A APPENDIX

A.1 PROOFS

Proof of Proposition 2.3. The proof of Proposition 2.3 is provided in Delon et al. (2010) for the optimal coupling for any pair of probability measures on $\mathbb{S}^1$. For the particular and enlightening case of discrete probability measures on $\mathbb{S}^1$, we refer the reader to Rabin et al. (2011).

For completeness, notice that the relation between x_0 and α hols by changing variables, using 1-periodicity of μ and ν and Definition 2.2 (see also Bonet et al. (2023, Proposition 1)):

$$\begin{aligned}
&\int_0^1 h(|F^{-1}_{\mu,x_0}(x) - F^{-1}_{\nu,x_0}(x)|_{\mathbb{R}})\,dx \\
&\qquad = \int_0^1 h(|(F_\mu(\cdot + x_0) - F_\mu(x_0))^{-1}(x) - (F_\nu(\cdot + x_0) - F_\nu(x_0))^{-1}(x)|_{\mathbb{R}})\,dx \\
&\qquad = \int_0^1 h(|(F_\mu - F_\mu(x_0))^{-1}(x) - (F_\nu - F_\nu(x_0))^{-1}(x)|_{\mathbb{R}})\,dx \\
&\qquad = \int_0^1 h(|F_\mu^{-1}(x + F_\mu(x_0)) - F_\nu^{-1}(x + F_\nu(x_0))|_{\mathbb{R}})\,dx \\
&\qquad = \int_0^1 h(|F_\mu^{-1}(x + \underbrace{F_\mu(x_0) - F_\nu(x_0)}_{\alpha}) - F_\nu^{-1}(x)|_{\mathbb{R}})\,dx
\end{aligned}$$

In particular, if $h(x) = |x|^2$, and $\mu = Unif(\mathbb{S}^1)$, then

$$\begin{aligned}
COT_2(\mu,\nu) &= \inf_{\alpha\in\mathbb{R}} \int_0^1 |F_\mu^{-1}(x+\alpha) - F_\nu^{-1}(x)|^2_{\mathbb{R}}\,dx \\
&= \inf_{\alpha\in\mathbb{R}} \int_0^1 |x + \alpha - F_\nu^{-1}(x)|^2\,dx \\
&= \inf_{\alpha\in\mathbb{R}} \left(\int_0^1 |F_\nu^{-1}(x) - x|^2\,dx - 2\alpha \int_0^1 (F_\nu^{-1}(x) - x)\,dx + \alpha^2 \right) \\
&= \inf_{\alpha\in\mathbb{R}} \left(\int_0^1 |F_\nu^{-1}(x) - x|^2\,dx - 2\alpha \left(\int_0^1 x\,d\nu(x) - \frac{1}{2} \right) + \alpha^2 \right) \\
&= \inf_{\alpha\in\mathbb{R}} \left(\int_0^1 |F_\nu^{-1}(x) - x|^2\,dx - 2\alpha \left(\mathbb{E}(\nu) - \frac{1}{2} \right) + \alpha^2 \right) \\
&= \int_0^1 |F_\nu^{-1}(x) - x|^2\,dx - 2\underbrace{\left(\mathbb{E}(\nu) - \frac{1}{2} \right)^2}_{\alpha_{\mu,\nu}}.
\end{aligned}$$

Therefore, in this case, the minimizer $\alpha_{\mu,\nu}$ of equation 8 is unique and has the closed-form $\alpha_{\mu,\nu} = \mathbb{E}(\nu) - 1/2$. □

Proof of Remark 2.4. We will show that, *in general*, the minimizer $\alpha_{\mu,\nu}$ of equation 8 is unique. Our arguments are based on the paper Delon et al. (2010). Specifically, the role played by the function $(F^\theta)^{-1}$, where $F^\theta(x) = F(x) + \theta$ in Delon et al. (2010) is substituted by $F^{-1}(x - \alpha)$ in our case (i.e., our parameter α correspond to $-\theta$ in the mentioned paper).

Our hypotheses are the following:

1. Let $c(x,y) := h(|x-y|)$ for $h : \mathbb{R} \to \mathbb{R}_+$ strictly convex (for example $h(x) = |x|^p$ with $p > 1$), or, more generally, let $c : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ satisfying the *Monge condition* or the *(strict) cyclical monotonicity condition:*

$$c(u_1,v_1) + c(u_2,v_2) < c(u_1,v_2) + c(u_2,v_1) \qquad \forall\, u_1 < u_2,\ v_1 < v_2. \tag{23}$$

2. Let μ, ν be two probability measures absolutely continuous with respect to the Lebesgue measure on $\mathbb{S}^1$.

The idea of the proof will rely on showing that the cost function

$$\text{Cost}(\alpha) := \int_0^1 c(F_\mu^{-1}(x), F_\nu^{-1}(x-\alpha))\, dx \tag{24}$$

is strictly convex and continuous (as a function on α), and so it has a unique global minimum (that we will denote by $\alpha_{\mu,\nu}$).

Let

$$c_{\mu,\nu}(x,y) := c(F_\mu^{-1}(x), F_\nu^{-1}(y)) \tag{25}$$

Under the above conditions, it holds that $c_{\mu,\nu}(\cdot,\cdot)$ satisfies the Monge condition equation 23. Then, to prove strictly convexity of $\text{Cost}(\alpha)$ it is sufficient to show that

$$\text{Cost}\left(\frac{\alpha'+\alpha''}{2}\right) < \frac{\text{Cost}(\alpha') + \text{Cost}(\alpha'')}{2} \tag{26}$$

Assume $\alpha' \leq \alpha''$ and let $\overline{\alpha} := \frac{\alpha'+\alpha''}{2}$. On the one hand, since

$$\begin{aligned}
\text{Cost}\,(\overline{\alpha}) &= \int_0^1 c_{\mu,\nu}(x, x-\overline{\alpha})\, dx \\
&= \int_{\alpha''-\overline{\alpha}}^{1+\alpha''-\overline{\alpha}} c_{\mu,\nu}(x, x-\overline{\alpha})\, dx \\
&= \int_0^1 c_{\mu,\nu}(y+\alpha''-\overline{\alpha}, y-\alpha')\, dy,
\end{aligned}$$

where in the last line we used the change of variables $y = x+\alpha''-\overline{\alpha}$ and the fact that $2\overline{\alpha}-\alpha'' = \alpha'$. Therefore,

$$2\,\text{Cost}\,(\overline{\alpha}) = \int_0^1 c_{\mu,\nu}(z, z-\overline{\alpha})\, dz + \int_0^1 c_{\mu,\nu}(z+\alpha''-\overline{\alpha}, z-\alpha')\, dz. \tag{27}$$

