different classes. On the basis of these classes, we train and test linear support vector machines (SVMs) to study the linear separability after our spherical transformations. To be more precise, let f_ν be the density function of a specific datum. This specimen is now transformed into

$$\left(\mathrm{CDT}_{u_{\mathbb{I}}}[\mathcal{V}_\psi f_\nu]\right)_{\psi \in \mathbb{T}} \quad \text{and} \quad \left(\mathrm{cCDT}_{u_{\mathbb{T}}}[\mathcal{W}_{\alpha,\beta} f_\nu]\right)_{\Phi(\alpha,\beta)\in\mathbb{S}^2},$$

where $u_\bullet$ denotes the uniform measure. For the numerical implementation, we use the quadrature points $\boldsymbol{\xi}^m$ with $N = 44$ from Section 7.1.

Training and testing of the SVMs is here based on 10-fold cross-validations, i.e., the dataset is divided in 10 subsets containing equally many samples of each class, the training is performed on 9 subsets, and the testing on the remaining. The procedure is repeated 10 times such that each subset serves one time as testing set. Before training, the dimension of the training set is reduced to 50 using a principle component analysis. The success rates of the trained SVMs are given in Table 2. Although the vertical slice transform cannot distinguish between the upper and lower hemisphere, the $\mathcal{V}$-CDT approach yields high-quality linear separators between the classes of single and mixture vMF densities. The $\mathcal{V}$-CDT approach only fails in experiment #3, which is not surprising since the samples from the second class are seen as single vMF densities by $\mathcal{V}$. The $\mathcal{W}$-CDT approach is useless in #1 and #2, but significantly increases the separability between single and symmetrized vMF distributions. These first simulations show that both spherical transforms can increase the linear separability between certain classes.

Table 1: Created datasets to study the distinctiveness of linear SVMs with respect to the CDT of the vertical slice/semicircle transform respectively. Each datum consists of a single or the equally weighted mixture of two vMF distributions with fixed concentration $\kappa = 50$ and randomly generated $\boldsymbol{\eta} \in \mathbb{S}^2$.

dataset	1st class	2nd class
#1	single vMF distributions	mixtures of two vMFs, means with fixed distance $\pi/2$
#2	single vMF distributions	mixtures of two vMFs
#3	single vMF distributions	mixtures of two vMFs, means mirrored at equatorial plane
#4	single vMF distributions	mixtures of two vMFs, means mirrored at ξ_3 axis
#5	mixtures of two vMFs, means mirrored at equatorial plane	mixtures of two vMFs, means mirrored at ξ_3 axis

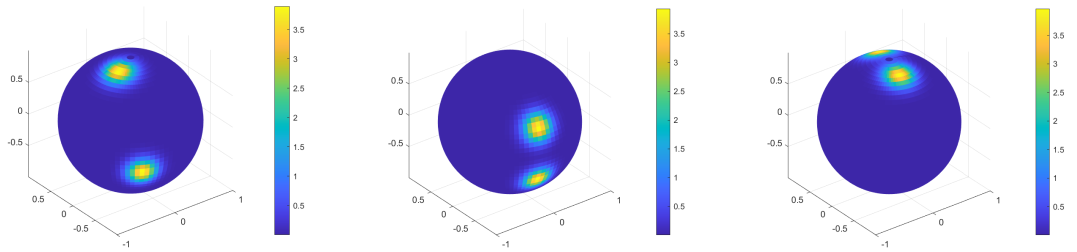


Figure 6: Examples from the generated datasets. From left to right: The means are generated with fixed distance $\pi/2$, mirrored at the equator, and mirrored at the ξ_3 axis.

Acknowledgements

We thank Sophie Mildenberger for creating an illustration of the semicircle transform. We gratefully acknowledge the funding by the German Research Foundation (DFG): STE 571/19-1, project number 495365311, within the SFB F68: "Tomography Across the Scales" as well as the BMBF under the project "VI-Screen" (13N15754).

A. Proof of Theorem 3.7

Any measure $\mu \in \mathcal{M}_{\mathrm{sym}}(\mathbb{S}^2)$ is uniquely determined by its application on $C_{\mathrm{sym}}(\mathbb{S}^2)$, since, by the definition (28), we have for any $f \in C(\mathbb{S}^2)$ that $\langle \mu, f \rangle = \frac{1}{2} \langle \mu, f + \check{f} \rangle$ and $f + \check{f} \in$

Table 2: Success rates of linear SVMs trained and tested directly on the density distributions (—/—), the CDT and vertical sliced transformed densities ($\mathcal{V}$-CDT), and the cCDT semicicle transformed densities ($\mathcal{W}$-CDT). Mean accuracy and standard deviation are computed with respect to 10-fold cross-validations of the datasets in Table 1.

dataset	#1	#2	#3	#4	#5
—/—	0.555 ± 0.123	0.545 ± 0.145	0.570 ± 0.079	0.585 ± 0.085	0.570 ± 0.136
$\mathcal{V}$-CDT	$\mathbf{0.985} \pm 0.024$	$\mathbf{0.985} \pm 0.024$	0.435 ± 0.111	$\mathbf{0.995} \pm 0.016$	0.995 ± 0.016
$\mathcal{W}$-CDT	0.465 ± 0.097	0.565 ± 0.125	$\mathbf{0.865} \pm 0.085$	0.945 ± 0.037	$\mathbf{1.000} \pm 0.000$

