# Parallelly Sliced Optimal Transport on Spheres and on the Rotation Group

Michael Quellmalz[1] Léo Buecher[1,2] Gabriele Steidl[1]

May 22, 2024

**Abstract**

Sliced optimal transport, which is basically a Radon transform followed by one-dimensional optimal transport, became popular in various applications due to its efficient computation. In this paper, we deal with sliced optimal transport on the sphere $\mathbb{S}^{d-1}$ and on the rotation group SO(3). We propose a parallel slicing procedure of the sphere which requires again only optimal transforms on the line. We analyze the properties of the corresponding parallelly sliced optimal transport, which provides in particular a rotationally invariant metric on the spherical probability measures. For SO(3), we introduce a new two-dimensional Radon transform and develop its singular value decomposition. Based on this, we propose a sliced optimal transport on SO(3).

As Wasserstein distances were extensively used in barycenter computations, we derive algorithms to compute the barycenters with respect to our new sliced Wasserstein distances and provide synthetic numerical examples on the 2-sphere that demonstrate their behavior for both the free and fixed support setting of discrete spherical measures. In terms of computational speed, they outperform the existing methods for semicircular slicing as well as the regularized Wasserstein barycenters.

## 1. Introduction

Optimal transport (OT) deals with the problem of finding the most efficient way to transport probability measures. The Wasserstein distance is a metric on the space of probability measures and has received much attention [55, 67, 75], e.g., for neural gradient flows [4, 26, 35, 45] in machine learning. Since OT on multi-dimensional domains is hard to compute, there exist different modifications that allow an efficient computation, such as the entropic regularization that yields the Sinkhorn algorithm [7, 21, 42, 55]. Sliced OT on Euclidean spaces utilizes the Radon transform to reduce the problem to the real line [16, 53, 62, 67], where OT possesses an analytic solution that can be computed efficiently. The notion of sliced OT can be generalized to other Radon-like transforms [43]. Specific geometries have been considered such as spheres [13, 58], manifolds with constant negative curvature [15], or separable Hilbert spaces [33].

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

[2]CentraleSupélec, Université Paris-Saclay, leo.buecher@student-cs.fr

In this paper, we are interested in sliced OT on special manifolds. A slicing approach on Riemannian manifolds based on eigenfunctions of the Laplacian was proposed in [66]. OT on the sphere has been intensely studied, e.g., the computation of Wasserstein barycenters [69, 70], the regularity of optimal maps [48], isometric rigidity of Wasserstein spaces [29] a connection with a Monge–Ampère type equation [32, 50, 78], or a variational framework [20]. Sliced OT was generalized to spheres in two different ways: Bonet et al. [14] introduced a slicing along semicircles to reduce the OT problem to one-dimensional circle, see Figure 1 right. This requires only OT computations on circles which was examined in [23]. The respective sliced Wasserstein distance is a metric in the space of probability measures on the 2-sphere [58]. Note that Radon transforms on such semicircles have been considered before in [30, 38], providing an extension of the Funk–Radon transform [28, 34, 39, 49, 60]. A second approach [58] of sliced spherical Wasserstein distances is based on the vertical slice transform [40, 64, 79]. This yields a family of measures on the unit interval instead of the circle and is therefore faster to compute than the first approach. However, this vertical slicing approach provides only a metric for even measures on the 2-sphere, i.e. the same values are taken on the upper and lower hemisphere. This is a serious restriction for practical applications.

In this paper, we generalize the vertically sliced OT from even measures to arbitrary probability measures by constructing a so-called parallelly sliced OT, see Figure 1 for an illustration. We provide a method for spheres $\mathbb{S}^{d-1}$ in general dimensions $d$. The key advantages are that the respective sliced Wasserstein distance is a rotationally invariant metric on the spherical probability measures, and that it is faster to compute than the semicircular sliced Wasserstein distance since we project on intervals instead of circles. Our numerical tests indicate a speedup between 40 and 100 times. In Theorem 3.7, we prove estimates between the spherical Wasserstein distance and its parallelly sliced version.

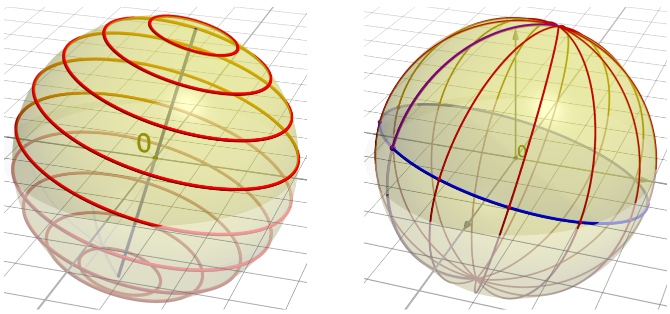


Figure 1: Our proposed parallel slicing (left): each red circle is projected to a single point on the blue line segment. The vertical slicing [58] can be regarded as special case that keeps the blue line direction fixed. Semicircular slicing from [13] (right): each red semicircle is projected to one point on the blue circle.

Furthermore, we consider OT in the group SO(3) of three-dimensional rotation matrices, which has applications in the synchronization of probability measures on rotations [11]. A Radon transform along one-dimensional geodesics of SO(3) was proposed in [37, 72], but, for the purpose of OT, we require slicing along two-dimensional submanifolds of SO(3). Therefore, we develop a new two-dimensional Radon transform on SO(3), including its singular value decomposition and adjoint operator. This paves the way to prove that the corresponding sliced Wasserstein distance fulfills the metric properties on the set of probability measures on SO(3).

Barycenter computations with respect to Wasserstein distances and their sliced variants are of increasing interest [16, 55, 62]. Therefore, we deal with barycenters with respect to our new sliced Wasserstein distances and describe their computation both in the free and fixed support discrete setting, as well as so-called Radon Wasserstein barycenter. As proof of the concept, we give numerical examples of barycenter computations on the 2-sphere. We compare our approach with the semicircular slicing [13] as well as the entropy-regularized Wasserstein barycenter computed with PythonOT [27].

**Outline of the paper.** We provide the basic preliminaries on OT and the manifolds $\mathbb{S}^{d-1}$ and SO(3) in Section 2. The parallel slice transform for functions and measures, and the corresponding parallelly sliced Wasserstein $p$-distances are introduced in Section 3. In Section 4, we generate sliced Wasserstein distances on SO(3) based on our new two-dimensional Radon transform on SO(3). Barycenter computations are examined in two different ways, namely sliced Wasserstein barycenters and Radon Wasserstein barycenters in Section 5. In the former case, we deal both with free and fixed discretization. In Section 6, we demonstrate by synthetic numerical examples the performance of our barycenter algorithms on the 2-sphere. In particular, we compare our parallel slicing approach with the slicing method of Bonet et al. [14]. Here some theoretically expected phenomena are illustrated. Finally, conclusions are drawn in Section 7. The appendix contains technical proofs.

# 2. Preliminaries

In this section, we provide the notation and necessary preliminaries on OT, in particular on the interval, and harmonic analysis on the unit sphere on $\mathbb{R}^d$.

## 2.1. Measures and OT

Let $\mathbb{X}$ be a compact Riemannian manifold with metric $d\colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, and let $\mathcal{B}(\mathbb{X})$ be the Borel $\sigma$-algebra induced by $d$. We denote by $\mathcal{M}(\mathbb{X})$ 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 the space of continuous functions $C(\mathbb{X})$. Let $\mathbb{Y}$ be another compact manifold and $T\colon \mathbb{X} \to \mathbb{Y}$ be measurable. For $\mu \in \mathcal{M}(\mathbb{X})$, we define the *push-forward measure* $T_{\#}\mu := \mu \circ T^{-1} \in \mathcal{M}(\mathbb{Y})$.

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

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

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

*Wasserstein barycenter* of $\mu_i \in \mathcal{P}^2(\mathbb{X})$, $i \in [\![M]\!] \coloneqq \{1, \dots, M\}$, is the minimizer

$$\operatorname{Bary}_X^{\mathrm{W}}(\mu_i, \lambda_i)_{i=1}^M \coloneqq \operatorname*{arg\,min}_{\nu \in \mathcal{P}(\mathbb{X})} \sum_{i=1}^M \lambda_i \, \mathrm{W}_2^2(\nu, \mu_i), \tag{2}$$

see [3]. The Wasserstein barycenter of absolutely continuous measures is unique [41].

**OT on the Interval** If $\mathbb{X}$ is the *unit interval* $\mathbb{I} \coloneqq [-1, 1]$ with the distance $d(x, y) = |x - y|$, the OT between two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{I})$ can be computed easily [55,67,75] 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* $F_\mu^{-1}(r) \coloneqq \min\{x \in \mathbb{I} \mid F_\mu(x) \geq r\}$, $r \in [0, 1]$, is non-decreasing and left continuous. The $p$-Wasserstein distance (1) between $\mu, \nu \in \mathcal{P}^p(\mathbb{I})$ now equals $\mathrm{W}_p(\mu, \nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L^p([0,1])}$. 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 OT plan $\pi$ 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 *OT map* $T^{\mu,\nu}$, the Wasserstein space $\mathcal{P}^p(\mathbb{I})$ can be isometrically embedded into $L^p_\omega(\mathbb{I})$ with $\omega \in \mathcal{P}_{\mathrm{ac}}(\mathbb{I})$ [8, 44, 54], where $L^p_\omega(\mathbb{I})$ consists of all $p$-integrable functions with respect to $\omega$. For a 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

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

The CDT is in fact a mapping from $\mathcal{P}^p(\mathbb{I})$ into the tangent space of $\mathcal{P}^p(\mathbb{I})$ at $\omega$, see [5, § 8.5]. Due to the relation to the OT 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 positive density functions $f_\mu$ and $f_\omega$, then, by the transformation formula for push-forward measures, $f_\mu$ can be recovered by

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

For $\mu_i \in \mathcal{P}(\mathbb{I})$ and an arbitrary reference $\omega \in \mathcal{P}_{\mathrm{ac}}(\mathbb{I})$, the Wasserstein barycenter (2) has the form [44]

$$\operatorname{Bary}_{\mathbb{I}}(\mu_i, \lambda_i)_{i=1}^M = \mathrm{CDT}_\omega^{-1}\left(\sum_{i=1}^M \lambda_i \, \mathrm{CDT}_\omega[\mu_i]\right). \tag{5}$$

**Sliced OT** Given another compact space $\mathbb{D}$ with a probability measure $u_{\mathbb{D}}$ and a slicing operator $\mathcal{S}_\psi \colon \mathbb{X} \to \mathbb{R}$ for all $\psi \in \mathbb{D}$, we define the *sliced $p$-Wasserstein distance*

$$\mathrm{SW}_p^p(\mu, \nu) \coloneqq \int_{\mathbb{D}} \mathrm{W}_p^p((\mathcal{S}_\psi)_\# \mu, (\mathcal{S}_\psi)_\# \nu) \, \mathrm{d}u_{\mathbb{D}}(\psi). \tag{6}$$

Sliced Wasserstein distances on the Euclidean space $\mathbb{X} = \mathbb{R}^d$ with the slicing operator $\mathcal{S}_{\boldsymbol{\psi}}^{\mathbb{R}^d} \coloneqq \langle \boldsymbol{\psi}, \cdot \rangle$ for $\boldsymbol{\psi} \in \mathbb{D} = \mathbb{S}^{d-1}$ are well known [16, 62, 67]. Sliced OT is closely related with the Radon transform

$$\mathcal{R}_{\boldsymbol{\psi}} \colon \mathcal{P}(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}), \quad \mu \mapsto (\mathcal{S}_{\boldsymbol{\psi}}^{\mathbb{R}^d})_\# \mu.$$

The Radon transform is often defined for functions on $\mathbb{R}^d$ via an integral, see [52].

## 2.2. Sphere

Let $d \in \mathbb{N}$ with $d \geq 3$. We define the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$ by

$$\mathbb{S}^{d-1} \coloneqq \{\boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| = 1\},$$

and denote the canonical unit vectors by $\boldsymbol{e}^j \in \mathbb{R}^d$ for $j \in [\![d]\!] \coloneqq \{1, \dots, d\}$. The geodesic distance on the sphere $\mathbb{S}^{d-1}$ reads as

$$d(\boldsymbol{\xi}, \boldsymbol{\eta}) \coloneqq \arccos(\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle), \qquad \forall \boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{S}^{d-1}, \tag{7}$$

and we denote the volume of $\mathbb{S}^{d-1}$ by

$$\left|\mathbb{S}^{d-1}\right| \coloneqq \int_{\mathbb{S}^{d-1}} \mathrm{d}\sigma_{\mathbb{S}^{d-1}} = \frac{2\pi^{d/2}}{\Gamma(d/2)}, \tag{8}$$

where $\sigma_{\mathbb{S}^{d-1}}$ is the surface measure on $\mathbb{S}^{d-1}$. Normalizing $\sigma_{\mathbb{S}^{d-1}}$ yields the *uniform measure* $u_{\mathbb{S}^{d-1}} \coloneqq |\mathbb{S}^{d-1}|^{-1} \sigma_{\mathbb{S}^{d-1}}$. We can write any vector $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$ as

$$\boldsymbol{\xi} = \begin{pmatrix} \sqrt{1-t^2}\, \boldsymbol{\eta} \\ t \end{pmatrix} \qquad \text{for} \quad \boldsymbol{\eta} \in \mathbb{S}^{d-2},\ t \in \mathbb{I},$$

then the surface measure on the sphere $\mathbb{S}^{d-1}$ decomposes as [6, (1.16)]

$$\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\xi}) = \mathrm{d}\sigma_{\mathbb{S}^{d-2}}(\boldsymbol{\eta})\, (1-t^2)^{\frac{d-3}{2}}\, \mathrm{d}t. \tag{9}$$

For $\mathbb{S}^2$, we denote the bijective spherical coordinate transform by

$$\Phi\colon [0, 2\pi) \times (0, \pi) \cup \{0\} \times \{0, \pi\} \to \mathbb{S}^2, \quad (\varphi, \theta) \mapsto (\cos\varphi \sin\theta, \sin\varphi \sin\theta, \cos\theta)^\top. \tag{10}$$

**Spherical harmonics** Let $n \in \mathbb{N}_0$. We denote by $\mathbb{Y}_{n,d}$ the space of all polynomials $f\colon \mathbb{R}^d \to \mathbb{C}$ which are harmonic, i.e., the Laplacian $\Delta f$ vanishes everywhere, and homogeneous of degree $n$, i.e., $f(\alpha \boldsymbol{x}) = \alpha^n f(\boldsymbol{x})$ for all $\alpha \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^d$, restricted to the sphere $\mathbb{S}^{d-1}$. Setting

$$N_{n,d} \coloneqq \dim(\mathbb{Y}_{n,d}) = \frac{(2n+d-2)\,(n+d-3)!}{n!\,(d-2)!}, \tag{11}$$

we call an orthonormal basis $\{Y^k_{n,d} \mid k \in [\![N_{n,d}]\!]\}$ of $\mathbb{Y}_{n,d}$ a basis of *spherical harmonics* on $\mathbb{S}^{d-1}$ of degree $n$, cf. [6]. Then $\{Y^k_{n,d} \mid n \in \mathbb{N}_0,\, k \in [\![N_{n,d}]\!]\}$ forms an orthonormal basis of $L^2(\mathbb{S}^{d-1})$. In particular, we can write any $f \in L^2(\mathbb{S}^{d-1})$ as spherical Fourier series

$$f = \sum_{n=0}^{\infty} \sum_{k=1}^{N_{n,d}} \hat{f}^k_{n,d}\, Y^k_{n,d}, \quad \text{where} \quad \hat{f}^k_{n,d} \coloneqq \left\langle f, Y^k_{n,d} \right\rangle_{L^2(\mathbb{S}^{d-1})}.$$

The *Legendre polynomial* $P_{n,d}$ of degree $n \in \mathbb{N}_0$ in dimension $d \geq 2$ is given by [6, (2.70)]

$$P_{n,d}(t) \coloneqq (-1)^n \frac{(d-3)!!}{(2n+d-3)!!} (1-t^2)^{\frac{3-d}{2}} \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^n (1-t^2)^{n+\frac{d-3}{2}}, \qquad t \in [-1, 1].$$

Up to normalization, the Legendre polynomials are equal to the Gegenbauer or ultraspherical polynomials, see [6, (2.145)]. The normalized Legendre polynomials

$$\widetilde{P}_{n,d}(t) \coloneqq \sqrt{\frac{N_{n,d}\,|\mathbb{S}^{d-2}|}{|\mathbb{S}^{d-1}|}}\, P_{n,d}(t) = \frac{\sqrt{(2n+d-2)\,(n+d-3)!}}{2^{(d-2)/2}\,\sqrt{n!}\,\Gamma(\frac{d-1}{2})}\, P_{n,d}(t) \tag{12}$$

satisfy the orthonormality relation $\int_{-1}^{1} \widetilde{P}_{n,d}(t)\,\overline{\widetilde{P}_{m,d}(t)}\,(1-t^2)^{\frac{d-3}{2}}\,\mathrm{d}t = \delta_{n,m}$, where $\delta$ denotes the Kronecker symbol.

## 2.3. Rotation Group

We define the *rotation group*

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

We are especially interested in the 3D rotation group SO(3). Every rotation matrix can be written as the rotation around an axis $\boldsymbol{n} \in \mathbb{S}^2$ with an angle $\omega \in \mathbb{T} \coloneqq \mathbb{R}/(2\pi\mathbb{Z})$ , i.e.,

$$\mathrm{R}_{\boldsymbol{n}}(\omega) \coloneqq (1-\cos\omega)\,\boldsymbol{n}\boldsymbol{n}^\top + \begin{pmatrix} \cos\omega & -n_3\sin\omega & n_2\sin\omega \\ n_3\sin\omega & \cos\omega & n_1\sin\omega \\ -n_2\sin\omega & n_1\sin\omega & \cos\omega \end{pmatrix} \in \mathrm{SO}(3). \tag{13}$$

Furthermore, we consider the *Euler angle* parameterization

$$\Psi\colon \mathbb{T}\times[0,\pi]\times\mathbb{T} \to \mathrm{SO}(3), \quad \Psi(\alpha,\beta,\gamma) \coloneqq \mathrm{R}_{\boldsymbol{e}^3}(\alpha)\mathrm{R}_{\boldsymbol{e}^2}(\beta)\mathrm{R}_{\boldsymbol{e}^3}(\gamma). \tag{14}$$

The rotationally invariant Lebesgue measure $\sigma_{\mathrm{SO}(3)}$ on SO(3) is given by

$$\int_{\mathrm{SO}(3)} f(\boldsymbol{Q})\,\mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{Q}) = 2\int_0^\pi \int_{\mathbb{S}^2} f(\mathrm{R}_{\boldsymbol{\xi}}(\omega))(1-\cos(\omega))\,\mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi})\,\mathrm{d}\omega, \tag{15}$$

see [36, p. 8], and 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

$$D_n^{k,j}(\Psi(\alpha,\beta,\gamma)) \coloneqq \mathrm{e}^{-\mathrm{i}k\alpha}\, d_n^{k,j}(\cos\beta)\,\mathrm{e}^{-\mathrm{i}j\gamma}, \tag{16}$$

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 [73, chap. 4]. The rotational harmonics satisfy the orthogonality relations

$$\int_{\mathrm{SO}(3)} D_n^{j,k}(\boldsymbol{Q}) D_{n'}^{j',k'}(\boldsymbol{Q})\,\mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{Q}) = \frac{8\pi^2}{2n+1}\delta_{n,n'}\delta_{k,k'}\delta_{j,j'} \tag{17}$$

and

$$\int_0^\pi d_n^{j,k}(\beta) d_{n'}^{j,k}(\beta)\,\sin(\beta)\,\mathrm{d}\beta = \frac{2}{2n+1}\delta_{n,n'} \tag{18}$$

for all $n, n' \in \mathbb{N}_0$, $j, k = -n,\dots,n$, and $j', k' = -n',\dots,n'$. The normalized rotational harmonics

$$\widetilde{D}_n^{j,k} \coloneqq \sqrt{\frac{2n+1}{8\pi^2}}\, D_n^{j,k}, \qquad n \in \mathbb{N}_0,\ j,k = -n,\dots,n, \tag{19}$$

form an orthonormal basis of $L^2(\mathrm{SO}(3))$.

# 3. Sliced OT on the Sphere

We present a slicing approach for OT on the sphere $\mathbb{X} = \mathbb{S}^{d-1}$ for $d \geq 3$. We define the *parallel slicing operator* for a fixed $\boldsymbol{\psi} \in \mathbb{S}^{d-1}$ by

$$\mathcal{S}_{\boldsymbol{\psi}}^{\mathbb{S}^{d-1}} : \mathbb{S}^{d-1} \to \mathbb{I}, \qquad \mathcal{S}_{\boldsymbol{\psi}}^{\mathbb{S}^{d-1}}(\boldsymbol{\xi}) \coloneqq \langle \boldsymbol{\xi}, \boldsymbol{\psi} \rangle . \tag{20}$$

We will omit the superscript of $\mathcal{S}_{\boldsymbol{\psi}}$ if no confusion arises. The corresponding slice is the $(d-2)$-dimensional *subsphere*

$$C_{\boldsymbol{\psi}}^t \coloneqq \mathcal{S}_{\boldsymbol{\psi}}^{-1}(t) = \{\boldsymbol{\xi} \in \mathbb{S}^{d-1} \mid \mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{\xi}) = t\} = \{\boldsymbol{\xi} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{\psi}, \boldsymbol{\xi} \rangle = t\}, \qquad t \in \mathbb{I}, \tag{21}$$

which is the intersection of $\mathbb{S}^{d-1}$ and the hyperplane of $\mathbb{R}^d$ with normal $\boldsymbol{\psi}$ and distance $t$ from the origin.

