Solving Discrete (Semi) Unbalanced Optimal Transport with Equivalent Transformation Mechanism and KKT-Multiplier Regularization

NeurIPS

Weiming Liu, Xinting Liao, Jun Dan, Fan Wang, Hua Yu, Junhao Dong, Shunjie Dong, Lianyong Qi, Yew-Soon Ong

2025

Overview

Discrete Semi-Unbalanced Optimal Transport (Background)

(Definition) Discrete Semi-Unbalanced Optimal Transport

$$\begin{aligned}
&\min_{\pi_{ij} \geq 0} J_{\mathrm{SemiUOT}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a} \right), \\
&\text{s.t. } \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}.
\end{aligned} \tag{1}$$

- $\boldsymbol{X} \in \mathbb{R}^{M \times D}$ and $\boldsymbol{Z} \in \mathbb{R}^{N \times D}$ denote source and target domains, where M, N denote the number of samples and D denotes the data dimension.
- $\boldsymbol{a} \in \mathbb{R}^{M \times 1}$ and $\boldsymbol{b} \in \mathbb{R}^{N \times 1}$ denote the mass weights for source and target domains.
- SemiUOT is set to measure the minimum cost among data samples $\boldsymbol{X}$ and $\boldsymbol{Z}$, meanwhile filtering out the outliers by relaxing one of marginal constraints.

Equivalent Transformation Mechanism for SemiUOT

Equivalent Transformation Mechanism (ETM) investigates the problem of SemiUOT from the perspective of marginal probability distribution.

Principles of Equivalent Transformation Mechanism (ETM) for SemiUOT

The Fenchel-Lagrange multipliers form of J_{SemiUOT} is given:

$$\min_{\boldsymbol{f},\boldsymbol{g},\zeta} \left[\tau \sum_{i=1}^{M} a_i \exp\left(-\frac{f_i + \zeta}{\tau}\right) - \sum_{j=1}^{N} b_j (g_j - \zeta) \right] \quad s.t. \begin{cases} f_i + g_j + s_{ij} = C_{ij}, \\ s_{ij} \geq 0 \end{cases}$$

where $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{s}$ and ζ denote Lagrange multipliers. Moreover, SemiUOT problem can be further transformed into OT with marginal constraints as follows:

$$\min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{P}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle, \quad s.t.\ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\frac{\boldsymbol{f}^* + \zeta^*}{\tau}\right) = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}.$$

Equivalent Transformation Mechanism for SemiUOT

(Definition) Approximate SemiUOT Equation

$$\min_{\widehat{\boldsymbol{f}},\zeta} \widehat{L}_{\mathrm{P}} = \tau \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{f}_i + \zeta}{\tau}\right) + \sum_{j=1}^{N} b_j \left[\epsilon \log\left[\sum_{k=1}^{M} \exp\left(\frac{\widehat{f}_k - C_{kj}}{\epsilon}\right)\right] + \zeta\right]. \quad (2)$$

To accelerate the optimization process, we consider making a smooth approximation on replacing $\inf(\cdot)$ with $g_j = \inf_{k\in[M]}[C_{kj} - f_k] \approx -\epsilon \log[\sum_{k=1}^{M} e^{\frac{f_k - C_{kj}}{\epsilon}}]$.

Equivalent Transformation Mechanism for SemiUOT

Calculation for Approximate SemiUOT Equation

ETM-Approx aims to solve the following equation for each $\widehat{f}_s$:

$$\frac{\partial \widehat{L}_{\mathrm{P}}}{\partial \widehat{f}_s} = -a_s \exp\left(-\frac{\widehat{f}_s + \zeta}{\tau}\right) + \exp\left(\frac{\widehat{f}_s}{\epsilon}\right) \sum_{j=1}^{N} \left[\frac{b_j \exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{\widehat{f}_k - C_{kj}}{\epsilon}\right)}\right] = 0. \quad (3)$$