$C_{\mathrm{sym}}(\mathbb{S}^2)$, where $\check{f}(\boldsymbol{\xi}) = f(\xi_1, \xi_2, -\xi_3)$. Let $\mu, \nu \in \mathcal{M}_{\mathrm{sym}}(\mathbb{S}^2)$ such that $\mathcal{V}\mu = \mathcal{V}\nu$. Then we obtain for $g \in C(\mathbb{T} \times \mathbb{I})$ by Proposition 3.5 that

$$\langle \mu, \mathcal{V}^* g \rangle = \langle \nu, \mathcal{V}^* g \rangle .$$

Hence, the claim $\mu = \nu$ holds true if $\{\mathcal{V}^* g : g \in C(\mathbb{T} \times \mathbb{I})\}$ is a dense subset of $C_{\mathrm{sym}}(\mathbb{S}^2)$. To show this, let $s > 2$ and $f \in H^s_{\mathrm{sym}}(\mathbb{S}^2)$, which is dense in $C_{\mathrm{sym}}(\mathbb{S}^2)$. Here we denote by $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ the subset of even functions of the Sobolev space $H^s(\mathbb{S}^2)$, see (10). Since $\mathcal{V}$ is injective by Theorem 3.2, we have $f = \mathcal{V}^* g$ if and only if $\mathcal{V} f = \mathcal{V}\mathcal{V}^* g$. In the following, we show that

$$g \coloneqq (\mathcal{V}\mathcal{V}^*)^{-1} \mathcal{V} f$$

is continuous on $\mathbb{T} \times \mathbb{I}$, then we obtain $f = \mathcal{V}^* g$, which shows the assertion. We proceed in a similar manner as for the proof of Sobolev's embedding theorem, cf. [56, lem. 6.14].

Recall the right singular functions B_n^k of $\mathcal{V}$ from (20). Since $\mathcal{V}^*$ has the same singular functions as $\mathcal{V}$ and the conjugate singular values $\overline{\mathrm{v}_n^k} = \mathrm{v}_n^k$, we have by Theorem 3.2

$$\mathcal{V}\mathcal{V}^* h = \sum_{n=0}^{\infty} \sum_{\substack{k=-n \\ n+k \text{ even}}}^{n} \left| \mathrm{v}_n^k \right|^2 \left\langle h, B_n^k \right\rangle_{L^2(\mathrm{SO}(3))} B_n^k, \qquad \forall h \in L^2(\mathbb{T} \times \mathbb{I}).$$

Hence, again by the singular value decomposition of $\mathcal{V}$, we have

$$(\mathcal{V}\mathcal{V}^*)^{-1} \mathcal{V} f = \sum_{n=0}^{\infty} \sum_{\substack{k=-n \\ n+k \text{ even}}}^{n} \frac{1}{\mathrm{v}_n^k} \left\langle f, Y_n^k \right\rangle_{L^2(\mathbb{S}^2)} B_n^k. \tag{49}$$

We want to show that the right hand side of (49) converges uniformly on $C(\mathbb{T} \times \mathbb{I})$. Let $(\psi, t) \in \mathbb{T} \times \mathbb{I}$. As the Legendre polynomials satisfy $|P_n(t)| \le 1$ for all $t \in \mathbb{I}$, cf. [33, § 8.917], we have

$$\left| B_n^k(\psi, t) \right| = \left| \sqrt{\frac{2n+1}{4\pi}}\, P_n(t)\, \mathrm{e}^{\mathrm{i}k\psi} \right| \le \sqrt{\frac{2n+1}{4\pi}}.$$

Let $N \in \mathbb{N}$. Using the bound (19) on the singular values of $\mathcal{V}$, we see that there exists $C > 0$ such that

$$\begin{aligned}
&\left| \sum_{n=0}^{\infty} \sum_{\substack{k=-n \\ n+k \text{ even}}}^{n} \frac{1}{\mathrm{v}_n^k} \left\langle f, Y_n^k \right\rangle_{L^2(\mathbb{S}^2)} B_n^k(\psi, t) - \sum_{n=0}^{N-1} \sum_{\substack{k=-n \\ n+k \text{ even}}}^{n} \frac{1}{\mathrm{v}_n^k} \left\langle f, Y_n^k \right\rangle_{L^2(\mathbb{S}^2)} B_n^k(\psi, t) \right| \\
&\le C \sum_{n=N}^{\infty} \sum_{\substack{k=-n \\ n+k \text{ even}}}^{n} \left(n + \tfrac{1}{2}\right) \left| \left\langle f, Y_n^k \right\rangle_{L^2(\mathbb{S}^2)} \right| \\
&\le C \sqrt{\sum_{n=N}^{\infty} \sum_{\substack{k=-n \\ n+k \text{ even}}}^{n} \left(n + \tfrac{1}{2}\right)^{2s} \left| \langle f, Y_n^k \rangle_{L^2(\mathbb{S}^2)} \right|^2} \sqrt{\sum_{n=N}^{\infty} \sum_{k=-n}^{n} \left(n + \tfrac{1}{2}\right)^{2-2s}},
\end{aligned}$$

where the last line follows by the Cauchy–Schwarz inequality. In the last equation, the first root converges to zero for $N \to \infty$ since the Sobolev norm $\|f\|_{H^s(\mathbb{S}^2)}$ is finite, and the term under the second root,

$$\sum_{n=N}^{\infty} \sum_{k=-n}^{n} \left(n+\tfrac{1}{2}\right)^{2-2s} = 4 \sum_{n=N}^{\infty} \left(n+\tfrac{1}{2}\right)^{3-2s},$$

also converges since $s > 2$. Hence, the right-hand side of (49), whose summands are continuous themselves, converges uniformly to a continuous function on $\mathbb{T} \times \mathbb{I}$, which finally implies that g is continuous. ■

