1. Supplementary Materials

1.1. Proofs

Here we include the proof for the C1 and C2 conditions covered in Section 3. Recall that μ_i represents a probability measure, μ_i^θ represents $g_{\theta\#}\mu_i$, where $g_\theta : \mathbb{R}^d \to \mathbb{R}$ with some regularity constraints, and that we define the cumulative distribution transform (CDT) of μ_i^θ as

$$\phi_\theta(\mu_i) \coloneqq (T_i^\theta - id),$$

where T_i^θ is the Monge map/coupling, and id denotes the identity function. For a fixed θ, here we prove that $\phi^\theta(\mu_i)$ satisfies the following conditions:

C1. The weighted ℓ_2-norm of the embedded slice, $\phi_\theta(\mu_i)$, satisfies:

$$\begin{aligned} \|\phi_\theta(\mu_i)\|_{\mu_0^\theta,2} &= \left(\int_{\mathbb{R}} \|\phi_\theta(\mu_i(t))\|_2^2 d\mu_0^\theta(t)\right)^{\frac{1}{2}} \\ &= \mathcal{W}_2(\mu_i^\theta, \mu_0^\theta), \end{aligned}$$

Proof. We start by writing the squared distance:

$$\begin{aligned} &\|\phi_\theta(\mu_i)\|_{\mu_0^\theta,2}^2 = \int_{\mathbb{R}} \|\phi_\theta(\mu_i(t))\|_2^2 d\mu_0^\theta(t) \\ &= \int_{\mathbb{R}} \|T_i^\theta(t) - t\|_2^2 d\mu_0^\theta(t) \\ &= \int_{\mathbb{R}} \|(F_{\mu_i^\theta}^{-1} \circ F_{\mu_0^\theta})(t) - t\|_2^2 d\mu_0^\theta(t) \\ &= \int_0^1 \|F_{\mu_i^\theta}^{-1}(\tau) - F_{\mu_0^\theta}^{-1}(\tau)\|_2^2 d\tau \\ &= \mathcal{W}_2^2(\mu_i^\theta, \mu_0^\theta) \end{aligned}$$

where we used the definition of the one-dimensional Monge map, $T_i^\theta = F_{\mu_i^\theta}^{-1} \circ F_{\mu_0^\theta}$, and the change of variable $\tau = F_{\mu_0^\theta}(t)$. The corollary, $\|\phi_\theta(\mu_0)\|_{\mu_0^\theta,2} = 0$, is trivial as $\mathcal{W}_2^2(\mu_0^\theta, \mu_0^\theta) = 0$. □

C2. The weighted ℓ_2 distance satisfies:

$$\|\phi_\theta(\mu_i) - \phi_\theta(\mu_j)\|_{\mu_0^\theta,2} = \mathcal{W}_2(\mu_i^\theta, \mu_j^\theta). \quad (1)$$

Proof. Similar to the previous proof:

$$\begin{aligned} &\|\phi_\theta(\mu_i) - \phi_\theta(\mu_j)\|_{\mu_0^\theta,2}^2 = \int_{\mathbb{R}} \|\phi_\theta(\mu_i(t)) - \phi_\theta(\mu_j(t))\|_2^2 d\mu_0^\theta(t) \\ &= \int_{\mathbb{R}} \|T_i^\theta(t) - T_j^\theta(t)\|_2^2 d\mu_0^\theta(t) \\ &= \int_{\mathbb{R}} \|(F_{\mu_i^\theta}^{-1} \circ F_{\mu_0^\theta})(t) - (F_{\mu_j^\theta}^{-1} \circ F_{\mu_0^\theta})(t)\|_2^2 d\mu_0^\theta(t) \\ &= \int_0^1 \|F_{\mu_i^\theta}^{-1}(\tau) - F_{\mu_j^\theta}^{-1}(\tau)\|_2^2 d\tau = \mathcal{W}_2^2(\mu_i^\theta, \mu_j^\theta) \end{aligned}$$

where again we used the definition of the one-dimensional Monge map, $T_i^\theta = F_{\mu_i^\theta}^{-1} \circ F_{\mu_0^\theta}$, and the change of variable $\tau = F_{\mu_0^\theta}(t)$. □