We can adopt fixed-point iteration method for solving Eq.(3) at the ℓ-th iteration:

$$\widehat{f}_s^{\ell+1} = \nu \left[\log\left(a_s \exp\left(-\frac{\zeta}{\tau}\right)\right) - \log\left[\sum_{j=1}^{N}\left(\frac{b_j}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \exp\left(-\frac{C_{sj}}{\epsilon}\right)\right)\right]\right] \quad (4)$$

Equivalent Transformation Mechanism for SemiUOT

(Remark 1) Convergence on ETM-Approx

ETM-Approx can reach the linear convergence rate via the fixed-point iteration shown as $\mathcal{O}(NM\log(1/\varepsilon_{\text{err}}))$ where $\varepsilon_{\text{err}} = ||\widehat{\boldsymbol{f}} - \widehat{\boldsymbol{f}}^*||_\infty$ and $\widehat{\boldsymbol{f}}^*$ denotes the optimal solution.

(Remark 2) Connections between ETM-Approx and ETM-Refine

$$\text{SemiUOT} \xrightarrow[\text{ETM Approx}]{\text{Fixed Point Iteration}} \widehat{\boldsymbol{f}}^* \xrightarrow[\text{ETM Refine}]{\text{BFGS}} \boldsymbol{f}^* \tag{5}$$

ETM-Refine utilizes the strength of ETM-Approx in efficient computation for the exact results.

Equivalent Transformation Mechanism for UOT

(Definition) Discrete Unbalanced Optimal Transport

$$\min_{\pi_{ij}\geq 0} J_{\mathrm{UOT}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau_a \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \tau_b \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}) \tag{6}$$

Principles of Equivalent Transformation Mechanism (ETM) for UOT

The Fenchel-Lagrange multipliers form of J_{UOT} is given:

$$\min_{\boldsymbol{u},\boldsymbol{v},\zeta} \left[\tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) + \tau_b \sum_{j=1}^{N} b_j \exp\left(-\frac{v_j - \zeta}{\tau_b}\right) \right], \; s.t. \begin{cases} u_i + v_j + s_{ij} = C_{ij}, \\ s_{ij} \geq 0, \end{cases}$$

where $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{s}$ and ζ denote Lagrange multipliers. UOT problem can be further transformed into OT with marginal constraints as follows:

$$\min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\mathrm{U}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle, \quad s.t. \; \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\frac{\boldsymbol{u}^* + \zeta^*}{\tau_a}\right), \; \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \odot \exp\left(-\frac{\boldsymbol{v}^* - \zeta^*}{\tau_b}\right)$$

Equivalent Transformation Mechanism for UOT

(Definition) Exact UOT Equation

$$\min_{\boldsymbol{u},\zeta} L_{\mathrm{U}} = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i+\zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(\frac{\sup_{k\in[M]}\left(u_k - C_{kj}\right)}{\tau_b}\right) \tag{7}$$

Since the optimization problem is convex, we can also utilize block gradient descent to optimize the problem.

$$\begin{aligned} &\min_{\boldsymbol{\pi}\geq 0} J_{\mathrm{U}}^{u} = \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau_a \mathrm{KL}\left(\boldsymbol{\pi}1_N \| \boldsymbol{a}\right), \\ &s.t. \begin{cases} \text{(Constraint)}: \ \boldsymbol{\pi}^{\top}\mathbf{1}_M = \boldsymbol{b}\odot \exp\left(-\dfrac{\boldsymbol{v}-\zeta}{\tau_b}\right) = \boldsymbol{\beta} \\ \text{(Optional)}: \ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a}\odot \exp\left(-\dfrac{\boldsymbol{u}+\zeta}{\tau_a}\right) = \boldsymbol{\alpha} \end{cases}. \end{aligned} \tag{8}$$