B. Proof of Theorem 4.3

1. First we show for $n \in \mathbb{N}_0$, $k \in \{-n, \dots, n\}$, and $\boldsymbol{Q} \in \mathrm{SO}(3)$ that

$$\mathcal{W}Y_n^k(\boldsymbol{Q}) = \sum_{j=-n}^{n} \lambda_n^j \, \overline{D_n^{k,j}(\boldsymbol{Q})}, \tag{50}$$

which implies (30). By (29), we have

$$\mathcal{W}Y_n^k(\boldsymbol{Q}) = \frac{1}{4\pi} \int_0^\pi Y_n^k(\boldsymbol{Q}\,\Phi(0,\vartheta)) \sin\vartheta \,\mathrm{d}\vartheta \overset{(14)}{=} \frac{1}{4\pi} \sum_{j=-n}^{n} D_n^{j,k}(\boldsymbol{Q}^\top) \int_0^\pi Y_n^j(\Phi(0,\vartheta)) \sin\vartheta \,\mathrm{d}\vartheta.$$

Noting that $D_n^{j,k}(\boldsymbol{Q}^\top) = \overline{D_n^{-k,-j}(\boldsymbol{Q})}$ by [85, § 4.4] and performing the substitution $z = \cos\vartheta$, we see that (50) holds with

$$\lambda_n^j = \frac{1}{4\pi} \int_{-1}^{1} Y_n^j(\Phi(0, \arccos z)) \,\mathrm{d}z \overset{(7)}{=} \frac{1}{4\pi} \sqrt{\frac{2n+1}{4\pi} \frac{(n-j)!}{(n+j)!}} \int_{-1}^{1} P_n^j(z) \,\mathrm{d}z. \tag{51}$$

If $n = 0$, then also $j = 0$ and we have $P_0^0 = 1$, which implies that $\lambda_0^0 = 2(4\pi)^{-1/2}$. Let $n \in \mathbb{N}$. If $j = 0$, then P_n^0 is the Legendre polynomial of degree n and thus we have $\int_{-1}^{1} P_n^0(z)\,\mathrm{d}z = 0$ for $n \geq 1$. If $n + j$ is odd, then P_n^j is an odd function and hence its integral (51) vanishes. Let us compute (51) for $n = j \in \mathbb{N}_0$. The substitution $z = \cos\vartheta$ and (8) yield

$$\begin{aligned} \int_{-1}^{1} P_n^n(z)\,\mathrm{d}z &= 2(-1)^n (2n-1)!! \int_0^1 (1-z^2)^{n/2} \,\mathrm{d}z \\ &= 2(-1)^n (2n-1)!! \int_0^{\pi/2} (\sin\vartheta)^{n+1} \,\mathrm{d}\vartheta \\ &= (-1)^n (2n-1)!! \frac{n!!}{(n+1)!!} \begin{cases} 2: & n \text{ even}, \\ \pi: & n \text{ odd}, \end{cases} \end{aligned} \tag{52}$$

where the last equality follows by [33, §3.621]. Let $n + j$ be even and $n \geq 2$, $j \geq 0$. We are going to use two recurrence relations from [33, §8.731]. First, we compute the integral of the relation

$$(n-j)P_n^j(z) = (z^2-1)\partial_z P_{n-1}^j(z) + nz P_{n-1}^j(z).$$

Using integration by parts and noting that $P_{n-1}^j(1) = -P_{n-1}^j(-1)$ yields

$$(n-j)\int_{-1}^{1} P_n^j(z)\,\mathrm{d}z = (n-2)\int_{-1}^{1} zP_{n-1}^j(z)\,\mathrm{d}z. \tag{53}$$

Second, inserting the integral of the recurrence relation

$$(2n-1)zP_{n-1}^j(z) = (n-j)P_n^j(z) + (n+j-1)P_{n-2}^j(z)$$

into (53) results in

$$(n-j)\int_{-1}^{1} P_n^j(z)\,\mathrm{d}z = \frac{n-2}{2n-1}\int_{-1}^{1}\Big((n-j)P_n^j(z) + (n+j-1)P_{n-2}^j(z)\Big)\,\mathrm{d}z$$

and thus

$$(n-j)(n+1)\int_{-1}^{1} P_n^j(z)\,\mathrm{d}z = (n-2)(n+j-1)\int_{-1}^{1} P_{n-2}^j(z)\,\mathrm{d}z.$$

Hence, we obtain by (52) that

$$\begin{aligned}\int_{-1}^{1} P_n^j(z)\,\mathrm{d}z &= \frac{(n-2)(n+j-1)}{(n-j)(n+1)}\int_{-1}^{1} P_{n-2}^j(z)\,\mathrm{d}z\\ &= \frac{(n-2)!!}{(j-2)!!}\,\frac{(n+j-1)!!}{(2j-1)!!}\,\frac{1}{(n-j)!!}\,\frac{(j+1)!!}{(n+1)!!}\int_{-1}^{1} P_j^j(z)\,\mathrm{d}z\\ &= (-1)^j\,\frac{j!!(n-2)!!(n+j-1)!!}{(j-2)!!(n-j)!!(n+1)!!}\begin{cases}2: & n \text{ even},\\ \pi: & n \text{ odd}.\end{cases}\end{aligned}$$

Together with (51) and (9), this implies (32).

