Presentation at ICLR ’25

MMD-Regularized Unbalanced Optimal Transport

Piyushi Manupriya, J. SakethaNath, Pratik Jawanpuria*

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Accepted at **Transactions on Machine Learning Research**

TMLR Jan ’24

*Work done while at Microsoft, INDIA.

Introduction to Optimal Transport

- Given c, a cost function over $\mathcal{X} \times \mathcal{X}$, **Kantorovich Optimal Transport (OT)** formulation between $s_0, t_0 \in \mathcal{R}_1^+(\mathcal{X})$:

$$\inf_{\pi \in \mathcal{R}_1^+(\mathcal{X} \times \mathcal{X})} \int c \, \mathrm{d}\pi \ \text{ s.t. } \ \pi_1 = s_0 \text{ and } \pi_2 = t_0,$$

where π_1, π_2 are marginal densities of π.

- *Optimal value*, whenever c is a metric, gives the *Wasserstein metric*.

$\mathcal{X}$: space where supports of the distributions lie.
$\mathcal{R}_1^+(\mathcal{X})$: set of probability measures over $\mathcal{X}$.

Unbalanced Optimal Transport

- **Unbalanced OT (UOT)** between $s_0, t_0 \in \mathcal{R}^+(\mathcal{X})$:

$$\inf_{\pi \in \mathcal{R}^+(\mathcal{X} \times \mathcal{X})} \int c \, \mathrm{d}\pi + \lambda D(\pi_1, s_0) + \lambda D(\pi_2, t_0)$$

where $\lambda > 0$ and D is a divergence over measures

Unbalanced Optimal Transport

- **Unbalanced OT (UOT)** between $s_0, t_0 \in \mathcal{R}^+(\mathcal{X})$:

$$\inf_{\pi \in \mathcal{R}^+(\mathcal{X} \times \mathcal{X})} \int c \, \mathrm{d}\pi + \lambda D(\pi_1, s_0) + \lambda D(\pi_2, t_0)$$

where $\lambda > 0$ and D is a divergence over measures

- **Generalized Wasserstein distance** between $s_0, t_0 \in \mathcal{R}^+(\mathcal{X})$:

$$\inf_{\pi \in \mathcal{R}^+(\mathcal{X} \times \mathcal{X})} \int c \, \mathrm{d}\pi + \lambda \mathrm{TV}(\pi_1, s_0) + \lambda \mathrm{TV}(\pi_2, t_0)$$

Piccoli & Rossi, 2016

- **KL-Regularized UOT** between $s_0, t_0 \in \mathcal{R}^+(\mathcal{X})$:

$$\inf_{\pi \in \mathcal{R}^+(\mathcal{X} \times \mathcal{X})} \int c \, \mathrm{d}\pi + \lambda \mathrm{KL}(\pi_1, s_0) + \lambda \mathrm{KL}(\pi_2, t_0)$$

Study of metricity: *Liero et al. 2018*
Computational aspects: *Chizat et al. 2018*
Improvements in various downstream ML applications

Proposed MMD-Regularized Unbalanced Optimal Transport

- We propose UOT formulation with Maximum Mean Discrepancy (MMD) regularization.
- **MMD-UOT** between $s_0, t_0 \in \mathcal{R}^+(\mathcal{X})$: $\inf_{\pi \in \mathcal{R}^+(\mathcal{X}\times\mathcal{X})} \int c \, \mathrm{d}\pi + \lambda \mathrm{MMD}_{\mathrm{k}}(\pi_1, s_0) + \lambda \mathrm{MMD}_{\mathrm{k}}(\pi_2, t_0)$
- Key research questions we focus on:
  - ❑ Does MMD-UOT lead to a valid metric?
  - ❑ Statistical complexity to estimate MMD-UOT?
  - ❑ Computational complexity?

$\mathcal{X}$: space where supports of the distributions lie.
$\mathcal{R}^+(\mathcal{X})$: subset of non-negative measures over $\mathcal{X}$.
$\lambda > 0$
k: kernel used.