On the other hand, by repeating the same idea we have,

$$\begin{aligned}
\text{Cost}(\alpha'') &= \int_0^1 c_{\mu,\nu}(x, x-\alpha'')\, dx = \int_{\alpha''-\overline{\alpha}}^{1+\alpha''-\overline{\alpha}} c_{\mu,\nu}(x, x-\alpha'')\, dx \\
&= \int_0^1 c_{\mu,\nu}(y+\alpha''-\overline{\alpha}, y-\overline{\alpha})\, dy,
\end{aligned}$$

and so,

$$\text{Cost}(\alpha') + \text{Cost}(\alpha'') = \int_0^1 c_{\mu,\nu}(z, z-\alpha')\, dz + \int_0^1 c_{\mu,\nu}(z+\alpha''-\overline{\alpha}, z-\overline{\alpha})\, dz. \tag{28}$$

Since $u_1(z) := z < z+\alpha''-\overline{\alpha} =: u_2(x)$ and $v_1(z) := z-\overline{\alpha} < z-\alpha' =: v_2(z)$, we have that

$$c(u_1(z), v_1(z)) + c(u_2(z), v_2(z)) < c(u_1(z), v_2(z)) + c(u_2(z), v_1(z))$$

because $c_{\mu,\nu}(\cdot,\cdot)$ satisfies Monge condition equation 23. Thus, from equation 27 and equation 28 we obtain

$$\begin{aligned}
2\,\text{Cost}(\overline{\alpha}) &= \int_0^1 c_{\mu,\nu}(u_1(z), v_1(z))\, dz + \int_0^1 c_{\mu,\nu}(u_2(z), v_2(z))\, dz \\
&< \int_0^1 c_{\mu,\nu}(u_1(z), v_2(z))\, dz + \int_0^1 c_{\mu,\nu}(u_2(z), v_1(z))\, dz = \text{Cost}(\alpha') + \text{Cost}(\alpha''),
\end{aligned}$$

and so equation 26 holds. The continuity of $\alpha \mapsto \text{Cost}(\alpha)$ holds as the integral in equation 24 is finite for all α.

□

Remark A.1. *We mention that in the case that $c(x,y) = h(x-y)$ for $h(x) = |x|$ (studied for example in Cabrelli & Molter (1998); Hundrieser et al. (2022)) we have convexity but not strictly convexity. However, the authors in Hundrieser et al. (2022) prove that a closed-formula for a minimizer $\alpha_{\mu,\nu}$ of equation 8:*

$$\alpha_{\mu,\nu} = \min\left\{\arg\min_{u\in\mathbb{R}} \int_0^1 |(F_\mu - F_\nu)(t) - u| dt\right\}$$

called the Level Median of the function $F_\mu - F_\nu$. To show that, it is used the fact that in this case equation 8 takes the form

$$COT_h(\mu,\nu) = \inf_{\alpha\in\mathbb{R}} \int_0^1 |F_\mu(t) - F_\nu(t) - \alpha|\, dt.$$

Proof of Theorem 2.5.

1. First, we will show that the map M_μ^ν given by equation 14 satisfies $(M_\mu^\nu)_\# \mu = \nu$. Here μ and ν are the extended measures form $\mathbb{S}^1$ to $\mathbb{R}$ having CDFs equal to F_μ and F_ν, respectively, defined by equation 3 and 4. By choosing the system of coordinates $\widetilde{x} \in [0,1)$ that starts at x_{cut} (see Figure 7) then,

$$M_\mu^\nu(\widetilde{x}) = F^{-1}_{\nu,x_{cut}} \circ F_{\mu,x_{cut}}(\widetilde{x})$$

(see equation 12). Let $\mu_{x_{cut}}$ and $\nu_{x_{cut}}$ be the (1-periodic) measures on $\mathbb{R}$ having CDFs $F_{\mu,x_{cut}}$ and $F_{\nu,x_{cut}}$, respectively, i.e., $F_{\nu,x_{cut}}(\widetilde{x}) = \mu_{x_{cut}}([0,\widetilde{x}))$ (analogously for $\nu_{x_{cut}}$). That is, we have unrolled μ and ν from $\mathbb{S}^1$ to $\mathbb{R}$, where the origin $0 \in \mathbb{R}$ corresponds to $x_{cut} \in \mathbb{S}^1$ (see Figure 1). Thus, a classic computation yields

$$(F^{-1}_{\nu,x_{cut}} \circ F_{\mu,x_{cut}})_\# \mu_{x_{cut}} = (F^{-1}_{\nu,x_{cut}})_\# \left((F_{\mu,x_{cut}})_\# \mu_{x_{cut}}\right) = (F^{-1}_{\nu,x_{cut}})_\# \mathcal{L}_{\mathbb{S}^1} = \nu_{x_{cut}}$$

where $\mathcal{L}_{\mathbb{S}^1} = Unif(\mathbb{S}^1)$ denotes the Lebesgue measure on the circle. We used that $(F_\mu)_\# \mu = \mathcal{L}_{\mathbb{S}^1}$ as μ does not give mass to atoms, and so, if we change the system of coordinates we also have $(F_{\mu,x_{cut}})_\# \mu_{x_{cut}} = \mathcal{L}_{\mathbb{S}^1}$.

Finally, we have to switch coordinates. Let

$$z(\widetilde{x}) := \widetilde{x} + x_{cut}$$

(that is, $z(\widetilde{x}) = x$). To visualize this, see Figure 7. It holds that

$$z_\# \nu_{cut} = \nu \tag{29}$$

(where we recall that ν is the extended measure form $\mathbb{S}^1$ to $\mathbb{R}$ having CDF equal to F_μ as in equation 3 and 4). Let us check this fact for intervals:

$$\begin{aligned} z_\# \nu_{x_{cut}}([a,b]) &= \nu_{x_{cut}}(z^{-1}([a,b])) = \nu([z^{-1}(a), z^{-1}(b)]) \\ &= \nu_{x_{cut}}([a - x_{cut}, b - x_{cut}]) \\ &= F_{\nu,x_{cut}}(b - x_{cut}) - F_{\nu,x_{cut}}(a - x_{cut}) \\ &= F_\nu(b) - F_\nu(x_{cut}) - (F_\nu(a) - F_\nu(x_{cut})) \\ &= F_\nu(b) - F_\nu(a) \\ &= \nu([a,b]). \end{aligned}$$

Besides, it holds that

$$F_{\mu,x_{cut}}(\cdot - x_{cut})_\# \mu = Unif(\mathbb{S}^1), \tag{30}$$

in the sense that it is the Lebesgue measure on $\mathbb{S}^1$ extended periodically (with period 1) to the real line, which we denote by $\mathcal{L}_{\mathbb{S}^1}$. Let us sketch the proof for intervals. First, notice that $F_{\mu,x_{cut}}(x - x_{cut}) = F_\mu(x) - F_\mu(x_{cut})$ and so its inverse is $y \mapsto F_\mu^{-1}(y + x_{cut})$. Therefore,

$$\begin{aligned} (F_{\mu,x_{cut}}(\cdot - x_{cut}))_\# \mu\,([a,b]) &= \mu\left([F_\mu^{-1}(a + x_{cut}), F_\mu^{-1}(b + x_{cut})]\right) \\ &= F_\mu(F_\mu^{-1}(a + x_{cut})) - F_\mu(F_\mu^{-1}(b + x_{cut})) = b - a. \end{aligned}$$

