Supplementary Materials for "Wasserstein Principal Component Analysis for Circular Measures"

Mario Beraha
Department of Mathematics, Politecnico di Milano
and
Matteo Pegoraro
Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark

February 19, 2024

Abstract

This Supplementary Material contains additional methodological details, mathematical proofs, and numerical results supporting the accompanying paper.

A Technical Preliminaries

A.1 Measure Theoretic Preliminaries

Let (M, g) be a Riemannian manifold of dimension n, with TM being its tangent bundle and TM^* its cotangent bundle. We know by definition that g is a section $g : M \to (TM \otimes TM)^*$ and the volume form $\omega : M \to \wedge^n (TM)^*$ is defined locally by $\omega = \mid \det(g) \mid^{1/2} dx_1 \wedge \ldots \wedge dx_n$.

Let $\mathcal{L}$ be the Lebesgue measure on $\mathbb{R}^n$, we consider the σ-algebra generate by all sets A such that $\varphi(A \cup U)$ is in the Lebesgue σ-algebra of $\mathbb{R}^n$ for some chart (U, φ). Then we indicate with $\mathcal{L}_M$ the Riemann-Lebesgue volume measure, i.e. the measure on M such that for every chart (U, φ) and $A \subset U$ contained in the σ-algebra just define:

$$\mathcal{L}_M(A) = \int_{\varphi(A)} \mid \det(g(\varphi^{-1})) \mid^{1/2} d\mathcal{L} \tag{1}$$

Note that, in general, $\varphi \# \mathcal{L}_M \neq \mathcal{L}$.

Consider $h : M \to \mathbb{R}$ such that $\mathrm{supp}(h) \subset U$, with (U, φ) being a chart, we can integrate h as follows:

$$\int_M h d\mu = \int_U h d\mu = \int_U h f_\mu d\mathcal{L}_M = \int_{z(U)} \mid \det(g(\varphi^{-1})) \mid^{1/2} h(\varphi^{-1}) f_\mu(z^{-1}) d\mathcal{L}. \tag{2}$$

The general case is defined in a natural way through a partition of unity.

Now we can consider a measure μ on M, with density function f_μ wrt $\mathcal{L}_M$, that is:

$$\mu(A) = \int_A f_\mu d(\mathcal{L}_m) = \int_{\varphi(A)} \mid \det(g(\varphi^{-1})) \mid^{1/2} f_\mu(\varphi^{-1}) d\mathcal{L}. \tag{3}$$

Lastly, if μ doesn't have a density function wrt $\mathcal{L}_M$, to integrate some function against μ we pick a weak converging sequence $\mu_n \rightharpoonup \mu$ such that, for every n, μ_n has a density function and extend the definition taking the limit of the integrals.

A.2 McCann's Result

Let us recall the definition of c-concavity. Let $c : M \times M \to \mathbb{R} \cup +\infty$. For a function $\psi : M \to \mathbb{R} \cup \{-\infty\}$ define its c-transform $\psi^{c+} : M \to \mathbb{R} \cup \{-\infty\}$ as

$$\psi^{c+}(x) = \inf_{y \in M} c(x, y) - \psi(y).$$

Note that this generalises the Legendre transform, which is recovered when $M = \mathbb{R}^d$ and $c(x, y) = \langle x, y\rangle$.

Definition A.1. *A function $\phi : M \to \mathbb{R} \cup \{-\infty\}$ is c-concave if its not identically $-\infty$ and there exists $\psi : M \to \mathbb{R} \cup \{-\infty\}$ such that*

$$\phi = \psi^{c+}$$

Given $\mu \in \mathcal{W}_2(M)$ and $U \subset M$ open we define $S(U) = \{v : U \to TM \mid \pi \circ v = \mathrm{Id}_U\}$ be the sheaf of local sections of the tangent bundle of M, that is the vector space of tangent vector fields on U. Whenever U is a local trivialisation of the tangent bundle, we may use the notation $v_z := v(z) \in T_zM$ for $v \in S(U)$. Now we can define the following sheaf of functions:

$$L^2_\mu(U) = \{v \in S(U) \mid \int \|v_z\|^2 d\mu(z) < \infty\}, \tag{4}$$

where $\|v_z\|^2$ stands for $g(v_z, v_z)$.

For any $v \in L^2_\mu(U)$ we can consider the map $\exp(v)$ defined as $\exp(v)(z) := \exp_z(v_z)$, $z \in U$. McCann (2001) proved that if μ is absolutely continuous with respect to the volume measure on M, the unique optimal plan γ^o between μ and ν is induced by a map, i.e. we have $T : M \to M$, inducing $(\mathrm{Id}, T) : M \to M \times M$, such that $\gamma^o = (\mathrm{Id}, T)\#\nu$. Moreover, the map T has the form $T = \exp(-\nabla\phi)$ where ϕ is a d^2-concave function (Gigli, 2011).

A.3 More details on $\mathbb{S}_1$

We call $\exp_c : (\mathbb{R}, +) \to (\mathbb{S}_1, \cdot)$ the map defined as $\exp_c(x) = e^{ix}$ and view $\mathbb{S}_1 \subset \mathbb{C}$ with the multiplication operation $e^{ix} \cdot e^{iy} = e^{i(x+y)}$. Thus $\exp_c$ is a group morphism: $\exp_c(x + y) = \exp_c(x) \cdot \exp_c(y)$. Similarly $log_c : \mathbb{S}_1 \to \mathbb{R}$ given by $\log_c(z) = x \in [0, 2\pi)$ such that $z = e^{ix}$. Clearly log_c is right inverse of $\exp_c$ i.e. $\exp_c \circ \log_c = \mathrm{Id}_{\mathbb{S}_1}$.

In a similar fashion, we use the projection on the quotient $\pi_c : \mathbb{R} \to \mathbb{R}/2\pi\mathbb{Z}$ which is a map of groups. This is an alternative, though equivalent representation of $\mathbb{S}_1$ in the following sense: we have that $\exp_c(x) = \exp_c(\pi_c(x))$ and so $\log_c = (\exp_c \circ \pi_c)^{-1}$ are group isomorphisms. We employ the metric induced by this alternative representation of the circle in the proofs - see Equation (7).