MMD-UOT: Duality Result

MMD-UOT between $s_0, t_0 \in \mathcal{R}^+(\mathcal{X})$:

$$U_{c,k,\lambda}(s_0, t_0) \;:=\; \inf_{\pi \in \mathcal{R}^+(\mathcal{X}\times\mathcal{X})} \int c \,\mathrm{d}\pi + \lambda \mathrm{MMD_k}(\pi_1, s_0) + \lambda \mathrm{MMD_k}(\pi_2, t_0)$$

Dual Reformulation: Whenever c is a (continuous) ground metric over a compact domain,

$$U_{c,k,\lambda}(s_0, t_0) = \max_{f \in G(k,\lambda) \cap W_c} \int_{\mathcal{X}} f \,\mathrm{d}s_0 - \int_{\mathcal{X}} f \,\mathrm{d}t_0$$

where $G(k,\lambda) := \{f \in H_k : ||f||_k \le \lambda\}$

$W_c := \{f \in C: ||f||_L \le 1\}$.

$\mathcal{X}$: space where supports of the distributions lie.
H_k : Reproducing Kernel (k) Hilbert Space
C: set of continuous functions

MMD-UOT: Metric Properties

Dual Reformulation: Whenever c is a (continuous) ground metric over a compact domain,

$$U_{c,k,\lambda}(s_0, t_0) = \max_{f \in G(k,\lambda) \cap W_c} \int_{\mathcal{X}} f \, \mathrm{d}s_0 - \int_{\mathcal{X}} f \, \mathrm{d}t_0$$

Corollary:

- $U_{c,k,\lambda}$ belongs to the family of Integral Probability Metrics (*Muller, 1997*)
  - with the generating set $G(k,\lambda) \cap W_c$
- By showing positive definiteness of $U_{c,k,\lambda}$ (with a characteristic kernel), we conclude $U_{c,k,\lambda}$ is a norm-induced metric.

$\mathcal{X}$: space where supports of the distributions lie.
H_k : Reproducing Kernel (k) Hilbert Space
C: set of continuous functions
$G(k,\lambda) \coloneqq \{f \in H_k : ||f||_k \leq \lambda\}$
$W_c \coloneqq \{f \in C : ||f||_L \leq 1\}$

MMD-UOT: Sample Efficiency

Dual Reformulation: Whenever c is a (continuous) ground metric over a compact domain,

$$U_{c,k,\lambda}(s_0, t_0) = \max_{f \in G(k,\lambda) \cap W_c} \int_{\mathcal{X}} f \, \mathrm{d}s_0 - \int_{\mathcal{X}} f \, \mathrm{d}t_0$$

Corollary: With $\hat{s}_m, \hat{t}_m$ as the corresponding empirical measures over m samples,

$U_{c,k,\lambda}(\hat{s}_m, \hat{t}_m) \rightarrow U_{c,k,\lambda}(s_0, t_0)$ with a rate same as $\mathrm{MMD}(\hat{s}_m, \hat{t}_m) \rightarrow \mathrm{MMD}(s_0, t_0)$.

Using a normalized kernel, sample complexity: $\boldsymbol{O}\left(\boldsymbol{m}^{-\frac{1}{2}}\right)$.

$\mathcal{X}$: space where supports of the distributions lie.
H_k : Reproducing Kernel (k) Hilbert Space
C: set of continuous functions
$G(k, \lambda) := \{f \in H_k : ||f||_k \leq \lambda\}$
$W_c := \{f \in C : ||f||_L \leq 1\}$

MMD-UOT: Consistent Estimator

$$U_{c,k,\lambda}(s_0, t_0) := \min_{\pi \in \mathcal{R}^+(\mathcal{X}\times\mathcal{X})} \int c \, \mathrm{d}\pi + \lambda \mathrm{MMD}_k(\pi_1, s_0) + \lambda \mathrm{MMD}_k(\pi_2, t_0)$$

$o\left(m^{-\frac{1}{2}}\right)$

$o\left(m^{-\frac{1}{2}}\right)$

$$\min_{\pi \in \mathcal{R}^+(\mathcal{X}\times\mathcal{X})} \int c \, \mathrm{d}\pi + \lambda \mathrm{MMD}_k(\pi_1, \hat{s}_m) + \lambda \mathrm{MMD}_k(\pi_2, \hat{t}_m)$$

$$\min_{\pi \in \mathbb{R}_+^{m\times m}} \int c \, \mathrm{d}\pi + \lambda \mathrm{MMD}_k(\pi_1, \hat{s}_m) + \lambda \mathrm{MMD}_k(\pi_2, \hat{t}_m)$$

Such finite-parameterization bounds are not studied for prior OT variants.