In this section, we first analyze the respective Radon transform for functions and measures on $\mathbb{S}^{d-1}$, and then show that the sliced Wasserstein distance is a metric on $\mathcal{P}(\mathbb{S}^{d-1})$.

## 3.1. Parallel Slice Transform of Functions

For $f \colon \mathbb{S}^{d-1} \to \mathbb{R}$, we define the *parallel slice transform*

$$\mathcal{U}f(\boldsymbol{\psi}, t) \coloneqq \begin{cases} \dfrac{1}{|\mathbb{S}^{d-1}| \sqrt{1-t^2}} \displaystyle\int_{C_{\boldsymbol{\psi}}^t} f(\boldsymbol{\xi}) \, \mathrm{ds}(\boldsymbol{\xi}), & \boldsymbol{\psi} \in \mathbb{S}^{d-1},\ t \in (-1,1), \\ \dfrac{1}{2}\, \delta_{d,3}\, f(\pm\boldsymbol{\psi}), & \boldsymbol{\psi} \in \mathbb{S}^{d-1},\ t = \pm 1, \end{cases} \tag{22}$$

where ds denotes the $(d-2)$-dimensional volume element on $C_{\boldsymbol{\psi}}^t$ and $\delta$ is the Kronecker symbol. The second line in (22) ensures the continuity of $\mathcal{U}f$ if $f$ is continuous. For fixed $\boldsymbol{\psi} \in \mathbb{S}^{d-1}$, the (normalized) restriction

$$\mathcal{U}_{\boldsymbol{\psi}} f \coloneqq |\mathbb{S}^{d-1}| \, \mathcal{U}f(\boldsymbol{\psi}, \cdot)$$

belongs to the class of convolution operators [59], and is known as the spherical section transform [63], translation [22] or shift operator [65]. The chosen normalization will become clear in Proposition 3.5. The following proposition was shown, e.g., in [57, Cor. 3.3].

**Proposition 3.1** (Integration in $t$)**.** *For every $f \in L^1(\mathbb{S}^{d-1})$ and $\boldsymbol{\psi} \in \mathbb{S}^{d-1}$, we have $\mathcal{U}_{\boldsymbol{\psi}} \in L^1(\mathbb{I})$ and*

$$\int_{\mathbb{I}} \mathcal{U}_{\boldsymbol{\psi}} f(t) \, \mathrm{d}t = \int_{\mathbb{S}^{d-1}} f(\boldsymbol{\xi}) \, \mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\xi}). \tag{23}$$

**Theorem 3.2** (Positivity)**.** *Let $f \in C(\mathbb{S}^{d-1})$. Then we have $f(\boldsymbol{\xi}) \geq 0$ for all $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$ if and only if $\mathcal{U}f(\boldsymbol{\psi}, t) \geq 0$ for all $\boldsymbol{\psi} \in \mathbb{S}^{d-1}$ and $t \in \mathbb{I}$.*

*Proof.* Let $f \in C(\mathbb{S}^{d-1})$. If $f \geq 0$ everywhere, then $\mathcal{U}f$ is non-negative everywhere as the integral of a non-negative function. Conversely, let $\boldsymbol{\eta} \in \mathbb{S}^{d-1}$ such that $f(\boldsymbol{\eta}) = -\delta < 0$. By continuity, there exists $\varepsilon > 0$ such that $f(\boldsymbol{\xi}) < -\delta/2$ for all $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$ with $d(\boldsymbol{\xi}, \boldsymbol{\eta}) \leq \varepsilon$. Because all points in $C_{\boldsymbol{\eta}}^{\cos(\varepsilon)}$ have spherical distance $\varepsilon$ to $\boldsymbol{\eta}$, we conclude that

$$\mathcal{U}f(\boldsymbol{\eta}, \cos(\varepsilon)) = \frac{1}{|\mathbb{S}^{d-1}| \sin(\varepsilon)} \int_{C_{\boldsymbol{\eta}}^{\cos(\varepsilon)}} f(\boldsymbol{\xi}) \, \mathrm{ds}(\boldsymbol{\xi}) < 0. \qquad \square$$

There is no analogue to Theorem 3.2 for the vertical slice nor for the semicircle transform considered in [14, 58] since one can always construct a function that is negative on a small ball and positive outside such that either transform is non-negative everywhere.

**Theorem 3.3** (Singular value decomposition)**.** *For each $n \in \mathbb{N}_0$, let $\{Y_{n,d}^k \mid k \in [\![N_{n,d}]\!]\}$ be an orthonormal basis of $\mathbb{Y}_{n,d}$. Then (22) is a compact operator $\mathcal{U}\colon L^2(\mathbb{S}^{d-1}) \to L^2_{w_d}(\mathbb{S}^{d-1}\times\mathbb{I})$, where $L^2_{w_d}(\mathbb{S}^{d-1}\times\mathbb{I})$ is the space of square integrable functions with the weighted norm $g \mapsto (\int_{\mathbb{S}^{d-1}\times\mathbb{I}} g(\boldsymbol{\xi},t)(1-t^2)^{(3-d)/2}\,\mathrm{d}(\boldsymbol{\xi},t))^{1/2}$, with the singular value decomposition*

$$\mathcal{U}Y_{n,d}^k(\boldsymbol{\psi},t) = \lambda_{n,d}^{\mathcal{U}}\, Y_{n,d}^k(\boldsymbol{\psi})\, \widetilde{P}_{n,d}(t)\, (1-t^2)^{\frac{d-3}{2}}, \qquad \forall n \in \mathbb{N}_0,\ k \in [\![N_{n,d}]\!], \tag{24}$$

*where $\widetilde{P}_{n,d}$ are given in (12) and the singular values are*

$$\lambda_{n,d}^{\mathcal{U}} \coloneqq \frac{2^{\frac{d}{2}}\,\sqrt[4]{\pi}\,\sqrt{n!}\,\Gamma(\frac{d}{2})}{\sqrt{(2n+d-2)\,(n+d-3)!}}. \tag{25}$$

*Proof.* The theorem basically follows from the generalized Funk–Hecke formula [10, (4.2.10)], which states for $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$, $t \in (-1,1)$, and $Y_{n,d} \in \mathbb{Y}_{n,d}(\mathbb{S}^{d-1})$ that

$$\frac{1}{|\mathbb{S}^{d-2}|\,(1-t^2)^{\frac{d-2}{2}}} \int_{\langle \boldsymbol{\xi},\boldsymbol{\eta}\rangle = t} Y_{n,d}(\boldsymbol{\eta})\,\mathrm{d}\sigma_{\mathbb{S}^{d-2}}(\boldsymbol{\eta}) = \frac{P_{n,d}(t)}{P_{n,d}(1)}\, Y_{n,d}(\boldsymbol{\xi}).$$

Inserting the normalization (12) yields (24) with

$$\lambda_{n,d}^{\mathcal{U}} = \sqrt{\frac{|\mathbb{S}^{d-2}|}{|\mathbb{S}^{d-1}|\,N_{n,d}}} \overset{(8),(11)}{=} \sqrt{\frac{\pi\,\Gamma(\frac{d}{2})\,n!\,(d-2)!}{\Gamma(\frac{d-1}{2})\,(2n+d-2)\,(n+d-3)!}}.$$

The Legendre duplication formula $\Gamma(\frac{d-1}{2})\,\Gamma(\frac{d}{2}) = 2^{-d}\sqrt{\pi}\,\Gamma(d-1)$ yields (25). Expanding the product (11) asymptotically for $n \to \infty$, we have

$$N_{n,d} = (2n+d-2)\,\frac{(n+d-3)\,(n+d-2)\cdots(n+1)}{(d-2)!} = \frac{2}{(d-2)!}\left(n^{d-2} + \mathcal{O}(n^{d-3})\right)$$

and hence the singular values $\lambda_{n,d}^{\mathcal{U}}$ converge to zero. Together with the orthonormality of (12), we deduce that (24) is a singular value decomposition. □

**Theorem 3.4** (Adjoint)**.** *Let $1 \le p,q \le \infty$ with $1/p + 1/q = 1$. For $1 \le p < \infty$, the adjoint $\mathcal{U}^*\colon L^q(\mathbb{S}^{d-1}\times\mathbb{I}) \to L^q(\mathbb{S}^{d-1})$ of $\mathcal{U}\colon L^p(\mathbb{S}^{d-1}) \to L^p(\mathbb{S}^{d-1}\times\mathbb{I})$ is given by*

$$\mathcal{U}^* g(\boldsymbol{\xi}) = \frac{1}{|\mathbb{S}^{d-1}|} \int_{\mathbb{S}^{d-1}} g(\langle \boldsymbol{\xi},\boldsymbol{\psi}\rangle)\,\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\psi}), \tag{26}$$

*and the adjoint $\mathcal{U}_{\boldsymbol{\psi}}^*\colon L^q(\mathbb{I}) \to L^q(\mathbb{S}^{d-1})$ of $\mathcal{U}_{\boldsymbol{\psi}}\colon L^p(\mathbb{S}^{d-1}) \to L^p(\mathbb{I})$ by*

$$\mathcal{U}_{\boldsymbol{\psi}}^* g(\boldsymbol{\xi}) = g(\langle \boldsymbol{\xi},\boldsymbol{\psi}\rangle) \tag{27}$$

*for all $\boldsymbol{\xi} \in \mathbb{S}^{d-1}$. Both adjoint operators map continuous functions to continuous functions.*

*Proof.* Let $f \in L^p(\mathbb{S}^{d-1})$ and $g \in L^q(\mathbb{S}^{d-1} \times \mathbb{I})$. We have

$$\begin{aligned}
\langle \mathcal{U} f, g\rangle &= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{I}} \mathcal{U} f(\boldsymbol{\psi}, t)\, g(\boldsymbol{\psi}, t) \,\mathrm{d}t \,\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\psi}) \\
&\overset{(22)}{=} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{I}} \frac{1}{|\mathbb{S}^{d-1}| \sqrt{1-t^2}} \int_{C_{\boldsymbol{\psi}}^t} f(\boldsymbol{\xi})\, g(\boldsymbol{\psi}, t) \,\mathrm{d}s(\boldsymbol{\xi}) \,\mathrm{d}t \,\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\psi}) \\
&\overset{(21)}{=} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{I}} \frac{1}{|\mathbb{S}^{d-1}| \sqrt{1-t^2}} \int_{C_{\boldsymbol{\psi}}^t} f(\boldsymbol{\xi})\, g(\boldsymbol{\psi}, \langle \boldsymbol{\psi}, \boldsymbol{\xi}\rangle) \,\mathrm{d}s(\boldsymbol{\xi}) \,\mathrm{d}t \,\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\psi}) \\
&\overset{(23)}{=} \frac{1}{|\mathbb{S}^{d-1}|} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} f(\boldsymbol{\xi})\, g(\boldsymbol{\psi}, \langle \boldsymbol{\psi}, \boldsymbol{\xi}\rangle) \,\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\xi}) \,\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\psi}).
\end{aligned}$$

The adjoint of $\mathcal{U}_{\boldsymbol{\psi}}$ can be established analogously without the outer integral. The continuity follows from Lebesgue's dominated convergence theorem. □

### 3.2. Parallel Slice Transform of Measures

We extend the definition (22) to measures as pushforward of the slicing operator (20). For $\boldsymbol{\psi} \in \mathbb{S}^{d-1}$, we define

$$\mathcal{U}_{\boldsymbol{\psi}} \colon \mathcal{M}(\mathbb{S}^{d-1}) \to \mathcal{M}(\mathbb{I}), \quad \mu \mapsto (\mathcal{S}_{\boldsymbol{\psi}})_{\#}\mu = \mu \circ \mathcal{S}_{\boldsymbol{\psi}}^{-1}. \tag{28}$$

and $\mathcal{U} \colon \mathcal{M}(\mathbb{S}^2) \to \mathcal{M}(\mathbb{T} \times \mathbb{I})$ by

$$\mathcal{U}\mu \coloneqq T_{\#}(u_{\mathbb{S}^{d-1}} \times \mu) \quad \text{with} \quad T(\boldsymbol{\psi}, \boldsymbol{\xi}) \coloneqq (\boldsymbol{\psi}, \mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{\xi})). \tag{29}$$

**Proposition 3.5** (Connection with adjoint)**.** *Let $\mu \in \mathcal{M}(\mathbb{S}^{d-1})$. The transforms (29) and (28) satisfy*

$$\begin{aligned}
\langle \mathcal{U}\mu, g\rangle &= \langle \mu, \mathcal{U}^* g\rangle \quad \textit{for all } g \in C(\mathbb{S}^{d-1} \times \mathbb{I}) \quad \textit{and} \\
\langle \mathcal{U}_{\boldsymbol{\psi}}\mu, g\rangle &= \langle \mu, \mathcal{U}_{\boldsymbol{\psi}}^* g\rangle \quad \textit{for all } g \in C(\mathbb{I}),\ \boldsymbol{\psi} \in \mathbb{S}^{d-1}
\end{aligned}$$

*with the adjoint operators from (26) and (27).*

*Proof.* For $g \in C(\mathbb{S}^{d-1} \times \mathbb{I})$, we have by the definition in (29)

$$\begin{aligned}
\langle \mathcal{U}\mu, g\rangle &= \int_{\mathbb{S}^{d-1} \times \mathbb{I}} g(\boldsymbol{\psi}, t) \,\mathrm{d}T_{\#}(u_{\mathbb{S}^{d-1}} \times \mu)(\boldsymbol{\psi}, t) \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} g(\boldsymbol{\psi}, \langle \boldsymbol{\psi}, \boldsymbol{\xi}\rangle) \,\mathrm{d}u_{\mathbb{S}^{d-1}}(\boldsymbol{\psi}) \,\mathrm{d}\mu(\boldsymbol{\xi}) = \langle \mu, \mathcal{U}^* g\rangle.
\end{aligned}$$

For $g \in C(\mathbb{I})$, and fixed $\boldsymbol{\psi} \in \mathbb{S}^{d-1}$,

$$\langle \mathcal{U}_{\boldsymbol{\psi}}\mu, g\rangle = \int_{\mathbb{I}} g(t) \,\mathrm{d}(\mathcal{S}_{\boldsymbol{\psi}})_{\#}\mu(t) = \int_{\mathbb{S}^{d-1}} g(\langle \boldsymbol{\psi}, \boldsymbol{\xi}\rangle) \,\mathrm{d}\mu(\boldsymbol{\xi}) = \langle \mu, \mathcal{U}_{\boldsymbol{\psi}}^* g\rangle. \qquad \square$$

The last proposition provides an alternative way of defining $\mathcal{U}$ for measures, similarly to what was done for the Radon transform, e.g. in [12].

For absolutely continuous measures with respect to the surface measure $\sigma_{\mathbb{S}^{d-1}}$, the measure- and function-valued transforms coincide.

**Proposition 3.6** (Absolutely continuous measures)**.** *For $f \in L^1(\mathbb{S}^{d-1})$, we have*

$$\mathcal{U}[f\sigma_{\mathbb{S}^{d-1}}] = (\mathcal{U}f)\,\sigma_{\mathbb{S}^{d-1}\times\mathbb{I}} \quad \textit{and} \quad \mathcal{U}_{\boldsymbol{\psi}}[f\sigma_{\mathbb{S}^{d-1}}] = (\mathcal{U}_{\boldsymbol{\psi}}f)\,\sigma_{\mathbb{I}} \qquad \forall \boldsymbol{\psi} \in \mathbb{S}^{d-1}.$$

*In particular, the transformed measures are again absolutely continuous.*

*Proof.* Let $g \in C(\mathbb{T}\times\mathbb{I})$. The first identity follows from Proposition 3.5 by

$$\begin{aligned}\langle \mathcal{U}[f\sigma_{\mathbb{S}^{d-1}}], g\rangle = \langle f\sigma_{\mathbb{S}^{d-1}}, \mathcal{U}^* g\rangle &= \int_{\mathbb{S}^{d-1}} f(\boldsymbol{\xi})\,\mathcal{U}^* g(\boldsymbol{\xi})\,\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\xi}) \\ &= \int_{-1}^{1}\int_{\mathbb{S}^{d-1}} \mathcal{U}f(\boldsymbol{\psi},t)\,g(\boldsymbol{\psi},t)\,\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\boldsymbol{\psi})\,\mathrm{d}t = \langle (\mathcal{U}f)\,\sigma_{\mathbb{T}\times\mathbb{I}}, g\rangle.\end{aligned}$$

The identity for $\mathcal{U}_{\boldsymbol{\psi}}$ follows analogously. □

### 3.3. Spherical Sliced Wasserstein Distance

For $p \in [1,\infty)$ and $\mu,\nu \in \mathcal{P}(\mathbb{S}^{d-1})$, the *parallel-sliced spherical Wasserstein distance*

$$\mathrm{PSW}_p^p(\mu,\nu) \coloneqq \int_{\mathbb{S}^{d-1}} \mathrm{W}_p^p(\mathcal{U}_{\boldsymbol{\psi}}\mu, \mathcal{U}_{\boldsymbol{\psi}}\nu)\,\mathrm{d}u_{\mathbb{S}^{d-1}}(\boldsymbol{\psi}) \tag{30}$$

is the mean value of Wasserstein distances on the unit interval $\mathbb{I}$. Since the geodesic distance (7) on the sphere is rotationally invariant, i.e., $d(\boldsymbol{Q}\boldsymbol{\xi},\boldsymbol{Q}\boldsymbol{\eta}) = d(\boldsymbol{\xi},\boldsymbol{\eta})$ for all rotations $\boldsymbol{Q} \in \mathrm{SO}(d)$, the spherical Wasserstein distance inherits this property, $\mathrm{W}_p(\mu,\nu) = \mathrm{W}_p(\mu\circ\boldsymbol{Q},\nu\circ\boldsymbol{Q})$.

**Theorem 3.7** (Metric)**.** *For every $p \in [1,\infty)$, the sliced spherical Wasserstein distance $\mathrm{PSW}_p$ is a metric on $\mathcal{P}(\mathbb{S}^{d-1})$, which induces the same topology as the spherical Wasserstein distance $\mathrm{W}_p$. There exist constants $c_{d,p}, C_{d,p} > 0$ such that*

$$c_{d,p}\,\mathrm{PSW}_p(\mu,\nu) \le \mathrm{W}_p(\mu,\nu) \le C_{d,p}\,\mathrm{PSW}_p(\mu,\nu)^{\frac{1}{p(d+1)}}, \qquad \forall \mu,\nu \in \mathcal{P}(\mathbb{S}^{d-1}). \tag{31}$$

*It is rotationally invariant in the sense that for every $\mu,\nu \in \mathcal{P}(\mathbb{S}^2)$ and $\boldsymbol{Q} \in \mathrm{SO}(d)$, we have*

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

The proof is given in Appendix A. Even though the Wasserstein $\mathrm{W}_p$ and sliced Wasserstein metric $\mathrm{PSW}_p$ on $\mathcal{P}(\mathbb{S}^{d-1})$ are topologically equivalent, they are not bilipschitz equivalent. As $\mathbb{S}^{d-1}$ is a compact set, the $p$-Wasserstein metrics for all $p \in [1,\infty)$ induce the same topology on $\mathcal{M}(\mathbb{S}^{d-1})$, see [67, § 5.2]. We conclude this section by mentioning another slicing approach.

**Remark 3.8** (Semicircular slices)**.** A different approach due to [13, 14] uses slicing along semicircles and maps to the one-dimensional torus $\mathbb{T} \coloneqq \mathbb{R}/(2\pi)$. We mention here the case $d = 3$ following [58]. We denote the components of the inverse of the bijective spherical coordinate transform (10) by $\Phi^{-1}(\boldsymbol{\xi}) = (\mathrm{azi}(\boldsymbol{\xi}), \mathrm{zen}(\boldsymbol{\xi}))$, where $\Psi$ denotes the transformation in Euler angles (14). For $\boldsymbol{\psi} = \Phi(\varphi,\theta) \in \mathbb{S}^2$, we define the slicing operator

$$\mathcal{A}_{\boldsymbol{\psi}} \colon \mathbb{S}^2 \to \mathbb{T},\ \boldsymbol{\xi} \mapsto \mathrm{azi}(\Psi(\varphi,\theta,0)^\top\boldsymbol{\xi}).$$

The semicircular sliced Wasserstein distance

$$\mathrm{SSW}_p^p(\mu,\nu) \coloneqq \int_{\mathbb{S}^2} \mathrm{W}_p^p((\mathcal{A}_{\boldsymbol{\psi}})_\#\mu, (\mathcal{A}_{\boldsymbol{\psi}})_\#\nu)\,\mathrm{d}u_{\mathbb{S}^2}(\boldsymbol{\psi})$$

is also a rotationally invariant metric, see [58], but an equivalence as in (31) is not known. Furthermore, it requires solving the one-dimensional OT on the torus, which is more difficult than on an interval, see [23, 61, 62]. □

# 4. Sliced OT on SO(3)

We present an approach to generate sliced Wasserstein distances on $\mathbb{X} = \mathrm{SO}(3)$. We denote the angle of the rotation $\boldsymbol{Q} \in \mathrm{SO}(3)$ by

$$\angle(\boldsymbol{Q}) \coloneqq \arccos \frac{-1 + \mathrm{trace}(\boldsymbol{Q})}{2} \in [0, \pi].$$

We take $\mathbb{D} = \mathrm{SO}(3)$ and the slicing operator

$$d_{\boldsymbol{Q}} \colon \mathrm{SO}(3) \to [0, \pi], \quad \boldsymbol{P} \mapsto \angle(\boldsymbol{Q}^\top \boldsymbol{P}), \qquad \forall \boldsymbol{Q} \in \mathrm{SO}(3). \tag{32}$$

The respective slice $d_{\boldsymbol{Q}}^{-1}(\omega)$ can be parameterized as follows.

