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

Weiming Liu[1], **Xinting Liao**[2*], **Jun Dan**[2], **Fan Wang**[2], **Hua Yu**[3],

Junhao Dong[3], **Shunjie Dong**[4], **Lianyong Qi**[5,6], **Yew-Soon Ong**[3,7]

[1] ByteDance Inc., [2] Zhejiang University,
[3] College of Computing and Data Science, Nanyang Technological University, Singapore,
[4] Shanghai Jiao Tong University, [5] China University of Petroleum (East China),
[6]Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software,
[7] Centre for Frontier AI Research, Institute of High Performance Computing,
Agency for Science, Technology and Research, Singapore
lwming95@gmail.com, {xintingliao, danjun, fanwang97}@zju.edu.cn,
{junhao003, yu_hua, asysong}@ntu.edu.sg, sjdong@sjtu.edu.cn, lianyongqi@gmail.com

Abstract

Semi-Unbalanced Optimal Transport (SemiUOT) shows great promise in matching two discrete probability measures by relaxing one of the marginal constraints. Previous SemiUOT solvers often incorporate an entropy regularization term, inevitably resulting in inaccurate matching solutions. To address this issue, we propose an Equivalent Transformation Mechanism (ETM) approach to determine the marginal probability distributions of SemiUOT with KL divergence. Furthermore, we validate the generalization capability of ETM by exploiting the marginal probability distributions of Unbalanced Optimal Transport (UOT). ETM is able to determine the exact marginal probabilities of both SemiUOT and UOT, based on which we can transform the SemiUOT/UOT into classic Optimal Transport (OT) problem. Moreover, we propose a KKT-Multiplier regularization term combined with Multiplier Regularized Optimal Transport (MROT) to achieve more accurate matching results. We conduct extensive experiments to demonstrate the effectiveness of our proposed methods in addressing SemiUOT and UOT problems[2].

1 Introduction

Optimal Transport (OT) technique is a powerful tool for matching and discerning two distinct probability distributions. Nowadays, OT has multiple successful applications in traditional machine learning [37, 35, 112, 22, 85, 62], unsupervised clustering [4, 15], domain adaptation [25, 23, 84, 60, 61], diffusion [49, 56], generative modeling [50, 79, 100, 44] and many others. Nevertheless, directly solving OT distances could have relatively high computation cost with around super-cubic time. Although one can adopt entropy-based Sinkhorn algorithm [24] for solving OT efficiently, it still suffers from the dilemma of dense and inaccurate solutions [58, 63, 31]. Moreover, classic OT strictly assumes that the probability masses on both source and target domains should be equal. It further hurdles the generalization of OT when the data samples inherit noise or outliers.

[*] Corresponding author.

[2] The demo code is provided: https://github.com/XeniaLLL/ETM.git

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

Recently, Unbalanced Optimal Transport (UOT) [6, 19, 93, 92, 96, 59] and Semi-Unbalanced Optimal Transport (SemiUOT) [52] have become more attractive in adapting outliers since they allow relaxing marginal constraints for transportation results. This advantage makes SemiUOT and UOT powerfully applicable in transfer learning [101, 73, 80, 29, 26–28], computer vision [9, 30, 21, 74, 16, 67, 110], structure data exploration [90], natural language processing [3], and other areas. SemiUOT and UOT relax the strict OT mass equality constraints by introducing relaxation terms defined by Kullback-Leibler (KL) divergence [83], ℓ_1 norm [10], or ℓ_2 norm [8], whose effects are controlled by a coefficient τ. Meanwhile, KL divergence is the most commonly-used in real practice [95]. Previous solvers always involves extra regularization terms, i.e., entropy regularization term and proximal point term [34], for tackling Semi and UOT problems. Meanwhile adding additional entropy terms will lead to dense and inaccurate matching solutions. Latest, [17] and [78] further reconsider solving UOT problem with majority maximization algorithm without the requirements of regularization terms. However, these methods are sensitive to the choice of τ, i.e., providing sparse and accurate matching solutions when τ is small, but unsatisfying solutions when τ is large. Therefore, it is quite challenging to efficiently achieve accurate matching solutions for both SemiUOT and UOT problems.

In this paper, we propose a new method, i.e., ***Equivalent Transformation Mechanism*** (ETM), which directly finds the exact marginal probabilities of discrete SemiUOT and UOT with KL divergence. Specifically, ETM first finds the marginal probability distributions for SemiUOT and UOT problems based on Karush-Kuhn-Tucker (KKT) conditions and their dual forms. This induces a new insight for understanding SemiUOT and UOT problems, i.e., *We can transform SemiUOT and UOT problems into classic OT problems based on adjusting initial marginal weights via ETM*. We propose ETM-Refine to achieve exact marginal probabilities without needing overwhelming computation. Specifically, ETM-Refine first seeks the approximate results of marginal distributions via the fixed-point iteration with the smoothness function, and applies quasi-Newton based iterative methods to obtain exact results within quite a few steps. This competitively reduces the computation burden to obtain accurate matching results on SemiUOT/UOT. Beyond solving the marginal distribution, *we also discover that the KKT multipliers provide valuable guidance for addressing the OT problem, which is transformed from the SemiUOT or UOT problem with adjusted marginal weights.* Therefore we further propose ***Multiplier Regularized Optimal Transport*** (MROT) for achieving more sparse and accurate OT matching solutions. We summarize our contributions: (1) To our best knowledge, we first propose both exact and approximate solutions for ETM on 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. (2) We first innovatively propose multiplier constraint terms to establish MROT for achieving more accurate results. (3) We conduct extensive experiments on both synthetic and real-world datasets to evaluate the performance of proposed ETM.

2 Preliminary

We first provide a brief introduction of SemiUOT and UOT. Let us consider two sets of data samples $\boldsymbol{X} \in \mathbb{R}^{M\times D}$ and $\boldsymbol{Z} \in \mathbb{R}^{N\times D}$ in source and target domains, where M, N denote the number of samples and D denotes the data dimension. Samples in each domain have corresponding prior-given mass weights, i.e., $\boldsymbol{a} \in \mathbb{R}^{M\times 1}$ for source domain and $\boldsymbol{b} \in \mathbb{R}^{N\times 1}$ for target domain. The semi-unbalanced optimal transport problem is set to measure the minimum cost among data samples $\boldsymbol{X}$ and $\boldsymbol{Z}$, meanwhile filtering out the noise and outliers by relaxing one of marginal constraints:

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

where $\boldsymbol{C} \in \mathbb{R}^{M\times N}$ denotes the pairwise distance matrix. Meanwhile $\boldsymbol{\pi} \in \mathbb{R}^{M\times N}$ denotes the coupling matching matrix among the data samples $\boldsymbol{X}$ and $\boldsymbol{Z}$. τ denotes the hyper parameter and $\mathrm{KL}(\cdot)$ denotes the commonly-used KL divergence. SemiUOT relaxes the constraint $\boldsymbol{\pi}\mathbf{1}_N \neq \boldsymbol{a}$ and keep the constraint $\boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}$. Likewise, the unbalanced optimal transport problem [83] further relaxes both two marginal constraints, i.e., $\boldsymbol{\pi}\mathbf{1}_N \neq \boldsymbol{a}$ and $\boldsymbol{\pi}^\top \mathbf{1}_M \neq \boldsymbol{b}$ as shown:

$$\min_{\pi_{ij}\geq 0} J_{\mathrm{UOT}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau_a \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \tau_b \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}), \tag{2}$$

where τ_a and τ_b denote the balanced hyper parameters. Previous researches always add an entropy regularization term to enhance the scalability of solving $\boldsymbol{\pi}^*$ of SemiUOT and UOT. However, it still suffers from the dense and inaccurate solution dilemma in real practice.

3 Methodology

In this section, we will first introduce *Equivalent Transformation Mechanism* (ETM), which investigates the problem of SemiUOT/UOT from the perspective of marginal probability distribution. Then we illustrate the effect of *Multiplier Regularized Optimal Transport* (MROT) that finds out the accurate solutions of $\boldsymbol{\pi}^*$ for SemiUOT and UOT, with the guidance of *KKT-Multiplier Regularization*.

3.1 Equivalent Transformation Mechanism

ETM aims to seek the exact marginal distributions with satisfyingly efficient computation. Previous methods [83, 20] always directly adopted entropy-based regularization term into tackling SemiUOT and UOT problems. Although such approaches can provide fast computation speed, it will lead to relatively ambiguous and dense solutions that does not match most of the situations in real practice [55, 91]. In this section, we propose ETM-based methods to determine the marginal probabilities of source data samples in SemiUOT, accompanied by detailed illustrations. We then extend the ETM-based method to address the more complex UOT problem.