Lastly for any $z \in \mathbb{S}_1$ we define shifted versions of the maps $\exp_c$ and $\log_c$, centred in z: $\exp_z(x) = \exp_c(x + \log_c(z))$ and $\log_z(z') = x'$ with $x' \in [-\pi/2, \pi/2)$ such that $\exp_z(x') = z'$. If we call $W := (-\pi/2, \pi/2)$ and $V_z := \mathbb{S}_1 - \{-z\}$ then for every $z \in \mathbb{S}_1$, the couple $(V_z, \log_z)$ is a local chart which gives a homeomorphism on W. With this differential structure $\mathbb{S}_1$ is a Lie Group and its tangent bundle is $T\mathbb{S}_1 = \{(x, v) \mid x \in \mathbb{S}_1 \text{ and } v \in T_x\mathbb{S}_1\} \simeq \mathbb{S}_1 \times \mathbb{R}$. We call 1 the point $(1, 0)$ which gives the neutral element in $\mathbb{S}_1$.

We can make the following observations: the Riemannian metric $g : \mathbb{S}_1 \to (T\mathbb{S}_1 \otimes T\mathbb{S}_1)^*$ is induced by the embedding $\mathbb{S}_1 \hookrightarrow \mathbb{C} \simeq \mathbb{R}^2$. In local coordinates (V_1, log_1) centred in $1 \in \mathbb{S}_1$ the embedding is $x \in (-\pi/2, \pi/2) \mapsto e^{ix} = (\cos(x), \sin(x))$. The map between tangent spaces is therefore $d/dx \mapsto -\sin(x)d/dx_1 + \cos(x)d/dx_2$. Let $g_{\mathbb{E}}$ be the euclidean metric in $\mathbb{R}^2$. In canonical coordinates of $\mathbb{R}^2$, x_1, x_2, this metric is clearly given by the quadratic form $\mathrm{Id}_{2\times 2}$. Thus $g(d/dx) = g_E(-\sin(x), \cos(x)) = \sin^2(x) + \cos^2(x) = 1$. In other words in local coordinates $(V_1, \log_1)$, $g(v, w) = v \cdot w$ for $v, w \in T_1\mathbb{S}_1 \simeq \mathbb{R}$. With this metric, $\exp_z : T_z\mathbb{S}_1 \to \mathbb{S}_1$ and $\log_z : \mathbb{S}_1 \to T_z\mathbb{S}_1$ are respectively the Riemannian exponential and logarithm.

A.4 Vector Fields on $\mathbb{S}_1$

Consider now $\mu \in \mathcal{P}(\mathbb{S}_1)$ and a real valued function $f : \mathbb{S}_1 \to \mathbb{R}$ which we would like to integrate against μ. We saw that $det(g) \equiv 1$ and the volume form ω locally is $\omega = dx$. This immediately implies that $\mathcal{L}_{\mathbb{S}_1} = \exp_c \#\mathcal{L}$ or, equivalenty, $\log_c \#\mathcal{L}_{\mathbb{S}_1} = \mathcal{L}$. Thus:

$$\int_{\mathbb{S}_1} f(z) d\mathcal{L}_{\mathbb{S}_1}(z) = \int_{[-\pi/2, \pi/2)} f(\exp_c(x)) d\mathcal{L}(x) \tag{5}$$

Along this line, the sheaf of tangent vector fields L^2_μ can be easily transported on $[0, 2\pi)$ with the change of variables:

$$\int_{\mathbb{S}_1} \mid v_z \mid^2 d\mu(z) = \int_{[0,2\pi)} \mid v(\exp_c(x)) \mid^2 d \log_c \#\mu(x) \tag{6}$$

In fact from Equation (3) plus the observation that $\mathcal{L}_{\mathbb{S}_1} = \exp_c \#\mathcal{L}$ we have that a measure μ on $\mathbb{S}_1$ is equivalently represented by the measure $\log_c \#\mu$ extended to a periodic measure $\widetilde{\mu}$ on $\mathbb{R}$ as follows: for any measurable A and $p \in \mathbb{Z}$, $\mu(\exp_c(A)) =: \widetilde{\mu}(A) = \widetilde{\mu}(A+p)$

B Proofs

Let us introduce some notation, motivated by Appendix A.3.

$$d_{\mathbb{Z}}(x, y)^2 := \inf_{p \in 2\pi\mathbb{Z}} (x - y - p)^2 \leq (x - y)^2 \tag{7}$$

Note that $d_R(z, z') = d_{\mathbb{Z}}(\log_c(z), \log_c(z'))$

B.1 Proof of Proposition 2

Proof. The proof follows from the notion of locally optimal plans in Delon et al. (2010). Let γ_θ be the transport plan that takes an element of mass from position $F_{\widetilde{\nu}}^{-}(u)$ to position $(F_{\widetilde{\mu}}^{\theta})^{-}(u)$. Then γ_θ is locally optimal and the associated cost is

$$C_{[\mu,\nu]}(\theta) = \int_0^1 \left(F_{\widetilde{\mu}}^{-}(u) - (F_{\widetilde{\nu}}^{\theta})^{-}(u)\right)^2 \mathrm{d}u$$

The (global) optimal plan is associated to $\theta^* = \operatorname{argmin} C(\theta) = W_2^2(\mu,\nu)$. To recover the optimal transport map we operate the change of variables $x = F_{\widetilde{\mu}}^{-}(u)$, which yields:

$$W_2^2(\mu,\nu) = \int_0^{2\pi} \left(T_{\widetilde{\mu}}^{\widetilde{\nu}}(x) - x\right)^2 \mathrm{d}\widetilde{\mu}(x) \geq \tag{8}$$

$$\int_0^{2\pi} d_{\mathbb{Z}}^2(T_{\widetilde{\mu}}^{\widetilde{\nu}}(x), x)\mathrm{d}\widetilde{\mu}(x) = \tag{9}$$

$$\int_{\mathbb{S}_1} d_R^2(\exp_c(T_{\widetilde{\mu}}^{\widetilde{\nu}}(\log_c(z))), z)^2 \mathrm{d}\mu(z) \geq W_2^2(\mu,\nu) \tag{10}$$

where the first equality follows by defining $T_{\widetilde{\mu}}^{\widetilde{\nu}} := (F_{\widetilde{\nu}}^{\theta^*})^{-} \circ F_{\widetilde{\mu}}$, while the last equality is obtained with $z = \exp_c(x)$ and the properties of $d_{\mathbb{Z}}$. $\square$

B.2 Proof of Theorem 3

Proof. First we observe that:

$$(F_{\widetilde{\nu}}^{\theta^*})^{-}\left(F_{\widetilde{\mu}}(u+2\pi p)\right) = (F_{\widetilde{\nu}}^{\theta^*})^{-}\left(F_{\widetilde{\mu}}(u)+p\right) = (F_{\widetilde{\nu}})^{-}\left(F_{\widetilde{\mu}}(u)+p-\theta^*\right) = (F_{\widetilde{\nu}}^{\theta^*})^{-}\left(F_{\widetilde{\mu}}(u)\right)+2\pi p$$

which means that $\widetilde{T}(u+2\pi p) = \widetilde{T}(u) + 2\pi p$ for every $p \in \mathbb{Z}$.