MMD-UOT: Computational Aspects

With an equivalent squared-MMD regularization,

$$\min_{\pi \in \mathbb{R}_+^{m\times m}} \int c \, \mathrm{d}\pi + \lambda \mathrm{MMD}_k^2(\pi_1, \hat{s}_m) + \lambda \mathrm{MMD}_k^2(\pi_2, \hat{t}_m)$$

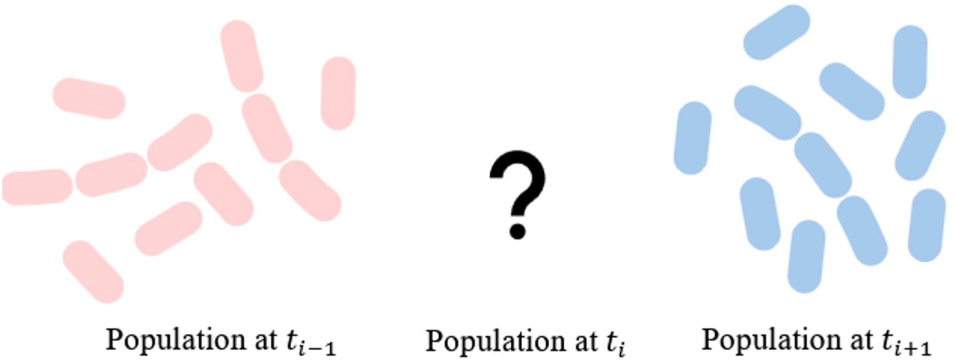
we show the objective is L-smooth

& employ accelerated PGD with step size $\left(\frac{1}{L}\right)$.

Resulting computational complexity: $\widetilde{O}\left(\frac{m^2}{\sqrt{\epsilon}}\right)$ with ϵ as the optimality gap.

Experimental Results: Single Cell RNA Sequencing

Setting: MMD-UOT based interpolating barycenter to predict the cell population



Timestep	MMD	ϵ-KLOT	Proposed MMD-UOT
t_1	0.375	0.391	**0.334**
t_2	0.190	0.184	**0.179**
t_3	0.125	0.138	**0.116**
Avg.	0.230	0.238	**0.210**

Distance between the predicted & the ground truth in terms of MMD.

ϵ-KLOT: KL-Regularized UOT with entropic regularization.

Experimental Results: Unsupervised Domain Adaptation

Setting: MMD-UOT based loss for Unsupervised Domain Adaptation following the setup in Fatras et al., [2021].

Method	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P
ResNet-50	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9
DANN [Ganin et al., 2016]	44.3	59.8	69.8	48.0	58.3	63.0	49.7	42.7	70.6	64.0	51.7	78.3
CDAN-E [Long et al., 2017]	52.5	71.4	76.1	59.7	69.9	71.5	58.7	50.3	77.5	70.5	57.9	83.5
DEEPJDOT [Damodaran et al., 2018]	50.7	68.7	74.4	59.9	65.8	68.1	55.2	46.3	73.8	66.0	54.9	78.3
ALDA [Chen et al., 2020]	52.2	69.3	76.4	58.7	68.2	71.1	57.4	49.6	76.8	70.6	57.3	82.5
ROT [Balaji et al., 2020]	47.2	71.8	76.4	58.6	68.1	70.2	56.5	45.0	75.8	69.4	52.1	80.6
ϵ-KLOT (JUMBOT) [Fatras et al., 2021]	55.2	75.5	80.8	65.5	74.4	74.9	65.2	52.7	79.2	73.0	59.9	83.4
BombOT [Nguyen et al., 2022]	56.2	75.2	80.5	65.8	74.6	75.4	66.2	53.2	80.0	74.2	**60.1**	83.3
MMD-UOT	**56.5**	**77.2**	**82.0**	**70.0**	**77.1**	**77.8**	**69.3**	**55.1**	**82.0**	**75.5**	59.3	**84.0**

Accuracies for adaptation from source → target domains on Office-Home dataset.

Summary & Open Problems

Summary of Contributions

Statistically efficient variants of Kantorovich-Wasserstein
Computationally efficient solver
Improvements in downstream ML applications

More results in our paper.

Open Problems

On expanding the proof of metricity
On finding settings with closed form solutions

More in our paper.

Code:
https://github.com/Piyushi-0/MMD-reg-OT

Other links:

Contacts:

cs18m20p100002@iith.ac.in
saketha@cse.iith.ac.in
pratik.iitb@gmail.com