Neural Monge map estimation and its applications

Jiaojiao Fan[1], Shu Liu[1], Shaojun Ma, Hao-min Zhou, Yongxin Chen

Georgia Tech

UCLA

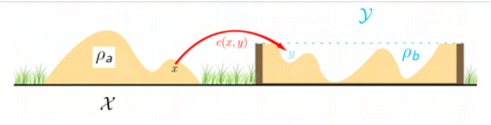
[1]Equal contribution

Monge formulation of Optimal transport (OT)

Our goal: Compute the Monge map T_*

$$T_* = \underset{T:\mathcal{X}\to\mathcal{Y}, T_\sharp \rho_a = \rho_b}{\arg\min} \int_{\mathcal{X}} c(x, T(x))\rho_a dx$$

Here we define $T_\sharp \rho_a$ as $T_\sharp \rho_a(E) = \rho_a(T^{-1}(E))$ for any measurable $E \subset \mathcal{X}$. One can treat $\mathcal{X}$, $\mathcal{Y}$ as Euclidean spaces $\mathbb{R}^n$, $\mathbb{R}^m$(n, m are not necessarily equal).



Explanation of Monge problem[2]

- Many applications in the generative model, multi-agent optimal control, computer vision, etc.

[2] https://medium.com/analytics-vidhya/introduction-to-optimal-transport-fd1816d51086

Saddle point scheme

Existing works on L^1, L^2 OT problems with costs $|x-y|$, $|x-y|^2$

Q: How to deal with OT problems with **general** cost?

A: Introduce the *Lagrange Multiplier* $f(\cdot)$ for $T_\sharp \rho_a = \rho_b$, and formulate the **saddle scheme**

$$\sup_{f \in C_b(\mathcal{Y})} \inf_{T \in \mathcal{M}(\mathcal{X},\mathcal{Y})} \mathcal{L}(T, f) \tag{1}$$

$C_b(\mathcal{Y})$ denotes the space of bounded continuous functions on $\mathcal{Y}$

$\mathcal{M}(\mathcal{X},\mathcal{Y})$ denotes the space of measurable map $T : \mathcal{X} \to \mathcal{Y}$

$$\mathcal{L}(T, f) = \int_{\mathcal{X}} \left[c(x, T(x)) - f(T(x))\right] \rho_a dx + \int_{\mathcal{Y}} f(y) \rho_b dy$$

We want to compute the saddle point $(\hat{T}, \hat{f})$ of (1), i.e.

$$\hat{T} \in \operatorname*{argmin}_{T \in \mathcal{M}(\mathcal{X},\mathcal{Y})} \mathcal{L}(T, \hat{f}) \qquad \hat{f} \in \operatorname*{argmax}_{f \in C_b(\mathcal{Y})} \mathcal{L}(\hat{T}, f)$$

Algorithm

Parametrize T and f by neural networks T_θ, f_η. Consider

$$\max_\eta \min_\theta \mathcal{L}(T_\theta, f_\eta) := \frac{1}{N}\sum_{k=1}^{N} c(X_k, T_\theta(X_k)) - f_\eta(T_\theta(X_k)) + f_\eta(Y_k). \tag{2}$$

Algorithm 1 Computing the Monge map from ρ_a to ρ_b

1: **Input**: Marginal distributions ρ_a and ρ_b, Batch size N, Cost function $c(x, y)$.
2: Initialize T_θ, f_η.
3: **for** K steps **do**
4: Sample $\{X_k\}_{k=1}^N \sim \rho_a$. Sample $\{Y_k\}_{k=1}^N \sim \rho_b$.
5: Update θ to decrease (2) for K_1 steps.
6: Update η to increase (2) for K_2 steps.
7: **end for**
8: **Output**: The transport map T_θ.

Comparison between our method and W-GAN

$$\underbrace{\max_f \min_T \int f(y)\rho_b(y)dy - \int f(T(x))\rho_a(x)dx + \int c(X, T(x))\rho_a(x)dx}_{\text{general Wasserstein distance } C_{\text{Monge}}(\rho_a,\rho_b)} \quad \text{Our method}$$

$$\min_G \underbrace{\max_{\|D\|_{\text{Lip}\leq 1}} \int D(y)\rho_b(y)dy - \int D(G(x))\rho_a(x)dx}_{1-\text{Wasserstein distance } W_1(G_\sharp\rho_a,\rho_b)} \quad \text{Wasserstein GAN}$$

- Our method: The optimal value is $C_{\text{Monge}}(\rho_a, \rho_b)$.