2. We show that $\{Z_n^k : n\in\mathbb{N}_0, k=-n,\dots,n\}$ forms an orthonormal system in $L^2(\mathrm{SO}(3))$. The orthogonality follows from the orthogonality relation (15) of the rotational harmonics and the fact that Z_n^k for different indices $(n,k)\neq(n',k')$ contains disjoint linear combinations of rotational harmonics. Since $\lambda_n^n \neq 0$, this also yields that $\mathrm{w}_n \neq 0$ for any $n\in\mathbb{N}_0$, and hence the normalization follows by definition of Z_n^k.

3. We show the bound (33). Let $n\in\mathbb{N}_0$. The squared singular values of the operator $\mathcal{W}$ are

$$(\mathrm{w}_n)^2 = \left\|\mathcal{W}Y_n^k\right\|_{L^2(\mathbb{S}^2)}^2 = \sum_{j=-n}^{n}\left|\lambda_n^j\right|^2\left\|D_n^{k,j}\right\|_{L^2(\mathrm{SO}(3))}^2 \overset{(15)}{=} \sum_{j=1}^{n}\left|\lambda_n^j\right|^2\frac{16\pi^2}{2n+1}, \tag{54}$$

where we used the identity $|\lambda_n^j| = |\lambda_n^{-j}|$. We have by (32) that

$$\begin{aligned}(\mathrm{w}_n)^2 &= \frac{1}{4\pi}\sum_{\substack{j=1\\ n+j \text{ even}}}^{n}\frac{(n-j)!}{(n+j)!}\left(\frac{j\,(n-2)!!\,(n+j-1)!!}{(n-j)!!\,(n+1)!!}\right)^2\begin{cases}4: & n \text{ even},\\ \pi^2: & n \text{ odd}.\end{cases}\\ &= \frac{1}{4\pi}\left(\frac{(n-2)!!}{(n+1)!!}\right)^2\sum_{\substack{j=1\\ n+j \text{ even}}}^{n} j^2\,\frac{(n-j-1)!!}{(n-j)!!}\,\frac{(n+j-1)!!}{(n+j)!!}\begin{cases}4: & n \text{ even},\\ \pi^2: & n \text{ odd}.\end{cases}\end{aligned}$$

We use the fact from [41, p. 9] that

$$\frac{(2m-1)!!}{(2m)!!} = c_m (2m+1)^{-1/2}, \quad \text{where } \sqrt{\tfrac{2}{\pi}} \le c_m \le 1, \qquad \forall m \in \mathbb{N}.$$

We perform the proof for the case that $n = 2m$ is even, the case of odd n is completely analogous. We have

$$\begin{aligned}
(\mathrm{w}_{2m})^2 &= \frac{1}{\pi} \left(\frac{(2m-2)!!}{(2m+1)!!} \right)^2 \sum_{j=1}^{m} (2j)^2 \, \frac{(2m-2j-1)!!}{(2m-2j)!!} \, \frac{(2m+2j-1)!!}{(2m+2j)!!} \\
&= \frac{(2m+1)}{\pi\, c_m^2 \, (2m+1)^2 \, (2m)^2} \sum_{j=1}^{m} \frac{4j^2 \, c_{m-j} \, c_{m+j}}{\sqrt{2m-2j+1} \, \sqrt{2m+2j+1}} \\
&= \frac{1}{\pi\, c_m^2 \, (2m+1) \, (2m)^2} \sum_{j=1}^{m} \frac{4j^2 \, c_{m-j} \, c_{m+j}}{\sqrt{(2m+1)^2 - (2j)^2}}.
\end{aligned}$$

Taking into account the bounds on c_m and noting that the summands increase monotonic with j, we replace the sum by an integral plus the last summand and obtain the upper bound

$$\begin{aligned}
(\mathrm{w}_{2m})^2 &\le \frac{1}{8(2m+1)\, m^2} \left(\int_0^{m+1/2} \frac{(2x)^2}{\sqrt{(2m+1)^2 - (2x)^2}} \, \mathrm{d}x + \frac{4m^2}{\sqrt{4m+1}} \right) \\
&= \frac{1}{8(2m+1)\, m^2} \left(\frac{\pi (2m+1)^2}{8} + \frac{4m^2}{\sqrt{4m+1}} \right) \in \mathcal{O}(m^{-1}).
\end{aligned}$$

For the lower bound, we analogously see that

$$\begin{aligned}
(\mathrm{w}_{2m})^2 &\ge \frac{1}{2\pi^2 \, (2m+1)\, m^2} \sum_{j=1}^{m} \frac{4j^2}{\sqrt{(2m+1)^2 - (2j)^2}} \\
&\ge \frac{1}{2\pi^2 \, (2m+1)\, m^2} \int_1^m \frac{4j^2 \, \mathrm{d}j}{\sqrt{(2m+1)^2 - (2j)^2}} \\
&= \frac{1}{2\pi^2 \, (2m+1)\, m^2} \left(\frac{(2m+1)^2}{4} \arcsin\left(\frac{2m}{2m+1} \right) - \frac{m}{2} \sqrt{4m+1} \right)
\end{aligned}$$

can be bounded from below by a positive multiple of m^{-1} for $m \to \infty$. ■

C. Proof of Theorem 4.10

Let $\mu, \nu \in \mathcal{M}(\mathbb{S}^2)$ such that $\mathcal{W}\mu = \mathcal{W}\nu$. By Proposition 4.7, we have

$$\langle \mu, \mathcal{W}^* g \rangle = \langle \nu, \mathcal{W}^* g \rangle, \qquad \forall g \in C(\mathrm{SO}(3)).$$