**Proposition 4.1** (Parameterization)**.** *Let $\boldsymbol{Q} \in \mathrm{SO}(3)$ and $\omega \in [0, \pi]$. Then*

$$d_{\boldsymbol{Q}}^{-1}(\omega) = \{\boldsymbol{A} \in \mathrm{SO}(3) \mid d_{\boldsymbol{Q}}(\boldsymbol{A}) = \omega\} = \{\boldsymbol{Q}\mathrm{R}_{\boldsymbol{\xi}}(\omega) \mid \boldsymbol{\xi} \in \mathbb{S}^2\},$$

*where $\mathrm{R}_{\boldsymbol{\xi}}(\omega)$ is the rotation with axis $\boldsymbol{\xi}$ and angle $\omega$, see* (13)*.*

*Proof.* We have $\angle(\mathrm{R}_{\boldsymbol{\eta}}(\omega)) = \omega$ for all $\boldsymbol{\eta} \in \mathbb{S}^2$ and $\omega \in [0, \pi]$. Let $\boldsymbol{A} \in \mathrm{SO}(3)$. We write $\boldsymbol{Q}^\top \boldsymbol{A}$ in axis–angle form (13) as $\boldsymbol{Q}^\top \boldsymbol{A} = \mathrm{R}_{\boldsymbol{\xi}}(\sigma)$ with $\sigma = \angle(\boldsymbol{Q}^\top \boldsymbol{A})$ and some $\boldsymbol{\xi} \in \mathbb{S}^2$. Then we have $\boldsymbol{A} \in d_{\boldsymbol{Q}}^{-1}(\omega)$ if and only if $\omega = \angle(\boldsymbol{Q}^\top \boldsymbol{A}) = \sigma$, which shows the claim. $\square$

## 4.1. A Two-Dimensional Radon Transform on SO(3)

Let $f \in L^1(\mathrm{SO}(3))$. We define the *Radon transform* on the rotation group for any $\boldsymbol{Q} \in \mathrm{SO}(3)$ and $\omega \in [0, \pi]$ by

$$\mathcal{T} f(\boldsymbol{Q}, \omega) \coloneqq \frac{1}{4\pi^2}(1 - \cos(\omega)) \int_{\mathbb{S}^2} f(\boldsymbol{Q}\mathrm{R}_{\boldsymbol{\xi}}(\omega)) \,\mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}), \tag{33}$$

and its restriction $\mathcal{T}_{\boldsymbol{Q}} f(\omega) \coloneqq 8\pi \mathcal{T}(\boldsymbol{Q}, \omega)$. By Proposition 4.1, the domain of integration of $f$ is the slice $d_{\boldsymbol{Q}}^{-1}(\omega)$. With (15), we obtain

$$\int_{\mathrm{SO}(3)} f(\boldsymbol{A}) \,\mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{A}) = \int_0^\pi \mathcal{T}_{\boldsymbol{Q}} f(\omega) \,\mathrm{d}\omega, \qquad \forall f \in L^1(\mathrm{SO}(3)),$$

which serves as analogue to (23). By Fubini's theorem, the last equation implies that $\mathcal{T}_{\boldsymbol{Q}} f \in L^1([0, \pi])$ and $\mathcal{T} f \in L^1(\mathrm{SO}(3) \times [0, \pi])$ if $f \in L^1(\mathrm{SO}(3))$.

**Theorem 4.2** (Singular value decomposition)**.** *The Radon transform* (33) *is one-to-one*

$$\mathcal{T} \colon L^2(\mathrm{SO}(3)) \to L^2(\mathrm{SO}(3) \times [0, \pi])$$

*and has the singular value decomposition*

$$\mathcal{T} \widetilde{D}_n^{j,k} = \lambda_n^{\mathcal{T}} F_n^{j,k}, \qquad \forall n \in \mathbb{N}_0,\ j, k \in \{-n, \dots, n\} \tag{34}$$

*with the orthonormal basis $\widetilde{D}_n^{j,k}$ of $L^2(\mathrm{SO}(3))$, see* (19)*, the singular values*

$$\lambda_0^{\mathcal{T}} = \sqrt{\frac{3}{2}}\pi^{-1/2} \quad \text{and} \quad \lambda_n^{\mathcal{T}} \coloneqq \frac{1}{(2n+1)\sqrt{\pi}} \qquad \forall n \in \mathbb{N},$$

*and the set of orthonormal functions on $L^2(\mathrm{SO}(3) \times [0,\pi])$ defined by*

$$F_n^{j,k}(\boldsymbol{Q},\omega) := \begin{cases} \frac{2}{\sqrt{\pi}}\, \widetilde{D}_n^{j,k}(\boldsymbol{Q})\, \sin\big((n+\frac{1}{2})\omega\big)\, \sin(\frac{\omega}{2}), & n \neq 0, \\ \sqrt{\frac{8}{3\pi}}\, \widetilde{D}_0^{0,0}(\boldsymbol{Q})\, \big(\sin(\frac{\omega}{2})\big)^2, & n = 0, \end{cases} \qquad (\boldsymbol{Q},\omega) \in \mathrm{SO}(3) \times [0,\pi].$$

The proof is postponed to Appendix B. Restricting $\mathcal{T}(\cdot,\omega)$ to a fixed radius $\omega$, we obtain the following injectivity result, which is of a similar structure as for the spherical cap [71] or spherical slice transform [68] on $\mathbb{S}^2$.

**Corollary 4.3** (Injectivity for fixed $\omega$)**.** *For fixed $\omega \in (0,\pi)$, the Radon transform $\mathcal{T}(\cdot,\omega)$ is injective as operator $L^2(\mathrm{SO}(3)) \to L^2(\mathrm{SO}(3))$ if and only if $\omega/\pi = p/q$ where $p/q$ is a reduced fraction and $p \in \mathbb{N}$ is even and $q \in \mathbb{N}$ is odd.*

*Proof.* As a direct consequence of Theorem 4.2 and the orthonormality of the rotational harmonics, the restricted transformation has the eigenvalue decomposition

$$\mathcal{T}\widetilde{D}_n^{j,k}(\boldsymbol{Q},\omega) = \frac{2}{(2n+1)\pi} \sin((n+\tfrac{1}{2})\omega) \sin(\tfrac{\omega}{2})\, \widetilde{D}_n^{j,k}(\boldsymbol{Q}), \qquad \forall \boldsymbol{Q} \in \mathrm{SO}(3).$$

It is injective if and only if all eigenvalues $\frac{2}{(2n+1)\pi} \sin((n+\frac{1}{2})\omega) \sin(\frac{\omega}{2})$ are non-zero. The $n$-th eigenvalue is zero if and only if $(n+1/2)\omega \in \pi\mathbb{N}$, i.e., there exists $k \in \mathbb{N}$ such that $\omega/\pi = 2k/(2n+1)$, the quotient of an even and an odd integer. Since this fraction can only be reduced with an odd factor, also the reduced fraction consists of an even divided by an odd integer. $\square$

**Theorem 4.4** (Adjoint)**.** *Let $\boldsymbol{Q} \in \mathrm{SO}(3)$. The adjoint of $\mathcal{T}_{\boldsymbol{Q}} \colon L^2(\mathrm{SO}(3)) \to L^2([0,\pi])$ is given by*

$$\mathcal{T}_{\boldsymbol{Q}}^* g(\boldsymbol{A}) = g(d_{\boldsymbol{Q}}(\boldsymbol{A})), \qquad \forall \boldsymbol{A} \in \mathrm{SO}(3),$$

*and the adjoint of $\mathcal{T} \colon L^2(\mathrm{SO}(3)) \to L^2(\mathrm{SO}(3) \times [0,\pi])$ is*

$$\mathcal{T}^*(\boldsymbol{A}) = \frac{1}{8\pi^2} \int_{\mathrm{SO}(3)} g(\boldsymbol{Q}, d_{\boldsymbol{Q}}(\boldsymbol{A}))\, \mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{Q}), \qquad \forall \boldsymbol{A} \in \mathrm{SO}(3).$$

*Proof.* Let $f \in C(\mathrm{SO}(3))$, $g \in C([0,\pi])$, and $\boldsymbol{Q} \in \mathrm{SO}(3)$. We have with the substitution $\boldsymbol{A} = \boldsymbol{Q}\mathrm{R}_{\boldsymbol{\xi}}(\omega)$ and (15)

$$\begin{aligned} \int_{\mathrm{SO}(3)} f(\boldsymbol{A}) g(d_{\boldsymbol{Q}}(\boldsymbol{A}))\, \mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{A}) &= 2 \int_0^\pi \int_{\mathbb{S}^2} f(\boldsymbol{Q}\mathrm{R}_{\boldsymbol{\xi}}(\omega)) g(\omega)(1-\cos(\omega))\, \mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi})\, \mathrm{d}\omega \\ &= \int_0^\pi \mathcal{T}_{\boldsymbol{Q}} f(\omega) g(\omega)\, \mathrm{d}\omega. \end{aligned}$$

The second claim follows analogously by considering $g \in C(\mathrm{SO}(3) \times [0,\pi])$ and integration over $\boldsymbol{Q} \in \mathrm{SO}(3)$. $\square$

## 4.2. Slicing of Measures

We generalize the definition (33) to a measure $\mu \in \mathcal{M}(\mathrm{SO}(3))$ via the pushforward of the slicing operator (32), i.e.,

$$\begin{aligned} \mathcal{T}_{\boldsymbol{Q}}\mu &\coloneqq (d_{\boldsymbol{Q}})_{\#}\mu \in \mathcal{P}([0,\pi]) \quad \text{and} \\ \mathcal{T}\mu &\coloneqq T_{\#}(u_{\mathrm{SO}(3)} \times \mu) \in \mathcal{P}(\mathrm{SO}(3) \times [0,\pi]), \qquad \text{with} \quad T(\boldsymbol{Q}, \boldsymbol{A}) = (\boldsymbol{Q}, d_{\boldsymbol{Q}}(\boldsymbol{A})). \end{aligned}$$

**Proposition 4.5** (Connection with adjoint)**.** *Let $\boldsymbol{Q} \in \mathrm{SO}(3)$ and $\mu \in \mathcal{M}(\mathrm{SO}(3))$. Then*

$$\begin{aligned} \langle \mathcal{T}_{\boldsymbol{Q}}\mu, g\rangle &= \langle \mu, \mathcal{T}_{\boldsymbol{Q}}^{*} g\rangle, \qquad &\forall g \in C([0,\pi]), \\ \langle \mathcal{T}\mu, g\rangle &= \langle \mu, \mathcal{T}^{*} g\rangle, \qquad &\forall g \in C(\mathrm{SO}(3) \times [0,\pi]). \end{aligned}$$

**Proposition 4.6** (Absolutely continuous measures)**.** *Let $\mu \in \mathcal{M}(\mathrm{SO}(3))$ be absolutely continuous, i.e., there exists a density function $f \in L^1(\mathrm{SO}(3))$ such that $\mu = f\sigma_{\mathrm{SO}(3)}$. Then*

$$\begin{aligned} \mathcal{T}_{\boldsymbol{Q}}\mu &= (\mathcal{T}_{\boldsymbol{Q}} f)\sigma_{[0,\pi]}, \quad \textit{and} \\ \mathcal{T}\mu &= (\mathcal{T} f)\sigma_{\mathrm{SO}(3)\times[0,\pi]}. \end{aligned}$$

The last two propositions can be proven analogously to Propositions 3.5 and 3.6.

**Theorem 4.7** (Injectivity)**.** *The Radon transform $\mathcal{T}\colon \mathcal{M}(\mathrm{SO}(3)) \to \mathcal{M}(\mathrm{SO}(3) \times [0,\pi])$ is injective.*

The proof, which uses the singular value decomposition, is given in Appendix B. Although the slicing operator and therefore the transform $\mathcal{T}$ could be easily generalized to $\mathrm{SO}(d)$, this does not apply for the singular value decomposition (34) as the harmonic analysis on $\mathrm{SO}(d)$ becomes much more evolved for $d > 3$, cf. [74, Chapter IX].

## 4.3. Sliced Wasserstein Distance on SO(3)

Let $p \in [1,\infty)$. We define the sliced Wasserstein distance between measures $\mu, \nu \in \mathcal{P}(\mathrm{SO}(3))$ by

$$\mathrm{SOSW}_p^p(\mu,\nu) \coloneqq \int_{\mathrm{SO}(3)} \mathrm{W}_p^p\left(\mathcal{T}_{\boldsymbol{Q}}\mu, \mathcal{T}_{\boldsymbol{Q}}\nu\right) \mathrm{d}u_{\mathrm{SO}(3)}(\boldsymbol{Q}), \tag{35}$$

which is the mean of Wasserstein distances on the interval $[0,\pi]$.

**Theorem 4.8.** *Let $p \in [1,\infty)$. The sliced Wasserstein distance (35) is a metric on $\mathcal{P}(\mathrm{SO}(3))$ that is invariant to rotations, i.e., for any $\boldsymbol{A} \in \mathrm{SO}(3)$ and $\mu, \nu \in \mathcal{P}(\mathrm{SO}(3))$, we have*

$$\mathrm{SOSW}_p^p(\mu(\boldsymbol{A}\cdot), \nu(\boldsymbol{A}\cdot)) = \mathrm{SOSW}_p^p(\mu,\nu).$$

*Proof.* The positive definiteness is due to Theorem 4.7, while the symmetry and triangular inequality follow from the respective properties of the Wasserstein distance on $[0,\pi]$. By definition in (35), we have

$$\mathrm{SOSW}_p^p(\mu(\boldsymbol{A}\cdot), \nu(\boldsymbol{A}\cdot)) = \int_{\mathrm{SO}(3)} \mathrm{W}_p^p\left(\mu \circ \boldsymbol{A} \circ d_{\boldsymbol{Q}}^{-1}, \nu \circ \boldsymbol{A} \circ d_{\boldsymbol{Q}}^{-1}\right) \mathrm{d}u_{\mathrm{SO}(3)}(\boldsymbol{Q}).$$

Let $\omega \in [0,\pi]$. We have $\boldsymbol{B} \in \boldsymbol{A} \circ d_{\boldsymbol{Q}}^{-1}(\omega)$ if and only if $\omega = d_{\boldsymbol{Q}}(\boldsymbol{A}^\top \boldsymbol{B}) = d_{\boldsymbol{AQ}}(\boldsymbol{B})$, cf. (32). Hence

$$\mathrm{SOSW}_p^p(\mu(\boldsymbol{A}\cdot), \nu(\boldsymbol{A}\cdot)) = \int_{\mathrm{SO}(3)} \mathrm{W}_p^p\left((d_{\boldsymbol{AQ}})_\# \mu, (d_{\boldsymbol{AQ}})_\# \nu\right) \mathrm{d}u_{\mathrm{SO}(3)}(\boldsymbol{AQ}) = \mathrm{SOSW}_p^p(\mu, \nu). \quad \square$$

In Appendix C, we provide a relation between the sliced Wasserstein distances on SO(3) and on $\mathbb{S}^3$.

## 5. Barycenter algorithms

There exist two approaches to compute barycenters of measures using 1D Wasserstein distances along projected measures, namely sliced Wasserstein barycenters and Radon Wasserstein barycenters, cf. [16]. We adapt them to our slicing in Section 5.1 and 5.2, respectively.

### 5.1. Sliced Wasserstein barycenters

For sliced Wasserstein barycenters, we replace in (2) the Wasserstein distance $\mathrm{W}_2$ by its sliced counterpart. In particular, with the general notion of slicing in (6), the *sliced Wasserstein barycenter* of given measures $\mu_i \in \mathcal{P}(\mathbb{X})$, $i \in [\![M]\!]$, and $\boldsymbol{\lambda} \in \Delta_M$, is defined by

$$\mathrm{Bary}_{\mathbb{X}}^{\mathrm{SW}}(\mu_i, \lambda_m)_{i=1}^M \coloneqq \operatorname*{arg\,min}_{\nu \in \mathcal{P}(\mathbb{X})} \sum_{i=1}^M \lambda_i \, \mathrm{SW}_2^2(\nu, \mu_i).$$

**Remark 5.1.** Although the different slicing approaches often yield similar barycenters, as we will see in the numerics, they differ considerably in the extreme case of two antipodal point measures on the sphere $\mathbb{S}^2$. Denote by $\delta_{\boldsymbol{\xi}}$ the Dirac measure at $\boldsymbol{\xi} \in \mathbb{S}^2$. The Wasserstein barycenter $\mathrm{Bary}_{\mathbb{S}^2}^{\mathrm{W}}$ of $\delta_{\boldsymbol{e}^3}$ and $\delta_{-\boldsymbol{e}^3}$ with equal weights $\lambda_1 = \lambda_2 = 1/2$ consists of the measures $\nu \in \mathcal{P}(\mathbb{S}^2)$ with support on the equator. However, all measures in $\mathcal{P}(\mathbb{S}^2)$ are parallelly sliced Wasserstein barycenters $\mathrm{Bary}_{\mathbb{S}^2}^{\mathrm{PSW}}$. For the semicircular slices of Remark 3.8, we can show that the uniform distribution on the equator is a candidate for the barycenter $\mathrm{Bary}_{\mathbb{S}^2}^{\mathrm{SSW}}$, while $u_{\mathbb{S}^2}$ is not. The details are provided in Appendix D. $\square$

We consider two types of discretization to compute sliced Wasserstein barycenters. Free-support barycenters are based on a Lagrangian discretization: measures are represented by samples, and the minimization is carried out over the coordinates of those samples, see [16, 55, 62]. Fixed-support barycenters are based on a Eulerian discretization: a fixed grid is considered for all the measures which are represented by the weights given to each grid point, and the minimization is carried out over those weights, cf. [9, 18].

#### 5.1.1. Free-support discretization

For $X = (X_k)_{k=1}^N \in \mathbb{X}^N$, we note $\mu_X \coloneqq \frac{1}{N} \sum_{k=1}^N \delta_{X_k}$. We consider $M$ discrete measures $\mu_{Y^{(i)}} \in \mathcal{P}(\mathbb{X})$ with $Y^{(i)} \in \mathbb{X}^N$ for all $i \in [\![M]\!]$. The aim is to compute a discrete barycenter of these measures, i.e., we want to minimize the functional

$$\mathcal{E} \colon \mathbb{X}^N \to \mathbb{R}, \qquad X \mapsto \sum_{i \in [\![M]\!]} \lambda_i \, \mathrm{SW}_2^2(\mu_X, \mu_{Y^{(i)}}), \tag{36}$$

which is not convex in general, via a stochastic gradient descent, as it has been applied in the Euclidean space $\mathbb{X} = \mathbb{R}^d$ in [62, sect. 3.2] and [55, sect. 10.4]. Since $X$ is on a manifold $\mathbb{X}$, the gradient is in its tangent space $T_x\mathbb{X}$ at $x \in \mathbb{X}$. By Whitney's embedding theorem [47, thm. 6.15], we can assume $\mathbb{X}$ to be embedded in Euclidean space $\mathbb{R}^d$ for sufficiently large $d$. If $f\colon D \to \mathbb{R}$ is differentiable on an open set $D \subset \mathbb{R}^d$ with $D \supset \mathbb{X}$, then the Riemannian gradient of the restriction of $f$ to $\mathbb{X}$ is given by $\nabla f(x) = \mathrm{proj}_{T_x\mathbb{X}}(\nabla_{\mathbb{R}^d} f(x))$, where $\nabla_{\mathbb{R}^d}$ denotes the gradient in Euclidean space and $\mathrm{proj}_{T_x\mathbb{X}}$ the orthogonal projection to the tangent space, see [2, sect. 3.6.1]. The gradient of the functional (36) can be expressed as follows.