ETM for SemiUOT. To start with, we illustrate the ETM for transforming SemiUOT to classic OT, and introduce three types of ETM-based methods, i.e., ETM-Exact, ETM-Approx, and ETM-Refine, for determining the marginal probability distributions. Specifically, ETM-Exact directly computes the dual variables via iterative methods, e.g., LBFGS, achieving the exact results with an overwhelming computation burden. While ETM-Approx is a variant of ETM for SemiUOT by replacing the infimum with its smoothness approximation. Then we newly propose a fixed-point iteration method to solve the optimization problem efficiently. To further figure out the exact results, ETM-Refine applies ETM-Exact with quite a few iterations to solve the exact SemiUOT, by taking the approximate results as starting points. ETM-Refine shows competitive performance while maintaining efficient computation. By utilizing the methods above, one can transform SemiUOT into classic optimal transport problem by adjusting initial marginal weights. In the following, we will introduce the deduction and optimization details for the proposed ETM-based method on SemiUOT.

Proposition 1. (Principles of Equivalent Transformation Mechanism for SemiUOT) *Given SemiUOT with KL-Divergence J_{SemiUOT}, one can obtain its Fenchel-Lagrange multipliers form as:*

$$\min_{\boldsymbol{f},\boldsymbol{g},\zeta}\left[\tau\sum_{i=1}^{M}a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right)-\sum_{j=1}^{N}b_j(g_j-\zeta)\right]\qquad s.t.\ f_i+g_j+s_{ij}=C_{ij},\quad s_{ij}\geq 0. \tag{3}$$

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

$$\min_{\boldsymbol{\pi}\geq 0}\mathcal{J}_{\text{P}}=\langle\boldsymbol{C},\boldsymbol{\pi}\rangle,\qquad s.t.\ \boldsymbol{\pi}\mathbf{1}_N=\boldsymbol{a}\odot\exp\left(-\frac{\boldsymbol{f}^*+\zeta^*}{\tau}\right)=\boldsymbol{\alpha},\quad\boldsymbol{\pi}^\top\mathbf{1}_M=\boldsymbol{b}. \tag{4}$$

When $\tau\to\infty$, the source marginal probability is given as $\boldsymbol{\pi}\mathbf{1}_N=\omega_{\text{L}}\boldsymbol{a}$ and $\omega_{\text{L}}=\langle\boldsymbol{b},\mathbf{1}_N\rangle/\langle\boldsymbol{a},\mathbf{1}_M\rangle$.

The proof of Proposition 1 can be found in Appendix B. We can observe that transforming SemiUOT into classic OT is to elementally adjust the initial weights of data samples by $\exp(-(\boldsymbol{f}^*+\zeta^*/\tau))$. To further simplify the calculation by reducing variable $\boldsymbol{g}$, we set $g_j=\inf_{k\in[M]}(C_{kj}-f_k)$ according to the c-transform theorem [103]. Hence we only need to optimize $\boldsymbol{f}$ and ζ without additional constraints:

$$\min_{\boldsymbol{f},\zeta}L_{\text{P}}=\tau\sum_{i=1}^{M}a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right)-\sum_{j=1}^{N}\left[\inf_{k\in[M]}[C_{kj}-f_k]-\zeta\right]b_j, \tag{5}$$

We refer to L_{P} as the newly proposed *Exact SemiUOT Equation*. To solve this exact SemiUOT in Eq.(5), we initialize $\zeta=0$ for the optimization and introduce an iterative method derived from LBFGS [109]. Specifically, we first fix ζ then adopting LBFGS method to reach optimal results of $\boldsymbol{f}^\ell$ and $g_j^\ell=\inf_{k\in[M]}(C_{kj}-f_k^\ell)$ at the ℓ-th iteration. Then we optimize $\zeta=\tau[\log(\sum_{i=1}^{M}a_i\exp(-f_i^\ell/\tau))-\log(\sum_{j=1}^{N}b_j)]$ which is obtained by considering $\nabla_\zeta L_{\text{P}}=0$ and it guarantees $\sum_{i=1}^{M}a_ie^{-(f_i^\ell+\zeta)/\tau}=\sum_{j=1}^{N}b_j$. We iteratively update L_{U} to reach the optimal solutions on ζ^*, $\boldsymbol{f}^*$ and $\boldsymbol{g}^*$. We refer to the entire optimization scheme as ETM-Exact approach for Eq.(5).

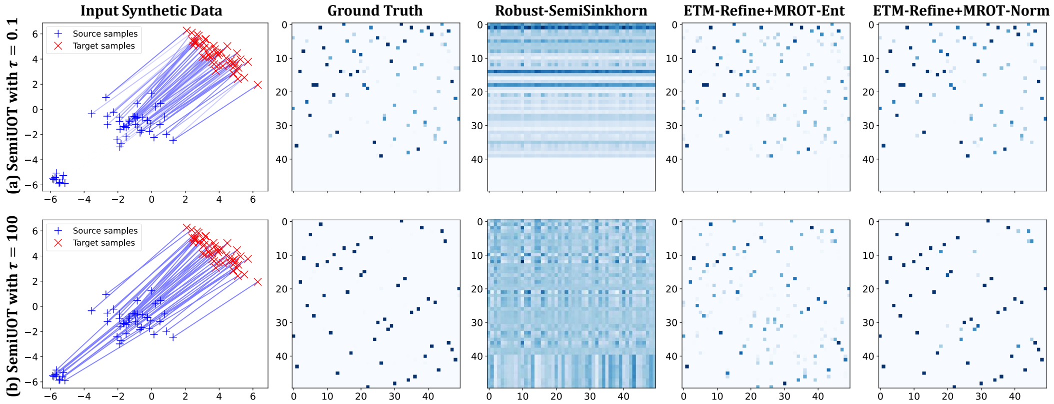


Figure 1: The SemiUOT matching solutions on $\boldsymbol{\pi}^*$ when $\tau = 0.1$ or $\tau = 100$ among the Robust-SemiSinkhorn [52] and our proposed ETM-Refine + MROT-Ent, ETM-Refine + MROT-Norm with $\eta_G = 10^2$ and $\epsilon = 10^{-2}$. We set $\eta_{\mathrm{Reg}} = 0.1$ for entropy or L_2-norm regularization term.

Although L_{P} is convex and has unique solutions, the presence of $\inf(\cdot)$ renders it a non-smooth function, preventing it from efficient optimization [1]. To further accelerate the optimization process, we consider making a smooth approximation on replacing $\inf(\cdot)$ with $\mathrm{LogSumExp}(\cdot)$ as $\inf_{k\in[M]}[C_{kj} - f_k] \approx -\epsilon\log[\sum_{k=1}^M e^{\frac{f_k - C_{kj}}{\epsilon}}]$ since $|\epsilon\log[\sum_{k=1}^M e^{\frac{f_k - C_{kj}}{\epsilon}}] - \sup_{k\in[M]}[f_k - C_{kj}]| \leq \epsilon\log M$ following [75]. Here $\epsilon > 0$ denotes the balanced hyperparameter between accuracy and function smoothness. Smaller ϵ (e.g., $\epsilon \to 0$) could lead to more accurate but less smooth solutions. Then we can obtain the proposed *Approximate SemiUOT Equation* as $\widehat{L}_{\mathrm{P}}$ by replacing $\inf(\cdot)$ with the smoothness term for $\widehat{f}$:

$$\min_{\widehat{\boldsymbol{f}},\zeta} \widehat{L}_{\mathrm{P}} = \tau\sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{f}_i + \zeta}{\tau}\right) + \sum_{j=1}^{N} b_j\left[\epsilon\log\left[\sum_{k=1}^{M}\exp\left(\frac{\widehat{f}_k - C_{kj}}{\epsilon}\right)\right] + \zeta\right]. \tag{6}$$

Proposition 2. (Calculation for Approximate SemiUOT Equation) *Given Approximate SemiUOT equation $\widehat{L}_{\mathrm{P}}$, it can be optimized via Equivalent Transformation Mechanism with Approximation (ETM-Approx). That is, ETM-Approx aims to solve the following equation for each $\widehat{f}_s$:*

$$\frac{\partial\widehat{L}_{\mathrm{P}}}{\partial\widehat{f}_s} = -a_s\exp\left(-\frac{\widehat{f}_s + \zeta}{\tau}\right) + \exp\left(\frac{\widehat{f}_s}{\epsilon}\right)\sum_{j=1}^{N}\left[\frac{b_j\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\sum_{k=1}^{M}\exp\left(\frac{\widehat{f}_k - C_{kj}}{\epsilon}\right)}\right] = 0. \tag{7}$$

Specifically, we can adopt fixed-point iteration method for solving Eq.(7) *at the ℓ-th iteration:*

$$\widehat{f}_s^{\ell+1} = \nu\left[\log\left(a_s\exp\left(-\frac{\zeta}{\tau}\right)\right) - \log\left[\sum_{j=1}^{N}\left(\frac{b_j}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})}\exp\left(-\frac{C_{sj}}{\epsilon}\right)\right)\right]\right], \quad \forall s \in [1, M], \tag{8}$$

where $\nu = \tau\epsilon/(\tau + \epsilon)$ for simplification and $\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})$ denotes the corresponding calculation as shown $\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell}) = \sum_{k=1}^{M}\exp((\widehat{f}_k^{\ell} - C_{kj})/\epsilon)$. The proposed procedure can converge with a theoretical guarantee. Finally, updating variable ζ by further considering $\nabla_{\zeta}\widehat{L}_{\mathrm{P}} = 0$ via $\zeta = \tau[\log(\sum_{i=1}^{M} a_i\exp(-\widehat{f}_i^*/\tau)) - \log(\sum_{j=1}^{N} b_j)]$, finally achieving optimal results $\widehat{\boldsymbol{f}}^*$ and ζ^*.

