

Generative Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

Article analysis: Neural Optimal Transport

Paul Barbier and Bastien Le Chenadec

March 28, 2024

Overview

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

1 Introduction

2 Optimal Transport

- The Problem
- Weak OT Duality

3 NOT

- Weak dual OT reformulation
- SGAD

4 Experiments

- Synthetic data
- Large-scale dataset

5 Conclusion

Introduction

Generative Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal Transport

The Problem
Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data
Large-scale dataset

Conclusion

Optimal Transport Problem

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction
Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments
Synthetic data
Large-scale dataset

Conclusion

- μ, ν probability distributions on \mathcal{X}, \mathcal{Y} .
- Cost function: $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$.
- Problem: finding a **transport map** $T^* : \mathcal{X} \rightarrow \mathcal{Y}$ such that:

$$T^* \in \operatorname{Argmin}_{T\#\mu=\nu} \int_{\mathcal{X}} c(x, T(x)) d\mu(x) \quad (1)$$

where $(T\#\mu)(A) = \mu(T^{-1}(A))$ is the pushforward measure.
Hence, one can define:

$$\text{Cost}(\mu, \nu) = \int_{\mathcal{X}} c(x, T^*(x)) d\mu(x) \quad (2)$$

Kantorovich formulation: a relaxed OT problem

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

- Let $\Pi(\mu, \nu)$ the set of distributions on $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν .
- Problem: find a **transport plan** $\pi^* \in \Pi(\mu, \nu)$ such that:

$$\pi^* \in \operatorname{Argmin}_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) \quad (3)$$

and we can define:

$$\text{Cost}(\mu, \nu) = \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi^*(x, y) \quad (4)$$

Weak OT Duality

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data
Large-scale dataset

Conclusion

For weak OT cost C and a sufficiently regular function f ,

$$f^C(x) = \inf_{\rho \in \mathcal{P}(\mathcal{Y})} \left\{ C(x, \rho) - \int_{\mathcal{Y}} f(y) d\rho(y) \right\} \quad (5)$$

where $\mathcal{P}(\mathcal{Y})$ denotes the set of probability distributions on \mathcal{Y} .

Then, the dual form of the weak OT problem writes:

$$f^* \in \operatorname{Argmax}_f f^C(x) d\mu(x) + \int_{\mathcal{Y}} f(y) d\nu(y) \quad (6)$$

and the cost can be defined as follows

$$\text{Cost}(\mu, \nu) = \int_{\mathcal{X}} f^{*C}(x) d\mu(x) + \int_{\mathcal{Y}} f^*(y) d\nu(y) \quad (7)$$

Weak dual OT reformulation

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem
Weak OT Duality

NOT

Weak dual OT
reformulation

SgAD

Experiments

Synthetic data
Large-scale dataset

Conclusion

- $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{Y} \subset \mathbb{R}^m$ and $\mathcal{Z} \subset \mathbb{R}^d$.
- Let $\rho \in \mathcal{P}(\mathcal{Z})$ with some basic assumptions.

Then, we have:

$$f^C(x) = \inf_t \left\{ C(x, T\#\rho) - \int_{\mathcal{Z}} f(t(z)) d\rho(z) \right\} \quad (8)$$

which leads to the maximin problem:

$$\text{Cost}(\mu, \nu) = \sup_f \inf_T \mathcal{L}(f, T) \quad (9)$$

where $\mathcal{L}(f, T) =$

$$\int_{\mathcal{Y}} f d\nu + \int_{\mathcal{X}} \left(C(x, T(x, \cdot)\#\rho) - \int_{\mathcal{Z}} f(T(x, z)) d\rho(z) \right) d\mu(x)$$

The trick: noise outsourcing

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem
Weak OT Duality

NOT
Weak dual OT
reformulation
SGAD

Experiments
Synthetic data
Large-scale dataset

Conclusion

- Here, they introduce $T: \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{Y}$.
- Trick known as **noise outsourcing**:

Theorem

If X and Y are random variables in suitable spaces \mathcal{X} and \mathcal{Y} , then there exists $\eta \sim \mathcal{U}([0, 1])$ with $\eta \perp\!\!\!\perp X$ and a function $h: [0, 1] \times \mathcal{X} \rightarrow \mathcal{Y}$ such that $(X, Y) = (X, h(\eta, X))$ almost surely.

