SPOTLIGHT

ICLR

Towards Marginal Fairness Sliced Wasserstein Barycenter

Khai Nguyen*[1], Hai Nguyen*[2] , Nhat Ho[1]

[1]Department of Statistics and Data Sciences, University of Texas at Austin
[2]Qualcomm AI Research

TEXAS
The University of Texas at Austin

Qualcomm

*: Equal Contributions

Sliced Wasserstein Distance

Let μ, ν be two probability measures that has supports in $\mathbb{R}^d$, the sliced Wasserstein distance of order $p \geq 1$:

$$SW_p(\mu,\nu) = \left(\mathbb{E}_{\theta\sim\mathcal{U}(S^{d-1})}\left[W_p^p(\theta\sharp\mu, \theta\sharp\nu)\right]\right)^{\frac{1}{p}}$$

where $\mathbb{S}^{d-1}$ is the unit-hypersphere, the Wasserstein distance on one-dimension has the closed-form:

$$\mathbb{W}_p(\mu,\nu) = \left(\int_0^1 |F_\mu^{-1}(z) - F_\nu^{-1}(z)|^p \, dz\right)^{\frac{1}{p}}$$

and $\theta\sharp\mu$ denotes the pushforward probability measure of μ through the function $T_\theta : \mathbb{R}^d \to \mathbb{R}$ with $T_\theta(x) = \theta^\top x$, and F^{-1} denotes the inverse CDF.

Sliced Wasserstein Barycenter

For $K \geq 2$ marginals $\mu_1, \dots, \mu_K \in \mathcal{P}_p(\mathbb{R}^d)$

$$\min_\mu \mathcal{F}(\mu) = \sum_{k=1}^{K} \omega_k \, SW_p^p(\mu, \mu_k)$$

uniform case: $\omega_1 = \cdots = \omega_K = \frac{1}{K}$

Let μ_ϕ be parameterized, then estimate gradient of barycenter by:

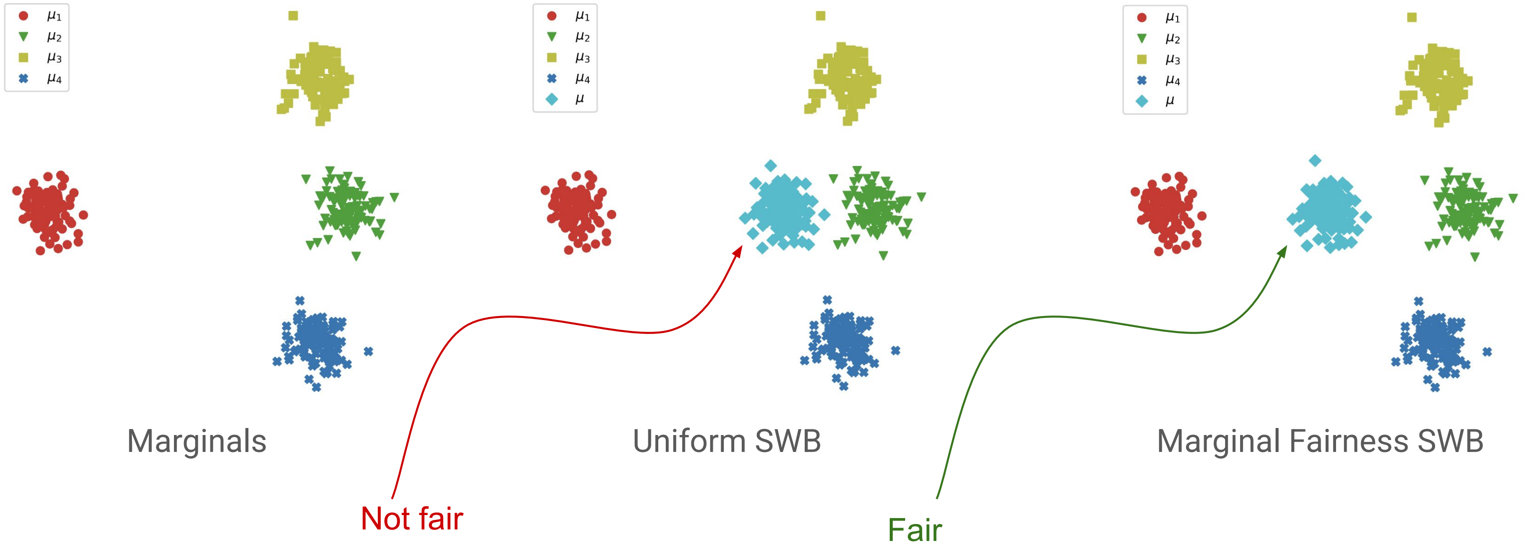
$$\nabla_\phi \mathbf{SW}_p^p(\mu_\phi, \mu_k) = \nabla_\phi \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^{d-1})} \left[\mathbf{W}_p^p(\theta \sharp \mu_\phi, \theta \sharp \mu_k)\right] = \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^{d-1})} \left[\nabla_\phi \mathbf{W}_p^p(\theta \sharp \mu_\phi, \theta \sharp \mu_k)\right]$$