By Theorem 2 we know that $\widetilde{T}$ is an optimal transport map if and only if $\theta = \theta^*$ as in Equation (8). Define:

$$C_{[\mu,\nu]}(\theta) = \int_0^1 \left(F_{\widetilde{\mu}}^{-}(u) - (F_{\widetilde{\nu}}^{\theta})^{-}(u)\right)^2 \mathrm{d}u$$

Delon et al. (2010) prove that the map $\theta \mapsto C_{[\mu,\nu]}(\theta)$ is strictly convex if μ is a.c.. Thus θ^* is the unique stationary point of the function. For this reason we compute the derivative of $C_{[\mu,\nu]}$ in θ, knowing that θ^* is the only value such that $(C_{[\mu,\nu]})' = 0$. Thanks to Leibniz rule we can write:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\theta} C_{[\mu,\nu]}(\theta) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\theta} \left(F_{\widetilde{\mu}}^{-}(u) - (F_{\widetilde{\nu}})^{-}(u-\theta)\right)^2 \mathrm{d}u \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\theta} \left(F_{\widetilde{\mu}}^{-}(u+\theta) - (F_{\widetilde{\nu}})^{-}(u)\right)^2 \mathrm{d}u \\
&= \int_0^1 2\left(F_{\widetilde{\mu}}^{-}(u+\theta) - (F_{\widetilde{\nu}})^{-}(u)\right) \frac{1}{f_{\widetilde{\mu}}(F_{\widetilde{\mu}}^{-}(u+\theta))} \mathrm{d}u
\end{aligned}$$

with the change of variables $v = F_{\widetilde{\mu}}^{-}(u+\theta)$, which entails $F_{\widetilde{\mu}}(v) - \theta = u$ and $\mathrm{d}u = \mathrm{d}\widetilde{\mu}(v) = f_{\widetilde{\mu}}(v)\mathrm{d}v$ we obtain

$$= \int_{v_0}^{2\pi+v_0} 2\left(v - \widetilde{T}(v)\right)\frac{f_{\widetilde{\mu}}(v)}{f_{\widetilde{\mu}}(v)}\mathrm{d}v = -2\int_{v_0}^{2\pi}\left(\widetilde{T}(v) - v\right)\mathrm{d}v - 2\int_{2\pi}^{2\pi+v_0}\left(\widetilde{T}(v) - v\right)\mathrm{d}v$$

with $v_0 = F_{\widetilde{\mu}}^{-}(\theta)$.

Via the change of variables $u = v - 2\pi$ we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}C_{[\mu,\nu]}(\theta) = -2\int_0^{2\pi}\left(\widetilde{T}(u) - 2\pi - (u - 2\pi)\right)\mathrm{d}v = -2\int_0^{2\pi}\left(\widetilde{T}(u) - u\right)\mathrm{d}u.$$

Optimality follows if and only if such quantity is equal to zero and thus:

$$0 = \int_0^{2\pi}\left(\widetilde{T}(u) - u\right)\mathrm{d}u.$$

□

B.3 Proof of Theorem 4

Proof. Observe that T is an optimal transport map, then defining

$$\widetilde{T}(x) := \log_c \circ T \circ \exp_c,\ x \in (0, 2\pi), \qquad \widetilde{T}(x+p) = \widetilde{T}(x) + p,\ p \in 2\pi\mathbb{Z}$$

clearly satisfies the monotonicity and "periodicity" requirements. Moreover, (11) is satisfied by Theorem 3. To prove that $|\widetilde{T}(x) - x| < \pi/2$ note that this is equivalent to $\inf_{p\in 2\pi\mathbb{Z}} \mid \widetilde{T}(x) - x - p \mid = \mid \widetilde{T}(x) - x \mid$. Since $T : \mathbb{S}_1 \to \mathbb{S}_1$ is an optimal transport map from μ to ν we have:

$$\begin{aligned} W_2(\mu,\nu) &= \int_{\mathbb{S}_1} d_R(T(z), z)^2 d\mu \\ &= \int_{[0,2\pi]} d_R(T(\exp_c(x)), \exp_c(x))^2 d(\log_c \#\mu)(x) \\ &= \int_{[0,2\pi]} d_{\mathbb{Z}}(\widetilde{T}(x), x)^2 d\widetilde{\mu}(x) \end{aligned}$$

where the first equality is obtained via the definition of optimal transport map, the second through the change of variables $z = \exp_{c|[0,2\pi]}(x)$, and in the last one we use the definition of $\widetilde{T}$, $\widetilde{\mu}$ and the properties of $d_{\mathbb{Z}}$. As already noted, we have $\inf_{p\in\mathbb{Z}} \mid \widetilde{T}(x) - x - p \mid \leq \mid \widetilde{T}(x) - x \mid$. If the strict inequality holds for some $A \subset [0, 2\pi]$ with $\widetilde{\mu}(A) > 0$ then also the integrals on $[0, 2\pi]$ must be different, and the thesis follows.

To prove the reverse statement, it suffices to prove that $\widetilde{T}(v)$ can be written as $F_{\widetilde{\nu}}^{-}(F_{\widetilde{\mu}}(v) + \theta)$, which is equivalent to saying that

$$F_{\widetilde{\nu}}^{-}(v) = \widetilde{T}\left((F_{\widetilde{\mu}}(\cdot) + \theta)^{-1}(v)\right) = \widetilde{T}(F_{\widetilde{\mu}}^{-}(v - \theta)).$$

Define $G_{\widetilde{\nu}} := \widetilde{T} \circ F_{\widetilde{\mu}}^{-}$, then of course $\widetilde{T} = G_{\widetilde{\nu}} \circ F_{\widetilde{\mu}}$. We show that $G_{\widetilde{\nu}}(u) \equiv F_{\widetilde{\nu}}^{-}(u + \theta)$. We have that, for $x \in [0, 2\pi)$

$$F_{\widetilde{\nu}}(x) = \widetilde{\nu}([0,x]) = \widetilde{\mu}(\widetilde{T}^{-1}([0,x])) = \widetilde{\mu}([\widetilde{T}^{-1}(0), \widetilde{T}^{-1}(x)]) = F_{\widetilde{\mu}}(\widetilde{T}^{-1}(x)) - F_{\widetilde{\mu}}(\widetilde{T}^{-1}(0))$$

and observe that $F_{\widetilde{\nu}}(x) \leq 1$ thanks to $|\widetilde{T}(x) - x| < \pi/2$. Hence, the pushforward of $\widetilde{\mu}$ on $\mathbb{S}_1$ gives a valid probability measure. Taking the inverse of $F_{\widetilde{\nu}}$

$$F_{\widetilde{\nu}}^{-}(u) = \left(F_{\widetilde{\mu}}(\widetilde{T}^{-1}(\cdot)) - F_{\widetilde{\mu}}(\widetilde{T}^{-1}(0))\right)^{-}(u) = \widetilde{T} \circ F_{\widetilde{\mu}}^{-}(u + F_{\widetilde{\mu}}(\widetilde{T}^{-1}(0)))$$

and setting $-\theta = F_{\widetilde{\mu}}(\widetilde{T}^{-1}(0))$ yields the result. □

B.4 Proof of Theorem 6

To prove item (ii), we will need the two following preliminary lemmas.