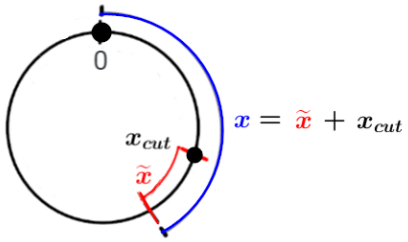

Figure 7: The unit circle (black) can be parametrized as $[0, 1)$ in many different ways. In the figure, we marked in black the North Pole as 0. The canonical parametrization of $\mathbb{S}^1$ identifies the North Pole with 0. Then, also in black, we pick a point x_{cut}. The distance in blue x that starts at 0 equals the distance in red $\widetilde{x}$ that starts at x_{cut} plus the corresponding starting point x_{cut}. This allows us to visualize the change of coordinates given by equation 13.

Finally,

$$\begin{aligned}
(M_\mu^\nu)_\# \mu &= \left(F^{-1}_{\nu,x_{cut}}(F_{\mu,x_{cut}}(\cdot - x_{cut}) + x_{cut})\right)_\# \mu \\
&= \left(z(F^{-1}_{\nu,x_{cut}}(F_{\mu,x_{cut}}(\cdot - x_{cut})))\right)_\# \mu \\
&= z_\#(F^{-1}_{\nu,x_{cut}})_\#(F_{\mu,x_{cut}}(\cdot - x_{cut}))_\#\mu \\
&= z_\#(F^{-1}_{\nu,x_{cut}})_\#\mathcal{L}_{\mathbb{S}^1} \qquad \text{(by equation 30)} \\
&= z_\#\nu_{x_{cut}} \\
&= \nu \qquad \text{(by equation 29).}
\end{aligned}$$

Now, let us prove that M_μ^ν is optimal.

First, assume that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{S}^1$ and let f_μ denote its density function. We will use the change of variables

$$\begin{cases} u = F_{\mu,x_{cut}}(x - x_{cut}) = F_\mu(x) - F_\mu(x_{cut}) \\ du = f_\mu(x)\, dx. \end{cases}$$

So,

$$\begin{aligned}
\int_0^1 h(|M_\mu^\nu(x) - x|_\mathbb{R})\, d\mu(x) &= \int_0^1 h(|F^{-1}_{\nu,x_{cut}}(F_{\mu,x_{cut}}(x - x_{cut})) - (x - x_{cut})|_\mathbb{R})\, \underbrace{f_\mu(x)dx}_{d\mu(x)} \\
&= \int_{-x_{cut}}^{1-x_{cut}} h(|F^{-1}_{\nu,x_{cut}}(u) - F^{-1}_{\mu,x_{cut}}(u)|_\mathbb{R})\, du \\
&= \int_0^1 h(|F^{-1}_{\nu,x_{cut}}(u) - F^{-1}_{\mu,x_{cut}}(u)|_\mathbb{R})\, du \\
&= COT_h(\mu, \nu).
\end{aligned}$$

Now, let us do the proof in general:

$$\begin{aligned}
\int_0^1 h(|M_\mu^\nu(x) - x|_\mathbb{R})\, d\mu(x) &= \int_0^1 h(|F^{-1}_{\nu,x_{cut}}(F_{\mu,x_{cut}}(x - x_{cut})) - (x - x_{cut})|_\mathbb{R})\, d\mu(x) \\
&= \int_0^1 h(|F^{-1}_{\nu,x_{cut}}(y) - F^{-1}_{\mu,x_{cut}}(y)|_\mathbb{R})\, d(F_{\mu,x_{cut}}(\cdot - x_{cut}))_\#\mu(y) \\
&= \int_0^1 h(|F^{-1}_{\nu,x_{cut}}(u) - F^{-1}_{\mu,x_{cut}}(u)|_\mathbb{R})\, du \\
&= COT_h(\mu, \nu).
\end{aligned}$$

In the last equality we have used that $F_{\mu,x_{cut}}(\cdot - x_{cut})_\#\mu$ is the Lebesgue measure (see equation 30).

2. Using the definition of the generalized inverse (quantile function), we have

$$\begin{aligned}
M_\mu^\nu(t) &= F^{-1}_{\nu,x_{cut}}(F_{\mu,x_{cut}}(x - x_{cut})) + x_{cut} \\
&= \inf\{x' : F_{\nu,x_{cut}}(x') > F_{\mu,x_{cut}}(x - x_{cut})\} + x_{cut} \\
&= \inf\{x' : F_\nu(x' + x_{cut}) - F_\nu(x_{cut}) > F_\mu(x) - F_\mu(x_{cut})\} + x_{cut} \\
&= \inf\{x' : F_\nu(x' + x_{cut}) > F_\mu(x) - F_\mu(x_{cut}) + F(x_{cut})\} + x_{cut} \\
&= \inf\{x' : F_\nu(x' + x_{cut}) > F_\mu(x) - \alpha_{\mu,\nu}\} + x_{cut} \\
&= \inf\{y - x_{cut} : F_\nu(y) > F_\mu(x) - \alpha_{\mu,\nu}\} + x_{cut} \\
&= \inf\{y : F_\nu(y) > F_\mu(x) - \alpha_{\mu,\nu}\} + x_{cut} - x_{cut} \\
&= F_\nu^{-1}(F_\mu(x) - \alpha_{\mu,\nu}).
\end{aligned}$$

3. This part follows from the previous item as the right-hand side of equation 16 does not depend on any minimizer of equation 7.

4. From (McCann, 2001, Theorem 13), there exists a unique optimal Monge map for the optimal transport problem on the unit circle. Therefore, by using Remark 2.4, M_μ^ν is the unique optimal transport map from μ to ν. For the quadratic case $h(x) = |x|^2$, we refer for example to Santambrogio (2015, Th. 1.25, Sec. 1.3.2)). Moreover, in this particular case, there exists a function φ such that $M_\mu^\nu(x) = x - \nabla\varphi(x)$, where φ is a *Kantorovich potential* (that is, a solution to the dual optimal transport problem on $\mathbb{S}^1$) and the sum is modulo $\mathbb{Z}$.