Equivalent Transformation Mechanism for UOT

Specifically, we first fix $\widehat{\boldsymbol{v}}^l$ and optimize variable $\widehat{\boldsymbol{u}}^l$ at the l-th iteration by replacing the original marginal probability $\boldsymbol{b}$ with β to transform UOT into SemiUOT problem:

$$\min_{\widehat{\boldsymbol{u}}} \widehat{L}_{\mathrm{U}}^{u} = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i + \zeta}{\tau_a}\right) + \sum_{j=1}^{N} \beta_j \left[\epsilon \log\left[\sum_{k=1}^{M} \exp\left(\frac{\widehat{u}_k - C_{kj}}{\epsilon}\right)\right] + \zeta\right]. \tag{9}$$

where $\boldsymbol{b} \odot \exp(-(\widehat{\boldsymbol{v}} - \zeta)/\tau_b) = \boldsymbol{\beta}$. It is equivalent to solve the equation by taking the differentiation w.r.t. on $\widehat{u}_s$ over $\widehat{L}_{\mathrm{U}}^{u}$ and set it 0:

$$\frac{\partial \widehat{L}_{\mathrm{U}}^{u}}{\partial \widehat{u}_s} = -a_s \exp\left(-\frac{\widehat{u}_s + \zeta}{\tau_a}\right) + \exp\left(\frac{\widehat{u}_s}{\epsilon}\right) \sum_{j=1}^{N} \left[\frac{\beta_j \exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{\widehat{u}_k - C_{kj}}{\epsilon}\right)}\right] = 0. \tag{10}$$

Thus the problem can be solved via the fixed-point iteration method as well.

Multiplier Regularized Optimal Transport

(Remark 3) Relation between SemiUOT/UOT and OT

Given any UOT/SemiUOT with KL divergence, we can transfer the original problem into classic optimal transport via adopting proposed ETM approach flexibly.

$$(\mathrm{UOT}, \mathrm{SemiUOT}) \xrightarrow{\text{ETM Method}} \mathrm{OT} \xrightarrow{\text{OT Solver}} \boldsymbol{\pi}^* \tag{11}$$

(Remark 4) KKT conditions

Multipliers $\boldsymbol{s}$ indicate the value of $\boldsymbol{\pi}$, i.e., (case 1) $s_{ij} > 0$ when $\pi_{ij} = 0$ and (case 2) $s_{ij} = 0$ when $\pi_{ij} > 0$ according to the KKT conditions.

Multiplier Regularized Optimal Transport

Multiplier Regularized Optimal Transport (MROT)

Given any OT with multiplier $\boldsymbol{s}$, one can obtain accurate solution $\boldsymbol{\pi}^*$ via proposed KKT-multiplier regularization term $\mathcal{G}(\boldsymbol{\pi}, \boldsymbol{s}) = \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle$, which formulates Multiplier Regularized Optimal Transport (MROT):

$$\min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{G}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle + \eta_{\mathrm{Reg}} \mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}), \quad s.t.\ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{\beta}, \tag{12}$$

where $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi})$ denotes the regularization term on $\boldsymbol{\pi}$. Ideally, η_G should be set as a relatively large number. Meanwhile the dual form of MROT is given as:

$$\max_{\boldsymbol{\psi}, \boldsymbol{\phi}} L_{\mathrm{G}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\mathrm{Reg}} \mathcal{L}^*_{\mathrm{Reg}}((\psi_i + \phi_j - C_{ij} - \eta_G s_{ij})/\eta_{\mathrm{Reg}}), \tag{13}$$

Experiments (Synethic Data)