The proof of Proposition 2 can be found in Appendix C. Generally, Proposition 2 outlines the optimization procedure using the newly proposed ETM-Approx approach for addressing the Approximate Semi-UOT Equation. We can observe that the ETM-Approx approach is easy to compute and implement, while avoiding complex calculations (e.g., searching the descent direction and finding the step size) and not requiring a large amount of storage space against previous methods. Therefore, the ETM-Approx approach is an efficient method for determining the result of $\widehat{\boldsymbol{f}}^*$ and $\widehat{g}_j^* = -\epsilon\log[\sum_{k=1}^{M}\exp((\widehat{f}_k^* - C_{kj})/\epsilon)]$, transforming SemiUOT into the optimal transport problem.

Remark 1. *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.*

Moreover, we can finally figure out the exact optimal solution $\boldsymbol{f}^*$ via the approximate optimal solution $\widehat{\boldsymbol{f}}^*$ on $\widehat{L}_{\text{P}}$ using Proposition 2. That is, if we directly optimize L_{P} from a randomly initial point, we could spend more time on quasi-Newton gradient descent (e.g., LBFGS) for reaching $\boldsymbol{f}^*$. Since $\widehat{\boldsymbol{f}}^*$ is close to $\boldsymbol{f}^*$, it should be more efficient to use $\widehat{\boldsymbol{f}}^*$ as the initial guess for optimizing $\boldsymbol{f}^*$ via ETM-Exact which has super-linear convergence rate [46, 47, 86, 109, 39]. And we regard the whole procedure as ETM-Refine method with time complexity of $\mathcal{O}(NM\log(1/\varepsilon_{\text{err}}) + NM(\log M)d_T)$ where d_T denotes the number of iterations. ETM-Refine utilizes the strength of ETM-Approx in efficient computation for the exact results. In summary, we can utilize ETM-based methods to transform SemiUOT into classic OT problem. We illustrate the optimization details in Alg.1 and Appendix D.

Algorithm 1 The algorithm of ETM-Based method on SemiUOT

```
Input: $\boldsymbol{C}$: cost matrix; $\boldsymbol{a}, \boldsymbol{b}$: initial marginal probability; $\tau, \epsilon$: Hyper parameters.
  Randomly initialize the value of $\boldsymbol{f}^{\text{init}}$.
  Choose ETM-Exact, ETM-Approx or ETM-Refine on SemiUOT for optimization.
(1) Function: ETM-Exact on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \boldsymbol{f}^{t=0} = \boldsymbol{f}^{\text{init}}$)
  Optimize $\boldsymbol{f}$ via L-BFGS algorithm on $L_{\text{P}}$.
  Optimize $\boldsymbol{g}$ via $g_j = \inf_{k\in[M]}(C_{kj} - f_k^t)$.
  Optimize $\zeta$ via $\zeta = \tau[\log(\sum_{i=1}^M a_i \exp(-f_i/\tau)) - \log(\sum_{j=1}^N b_j)]$.
Return: The optimal solutions of $\boldsymbol{f}^*$, $\boldsymbol{g}^*$ and $\zeta^*$.
(2) Function: ETM-Approx on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \widehat{\boldsymbol{f}}^{t=0} = \boldsymbol{f}^{\text{init}}$)
  Optimize $\widehat{\boldsymbol{f}}$ via Proposition 2 on $\widehat{L}_{\text{P}}$.
  Optimize $\widehat{\boldsymbol{g}}$ via $\widehat{g}_j = -\epsilon\log[\sum_{k=1}^M \exp((\widehat{f}_k - C_{kj})/\epsilon)]$.
  Optimize $\zeta$ via $\zeta = \tau[\log(\sum_{i=1}^M a_i \exp(-\widehat{f}_i/\tau)) - \log(\sum_{j=1}^N b_j)]$.
Return: The optimal solutions of $\widehat{\boldsymbol{f}}^*$, $\widehat{\boldsymbol{g}}^*$ and $\zeta^*$.
(3) Function: ETM-Refine on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \widehat{\boldsymbol{f}}^{t=0} = \boldsymbol{f}^{\text{init}}$)
  Obtain $\widehat{\boldsymbol{f}}^*$ = ETM-Approx on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \widehat{\boldsymbol{f}}^{t=0} = \boldsymbol{f}^{\text{init}}$).
  Obtain $\boldsymbol{f}^*$ = ETM-Exact on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \boldsymbol{f}^{t=0} = \widehat{\boldsymbol{f}}^*$).
Return: The optimal solutions of $\boldsymbol{f}^*$, $\boldsymbol{g}^*$ and $\zeta^*$.
```

ETM for UOT. We have obtained the marginal probability of SemiUOT via tackling Proposition 1 with proposed ETM-based method. In this section, we will further extend ETM for solving the marginal probability on UOT, which is also a commonly existing optimization problem.

Proposition 3. (Principles of Equivalent Transformation Mechanism for UOT) *Given UOT with KL-Divergence* J_{UOT}, *its Fenchel-Lagrange multipliers form is given:*

$$\min_{\boldsymbol{u},\boldsymbol{v},\zeta}\left[\tau_a\sum_{i=1}^M a_i\exp\left(-\frac{u_i+\zeta}{\tau_a}\right) + \tau_b\sum_{j=1}^N b_j\exp\left(-\frac{v_j-\zeta}{\tau_b}\right)\right], \quad s.t. \begin{cases} u_i + v_j + s_{ij} = C_{ij}, \\ s_{ij} \geq 0, \end{cases} \tag{9}$$

where $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{s}$ *and* ζ *denote Lagrange multipliers. Moreover, UOT problem can also be transformed into classic optimal transport as follows:*

$$\min_{\boldsymbol{\pi}\geq 0}\mathcal{J}_{\text{U}} = \langle\boldsymbol{C},\boldsymbol{\pi}\rangle, \quad s.t.\ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a}\odot\exp\left(-\frac{\boldsymbol{u}^*+\zeta^*}{\tau_a}\right) = \boldsymbol{\alpha},\ \boldsymbol{\pi}^\top\mathbf{1}_M = \boldsymbol{b}\odot\exp\left(-\frac{\boldsymbol{v}^*-\zeta^*}{\tau_b}\right) = \boldsymbol{\beta}. \tag{10}$$

Note that when $\tau_a, \tau_b \to \infty$, *the source and target marginal probabilities can be determined as* $\boldsymbol{\pi}\mathbf{1}_N = \sqrt{\omega_{\text{L}}}\boldsymbol{a}$ *and* $\boldsymbol{\pi}^\top\mathbf{1}_M = \boldsymbol{b}/\sqrt{\omega_{\text{L}}}$ *where* $\omega_{\text{L}} = \langle\boldsymbol{b},\mathbf{1}_N\rangle/\langle\boldsymbol{a},\mathbf{1}_M\rangle$ *respectively.*

The proof of Proposition 3 can be found in Appendix E. Likewise, we set $v_j = \inf_{k\in[M]}(C_{kj} - u_k)$ by the c-transform theorem [103] to simplify the calculation. Hence we obtain *Exact UOT Equation*:

$$\min_{\boldsymbol{u},\zeta} L_{\text{U}} = \tau_a\sum_{i=1}^M a_i\exp\left(-\frac{u_i+\zeta}{\tau_a}\right) + \tau_b\exp\left(\frac{\zeta}{\tau_b}\right)\sum_{j=1}^N b_j\exp\left(\frac{\sup_{k\in[M]}(u_k - C_{kj})}{\tau_b}\right). \tag{11}$$

We first fix ζ then adopting LBFGS approach to optimize L_{U} on $\boldsymbol{u}$. Then we futher optimize $\zeta = \kappa[\log(\sum_{i=1}^M a_i\exp(-u_i^\ell/\tau_a)) - \log(\sum_{j=1}^N b_j\exp(-v_j^\ell/\tau_b))]$ at the ℓ-th iteration where $v_j^\ell =$

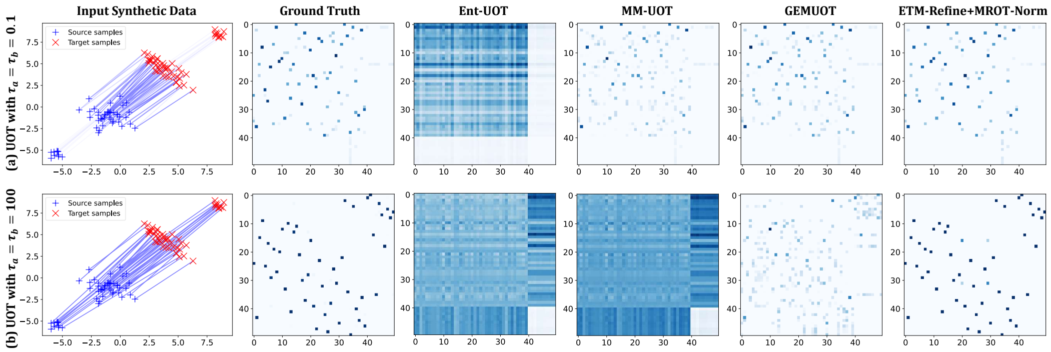


Figure 2: Results of $\boldsymbol{\pi}^*$ on UOT when $\tau_a = \tau_b = 0.1$ or $\tau_a = \tau_b = 100$ among Ent-UOT [83], MM-UOT [17], GEMUOT [78] and ETM-Refine+MROT-Norm with $\eta_G = 10^2$ and $\eta_{\text{Reg}} = 0.1$.