**Theorem 5.2.** *Let $\mathbb{X}$ be a smooth submanifold of $\mathbb{R}^d$ and $X, Y \in \mathbb{X}^N$ consist of pairwise distinct points $X_k$, $k \in [\![N]\!]$. Assume that the slicing operator $\mathcal{S}_{\boldsymbol{\psi}}\colon \mathbb{X} \to \mathbb{R}$ is differentiable for all $\boldsymbol{\psi} \in \mathbb{D}$, that $(\boldsymbol{x}, \boldsymbol{\psi}) \mapsto \mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x})$ is bounded on $\mathbb{X} \times \mathbb{D}$, and that $\boldsymbol{\psi} \mapsto \nabla \mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x})$ is integrable on $\mathbb{D}$ uniformly for every $\boldsymbol{x} \in \mathbb{X}$. Furthermore, assume that $\mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x}_1) \neq \mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x}_2)$ for $u_{\mathbb{D}}$-almost every $\boldsymbol{\psi} \in \mathbb{D}$ if $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$. Then the gradient of the sliced Wasserstein distance between the two measures $\mu_X$ and $\mu_Y$ with respect to $X$ reads*

$$\nabla_{X_k} \mathrm{SW}_2^2(\mu_X, \mu_Y) = \frac{2}{N} \int_{\mathbb{D}} \left[\mathcal{S}_{\boldsymbol{\psi}}\left(X_k\right) - \mathcal{S}_{\boldsymbol{\psi}}\left(Y_{\sigma_{Y,\boldsymbol{\psi}} \circ \sigma^{-1}_{X,\boldsymbol{\psi}}(k)}\right)\right] \nabla \mathcal{S}_{\boldsymbol{\psi}}(X_k) \,\mathrm{d}u_{\mathbb{D}}(\boldsymbol{\psi}), \tag{37}$$

*where $\sigma_{X,\boldsymbol{\psi}}\colon [\![N]\!] \to [\![N]\!]$ is a permutation which sorts $\mathcal{S}_{\boldsymbol{\psi}}(X_k)$, i.e.,*

$$\mathcal{S}_{\boldsymbol{\psi}}(X_{\sigma_{X,\boldsymbol{\psi}}(1)}) \le \mathcal{S}_{\boldsymbol{\psi}}(X_{\sigma_{X,\boldsymbol{\psi}}(2)}) \le \ldots \le \mathcal{S}_{\boldsymbol{\psi}}(X_{\sigma_{X,\boldsymbol{\psi}}(N)}).$$

*Proof.* Let $\boldsymbol{\psi} \in \mathbb{D}$. For the sake of simplicity, we write $\mathcal{S}_{\boldsymbol{\psi}}(X) \coloneqq (\mathcal{S}_{\boldsymbol{\psi}}(X_k))_{1 \le k \le N} \in \mathbb{R}^N$. The pseudo-inverse of the cumulative density function of $\mu_{\mathcal{S}_{\boldsymbol{\psi}}(X)}$ is written, for $r \in (0,1)$,

$$\begin{aligned} F^{-1}_{\mu_{\mathcal{S}_{\boldsymbol{\psi}}(X)}}(r) &= \min\left\{x \in \mathbb{R} \cup \{-\infty\} \;\middle|\; \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{[\mathcal{S}_{\boldsymbol{\psi}}(X_k),1]}(x) \ge r\right\} \\ &= \mathcal{S}_{\boldsymbol{\psi}}\left(X_{\sigma_{X,\boldsymbol{\psi}}(\lceil rN \rceil)}\right). \end{aligned}$$

Then

$$\begin{aligned} \mathrm{W}_2^2\left((\mathcal{S}_{\boldsymbol{\psi}})_{\#}\mu_X, (\mathcal{S}_{\boldsymbol{\psi}})_{\#}\mu_Y\right) &= \mathrm{W}_2^2\left(\mu_{\mathcal{S}_{\boldsymbol{\psi}}(X)}, \mu_{\mathcal{S}_{\boldsymbol{\psi}}(Y)}\right) \\ &= \int_{[0,1]} \left|F^{-1}_{\mu_{\mathcal{S}_{\boldsymbol{\psi}}(X)}}(r) - F^{-1}_{\mu_{\mathcal{S}_{\boldsymbol{\psi}}(Y)}}(r)\right|^2 \mathrm{d}r \\ &= \sum_{k=1}^{N} \frac{1}{N}\left|\mathcal{S}_{\boldsymbol{\psi}}\left(X_{\sigma_{X,\boldsymbol{\psi}}(k)}\right) - \mathcal{S}_{\boldsymbol{\psi}}\left(Y_{\sigma_{Y,\boldsymbol{\psi}}(k)}\right)\right|^2 \\ &= \frac{1}{N}\sum_{k=1}^{N}\left|\mathcal{S}_{\boldsymbol{\psi}}\left(X_k\right) - \mathcal{S}_{\boldsymbol{\psi}}\left(Y_{\sigma_{Y,\boldsymbol{\psi}} \circ \sigma^{-1}_{X,\boldsymbol{\psi}}(k)}\right)\right|^2, \end{aligned}$$

is bounded and, thus, integrable with respect to $\boldsymbol{\psi}$ on $(\mathbb{D}, u_{\mathbb{D}})$. Its gradient is given by

$$\nabla_{X_k} \mathrm{W}_2^2\left((\mathcal{S}_{\boldsymbol{\psi}})_{\#}\mu_X, (\mathcal{S}_{\boldsymbol{\psi}})_{\#}\mu_Y\right) = \left[\mathcal{S}_{\boldsymbol{\psi}}\left(X_k\right) - \mathcal{S}_{\boldsymbol{\psi}}\left(Y_{\sigma_{Y,\boldsymbol{\psi}} \circ \sigma^{-1}_{X,\boldsymbol{\psi}}(k)}\right)\right] \nabla \mathcal{S}_{\boldsymbol{\psi}}(X_k) \tag{38}$$

for $u_{\mathbb{D}}$-almost every $\boldsymbol{\psi} \in \mathbb{D}$. Indeed, we assume that $X_k$, $k \in [\![N]\!]$, are pairwise distinct, so $\mathcal{S}_{\boldsymbol{\psi}}(X_k)$ are $u_{\mathbb{D}}$-almost surely pairwise distinct, so $\sigma_{X,\boldsymbol{\psi}}$ is $u_{\mathbb{D}}$-almost surely uniquely defined, and constant in the neighborhood of $X$. Hence, (38) is integrable on $\mathbb{D}$ since $\nabla\mathcal{S}_\psi$ is and the rest is bounded by assumption. Therefore, similarly to [16], we have

$$\begin{aligned}\mathrm{SW}_p^p(\mu_Y,\mu_X) &= \int_{\mathbb{D}} \mathrm{W}_p^p\left((\mathcal{S}_{\boldsymbol{\psi}})_{\#}\mu_X,(\mathcal{S}_{\boldsymbol{\psi}})_{\#}\mu_Y\right)\mathrm{d}u(\boldsymbol{\psi})\\ &= \frac{1}{N}\int_{\mathbb{D}}\sum_{k=1}^{N}\left|\mathcal{S}_{\boldsymbol{\psi}}\left(X_k\right)-\mathcal{S}_{\boldsymbol{\psi}}\left(Y_{\sigma_{Y,\boldsymbol{\psi}}\circ\sigma_{X,\boldsymbol{\psi}}^{-1}(k)}\right)\right|^p\mathrm{d}u(\boldsymbol{\psi})\end{aligned}$$

and

$$\begin{aligned}\nabla_{X_k}\mathrm{SW}_2^2(\mu_Y,\mu_X) &= \int_{\mathbb{D}}\nabla_{X_k}\mathrm{W}_2^2\left(\mu_{\mathcal{S}_{\boldsymbol{\psi}}(Y)},\mu_{\mathcal{S}_{\boldsymbol{\psi}}(X)}\right)\mathrm{d}u(\boldsymbol{\psi})\\ &= \frac{2}{N}\int_{\mathbb{D}}\left[\mathcal{S}_{\boldsymbol{\psi}}\left(X_k\right)-\mathcal{S}_{\boldsymbol{\psi}}\left(Y_{\sigma_{Y,\boldsymbol{\psi}}\circ\sigma_{X,\boldsymbol{\psi}}^{-1}(k)}\right)\right]\nabla\mathcal{S}_{\boldsymbol{\psi}}(X_k)\,\mathrm{d}u(\boldsymbol{\psi}).\end{aligned}$$ □

We discretize (37) over $\boldsymbol{\psi}$ via considering $(\boldsymbol{\psi}_q)_{q=1}^P$ distributed according to the uniform measure $u_{\mathbb{D}}$ to get a numerical approximation. This enables us to devise a stochastic gradient descent algorithm with initialization $X^0\in\mathbb{X}^N$ and whose step $l\in\mathbb{N}$ is given by

$$\begin{aligned}X_k^{l+1} &= \exp_{X_k^l}^{\mathbb{X}}(-\tau_l\nabla\mathcal{E}(X^l)_k)\\ &= \exp_{X_k^l}^{\mathbb{X}}\left(-\tau_l\sum_{i=1}^{M}\frac{2\lambda_i}{NP}\sum_{q=1}^{P}\left[\mathcal{S}_{\boldsymbol{\psi}_q}(X_k^l)-\mathcal{S}_{\boldsymbol{\psi}_q}\left(Y^{(i)}_{\sigma_{Y^{(i)},\boldsymbol{\psi}_q}\circ\sigma_{X^l,\boldsymbol{\psi}_q}^{-1}(k)}\right)\right]\nabla\mathcal{S}_{\boldsymbol{\psi}}(X_k^l)\right)\end{aligned}\tag{39}$$

for every $k\in[\![N]\!]$, where $\exp_x^{\mathbb{X}}$ denotes exponential map, which maps a subset of the tangent space $T_x\mathbb{X}$ to the manifold $\mathbb{X}$, and $\tau_l>0$ is the step size, also known as the learning rate.

**Remark.** The last theorem can be extended to the case where both measures have different numbers of points. The squared Wasserstein distance between $\mu_x$ and $\mu_y$, with $x=(x_i)_{1\le i\le N}\in\mathbb{R}^N$ and $y=(y_j)_{1\le j\le M}\in\mathbb{R}^M$ two sorted lists of real numbers, is not anymore $\frac{1}{N}\|x-y\|_{\mathbb{R}^N}^2$ but is of the shape $\sum_{i,j}\pi_{i,j}(x_i-y_j)^2$ with the transport plan $\pi\in\mathbb{R}^{N\times M}$ that depends only on $N$ and $M$ (not $x$ nor $y$). Because the matrix $\pi$ is sparse with support close to the diagonal $\frac{i}{j}\approx\frac{N}{M}$, we can generalize our algorithm while keeping its complexity $\mathcal{O}((N+M)\log(N+M))$. □

**5.1.1.1. Application to PSW on the sphere** We look at the stochastic gradient descent step for the sphere with the parallel slicing operator (20). Let $\boldsymbol{x},\boldsymbol{\psi}\in\mathbb{S}^{d-1}$. The projection to the tangential plane $T_{\boldsymbol{x}}\mathbb{S}^{d-1}=\{\boldsymbol{v}\in\mathbb{R}^d\mid\langle\boldsymbol{v},\boldsymbol{x}\rangle=0\}$ is

$$\mathrm{proj}_{T_{\boldsymbol{x}}\mathbb{S}^{d-1}}\colon\mathbb{R}^d\to T_{\boldsymbol{x}}\mathbb{S}^{d-1},\;\boldsymbol{v}\mapsto\boldsymbol{v}-\langle\boldsymbol{x},\boldsymbol{v}\rangle\,\boldsymbol{x},$$

and we have $\nabla\mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x})=\mathrm{proj}_{T_{\boldsymbol{x}}\mathbb{S}^{d-1}}(\boldsymbol{\psi})$. As a consequence, the stochastic gradient descent step (39) is

$$X_k^{l+1}=\exp_{X_k^l}^{\mathbb{S}^{d-1}}\left[-\tau_l\,\mathrm{proj}_{T_{X_k^l}\mathbb{S}^{d-1}}\left(\sum_{i=1}^{M}\frac{2\lambda_i}{NP}\sum_{q=1}^{P}\left\langle\boldsymbol{\psi}_q,\;X_k^l-Y^{(i)}_{\sigma_{Y^{(i)},\boldsymbol{\psi}_q}\circ\sigma_{X^l,\boldsymbol{\psi}_q}^{-1}(k)}\right\rangle\boldsymbol{\psi}_q\right)\right]\tag{40}$$

with the exponential map $\exp_{\boldsymbol{x}}^{\mathbb{S}^{d-1}}(\boldsymbol{v}) := \cos(\|\boldsymbol{v}\|)\boldsymbol{x} + \sin(\|\boldsymbol{v}\|)\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$.

The numerical complexity of this step is $\mathcal{O}(MN(\log(N)+d)P)$ with $M$ the number of measures, $N$ the number of points in each measure and $P$ the number of directions (i.e. of projections or slices), since the sorting has complexity $\mathcal{O}(N\log(N))$ and the dimension $d$ comes in only due to the computation of the inner product in $\mathbb{R}^d$ and the generation of uniform samples on $\mathbb{S}^{d-1}$.

**5.1.1.2. Application to SOSW on the rotation group** Let us now look at the case $\mathbb{X} = \mathrm{SO}(3)$ with the slicing operator $\mathcal{S}_{\boldsymbol{\psi}}^{\mathrm{SO}(3)}(\boldsymbol{R}) = \mathrm{trace}(\boldsymbol{R}^\top\boldsymbol{\psi})$ for $\boldsymbol{\psi} \in \mathbb{D} = \mathrm{SO}(3)$ and $\boldsymbol{R} \in \mathrm{SO}(3)$. In order to avoid numerical instabilities due to the unboundedness of the derivative of the arccosine, we take here a monotone transformation of the slicing operator (32) in order to simplify computation while keeping the properties of SOSW. Drawing random directions $\boldsymbol{\psi} = \Psi(\alpha,\beta,\gamma) \in \mathrm{SO}(3)$ can be done by randomly generating Euler angles $\alpha,\gamma \sim u_{[0,2\pi]}$ and $\beta = \arccos t$ with $t \sim u_{[-1,1]}$. The gradient descent step (39) is similar to the case of the sphere. The tangent space is $T_{\boldsymbol{R}}\mathrm{SO}(3) = \{\boldsymbol{A} \in \mathbb{R}^{3\times3} \mid \boldsymbol{R}^\top\boldsymbol{A} = -\boldsymbol{A}^\top\boldsymbol{R}\}$. Utilizing that the orthogonal projection on the set of skew-symmetric matrices is given by $\boldsymbol{A} \mapsto \frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^\top)$, one can show similarly that the projection to $T_{\boldsymbol{R}}\mathrm{SO}(3)$ is given by

$$\mathrm{proj}_{T_{\boldsymbol{R}}\mathrm{SO}(3)} \colon \mathbb{R}^{3\times3} \to T_{\boldsymbol{R}}\mathrm{SO}(3), \quad \boldsymbol{A} \mapsto \tfrac{1}{2}(\boldsymbol{A} - \boldsymbol{R}\boldsymbol{A}^\top\boldsymbol{R}).$$

The exponential map is given by [77, 3.37]

$$\exp_{\boldsymbol{R}}^{\mathrm{SO}(3)} \colon T_{\boldsymbol{R}}\mathrm{SO}(3) \to \mathrm{SO}(3), \quad \boldsymbol{A} \mapsto \boldsymbol{R}\exp(\boldsymbol{R}^\top\boldsymbol{A}).$$

In order to avoid the computation of the matrix exponential, one could replace the exponential map by a retraction, see [2, Sect. 4.1].

### 5.1.2. Fixed-support discretization

Fixed-support sliced Wasserstein barycenters correspond to a Eulerian discretization. As opposed to Section 5.1.1, the support is the same, fixed set for all measures (including the barycenter), and we minimize over the weights of the Dirac measures. We first study the 1-dimensional case of fixed-support OT.

**Theorem 5.3.** *Let $\{t_j \mid j \in [\![N]\!]\} \subset \mathbb{R}$ with $t_1 < t_2 < \dots < t_N$. Further, let $\boldsymbol{w}, \boldsymbol{v} \in \Delta_N$ and the discrete probability measures $\mu_{\boldsymbol{w}} = \sum_{j=1}^N w_j\delta_{t_j}$ and $\mu_{\boldsymbol{v}} = \sum_{j=1}^N v_j\delta_{t_j}$. We introduce the partial sums $\tilde{\boldsymbol{w}} = (\sum_{j=1}^k w_j)_{k=1}^N$ and $\tilde{\boldsymbol{v}} = (\sum_{j=1}^k v_j)_{k=1}^N$ in $\mathbb{R}^N$ as well as the vectors $\boldsymbol{y} = (\tilde{w}_1,\dots,\tilde{w}_{N-1},\tilde{v}_1,\dots,\tilde{v}_{N-1}) \in \mathbb{R}^{2N-2}$ and $\boldsymbol{z} = (u_{\sigma(j)})_{j=1}^{2N-2}$ with $\sigma$ a permutation such that $u_{\sigma(1)} \le u_{\sigma(2)} \le \dots \le u_{\sigma(2N-2)}$. We set*

$$a_j = \left|t_{\min\{k \mid \tilde{w}_k \ge z_{j+1}\}} - t_{\min\{k \mid \tilde{v}_k \ge z_{j+1}\}}\right|^p \quad \text{for } j \in [\![2N-3]\!],$$

*and $a_0 = a_{2N-2} = 0$. Then the gradient of $\mathrm{W}_p^p(\mu_{\boldsymbol{w}},\mu_{\boldsymbol{v}})$ in $\Delta_N$ with respect to $\boldsymbol{w}$ is given almost everywhere by*

$$\nabla_{\boldsymbol{w}} \mathrm{W}_p^p(\mu_{\boldsymbol{w}},\mu_{\boldsymbol{v}}) = \mathrm{proj}_H\left(\left(\sum_{k=j}^{N-1}\left(a_{\sigma^{-1}(k)-1} - a_{\sigma^{-1}(k)}\right)\right)_{j=1}^{N}\right), \tag{41}$$

*where* $\mathrm{proj}_H\colon \mathbb{R}^N \to H$, $\boldsymbol{x} \mapsto \boldsymbol{x} - \langle \boldsymbol{x}, \mathbb{1}\rangle\, \mathbb{1}$ *is the orthogonal projection on the hyperplane* $H = \{\boldsymbol{x} \in \mathbb{R}^N \mid \langle \boldsymbol{x}, \mathbb{1}\rangle = 0\}$.

*Proof.* The pseudo-inverse of the cumulative distribution function of $\mu_{\boldsymbol{w}}$ is

$$F_{\mu_{\boldsymbol{w}}}^{-1}(r) = \min\left\{ s \in \mathbb{R} \;\middle|\; \sum_{i=1}^N w_i \mathbb{1}_{[t_i,+\infty)}(s) \geq r \right\} = t_{\min\{k \mid \tilde{w}_k \geq r\}}, \quad \forall r \in (0,1),$$

and analogously for $\boldsymbol{v}$. We thus have

$$\begin{aligned} \mathrm{W}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}}) &= \int_0^1 \left|F_{\mu_{\boldsymbol{w}}}^{-1}(r) - F_{\mu_{\boldsymbol{v}}}^{-1}(r)\right|^p \,\mathrm{d}r \\ &= \int_0^1 \left|t_{\min\{k \mid \tilde{w}_k \geq r\}} - t_{\min\{k \mid \tilde{v}_k \geq r\}}\right|^p \,\mathrm{d}r \\ &= \sum_{j=1}^{2N-3} \left|t_{\min\{k \mid \tilde{w}_k \geq z_{j+1}\}} - t_{\min\{k \mid \tilde{v}_k \geq z_{j+1}\}}\right|^p (z_{j+1} - z_j) \\ &= \sum_{j=1}^{2N-3} a_j (z_{j+1} - z_j), \end{aligned}$$

where we have ignored the segment $(0, z_1]$ on which the integrand is 0, and the segment $[z_{2N-2}, 1)$ on which the integrand is 0 or which is empty if $z_{2N-2} = 1$.

We extend the map $\boldsymbol{w} \mapsto \mathrm{W}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}})$ from $\Delta_N$ to some neighborhood in the Euclidean space $\mathbb{R}^N$ by making it constant in the last component $w_N$, then $\frac{\mathrm{d}}{\mathrm{d}w_N} \mathrm{W}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}}) = 0$. For $k \in [\![N-1]\!]$, since $z_j = \tilde{w}_{\sigma(j)}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{w}_k} \mathrm{W}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}}) = a_{\sigma^{-1}(k)-1} - a_{\sigma^{-1}(k)}, \quad \text{a.e.}$$

Indeed, $a_j$ is almost everywhere constant with respect to $\tilde{w}_k$. It is justified using the fact that the $z_j$ are almost surely pairwise distinct. So, if $\tilde{w}_k = z_{j+1}$, then $z_{j+1}$ "moves" (a.s.) with $\tilde{w}_k$, such that $\{k' \mid \tilde{w}_{k'} \geq z_{j+1}\}$ does not change. If $\tilde{w}_k \neq z_{j+1}$, it is obvious that $\{k' \mid \tilde{w}_{k'} \geq z_{j+1}\}$ is locally constant with respect to $\tilde{w}_k$. The definition of $\tilde{\boldsymbol{w}}$ gives a bijection between $(w_j)_{j=1}^{N-1}$ and $(\tilde{w}_k)_{k=1}^{N-1}$, so we obtain for any $j \in [\![N-1]\!]$

$$\frac{\mathrm{d}}{\mathrm{d}w_j} \mathrm{W}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}}) = \sum_{k=1}^N \frac{\mathrm{d}\tilde{w}_k}{\mathrm{d}w_j} \frac{\mathrm{d}}{\mathrm{d}\tilde{w}_k} \mathrm{W}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}}) = \sum_{k=j}^{N-1} \frac{\mathrm{d}}{\mathrm{d}\tilde{w}_k} \mathrm{W}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}}),$$

The projection onto the hyperplane $H$ yields the Riemannian gradient in $\Delta_N$. □

We now come to the sliced OT on $\mathbb{X} \in \{\mathbb{S}^{d-1}, \mathrm{SO}(3)\}$. We consider a family $X = (\boldsymbol{x}_j)_{j=1}^N \in \mathbb{X}^N$ of points on $\mathbb{X}$ representing the fixed support and measures of the form

$$\mu_{\boldsymbol{w}} = \sum_{j=1}^N w_j \delta_{x_j}$$

with some weight vector $\boldsymbol{w} \in \Delta_N$. Let $\boldsymbol{v}^{(i)} \in \Delta_N$ for $i \in [\![M]\!]$ be given probability vectors of the measures whose barycenter we want to compute. Therefore, we minimize

$$\mathcal{E}\colon \Delta_N \to \mathbb{R}, \qquad \boldsymbol{w} \mapsto \sum_{i=1}^{M} \lambda_i \, \mathrm{SW}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}^{(i)}}).$$

We have, for $i \in [\![M]\!]$,

$$\begin{aligned} \mathrm{SW}_p^p(\mu_w, \mu_{v^{(i)}}) &= \int_{\mathbb{D}} \mathrm{W}_p^p \left( (\mathcal{S}_{\boldsymbol{\psi}})_\# \mu_{\boldsymbol{w}}, (\mathcal{S}_{\boldsymbol{\psi}})_\# \mu_{\boldsymbol{v}^{(i)}} \right) \mathrm{d}u_{\mathbb{D}}(\boldsymbol{\psi}) \\ &= \int_{\mathbb{D}} \mathrm{W}_p^p \left( \sum_{j=1}^{N} w_j \delta_{\mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x}_j)}, \sum_{j=1}^{N} v_j^{(i)} \delta_{\mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x}_j)} \right) \mathrm{d}u_{\mathbb{D}}(\boldsymbol{\psi}), \end{aligned}$$

and therefore

$$\nabla_{\boldsymbol{w}} \mathrm{SW}_p^p(\mu_{\boldsymbol{w}}, \mu_{\boldsymbol{v}^{(i)}}) = \int_{\mathbb{D}} \nabla_{\boldsymbol{w}} \mathrm{W}_p^p \left( \sum_{j=1}^{N} w_j \delta_{\mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x}_j)}, \sum_{j=1}^{N} v_j^{(i)} \delta_{\mathcal{S}_{\boldsymbol{\psi}}(\boldsymbol{x}_j)} \right) \mathrm{d}u_{\mathbb{D}}(\boldsymbol{\psi}).$$

As the integrand of the last equation is handled in Theorem 5.3, we can thus compute $\nabla \mathcal{E}$ and devise a stochastic gradient descent, as synthesized in Algorithm 1, where we use $P$ projections per descent step as above.