5. The identity $(M_\mu^\nu)^{-1} = (M_\nu^\mu)$ holds from the symmetry of the cost equation 5 that one should optimize. Also, it can be verified using equation 16 and the fact that from equation 9 $\alpha_{\mu,\nu} = -\alpha_{\nu,\mu}$:

$$\begin{aligned}
M_\nu^\mu \circ M_\mu^\nu(x) &= F_\mu^{-1}\left(F_\nu(F_\nu^{-1}(F_\mu(x) - \alpha_{\mu,\nu})) - \alpha_{\nu,\mu}\right) \\
&= F_\mu^{-1}\left(F_\mu(x) - \alpha_{\mu,\nu} + \alpha_{\mu,\nu}\right) = x.
\end{aligned}$$

□

Proposition A.2 (Properties of the LCOT-Embedding)**.** *Let $\mu \in \mathcal{P}(\mathbb{S}^1)$ be absolutely continuous with respect to the Lebesgue measure on $\mathbb{S}^1$, and let $\nu \in \mathcal{P}(\mathbb{S}^1)$.*

1. $\widehat{\mu}^{\mu,h} \equiv 0$.
2. $\widehat{\nu}^{\mu,h}(x) \in [-0.5, 0.5]$ *for every* $x \in [0, 1)$.
3. *Let $\nu_1, \nu_2 \in \mathcal{P}(\mathbb{S}^1)$ with ν_1 that does not give mass to atoms, then the map*

$$M := (\widehat{\nu_2}^{\mu,h} - \widehat{\nu_1}^{\mu,h}) \circ ((\widehat{\nu_1}^{\mu,h} + \mathrm{id})^{-1}) + \mathrm{id}, \tag{31}$$

satisfies $M_\# \nu_1 = \nu_2$ (however, it is not necessarily an optimal circular transport map).

Proof of Proposition A.2.

1. It trivially holds that the optimal Monge map from the distribution μ to itself is the identity id, or equivalently, that the optimal displacement is zero for all the particles.

2. It holds from the fact of being the optimal displacement, that is,

$$COT_h(\mu, \nu) = \inf_{M:\, M_\#\mu = \nu} \int_{\mathbb{S}^1} h(|M(x) - x|_{\mathbb{S}^1})\, d\mu(x) = \int_{\mathbb{S}^1} h(|\widehat{\nu}^{\mu,h}(x)|_{\mathbb{S}^1})\, d\mu(x),$$

and from the fact that $|z|_{\mathbb{S}^1}$ is at most 0.5.

3. We will use that $\widehat{\nu}^{\mu,h} = M_\mu^\nu - \mathrm{id}$, and that $(M_\mu^\nu)^{-1} = M_\nu^\mu$:

$$\begin{aligned} M(x) &= (\widehat{\nu_2}^{\mu,h} - \widehat{\nu_1}^{\mu,h}) \circ M_{\nu_1}^{\mu}(x) + x \\ &= (M_{\mu}^{\nu_2} - M_{\mu}^{\nu_1}) \circ M_{\nu_1}^{\mu}(x) + x \\ &= M_{\mu}^{\nu_2} \circ M_{\nu_1}^{\mu}(x) - x + x \\ &= M_{\mu}^{\nu_2} \circ M_{\nu_1}^{\mu}(x). \end{aligned}$$

Finally, notice that

$$(M_{\mu}^{\nu_2} \circ M_{\nu_1}^{\mu})_{\#}\nu_1 = (M_{\mu}^{\nu_2})_{\#}\left((M_{\nu_1}^{\mu})_{\#}\nu_1\right) = (M_{\mu}^{\nu_2})_{\#}\mu = \nu_2.$$

□

Now, we will proceed to prove Theorem 3.6. By having this result, it is worth noticing that $LCOT_{\mu,p}(\cdot,\cdot)^{1/p}$ endows $\mathcal{P}(\mathbb{S}^1)$ with a metric-space structure. The proof is based on the fact that we have introduced an explicit embedding and then we have considered an L^p-distance. It will follow that we have defined a kernel distance (that is in fact positive semidefinite).

Proof of Theorem 3.6. From equation 20, it is straightforward to prove the symmetric property and non-negativity.

If $\nu_1 = \nu_2$, by the uniqueness of the optimal COT map (see Theorem 2.5, Part 3), we have $\widehat{\nu_1}^{\mu,h} = \widehat{\nu_2}^{\mu,h}$. Thus, $LCOT_{\mu,h}(\nu_1,\nu_2) = 0$.

For the reverse direction, if $LCOT_{\mu,h}(\nu^1,\nu^2) = 0$, then

$$h(\min_{k\in\mathbb{Z}}\{|\widehat{\nu_1}^{\mu,h}(x) - \widehat{\nu_2}^{\mu,h}(x) + k|\}) = 0 \qquad \mu - \text{a.s.}$$

Thus,

$$\widehat{\nu_1}^{\mu,h}(x) \equiv_1 \widehat{\nu_2}^{\mu,h}(x) \qquad \mu - \text{a.s.}$$

(where $\equiv_1$ stands for the equality modulo $\mathbb{Z}$). That is,

$$M_{\mu}^{\nu_1}(x) = \widehat{\nu_1}^{\mu,h}(x) + x \equiv_1 \widehat{\nu_2}^{\mu,h}(x) + x = M_{\mu}^{\nu_2}(x) \qquad \mu \text{ a.s.}$$

Let $S \subseteq [0,1)$ denote the set of x such that the equation above holds, we have $\mu(S) = 1, \mu(\mathbb{S}^1 \backslash S) = 0$. Equivalently, for any (measurable) $B \subseteq \mathbb{S}^1$, $\mu(B \cap S) = \mu(B)$. Pick any Borel set $A \subseteq \mathbb{S}^1$, we have:

$$\begin{aligned} \nu_1(A) &= \mu\left((M_{\mu}^{\nu_1})^{-1}(A)\right) \\ &= \mu\left((M_{\mu}^{\nu_1})^{-1}(A) \cap S\right) \\ &= \mu\left((M_{\mu}^{\nu_2})^{-1}(A) \cap S\right) \\ &= \mu((M_{\mu}^{\nu_2})^{-1}(A)) \\ &= \nu_2(A) \end{aligned} \tag{32}$$

where the first and last equation follows from the fact $M_{\mu}^{\nu_1}, M_{\mu}^{\nu_2}$ are push forward mapping from μ to ν_1, ν_2 respectively.

Finally, we verify the triangular inequality. Here we will use that $h(x) = |x|^p$, for $1 \leq p < \infty$. Let $\nu_1, \nu_2, \nu_3 \in \mathcal{P}(\mathbb{S}^1)$,

$$\begin{aligned} LCOT_{\mu,p}(\nu_1,\nu_2)^{1/p} &= \left(\int_0^1 (|\widehat{\nu_1}(t) - \widehat{\nu_2}(t)|_{\mathbb{S}^1})^p \, d\mu(t)\right)^{1/p} \\ &\leq \left(\int_0^1 (|\widehat{\nu_1}(t) - \widehat{\nu_3}(t)|_{\mathbb{S}^1} + |\widehat{\nu_3}(t) - \widehat{\nu_2}(t)|_{\mathbb{S}^1})^p \, d\mu(t)\right)^{1/p} \\ &\leq \left(\int_0^1 |\widehat{\nu_1}(t) - \widehat{\nu_3}(t)|_{\mathbb{S}^1}^p \, d\mu(t)\right)^{1/p} + \left(\int_0^1 |\widehat{\nu_3}(t) - \widehat{\nu_2}(t)|_{\mathbb{S}^1}^p \, d\mu(t)\right)^{1/p} \\ &= LCOT_{\mu,p}(\nu_1,\nu_3)^{1/p} + LCOT_{\mu,p}(\nu_2,\nu_3)^{1/p} \end{aligned}$$

where the last inequality holds from Minkowski inequality. □

A.2 UNDERSTANDING THE RELATION BETWEEN THE MINIMIZERS OF EQUATION 7 AND EQUATION 8

We briefly revisit the discussion in Section equation 2.3.1, specifically in Remark equation 2.4, concerning the optimizers x_{cut} and $\alpha_{\mu,\nu}$ of equation 7 and equation 8, respectively.