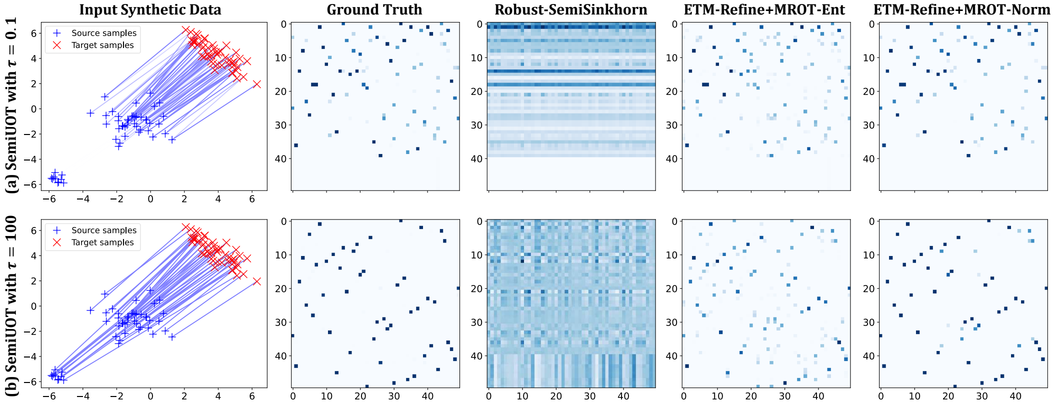


Figure: The SemiUOT matching solutions on π^* when $\tau = 0.1$ or $\tau = 100$ among the Robust-SemiSinkhorn and our proposed methods. We set $\eta_{\text{Reg}} = 0.1$ for entropy or L_2-norm regularization term.

Experiments (Synethic Data)

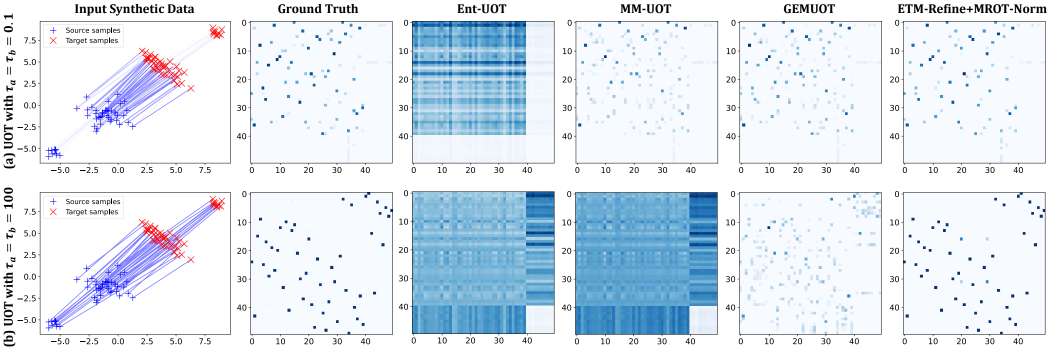


Figure: Results of π^* on UOT when $\tau_a = \tau_b = 0.1$ or $\tau_a = \tau_b = 100$ among Ent-UOT, MM-UOT, GEMUOT and ETM-Refine+MROT-Norm with $\eta_G = 10^2$ and $\eta_{\text{Reg}} = 0.1$.

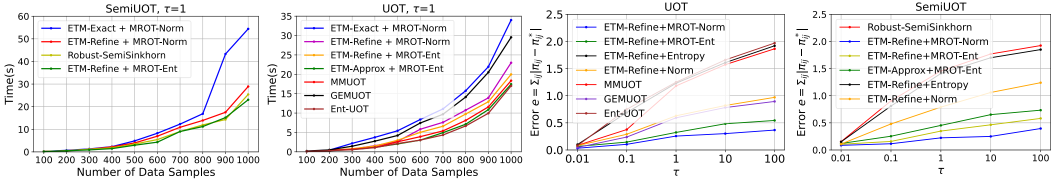
Experiments (Domain Adaptation)

Table: Experimental results on partial domain adaptation

Method	Avg on Office-Home	Avg on ImageCLEF
JUMBOT	75.5	87.6
m-POT	78.0	90.4
MOT	80.6	93.6
MOT+UOT(ETM+MROT-Ent)	81.8	94.0
MOT+UOT(ETM+MROT-Norm)	82.1	94.4
MOT+SemiUOT(ETM+MROT-Ent)	84.8	94.8
MOT+SemiUOT(ETM+MROT-Norm)	85.2	95.2

- Our proposed methods can achieve state-of-the-art performances on domain adaptation tasks (e.g., partial domain adaptation).