The complexity of the computation of the gradient (41) is $\mathcal{O}(N \log N)$, as we need to sort the points $(t_j)_{j=1}^N$ and the vector $\boldsymbol{u}$, and all other operations are done in linear time. The projection $\mathrm{proj}_{\Delta_N}\colon \mathbb{R}^N \to \Delta_N$ on the probability simplex can be computed in complexity $\mathcal{O}(N \log(N))$ using the algorithm from [76], see also [19] for further numerical approaches. Therefore, one iteration of Algorithm 1 has the arithmetic complexity $\mathcal{O}(MN(\log(N){+}d)P)$. However, we want to point out that the number $N$ of points of a fixed grid generally grows with the dimension $d$.

### 5.2. Radon Wasserstein barycenters

*Radon Wasserstein barycenters* are obtained by first computing the 1D barycenter (5) for every slice, stacking them together, and then applying the pseudoinverse of the respective Radon transform, cf. [16, 43]. Denoting by $\mathcal{Z}\colon \mathcal{P}(\mathbb{X}) \to \mathcal{P}(\mathbb{D} \times \mathbb{I})$ the generalized slicing transformation, we set for $\mu_m \in \mathcal{P}_{\mathrm{ac}}(\mathbb{X})$ the Radon barycenter

$$\mathrm{Bary}_{\mathbb{X}}^{\mathcal{Z}}(\mu_m, \lambda_m)_{m=1}^M \coloneqq \mathcal{Z}^\dagger \left( \left( \mathrm{Bary}_{\mathbb{I}}(\lambda_m, (\mathcal{S}_{\boldsymbol{\psi}})_\# \mu_m)_{m=1}^M \right)_{\boldsymbol{\psi} \in \mathbb{D}} \right), \tag{42}$$

where $\mathcal{Z}^\dagger$ is the pseudoinverse whose argument is viewed as a density function on $\mathbb{D} \times \mathbb{I}$. In general, it is not clear if the pseudoinverse yields again a nonnegative function, which then gives a probability density. On $\mathbb{S}^{d-1}$, it is fulfilled for the parallel slicing $\mathcal{V}$ by Theorem 3.2, but not for the semicircular slicing, see [58, sect. 6.2].

The discretization is based on a fixed support. We describe the parallel slicing case $\mathcal{Z} = \mathcal{U}$ analogously to the semicircular case $\mathcal{Z} = \mathcal{W}$ from [58]. Let $\boldsymbol{\psi}_p \in \mathbb{S}^{d-1}$, $p \in [\![P]\!]$, be the nodes of a quadrature rule with weights $\boldsymbol{w} \in \Delta_P$, and some grid $t_\ell$, $\ell \in [\![L]\!]$, on the interval $\mathbb{I}$.

**Algorithm 1** Fixed support sliced Wasserstein Barycenter algorithm

**Require:** Support $X = (\boldsymbol{x}_j)_{1\le j\le N} \in \mathbb{X}^N \subset \mathbb{R}^{N\times d}$, weights $\boldsymbol{v}^{(i)} \in \Delta_N, \forall i \in [\![M]\!]$, step size $\tau > 0$, initialization $\boldsymbol{w}^0 \in \Delta_N$, number $P$ of slices

**for** $l = 1, ...$ **do**

$(\boldsymbol{\psi}_q)_{1\le q\le P} \leftarrow$ `generate_uniform_samples_on_` $\mathbb{D}(P)$

**for** $i \in [\![M]\!]$ and $q \in [\![P]\!]$ **do**

$(s_j)_{1\le j\le N} \leftarrow (\mathcal{S}_{\boldsymbol{\psi}_q}(\boldsymbol{x}_j))_{1\le j\le N}$

$\boldsymbol{\rho} \leftarrow \texttt{argsort}\left((s_j)_{1\le j\le N}\right)$

$\boldsymbol{t} \leftarrow \left(s_{\rho(j)}\right)_{1\le j\le N}$ ▷ sorted

$\tilde{\boldsymbol{w}} \leftarrow \texttt{cumsum}\left(w^l_{\rho(j)}\right)_{1\le j\le N-1}$ ▷ cumulative sum

$\tilde{\boldsymbol{v}} \leftarrow \texttt{cumsum}\left(v^{(i)}_{\rho(j)}\right)_{1\le j\le N-1}$

$\boldsymbol{u} \leftarrow \texttt{concatenate}(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{v}})$

$\texttt{mw} \leftarrow (1, ..., 1, 0, ..., 0) \in \mathbb{R}^{2N-2}$

$\boldsymbol{\sigma} \leftarrow \texttt{argsort}(\boldsymbol{u})$

$\boldsymbol{z} \leftarrow (u_{\sigma(j)})_{1\le j\le 2N-2}$

$\texttt{mw} \leftarrow (\texttt{mw}_{\sigma(j)})_{1\le j\le 2N-2}$

$\texttt{idx\_w} \leftarrow \texttt{cumsum}(\texttt{mw}) + 1$

$\texttt{idx\_v} \leftarrow \texttt{cumsum}(1 - \texttt{mw}) + 1$

$\boldsymbol{a} \leftarrow (|t_{\texttt{idx\_w}(j)} - t_{\texttt{idx\_v}(j)}|^p)_{1\le j\le 2N-2}$

$a_0 \leftarrow 0$

$\tilde{\boldsymbol{g}} \leftarrow \left(a_{\sigma^{-1}(k)-1} - a_{\sigma^{-1}(k)}\right)_{1\le k\le N-1}$

$\boldsymbol{g} \leftarrow \texttt{cumsum\_adj}(\tilde{\boldsymbol{g}})$ ▷ cumulative sum from right to left

$g_N = 0$

$\texttt{grad}_{i,q} = (g_{\rho^{-1}(j)})_{1\le j\le N}$

**end for**

$\texttt{grad}_{\mathbb{R}^N} \leftarrow \frac{1}{P}\sum_{i=1}^M \lambda_i \sum_{q=1}^P \texttt{grad}_{i,q}$

$\texttt{grad}_{\Delta_N} \leftarrow \mathrm{proj}_H(\texttt{grad}_{\mathbb{R}^N}) \qquad = \texttt{grad}_{\mathbb{R}^N} - \frac{1}{\sqrt{N}}\langle \texttt{grad}_{\mathbb{R}^N}, \mathbb{1}_N\rangle$

$w^{l+1} \leftarrow \mathrm{proj}_{\Delta_N}(w^l - \tau_l \texttt{grad}_{\Delta_N})$

**end for**

**Output:** $w^{l+1}$

We denote the density function of $\mu_m$ by $f^{\mu_m}$ and assume that we are given $f^{\mu_m}(\boldsymbol{\psi}_p)$ for $p \in [\![P]\!]$. Firstly, we approximate $\mathcal{U}_{\boldsymbol{\psi}_p}(t_\ell)$ via the singular value decomposition (24) for fixed truncation degree $D \in \mathbb{N}$ by

$$\mathcal{U}_{\boldsymbol{\psi}_p} f^{\mu_m}(t_\ell) \approx \sum_{n=0}^{D} \sum_{j=1}^{N_{n,d}} \lambda_{n,d}^{\mathcal{U}} Y_{n,d}^{j}(\boldsymbol{\psi}_p) \widetilde{P}_{n,d}(t_\ell) \sum_{i=1}^{P} f^{\mu_m}(\boldsymbol{\psi}_i) \overline{Y_{n,d}^{j}(\boldsymbol{\psi}_i)} w_i.$$

Secondly, we compute the density of the one-dimensional barycenters (5) of the measures $\mathcal{U}_{\boldsymbol{\psi}_p}\mu_m$. In particular, we set $g(\boldsymbol{\psi}_p, \cdot)$ as the density function of

$$\mathrm{CDT}_\omega^{-1}\left(\sum_{m=1}^{M} \lambda_m \,\mathrm{CDT}_\omega[\mathcal{U}_{\boldsymbol{\psi}_p}\mu_m]\right).$$

Using again the singular value decomposition Theorem 4.2, we note that the *Moore–Penrose pseudoinverse* [25] of $\mathcal{U}$ is given by

$$\mathcal{U}^\dagger \colon \mathrm{Range}(\mathcal{U}) \oplus \mathrm{Range}(\mathcal{U})^\perp \to L^2(\mathbb{S}^{d-1}), \quad \mathcal{U}^\dagger g = \sum_{n=0}^{\infty} \sum_{k=1}^{N_{n,d}} \frac{1}{\lambda_{n,d}^{\mathcal{U}}} \left\langle g, Y_{n,d}^{k} \widetilde{P}_{n,d} \right\rangle Y_{n,d}^{k}.$$

Finally, we discretize the Moore–Penrose pseudoinverse to approximate the density of the desired barycenter $\mathrm{Bary}_{\mathbb{S}^{d-1}}^{\mathcal{U}}$ by

$$\mathcal{U}^\dagger g(\boldsymbol{\psi}_p) \approx \sum_{n=0}^{D} \sum_{j=1}^{N_{n,d}} \frac{1}{\lambda_{n,d}^{\mathcal{U}}} Y_{n,d}^{j}(\boldsymbol{\psi}_p) \frac{1}{L} \sum_{i=1}^{P} \sum_{\ell=1}^{L} g(\boldsymbol{\psi}_i, t_\ell) \overline{Y_{n,d}^{j}(\boldsymbol{\psi}_i)} \widetilde{P}_{n,d}(t_\ell) w_i.$$

We analyze the complexity for $\mathbb{S}^2$. The sums over $n$ and $j$ constitute a nonuniform spherical Fourier transform, and the sum over $i$ its adjoint, which can both be computed efficiently in $\mathcal{O}(D^2 \log(D) + P)$ steps, see [46] and [56, sect. 9.6]. The CDT (3) and its inverse (4) have linear complexity and can be computed with the algorithm [44]. Therefore, we obtain an overall complexity of $\mathcal{O}((D^2 \log(D) + P)LM)$. Since the number of points is usually $P \sim D^2$, the complexity grows slower than for the algorithms of Section 5.1.

# 6. Numerical results

In this section, we present numerical computations of sliced barycenters between two measures on the sphere. and visualize them $\mathbb{S}^2$. On general $\mathbb{S}^{d-1}$, we only look at free-support barycenters, which do not require grids that become unhandy in higher dimension, and test their convergence behavior and execution times. We compare two notions of slicing: our parallel slicing (20) and the semicircular slicing of Remark 3.8, each for the two different notions of free and fixed support sliced Wasserstein barycenters from Section 5.1 and the Radon barycenters from Section 5.2. Further, we compare the fixed-support and Radon barycenters with the entropy-regularized version [55] of the Wasserstein barycenter (2) computed with PythonOT [27]. Finally, we consider the barycenters on SO(3) in Section 6.4. Our code is available online.[1]

[1] https://github.com/leo-buecher/Sliced-OT-Sphere

### 6.1. Free-support sliced Wasserstein barycenters on the sphere

For the free-support discretization of Section 5.1.1, we compute the parallelly sliced Wasserstein barycenter (PSB) $\mathrm{Bary}_{\mathbb{S}^2}^{\mathrm{PSW}}$ with the stochastic gradient descent (40), and the semicircular sliced Wasserstein barycenter (SSB) $\mathrm{Bary}_{\mathbb{S}^2}^{\mathrm{SSW}}$, see Remark 3.8, with the algorithm [14], which uses a similar gradient descent scheme.[2]

The von Mises–Fisher (vMF) distribution with center $\boldsymbol{\eta} \in \mathbb{S}^2$ and concentration $\kappa > 0$ has the density function

$$f(\boldsymbol{\xi}) = \frac{\kappa}{4\pi \sinh \kappa} \mathrm{e}^{\kappa \langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle}, \qquad \boldsymbol{\xi} \in \mathbb{S}^2. \tag{43}$$

In the first setting, we consider the sliced Wasserstein barycenters of two vMF distributions with centers on the equator shifted by 90° and concentration $\kappa = 100$, represented by $N = 200$ samples each. We use 1000 iterations, $P = 500$ slices, a step size of $\tau = 40$ for the PSB algorithm and 80 for the SSB algorithm, and we take samples of the uniform distribution $u_{\mathbb{S}^2}$ as initialization $X^0$. The computed sliced Wasserstein barycenters are displayed in Figure 2. We observe that the SSB is slightly more extended toward the poles.

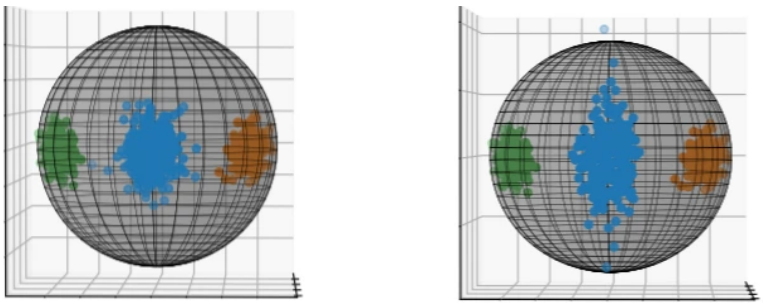


Figure 2: PSB (left) and SSB (right) of two vMF distributions. The green and orange points represent the two input measures, and the blue points the barycenter.

In the second setting, we consider vMF distributions that are highly concentrated near the poles with $\kappa = 400$, to illustrate the observations of Remark 5.1. The resulting sliced Wasserstein barycenters are shown in Figure 3. The PSB is seemingly uniform on the sphere, which corresponds to the initial distribution $X^0$. This is coherent with the observation that all measures on the sphere are PSB of two antipodal Dirac measures. Conversely, the SSB is supported on a ring around the equator.

The third setting is with two "croissant" measures spanning from the South Pole to the North Pole, and rotated from each other by an angle of 120°. The resulting PSB and SSB in Figure 4 are quite similar, which illustrates the fact that in many cases the two notions of barycenters seem more or less to coincide.

**Convergence** We study the convergence of the algorithms using the same measures as in Figure 2, but with $N = 50$ samples. Figure 5 shows the evolution of the loss function (36) and the step norm, which is the $L^2$ norm of the step $X^{l+1} - X^l$, depending on the iteration. We show the results on $\mathbb{S}^{d-1}$ for $d = 3$ and $d = 10$, where we use $\tau = 20$ for PSB and $\tau = 50$ for SSB. The two losses are rescaled so that the initial loss coincides, as they cannot be

[2] See the code https://github.com/clbonet/Spherical_Sliced-Wasserstein

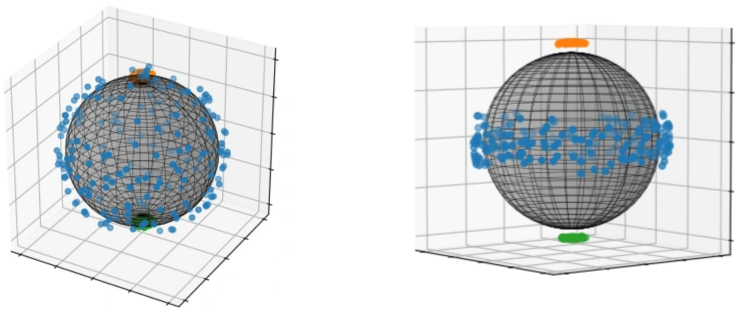


Figure 3: PSB (left) and SSB (right) of highly concentrated measures. The green and orange points represent the two input measures, and the blue points the barycenter.

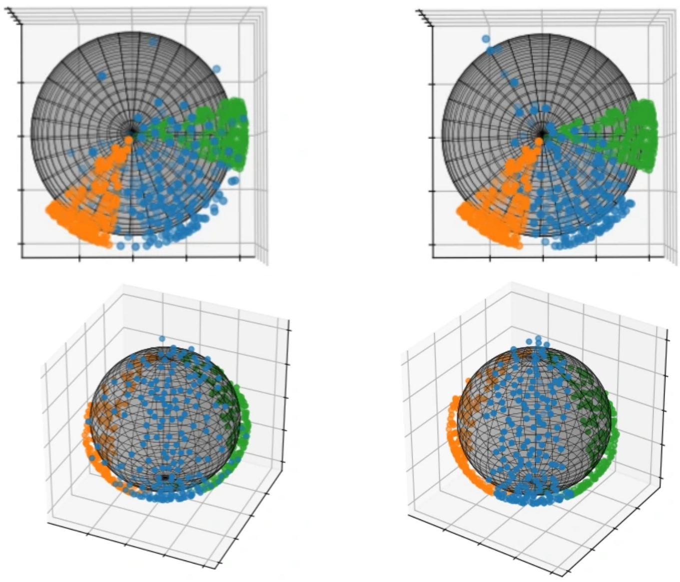


Figure 4: PSB (left) and SSB (right) for 2 "croissants" measures. The green and orange points represent the two input measures, and the blue points represent the barycenter.

compared in absolute value. Despite this, these loss evolutions remain difficult to compare, as they highly depend on the chosen step size $\tau$. However, we observed that for appropriate step sizes $\tau$ (i.e. that avoid oscillatory behavior around the minimizer), the PSB algorithm converges faster on $\mathbb{S}^2$. On $\mathbb{S}^9$, both algorithms show a similar speed of convergence. This change in dimension might be explained by the fact that a uniform measure on $\mathbb{S}^{d-1}$ (which is our initial distribution) is projected by the semicircular slicing to a uniform measure on the circle, while the parallel slicing becomes more concentrated around 0 on the unit interval $\mathbb{I}$ for larger $d$.

**Complexity** All tests were performed on an Intel Core i7-10700 with 32 GB memory. Figure 6 shows the execution time depending on the number $N$ of points of each given measure and the number $P$ of slices, both for a fixed number of 20 iterations without stopping criterion. We observe that the PSB algorithm is between 40 and 100 times faster than the SSB

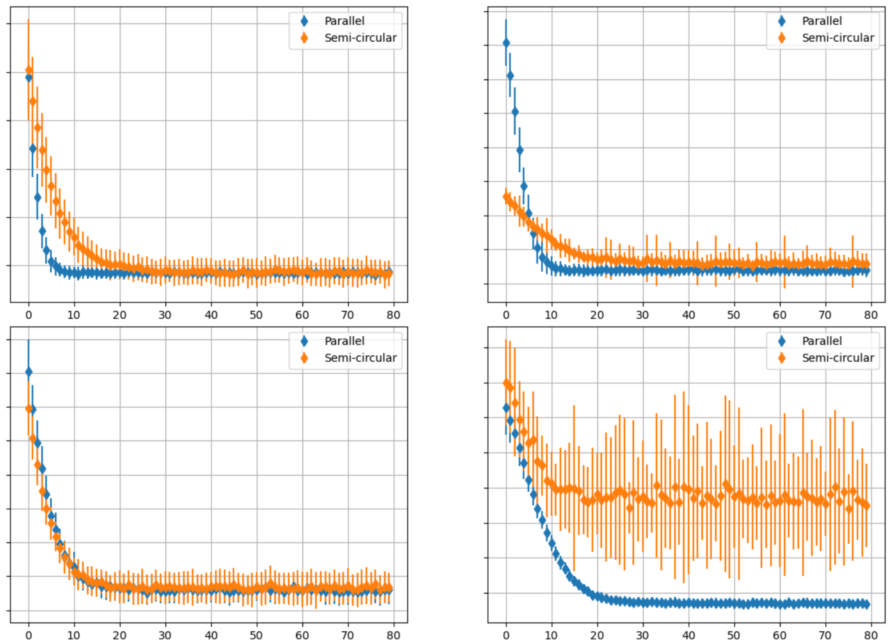



Figure 5: Loss function (left) and step norm (right) on $\mathbb{S}^{d-1}$ for $d = 3$ (top row) and $d = 10$ (bottom row) depending on the number of iterations. Reported are the mean and standard deviation (vertical lines) over 100 runs.

algorithm. This comes from the fact that the SSB requires to solve an OT problem on the torus $\mathbb{T}$, which is more difficult than on the real line. In terms of the evolution along $N$, the two algorithms (PSB, SSB) have a complexity of $O(PN(\log(N) + d))$, which is coherent with our observations. The dependence on $P$ is linear, but experiments with bigger values of $P$ would be required to confirm this.