Lemma B.1. *Suppose we have $W_2(\nu, \nu_n) \to 0$ in $\mathcal{W}_2(\mathbb{S}_1)$ with ν, ν_n being a.c. wrt $\mathcal{L}_{\mathbb{S}_1}$ (for every n). Then $W_2(\log_c \#\nu, \log_c \#\nu_n) \to 0$ in $\mathcal{W}_2(\mathbb{R})$.*

Proof. From Theorem 7.12 in Villani (2003), convergence in the Wasserstein metric is equivalent to weak convergence plus the tightness condition: there exist x_0 such that

$$\lim_{R\to+\infty} \limsup_{k\to+\infty} \int_{d(x,x_0)>R} d(x, x_0)^p d \log_c \#\nu_k(x)$$

Observe that each measure $\log_c \#\nu_k$ is supported on $[0, 2\pi]$ so that the condition is always met. Hence, we just need to show that the sequence $\log_c \#\nu_k$ converges weakly. For measures on the real line, weak convergence is equivalent of pointwise convergence of the associated distribution functions at continuity points. That is, letting $F_k(x) := \log_c \#\nu_k([0, x))$ and $F(x) := \log_c \#\nu([0, x))$, it must hold that

$$F_k(x) \to F(x), \qquad \text{all } x \text{ such that } F(x) \text{ is continuous} \tag{11}$$

Observe that $F_k(x) = \nu_k(\exp_c([0, x)))$ by definition. By Portmanteau's theorem, for any x such that $\nu(\{\exp_c(x)\}) = 0$ we have that $\nu_k(\exp_c([0, x))) \to \nu_k(\exp_c([0, x)))$ which easily implies (11)

□

Lemma B.2. *Suppose we have $W_2(\nu, \nu_n) \to 0$ with μ, ν, ν_n being a.c. wrt $\mathcal{L}_{\mathbb{S}_1}$ (for every n). Then $\| \log_\mu(\nu_n) - \log_\mu(\nu) \|_{L^2_\mu} \to 0$.*

Proof. By Theorem B.1 we have $F_{\widetilde{\nu}_n}(x) \to F_{\widetilde{\nu}}(x)$ and the same for the quantile functions. As a consequence $C_{[\nu_n,\mu]}(\theta) \to C_{[\nu,\mu]}(\theta)$.

Thus consider $\theta_n = \arg\min C_{[\nu_n,\mu]}$. By the discussion in Section 3.2 and in particular Equation (10), we have that the minimisation domain of $C_{[\nu_n,\mu]}$ can be restricted to a sufficiently large compact interval K. Since $\{\theta_n\} \subset K$ compact, we can consider a converging subsequence which we still call $\{\theta_n\}$ with an abuse of notation. Let $\theta_n \to \theta^*$. Recall that $C_{[\nu_n,\mu]}$ (Delon et al., 2010) is strictly convex. Thus, by standard arguments, we conclude that $\theta^* = \arg\min C_{[\nu,\mu]}$.

Now consider:

$$\left| \left(F_{\widetilde{\mu}}(F_{\widetilde{\nu}_n}^{-}(u + \theta_n))\right) - \left(F_{\widetilde{\mu}}(F_{\widetilde{\nu}}^{-}(u + \theta^*))\right) \right| \leq \tag{12}$$

$$\left| \left(F_{\widetilde{\mu}}(F_{\widetilde{\nu}_n}^{-}(u + \theta_n))\right) - \left(F_{\widetilde{\mu}}(F_{\widetilde{\nu}}^{-}(u + \theta_n))\right) \right| + \left| \left(F_{\widetilde{\mu}}(F_{\widetilde{\nu}}^{-}(u + \theta_n))\right) - \left(F_{\widetilde{\mu}}(F_{\widetilde{\nu}}^{-}(u + \theta^*))\right) \right| \tag{13}$$

Which implies the pointwise convergence $T^{\widetilde{\mu}}_{\widetilde{\nu}_n}(u) \to T^{\widetilde{\mu}}_{\widetilde{\nu}}(u)$: both addends in the last sum go to 0. Since these maps are continuous and bounded on $[0, 2\pi]$ we have uniform convergence and strong convergence. The strong convergence in the image of $\log_\mu$ then follows. □

We are now ready to prove Theorem 6.

Proof.

1. To check the continuity of $\exp_\mu$, consider $T^{\nu_1}_\mu \times T^{\nu_2}_\mu : \mathbb{S}_1 \to \mathbb{S}_1 \times \mathbb{S}_1$ and induce the transport plan $\gamma = (T^{\nu_1}_\mu, T^{\nu_2}_\mu)\#\mu$. Then we have:

$$\begin{aligned} W_2^2(\nu_1, \nu_2) &\leq \int_{\mathbb{S}_1 \times \mathbb{S}_1} d_R(z, w)^2 d\gamma(dzdw) \\ &= \int_{\mathbb{S}_1} d_R(T^{\nu_1}_\mu(z), T^{\nu_2}_\mu(z))^2 d\mu(dz) \\ &\leq \int_{[0,2\pi]} d_{\mathbb{Z}}(T^{\widetilde{\nu}_1}_{\widetilde{\mu}}(x), T^{\widetilde{\nu}_2}_{\widetilde{\mu}}(x))^2 d\widetilde{\mu}(dx) \\ &\leq \| T^{\widetilde{\nu}_1}_{\widetilde{\mu}} - T^{\widetilde{\nu}_2}_{\widetilde{\mu}} \|^2_{L^2_{\widetilde{\mu}}([0,2\pi])} = \| \log_\mu(\nu_1) - \log_\mu(\nu_2) \|^2_{L^2_\mu}, \end{aligned}$$

   where the last identity is obtained thanks to $\log_c \#\mu = \widetilde{\mu}$ on $[0, 2\pi]$.