Monte Carlo estimator:

$$\nabla_\phi SW_p^p(\mu_\phi, \mu_k) \approx \frac{1}{L} \sum_{l=1}^{L} \nabla_\phi W_p^p(\theta_l \sharp \mu_\phi, \theta_l \sharp \mu_k)$$

Motivations

Objective: Find a barycenter that minimizes the distances to marginals while having equal distances to marginals at the same time.



Marginal Fairness Sliced Wasserstein Barycenter

For $K \geq 2$ marginals $\mu_1, \ldots, \mu_K \in \mathcal{P}_p(\mathbb{R}^d)$ and admissible $\epsilon \geq 0$

$$\min_\mu \frac{1}{K} \sum_{k=1}^K SW_p^p(\mu, \mu_k)$$

$$\text{s.t.} \quad \frac{2}{(K-1)K} \sum_{i=1}^{K-1} \sum_{j=i+1}^K |SW_p^p(\mu, \mu_i) - SW_p^p(\mu, \mu_j)| \leq \epsilon.$$

➡ Lagrange Multiplier: *Marginal Fairness Sliced Wasserstein barycenter (MFSWB)*

$$\mathcal{L}(\mu, \lambda) = \frac{1}{K} \sum_{k=1}^K \text{SW}_p^p(\mu, \mu_k) + \frac{2\lambda}{(K-1)K} \sum_{i=1}^{K-1} \sum_{j=i+1}^K |\text{SW}_p^p(\mu, \mu_i) - \text{SW}_p^p(\mu, \mu_j)| - \lambda\epsilon$$

Marginal Fairness Sliced Wasserstein Barycenter

surrogate Marginal Fairness Sliced Wasserstein Barycenter (s-MFSWB)

$$\min_\mu \mathcal{SF}(\mu;\mu_{1:K}); \quad \mathcal{SF}(\mu;\mu_{1:K}) = \max_{k\in\{1,\ldots,K\}} SW_p^p(\mu,\mu_k).$$

unbiased surrogate Marginal Fairness Sliced Wasserstein Barycenter (us-MFSWB)

$$\min_\mu \mathcal{USF}(\mu;\mu_{1:K}); \quad \mathcal{USF}(\mu;\mu_{1:K}) = \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\max_{k\in\{1,\ldots,K\}} W_p^p(\theta_H^{\mu_k},\theta_H^{\mu_{kk}})\right].$$

Marginal Fairness Sliced Wasserstein Barycenter

energy-based surrogate Marginal Fairness Sliced Wasserstein Barycenter (es-MFSWB)

$$\min_\mu \mathcal{ESF}(\mu; \mu_{1:K}); \quad \mathcal{ESF}(\mu; \mu_{1:K}) = \mathbb{E}_{\theta \sim \sigma(\theta;\mu,\mu_{1:K})} \left[\max_{k\in\{1,\ldots,K\}} W_p^p(\theta\sharp\mu, \theta\sharp\mu_k)\right].$$

Energy-based Slicing distribution instead of Uniform distribution

$$f_\sigma(\theta; \mu, \mu_{1:K}) \propto \exp\left(\max_{k\in\{1,\ldots,K\}} W_p^p(\theta\sharp\mu, \theta\sharp\mu_k)\right).$$