  W-GAN: The ideal optimal value is **0**

- Our method: Computes for **optimal** map T_* s.t. $T_{*\sharp}\rho_a = \rho_b$, and **minimizes** the transport cost

  W-GAN: Computes for **feasible** map G s.t. $G_\sharp\rho_a = \rho_b$

Theoretical guarantee

Theorem (Existence of saddle point & its consistency with Monge map)

We consider the saddle problem (1) *on* $\mathcal{X} = \mathbb{R}^n, \mathcal{Y} = \mathbb{R}^m$.

Assume that ρ_a, ρ_b *satisfy*

- ρ_a, ρ_b *are compactly supported Borel probability distributions on* $\mathbb{R}^n, \mathbb{R}^m$;
- ρ_a *is absolute continuous with respect to the Lebesgue measure on* $\mathbb{R}^n$.

Assume the cost $c(\cdot, \cdot)$ *satisfies*

- $c \in \mathcal{C}^1(\mathcal{X} \times \mathcal{Y})$;
- *Fix* $x \in \mathbb{R}^n$, $\nabla_x c(x, \cdot) : \mathbb{R}^m \ni y \mapsto \nabla_x c(x, y) \in \mathbb{R}^n$ *is an injective map;*
- *There exists a finite constant* $\underline{c}$ *such that* $c \geq \underline{c}$.

Then the saddle point of $\mathcal{L}(T, f)$ *exists. Furthermore, if* $(\hat{T}, \hat{f})$ *is a saddle point of* $\mathcal{L}(T, f)$*, then* $\hat{T}$ *is the Monge map.*

This is a simplified version of Theorem 2 and Corollary 1 from our paper.

Posterior error estimation via duality gaps

Consider solving saddle point problem on $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$. Suppose at a certain optimization stage, we obtain (T, f)

Theorem (*Posterior* error estimation via duality gaps)

Assume that

- $\nabla^2_{xy} c(x, y)$, *as a* $d \times d$ *matrix, is invertible for all* x, y;
- $\nabla^2_{yy} c(x, y)$ *is independent of* x;
- f *is* $c-$*concave function on* $\mathbb{R}^d$;
- *And some other standard conditions on* ρ_a, ρ_b *and* c *hold;*

Denote the duality gaps:

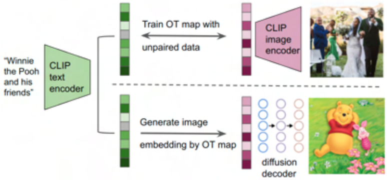
$$\mathcal{E}_1(T, f) = \mathcal{L}(T, f) - \inf_{\widetilde{T}} \mathcal{L}(\widetilde{T}, f), \quad \mathcal{E}_2(f) = \sup_{\widetilde{f}} \inf_{\widetilde{T}} \mathcal{L}(\widetilde{T}, \widetilde{f}) - \inf_{\widetilde{T}} \mathcal{L}(\widetilde{T}, f).$$

Denote T_* *as the Monge map of the OT problem. Then*

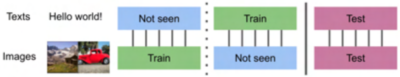
$$\|T - T_*\|_{L^2(\beta \rho_a)} \leq \sqrt{2(\mathcal{E}_1(T, f) + \mathcal{E}_2(f))},$$

where $\beta(\cdot) > 0$ *is a positive weight function depending on* c, T_*, *and* f.

Experiment 1: Unpaired text to image generation



Pipeline motivated by DALL·E2.



Unpaired data generation process

- ρ_a: distribution of text encoding $x \in \mathbb{R}^{77\times768}$ $(x \neq 0)$ from CLIP model;
- ρ_b: distribution of image embedding $y \in \mathbb{R}^{768}$.
- Choose transport cost as negative cosine similarity between Rx and y:

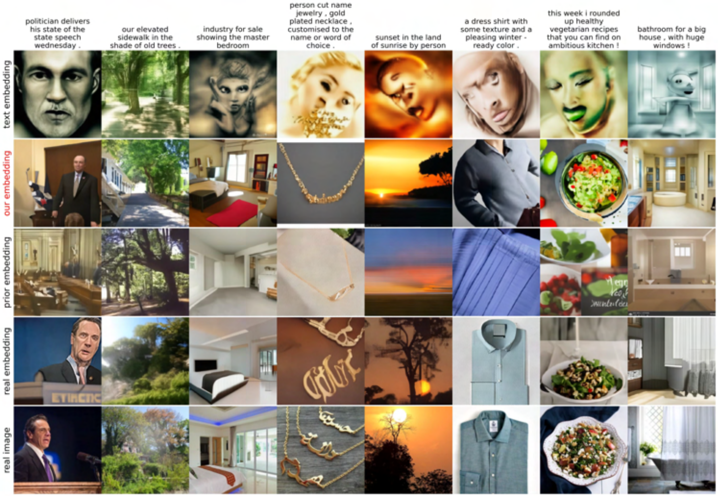
$$c(x, y) = -\frac{\langle Rx, y\rangle}{\|Rx\|_2\|y\|_2}.$$

The frozen matrix $R : \mathbb{R}^{77\times768} \to \mathbb{R}^{768}$ is extracted from a linear layer of CLIP model and it projects the text encoding x to the same dimension as image embedding y.

Experiment 1: Unpaired text to image generation

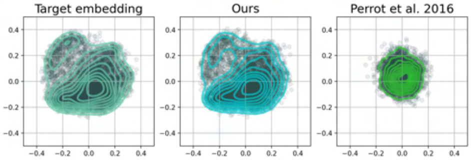
Evaluation of our method on the Laion art dataset

Experiment 1: Unpaired text to image generation

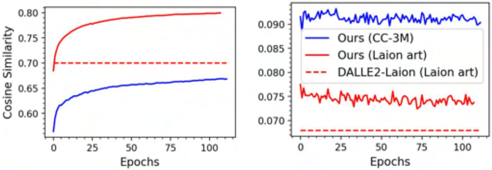


Evaluation of our method on the Conceptual Captions 3M (CC-3M) dataset

Experiment 1: Unpaired text to image generation



(Verification on $T_\sharp \rho_a \approx \rho_b$) Target image embeddings ρ_b (Left), fitted measure of generated embeddings by our method $T_\sharp \rho_a$, results of non-linear kernel map given by Perrot et al. (2016).



(Verification on the optimality of computed T) Averaged cosine similarity between the generated image embeddings and (Left) the ground truth image embeddings or (Right) the unrelated text embeddings.

Experiment 2: Unpaired image inpaintinting

- ρ_a: distribution of images with masked faces
- ρ_b: distribution of images with intact faces
- Choose the cost function to be a mean squared error (MSE) in the *unmasked area*

$$c(x, y) = \alpha \cdot \frac{\|x \odot M(x) - y \odot M(x)\|_2^2}{n},$$

$M(x)$ is a binary mask with the same size as the image. M takes the value 1 in the unmasked region and 0 in the masked region. $\odot$ represents the point-wise multiplication, α is a tunable coefficient, and n is the dimension of x.

Unpaired image inpainting on **test** dataset of CelebA 128×128. We take the composite image $G(x) = T(x) \odot M^C + x \odot M$ $(M^C = \mathbf{1} - M)$ as the output image

Experiment 2: Unpaired image inpaintinting

Masked images

Real images

Our $T(x)$

Unpaired image inpainting on the **test** dataset of CelebA 64×64.

Experiment 3: Population transportation on Earth

- ρ_a: current distribution of the population on Earth;
- ρ_b: uniform distribution of population over the landmass on Earth.
- Choose the cost function as the *geodesic distance* ($\lambda = 1$) on the sphere

$$c_\lambda((\theta_1, \phi_1), (\theta_2, \phi_2)) = \arccos(\lambda(\sin\phi_1 \sin\phi_2 \cos(\theta_1 - \theta_2) + \cos\phi_1 \cos\phi_2)).$$

Here we represent the distance function on sphere under the spherical coordinate (θ, ϕ) (fix radius $r = 1$). In practice, we use the approximation versions of c_λ ($\lambda < 1$) to relieve the gradient blow-up of $\arccos(\cdot)$ near ± 1.

Left: Samples of $T_\sharp \rho_a$ (green) and samples of ρ_a (blue)
Right: Transport map with cost c_λ, $\lambda = 0.99$

Thank you!

Contact information:

Jiaojiao Fan jiaojiaofan@gatech.edu

Shu Liu shuliu@math.ucla.edu

Shaojun Ma shaojunma@gatech.edu

Hao-min Zhou hmzhou@math.gatech.edu

Yongxin Chen yongchen@gatech.edu