Weak OT: A summary

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

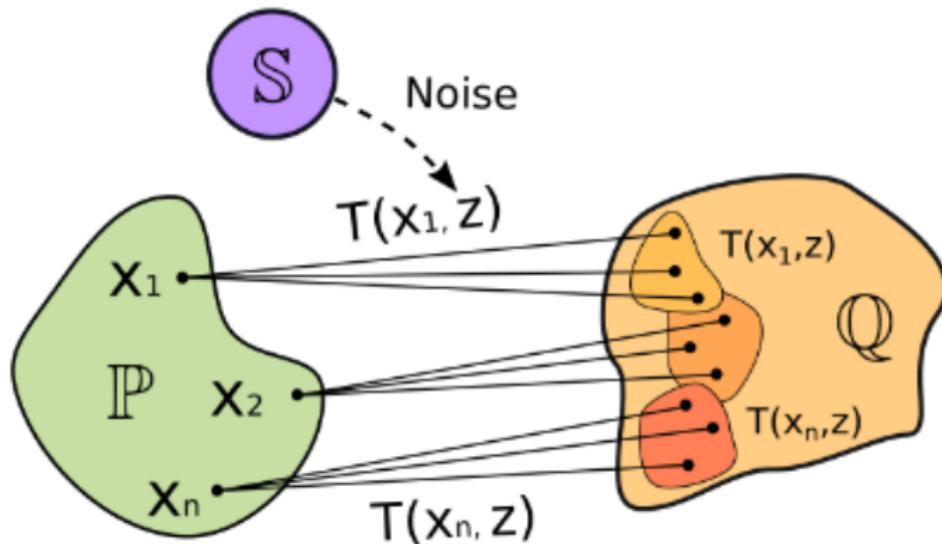


Figure: Transportation map with outsourced noise z (taken from [KSB22])

Stochastic Gradient Ascent Descent (SGAD)

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction
Optimal
Transport
The Problem
Weak OT Duality

NOT
Weak dual OT
reformulation
SGAD

Experiments
Synthetic data
Large-scale dataset

Conclusion

Algorithm Stochastic Gradient Ascent Descent algorithm

- 1: **Input:** distributions μ, ν, ρ accessible by samples, mapping network $T_\theta : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^d$, potential network $f_\omega : \mathbb{R}^n \rightarrow \mathbb{R}$, number of inner iterations K_T , (weak) cost $C : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \rightarrow \mathbb{R}$, empirical estimator $\hat{C}(x, T(x, Z))$ for the cost
 - 2: **Output:** learned stochastic OT map T_θ representing an OT plan between distributions μ, ν
 - 3: **repeat**
 - 4: Sample batches $Y \sim \nu$, $X \sim \mu$, for each $x \in X$ sample batch $Z_x \sim \rho$
 - 5: $\mathcal{L}_f \leftarrow \frac{1}{|X|} \sum_{x \in X} \frac{1}{|Z_x|} \sum_{z \in Z_x} f_\omega(T_\theta(x, z)) - \frac{1}{|Y|} \sum_{y \in Y} f_\omega(y)$
 - 6: Update ω by using $\frac{\partial \mathcal{L}_f}{\partial \theta}$
 - 7: **for** $k_T = 1, 2, \dots, K_T$ **do**
 - 8: Sample batch $X \sim \mu$, for each $x \in X$ sample batch $Z_x \sim \rho$
 - 9: $\mathcal{L}_T \leftarrow \frac{1}{|X|} \sum_{x \in X} [\hat{C}(x, T_\theta(x, Z_x)) - \frac{1}{|Z_x|} \sum_{z \in Z_x} f_\omega(T_\theta(x, z))]$
 - 10: Update θ by using $\frac{\partial \mathcal{L}_T}{\partial \theta}$
 - 11: **until** converged
-

Synthetic data

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

- 2D standard normal distribution for \mathcal{X} .
- Moon distribution for \mathcal{Y} .
- Weak OT cost for $\gamma = 1$:

$$C(x, \mu) = \int_{\mathcal{Y}} \frac{1}{2} \|x - y\|^2 d\mu(y) - \frac{1}{2} \text{Var}(\mu) = \frac{1}{2} \|x - \int_{\mathcal{Y}} y d\mu(y)\|^2 \quad (10)$$

The parameters:

- simple feedforward model.
 - 2 hidden layers (100 units each).
 - ReLU activation.
- Adam optimizer.

Results: the fitted distribution

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

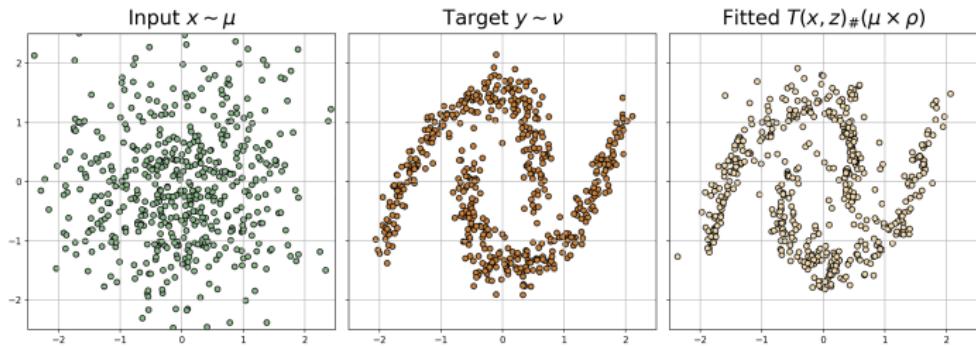


Figure: Synthetic data experiment. Left: input gaussian distribution. Middle: target distribution. Right: learned transport of the input distribution to the target distribution.