Assuming minimizers exist for equation 7 and equation 8, we first explain why we adopt the terminology *"cutting point"* (x_{cut}) for a minimizer of equation 7 and not for the minimizer $\alpha_{\mu,\nu}$ of equation 8. On the one hand, the cost function presented in 7 is given by

$$\text{Cost}(x_0) := \int_0^1 h(|F^{-1}_{\mu,x_0}(x) - F^{-1}_{\nu,x_0}(x)|_{\mathbb{R}})\, dx. \tag{33}$$

We seek to minimize over $x_0 \in [0,1) \sim \mathbb{S}^1$, aiming to find an optimal x_0 that affects both CDFs F_μ and F_ν. By looking at the cost 33, for each fixed x_0, we change the system of reference by adopting x_0 as the origin. Then, once an optimal x_0 is found (called x_{cut}), it leads to the optimal transportation displacement, providing a change of coordinates to unroll the CDFs of μ and ν into $\mathbb{R}$ and allowing the use the classical Optimal Transport Theory on the real line (see Section equation 2.3.2 and the proofs in Appendix equation A.1). On the other hand, the cost function in 8 is

$$\text{Cost}(\alpha) := \int_0^1 h(|F^{-1}_{\mu}(x+\alpha) - F^{-1}_{\nu}(x)|_{\mathbb{R}})\, dx,$$

and the minimization runs over every real number α. Here, the shift by α affects only one of the CDFs, not both. Therefore, it will not allow for a consistent change in the system of reference. This is why we do not refer to α as a cutting point in this paper, but we do refer to the minimizer of equation 7 as x_{cut}.

Finally, Figure 8 below is meant to provide a visualization of Remark 2.4, that is, to show through an example that, when minimizers for 7 and equation 8 do exist, while one could have multiple minimizers of 7, the minimizer of 8 is unique.

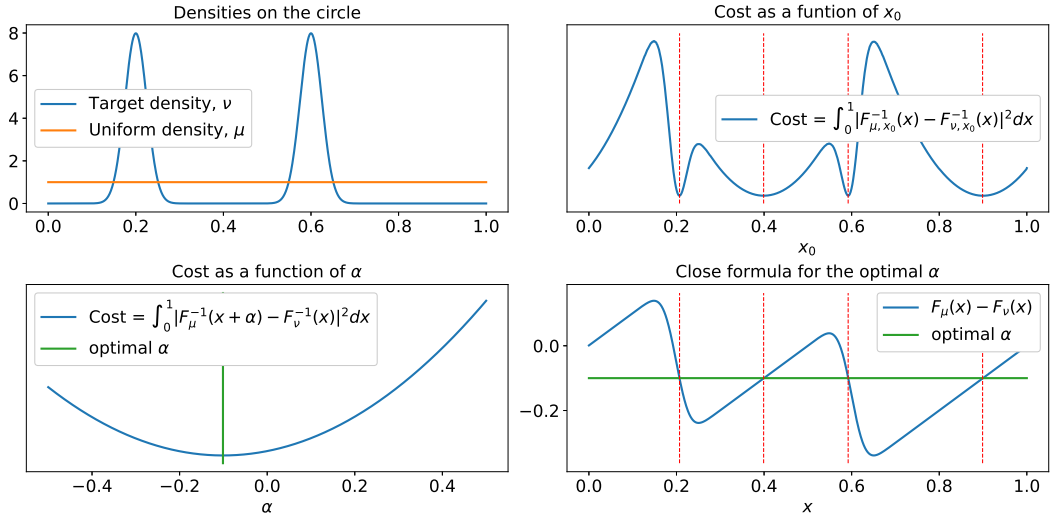

Figure 8: Top left: Uniform density, μ, and a random target density ν on $\mathbb{S}^1$. Top right: The circular transportation cost $\int_0^1 |F^{-1}_{\mu,x_0}(x) - F^{-1}_{\nu,x_0}(x)|^2\, dx$ is depicted as a function of the cut, x_0, showing that the optimization in equation 7 can have multiple minimizers. Bottom right: Following equation 9, we depict the difference between the two CDFs, $F_\mu(x) - F_\nu(x)$, for each $x \in [0,1) \sim \mathbb{S}^1$. As can be seen, for the optimal cuts (dotted red lines), the difference is constant, indicating that the optimal α for equation 8 is unique. Bottom left: The optimizer for the circular transportation cost in equation 8 is unique, and given that μ is the uniform measure, it has a closed-form solution $\mathbb{E}(\nu) - \frac{1}{2}$.

A.3 TIME COMPLEXITY OF LINEAR COT

In this section, we assume that we are given discrete or empirical measures.[2]

First, we mention that according to (Delon et al., 2010, Section 6), given two non-decreasing step functions F and G represented by

$$[\,[x_1,\dots,x_{N_1}],\ [F(x_1),\dots,F(x_{N_1})]\,] \qquad \text{and} \qquad [\,[y_1,\dots,y_{N_2}],\ [G(y_1),\dots,G(y_{N_2})]\,],$$

the computation of an integral of the form

$$\int c(F^{-1}(x), G^{-1}(x))\,dx$$

requires $\mathcal{O}(N_1+N_2)$ evaluations of a given cost function $c(\cdot,\cdot)$.