Experiments (Performance Analysis)

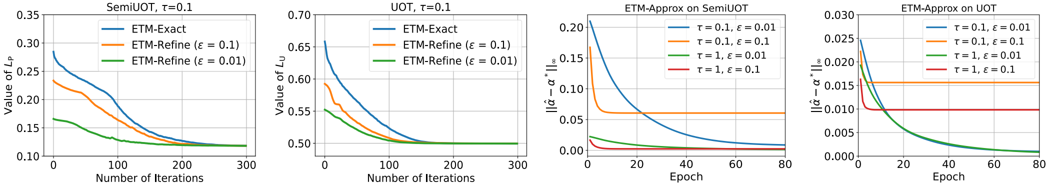


(a) The time consumption on SemiUOT **(b) The time consumption on UOT** **(c) The absolute error on SemiUOT** **(d) The absolute error on UOT**

Figure: The time consumption and computation error analysis on UOT and SemiUOT.

- ETM-Refine reaches a similar computation time with ETM-Approx, validating that it accelerates the process of finding the optimal $\boldsymbol{u}^*$ or $\boldsymbol{f}^*$ by utilizing $\widehat{\boldsymbol{u}}^*$ or $\widehat{\boldsymbol{f}}^*$.
- Meanwhile ETM-Refine with MROT-Norm can further reach more accurate solutions against MROT-Ent.

Experiments (Hyperparameters)



(a) The number iterations on training L_p **(b) The number iterations on training L_U** **(c) The computation error on SemiUOT** **(d) The computation error on UOT**

Figure: The effects on tuning different $\epsilon = \{0.01, 0.1\}$ on solving SemiUOT and UOT problems.

- Smaller ϵ could provide good approximation on UOT/SemiUOT, reducing the iteration steps for optimizing L_{U} and L_{P}.
- Larger values of ϵ may fail to reduce the computation error e_α when compared to smaller values of ϵ.

Experiments (Hyperparameters)

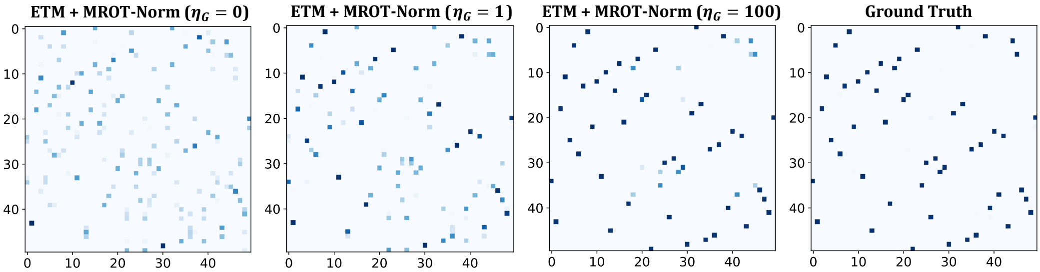


Figure: The matching results on ETM + MROT-Norm on SemiUOT with different values of $\eta_G = \{0, 1, 100\}$.

- We can conclude that choosing a larger value of η_G can fully utilize the knowledge provided by KKT multiplier and enhance the final results.

Conclusion

- In this talk, we mainly focus on the details of ETM-based methods for two problems, i.e., SemiUOT and UOT. After optimizing these problems, one can obtain the sample marginal probabilities and transfer SemiUOT/UOT into standard optimal transport problems.
- We first innovatively propose multiplier constraint terms to establish MROT (Multiplier Regularized Optimal Transport) for achieving more accurate results.
- We conduct extensive experiments on both synthetic and real-world datasets to evaluate the performance of proposed method.

Conclusion

Thanks for your attention

- Code demo is provided below. Try it if interested.

Colab Code for SemiUOT   Colab Code for UOT

- Contact Email: 21831010@zju.edu.cn or lwming95@gmail.com