Riemannian Convex Potential Maps: Supplementary Material

A. Manifold Operations

We briefly describe manifold operations, on a Riemannian manifold $\mathcal{M}$ with metric g, used in this paper. Specifically, we define the exponential map $\exp$ and the intrinsic manifold distance $d_{\mathcal{M}}$.

Exponential map. Let $x \in \mathcal{M}$, $v \in T_x\mathcal{M}$ and consider the unique geodesic $\gamma : [0, 1] \to \mathcal{M}$ such that $\gamma(0) = x$ and $\gamma'(0) = v$. The exponential map at x, $\exp_x : T_x\mathcal{M} \to \mathcal{M}$, is defined as

$$\exp_x(v) = \gamma(1). \tag{26}$$

Intrinsic distance. Define the length of a curve $\gamma : [0, 1] \to \mathcal{M}$ as

$$L(\gamma) = \int_0^1 \|\gamma'(t)\|_g \, dt, \tag{27}$$

where $\|\gamma'(t)\|_g$ means taking the norm of the velocity $\gamma'(t)$ at $T_{\gamma(t)}\mathcal{M}$ with respect to the metric g of the manifold $\mathcal{M}$. Then, the intrinsic distance $d_{\mathcal{M}}$ between $x, y \in \mathcal{M}$ is:

$$d_{\mathcal{M}}(x, y) = \inf_{\gamma} L(\gamma) \tag{28}$$

where the $\inf$ is over curves $\gamma : [0, 1] \to \mathcal{M}$ where $\gamma(0) = x$ and $\gamma(1) = y$. If $\mathcal{M}$ is *complete* (see *e.g.*, Hopf-Rinow Theorem) the intrinsic distance is realized by a geodesic.

Sphere. On the n-sphere $\mathcal{S}^n$, the exponential map and the intrinsic distance are provided as closed-form expressions. If $x, y \in \mathcal{S}^n$ and $v \in T_x\mathcal{S}^n$,

$$\exp_x(v) = x \cos(\|v\|) + \frac{v}{\|v\|} \sin(\|v\|) \tag{29}$$

$$d_{\mathcal{S}^n}(x, y) = \arccos(x^T y), \tag{30}$$

where $\|\cdot\|$ is the standard Euclidean norm.

Product manifolds. We now consider operations on product manifolds of the form $\mathcal{M} = \mathcal{M}_1 \times \ldots \times \mathcal{M}_l$. The squared intrinsic distance is simply

$$d^2_{\mathcal{M}}(x, y) = d^2_{\mathcal{M}_1}(x_1, y_1) + \ldots + d^2_{\mathcal{M}_l}(x_l, y_l). \tag{31}$$

Here $x = (x_1, \ldots, x_l)$, and $x_j \in \mathcal{M}_j, \;\; j \in [l]$ (and similarly for y). The exponential map on the product manifold is the cartesian product of exponential maps on the individual manifolds. An instantiation of such product that will be considered in experiments is the torus $\mathcal{S}^1 \times \mathcal{S}^1$. In that case, we can use eqs. (29) and (30) to get the exponential map and squared intrinsic distance in closed-form.

B. Proof of c-concavity of the multi-layer potential

Proof. The proof is by induction. Constant functions are c-concave, hence ψ_0 is c-concave. Also, $\psi_1 = (1 - w_0)\phi_0$ is c-concave by the assumption of convexity of the space of c-concave functions. Next, assuming $\psi_{k-1}(x)$ is c-concave, $\sigma(\psi_{k-1})$ is also c-concave (because σ preserves c-concavity), and $\psi_k(x)$ is c-concave because convex combinations of c-concave functions are c-concave. In conclusion, $\varphi = \psi_K$ is c-concave □

C. Additional experimental and implementation details

C.1. Synthetic Sphere

We conducted a hyper-parameter search over the parameters in table 3 to find the flows used in our demonstrations and experiments. We report results from the best hyper-parameters obtained by randomly sampling the space of parameters. The α values are initialized from $\mathcal{U}[\alpha_{\min}, \alpha_{\min} + \alpha_{\text{range}}]$. Also, γ_1 corresponds to the softing coefficient of the soft-min operation of discrete c-concave potentials, and γ_2 to the softing coefficient of the soft-min operation in the identity initialization (see sects. 5.1 and 5.3).