The claim holds if we can show that $\{\mathcal{W}^* g : g \in C(\mathrm{SO}(3))\}$ is a dense subset of $C(\mathbb{S}^2)$. Let $f \in H^s(\mathbb{S}^2)$ with $s > 2$, cf. (10), which is dense in $C(\mathbb{S}^2)$, see [6, p. 121]. We show that $g \coloneqq (\mathcal{W}\mathcal{W}^*)^{-1} \mathcal{W} f \in C(\mathrm{SO}(3))$, which also implies $f = \mathcal{W}^* g$ by the injectivity of $\mathcal{W}$. We proceed analogously to the proof of Sobolev's embedding theorem [6, p. 122]. Since $\mathcal{W}^*$ has the same singular functions as $\mathcal{W}$ and the singular values $\mathrm{w}_n = \overline{\mathrm{w}_n}$, Theorem 4.10 implies

$$g = \sum_{n\in\mathbb{N}_0} \sum_{k=-n}^{n} \frac{1}{\mathrm{w}_n} \left\langle f, Y_n^k \right\rangle Z_n^k.$$

Let $\boldsymbol{Q} \in \mathrm{SO}(3)$. We have by (31) and the Cauchy–Schwarz inequality

$$\left|Z_n^k(\boldsymbol{Q})\right|^2 \le |\mathrm{w}_n|^{-2} \left(\sum_{j=-n}^{n} |\lambda_n^j| \left|D_n^{k,j}(\boldsymbol{Q})\right| \right)^2 \overset{(54)}{\le} \frac{2n+1}{16\pi^2} \sum_{j=-n}^{n} \left|D_n^{k,j}(\boldsymbol{Q})\right|^2. \tag{55}$$

Again by the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} |g(\boldsymbol{Q})|^2 &\le \left(\sum_{n\in\mathbb{N}_0} \sum_{k=-n}^{n} |\mathrm{w}_n|^{-1} \left|\left\langle f, Y_n^k \right\rangle\right| \left|Z_n^k(\boldsymbol{Q})\right| \right)^2 \\ &\le \sum_{n\in\mathbb{N}_0} \sum_{k=-n}^{n} (n+\tfrac{1}{2})^{2s} \left|\left\langle f, Y_n^k \right\rangle\right|^2 \sum_{n\in\mathbb{N}_0} \sum_{k=-n}^{n} (n+\tfrac{1}{2})^{-2s} |\mathrm{w}_n|^{-2} \left|Z_n^k(\boldsymbol{Q})\right|^2 \\ &\overset{(55)}{\le} \|f\|_{H^s(\mathbb{S}^2)}^2 \sum_{n\in\mathbb{N}_0} (n+\tfrac{1}{2})^{-2s} |\mathrm{w}_n|^{-2} \frac{2n+1}{16\pi^2} \sum_{k=-n}^{n} \sum_{j=-n}^{n} \left|D_n^{k,j}(\boldsymbol{Q})\right|^2. \end{aligned}$$

Using the addition formula $\sum_{j,k=-n}^{n} |D_n^{k,j}(\boldsymbol{Q})|^2 = n+1$, see [40, p. 17], and the bound (33) of w_n, we see that the last sum converges uniformly in $\boldsymbol{Q}$. Since the basis functions Z_n^k are continuous, this implies the continuity of g, which proves the assertion. ■

References

[1] A. Abouelaz and R. Daher. Sur la transformation de Radon de la sphère S^d. *Bull. Soc. Math. France*, 121(3):353–382, 1993. URL: http://eudml.org/doc/87670.

[2] M. L. Agranovsky and B. Rubin. On two families of Funk-type transforms. *Anal. Math. Phys.*, 10(4):44, 2020. doi:10.1007/s13324-020-00388-x.

[3] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011. doi:10.1137/100805741.

[4] F. Altekrüger, J. Hertrich, and G. Steidl. Neural Wasserstein gradient flows for maximum mean discrepancies with Riesz kernels. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 664–690. PMLR, 2023. URL: https://proceedings.mlr.press/v202/altekruger23a.html.

[5] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, 2005. doi:10.1007/b137080.

[6] K. Atkinson and W. Han. *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. Springer, Heidelberg, 2012. doi:10.1007/978-3-642-25983-8.

[7] F. A. Ba and M. Quellmalz. Accelerating the Sinkhorn algorithm for sparse multi-marginal optimal transport via fast Fourier transforms. *Algorithms*, 15(9):311, 2022. doi:10.3390/a15090311.

[8] Y. Bai, B. Schmitzer, M. Thorpe, and S. Kolouri. Sliced optimal partial transport. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13681–13690, 2023. URL: https://openaccess.thecvf.com/content/CVPR2023/html/Bai_Sliced_Optimal_Partial_Transport_CVPR_2023_paper.html.

[9] A. Beck. *First-order methods in optimization*. Number 25 in MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia; Mathematical Optimization Society, Philadelphia, 2017. doi:10.1137/1.9781611974997.ch1.

[10] F. Beier, R. Beinert, and G. Steidl. Multi-marginal Gromov–Wasserstein transport and barycenters, 2022. arXiv:2205.06725.

[11] F. Beier, R. Beinert, and G. Steidl. On a linear Gromov–Wasserstein distance. *IEEE Trans. Image Process.*, 31:7292–7305, 2022. doi:10.1109/TIP.2022.3221286.

[12] F. Beier, J. von Lindheim, S. Neumayer, and G. Steidl. Unbalanced multi-marginal optimal transport. *J. Math. Imaging. Vis.*, 2022. doi:10.1007/s10851-022-01126-7.