Experiments

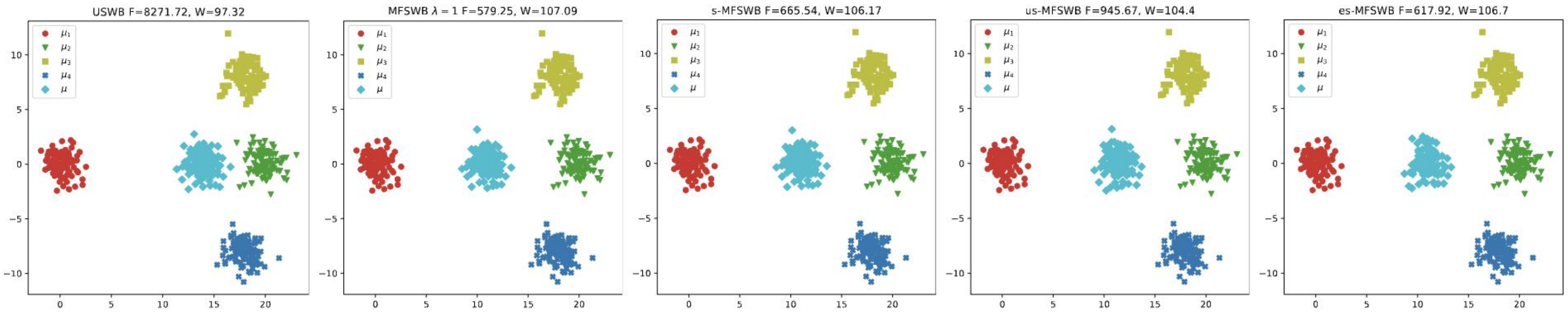
F metric: represents the marginal fairness degree of the barycenter

$$F = \frac{2}{K(K-1)} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} |W_p^p(\mu, \mu_i) - W_p^p(\mu, \mu_j)|$$

W metric: represents the centrality of the barycenter

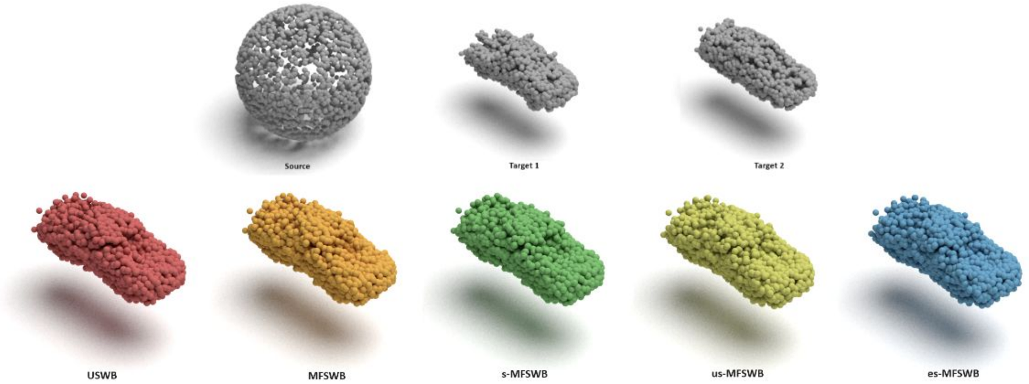
$$W = \frac{1}{K} \sum_{i=1}^{K} W_p^p(\mu, \mu_i)$$