2. To check the continuity of $\log_\mu$ instead, by an approximation argument we obtain sequential continuity of $\log_\mu$ at any measure $\nu \in \mathcal{W}_2(\mathbb{S}_1)$: consider $\nu_n \to \nu$, with ν_n a.c. measures. Then $\{\log_\mu(\nu_n)\}$ is a Cauchy sequence in L^2_μ, which is a complete metric space, and so it converges to a vector field v. Consider $\exp_\mu(v)$. By the continuity of $\exp_\mu$ we have $\nu_n \to \exp_\mu(v)$ which then entails $\exp_\mu(v) = \nu$.

   Lastly, sequential continuity in metric spaces implies continuity.

□

B.5 Proof of Proposition 7

Theorem 3.2 in Ambrosio et al. (2019) ensures that, in the hypotheses of the proposition, $\int_{\mathbb{S}_1} d_R(T^{\nu_n}_\mu(x), T^{\nu}_\mu(x))^2 d\mu(x) \to 0$.

Now we prove the following lemma which ends the proof.

Lemma B.3. *Suppose we have μ, ν, ν_n being a.c. wrt $\mathcal{L}_{\mathbb{S}_1}$ (for every n). And suppose that the following hold.*

$$\int_{\mathbb{S}_1} d_R(T^{\nu_n}_\mu(x), T^{\nu}_\mu(x))^2 d\mu(x) \to 0.$$

Then $\| \log_\mu(\nu_n) - \log_\mu(\nu) \|_{L^2_\mu} \to 0$.

Proof. To simplify the notation, call $G := \log_\mu(\nu)$ and $G_n := G := \log_\mu(\nu_n)$. So that $G(x) = \log_x(T^{\nu}_\mu(x))$ and $G_n(x) = \log_x(T^{\nu_n}_\mu(x))$. On top of that define $f_n : \mathbb{S}_1 \to \mathbb{R}$ as $f_n(x) = G(x) - G_n(x)$. Note that f_n is bounded.

We can write:

$$\int_{\mathbb{S}_1} d_R(T_\mu^{\nu_n}(x), T_\mu^\nu(x))^2 d\mu(x) = \int_{A_n^+} \mid f_n(x) \mid^2 d\mu(x) + \int_{A_n^-} \mid f_n(x) + 2\pi \mid^2 d\mu(x).$$

Where $A_n^+ = f_n^{-1}([-\pi, \pi])$ and $A_n^- = f_n^{-1}([\pi, 2\pi]) \cup f_n^{-1}([-2\pi, -\pi])$.
We want to prove that:

$$\int_{\mathbb{S}_1} \mid f_n(x) \mid^2 d\mu(x) = \int_{A_n^+} \mid f_n(x) \mid^2 d\mu(x) + \int_{A_n^-} \mid f_n(x) \mid^2 d\mu(x) \to 0,$$

and thus we need to work on the integral $\int_{A_n^-} \mid f_n(x) \mid^2 d\mu(x)$ as we already know $\int_{A_n^+} \mid f_n(x) \mid^2 d\mu(x) \to 0$.

We want to show that $\mu(A_n^-) \to 0$.

Reasoning by contradiction, suppose that there exist $\varepsilon > 0$ such that for every $N > 0$ there is $n > N$ satisfying $\mu(A_n^-) > \varepsilon$. For an ease of notation, instead of taking a subsequence $\{n_k\}$ such that $\mu(A_{n_k}^-) > \varepsilon$, we just suppose that it holds for every n.

Consider $c > 0$, $c \in \mathbb{R}$. The set $B_{n,c}^+ := f_n^{-1}([c, +\infty)) \cap A_n^-$ must satisfy $\mu(B_{n,c}^+) \to 0$. In fact:

$$\int_{A_n^-} \mid f_n(x) + 2\pi \mid^2 d\mu(x) = \int_{B_{n,c}^+} \mid f_n(x) + 2\pi \mid^2 d\mu(x) + \int_{B_{n,c}^-} \mid f_n(x) + 2\pi \mid^2 d\mu(x),$$

and $\int_{B_{n,c}^+} \mid f_n(x) + 2\pi \mid^2 d\mu(x) \geq c \cdot \mu(B_{n,c}^+)$.

If $x \in B_{n,c}^+$ then either 1) $G(x) \in [-\pi, -\pi + c]$ and $G_n(x) \in [\pi - c, \pi]$ or 2) $G_n(x) \in [-\pi, -\pi + c]$ and $G(x) \in [\pi - c, \pi]$. At least one between 1) or 2) must hold an infinite number of times. WLOG 1) holds a countable number of times. Again, instead of taking a subsequence $\{n_k\}$ such that 1) always holds, we just suppose that it holds for every n.

In other words, for any $c > 0$, $c \in \mathbb{R}$ we have that: for any n, for every $x \in B_{n,c}^+$, $G(x) \in [-\pi, -\pi + c]$. And $B_{n,c}^+$ is always a set of positive measure. By the compactness of $\mathbb{S}_1$ we can find a set B with positive measure such that $G(x) = -\pi$ for all $x \in B$. Thus T_μ^ν is an OTM that sends each point of a set of full measure, into its antipodal point on the circle. This is absurd as it violates the cyclical monotonicity condition. Thus $\mu(A_n^-) \to 0$.

The following concludes the proof.

$$\begin{aligned}\int_{\mathbb{S}_1} \mid f_n(x) \mid^2 d\mu(x) &= \int_{A_n^+} \mid f_n(x) \mid^2 d\mu(x) + \int_{A_n^-} \mid f_n(x) \mid^2 d\mu(x) \\ &\leq \int_{A_n^+} \mid f_n(x) \mid^2 d\mu(x) + 4\pi^2 \mu(A_n^-) \to 0.\end{aligned}$$

□

B.6 Proof of Proposition 8

Our proof relies on the multi-marginal formulation of the Wasserstein barycentre in Agueh and Carlier (2011); Panaretos and Zemel (2020). In the following, consider $\mu_1, \ldots, \mu_n$ in $\mathcal{W}(M)$ where (M, d) is connected and compact manifold. Let $\Pi(\mu_1, \ldots, \mu_n)$ be the set of probability measures on $M^n := M \times \cdots \times M$ having marginals $\mu_1, \ldots, \mu_n$. The multi-marginal problem is to minimise

$$G(\pi) = \frac{1}{2n^2} \int_{M^n} \sum_{i<j} d(x_i, x_j)^2 \mathrm{d}\pi(x_1, \ldots, x), \quad \pi \in \Pi(\mu_1, \ldots, \mu_n).$$