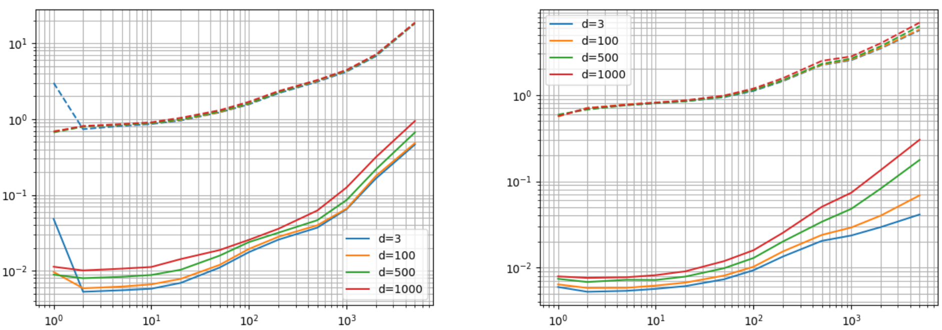


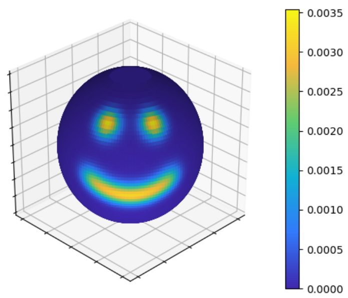

Figure 6: Execution time of the PSB (solid line) and SSB (dashed) algorithms with 20 iterations and different dimensions $d$.
Left: Dependence on the number $N$ of points with $P = 200$ slices.
Right: Dependence on the number $P$ of slices with $N = 40$ points.

The execution times barely grow with increasing dimension $d$ of the sphere. The only parts of the algorithm that depend on $d$ are the inner products in $\mathbb{R}^d$, which are well parallelized, and the generation of uniform samples on $\mathbb{S}^{d-1}$, which explains the stronger effect for larger number of slices $P$. While in Figure 6 the relative differences are higher for the generally faster PSB alrorithm, the absolute differences are usually larger for the SSB.

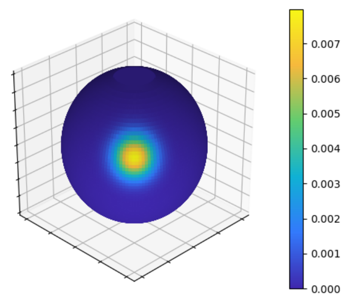
### 6.2. Fixed-support sliced Wasserstein barycenters

We test the fixed-support sliced barycenter algorithm from Section 5.1.2 and compare the resulting barycenters with the free-support ones. As input measures, we take a vMF distribution with $\kappa = 30$ and a "smiley" distribution, see Figure 7. We use the gradient descent Algorithm 1 with a grid of $150 \times 50$ points on the sphere, $P = 100$ slices, 500 iterations, the uniform distribution as initialization, and an empirically chosen step size $\tau = 0.005\,(1 + k/20)^{-1/2}$ in the $k$-th iteration. We noticed that the results highly depend on the step size.

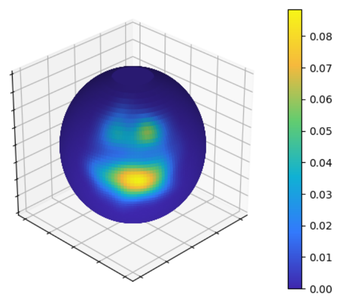
The resulting fixed-support PSB is displayed in Figure 7, along with the free-support PSB and SSB from Section 5.1.1 and the regularized Wasserstein barycenter from PythonOT with regularization parameter 0.05. For the free-support sliced Wasserstein barycenters (PSB and SSB), the two input measures are sampled $N = 200$ points and the resulting barycenters are represented using kernel density estimation [31] with the density of the vMF distributions (43) as kernel function. We notice that all barycenters look similarly.



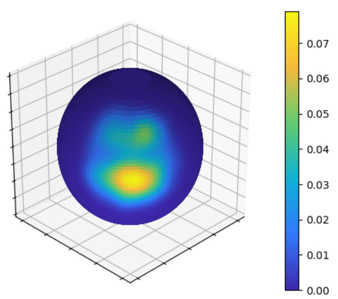

(a) First input measure



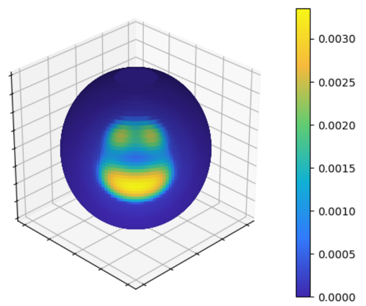

(b) Second input measure



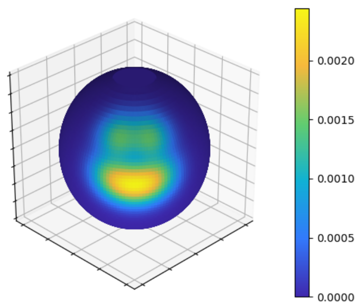

(c) Density estimation of the free-support PSB (0.05 s)




(d) Density estimation of the free-support SSB (15 s)




(e) Fixed-support PSB (107 s)

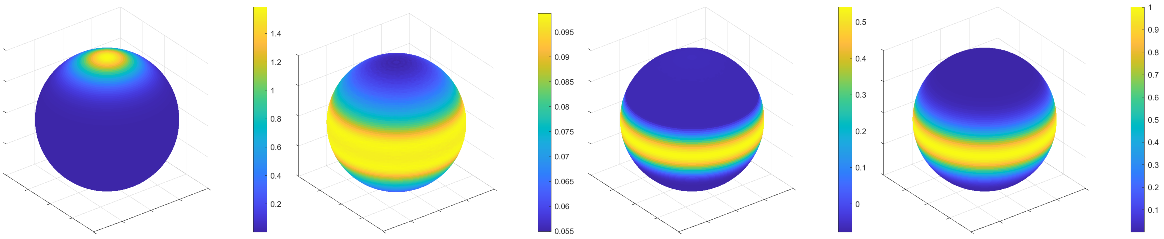



(f) Regularised Wasserstein barycenter (3.8 s)

Figure 7: Fixed-support and free-support barycenters between two input measures.

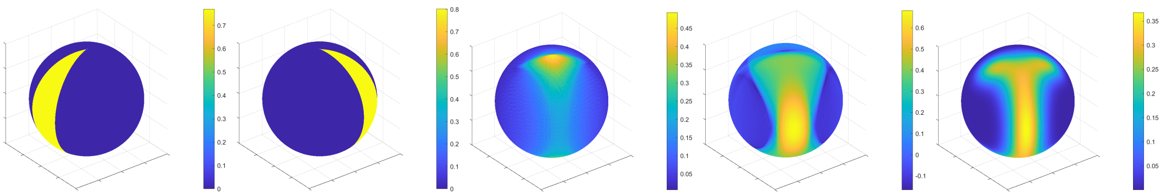
### 6.3. Radon Wasserstein barycenters on the sphere

We compare the Radon PSB $\mathrm{Bary}^{\mathcal{U}}_{\mathbb{S}^2}$ of Section 5.2 with the Radon SSB $\mathrm{Bary}^{\mathcal{W}}_{\mathbb{S}^2}$. For the latter, we apply the algorithm [58]. Both use the truncation degree $N = 120$ in (42) and $P = 29282$ slices, which equals the number of points on the $121 \times 242$ grid on $\mathbb{S}^2$. Figure 8 shows the barycenters of vMF distributions concentrated at the poles. As above, the Wasserstein barycenter is computed with the POT library with the regularization parameter 0.05 (or 0.01 for the “smiley” test). The regularized Wasserstein barycenter is somehow “between” the Radon PSB and SSB. Different from Figure 3, the Radon is concentrated on a ring around the equator, which might be explained by the fact that, other than in Remark 5.1, the measures are not Dirac measures. Furthermore, the computation of the Radon PSB is much faster than the Radon SSB. Moreover, we compare the Radon barycenters of the “croissant” shape in Figure 9 as well as the “smiley” in Figure 10. Again, the Wasserstein barycenter seems to be between the Radon PSB and SSB.



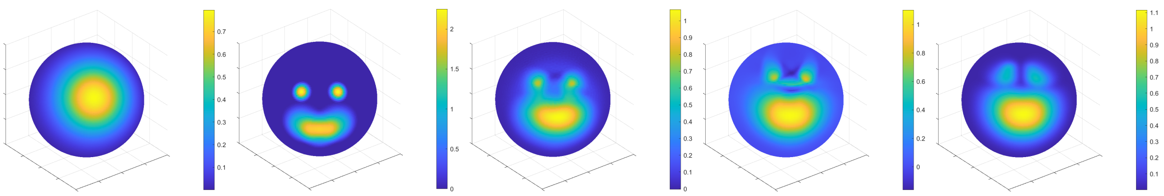
(a) First input distribution (b) Radon PSB (0.7 s) (c) Radon SSB (89 s) (d) Wasserstein barycenter (364 s)

Figure 8: First input measure (the second is the antipodal) and the Radon barycenters.



(a) First input distribution (b) Second input distribution (c) Radon PSB (d) Radon SSB (e) Wasserstein barycenter

Figure 9: Input measures and their Radon barycenters.



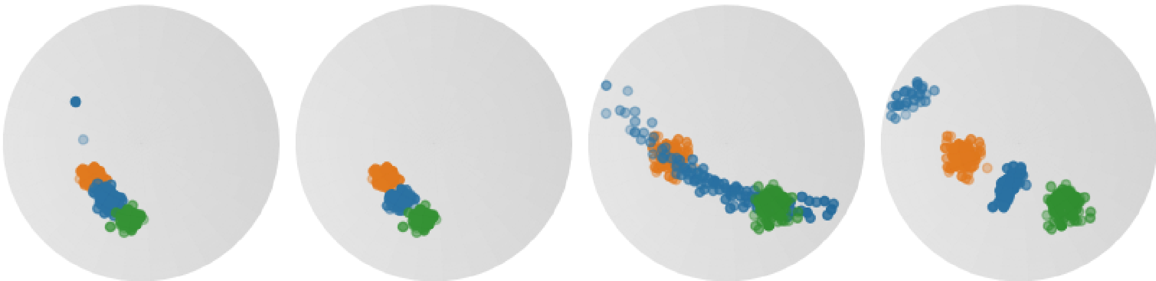
(a) First input distribution (b) Second input distribution (c) Radon PSB (d) Radon SSB (e) Wasserstein barycenter

Figure 10: Input measures and their Radon barycenters.

## 6.4. Sliced barycenters on the rotation group

We compute free-support sliced Wasserstein barycenters on SO(3) with the stochastic gradient descent method outlined in paragraph 5.1.1.2. Similar to the spherical setting, we start with two given measures with $N = 100$ points each. Since there is no other slicing approach on SO(3), we compare it with a Wasserstein barycenter that is obtained by computing an optimal transport map for the 1-Wasserstein distance (1) using PythonOT and applying the logarithm map to project it to the manifold SO(3).

We visualize a rotation $Q = \mathrm{R}_{\boldsymbol{n}}(\omega) \in \mathrm{SO}(3)$, see (13), as the point $\tan(\frac{\omega}{4})\,\boldsymbol{n} \in \mathbb{R}^3$, where $\omega \in [0, \pi]$ is the angle and $\boldsymbol{n} \in \mathbb{S}^2$ is the rotation axis, cf. [24, p. 1633]. The resulting free-support barycenters between two input measures are depicted in Figure 11. In (a) and (b), the measures are closer together, the angle between their centers is approximately 92 degree, and notions of barycenters are similar. In (c) and (d), the input measures are almost opposite to each other with a distance of 164 degrees, then the sliced barycenter is supported on a curve with high clustering between the inputs, and the projected Wasserstein barycenter is stronger localized around two locations.



(a) Sliced Wasserstein barycenter (b) Wasserstein barycenter (c) Sliced Wasserstein barycenter (d) Wasserstein barycenter

Figure 11: Free support barycenters on SO(3). Blue points represent the barycenter, while green and orange points represent the two input measures, which are closer together in (a) and (b) or further away in (c) and (d).

# 7. Conclusions

We investigated a new approach for sliced Wasserstein distance of spherical measures and proved that this parallel slicing provides a rotational invariant metric on $\mathcal{P}(\mathbb{S}^{d-1})$ that induces the same topology as the Wasserstein distance. We provided numerical algorithms for the computation of the respective sliced barycenters, both with free or fixed support, that are considerably faster than for the semicircular slicing, while producing comparable results in most cases, except when the input measures are highly concentrated around antipodal points.

Extending our method to the rotation group SO(3), we proved the metric properties of the proposed sliced Wasserstein distance based on a new Radon transform on SO(3) and its singular value decomposition. An extensive numerical evaluation of the latter transform is planned, but out of the scope of this paper. Further, it will be interesting to incorporate our slicing approach into gradient flows on $\mathbb{S}^{d-1}$ for $d \gg 2$.

#### Acknowledgements

We gratefully acknowledge the funding by the German Research Foundation (DFG): STE 571/19-1, project number 495365311, within the Austrian Science Fund (FWF) SFB 10.55776/F68: "Tomography Across the Scales". L.B. acknowledges funding from the TU Berlin, Institute of Mathematics, during his internship in 2023. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

# A. Metric properties of the parallelly sliced Wasserstein distance

In the following, we show bounds between the sliced Wasserstein distance $\mathrm{PSW}_p$ and the spherical Wasserstein distance $\mathrm{W}_p$ by a relation to the Euclidean case [17, Sect. 5.1].

**Lemma A.1.** *The geodesic distance (7) on the sphere and the Euclidean distance are related via*

$$\|\boldsymbol{\xi}-\boldsymbol{\eta}\| \le d(\boldsymbol{\xi},\boldsymbol{\eta}) \le \frac{\pi}{2}\|\boldsymbol{\xi}-\boldsymbol{\eta}\|, \qquad \forall \boldsymbol{\xi},\boldsymbol{\eta}\in\mathbb{S}^{d-1}.$$

*Proof.* We first show that

$$\sqrt{2-2x} \le \arccos(x) \le \frac{\pi}{2}\sqrt{2-2x}, \qquad \forall x\in\mathbb{I}. \tag{44}$$

For $x\in\mathbb{I}$, we have by [1, 4.4.2]

$$\arccos(x) = \int_x^1 \frac{1}{\sqrt{1-t^2}}\,\mathrm{d}t = \int_x^1 \frac{1}{\sqrt{1+t}\sqrt{1-t}}\,\mathrm{d}t \ge \int_x^1 \frac{1}{\sqrt{2-2t}}\,\mathrm{d}t = \sqrt{2-2x},$$

which is the first inequality of (44). Analogously, we have for $x\in[0,1]$ that

$$\arccos(x) = \int_x^1 \frac{1}{\sqrt{1-t^2}}\,\mathrm{d}t \le \int_x^1 \frac{1}{\sqrt{1-t}}\,\mathrm{d}t = 2\sqrt{1-x}.$$

Because arccos is convex on $[-1,0]$ and $\arccos(-1)=\pi$, we obtain the second inequality of (44) for all $x\in[-1,1]$. The assertion follows from the fact that

$$\|\boldsymbol{\xi}-\boldsymbol{\eta}\| = \sqrt{\|\boldsymbol{\xi}\|^2+\|\boldsymbol{\eta}\|^2-2\langle\boldsymbol{\xi},\boldsymbol{\eta}\rangle} = \sqrt{2-2\langle\boldsymbol{\xi},\boldsymbol{\eta}\rangle}$$

and $d(\boldsymbol{\xi},\boldsymbol{\eta}) = \arccos(\langle\boldsymbol{\xi},\boldsymbol{\eta}\rangle)$ for all $\boldsymbol{\xi},\boldsymbol{\eta}\in\mathbb{S}^{d-1}$ by (7). $\square$

We extend a spherical measure $\mu\in\mathcal{P}(\mathbb{S}^{d-1})$ to a measure $\tilde{\mu}\in\mathcal{P}(\mathbb{R}^d)$ that is supported on $\mathbb{S}^{d-1}$ by setting

$$\tilde{\mu}(B) \coloneqq \mu(B\cap\mathbb{S}^{d-1}), \qquad \forall B\in\mathcal{B}(\mathbb{R}^d).$$

**Lemma A.2.** *Let $\mu,\nu\in\mathcal{P}(\mathbb{S}^{d-1})$ be spherical measures with extensions $\tilde{\mu},\tilde{\nu}\in\mathcal{P}(\mathbb{R}^d)$. Then the following inequalities between the spherical Wasserstein distance $\mathrm{W}_p(\mu,\nu)$ and the Euclidean Wasserstein distance $\mathrm{W}_p(\tilde{\mu},\tilde{\nu})$ on $\mathbb{R}^d$ holds:*

$$\mathrm{W}_p(\tilde{\mu},\tilde{\nu}) \le \mathrm{W}_p(\mu,\nu) \le \frac{\pi}{2}\,\mathrm{W}_p(\tilde{\mu},\tilde{\nu}).$$

*Proof.* The Euclidean Wasserstein distance on $\mathbb{X} = \mathbb{R}^d$ is given in (1) with $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$. Any transport plan $\tilde{\gamma} \in \Pi(\tilde{\mu}, \tilde{\nu})$ is supported only on $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$. Hence, its restriction to $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ yields a transport plan in $\Pi(\mu, \nu)$. Conversely, a transport plan $\gamma \in \Pi(\mu, \nu)$ can be extended to $\mathbb{R}^d$ by setting it to zero outside the sphere. Hence, we have

$$\mathrm{W}_p^p(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}} d(\boldsymbol{x}, \boldsymbol{y}) \,\mathrm{d}\gamma(\boldsymbol{x}, \boldsymbol{y}) = \inf_{\tilde{\gamma} \in \Pi(\tilde{\mu},\tilde{\nu})} \int_{\mathbb{R}^d \times \mathbb{R}^d} d(\boldsymbol{x}, \boldsymbol{y}) \,\mathrm{d}\tilde{\gamma}(\boldsymbol{x}, \boldsymbol{y}).$$

The claim follows by Lemma A.1. □

*Proof of Theorem 3.7.* We first show (31) via applying the respective result from the Euclidean space. We briefly recall sliced OT on $\mathbb{R}^d$, see [16], with the slicing operator $\mathcal{S}_{\boldsymbol{\psi}}^{\mathbb{R}^d} : \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{x} \mapsto \langle \boldsymbol{\psi}, \boldsymbol{x} \rangle$ for $\boldsymbol{\psi} \in \mathbb{S}^{d-1}$ and the sliced Wasserstein distance

$$\mathrm{RSW}_p^p(\mu, \nu) \coloneqq \int_{\mathbb{S}^{d-1}} \mathrm{W}_p^p \left( (\mathcal{S}_{\boldsymbol{\psi}}^{\mathbb{R}^d})_\# \mu, (\mathcal{S}_{\boldsymbol{\psi}}^{\mathbb{R}^d})_\# \nu \right) \mathrm{d}u_{\mathbb{S}^{d-1}}(\boldsymbol{\psi}), \qquad \mu, \nu \in \mathcal{P}(\mathbb{R}^d).$$

Comparing the Euclidean sliced Wasserstein distance RSW with the spherical sliced Wasserstein distance (30), we see that

$$\mathrm{PSW}_p(\mu, \nu) = \mathrm{RSW}_p(\tilde{\mu}, \tilde{\nu}). \tag{45}$$

By [17, thm. 5.1.5], there exist constants $\tilde{c}_{d,p}, \tilde{C}_{d,p}$ such that for all measures $\tilde{\mu}, \tilde{\nu} \in \mathcal{P}(\mathbb{R}^d)$ which are supported in a ball of fixed radius $R > 0$ we have

$$\mathrm{RSW}_p(\tilde{\mu}, \tilde{\nu}) \le \tilde{c}_{d,p} \,\mathrm{W}_p(\tilde{\mu}, \tilde{\nu}) \le \tilde{C}_{d,p} R^{1 - \frac{1}{p(d+1)}} \,\mathrm{RSW}_p(\tilde{\mu}, \tilde{\nu})^{\frac{1}{p(d+1)}}.$$

As $\tilde{\nu}$ and $\tilde{\mu}$ are by construction supported in a ball of radius $1 + \varepsilon$ for any $\varepsilon > 0$, the validity of (31) follows by invoking (45) and Lemma A.2.