$\inf_{k\in[M]}(C_{kj} - u_k^\ell)$ and $\kappa = \tau_a\tau_b/(\tau_a + \tau_b)$ by considering $\nabla_\zeta L_{\text{U}} = 0$. Here we regard the above process as the ETM-Exact approach for solving UOT problem. Note that the non-smooth function $\sup(\cdot)$ will result in inefficient optimization. However, if we directly apply a similar function approximation to replace $\sup(\cdot)$ following Eq.(6), the optimization problem becomes quite complex, making it relatively difficult to determine the iterative solutions. Meanwhile, Proposition 2 enlightens us with a completely new ETM-Approx approach for optimizing UOT.

Optimization 1. (Calculation of ETM-Approx approach for UOT) Since the optimization problem in Eq.(9) is convex, we can also utilize block gradient descent to optimize the problem. Specifically, we first fix $\widehat{v}^l$ and optimize variable $\widehat{u}^l$ at the l-th iteration by replacing the original marginal probability $\boldsymbol{b}$ in Eq.(6) with $\boldsymbol{b} \odot \exp(-(\widehat{\boldsymbol{v}} - \zeta)/\tau_b) = \boldsymbol{\beta}$ accordingly to transform UOT into SemiUOT problem:

$$\min_{\widehat{\boldsymbol{u}}} \widehat{L}_{\text{U}}^u = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i + \zeta}{\tau_a}\right) + \sum_{j=1}^{N} \beta_j \left[\epsilon \log\left[\sum_{k=1}^{M} \exp\left(\frac{\widehat{u}_k - C_{kj}}{\epsilon}\right)\right] + \zeta\right]. \tag{12}$$

It is equivalent to solve the equation by taking the differentiation w.r.t. on $\widehat{u}_s$ over $\widehat{L}_{\text{U}}^u$ and set it 0:

$$\frac{\partial \widehat{L}_{\text{U}}^u}{\partial \widehat{u}_s} = -a_s \exp\left(-\frac{\widehat{u}_s + \zeta}{\tau_a}\right) + \exp\left(\frac{\widehat{u}_s}{\epsilon}\right) \sum_{j=1}^{N} \left[\frac{\beta_j \exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{\widehat{u}_k - C_{kj}}{\epsilon}\right)}\right] = 0. \tag{13}$$

Obviously, it is equivalent to replace $\boldsymbol{b}$ with $\boldsymbol{\beta}$ in Eq.(7) for solving Eq.(13). Then we can utilize the iteration step shown in Eq.(8) to obtain $\widehat{\boldsymbol{u}}^{l+1}$. After that we fix $\widehat{\boldsymbol{u}}^{l+1}$ and optimize variable $\widehat{\boldsymbol{v}}^{l+1}$ via $\widehat{v}_j^{l+1} = -\epsilon \log[\sum_{k=1}^{M} \exp((\widehat{u}_k^{l+1} - C_{kj})/\epsilon)]$. We can achieve the optimal solution on $\widehat{\boldsymbol{u}}^*$ and $\widehat{\boldsymbol{v}}^*$ via iteratively computing via the above procedure accordingly. Finally, we update variable ζ via considering $\zeta = (\tau_a\tau_b/(\tau_a + \tau_b))[\log(\sum_{i=1}^{M} a_i \exp(-\widehat{u}_i^*/\tau_a)) - \log(\sum_{j=1}^{N} b_j \exp(-\widehat{v}_j^*/\tau_b))]$. Due to the space limits, the deduction details are provided in Appendix F. Moreover, it has the time complexity of $\mathcal{O}(NM\log(1/\varepsilon_{\text{err}}))$ where $\varepsilon_{\text{err}} = ||\widehat{\boldsymbol{u}} - \widehat{\boldsymbol{u}}^*||_\infty$ with the linear convergence rate.

In summary, Optimization 1 for solving the UOT can be seen as an extension of Proposition 2 applied to SemiUOT, demonstrating the robust generalization capability of the proposed ETM method. Likewise, one can utilize $\widehat{\boldsymbol{u}}^*$ and $\widehat{\boldsymbol{v}}^*$ as the initial guess for solving Exact UOT Equation on Eq.(11) via ETM-Refine. Hence, ETM-based methods (i.e., ETM-Exact, ETM-Approx and ETM-Refine) transform UOT into classic optimal transport via computing the exact marginal distributions.

3.2 Multiplier Regularized Optimal Transport Induced by KKT-Multiplier Regularization

According to the Proposition 1-3 that discussed in Section 3.1, we have figured out the marginal probability distributions on both SemiUOT and UOT with commonly used KL Divergence via proposed ETM-based methods. Motivated by this, we can observe that the core mechanism of UOT/SemiUOT is carefully reweighting the weights of different samples accordingly. If the samples are noise or outliers, the corresponding weights will be much smaller than the corresponding weights among similar data samples. Therefore, UOT/SemiUOT has better adaptability than traditional OT that commonly treats all data samples equally. In this section, we will further exploit the matching results of $\boldsymbol{\pi}$ for SemiUOT and UOT using the following corollary:

Corollary 1. *Given any UOT/SemiUOT with KL divergence, we can transfer the original problem into classic optimal transport via adopting proposed ETM approach flexibly. We can further utilize existing OT solver for solving* $\boldsymbol{\pi}^*$ *as* $(\mathrm{UOT}, \mathrm{SemiUOT}) \xrightarrow{\text{ETM Method}} \mathrm{OT} \xrightarrow{\text{OT Solver}} \boldsymbol{\pi}^*$.

This observation provides us with entirely new unified insight into solving the matching results of $\boldsymbol{\pi}^*$ for SemiUOT and UOT. It is essential to utilize the proposed ETM-based method, as it offers a variety of OT solvers that yield more efficient and accurate results than directly optimizing UOT or SemiUOT. In general, one can further adopt Sinkhorn [24, 14], ℓ_2-norm term [8] or some other sparsification OT solver [58, 41] with different regularization terms to achieve the transportation $\boldsymbol{\pi}$.

Although some OT solvers (e.g., Sinkhorn [24]) could figure out $\boldsymbol{\pi}$ efficiently comparing to the linear programming with cubic time complexity [48], they often provide ambiguous results that may deviate significantly from the correct solutions [72, 58]. Hence it remains a challenge to efficiently find an accurate result for $\boldsymbol{\pi}^*$. Recalling the whole process of ETM method, we not only obtain the marginal probabilities, but also derive multipliers $\boldsymbol{s}$ which can be further utilized as guidance.

Corollary 2. *Given the optimal* $\boldsymbol{u}^*$ *and* $\boldsymbol{v}^*$ *in UOT via ETM-based method, one can obtain* $\boldsymbol{s}$ *on UOT by* $s_{ij} = \max(0, C_{ij} - u_i^* - v_j^*)$. *Likewise, the multipliers* $\boldsymbol{s}$ *on SemiUOT can be obtained via ETM-based method as* $s_{ij} = \max(0, C_{ij} - f_i^* - g_j^*)$. *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.*

The Corollary 2 demonstrates that the value of π_{ij} can be reflected via s_{ij}. This observation inspires us to further utilize such useful information in accurately calculating matching results $\boldsymbol{\pi}^*$.

Proposition 4. (The Definition and Usage of KKT-Multiplier Regularization) *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):*

$$\min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{G}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle + \eta_{\mathrm{Reg}} \mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}), \quad s.t.\ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{\beta}, \tag{14}$$

where $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi})$ *denotes the regularization term on* $\boldsymbol{\pi}$. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ *denote the final marginal probabilities obtained by ETM-based method, while* η_{Reg} *and* η_G *denote the hyperparameters. Ideally,* η_G *should be set as a relatively large number. Meanwhile the dual form of MROT is given as:*

$$\max_{\boldsymbol{\psi}, \boldsymbol{\phi}} L_{\mathrm{G}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\mathrm{Reg}} \mathcal{L}^*_{\mathrm{Reg}}((\psi_i + \phi_j - \widetilde{C}_{ij})/\eta_{\mathrm{Reg}}), \tag{15}$$

where $\widetilde{C}_{ij} = C_{ij} + \eta_G s_{ij}$, $\boldsymbol{\phi}$ *and* $\boldsymbol{\psi}$ *denote the Lagrange multipliers for MROT.* $\mathcal{L}^*_{\mathrm{Reg}}(\cdot)$ *denotes the conjugate function of* $\mathcal{L}_{\mathrm{Reg}}(\cdot)$ *and one can figure out the matching results of* $\boldsymbol{\pi}$ *via solving the following equation* $\nabla_{\pi_{ij}} \mathcal{L}_{\mathrm{Reg}}(\pi_{ij}) = (\psi_i + \phi_j - \widetilde{C}_{ij})/\eta_{\mathrm{Reg}}$.