Now, by considering the reference measure $\mu = Unif(\mathbb{S}^1)$ we will detail our algorithm for computing $LCOT(\nu_1,\nu_2)$. Let us assume that ν_1,ν_2 are two discrete probability measures on $\mathbb{S}^1$ having N_1 and N_2 masses, respectively. We represent these measures $\nu_i = \sum_{j=1}^{N_i} m_j^i \delta_{x_j^i}$ (that is, ν_1 has mass m_j^1 at location x_j^1 for $j=1,\dots,N_1$, and analogously for ν_2) as arrays of the form

$$\nu_i = [\,[x_1^i,\dots,x_{N_i}^i],\ [m_1^i,\dots,m_{N_i}^i]\,], \qquad i=1,2.$$

Algorithm to compute LCOT:

1. For $i=1,2$, compute $\alpha_{\mu,\nu_i} = \mathbb{E}(\nu_i) - 1/2$.
2. For $i=1,2$, represent $F_{\nu_i}(\cdot) + \alpha_{\mu,\nu_i}$ as the arrays
$$[\,[x_1^i,\dots,x_{N_i}^i],\ [c_1^i,\dots,c_{N_i}^i]\,]$$
where
$$c_1^i := m_1^i + \alpha_{\mu,\nu_i}, \qquad c_j^i := c_{j-1}^i + m_j^i, \quad \text{for } j=2,\dots,N_i.$$
3. Use that
$$F_\nu^{-1}(x-\alpha_{\mu,\nu}) = (F_\nu(\cdot)+\alpha_{\mu,\nu})^{-1}(x),$$
and the algorithm provided in (Delon et al., 2010, Section 6) mentioned above with $F = F_{\nu_1}(\cdot)+\alpha_{\mu,\nu_1}$ and $G = F_{\nu_2}(\cdot)+\alpha_{\mu,\nu_2}$ to compute
$$\begin{aligned} LCOT(\nu_1,\nu_2) &= \|\widehat{\nu_1} - \widehat{\nu_2}\|^2_{L^2(\mathbb{S}^1)} \\ &= \int_0^1 |\,\big(F_{\nu_1}^{-1}(x-\alpha_{\mu,\nu_1}) - x\big) - \big(F_{\nu_2}^{-1}(x-\alpha_{\mu,\nu_2}) - x\big)\,|^2_{\mathbb{S}^1}\,dx \\ &= \int_0^1 |\underbrace{(F_{\nu_1}(\cdot)+\alpha_{\mu,\nu_1})}_{F}{}^{-1}(x) - \underbrace{(F_{\nu_2}(\cdot)+\alpha_{\mu,\nu_2})}_{G}{}^{-1}(x)|^2_{\mathbb{S}^1}\,dx \end{aligned}$$

Each step requires $\mathcal{O}(N_1+N_2)$ operations. Therefore, the full algorithm to compute $LCOT(\nu_1,\nu_2)$ is of order $\mathcal{O}(N_1+N_2)$.

A.4 LCOT BARYCENTER

Although the following definition holds for any non-atomic reference measure $\mu \in \mathcal{P}(\mathbb{S}^1)$, for simplicity, we consider the reference measure as $\mu = Unif(\mathbb{S}^1)$.

Given N target measures $\nu_1,\dots,\nu_N \in \mathcal{P}(\mathbb{S}^1)$, as $LCOT_2(\cdot,\cdot)$ is a distance, their ***LCOT barycenter*** is defined by the measure ν_b such that

$$\nu_b = argmin_{\nu\in\mathcal{P}(\mathbb{S}^1)} \frac{1}{N}\sum_{j=1}^N LCOT_2(\nu,\nu_j) = argmin_{\nu\in\mathcal{P}(\mathbb{S}^1)} \frac{1}{N}\sum_{j=1}^N \|\widehat{\nu} - \widehat{\nu_j}\|^2_{L^2(\mathbb{S}^1)}.$$

[2]It is worth mentioning that for some applications, the LCOT framework can be also used for continuous densities, as in the case of the CDT Park et al. (2018).

In the embedding space, it can be shown that the minimizer of

$$argmin_{\widehat{\nu}} \frac{1}{N} \sum_{j=1}^{N} \|\widehat{\nu} - \widehat{\nu_j}\|^2_{L^2(\mathbb{S}^1)}$$

is given by the circular mean

$$\overline{\nu}(x) := \text{circle mean}(\{\widehat{\nu_1}(x), \dots \widehat{\nu_N}(x)\}) := \frac{1}{2\pi} \arg\tan\left(\frac{\sum_{i=1}^{N} \sin(2\pi\widehat{\nu_i}(x))}{\sum_{i=1}^{N} \cos(2\pi\widehat{\nu_i}(x))}\right).$$

For each $x \in [0, 1)$, the last value is the average of the angles $\{2\pi\widehat{\nu_1}(x), \dots, 2\pi\widehat{\nu_N}(x)\}$, which is then normalized to fall within the range $[-0.5, 0.5]$. By using the closed formula for the inverse of the LCOT Embedding provided by Proposition 3.5, we can go back to the measure space obtaining the *LCOT barycenter* between $\nu_1, \dots, \nu_N$ as

$$\nu_b = (\overline{\nu} + \text{id})_{\#}\mu. \tag{34}$$

In our experiments, we use the expression equation 34.

A.5 Extra Figures and Experiments

The following Figure 9 is from an experiment analogous to Experiment 1 but for a different family of measures (Figure 9 Left). We include it to have an intuition of how the LCOT behaves under translations and dilations of an initial von Mises density.

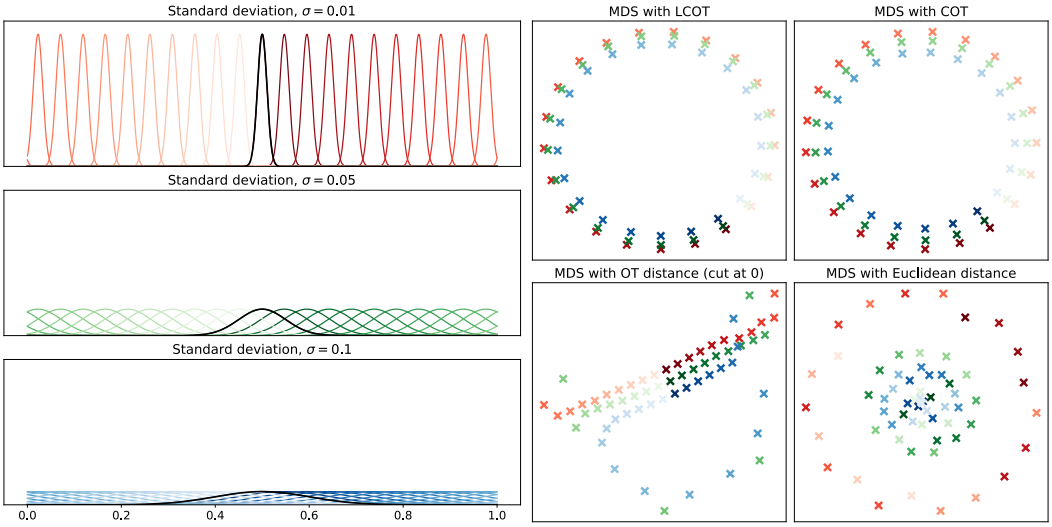

Figure 9: MDS for embedding classes of probability densities into an Euclidean space of dimension 2 where the original pair-wise distances (COT-distance, LOT-distance, Euclidean or L^2-distance) are preserved as well as possible.