[13] R. Beinert, C. Heiss, and G. Steidl. On assignment problems related to Gromov–Wasserstein distances on the real line. *SIAM J. Imaging Sci.*, 16(2):1028–1032, 2023. doi:10.1137/22M1497808.

[14] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015. doi:10.1137/141000439.

[15] C. Bonet, P. Berg, N. Courty, F. Septier, L. Drumetz, and M.-T. Pham. Spherical sliced-Wasserstein. In *International Conference on Learning Representations*, 2023. URL: https://openreview.net/forum?id=jXQ0ipgMdU.

[16] N. Bonneel and D. Coeurjolly. SPOT: Sliced partial optimal transport. *ACM Trans. Graph.*, 38(4):89:1–13, 2019. doi:10.1145/3306346.3323021.

[17] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. Sliced and Radon Wasserstein barycenters of measures. *J. Math. Imaging Vis.*, 51(1):22–45, 2015. doi:10.1007/s10851-014-0506-3.

[18] A. Chambolle and T. Pock. An introduction to continuous optimization for imaging. *Acta Numer.*, 25:161–319, 2016. doi:10.1017/S096249291600009X.

[19] L. Chapel, M. Z. Alaya, and G. Gasso. Partial optimal tranport with applications on positive-unlabeled learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33,

pages 2903–2913. Curran Associates, 2020. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/1e6e25d952a0d639b676ee20d0519ee2-Paper.pdf.

[20] S. Cohen, A. Terenin, Y. Pitcan, B. Amos, M. P. Deisenroth, and K. S. S. Kumar. Sliced multi-marginal optimal transport, 2021. arXiv:2102.07115.

[21] L. Condat. Fast projection onto the simplex and the l_1 ball. *Math. Program.*, 158(1–2):575–585, 2016. doi:10.1007/s10107-015-0946-6.

[22] L. Cui, X. Qi, C. Wen, N. Lei, X. Li, M. Zhang, and X. Gu. Spherical optimal transportation. *Computer-Aided Design*, 115:181—193, 2019. doi:0.1016/j.cad.2019.05.024.

[23] J. Delon, J. Salomon, and A. Sobolevski. Fast transport optimization for Monge costs on the circle. *SIAM J. Appl. Math.*, 70(7):2239–2258, 2010. doi:10.1137/090772708.

[24] T. Dumont, T. Lacombe, and F.-X. Vialard. On the existence of Monge maps for the Gromov–Wasserstein problem, 2023. arXiv:2210.11945.

[25] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and Its Applications*. Kluwer, Dodrecht, 1996.

[26] J. Fan, Q. Zhang, A. Taghvaei, and Y. Chen. Variational Wasserstein gradient flow. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 6185–6215. PMLR, 2022.

[27] A. Figalli. The optimal partial transport problem. *Arch. Ration. Mech. Anal.*, 195(2):533–560, 2010. doi:10.1007/s00205-008-0212-7.

[28] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, R. Tavenard, A. Tong, and T. Vayer. POT: Python optimal transport. *J. Mach. Learn. Res.*, 22(78):1–8, 2021. URL: http://jmlr.org/papers/v22/20-451.html.

[29] P. Funk. Über Flächen mit lauter geschlossenen geodätischen Linien. *Math. Ann.*, 74(2):278–300, 1913. doi:10.1007/BF01456044.

[30] W. Gangbo and A. Święch. Optimal maps for the multidimensional Monge-Kantorovich problem. *Comm. Pure Appl. Math.*, 51(1):23–45, 1998. doi:10.1002/(SICI)1097-0312(199801)51:1<23::AID-CPA2>3.0.CO;2-H.

[31] I. Gel'fand and G. Shmelev. Geometric structures of double bundles and their relation to certain problems in integral geometry. *Funct. Anal. Its Appl.*, 17:84—96, 1983. doi:10.1007/BF01083134.

[32] S. Gindikin, J. Reeds, and L. Shepp. Spherical tomography and spherical integral geometry. In E. T. Quinto, M. Cheney, and P. Kuchment, editors, *Tomography, Impedance Imaging, and Integral Geometry*, volume 30 of *Lectures in Appl. Math*, pages 83–92. American Mathematical Society, 1994.

[33] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, New York, 8th edition, 2014. doi:10.1016/C2010-0-64839-5.

[34] M. Gräf and D. Potts. Sampling sets and quadrature formulae on the rotation group. *Numer. Funct. Anal. Optim.*, 30:665–688, 2009. doi:10.1080/01630560903163508.

[35] M. Gräf and D. Potts. On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms. *Numer. Math.*, 119:699–724, 2011. doi:10.1007/s00211-011-0399-7.

[36] H. Groemer. On a spherical integral transformation and sections of star bodies. *Monatsh. Math.*, 126(2):117–124, 1998. doi:10.1007/BF01473582.

[37] S. Guan, B. Liao, Y. Du, and X. Yin. Vehicle type recognition based on Radon-CDT hybrid transfer learning. In *10th International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2019. doi:10.1109/ICSESS47205.2019.9040687.

[38] B. Hamfeldt and A. Turnquist. A convergence framework for optimal transport on the sphere. *Numer. Math.*, 151:627—657, 2022. doi:10.1007/s00211-022-01292-1.

[39] S. Helgason. *Integral Geometry and Radon Transforms*. Springer, New York, 2011. doi:10.1007/978-1-4419-6055-9.

[40] R. Hielscher. The Radon Transform on the Rotation Group–Inversion and Application to Texture Analysis. Dissertation, Technische Universität Bergakademie Freiberg, 2007.