Table 3. Hyper-parameter sweep for our sphere results

Adam	
learning rate	$[10^{-6}, 10^{-1}]$
β_1	[0.1, 0.3, 0.5, 0.7, 0.9]
β_2	[0.1, 0.3, 0.5, 0.7, 0.9, 0.99, 0.999]
Flow Hyper-parameters	
Nb. of Components y_i	[50, 1000]
$\alpha_{\min}$	$[10^{-5}, 10]$
α_{range}	$[10^{-3}, 1]$
γ_1	[0.01, 0.05, 0.1, 0.5]
γ_2	[None, 0.01, 0.05, 0.1, 0.5]

We now verify empirically whether the RPCM define diffeomorphisms in practice. We compute Jacobian log-determinants of the flow trained on the 4-modal density taken from Rezende et al. (2020) for 10^6 points uniformly

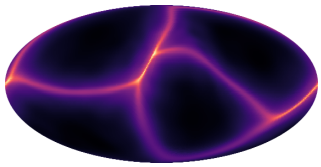

Figure 8. Jacobian log-determinants for points uniformly sampled on the sphere.

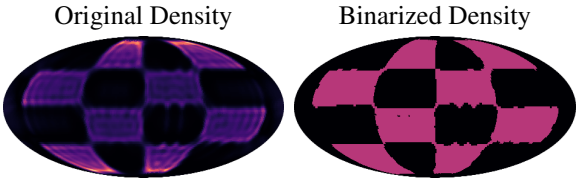

Figure 9. Binarized density of the sphere checkerboard

sampled on the sphere, and observe that all these are positive (see fig. 8).

Binarized checkerboard density. We found it difficult to visualize the learned density of our model on the checkerboard because a few regions have unusually high values that mess up the ranges of the colormap. For visualization purposes, we binarize the density values by taking the portion of the density greater than the uniform density. Figure 9 shows the original and binarized densities of our models.

C.2. Torus

Model. We provide details on the model used in the torus demonstration. The RCPM is composed of 6 single-layer blocks of 200 components, and the softing parameter is set to 0.5. Adam's learning rate is set to $6e^{-4}$ and β to $(0.9, 0.99)$.

Data. The target density is of a form inspired by the target densities in Rezende et al. (2020)):

$$p(\theta_1, \theta_2) = \frac{1}{3} \sum_{i=1}^{3} p_i(\theta_1, \theta_2) \tag{32}$$

$$p_i(\theta_1, \theta_2) \propto \exp\left[\cos(\theta_1 - a_i^1) + \cos(\theta_2 - a_i^2)\right] \tag{33}$$

where $a_1 = [4.18, 6.7], a_2 = [4.18, 4.7], a_3 = [4.18, 2.7]$, and $\theta_1, \theta_2 \in [0, 2\pi]$.

C.3. Continental Drift

Mapping estimation. We continue with details on the model used in the mapping estimation setting of the continental drift case study. The RCPM is composed of 7 blocks containing each 3 layers with 200 components, and the softing coefficient is set to 0.2. Adam's learning rate is set to $2e^{-3}$ and $\beta = (0.9, 0.99)$.

Transport geodesics. We now discuss the transport geodesics setting. The RCPM is composed of a single block (hence allowing to recover the optimal transport geodesics) containing 3 layers with 200 components, and the softing coefficient is set to $\gamma = 0.2$. Adam's learning rate is set to $2e^{-3}$ and $\beta = (0.9, 0.99)$.

Density estimation. Finally, we provide details on the model used in the density estimation setting. The RCPM is composed of 6 single-layer blocks containing each 400 components, and the softing coefficient is set to $6e^{-2}$. Adam's learning rate is set to $2e^{-3}$ and $\beta = (0.9, 0.99)$.

Data. The earth densities are obtained by leveraging the code from `https://github.com/cgarciae/point-cloud-mnist-2D` to turn Mollweide earth images into spherical point clouds, converting to Euclidean coordinates, and applying kernel density estimation to such point clouds both for visualization, and to get log-probabilities when they are required (e.g., in the mapping estimation setting, where access to log-probabilities from the base – old earth – is needed to train the model).