Improving Relational Regularized Autoencoders with Spherical Sliced Fused Gromov Wasserstein

Khai Nguyen[1], Son Nguyen[1], Nhat Ho[2], Tung Pham[1], Hung Bui[1]

[1]VinAI Research, Vietnam [2]University of Texas, Austin

Deterministic Relational Regularized Autoencoders

Relational regularized autoencoders (RAEs) is a special case of WAEs to overcome the under-regularization problem by using **fused Gromov Wasserstein** as the regularization to combine both direct comparison and relational comparison.

$$\min_{\theta,\phi,\gamma} \mathbb{E}_{p_x} \mathbb{E}_{q_\phi(z|x)} [d(x, G_\theta(z))] + \lambda D_{fgw}(q_\phi(z), p_\gamma(z))$$

Deterministic relational regularized autoencoder (DRAE) is a variant of RAEs that achieves the state-of-the-art generative quality and has a fast computational time.

Replace FGW by **sliced fused Gromov Wasserstein** (SFG):

$$SFG(\mu, \nu; \beta) := \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^{q-1})} [D_{fgw}(\theta \sharp \mu, \theta \sharp \nu; \beta)]$$

- $\mu, \nu \in \mathcal{P}(\mathbb{R}^q)$ and $\mathcal{U}(\mathbb{S}^{q-1})$ is the uniform distribution on the hypersphere of q dimension
- The expectation is approximated by Monte Carlo scheme with L samples (projections)
- When μ, ν are empirical distributions, $D_{fgw}(\theta \sharp \mu, \theta \sharp \nu; \beta)$ can be computed efficiently by **sorting** the projected supports.

Spherical Sliced Fused Gromov Wasserstein

We introduce **spherical sliced fused Gromov Wasserstein** (SSFG), a new discrepancy for the relational regularization.

$$SSFG(\mu, \nu; \beta, \kappa) := \max_{\epsilon \in \mathbb{S}^{q-1}} \mathbb{E}_{\theta \sim \text{vMF}(.|\epsilon,\kappa)} [D_{fgw}(\theta \sharp \mu, \theta \sharp \nu; \beta)]$$

- $\text{vMF}(.|\epsilon, \kappa)$ is the von-Mises Fisher distribution with the location parameter ϵ and the concentration parameter κ

SSFG finds the **best** von-Mises Fisher distribution that can **maximize** the expected 1-d FGW

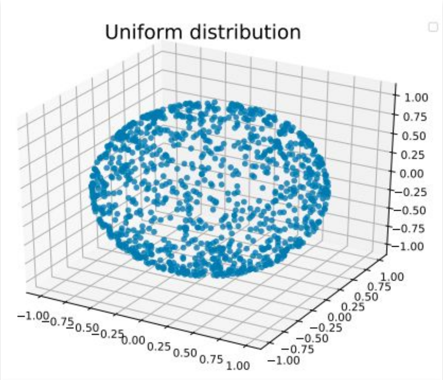
- The optimization can be solved by stochastic gradient ascent with the **reparameterization trick** and **sampling procedure** of the vMF the distribution.
- SSFG is a **pseudo distance** between two distributions since it satisfies non-negativity, symmetry, and the weak triangle inequality.

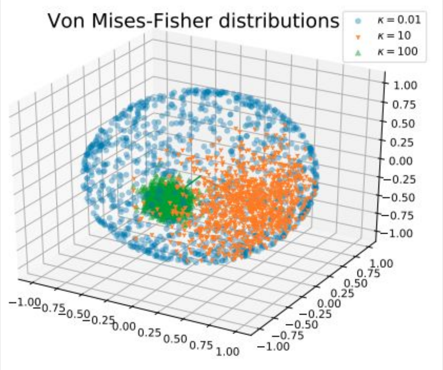
Spherical Sliced Fused Gromov Wasserstein