As shown in Kim and Pass (2017) (Theorem 2.4), minimising $G(\pi)$ is equivalent to minimising $F(\nu)$ in (16). Indeed, let $\bar{x} : M^n \to M$ such that

$$x_1, \ldots, x_n \mapsto \underset{z \in M}{\operatorname{argmin}} \sum_{i=1}^{n} d^2(x_i, z)$$

then the optimal multicoupling $\pi^o = \operatorname{argmin}_{\pi \in \Pi} G(\pi)$ gives the minimiser of the Frechét functional via the rule

$$\bar{\mu} := \bar{x} \# \pi^o. \tag{14}$$

Lemma B.4. *Let (M, d) be a connected compact Riemannian manifold whose exponential map $\exp_M$ is non expansive. Denote by $\log_M$ the associated logarithmic map. Let μ^* be an absolutely continuous measure in $\mathcal{W}_2(M)$ and $\mu_1, \ldots, \mu_n \in \mathcal{W}_2(M)$. Assume that, for any $i, j = 1, \ldots, n$,*

$$\| \log_{\mu^*}(\mu_i) - \log_{\mu^*}(\mu_j) \|^2_{L^2_{\mu^*}} = W_2^2(\mu_i, \mu_j),$$

where $T^{\mu_i}_{\mu^*} = \exp_M \circ \log_{\mu^*}(\mu_i)$. Letting $\bar{T} = n^{-1} \sum_{i=1}^{n} \log_{\mu^*}(\mu_i)$, then the Wasserestein barycentre of $\mu_1, \ldots, \mu_n$ is $\bar{T} \# \mu^*$.

Proof. Consider the multicoupling $\pi^* = (T^{\mu_1}_{\mu^*}, \ldots, T^{\mu_n}_{\mu^*}) \# \mu^*$. Of course, $\pi^* \in \Pi(\mu_1, \ldots, \mu_n)$ by construction, and $G(\pi^*) \geq G(\pi^o)$. We observe that, by the non-expansiveness of the exponential map,

$$\begin{aligned} \int_{M^n} d^2(x_i, x_j) \mathrm{d}\pi^*(x_1, \ldots, x_n) &= \int_M d^2(T^{\mu_i}_{\mu^*}(x), T^{\mu_j}_{\mu^*}(x)) \mathrm{d}\mu^*(x) \\ &\leq \| \log_{\mu^*}(\mu_i) - \log_{\mu^*}(\mu_j) \|^2_{L^2_{\mu^*}} \\ &= W_2^2(\mu_i, \mu_j), \end{aligned}$$

where the first inequality follows from the first point of Theorem 6 and the last equality by hypothesis. Hence, π^* minimises $G(\pi)$ and the result follows from (14). □

The proof of Theorem 8 follows from Theorem B.4 and the fact that the exponential map of $\mathbb{S}_1$ is indeed non-expansive, see (7).

C Additional Simulation

We report here an additional simulation study for the PCA, with the goal of showing the effect of the base point $\bar{\mu}$. We consider truncated Gaussian measures on $[0, 2\pi)$ (and extended periodically over $\mathbb{R}$), parametrised by the mean parameter m and scale parameter (square root of the variance) s. We simulate $n = 100$ datapoints by sampling $m \sim \mathcal{U}(0.3\pi, 1.7\pi)$ and $s \mid m \sim \mathcal{U}(\pi/20, \pi/10) I_{(0,2\pi)}(m \pm 3s)$. Figure 1 (top row) shows the data and the

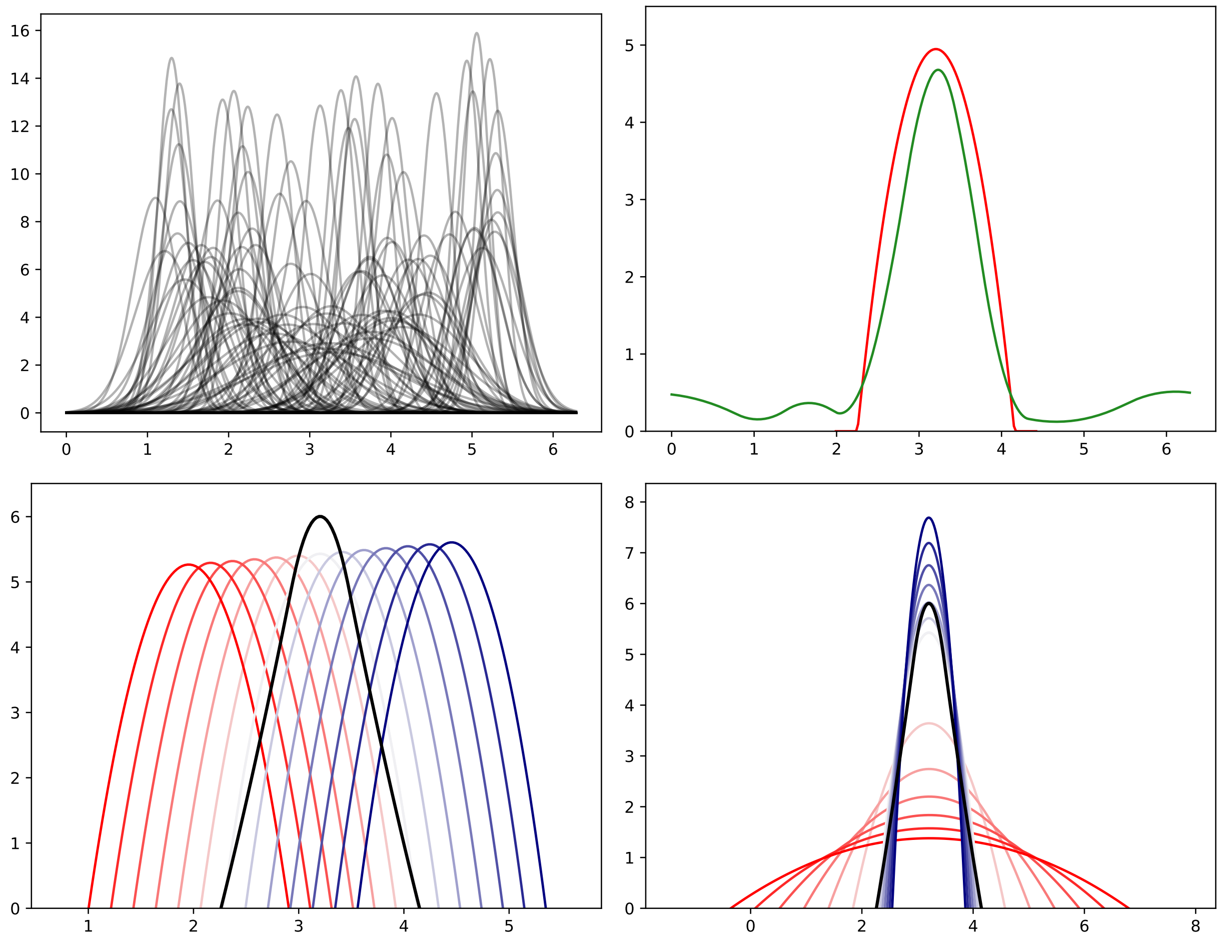

Figure 1: Top row: Simulated data (left) and comparison between Wasserstein barycentres (right): the red line is the barycentre in $\mathcal{W}_2(\mathbb{R})$ while the green one is the barycentre in $\mathcal{W}_2(\mathbb{S}_1)$. Bottom row: first two principal directions in $\mathcal{W}_2(\mathbb{R})$ for the Gaussian measures, the solid black line is the barycentre.