We provide the deduction of MROT in Appendix H. That is, minimizing $s_{ij}\pi_{ij}$ to 0 could result in $s_{ij}\pi_{ij} = 0$, which not only aligns with the KKT complementary condition, but also reweights the matching for more accurate results. Generally, MROT is orthogonal to adopting different kinds of regularization term $\mathcal{L}_{\mathrm{Reg}}(\cdot)$ for efficient optimization. For instance, one can use the widely adopted entropy regularization term $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}) = -\langle \boldsymbol{\pi}, \log(\boldsymbol{\pi}) - 1 \rangle$ to formulate Entropic Multiplier Regularized Optimal Transport (MROT-Ent), whose matching results satisfy $\pi_{ij} = \exp(-\eta_G s_{ij}/\eta_{\mathrm{Reg}}) \exp((\psi_i + \phi_j - C_{ij})/\eta_{\mathrm{Reg}})$. Obviously, involving the multipliers information $\boldsymbol{s}$ has achieved more accurate solutions. Specifically, the non-matching samples pairs will get lower value on π_{ij} since $\mathcal{G}(\boldsymbol{\pi}, \boldsymbol{s}) = \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle$ avoids rigorous results. Otherwise, the matching results on π_{ij} will mainly be determined by the transportation cost. Similarly, one can also adopt L_2-norm regularization term $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}) = \frac{1}{2}\langle \boldsymbol{\pi}, \boldsymbol{\pi} \rangle$ to formulate Sparse Multiplier Regularized Optimal Transport (MROT-Norm) with similar characteristics. The time complexity of MROT depends on the regularization term $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi})$, that is, MROT-Ent and MROT-Norm have a complexity of $\mathcal{O}(NMd_\pi)$ where d_π is the number of iterations. In conclusion, we can integrate the ETM-based methods with MROT method to solve the SemiUOT and UOT problems, achieving accurate results for both marginal probabilities and the matching solution π_{ij}.

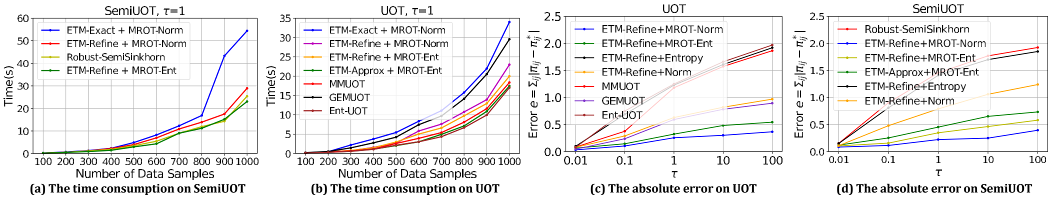
4 Experiments

4.1 Experimental setup

Datasets. We conduct experiments on both synthetic and real-world datasets to evaluate the methods. **(1) Synthetic Datasets.** We first conduct the experiments on the synthetic datasets. That is, we set

Table 1: Classification accuracy (%) on *Office-Home* for UDA and Partial UDA

Method for UDA	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
ResNet [43]	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
DeepJDOT [25]	50.7	68.6	74.4	59.9	65.8	68.1	55.2	46.3	73.8	66.0	54.9	78.3	63.5
ROT [5]	47.2	71.8	76.4	58.6	68.1	70.2	56.5	45.0	75.8	69.4	52.1	80.6	64.3
JUMBOT [34]	55.2	75.5	80.8	65.5	74.4	74.9	65.2	52.7	79.2	73.0	59.9	83.4	70.0
JUMBOT + UOT(MM-UOT)	56.3	76.2	81.6	66.0	75.3	75.1	66.4	52.9	79.2	73.8	60.7	84.1	70.6
JUMBOT + UOT(GEMUOT)	57.5	77.4	82.7	67.2	76.0	75.6	66.1	54.5	80.5	74.9	61.8	85.2	71.6
JUMBOT + UOT(ℓ_2-Norm Solver)	57.0	76.7	81.8	66.1	74.5	75.5	65.9	53.4	79.6	74.2	60.6	83.3	70.7
JUMBOT + UOT(Sparse Solver)	57.8	77.1	82.3	66.7	76.2	75.8	67.0	54.1	80.7	75.4	61.3	84.6	71.5
JUMBOT + UOT(ETM-Refine + MROT-Ent)	59.0	78.5	83.4	**68.7**	77.1	77.6	68.3	57.2	82.4	76.2	**62.5**	86.4	73.1
JUMBOT + UOT(ETM-Refine + MROT-Norm)	**59.4**	**78.7**	**84.1**	68.5	**77.3**	**78.5**	**68.6**	**57.9**	**82.8**	**76.3**	**62.5**	**86.5**	**73.4**
Method for Partial UDA	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
ResNet [43]	46.3	67.5	75.9	59.1	59.9	62.7	58.2	41.8	74.9	67.4	48.2	74.2	61.4
ETN [12]	59.2	77.0	79.5	62.9	65.7	75.0	68.3	55.4	84.4	75.7	57.7	84.5	70.5
JUMBOT [34]	62.7	77.5	84.4	76.0	73.3	80.5	74.7	60.8	85.1	80.2	66.5	83.9	75.5
AR [42]	67.4	85.3	90.0	77.3	70.6	85.2	79.0	64.8	89.5	80.4	66.2	86.4	78.3
m-POT [77]	64.6	80.6	87.2	76.4	77.6	83.6	77.1	63.7	87.6	81.4	68.5	87.4	78.0
MOT [65]	63.1	86.1	92.3	78.7	85.4	89.6	79.8	62.3	89.7	83.8	67.0	89.6	80.6
MOT + UOT(ETM + MROT-Ent)	65.2	87.3	92.8	79.5	86.4	91.0	80.8	64.5	90.7	84.5	67.9	90.4	81.8
MOT + UOT(ETM + MROT-Norm)	65.8	88.0	93.1	79.9	86.2	91.3	81.4	64.9	91.2	84.9	68.3	90.7	82.1
MOT + SemiUOT(Robust-SemiSinkhorn)	66.0	88.2	93.0	80.5	86.8	91.5	81.3	65.2	91.6	85.2	68.5	90.9	82.4
MOT + SemiUOT(ETM-Refine + MROT-Ent)	68.6	90.4	94.2	83.7	89.5	93.9	83.5	67.4	93.9	88.4	**71.8**	92.1	84.8
MOT + SemiUOT(ETM-Refine + MROT-Norm)	**69.1**	**90.7**	**94.6**	**84.0**	**90.3**	**94.0**	**83.8**	**67.9**	**94.4**	**88.5**	71.3	**93.6**	**85.2**



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

Figure 3: The time consumption and computation error analysis on UOT and SemiUOT.

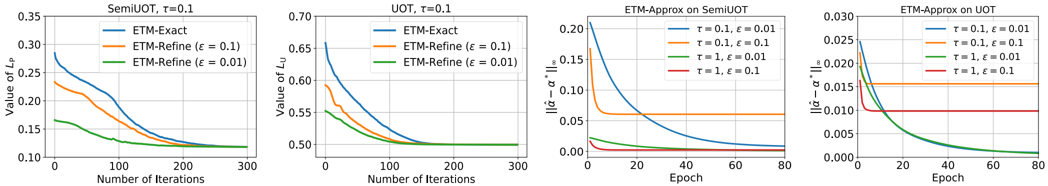
the source and target domain distributions as $\mathbb{P}_X = \mathcal{N}\left(\begin{bmatrix}-1\\-1\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$ and $\mathbb{P}_Z = \mathcal{N}\left(\begin{bmatrix}4\\4\end{bmatrix}, \begin{bmatrix}1 & -0.8\\-0.8 & 1\end{bmatrix}\right)$ following previous works [36, 17]. For the SemiUOT scenario, we set target distribution as $\mathbb{P}_Z$ then we sample 80% data from $\mathbb{P}_X$ with 20% outlier data to generate the source distribution. For the UOT scenario, we sample 80% data from $\mathbb{P}_X$ and $\mathbb{P}_Z$ accordingly while randomly sampling 20% outlier data for both $\mathbb{P}_X$ and $\mathbb{P}_Z$. **(2) Real-world Datasets.** We conduct the domain adaptation tasks on *Office-31* [88], *Office-Home* [102], and *ImageCLEF* [13]. More details are provided in Appendix I,J,K.