A.5.1 LCOT–Interpolation: Real Data Application

Hue transfer experiment: In Figures 10, 11, and 12 we interpolate the hue channel between pairs of images using LCOT interpolation (given by equation 22), and COT interpolation (given by equation 21). Given a pair of images, one is considered as the *source* and the other as the *target*. Each image of $M \times N$ pixels is represented using Hue, Saturation, and Value channels (HSV). We compute a density-normalized histogram of the Hue values across all pixels and consider this histogram as a circular density. Each bin represents at the same time a color variety (Hue value) and a point (angle) in the circle. Thus, displacements in the circle correspond to color conversions. Each interpolation (LCOT / COT) provides a curve of color conversions parametrized between $t = 0$ (source) and $t = 1$ (target). For three different pairs of images, Figures 10, 11, and 12 depict color-converted images using steps $t = 0.25, 0.5, 0.75$ for each interpolation type.

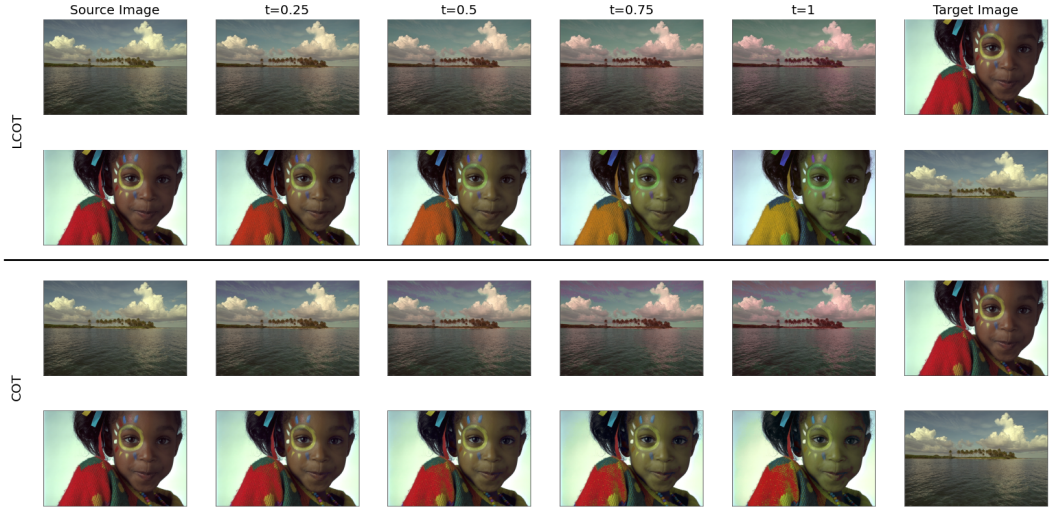

Figure 10: LCOT and COT color interpolations.

Figure 11: LCOT and COT color interpolations.

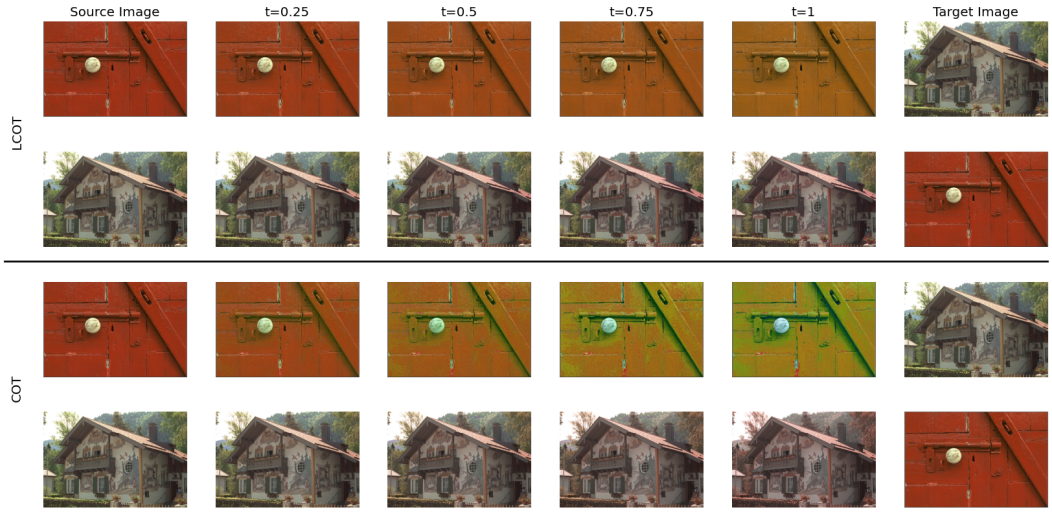

Figure 12: LCOT and COT color interpolations.

A.5.2 LCOT FOR HUE-BASED IMAGE RETRIEVAL

Given a data set of $N = 100$ flower images represented using hue, saturation, and value channels (HSV), we use the hue channel for image retrieval. For this, the hue information of an image is used as a primary feature for searching. We extract the hue component from the images of the data set and compute density-normalized histograms of the hue values across all pixels. We consider these histograms as circular densities $\{\nu_i\}_{i=1}^N$ (similarly to A.5.1).

For LCOT comparison and retrieval, we compute the LCOT transforms $\{\widehat{\nu}_i\}_{i=1}^N$. Given a new query image, we compute its hue histogram denoted σ. Then, we perform LCOT-matching, that is, we embed the input image in LCOT space by computing $\widehat{\sigma}$ so that we can calculate N squared Euclidean distances $\|\widehat{\sigma} - \widehat{\nu}_i\|_2^2$ (i.e., we are computing $LCOT_2(\sigma, \nu_i)$ for $i = 1, \ldots, N$). We sort the obtained LCOT-distance values in ascending order. In Figure 13 we show the four closest and furthest images recovered using this technique.

In Figure 14, we repeat the same experiment but using classic COT-matching for the same data set. As before, given a query image, we first compute its hue histogram denoted by σ. Then, we perform N COT-distances $COT_2(\sigma, \nu_i)$, for $i = 1, \ldots, N$, and We sort the obtained COT-distance values.

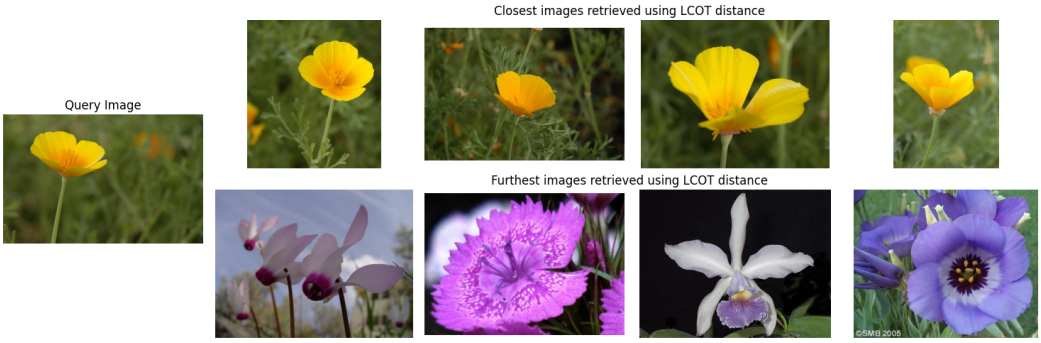

Figure 13: LCOT-approach for Hue-based image retrieval. The leftmost image is the original query image. In the upper row, we retrieve the 4 closest images in Hue space according to LCOT, while the bottom row shows the 4 furthest images with respect to LCOT-distance.