As a corollary of C1 and C2 we have:

$$\begin{aligned} &\mathcal{SW}_2^2(\mu_i, \mu_j) = \int_{\mathbb{S}^{d-1}} \mathcal{W}_2^2(\mu_i^\theta, \mu_j^\theta) d\sigma(\theta) = \\ &\int_{\mathbb{S}^{d-1}} \|\phi_\theta(\mu_i) - \phi_\theta(\mu_j)\|_{\mu_0^\theta,2}^2 d\sigma(\theta) \end{aligned}$$

1.2. Full Experimental Results

In Table 1 and 2, we provide the full experimental results with standard deviation. GeM and Cov do not have a std as their outputs are deterministic.

Table 1. Retrieval results of baselines and our approach on set retrieval on the three data sets (precision@k).

	Point MNIST 2D			ModelNet 40			Oxford 5k		
	Precision@k			Precision@k			Precision@k		
	k=4	k=8	k=16	k=4	k=8	k=16	k=4	k=8	k=16
GeM-1 (GAP)	0.10	0.10	0.10	0.14	0.14	0.14	0.29	0.25	0.22
GeM-2	0.29	0.29	0.29	0.30	0.28	0.25	0.38	0.31	0.27
GeM-4	0.39	0.39	0.38	0.34	0.31	0.28	0.09	0.09	0.09
Cov	0.25	0.25	0.25	0.45	0.43	0.41	0.35	0.30	0.26
FSPool	0.75±0.00	0.74±0.00	0.72±0.00	0.51±0.00	0.48±0.00	0.44±0.00	0.43±0.03	0.36±0.02	0.36±0.01
WE	**0.89±0.00**	**0.88±0.00**	**0.86±0.00**	**0.71±0.00**	**0.68±0.00**	**0.64±0.00**	**0.54±0.04**	**0.47±0.02**	**0.39±0.02**
SLOSH ($L = d$)	0.78±0.08	0.76±0.08	0.74±0.08	0.40±0.01	0.38±0.00	0.35±0.00	0.43±0.03	0.36±0.03	0.30±0.03
SLOSH ($L > d$)	**0.89±0.01**	**0.88±0.01**	**0.86±0.01**	**0.63±0.00**	**0.59±0.00**	**0.55±0.00**	**0.52±0.02**	**0.45±0.01**	**0.37±0.01**

Table 2. Retrieval results of baselines and our approach on set retrieval on the three data sets (accuracy).

	Point MNIST 2D			ModelNet 40			Oxford 5k		
	Accuracy			Accuracy			Accuracy		
	k=4	k=8	k=16	k=4	k=8	k=16	k=4	k=8	k=16
GeM-1 (GAP)	0.11	0.10	0.10	0.17	0.19	0.21	0.35	0.31	0.29
GeM-2	0.32	0.35	0.37	0.34	0.36	0.37	0.53	0.40	0.38
GeM-4	0.45	0.47	0.49	0.39	0.39	0.39	0.09	0.09	0.09
Cov	0.26	0.28	0.28	0.51	0.52	0.52	0.55	0.37	0.33
FSPool	0.80±0.00	0.81±0.01	0.81±0.01	0.58±0.00	0.58±0.01	0.58±0.01	0.50±0.04	0.53±0.06	0.44±0.05
WE	**0.92±0.00**	**0.92±0.00**	**0.92±0.00**	**0.76±0.01**	**0.76±0.00**	**0.75±0.01**	**0.70±0.08**	**0.68±0.04**	**0.61±0.05**
SLOSH ($L = d$)	0.82±0.06	0.83±0.06	0.82±0.06	0.46±0.01	0.47±0.01	0.46±0.00	0.54±0.04	0.53±0.07	0.45±0.07
SLOSH ($L > d$)	**0.92±0.01**	**0.92±0.01**	**0.92±0.01**	**0.68±0.01**	**0.67±0.00**	**0.65±0.00**	**0.69±0.03**	**0.67±0.03**	**0.61±0.03**