[41] R. Hielscher, D. Potts, and M. Quellmalz. An SVD in spherical surface wave tomography. In B. Hofmann, A. Leitao, and J. P. Zubelli, editors, *New Trends in Parameter Identification for Mathematical Models*, Trends in Mathematics, pages 121–144. Birkhäuser, 2018. doi:10.1007/978-3-319-70824-9_7.

[42] R. Hielscher and M. Quellmalz. Optimal mollifiers for spherical deconvolution. *Inverse Problems*, 31(8):085001, 2015. doi:10.1088/0266-5611/31/8/085001.

[43] R. Hielscher and M. Quellmalz. Reconstructing a function on the sphere from its means along vertical slices. *Inverse Probl. Imaging*, 10(3):711–739, 2016. doi:10.3934/ipi.2016018.

[44] J. Keiner, S. Kunis, and D. Potts. NFFT 3.5, C subroutine library. http://www.tu-chemnitz.de/~potts/nfft. Contributors: F. Bartel, M. Fenn, T. Görner, M. Kircheis, T. Knopp, M. Quellmalz, M. Schmischke, T. Volkmer, A. Vollrath.

[45] Y.-H. Kim and B. Pass. Wasserstein barycenters over Riemannian manifolds. *Adv. Math.*, 307:640–683, 2017. doi:10.1016/j.aim.2016.11.026.

[46] P. A. Knight. The Sinkhorn–Knopp algorithm: convergence and applications. *SIAM J. Matrix Anal. Appl.*, 30(1):261–275, 2008. doi:10.1137/060659624.

[47] S. Kolouri, K. Nadjahi, U. Simsekli, R. Badeau, and G. K. Rohde. Generalized sliced Wasserstein distances. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 2019.

[48] S. Kolouri, S. R. Park, and G. K. Rohde. The Radon cumulative distribution transform and its application to image classification. *IEEE Trans Image Process.*, 25(2):920–34, 2016. doi:10.1109/TIP.2015.2509419.

[49] A. Korotin, D. Selikhanovych, and E. Burnaev. Neural optimal transport. In *The Eleventh International Conference on Learning Representations*, 2023. URL: https://openreview.net/forum?id=d8CBRlWNkqH.

[50] S. Kunis and D. Potts. Fast spherical Fourier algorithms. *J. Comput. Appl. Math.*, 161:75–98, 2003. doi:10.1016/S0377-0427(03)00546-6.

[51] G. Loeper. Regularity of optimal maps on the sphere: the quadratic cost and the reflector antenna. *Arch. Rational Mech. Anal.*, 199(1):269–289, 2010. doi:10.1007/s00205-010-0330-x.

[52] A. K. Louis, M. Riplinger, M. Spiess, and E. Spodarev. Inversion algorithms for the spherical Radon and cosine transform. *Inverse Problems*, 27(3):035015, 2011. doi:10.1088/0266-5611/27/3/035015.

[53] E. Makai, H. Martini, and T. Ódor. On an integro-differential transform on the sphere. *Studia Sci. Math. Hungar.*, 38(1-4):299–312, 2001.

[54] A. T. T. McRae, C. J. Cotter, and C. J. Budd. Optimal-transport–based mesh adaptivity on the plane and sphere using finite elements. *SIAM J. Sci. Comput.*, 40(2):A1121–A1148, 2018. doi:10.1137/16M1109515.

[55] F. Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Found. Comput. Math.*, 11(4):417–487, 2011.

[56] V. Michel. *Lectures on Constructive Approximation: Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball.* Birkhäuser, New York, 2013. doi:10.1007/978-0-8176-8403-7.

[57] C. Moosmüller and A. Cloninger. Linear optimal transport embedding: provable Wasserstein classification for certain rigid transformations and perturbations. *Inf. Inference*, 12(1):363–389, 2022. doi:10.1093/imaiai/iaac023.

[58] K. Nguyen, T. Ren, H. Nguyen, L. Rout, T. M. Nguyen, and N. Ho. Hierarchical sliced Wasserstein distance. In *The Eleventh International Conference on Learning Representations*, 2023. URL: https://openreview.net/forum?id=CUOaVn6mYEj.

[59] V. P. Palamodov. Distributions and harmonic analysis. In *Commutative Harmonic Analysis III*, volume 72 of *Encyclopaedia Math. Sci.*, pages 1–127. Springer, Berlin, 1995.

[60] V. P. Palamodov. *Reconstruction from Integral Data.* Monographs and Research Notes in Mathematics. CRC Press, Boca Raton, FL, 2016. doi:10.1201/b19575.

[61] S. R. Park, S. Kolouri, S. Kundu, and G. K. Rohde. The cumulative distribution transform and linear pattern classification. *Appl. Comput. Harmon. Anal.*, 45(3):616–641, 2018. doi:10.1016/j.acha.2017.02.002.

[62] G. Peyré and M. Cuturi. Computational optimal transport. *Found. Trends Mach. Learn.*, 11(5-6):355–607, 2019. doi:10.1561/2200000073.

[63] G. Plonka, D. Potts, G. Steidl, and M. Tasche. *Numerical Fourier Analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser, Basel, 2018. doi:10.1007/978-3-030-04306-3.

[64] D. Potts, J. Prestin, and A. Vollrath. A fast algorithm for nonequispaced Fourier transforms on the rotation group. *Numer. Algorithms*, 52:355–384, 2009. doi:10.1007/s11075-009-9277-0.