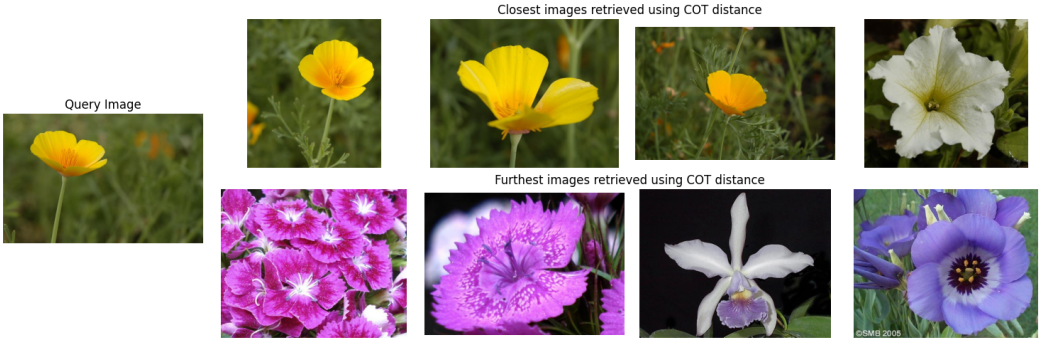

Figure 14: COT-distance for Hue-based image retrieval. The leftmost image is the original query image. In the upper row, we retrieve the 4 closest images in Hue space according to COT-distance, while the bottom row shows the 4 furthest images with respect to COT-distance.

In both figures, the retrieval of images with similar color content is evident. The advantage of using LCOT over COT is that when using COT each new image requires to solve N new circular optimal transport problems whereas LCOT only requires to solve one followed by N Euclidean distance calculations for comparison and sorting. For M queries we have to compute MN COT distances when using the COT approach (N for each query) but only solve $N + M$ COT problems when using LCOT (N for pre-processing the data set + one per query).

A.6 UNDERSTANDING THE EMBEDDING IN DIFFERENTIAL GEOMETRY

Our embedding $\nu \mapsto \widehat{\nu}$ as given by equation equation 17 aligns with the definition of the Logarithm function presented in (Sarrazin & Schmitzer, 2023, Definition 2.7). To be specific, for $\mu, \nu \in \mathbb{S}^1$ and the Monge mapping M_μ^ν, the Logarithm function as introduced in Sarrazin & Schmitzer (2023) is expressed as:

$$\mathcal{P}_2(\mathbb{S}^1) \ni \nu \mapsto \log_\mu^{COT}(\nu) \in L^2(\mathbb{S}^1, T\mathbb{S}^1; \mu).$$

Here, the tangent bundle of $\mathbb{S}^1$ is represented as

$$T\mathbb{S}^1 := \{(x, T_x(\mathbb{S}^1)) |\, x \in \mathbb{S}^1\},$$

where $T_x(\mathbb{S}^1)$ denotes the tangent space at the point $x \in \mathbb{S}^1$. The space $L^2(\mathbb{S}^1, T\mathbb{S}^1; \mu)$ is the set of vector fields on $\mathbb{S}^1$ with squared norms (based on the metric on $T\mathbb{S}^1$), that are μ-integrable. The function (vector field) $\log_\mu^{COT}(\nu)$ is defined as:

$$\log_\mu^{COT}(\nu) := (\mathbb{S}^1 \ni x \mapsto (x, v_x)) \in T\mathbb{S}^1,$$

where $v_x \mapsto T_x(\mathbb{S}^1)$ is the initial velocity of the unique constant speed geodesic curve $x \mapsto T_\mu^\nu(x)$.

The relation between $\log_\mu^{COT}(\nu)$ and $\widehat{\nu}$ in equation 19 can be established as follows: For any x in $\mathbb{S}^1$, the spaces $T_x(\mathbb{S}^1)$ and $\mathbb{S}^1$ can be parameterized by $\mathbb{R}$ and $[0, 1)$, respectively. Then, the unique constant speed curve $x \mapsto M_\mu^\nu(x)$ is given by:

$$x(t) := x + t(M_\mu^\nu(x) - x), \qquad \forall t \in [0, 1].$$

Then, the initial velocity is $M_\mu^\nu(x) - x$. Drawing from equation 15, Theorem 2.5, and Proposition A.2, we find $\widehat{\nu}(x) = M_\mu^\nu(x) - x$ for all x in $\mathbb{S}^1$, making $\widehat{\nu}$ and $\log_\mu^{COT}(\nu)$ equivalent.

However, it is important to note that while $\log_\mu^{COT}$ is defined for a generic (connected, compact, and complete[3]) manifold, it does not provide a concrete computational method for the embedding $\log_\mu^{COT}$. Our focus in this paper is on computational efficiency, delivering a closed-form formula.

Regarding the embedding space, in Sarrazin & Schmitzer (2023), the space $L^2(\mathbb{S}^1, T\mathbb{S}^1; \mu)$ is equipped with the L^2, induced by $T\mathbb{S}^1$. Explicitly, for any f belonging to $L^2(\mathbb{S}^1, T\mathbb{S}^1; \mu)$,

$$\|f\|^2 = \int_0^1 \|f(x)\|_x^2 dx = \int_0^1 |f(x)|^2 \, dx,$$

where $\|f(x)\|_x^2$ represents the norm square in the tangent space $T_x(\mathbb{S}^1)$ of the vector $f(x)$. By parameterizing $\mathbb{S}^1$ and $T_x(\mathbb{S}^1)$ as $[0, 1)$ and $\mathbb{R}$, respectively, this squared norm becomes $|f(x)|^2$. Consequently, $L^2(\mathbb{S}^1, T\mathbb{S}^1; \mu)$ becomes an inner product space, whereby the expression (polarization identity) $\|f + g\|^2 - \|f\|^2 - \|g\|^2$ establishes an inner product between f and g.

However, in this paper, the introduced embedding space $L^2(\mathbb{S}^1, d\mu)$ presented in equation 18. This space uses the L^2-norm on the circle, defined for each f in $L^2(\mathbb{S}^1, d\mu)$ as:

$$\|f\|_{L^2(\mathbb{S}^1; d\mu)}^2 = \int_{\mathbb{S}^1} |f(x)|_{\mathbb{S}^1}^2 \, d\mu.$$

Unlike the previous space, this does not induce an inner product (in fact, $|\cdot|_{\mathbb{S}^1}$ is not a norm). As such, throughout this paper, we term our embedding as a "linear embedding" rather than a "Euclidean embedding".

[3]In Sarrazin & Schmitzer (2023), the Riemannian manifold is not necessarily compact. However, the measures μ, ν must have compact support sets. For brevity, we have slightly overlooked this difference.