Experiments: Toy example barycenter of Gaussian

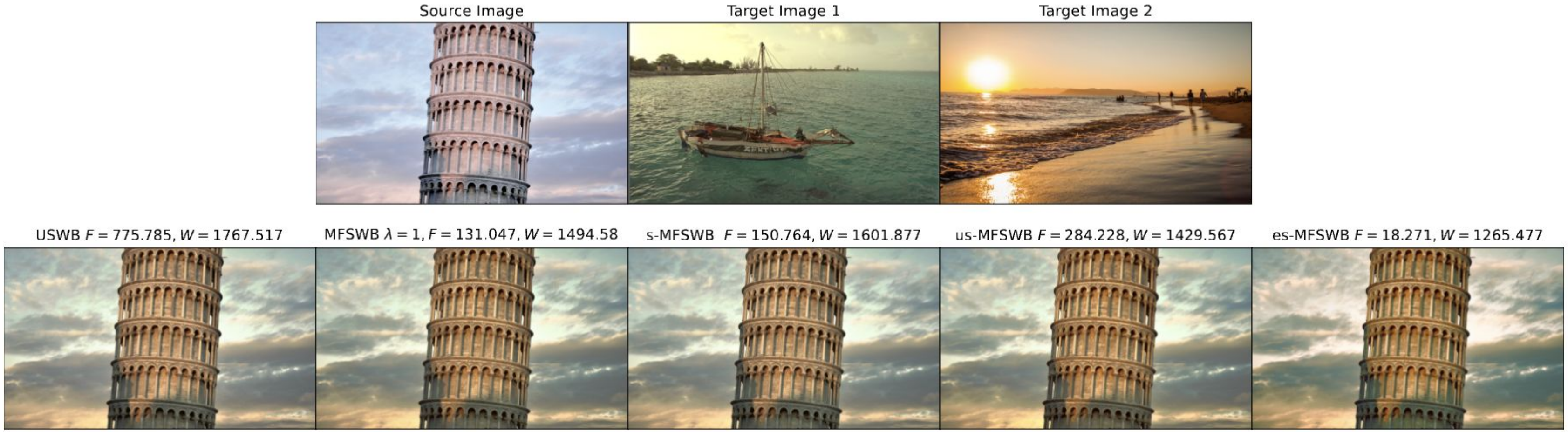


Experiments: 3D Point Cloud Averaging

Method	Iteration 0		Iteration 1000		Iteration 5000		Iteration 10000	
	F (↓)	W (↓)	F (↓)	W (↓)	F (↓)	W (↓)	F (↓)	W (↓)
USWB	252.24 ± 0.0	3746.05 ± 0.0	4.89 ± 0.28	85.72 ± 0.18	3.79 ± 0.32	45.37 ± 0.18	1.55 ± 0.48	39.81 ± 0.18
MFSWB $\lambda = 0.1$	252.24 ± 0.0	3746.05 ± 0.0	4.76 ± 0.27	84.86 ± 0.17	3.78 ± 0.2	45.2 ± 0.11	1.32 ± 0.22	39.73 ± 0.16
MFSWB $\lambda = 1$	252.24 ± 0.0	3746.05 ± 0.0	0.49 ± 0.2	79.08 ± 0.15	3.64 ± 0.26	44.71 ± 0.19	1.03 ± 0.06	39.45 ± 0.18
MFSWB $\lambda = 10$	252.24 ± 0.0	3746.05 ± 0.0	4.03 ± 2.43	$\mathbf{71.24 \pm 0.9}$	7.32 ± 2.5	45.21 ± 0.2	4.13 ± 2.48	42.56 ± 0.36
s-MFSWB	252.24 ± 0.0	3746.05 ± 0.0	2.52 ± 0.77	81.84 ± 0.14	4.01 ± 0.38	44.9 ± 0.13	1.15 ± 0.09	39.58 ± 0.17
us-MFSWB	252.24 ± 0.0	3746.05 ± 0.0	0.3 ± 0.18	78.69 ± 0.17	3.74 ± 0.26	44.38 ± 0.1	0.87 ± 0.18	39.26 ± 0.1
es-MFSWB	252.24 ± 0.0	3746.05 ± 0.0	$\mathbf{0.2 \pm 0.19}$	78.1 ± 0.16	$\mathbf{3.5 \pm 0.29}$	$\mathbf{44.37 \pm 0.08}$	$\mathbf{0.84 \pm 0.22}$	$\mathbf{39.18 \pm 0.08}$