Baselines. We first compare the proposed ETM-Refine with MROT method with the following state-of-the-art UOT/SemiUOT solvers on the synthetic datasets. (1) **Ent-UOT** [83] utilizes the entropy regularization term on tackling UOT problem. (2) **MM-UOT** [17] adopts majority maximization algorithm for solving UOT. (3) **GEMUOT** [78] adopts ℓ_2-norm term for reaching transport solutions on UOT which is the state-of-the-art approach. (4) **Robust-SemiSinkhorn** [52] adopts the entropy regularization term for solving SemiUOT problem. We also involve **DeepJDOT** [25], **ROT** [5], **JUMBOT** [34], **ETN** [12], **AR** [42], **m-POT** [77], **MOT** [65] as the model baselines for the real-world domain adaptation task. These model details are provided in Appendix.J.

Implemented details. For both synthetic and real-world datasets, we set $\epsilon = 0.01$ on both $\widehat{L}_{\mathrm{U}}$ and $\widehat{L}_{\mathrm{P}}$. We set $\eta_G = 10^2$ and $\eta_{\mathrm{Reg}} = 0.1$ for MROT in the calculation. The initial value of $\widehat{\boldsymbol{u}}^{(0)}$ and $\widehat{\boldsymbol{f}}^{(0)}$ as set as zero vectors. The initial sample weights are set to be equal, i.e., $a_i = 1/M$ and $b_j = 1/N$. And we adopt square Euclidean distance for the cost C_{ij}. Besides, we adopt the *same* framework and experimental settings of the UDA model JUMBOT [34] for unsupervised domain adaptation and the partial UDA model MOT [65] for partial unsupervised domain adaptation with the fair comparison. For all the experiments, we perform five random experiments and report the average results.

4.2 Performance and Extensive Analysis on Synthetic and Real-World Datasets

Performance on Synthetic Datasets. We sample 50 data samples on both source and target distributions for finding $\boldsymbol{\pi}^*$ on UOT/SemiUOT. We first set $\tau = \{0.1, 100\}$ on SemiUOT and the matching solutions are shown in Fig.1(a)-(b). Note that we randomly sample 20% of noise in the source datasets. We can observe that previous method **Robust-SemiSinkhorn** could lead to ambiguous matching results. Our proposed ETM-Refine with MROT+Ent can reach relatively accurate results even if τ is large (e.g., $\tau = 100$). More importantly, ETM-Refine with MROT-Norm can achieve more precise results comparing with ETM-Refine with MROT-Ent shown in Fig.1. Then we also set $\tau_a = \tau_b = \{0.1, 100\}$ on UOT and the matching solutions are shown in Fig.2(a)-(b). From that we can observe: (1) **Ent-UOT** could merely provide dense matching solutions which are inaccurate. (2) **MM-UOT** obtains relatively accurate solutions when τ is small. However, **MM-UOT**



(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 4: The effects on tuning different $\epsilon = \{0.01, 0.1\}$ with the loss descent curve and computation error $e_\alpha = ||\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*||_\infty$ for ETM-Approx method on solving SemiUOT and UOT problems.

Table 2: Classification accuracy (%) on *Office-31* and *ImageCLEF* for partial UDA

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg	I→P	P→I	I→C	C→I	C→P	P→C	Avg
ResNet [43]	75.6	96.3	98.1	83.4	83.9	85.0	87.1	78.3	86.9	91.0	84.3	72.5	91.5	84.1
ETN [12]	84.7	97.4	99.2	91.3	90.2	92.8	92.6	79.6	88.5	92.9	87.2	74.1	93.4	86.0
JUMBOT [34]	90.2	98.9	99.3	94.5	93.8	93.4	95.0	80.1	91.3	93.6	90.9	75.7	94.2	87.6
AR [42]	93.5	**100.0**	99.7	96.8	95.5	96.0	96.9	83.1	92.8	94.5	92.4	76.3	95.0	89.0
m-POT [77]	96.2	99.5	**100.0**	97.6	94.4	95.3	97.2	82.6	94.1	96.3	94.7	78.5	96.2	90.4
MOT [65]	99.3	**100.0**	**100.0**	98.7	96.1	96.4	98.4	87.7	95.0	98.0	95.0	87.0	98.7	93.6
MOT + UOT(ETM-Refine + MROT-Ent)	99.4	**100.0**	**100.0**	98.9	96.8	97.3	98.7	88.3	95.6	98.4	95.3	87.6	99.0	94.0
MOT + UOT(ETM-Refine + MROT-Norm)	99.6	**100.0**	**100.0**	99.2	97.3	97.7	99.0	88.7	95.9	98.7	95.8	88.0	99.1	94.4
MOT + SemiUOT(ETM-Refine + MROT-Ent)	99.7	**100.0**	**100.0**	99.4	97.8	98.4	99.2	89.1	96.2	99.2	96.1	88.5	99.4	94.8
MOT + SemiUOT(ETM-Refine + MROT-Norm)	**99.8**	**100.0**	**100.0**	**99.7**	**98.4**	**98.8**	**99.5**	**89.6**	**96.7**	**99.4**	**96.5**	**89.1**	**99.6**	**95.2**

cannot better handle the case when τ is large (e.g., $\tau = 100$) due to the deterioration of majority maximization algorithm. (3) **GEMUOT** can even reach more sparse matching solution against **Ent-UOT** and **MM-UOT** with the aid of ℓ_2-norm term. However, the matching results obtained from **GEMUOT** remain coarse and ambiguous, especially when τ is large. (4) Benefiting from multipliers $\boldsymbol{s}$, ETM-Refine with MROT-Norm achieves the most accurate results among existing UOT methods.

Performance on Real-World Datasets. We further conduct the experiments on the real-world datasets to validate our proposed method. The experimental UDA task results on *Office-Home* are shown in Table.1. We also directly adopt ℓ_2-norm [8] and sparse solver [58] on solving $\mathcal{J}_{\mathrm{U}}$. Meanwhile, our proposed ETM-Refine with MROT-Norm obtains the best performance, which indicates its efficacy for finding more accurate matching results on UDA. Then we adopt the same experimental setting as **MOT** to evaluate the partial UDA task where target label space is a subset of source label space. This makes it more challenging than classic UDA task [11, 66]. The partial UDA results on *Office-Home*, *Office-31* and *ImageCLEF* are also shown in Table.1 and Table.2. We can easily observe that all methods using EMT-Refine with MROT-Norm or MROT-Ent significantly improve the performance on partial UDA task. Especially, MOT + SemiUOT (ETM-Refine + MROT-Norm) achieves the best performance, benefiting from its powerful matching capability. UOT relaxes the dual transportation constraints, causing some target samples cannot be transported to the source domain. However, SemiUOT overcomes the mentioned issue while avoiding negative transfer in partial UDA, which boosts the model performance. We also conduct more domain adaptation experiments to verify the effects of ETM-Refine with MROT-Norm in Appendix.J, K, L.

Solver Comparison Analysis. To further analyze the proposed ETM-based methods with MROT, we conduct the solver comparison in terms of *computation time* and *computation error*. We first sample the same number of source/target data samples from $\mathbb{P}_X$ and $\mathbb{P}_Z$, respectively. As shown in Fig.3(a)-(b), we conduct the experiments on both UOT and SemiUOT with $\tau_a = \tau_b = \tau = 1$. We can conclude that ETM-Exact with MROT-Norm is most time-consuming, due to directly seeking the optimum from a random initial point. Meanwhile, 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 $\widehat{\boldsymbol{f}}^*$. Moreover, we calculate the absolute computation error between matching solution $\boldsymbol{\pi}$ learned by ETM with MROT and the standard UOT/SemiUOT solution with CVXPY as $\boldsymbol{\pi}^*$, i.e., $e = \sum_{i,j} ||\boldsymbol{\pi}_{ij} - \boldsymbol{\pi}^*_{ij}||_1$. We sample 500 number of data samples ranging from $\tau_a = \tau_b = \tau = \{0.01, 0.1, 1, 10, 100\}$ for calculation and the results are shown in Fig.3(c)-(d). We can observe that although ETM-Approx with MROT-Ent has the fastest computation speed, the provided results $\boldsymbol{\pi}$ still have the highest error compared to the ground truth $\boldsymbol{\pi}^*$. Meanwhile ETM-Refine with MROT-Norm can further reach more accurate solutions against MROT-Ent. Though regularization terms, e.g., entropy and norm, can be directly used in SemiUOT and UOT, they bring larger computation errors without the guidance of KKT-multiplier $\boldsymbol{s}$. **Robust-SemiSinkhorn** causes the largest computation error in solving SemiUOT. We also find that all existing UOT methods, i.e., **Ent-UOT**, **MM-UOT**, and **GEMUOT**, not only underperform in matching, but also cost more computation time. We can observe that ETM-Refine method achieves much better results, especially when τ is relatively large, which is consistent with our discovery in Fig.1-Fig.2.