barycentres in $\mathcal{W}_2(\mathbb{R})$ and $\mathcal{W}_2(\mathbb{S}_1)$. In particular, the barycentre in $\mathcal{W}_2(\mathbb{R})$ is unimodal and centred on the domain. Also the barycentre in $\mathcal{W}_2(\mathbb{S}_1)$ presents its tallest mode at the centre of the domain, but can be seen to be trimodal, with, in particular, a mode around the origin. As already pointed out in Section 5, this is due to the fact that the geodesics between some measures, force mass to travel across 0 around the circle.

The first two principal directions in $\mathcal{W}_2(\mathbb{R})$ are displayed in Figure 1 (bottom row), these clearly separate the effect of the location and the one of the scale as expected. When performing PCA in $\mathcal{W}_2(\mathbb{S}_1)$, the scale and location's effects are not clearly separated as shown in Figure 2 (top row): the mass located at the minor modes in the barycentre needs to be moved to match the unimodal measures we generated. In particular, moving along the first principal direction results in a less pronounced change in the location, compared to the PCA on $\mathbb{R}$ and in densities having a more evident mode around the origin. Note that the mode close to 0 tends to get closer to the main mode. Both these effects combine with the second principal direction (not shown in the plots), to approximate the unimodal distributions in the data set.

We also consider a different point $\bar{\mu}$ where to centre the PCA, namely one of the observations, displayed in the solid black line in Figure 2 (bottom). In this case, we set $\log_{\bar{\mu}}(\mu_0)$ equal to the empirical mean of $\log_{\bar{\mu}}(\mu_i)$. This is clearly unimodal and results in a significantly different first principal direction. Indeed, moving along this direction, the densities are unimodal and their location changes quite substantially, although they present a bit of skewness (see the red densities in Figure 2 (bottom). One could argue that this direction is more interpretable than the one found when centering the PCA at the barycentre. However, the reconstruction error using this direction is two times higher than the ones using the "original" direction from the barycentre. We argue that, when centering the PCA in other points than the barycentre, the interpretation of the principal directions as the main "sources of variability" might be misleading, as shown in this case. However, it is true that this procedure might yield practically relevant insights on the data set under analysis: in this case, it is true that the location of the measures changes significantly in the data.

Hence, if the ultimate goal of performing PCA is to gain insight on the data set, and the barycentre presents some features that are not displayed in the datapoints (as in this case, the barycentre is trimodal while all datapoints are unimodal), it might be worth to consider other candidate points where to centre the PCA. If the goal is to perform dimensionality reduction instead, we argue that the barycentre is the only sensible candidate to centre the PCA, as it will surely lead to smaller reconstruction errors, which are synonymous to a smaller loss of information in the reduced data.

D Additional Plots for the Eye Dataset

Figure 3 reports the refined clusters of the eye dataset obtained by cutting the dendrogram at height 0.3. When considering the refined clusters in Figure 3, we see that all clusters are characterised by slightly different shapes: going from left to right, from the first row to the second, the first one has a bump on the left side and is relatively flat on the right side; the second one has a bump on the left side and a smaller bump on the right side; the third one is more stretched along the vertical axis; the fourth one has an even bigger bump on the right side and is flatter on the right side compared to the first cluster; the fifth one is similar in shape to the second one, but has more mass in the lower-left area; the sixth one is much thicker in the upper section than all the other clusters; the last one is quite thick as well in the upper section but with a small bump on the right side and a flatter profile on the left side compared to the sixth cluster.

References

Agueh, M. and G. Carlier (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal. 43*(2), 904–924.

Ambrosio, L., F. Glaudo, and D. Trevisan (2019). On the optimal map in the 2-dimensional random matching problem. *arXiv preprint arXiv:1903.12153*.

Delon, J., J. Salomon, and A. Sobolevski (2010). Fast transport optimization for monge costs on the circle. *SIAM J. Appl. Math. 70*(7), 2239–2258.

Gigli, N. (2011). On the inverse implication of brenier-mccann theorems and the structure of $(P_2(M), W_2)$. *Meth. Appl. of Anal. 18*(2), 127–158.

Kim, Y.-H. and B. Pass (2017). Wasserstein barycenters over Riemannian manifolds. *Adv. Math. 307*, 640–683.

McCann, R. J. (2001). Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal. 11*(3), 589–608.

Panaretos, V. M. and Y. Zemel (2020). *An Invitation to Statistics in Wasserstein Space.* Springer Nature.