[65] D. Potts, G. Steidl, and M. Tasche. Fast algorithms for discrete polynomial transforms. *Math. Comput.*, 67:1577–1590, 1998. URL: http://www.jstor.org/stable/2584863.

[66] M. Quellmalz. A generalization of the Funk–Radon transform. *Inverse Problems*, 33(3):035016, 2017. doi:10.1088/1361-6420/33/3/035016.

[67] M. Quellmalz. *Reconstructing Functions on the Sphere from Circular Means*. Dissertation. Universitätsverlag Chemnitz, 2019. URL: https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa2-384068.

[68] M. Quellmalz. The Funk-Radon transform for hyperplane sections through a common point. *Anal. Math. Phys.*, 10(38), 2020. doi:10.1007/s13324-020-00383-2.

[69] M. Quellmalz, R. Hielscher, and A. K. Louis. The cone-beam transform and spherical convolution operators. *Inverse Problems*, 34(10):105006, 2018. doi:10.1088/1361-6420/aad679.

[70] M. Quellmalz, L. Weissinger, S. Hubmer, and P. D. Erchinger. A frame decomposition of the Funk-Radon transform. In L. Calatroni, M. Donatelli, S. Morigi, M. Prato, and M. Santacesaria, editors, *Scale Space and Variational Methods in Computer Vision, SSVM 2023*, volume 14009 of *Lecture Notes in Computer Science*, pages 42–54, Cham, 2023. Springer. doi:10.1007/978-3-031-31975-4_4.

[71] J. Rabin, J. Delon, and Y. Gousseau. Transportation distances on the circle. *J. Math. Imaging Vis.*, 41:147–167, 2011. doi:10.1007/s10851-011-0284-0.

[72] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In A. Bruckstein, B. ter Haar Romeny, A. Bronstein, and M. Bronstein, editors, *Scale Space and Variational Methods in Computer Vision, SSVM 2011*, Lecture Notes in Computer Science, pages 435–446, Berlin, 2012. Springer. doi:10.1007/978-3-642-24785-9_37.

[73] J. Ren, W. R. Wolf, and X. Mao. Model reduction of traveling-wave problems via Radon cumulative distribution transform. *Phys. Rev. Fluids*, 6(8):L082501, 2021. doi:10.1103/PhysRevFluids.6.L082501.

[74] B. Rubin. Generalized Minkowski-Funk transforms and small denominators on the sphere. *Fract. Calc. Appl. Anal.*, 3(2):177–203, 2000.

[75] B. Rubin. The vertical slice transform on the unit sphere. *Fractional Calculus and Applied Analysis*, 22(4):899–917, 2019. doi:10.1515/fca-2019-0049.

[76] B. Rubin. On the spherical slice transform. *Anal. Appl.*, 20(3):483—497, 2022. doi:10.1142/S021953052150024X.

[77] R. M. Rustamov and S. Majumdar. Intrinsic sliced Wasserstein distances for comparing collections of probability distributions on manifolds and graphs, 2020. arXiv:2010.15285.

[78] Y. Salman. Recovering functions defined on the unit sphere by integration on a special family of sub-spheres. *Anal. Math. Phys.*, 7(2):165–185, 2017. doi:10.1007/s13324-016-0135-7.

[79] F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, Cham, 2015. doi:10.1007/978-3-319-20828-2.

[80] R. Schneider. Functions on a sphere with vanishing integrals over certain subspheres. *J. Math. Anal. Appl.*, 26:381–384, 1969. doi:10.1016/0022-247X(69)90160-7.

[81] M. Shifat-E-Rabbi, X. Yin, A. H. M. Rubaiyat, S. Li, S. Kolouri, A. Aldroubi, J. M. Nichols, and G. K. Rohde. Radon cumulative distribution transform subspace modeling for image classification. *J. Math. Imaging Vis.*, 63:1185–1203, 2021. doi:10.1007/s10851-021-01052-0.

[82] M. Staib, S. Claici, J. M. Solomon, and S. Jegelka. Parallel streaming Wasserstein barycenters. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017. URL: https://proceedings.neurips.cc/paper/2017/hash/253f7b5d921338af34da817c00f42753-Abstract.html.

[83] K.-T. Sturm. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. *arXiv:1208.0434*, 2020. arXiv:1208.0434.

[84] M. Theveneau and N. Keriven. Stability of entropic Wasserstein barycenters and application to random geometric graphs, 2022. arXiv:2210.10535.

[85] D. Varshalovich, A. Moskalev, and V. Khersonskii. *Quantum Theory of Angular Momentum*. World Scientific Publishing, Singapore, 1988. doi:10.1142/0270.

[86] T. Vayer, R. Flamary, N. Courty, R. Tavenard, and L. Chapel. Sliced Gromov–Wasserstein. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, 2019. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/a9cc6694dc40736d7a2ec018ea566113-Paper.pdf.

[87] C. Villani. *Topics in Optimal Transportation*. Number 58 in Graduate Studies in Mathematics. American Mathematical Society, Providence, 2003. doi:10.1090/gsm/058.

[88] H. Weller, P. Browne, C. Budd, and M. Cullen. Mesh adaptation on the sphere using optimal transport and the numerical solution of a Monge–Ampère type equation. *J. Comput. Physics*, 308:102–123, 2016. doi:10.1016/j.jcp.2015.12.018.

[89] G. Zangerl and O. Scherzer. Exact reconstruction in photoacoustic tomography with circular integrating detectors II: Spherical geometry. *Math. Methods Appl. Sci.*, 33(15):1771–1782, 2010. doi:10.1002/mma.1266.