- SSFG is the **generalization** of SFG due to the interpolation property of the vMF distribution.
  - $\lim_{\kappa \to 0} SSFG(\mu, \nu; \beta, \kappa) = SFG(\mu, \nu; \beta)$
  - $\lim_{\kappa \to \infty} SSFG(\mu, \nu; \beta, \kappa) = \text{max-}SFG(\mu, \nu; \beta)$

$$\text{max-}SFG(\mu, \nu; \beta) := \max_{\theta \in \mathbb{S}^{d-1}} D_{fgw}(\theta\sharp\mu, \theta\sharp\nu; \beta)$$

➡ SSFG is a **interpolation** between SFG and max-SFG



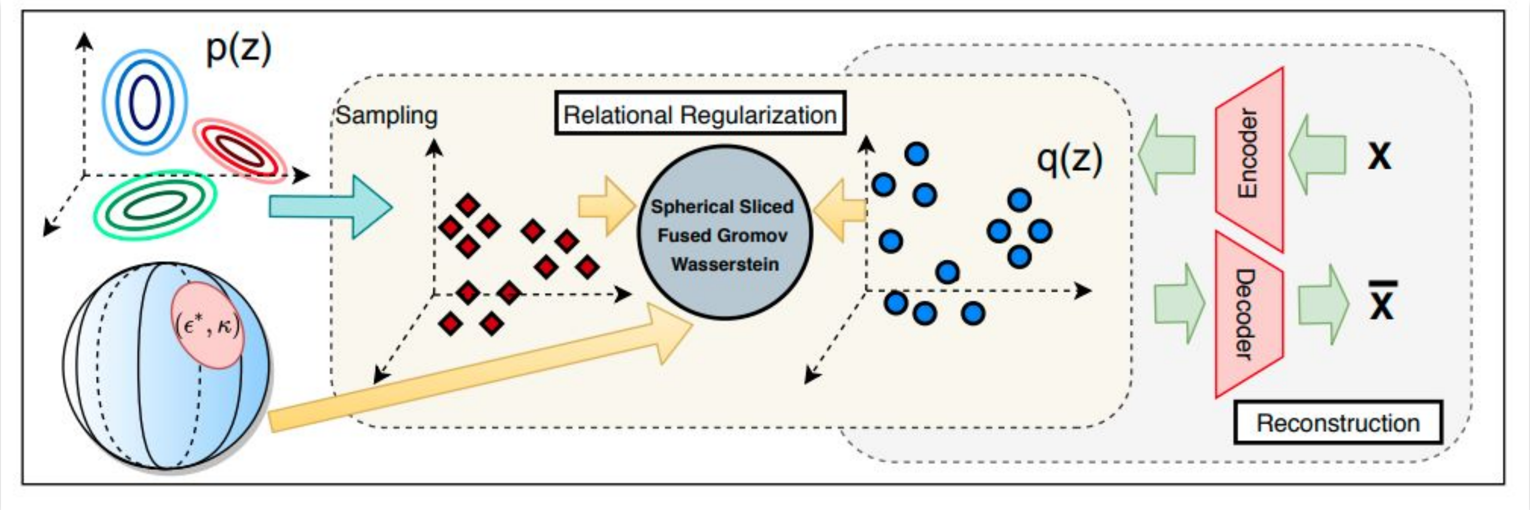


- We also have following inequality for any $\kappa > 0$
  - $\exp(-\kappa)C_q(\kappa)\text{SFG}(\mu, \nu; \beta) \leq \text{SSFG}(\mu, \nu; \beta, \kappa) \leq \exp(\kappa)C_q(\kappa)\text{SFG}(\mu, \nu; \beta)$
  - $\text{SSFG}(\mu, \nu; \beta, \kappa) \leq \text{max-SFG}(\mu, \nu; \beta)$

- SSFG does not suffer from **the curse of dimensionality** for the inference purposes

Spherical Deterministic Relational Regularized Autoencoder

By using SSFG for the regularization in WAEs, we obtain a new relational regularized autoencoder : **spherical deterministic relational regularized autoencoder** (s-DRAE)