Next we prove the metric properties. The symmetry and the triangle inequality follow from the corresponding properties of the Wasserstein distance and the $p$-norm on $\mathbb{S}^{d-1}$. The positive definiteness and the equivalence to the spherical Wasserstein distance follow from (31). The rotational invariance of PSW follows since $\mathcal{U}$ is rotationally invariant. □

## B. Proofs from Section 4

*Proof of Theorem 4.2.* Let $n \in \mathbb{N}_0$ and $j, k \in \{-n, \dots, n\}$. Using the product identity [36, cor. 2.11]

$$D_n^{j,k}(\boldsymbol{P}\boldsymbol{Q}) = \sum_{\ell=-n}^{n} D_n^{j,\ell}(\boldsymbol{P}) D_n^{\ell,k}(\boldsymbol{Q}) \qquad \forall \boldsymbol{P}, \boldsymbol{Q} \in \mathrm{SO}(3), \tag{46}$$

we have

$$\mathcal{T} D_n^{j,k}(\boldsymbol{Q}, \omega) = \frac{1}{4\pi^2}(1 - \cos(\omega)) \sum_{\ell=-n}^{n} D_n^{j,\ell}(\boldsymbol{Q}) \int_{\mathbb{S}^2} D_n^{\ell,k}(\mathrm{R}_{\boldsymbol{\xi}}(\omega)) \,\mathrm{d}\sigma_{\mathbb{S}^2}(\boldsymbol{\xi}).$$

We write $\boldsymbol{\xi} = \Phi(\varphi, \vartheta)$ with the spherical coordinates (10). Since

$$\mathrm{R}_{\Phi(\varphi,\vartheta)}(\omega) = \mathrm{R}_{\boldsymbol{e}^3}(\varphi) \mathrm{R}_{\boldsymbol{e}^2}(\vartheta) \mathrm{R}_{\boldsymbol{e}^3}(\varphi) \mathrm{R}_{\boldsymbol{e}^3}(\omega) \mathrm{R}_{\boldsymbol{e}^2}(-\vartheta) \mathrm{R}_{\boldsymbol{e}^3}(-\varphi) = \Psi(\varphi, \vartheta, 0)\, \Psi(\omega, -\vartheta, -\varphi),$$

we have by (16) and (46)

$$D_n^{j,k}(\mathrm{R}_{\Phi(\varphi,\vartheta)}(\omega)) = \sum_{\ell=-n}^{n} D_n^{j,\ell}(\Psi(\varphi,\vartheta,0))\,\mathrm{e}^{-\mathrm{i}\ell\omega}\,D_n^{\ell,k}(\Psi(0,-\vartheta,-\varphi)),$$

cf. [73, § 4.5]. Hence, we obtain

$$\mathcal{T}D_n^{j,k}(\boldsymbol{Q},\omega) = \frac{1}{4\pi^2}(1-\cos(\omega))$$
$$\sum_{\ell=-n}^{n} D_n^{j,\ell}(\boldsymbol{Q}) \sum_{m=-n}^{n} \int_0^{\pi}\int_{\mathbb{T}} D_n^{\ell,m}(\Psi(\varphi,\vartheta,0))\mathrm{e}^{-\mathrm{i}m\omega} D_n^{m,k}(\Psi(0,-\vartheta,-\varphi))\,\mathrm{d}\varphi\sin(\vartheta)\,\mathrm{d}\vartheta.$$

With the symmetry $D_n^{m,k}(\boldsymbol{Q}) = \overline{D_n^{k,m}(\boldsymbol{Q}^\top)}$ and (16), we see that

$$\mathcal{T}D_n^{j,k}(\boldsymbol{Q},\omega) = \frac{1}{4\pi^2}(1-\cos(\omega))$$
$$\sum_{\ell=-n}^{n} D_n^{j,\ell}(\boldsymbol{Q}) \sum_{m=-n}^{n} \mathrm{e}^{-\mathrm{i}m\omega} \int_0^{\pi}\int_{\mathbb{T}} \mathrm{e}^{-\mathrm{i}\ell\varphi} d_n^{\ell,m}(\cos(\vartheta)) d_n^{k,m}(\cos(\vartheta))\mathrm{e}^{\mathrm{i}m\varphi}\,\mathrm{d}\varphi\sin(\vartheta)\,\mathrm{d}\vartheta.$$

With the orthogonality of the exponentials and the d-functions in (18), we obtain

$$\mathcal{T}D_n^{j,k}(\boldsymbol{Q},\omega) = (1-\cos(\omega))\,\frac{1}{(2n+1)\pi} D_n^{j,k}(\boldsymbol{Q}) \sum_{\ell=-n}^{n} \mathrm{e}^{-\mathrm{i}\ell\omega}.$$

The expansion relation of the Dirichlet kernel

$$\sum_{\ell=-n}^{n} \mathrm{e}^{-\mathrm{i}\ell\omega} = \frac{\sin((n+\frac{1}{2})\omega)}{\sin(\frac{\omega}{2})}$$

and the half angle formula $(1-\cos(\omega)) = 2\sin(\omega/2)^2$ yield

$$\mathcal{T}D_n^{j,k}(\boldsymbol{Q},\omega) = \frac{2}{(2n+1)\pi} D_n^{j,k}(\boldsymbol{Q})\,\sin\left((n+\tfrac{1}{2})\omega\right)\,\sin(\tfrac{\omega}{2}),$$

which implies (34). The orthogonality of $F_n^{j,k}$ follows from the orthonormality (17) of the rotational harmonics $\widetilde{D}_n^{j,k}$. Using the identity $(\sin(\frac{\omega}{2}))^2 = (1+\cos(\omega))/2$ and the orthogonality of the cosine, we obtain

$$\int_0^{\pi} \left(\sin\left((n+\tfrac{1}{2})\omega\right)\right)^2 \left(\sin(\tfrac{\omega}{2})\right)^2\,\mathrm{d}\omega = \begin{cases} \pi/4, & n\in\mathbb{N}, \\ 3\pi/8, & n=0. \end{cases} \qquad \square$$

*Proof of Theorem 4.7.* This proof uses a similar structure as [58, Thm. 3.7]. Let $\mu,\nu \in \mathcal{M}(\mathrm{SO}(3))$ such that $\mathcal{T}\mu = \mathcal{T}\nu$. For $g \in C(\mathrm{SO}(3)\times\mathbb{I})$, we have by Proposition 3.5

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

We show that $\{\mathcal{T}^* g : g \in C(\mathrm{SO}(3)\times\mathbb{I})\}$ is a dense subset of $C(\mathrm{SO}(3))$, which implies $\mu = \nu$.

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

$$\|f\|_{H^s(\mathrm{SO}(3))}^2 \coloneqq \sum_{n=0}^{\infty} \left(n+\tfrac{1}{2}\right)^{2s} \sum_{j,k=-n}^{n} \frac{8\pi^2}{2n+1} |\langle f, D_n^{j,k}\rangle|^2. \tag{47}$$

Let $s > 2$, then $H^s(\mathrm{SO}(3))$ is dense in $C(\mathrm{SO}(3))$, cf. [36, Lem. 2.22]. Let $f \in H^s(\mathrm{SO}(3))$. Since $\mathcal{T}$ is injective on $L^2(\mathrm{SO}(3))$ by Theorem 4.2, we have $f = \mathcal{T}^* g$ if and only if $\mathcal{T} f = \mathcal{T}\mathcal{T}^* g$. In the following, we show that

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

is in $C(\mathrm{SO}(3) \times [0,\pi])$, then we obtain $f = \mathcal{T}^* g$, which shows that $H^s(\mathrm{SO}(3)) \subset \mathcal{T}^*(C(\mathrm{SO}(3) \times [0,\pi])$ and therefore the assertion.

Since $\mathcal{T}^*$ has the same singular functions as $\mathcal{T}$ and the conjugate singular values, we obtain by the singular value decomposition (34) that

$$(\mathcal{T}\mathcal{T}^*)^{-1} \mathcal{T} f = \sum_{n=0}^{\infty} \sum_{j,k=-n}^{n} \frac{1}{\lambda_n^{\mathcal{T}}} \left\langle f, D_n^{j,k} \right\rangle_{L^2(\mathrm{SO}(3))} F_n^{j,k}. \tag{48}$$

We want to show that the right-hand side of (48) converges uniformly on $C(\mathrm{SO}(3) \times [0,\pi])$. Let $(\boldsymbol{Q},\omega) \in \mathrm{SO}(3) \times [0,\pi]$ and $N \in \mathbb{N}$. Inserting $\lambda_n^{\mathcal{T}}$ from Theorem 4.2, we have

$$\left| \sum_{n=0}^{\infty} \sum_{j,k=-n}^{n} \frac{1}{\lambda_n^{\mathcal{T}}} \left\langle f, D_n^{j,k} \right\rangle_{L^2(\mathrm{SO}(3))} F_n^{j,k}(\boldsymbol{Q},\omega) - \sum_{n=0}^{N-1} \sum_{j,k=-n}^{n} \frac{1}{\lambda_n^{\mathcal{T}}} \left\langle f, D_n^{j,k} \right\rangle_{L^2(\mathrm{SO}(3))} F_n^{j,k}(\boldsymbol{Q},\omega) \right|$$

$$\le \frac{1}{2} \sum_{n=N}^{\infty} \sum_{j,k=-n}^{n} \left(n+\tfrac{1}{2}\right) \left| \left\langle f, D_n^{j,k} \right\rangle_{L^2(\mathbb{S}^2)} \right| \left| \widetilde{D}_n^{j,k}(\boldsymbol{Q}) \right|$$

$$\le \frac{1}{2} \sqrt{\sum_{n=N}^{\infty} \sum_{j,k=-n}^{n} \left(n+\tfrac{1}{2}\right)^{2s} \left| \left\langle f, D_n^{j,k} \right\rangle_{L^2(\mathbb{S}^2)} \right|^2} \sqrt{\sum_{n=N}^{\infty} \left(n+\tfrac{1}{2}\right)^{2-2s} \sum_{j,k=-n}^{n} \left| \widetilde{D}_n^{j,k}(\boldsymbol{Q}) \right|},$$

where we made use of the Cauchy–Schwarz inequality. In the last equation, the first part is bounded by the Sobolev norm (47). For the second part, the addition theorem [36, Thm. 2.14] yields

$$\sum_{n=N}^{\infty} \left(n+\tfrac{1}{2}\right)^{2-2s} \sum_{j,k=-n}^{n} \left| \widetilde{D}_n^{j,k}(\boldsymbol{Q}) \right| = \sum_{n=N}^{\infty} \left(n+\tfrac{1}{2}\right)^{2-2s} \frac{(2n+1)(n+1)}{8\pi^2} < \infty$$

since $s > 2$. Hence, the right-hand side of (48) converges uniformly to a continuous function on $\mathrm{SO}(3) \times [0,\pi]$, which finally implies that $g$ is continuous. $\square$

## C. Relation of the rotation group with the 3-sphere

We show a relation of the sliced Wasserstein distance (35) with an analogue of PSW on the sphere $\mathbb{S}^3$, making use of the representation of SO(3) via unit quaternions, see [51, § 2.6].

The algebra of quaternions $\mathbb{H}$ consists of vectors $\boldsymbol{q} = (q_0, q_1, q_2, q_3) = (q_0, \boldsymbol{q}') \in \mathbb{R}^4$, where $\boldsymbol{q}' = (q_1, q_2, q_3)$ is called the vector part, with the standard addition and the multiplication

$$\boldsymbol{q} \diamond \boldsymbol{r} := (r_0 q_0 - \boldsymbol{q}' \cdot \boldsymbol{r}', q_0 \boldsymbol{r}' + r_0 \boldsymbol{q}' + \boldsymbol{q}' \times \boldsymbol{r}'), \tag{49}$$

where $\cdot$ is the scalar product and $\times$ is the cross product in $\mathbb{R}^3$. The unit quaternions $\{q \in \mathbb{H} \mid q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1\}$ can be identified with $\mathbb{S}^3$. The inverse of $\boldsymbol{q} \in \mathbb{S}^3$ with respect to $\diamond$ is $\bar{\boldsymbol{q}} = (q_0, -q_1, -q_2, -q_3)$. The map

$$\phi \colon \mathbb{S}^3 \to \mathrm{SO}(3), \quad \boldsymbol{q} \mapsto \mathrm{R}_{\boldsymbol{q}'/\sqrt{1-q_0^2}}(2\arccos(q_0))$$

is surjective and satisfies $\phi^{-1}(\phi(\boldsymbol{q})) = \{\boldsymbol{q}, -\boldsymbol{q}\}$ for all $\boldsymbol{q} \in \mathbb{S}^3$. It is a homomorphism in the sense that $\phi(\boldsymbol{q} \diamond \boldsymbol{r}) = \phi(\boldsymbol{q})\phi(\boldsymbol{r})$. By [36, (2.6)], the integral on SO(3) is transformed via

$$\begin{aligned}\int_{\mathrm{SO}(3)} f(\boldsymbol{Q}) \,\mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{Q}) &= \int_{-1}^{1} f \circ \phi(q_0, \boldsymbol{q}') 4\sqrt{1-q_0^2} \,\mathrm{d}q_0 \,\mathrm{d}\sigma_{\mathbb{S}^2}\left(\sqrt{1-q_0^2}\boldsymbol{q}'\right) \\ &\overset{(9)}{=} 4 \int_{\mathbb{S}^3} f \circ \phi(\boldsymbol{q}) \,\mathrm{d}\sigma_{\mathbb{S}^3}(\boldsymbol{q}).\end{aligned} \tag{50}$$

We denote the set of even probability measures on SO(3) by

$$\mathcal{P}_{\mathrm{even}}(\mathbb{S}^3) := \{\mu \in \mathcal{P}(\mathbb{S}^3) \mid \mu(B) = \mu(-B) \,\forall B \in \mathcal{B}(\mathbb{S}^3)\}.$$

**Theorem C.1.** *Let $\mu, \nu \in \mathcal{P}(\mathrm{SO}(3))$ and $p \in [1, \infty)$. We define $c \colon \mathbb{I} \to [0, \pi]$, $c(t) := 2\arccos|t|$. Then*

$$\mathrm{SOSW}_p^p(\mu, \nu) = \int_{\mathbb{S}^3} \mathrm{W}_p^p\left(c_\# \mathcal{U}_{\boldsymbol{q}}(\mu \circ \phi), c_\# \mathcal{U}_{\boldsymbol{q}}(\nu \circ \phi)\right) \mathrm{d}u_{\mathbb{S}^3}(\boldsymbol{q}). \tag{51}$$

*Proof.* Let $\boldsymbol{\xi} \in \mathbb{S}^3$, $\boldsymbol{Q} \in \mathrm{SO}(3)$ and $q \in \phi^{-1}(\mathrm{SO}(3))$. By the multiplication-invariance of $\phi$, we have

$$d_{\boldsymbol{Q}} \circ \phi(\boldsymbol{\xi}) = \angle(\boldsymbol{Q}^\top \phi(\boldsymbol{\xi})) = \angle(\phi(\bar{\boldsymbol{q}} \diamond \boldsymbol{\xi})).$$

We note that since the pushforward $\phi_\#$ is bijective from $\mathcal{P}_{\mathrm{even}}(\mathbb{S}^3)$ to $\mathcal{P}(\mathrm{SO}(3))$, and its inverse is given by $\mu \circ \phi \in \mathcal{P}_{\mathrm{even}}(\mathbb{S}^3)$. Since $\angle(\phi(\boldsymbol{r})) = 2\arccos|r_0|$ for any $\boldsymbol{r} \in \mathbb{S}^3$, we obtain by (49)

$$d_{\boldsymbol{Q}} \circ \phi(\boldsymbol{\xi}) = 2\arccos\left|q_0 \xi_0 + \boldsymbol{q}' \cdot \boldsymbol{\xi}'\right| = 2\arccos|\boldsymbol{q} \cdot \boldsymbol{\xi}| = c \circ \mathcal{S}_{\boldsymbol{q}}(\boldsymbol{\xi}).$$

Since $\phi_\#(\mu \circ \phi) = \mu$, we obtain

$$\begin{aligned}\mathrm{SOSW}_p^p(\mu, \nu) &\overset{(35)}{=} \frac{1}{8\pi^2} \int_{\mathrm{SO}(3)} \mathrm{W}_p^p\left((d_{\boldsymbol{Q}} \circ \phi)_\#(\mu \circ \phi), (d_{\boldsymbol{Q}} \circ \phi)_\#(\nu \circ \phi)\right) \mathrm{d}\sigma_{\mathrm{SO}(3)}(\boldsymbol{Q}) \\ &\overset{(50)}{=} \frac{4}{8\pi^2} \int_{\mathbb{S}^3} \mathrm{W}_p^p\left((c \circ \mathcal{S}_{\boldsymbol{q}})_\#(\mu \circ \phi), (c \circ \mathcal{S}_{\boldsymbol{q}})_\#(\nu \circ \phi)\right) \mathrm{d}\sigma_{\mathbb{S}^3}(\boldsymbol{q}).\end{aligned}$$

□

The right-hand side of (51) mimics the parallelly sliced Wasserstein distance (30) on $\mathbb{S}^3$ between $\mu \circ \phi$ and $\nu \circ \phi$, except for the additional transformation $c$. We want to point out that this equivalence in Theorem C.1 holds only for even measures on $\mathbb{S}^3$, since we always have $\mu \circ \phi \in \mathcal{P}_{\mathrm{even}}(\mathbb{S}^3)$.

## D. Sliced Wasserstein distances for antipodal point measures

We study sliced Wasserstein barycenters of two antipodal Dirac measures $\mu_1 = \delta_{\boldsymbol{e}^3}$ and $\mu_2 = \delta_{-\boldsymbol{e}^3}$ on the sphere $\mathbb{S}^2$, as presented in Remark 5.1. This case exhibits some differences between the Wasserstein distance, the parallelly sliced Wasserstein distance and the semicircular sliced Wasserstein distance. We define the equator $C \coloneqq \{\boldsymbol{\xi} \in \mathbb{S}^2 \mid \xi_3 = 0\}$.

**Proposition D.1.** *The 2-Wasserstein barycenters of the two antipodal Dirac measures $\mu_1$ and $\mu_2$ are the probability measures $\nu \in \mathcal{P}(\mathbb{S}^2)$ whose support is included in the equator $C$.*

*Proof.* The Wasserstein barycenters of $\mu_1$ and $\mu_2$ on the sphere are the measures in $\mathcal{P}(\mathbb{S}^2)$ minimizing $\mathcal{E}_{\mathrm{W}}(\nu) \coloneqq \frac{1}{2}\,\mathrm{W}_2^2(\nu, \mu_1) + \frac{1}{2}\,\mathrm{W}_2^2(\nu, \mu_2)$. Let $\nu \in \mathcal{P}(\mathbb{S}^2)$. With the map $\mathrm{zen}\colon \mathbb{S}^2 \to [0, \pi]$, $\boldsymbol{\xi} \mapsto \arccos(\xi_3)$, we have

$$\mathrm{W}_2^2(\nu, \delta_{\boldsymbol{e}^3}) = \int_{\mathbb{S}^2} d_{\mathbb{S}^2}(\boldsymbol{\psi}, \boldsymbol{e}^3)^2 \,\mathrm{d}\nu(\boldsymbol{\psi}) = \int_0^\pi t^2 \,\mathrm{d}(\mathrm{zen})_\#\nu(t)$$

and similarly

$$\mathrm{W}_2^2(\nu, \delta_{-\boldsymbol{e}^3}) = \int_0^\pi (\pi - t)^2 \;\mathrm{d}(\mathrm{zen})_\#\nu(t).$$

Hence

$$\mathcal{E}_{\mathrm{W}}(\nu) = \frac{1}{2}\int_0^\pi \left[t^2 + (\pi - t)^2\right] \mathrm{d}(\mathrm{zen})_\#\nu(t).$$

The integrand has a unique minimizer $t = \frac{\pi}{2}$ and its minimum is $\frac{\pi^2}{4}$. Therefore, $\mathcal{E}_{\mathrm{W}}(\nu) \geq \frac{\pi^2}{4}$ for all $\nu \in \mathcal{P}(\mathbb{S}^2)$ with equality if and only if $(\mathrm{zen})_\#\nu(dt) = \delta_{\frac{\pi}{2}}$, i.e., if $\nu(C) = 1$. □

However, this observation does not apply to the parallelly sliced Wasserstein barycenters.