Results: a closer look at the map

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

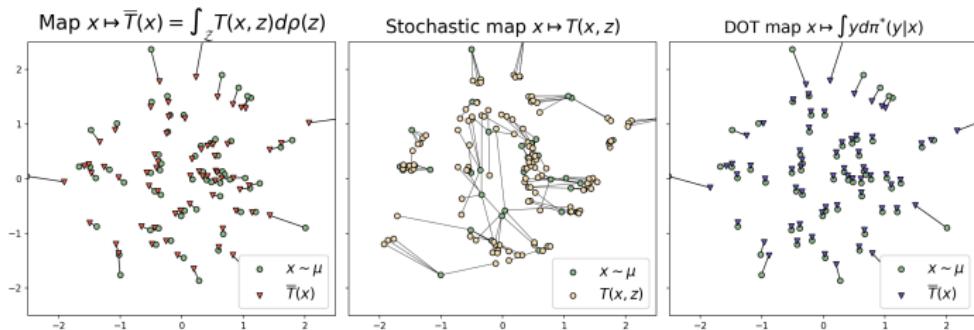


Figure: Synthetic data experiment. Left: average of the learned transport for different points. Middle: learned transport for a batch of points. Right: optimal transport plan obtained with the POT library.

Large-scale dataset

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

- CelebA dataset with 200k celebrity face images.
- CartoonSet with 100k avatar 2D generated images.



Figure: Images samples from CartoonSet100K

Large-scale experiment: the parameters

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

- Resnet to parametrise f .
- U-Net to parametrise the map T .
- image size: 128×128 .
- Batch size: 128.
- Adam optimizer.
- $\text{lr} = 1 \times 10^{-4}$.
- $k_T = 10$.

Results: it works!

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion



Figure: \mathcal{L}_f during training

Results: some generated avatars

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion



Figure: Iteration 1

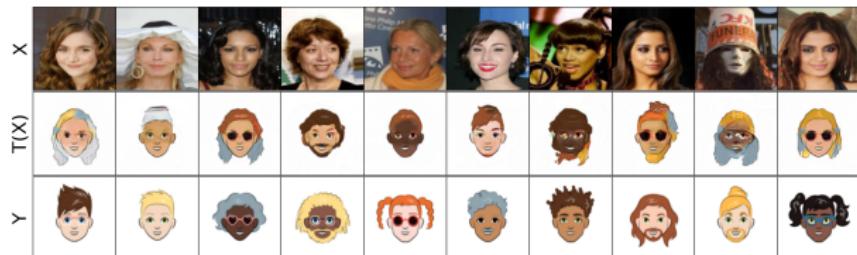


Figure: Iteration 5000

Results: an issue

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion

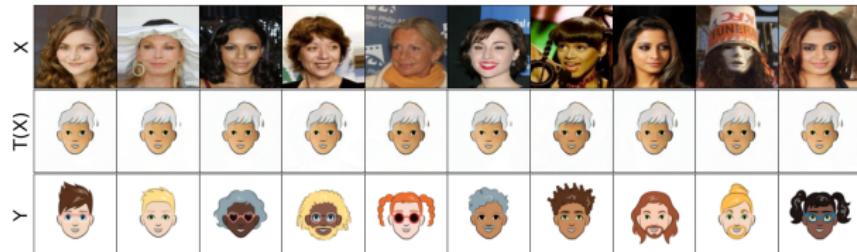


Figure: Iteration 13800

■ Mode collapse!

Conclusion

Generative
Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal
Transport

The Problem
Weak OT Duality

NOT

Weak dual OT
reformulation
SGAD

Experiments

Synthetic data
Large-scale dataset

Conclusion

- An algorithm to learn optimal transport maps.
- New approach to incorporate OT in ML.
- With good results on large-scale datasets!
- Good perspectives for generative models.

Generative Images

Paul Barbier
and Bastien
Le Chenadec

Introduction

Optimal Transport

The Problem

Weak OT Duality

NOT

Weak dual OT
reformulation

SGAD

Experiments

Synthetic data

Large-scale dataset

Conclusion



Alexander Korotin, Daniil Selikhanovich, and Evgeny Burnaev, *Neural optimal transport*, arXiv (Cornell University) (2022).