➡ s-DRAE is the **generalization** of DRAE and m-DRAE (uses max-SFG)

Power Spherical Sliced Fused Gromov Wasserstein

We introduce **power spherical sliced fused Gromov Wasserstein** (PSSFG), a new discrepancy that has the same property as SSFG but has faster computational time.

$$PSSFG(\mu, \nu; \beta, \kappa) := \max_{\epsilon \in \mathbb{S}^{q-1}} \mathbb{E}_{\theta \sim \text{PS}(.|\epsilon, \kappa)} [D_{fgw}(\theta \sharp \mu, \theta \sharp \nu; \beta)]$$

- $\text{PS}(.|\epsilon, \kappa)$ is the power spherical distribution with the location parameter ϵ and the concentration parameter κ

- PSSFG is **faster** than SSFG since the power spherical does not need rejection sampling algorithm to sample from like the vMF distribution (also lead to more stable sampling).
- PSSFG inherits **all** properties of SSFG such as metricity, interpolation, no curse of dimensionality
- Using PSSFG in WAEs creates **power spherical deterministic relational regularized autoencoder** (ps-DRAE)

Mixture variants of SSFG and PSSFG

Using the **mixture** of von-Mises Fisher (power spherical) distribution can lead to following variants

- **Mixture spherical sliced fused Gromov Wasserstein** (MSSFG)

$$MSSFG(\mu, \nu; \beta, \{\kappa\}_{i=1}^k, \{\alpha\}_{i=1}^k) := \max_{\{\epsilon\}_{i=1}^k \in \mathbb{S}^{q-1}} \mathbb{E}_{\theta \sim \text{MovMF}(.|\{\epsilon\}_{i=1}^k, \{\kappa\}_{i=1}^k, \{\alpha\}_{i=1}^k)} [D_{fgw}(\theta \sharp \mu, \theta \sharp \nu; \beta)]$$

- $\text{MovMF}(.|\{\epsilon\}_{i=1}^k, \{\kappa\}_{i=1}^k, \{\alpha\}_{i=1}^k) := \sum_{i=1}^k \alpha_i \text{vMF}(.| \epsilon_i, \kappa_i)$

- **Mixture power spherical sliced fused Gromov Wasserstein** (MPSSFG)

$$MPSSFG(\mu, \nu; \beta, \{\kappa\}_{i=1}^k, \{\alpha\}_{i=1}^k) := \max_{\{\epsilon\}_{i=1}^k \in \mathbb{S}^{q-1}} \mathbb{E}_{\theta \sim \text{MoPS}(.|\{\epsilon\}_{i=1}^k, \{\kappa\}_{i=1}^k, \{\alpha\}_{i=1}^k)} [D_{fgw}(\theta \sharp \mu, \theta \sharp \nu; \beta)]$$

- $\text{MoPS}(.|\{\epsilon\}_{i=1}^k, \{\kappa\}_{i=1}^k, \{\alpha\}_{i=1}^k) := \sum_{i=1}^k \alpha_i \text{PS}(.| \epsilon_i, \kappa_i)$

- The RAEs versions of MSSFG and MPSSFG are **mixture spherical DRAE** (ms-DRAE) and and **mixture power spherical DRAE** (mps-DRAE) respectively