Parameter sensitivity Analysis. We finally study the effects of hyper-parameters on model performance. We tune ϵ in range of $\epsilon \in \{0.01, 0.1\}$ and show the results in Fig.4(a)-(d). We can observe that smaller ϵ could provide good approximation on UOT/SemiUOT, reducing the iteration steps for optimizing L_{U} and L_{P}. Although ϵ could hardly affect the performance on ETM-Refine, larger value on ϵ could consume more iteration steps for solving L_{U} and L_{P} since the initial values are more random. Additionally, we collect the computation error $e_\alpha = ||\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*||_\infty$, which measures the discrepancy between the marginal probability learned via ETM-Approx $\widehat{\boldsymbol{\alpha}}$ and the ground truth $\boldsymbol{\alpha}^*$. Larger values of ϵ may fail to reduce the computation error e_α when compared to smaller values of ϵ. Hence we set $\epsilon = 0.01$ empirically and more experimental results can be found in Appendix.M, N.

5 Related Works

Unbalanced and Semi-Unbalanced Optimal Transport. (1) *Related works on UOT*: UOT with KL divergence has been widely investigated for dealing with diverse applications [82, 30, 94, 51, 40, 32, 106, 64]. Different types of UOT solutions can be distinguished in terms of using entropy regularization term or not. Involving entropy in UOT can enhance the model scalability, yet resulting in dense matching results [98, 5]. Latest, [17] further considers UOT without entropy terms by Majorization-Minimization (MM) [20, 99] or regularization path methods [68, 69, 57]. However, the nature of MM algorithm inherits inexact proximal point of KL term [105], which still causes dense mapping when τ becomes larger. Meanwhile regularization path methods could be quite slow in computation, especially when $\tau \to +\infty$. Furthermore, as the number of samples increases, it can lead to high storage space consumption, which can be problematic. Recently, [21] discovers a similar transformation between continuous UOT and classic OT problem. However, this discovery cannot directly extend to SemiUOT and UOT in discrete scenarios, and provides no hint to compute the exact marginal distributions and corresponding matching $\boldsymbol{\pi}$. (2) *Related works on SemiUOT*: SemiUOT with KL divergence only relaxes one of the marginal constraints comparing with UOT. [52] first fully investigated the corresponding problem and proposed Robust-SemiSinkhorn algorithm. Nevertheless, it still suffers from inaccurate matching solutions with entropy regularization term. Currently, there are only extremely few works for solving SemiUOT [71]. Therefore, how to efficiently provide accurate matching solutions on both discrete SemiUOT and UOT is still a challenging problem.

6 Conclusion

In this paper, we propose Equivalent Transformation Mechanism (ETM) approach with ETM-Exact, ETM-Approx, and ETM-Refine to solve the marginal probabilities of SemiUOT and UOT. We illustrate that the essence of SemiUOT/UOT is reweighting data samples accordingly and thus SemiUOT/UOT problem can be transformed into standard optimal transport. Moreover, we propose KKT-Multiplier Regularization with Multiplier Regularized Optimal Transport (MROT) to obtain more accurate solutions. We conduct extensive experiments to show the superior performance of ETM with MROT, on both synthetic and real-world datasets of different tasks and applications.

Acknowledgments and Disclosure of Funding

This work was supported by the National Natural Science Foundation of China (No. 62572486), Natural Science Foundation of Shandong Province (No. ZR2023MF007). This work is also partly supported by the National Research Foundation (NRF), Singapore, through the AI Singapore Programme under the project titled “AI-based Urban Cooling Technology Development”(Award No. AISG3-TC-2024-014-SGKR), partly supported by the National Research Foundation (NRF), Singapore, through the AI Singapore Programme under the project titled "Learning Assisted Human-AI Collaboration for Large-scale Practical Combinatorial Optimization" (AISG Award No: AISG3-RP-2022-031).

References

[1] Dongsheng An, Na Lei, Xiaoyin Xu, and Xianfeng Gu. Efficient optimal transport algorithm by accelerated gradient descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*,

volume 36, pages 10119–10128, 2022.

[2] Joshua D Angrist and Guido W Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90(430):431–442, 1995.

[3] Yuki Arase, Han Bao, and Sho Yokoi. Unbalanced optimal transport for unbalanced word alignment. *ACL*, 2023.

[4] Yuki Markus Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*, 2019.

[5] Yogesh Balaji, Rama Chellappa, and Soheil Feizi. Robust optimal transport with applications in generative modeling and domain adaptation. *Advances in Neural Information Processing Systems*, 33:12934–12944, 2020.

[6] Jean-David Benamou. Numerical resolution of an "unbalanced" mass transport problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 37(5):851–868, 2003.

[7] Artem Betlei, Eustache Diemert, and Massih-Reza Amini. Uplift modeling with generalization guarantees. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 55–65, 2021.

[8] Mathieu Blondel, Vivien Seguy, and Antoine Rolet. Smooth and sparse optimal transport. In *International conference on artificial intelligence and statistics*, pages 880–889. PMLR, 2018.

[9] Nicolas Bonneel and David Coeurjolly. Spot: sliced partial optimal transport. *ACM Transactions on Graphics (TOG)*, 38(4):1–13, 2019.

[10] Luis A Caffarelli and Robert J McCann. Free boundaries in optimal transport and monge-ampere obstacle problems. *Annals of mathematics*, pages 673–730, 2010.

[11] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 135–150, 2018.

[12] Zhangjie Cao, Kaichao You, Mingsheng Long, Jianmin Wang, and Qiang Yang. Learning to transfer examples for partial domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2985–2994, 2019.

[13] Barbara Caputo, Henning Müller, Jesus Martinez-Gomez, Mauricio Villegas, Burak Acar, Novi Patricia, Neda Marvasti, Suzan Üsküdarlı, Roberto Paredes, Miguel Cazorla, et al. Imageclef 2014: Overview and analysis of the results. In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction: 5th International Conference of the CLEF Initiative, CLEF 2014, Sheffield, UK, September 15-18, 2014. Proceedings 5*, pages 192–211. Springer, 2014.

[14] Guillaume Carlier. On the linear convergence of the multimarginal sinkhorn algorithm. *SIAM Journal on Optimization*, 32(2):786–794, 2022.

[15] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020.

[16] Wanxing Chang, Ye Shi, Hoang Tuan, and Jingya Wang. Unified optimal transport framework for universal domain adaptation. *Advances in Neural Information Processing Systems*, 35:29512–29524, 2022.

[17] Laetitia Chapel, Rémi Flamary, Haoran Wu, Cédric Févotte, and Gilles Gasso. Unbalanced optimal transport through non-negative penalized linear regression. *Advances in Neural Information Processing Systems*, 34:23270–23282, 2021.

[18] Liang Chen, Yihang Lou, Jianzhong He, Tao Bai, and Minghua Deng. Evidential neighborhood contrastive learning for universal domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6258–6267, 2022.

[19] Lenaic Chizat. *Unbalanced optimal transport: Models, numerical methods, applications*. PhD thesis, Université Paris sciences et lettres, 2017.

[20] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609, 2018.

[21] Jaemoo Choi, Jaewoong Choi, and Myungjoo Kang. Generative modeling through the semi-dual formulation of unbalanced optimal transport. *NeurIPS*, 2023.

[22] Ching-Yao Chuang, Stefanie Jegelka, and David Alvarez-Melis. Infoot: Information maximizing optimal transport. In *International Conference on Machine Learning*, pages 6228–6242. PMLR, 2023.

[23] Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. *Advances in neural information processing systems*, 30, 2017.

[24] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

[25] Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 447–463, 2018.

[26] Jun Dan, Weiming Liu, Mushui Liu, Chunfeng Xie, Shunjie Dong, Guofang Ma, Yanchao Tan, and Jiazheng Xing. Hogda: Boosting semi-supervised graph domain adaptation via high-order structure-guided adaptive feature alignment. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 11109–11118, 2024.

[27] Jun Dan, Weiming Liu, Xie Xie, Hua Yu, Shunjie Dong, and Yanchao Tan. Tfgda: Exploring topology and feature alignment in semi-supervised graph domain adaptation through robust clustering. *Advances in Neural Information Processing Systems*, 37:50230–50255, 2024.

[28] Jun Dan, Yang Liu, Jiankang Deng, Haoyu Xie, Siyuan Li, Baigui Sun, and Shan Luo. Topofr: A closer look at topology alignment on face recognition. *Advances in Neural Information Processing Systems*, 37:37213–37240, 2024.

[29] Jun Dan, Yang Liu, Haoyu Xie, Jiankang Deng, Haoran Xie, Xuansong Xie, and Baigui Sun. Transface: Calibrating transformer training for face recognition from a data-centric perspective. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 20642–20653, 2023.

[30] Henri De Plaen, Pierre-François De Plaen, Johan AK Suykens, Marc Proesmans, Tinne Tuytelaars, and Luc Van Gool. Unbalanced optimal transport: A unified framework for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3198–3207, 2023.