**Proposition D.2.** *All probability measures on the sphere are parallelly sliced Wasserstein barycenters of $\mu_1$ and $\mu_2$.*

*Proof.* Let $\nu \in \mathcal{P}(\mathbb{S}^2)$. We have

$$\begin{aligned}
\mathcal{E}_V(\nu) &\coloneqq \frac{1}{2}\,\mathrm{PSW}_2^2(\nu, \delta_{\boldsymbol{e}^3}) + \frac{1}{2}\,\mathrm{PSW}_2^2(\nu, \delta_{-\boldsymbol{e}^3}) \\
&= \frac{1}{2}\int_{\mathbb{S}^2} \Big[\mathrm{W}_2^2\left(\mathcal{U}_{\boldsymbol{\psi}}\nu, \mathcal{U}_{\boldsymbol{\psi}}\delta_{\boldsymbol{e}^3}\right) + \mathrm{W}_2^2\left(\mathcal{U}_{\boldsymbol{\psi}}\nu, \mathcal{U}_{\boldsymbol{\psi}}\delta_{-\boldsymbol{e}^3}\right)\Big] \,\mathrm{d}u_{\mathbb{S}^2}(\boldsymbol{\psi}).
\end{aligned}$$

Using that $\mathcal{U}_{\boldsymbol{\psi}}\delta_{\pm\boldsymbol{e}^3} = \delta_{\langle\boldsymbol{\psi}, \pm\boldsymbol{e}^3\rangle} = \delta_{\pm\psi_3}$ and $\mathcal{U}_{\boldsymbol{\psi}}\nu = (\mathcal{S}_{\boldsymbol{\psi}})_\#\nu$ for all $\boldsymbol{\psi} \in \mathbb{S}^2$, we have

$$\begin{aligned}
\mathcal{E}_V(\nu) &= \frac{1}{2}\int_{\mathbb{S}^2} \left[\int_{-1}^1 \left|t - \langle\boldsymbol{\psi}, \boldsymbol{e}^3\rangle\right|^2 \,\mathrm{d}(\mathcal{S}_{\boldsymbol{\psi}})_\#\nu(t) + \int_{-1}^1 \left|t + \langle\boldsymbol{\psi}, \boldsymbol{e}^3\rangle\right|^2 \,\mathrm{d}(\mathcal{S}_{\boldsymbol{\psi}})_\#\nu(t)\right] \mathrm{d}u_{\mathbb{S}^2}(\boldsymbol{\psi}) \\
&= \frac{1}{2}\int_{\mathbb{S}^2}\int_{-1}^1 \left[2t^2 + 2\left\langle\boldsymbol{\psi}, \boldsymbol{e}^3\right\rangle^2\right] \mathrm{d}(\mathcal{S}_{\boldsymbol{\psi}})_\#\nu(t)\; u_{\mathbb{S}^2}(d\boldsymbol{\psi}).
\end{aligned}$$

Using (9) and the rotation invariance of the spherical integral, we have

$$\int_{\mathbb{S}^2} \langle\boldsymbol{\psi}, \boldsymbol{\xi}\rangle^2 \,\mathrm{d}u_{\mathbb{S}^2}(\boldsymbol{\psi}) = \frac{1}{2}\int_{-1}^1 t^2 \,\mathrm{d}t = \frac{1}{3} \qquad \forall \boldsymbol{\psi} \in \mathbb{S}^2.$$

Hence, we obtain

$$\mathcal{E}_V(\nu) = \int_{\mathbb{S}^2}\int_{\mathbb{S}^2} \left[\langle\boldsymbol{\psi}, \boldsymbol{\xi}\rangle^2 + \left\langle\boldsymbol{\psi}, \boldsymbol{e}^3\right\rangle^2\right] \mathrm{d}u_{\mathbb{S}^2}(\boldsymbol{\psi}) \,\mathrm{d}\nu(\boldsymbol{\xi}) = \frac{2}{3}.$$

□

Let us now consider the case of the semicircular sliced Wasserstein distance. Such slicing operator is much harder to manipulate. Therefore, we did not manage to determine the semicircular sliced Wasserstein barycenters of $\mu_1$ and $\mu_2$, but we can show that the observation made in Proposition D.2 does not hold in the case of the SSW distance.

**Proposition D.3.** *The uniform probability measure on the sphere is not a semicircular sliced 2-Wasserstein barycenter. In particular, considering $\chi$, the uniform probability measure on the equator $C$, and $u_{\mathbb{S}^2}$ the uniform probability measure on $\mathbb{S}^2$, we have*

$$\tfrac{1}{2}\operatorname{SSW}_2^2(\mu_1,\chi)+\tfrac{1}{2}\operatorname{SSW}_2^2(\mu_2,\chi)<\tfrac{1}{2}\operatorname{SSW}_2^2(\mu_1,u_{\mathbb{S}^2})+\tfrac{1}{2}\operatorname{SSW}_2^2(\mu_2,u_{\mathbb{S}^2}).$$

*Proof.* Recall the circle $\mathbb{T}=\mathbb{R}/2\pi\mathbb{Z}$. For any $x\in\mathbb{R}$, we note $[x]=x+2\pi\mathbb{Z}$ the equivalence class of $x$ and for any $\gamma\in\mathbb{T}$ we note $\tilde{\gamma}$ its representative in $[0,2\pi[$. Let $\mu\in\mathcal{P}(\mathbb{S}^2)$. The semicircular sliced Wasserstein barycenters are the minimizers of the functional

$$\mathcal{E}_S(\mu)\coloneqq\frac{1}{2}\operatorname{SSW}_2^2(\mu,\delta_{\boldsymbol{e}^3})+\frac{1}{2}\operatorname{SSW}_2^2(\mu,\delta_{-\boldsymbol{e}^3}),$$

where

$$\operatorname{SSW}_2^2(\mu,\delta_{\boldsymbol{e}^3})=\int_{\mathbb{S}^2}\mathrm{W}_2^2(\mathcal{A}_{\boldsymbol{\psi}\#}\mu,\mathcal{A}_{\boldsymbol{\psi}\#}\delta_{\boldsymbol{e}^3})\,\mathrm{d}u_{\mathbb{S}^2}(\boldsymbol{\psi})$$

and the slicing operator $\mathcal{A}_{\boldsymbol{\psi}}$ is given in Remark 3.8. Let $\boldsymbol{\psi}=\Phi(\varphi,\theta)\in\mathbb{S}^2$. Since we integrate over $\mathbb{S}^2$, we can assume $\boldsymbol{\psi}\notin\{\pm\boldsymbol{e}^3\}$. Then we have $(\mathcal{A}_{\boldsymbol{\psi}})_\#\delta_{\boldsymbol{e}^3}=\delta_{[\pi]}$ and $(\mathcal{A}_{\boldsymbol{\psi}})_\#\delta_{-\boldsymbol{e}^3}=\delta_{[0]}$. Therefore,

$$\mathrm{W}_2^2(\mathcal{A}_{\boldsymbol{\psi}\#}\mu,\mathcal{A}_{\boldsymbol{\psi}\#}\delta_{\boldsymbol{e}^3})=\int_{\mathbb{T}}|\tilde{\gamma}-\pi|^2\,\mathrm{d}\mathcal{A}_{\boldsymbol{\psi}\#}\mu(t)=\int_{\mathbb{S}^2}\left|\tilde{\mathcal{A}}_{\boldsymbol{\psi}}(\boldsymbol{\xi})-\pi\right|^2\mathrm{d}\mu(\boldsymbol{\xi}).\tag{52}$$

Using spherical coordinates $\boldsymbol{\xi}=\Phi(\alpha,\beta)$, we see that

$$\mathcal{A}_{\boldsymbol{\psi}}(\boldsymbol{\xi})=\operatorname{azi}\big(\Psi(0,-\theta,-\varphi)\,\Phi(\alpha,\beta)\big)=\operatorname{azi}\big(\Psi(0,-\theta,-\varphi+\alpha)\,\Phi(0,\beta)\big)=\mathcal{A}_{\Phi(\varphi-\alpha,\theta)}(\Phi(0,\beta)).$$

Hence, with the substitution $\varphi\mapsto\varphi-\alpha$, we have

$$\begin{aligned}\operatorname{SSW}_2^2(\mu,\delta_{\boldsymbol{e}^3})&=\frac{1}{4\pi}\int_{\mathbb{S}^2}\int_0^\pi\int_0^{2\pi}\left|\tilde{\mathcal{A}}_{\Phi(\varphi-\alpha,\theta)}(\Phi(0,\beta))-\pi\right|^2\sin(\theta)\,\mathrm{d}\varphi\,\mathrm{d}\theta\,\mathrm{d}\mu(\Phi(\alpha,\beta))\\&=\frac{1}{4\pi}\int_{\mathbb{S}^2}\int_0^\pi\int_0^{2\pi}\left|\tilde{\mathcal{A}}_{\Phi(\varphi,\theta)}(\Phi(0,\beta))-\pi\right|^2\sin(\theta)\,\mathrm{d}\varphi\,\mathrm{d}\theta\,\mathrm{d}\mu(\Phi(\alpha,\beta))\\&=\frac{1}{4\pi}\int_{\mathbb{S}^2}\int_0^\pi\int_0^{2\pi}\left|\tilde{\mathcal{A}}_{\Phi(\varphi,\theta)}(\Phi(0,\beta))-\pi\right|^2\sin(\theta)\,\mathrm{d}\varphi\,\mathrm{d}\theta\,\mathrm{d}\operatorname{zen}_\#\mu(\beta).\end{aligned}$$

Introducing the functions

$$F_1\colon[0,\pi]\to\mathbb{R},\quad\beta\mapsto\frac{1}{4\pi}\int_0^{2\pi}\int_0^\pi\left|\tilde{\mathcal{A}}_{\Phi(-\varphi,\theta)}(\Phi(0,\beta))-\pi\right|^2\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi,$$

and

$$F\colon[0,\pi]\to\mathbb{R},\quad\beta\mapsto\tfrac{1}{2}F_1(\beta)+\tfrac{1}{2}F_1(\pi-\beta),$$

we obtain by using the symmetry that

$$\mathcal{E}_S(\mu) = \tfrac{1}{2}\,\mathrm{SSW}_p^p(\mu, \delta_{\boldsymbol{e}^3}) + \tfrac{1}{2}\,\mathrm{SSW}_p^p(\mu, \delta_{-\boldsymbol{e}^3}) = \int_0^\pi F(\beta)\,\mathrm{d}\,\mathrm{zen}_\#\,\mu(\beta).$$

Let $\beta \in [0, \pi]$. We study $f_\beta(\varphi, \theta) \coloneqq \tilde{\mathcal{A}}_{\Phi(-\varphi,\theta)}(\Phi(0, \beta))$. For $\theta \in (0, \pi)$ and $\varphi \in (0, 2\pi) \setminus \{\pi\}$, we have

$$\begin{aligned} f_\beta(\varphi, \theta) &= \mathrm{azi}\left(\mathrm{R}_{\boldsymbol{e}^2}(-\theta)\mathrm{R}_{\boldsymbol{e}^3}(\varphi)\Phi(0, \beta)\right) \\ &= \mathrm{azi}\left(\begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}\begin{bmatrix} \cos(\varphi) & -\sin(\varphi) & 0 \\ \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \sin(\beta) \\ 0 \\ \cos(\beta) \end{bmatrix}\right) \\ &= \mathrm{azi}\left(\begin{bmatrix} \cos(\theta)\cos(\varphi)\sin(\beta) - \sin(\theta)\cos(\beta) \\ \sin(\varphi)\sin(\beta) \\ \sin(\theta)\cos(\varphi)\sin(\beta) + \cos(\theta)\cos(\beta) \end{bmatrix}\right). \end{aligned}$$

We identify some symmetries. For $\varphi \in (0, 2\pi) \setminus \{\pi\}$, and $\theta \in (0, \pi)$, we have

$$f_\beta(2\pi - \varphi, \theta) = 2\pi - f_\beta(\varphi, \theta) \quad \text{or} \quad f_\beta(2\pi - \varphi, \theta) = f_\beta(\varphi, \theta).$$

In both cases, $(f_\beta(2\pi - \varphi, \theta) - \pi)^2 = (f_\beta(\varphi, \theta) - \pi)^2$. Moreover, for $\varphi, \theta \in (0, \pi)$,

$$f_\beta(\varphi, \pi - \theta) = f_\beta(\pi - \varphi, \theta) \quad \text{and} \quad f_{\pi-\beta}(\varphi, \theta) = \pi - f_\beta(\pi - \varphi, \theta). \tag{53}$$

Hence, we have

$$\begin{aligned} F_1(\beta) &= \frac{1}{\pi}\int_0^\pi \int_0^{\frac{\pi}{2}} (f_\beta(\varphi, \theta) - \pi)^2 \sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi \qquad \text{and} \\ F_1(\pi - \beta) &= \frac{1}{\pi}\int_0^\pi \int_0^{\frac{\pi}{2}} f_\beta(\varphi, \theta)^2 \sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi. \end{aligned}$$

Eventually, $F$ is given by

$$F(\beta) = \frac{1}{\pi}\int_0^\pi \int_0^{\frac{\pi}{2}} \left(f_\beta(\varphi, \theta) - \frac{\pi}{2}\right)^2 \sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi \; + \frac{\pi^2}{4},$$

where, for $\beta, \varphi, \theta \in (0, \pi)$, we have

$$f_\beta(\varphi, \theta) = \frac{\pi}{2} - \arctan\left(\cos(\theta)\cot(\varphi) - \frac{\sin(\theta)}{\sin(\varphi)}\cot(\beta)\right).$$

Let us now focus on the two particular cases of the theorem. Let $\chi$ be a measure supported by the equator. By the symmetry (53), we have

$$\begin{aligned} \mathcal{E}_S(\chi) = F\left(\frac{\pi}{2}\right) &= \frac{1}{\pi}\int_0^\pi \int_0^{\frac{\pi}{2}} \left(f_{\frac{\pi}{2}}(\varphi, \theta) - \frac{\pi}{2}\right)^2 \sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi \; + \frac{\pi^2}{4} \\ &= \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \left(\arctan\left(\frac{\tan(\varphi)}{\cos(\theta)}\right) - \frac{\pi}{2}\right)^2 \sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi \; + \frac{\pi^2}{4}. \end{aligned}$$

Since $\arctan\left(\frac{\tan(\varphi)}{\cos(\theta)}\right) > \varphi$ for any $\varphi, \theta \in (0, \frac{\pi}{2})$, and $t \mapsto \left(t - \frac{\pi}{2}\right)^2$ is strictly decreasing on $[0, \frac{\pi}{2}]$, we obtain

$$\mathcal{E}_S(\chi) < \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \left(\varphi - \frac{\pi}{2}\right)^2 \sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi \; + \frac{\pi^2}{4} = \frac{\pi^2}{12} + \frac{\pi^2}{4} = \frac{\pi^2}{3}.$$

For the uniform measure $u_{\mathbb{S}^2}$ on the sphere, we have $\mathcal{A}_{\boldsymbol{\psi}\#} u_{\mathbb{S}^2} = u_{\mathbb{T}}$ for any $\boldsymbol{\psi} \in \mathbb{S}^2$. From the first equality of (52), we obtain

$$\mathrm{SSW}_2^2(u_{\mathbb{S}^2}, \delta_{\boldsymbol{e}^3}) = \frac{1}{2\pi} \int_{\mathbb{S}^2} \int_0^{2\pi} |t - \pi|^2 \,\mathrm{d}t \,\mathrm{d}u_{\mathbb{S}^2}(\boldsymbol{\psi}) = \frac{\pi^2}{3}.$$

By symmetry, we have $\mathcal{E}_S(u) = \frac{\pi^2}{3}$, and therefore $\mathcal{E}_S(u) > \mathcal{E}_S(\chi)$. □

## References


[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions.* National Bureau of Standards, Washington, DC, USA, 1972.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, 2008. doi:10.1515/9781400830244.

[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*, 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] 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.

[9] 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.

[10] H. Berens, P. L. Butzer, and S. Pawelke. Limitierungsverfahren von Reihen mehrdimensionaler Kugelfunktionen und deren Saturationsverhalten. *Publ. Res. Inst. Math. Sci.*, 4:201–268, 1968. doi:10.2977/prims/1195194875.

[11] T. Birdal, M. Arbel, U. Şimşekli, and L. J. Guibas. Synchronizing probability measures on rotations via optimal transport. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1566–1576, 2020. doi:10.1109/CVPR42600.2020.00164.

[12] J. Boman and F. Lindskog. Support theorems for the Radon transform and Cramèr–Wold theorem. *J. Theor. Probab.*, 22:683–710, 2009. doi:10.1007/s10959-008-0151-0.

[13] C. Bonet. *Leveraging Optimal Transport via Projections on Subspaces for Machine Learning Applications.* PhD thesis, Université Bretagne Sud, 2023.

[14] 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.

[15] C. Bonet, L. Chapel, L. Drumetz, and N. Courty. Hyperbolic sliced-Wasserstein via geodesic and horospherical projections. In T. Doster, T. Emerson, H. Kvinge, N. Miolane, M. Papillon, B. Rieck, and S. Sanborn, editors, *Proceedings of 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)*, pages 334–370. PMLR, 2023. URL: https://proceedings.mlr.press/v221/bonet23a.html.

[16] 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.

[17] N. Bonnotte. *Unidimensional and Evolution Methods for Optimal Transportation*. PhD thesis, Université Paris Sud, 2013.

[18] S. Borgwardt. An LP-based, strongly-polynomial 2-approximation algorithm for sparse Wasserstein barycenters. *Oper. Res. Int. J.*, 22:1511–1551, 2022. doi:10.1007/s12351-020-00589-z.

[19] 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.

[20] 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.

[21] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL: https://papers.nips.cc/paper_files/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

[22] F. Dai and Y. Xu. *Approximation Theory and Harmonic Analysis on Spheres and Balls*. Springer Monographs in Mathematics. Springer, New York, 2013. doi:10.1007/978-1-4614-6660-4.

[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] M. Ehler, M. Gräf, S. Neumayer, and G. Steidl. Curve based approximation of measures on manifolds by discrepancy minimization. *Foundations of Computational Mathematics*, 21(6):1595–1642, 2021. doi:10.1007/s10208-021-09491-2.

[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*, pages 6185–6215. PMLR, 2022.

[27] 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.

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

[29] G. P. Gehér, A. Hrušková, T. Titkos, and D. Virosztek. Isometric rigidity of Wasserstein spaces over Euclidean spheres, 2023. arXiv:2308.05065.

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

[31] P. Hall, G. Watson, and J. Cabrera. Kernel density estimation with spherical data. *Biometrika*, 74(4):751–62, 1987. doi:10.1093/biomet/74.4.751.

[32] 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.

[33] R. Han. Sliced Wasserstein distance between probability measures on Hilbert spaces, 2023. arXiv:2307.05802.

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

[35] J. Hertrich, C. Wald, F. Altekrüger, and P. Hagemann. Generative sliced MMD flows with Riesz kernels. *ICLR*, 2024.

[36] R. Hielscher. The Radon Transform on the Rotation Group–Inversion and Application to Texture Analysis. Dissertation, Technische Universität Bergakademie Freiberg, 2007. URL: https://nbn-resolving.org/urn:nbn:de:bsz:105-3614018.

[37] R. Hielscher, D. Potts, J. Prestin, H. Schaeben, and M. Schmalz. The Radon transform on $SO(3)$: A Fourier slice theorem and numerical inversion. *Inverse Problems*, 24:025011, 2008. doi:10.1088/0266-5611/24/2/025011.

[38] 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*, pages 121–144. Birkhäuser, 2018. doi:10.1007/978-3-319-70824-9_7.

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

[40] 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.

[41] 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.

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

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] J. M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Grad. Texts in Math.* Springer, New York, 2012. doi:10.1007/978-1-4419-9982-5.

[48] 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.

[49] 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.

[50] 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.

[51] A. Morawiec. *Orientations and Rotations.* Springer, Berlin, 2004. doi:10.1007/978-3-662-09156-2.

[52] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction.* SIAM, Philadelphia, PA, 2000. doi:10.1137/1.9780898718324.fm.

[53] 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.

[54] 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.

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

[56] G. Plonka, D. Potts, G. Steidl, and M. Tasche. *Numerical Fourier Analysis.* Applied and Numerical Harmonic Analysis. Birkhäuser, Basel, 2nd edition, 2023. doi:10.1007/978-3-031-35005-4.

[57] 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.

[58] M. Quellmalz, R. Beinert, and G. Steidl. Sliced optimal transport on the sphere. *Inverse Problems*, 39(10):105005, 2023. doi:10.1088/1361-6420/acf156.

[59] 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.

[60] 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*, pages 42–54. Springer, 2023. doi:10.1007/978-3-031-31975-4_4.

[61] 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.

[62] 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*, pages 435–446. Springer, 2012. doi:10.1007/978-3-642-24785-9_37.

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

[64] 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.

[65] K. P. Rustamov. On approximation of functions on the sphere. *Izv. RAN. Ser. Mat.*, 57(5):127–148, 1993. doi:10.1070/IM1994v043n02ABEH001566.

[66] R. M. Rustamov and S. Majumdar. Intrinsic sliced Wasserstein distances for comparing collections of probability distributions on manifolds and graphs, 2023. URL: https://proceedings.mlr.press/v202/rustamov23a.html.

[67] 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.

[68] 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.

[69] 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.

[70] M. Theveneau and N. Keriven. Stability of entropic Wasserstein barycenters and application to random geometric graphs. In *29° Colloque sur le traitement du signal et des images*, pages 93–96. GRETSI - Groupe de Recherche en Traitement du Signal et des Images, 2023. URL: https://gretsi.fr/data/colloque/pdf/2023_keriven1083.pdf.

[71] P. Ungar. Freak theorem about functions on a sphere. *J. Lond. Math. Soc.*, 1(1):100–103, 1954. doi:10.1112/jlms/s1-29.1.100.

[72] K. G. van den Boogaart, R. Hielscher, J. Prestin, and H. Schaeben. Kernel-based methods for inversion of the Radon transform on SO(3) and their applications to texture analysis. *J. Comput. Appl. Math.*, 199:122–140, 2007. doi:10.1016/j.cam.2005.12.003.

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

[74] N. J. Vilenkin. *Special Functions and the Theory of Group Representations*. AMS, Providence, RI, 1968.

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

[76] W. Wang and M. A. Carreira-Perpinán. Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. *arXiv preprint arXiv:1309.1541*, 2013.

[77] F. W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Graduate Texts in Mathematics. Springer, New York, 1983. doi:10.1007/978-1-4757-1799-0.

[78] 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.

[79] 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.