Experiments : Image generation and reconstruction

Table: **Comparison between autoencoders**

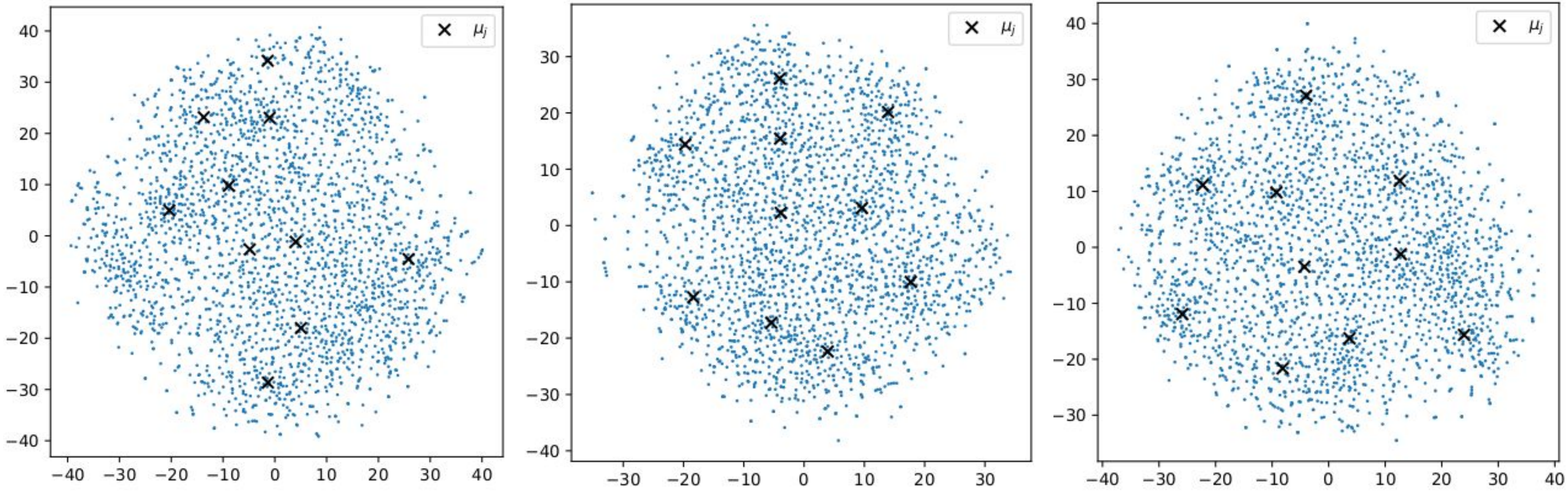
Method	MNIST FID	MNIST Reconstruction	CelebA FID	CelebA Reconstruction
VAE	71.55 ± 26.65	18.59 ± 2.22	59.99(*)	96.36(*)
GMVAE	75.68 ±11.95	18.19 ± 0.14	212.59 ± 18.15	97.77± 0.19
Vampprior	138.03 ± 34.09	29.98± 4.09	-	-
PRAE	100.25 ±41.72	16.20 ±3.14	52.20 (*)	**63.21(*)**
WAE	80.77 ± 11	11.53 ±0.33	52.07 (*)	63.83(*)
SWAE	80.28 ±19.22	14.12 ± 2.06	86.53 ± 2.49	89.71±2.15
DRAE	58.04 ± 20.74	14.07 ± 4.31	50.09 ±1.33	66.05 ± 2.56
m-DRAE (ours)	52.92 ±13.81	13.13 ± 0.33	49.05 ± 0.93	66.30 ± 0.22
s-DRAE (ours)	47.97 ± 13.83	11.17 ±1.73	46.63 ±0.83	66.62 ± 0.51
ps-DRAE (ours)	49.15 ± 12.93	11.71 ±1.21	48.21 ±1.02	66.31 ± 0.43
mps-DRAE (ours)	44.67 ±9.98	11.01 ±1.32	46.61 ± 1.01	66.23 ± 0.56
ms-DRAE (ours)	**43.57 ± 10.98**	**11.12 ±0.91**	**46.01 ±0.91**	65.91± 0.4

Experiments : Generated images

s-DRAE

ms-DRAE ($k = 50$)

Experiments : Latent space visualization



DRAE

s-DRAE

ms-DRAE ($k = 50$)

Summary

- Introducing a new family of sliced fused Gromov Wasserstein discrepancies
- Theoretical analysis (metricity, interpolation, curse of dimensionality)
- Introduce corresponding improved variants of RAEs
- Experimental results to show the favorable performance of new autoencoders

Email: v.khainb@vinai.io