Experiments: Color Harmonization



Experiments: Sliced Wasserstein Autoencoder

SW Autoencoder objective

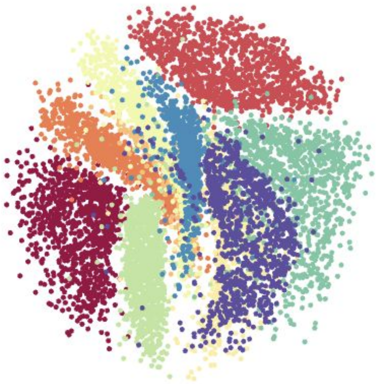
New added objective

$$\min_{\phi,\psi} \mathbb{E}\left[\frac{1}{KM}\sum_{k=1}^{K}\sum_{i=1}^{M} c(X_{ki}, g_\psi(f_\phi(X_{ki}))) + \kappa_1 SW_p^p\left(P_Z, P_{(f_\phi(X_k))_{k=1}^K}\right)\right.$$

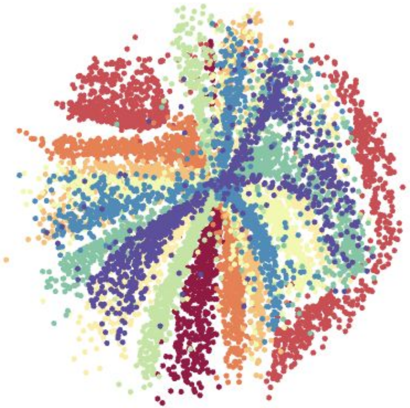
$$\left. + \kappa_2 \mathcal{B}\left(P_Z; P_{f_\phi(X_1)} : P_{f_\phi(X_K)}\right)\right]$$

Methods	RL (↓)	$W^2_{2,\text{latent}} \times 10^2$ (↓)	$W^2_{2,\text{image}} \times 10^2$ (↓)	$F \times 10^2$ (↓)	$W \times 10^2$ (↓)	F_{images} (↓)
SWAE	3.002	9.949	**26.572**	17.661	28.512	7.787
USWB	3.195	9.174	27.446	5.190	12.448	7.140
MFSWB $\lambda = 0.1$	**2.812**	8.981	26.636	17.206	28.734	7.846
MFSWB $\lambda = 1.0$	2.883	7.978	26.355	18.069	29.701	7.367
MFSWB $\lambda = 10.0$	3.801	8.497	26.658	18.501	28.768	7.950
s-MFSWB	3.170	**7.806**	28.277	2.037	8.699	7.419
us-MFSWB	2.833	8.720	27.939	2.072	**7.780**	**6.898**
es-MFSWB	3.056	9.154	28.012	**1.760**	7.268	7.485

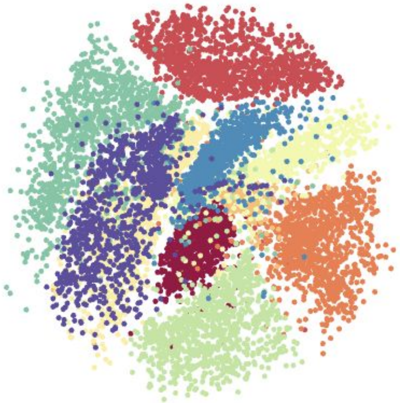
Experiments: Sliced Wasserstein Autoencoder



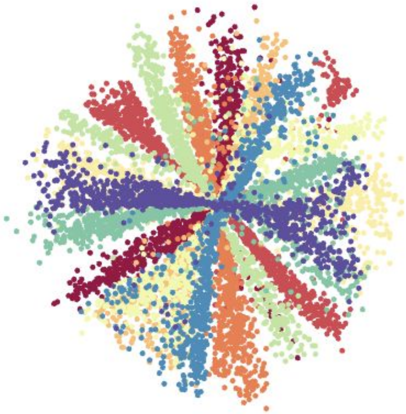
Vanilla version



use USWB (Uniform)



use MFSWB (Lagrange)



use es-MFSWB

Please check out the paper

Paper: https://openreview.net/pdf?id=NQqJPPCesd

Thank you for listening!

Hai Nguyen: namhai283287@gmail.com hainn@qti.qualcomm.com https://hainn2803.github.io/

currently seeking PhD opportunities. If interested, please feel free to reach out via my personal email 😄