Villani, C. (2003). *Topics in Optimal Transportation.*

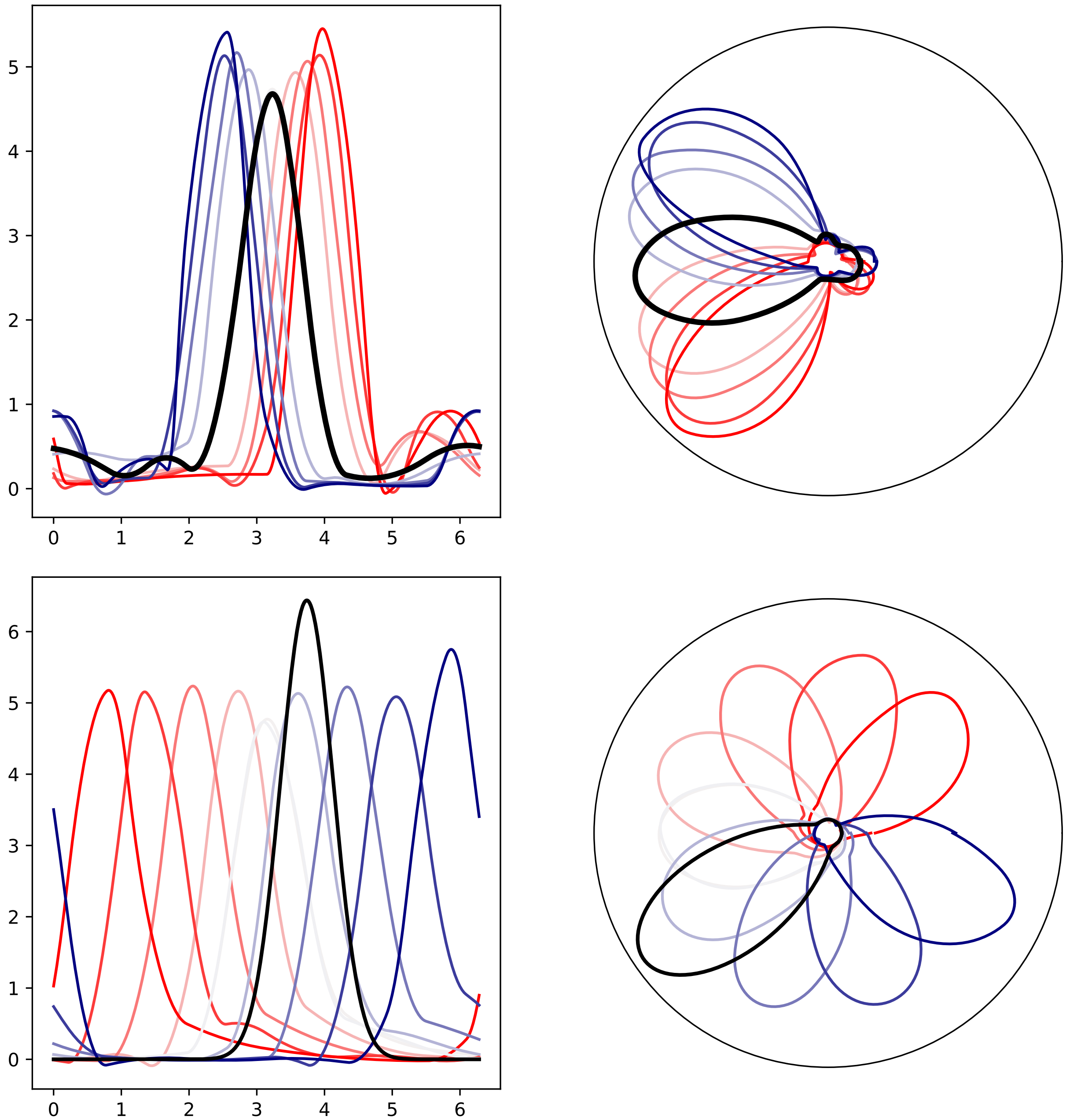

Figure 2: Top row: Bottom row:

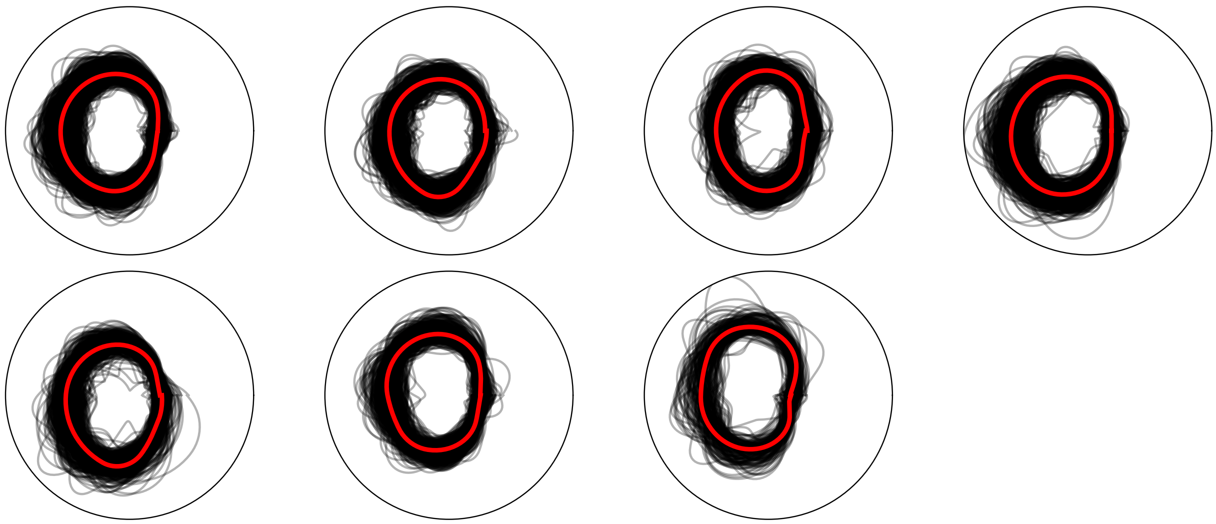

Figure 3: Eye dataset subdivided in seven clusters.

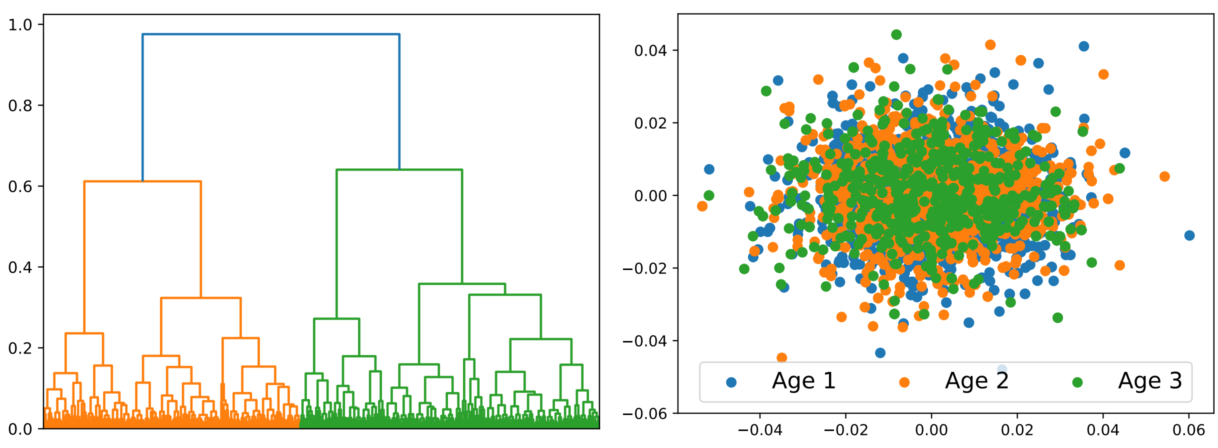

Figure 4: Hierarchical clustering dendrogram (left) and scatter plot of the scores along the first two principal directions, stratified by age group (colour) for the analysis of the OCT measurements dataset analysed in Section 6.