[31] Arnaud Dessein, Nicolas Papadakis, and Jean-Luc Rouas. Regularized optimal transport and the rot mover's distance. *The Journal of Machine Learning Research*, 19(1):590–642, 2018.

[32] Luca Eyring, Dominik Klein, Théo Uscidda, Giovanni Palla, Niki Kilbertus, Zeynep Akata, and Fabian J Theis. Unbalancedness in neural monge maps improves unpaired domain translation. In *The Twelfth International Conference on Learning Representations*, 2023.

[33] Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R Arabnia. A brief review of domain adaptation. *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020*, pages 877–894, 2021.

[34] Kilian Fatras, Thibault Séjourné, Rémi Flamary, and Nicolas Courty. Unbalanced minibatch optimal transport; applications to domain adaptation. In *International Conference on Machine Learning*, pages 3186–3197. PMLR, 2021.

[35] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690. PMLR, 2019.

[36] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *The Journal of Machine Learning Research*, 22(1):3571–3578, 2021.

[37] Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya, and Tomaso A Poggio. Learning with a wasserstein loss. *Advances in neural information processing systems*, 28, 2015.

[38] Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*, pages 567–583. Springer, 2020.

[39] Zhan Gao, Aryan Mokhtari, and Alec Koppel. Limited-memory greedy quasi-newton method with non-asymptotic superlinear convergence rate. *arXiv preprint arXiv:2306.15444*, 2023.

[40] Milena Gazdieva, Arip Asadulaev, Evgeny Burnaev, and Aleksandr Korotin. Light unbalanced optimal transport. *Advances in Neural Information Processing Systems*, 37:93907–93938, 2024.

[41] Aude Genevay, Marco Cuturi, Gabriel Peyré, and Francis Bach. Stochastic optimization for large-scale optimal transport. *Advances in neural information processing systems*, 29, 2016.

[42] Xiang Gu, Xi Yu, Jian Sun, Zongben Xu, et al. Adversarial reweighting for partial domain adaptation. *Advances in Neural Information Processing Systems*, 34:14860–14872, 2021.

[43] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[44] Yu Hua, Weiming Liu, Gui Xu, Yaqing Hou, Yew-Soon Ong, and Qiang Zhang. Deterministic-to-stochastic diverse latent feature mapping for human motion synthesis. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 22724–22734, 2025.

[45] J. J. Hull. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 16(5):550–554, 2002.

[46] Qiujiang Jin and Aryan Mokhtari. Exploiting local convergence of quasi-newton methods globally: adaptive sample size approach. *Advances in Neural Information Processing Systems*, 34:3824–3835, 2021.

[47] Qiujiang Jin and Aryan Mokhtari. Non-asymptotic superlinear convergence of standard quasi-newton methods. *Mathematical Programming*, 200(1):425–473, 2023.

[48] Jeff L Kennington and Richard V Helgason. *Algorithms for network programming*. John Wiley & Sons, Inc., 1980.

[49] Valentin Khrulkov, Gleb Ryzhakov, Andrei Chertkov, and Ivan Oseledets. Understanding ddpm latent codes through optimal transport. *ICLR*, 2023.

[50] Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. Neural optimal transport. *ICLR*, 2023.

[51] Khang Le, Huy Nguyen, Khai Nguyen, Tung Pham, and Nhat Ho. On multimarginal partial optimal transport: Equivalent forms and computational complexity. In *International Conference on Artificial Intelligence and Statistics*, pages 4397–4413. PMLR, 2022.

[52] Khang Le, Huy Nguyen, Quang M Nguyen, Tung Pham, Hung Bui, and Nhat Ho. On robust optimal transport: Computational complexity and barycenter computation. *Advances in Neural Information Processing Systems*, 34:21947–21959, 2021.

[53] Y. L. Lecun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. proc ieee. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[54] Guangrui Li, Guoliang Kang, Yi Zhu, Yunchao Wei, and Yi Yang. Domain consensus clustering for universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9757–9766, 2021.

[55] Mengyu Li, Jun Yu, Tao Li, and Cheng Meng. Importance sparsification for sinkhorn algorithm. *arXiv preprint arXiv:2306.06581*, 2023.

[56] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *ICLR*, 2023.

[57] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[58] Tianlin Liu, Joan Puigcerver, and Mathieu Blondel. Sparsity-constrained optimal transport. *ICLR*, 2023.

[59] Weiming Liu, Chaochao Chen, Xinting Liao, Mengling Hu, Jiajie Su, Yanchao Tan, and Fan Wang. User distribution mapping modelling with collaborative filtering for cross domain recommendation. In *Proceedings of the ACM Web Conference 2024*, pages 334–343, 2024.

[60] Weiming Liu, Jun Dan, Fan Wang, Xinting Liao, Junhao Dong, Hua Yu, Shunjie Dong, and Lianyong Qi. Distinguish then exploit: Source-free open set domain adaptation via weight barcode estimation and sparse label assignment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4927–4938, 2025.

[61] Weiming Liu, Xiaolin Zheng, Chaochao Chen, Jiajie Su, Xinting Liao, Mengling Hu, and Yanchao Tan. Joint internal multi-interest exploration and external domain alignment for cross domain sequential recommendation. In *Proceedings of the ACM web conference 2023*, pages 383–394, 2023.

[62] Weiming Liu, Xiaolin Zheng, Chaochao Chen, Jiahe Xu, Xinting Liao, Fan Wang, Yanchao Tan, and Yew-Soon Ong. Reducing item discrepancy via differentially private robust embedding alignment for privacy-preserving cross domain recommendation. In *Forty-first International Conference on Machine Learning*, 2024.

[63] Dirk A Lorenz, Paul Manns, and Christian Meyer. Quadratically regularized optimal transport. *Applied Mathematics & Optimization*, 83(3):1919–1949, 2021.

[64] Frederike Lübeck, Charlotte Bunne, Gabriele Gut, Jacobo Sarabia del Castillo, Lucas Pelkmans, and David Alvarez-Melis. Neural unbalanced optimal transport via cycle-consistent semi-couplings. *arXiv preprint arXiv:2209.15621*, 2022.

[65] You-Wei Luo and Chuan-Xian Ren. Mot: Masked optimal transport for partial domain adaptation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3531–3540. IEEE, 2023.

[66] You-Wei Luo, Chuan-Xian Ren, Dao-Qing Dai, and Hong Yan. Unsupervised domain adaptation via discriminative manifold propagation. *IEEE transactions on pattern analysis and*

machine intelligence, 44(3):1653–1669, 2020.

[67] Zhiheng Ma, Xing Wei, Xiaopeng Hong, Hui Lin, Yunfeng Qiu, and Yihong Gong. Learning to count via unbalanced optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2319–2327, 2021.

[68] Julien Mairal and Bin Yu. Complexity analysis of the lasso regularization path. *ICML*, 2012.

[69] Mathurin Massias, Alexandre Gramfort, and Joseph Salmon. Celer: a fast solver for the lasso with dual extrapolation. In *International Conference on Machine Learning*, pages 3315–3324. PMLR, 2018.

[70] JH Mathews. Numerical methods using matlab, 2004.

[71] Eduardo Fernandes Montesuma, Fred Maurice Ngole Mboula, and Antoine Souloumiac. Recent advances in optimal transport for machine learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[72] Eduardo Fernandes Montesuma, Fred Ngole Mboula, and Antoine Souloumiac. Recent advances in optimal transport for machine learning. *arXiv preprint arXiv:2306.16156*, 2023.

[73] Debarghya Mukherjee, Aritra Guha, Justin M Solomon, Yuekai Sun, and Mikhail Yurochkin. Outlier-robust optimal transport. In *International Conference on Machine Learning*, pages 7850–7860. PMLR, 2021.

[74] Kirill Neklyudov, Rob Brekelmans, Alexander Tong, Lazar Atanackovic, Qiang Liu, and Alireza Makhzani. A computational framework for solving wasserstein lagrangian flows. *arXiv preprint arXiv:2310.10649*, 2023.

[75] Yu Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming*, 103:127–152, 2005.

[76] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. *Nips Workshop on Deep Learning & Unsupervised Feature Learning*, 2011.

[77] Khai Nguyen, Dang Nguyen, Tung Pham, Nhat Ho, et al. Improving mini-batch optimal transport via partial transportation. In *International Conference on Machine Learning*, pages 16656–16690. PMLR, 2022.

[78] Quang Minh Nguyen, Hoang H Nguyen, Yi Zhou, and Lam M Nguyen. On unbalanced optimal transport: Gradient methods, sparsity and approximation error. *JMLR*, 2023.

[79] Derek Onken, Samy Wu Fung, Xingjian Li, and Lars Ruthotto. Ot-flow: Fast and accurate continuous normalizing flows via optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9223–9232, 2021.

[80] Matteo Pariset, Ya-Ping Hsieh, Charlotte Bunne, Andreas Krause, and Valentin De Bortoli. Unbalanced diffusion schr\" odinger bridge. *arXiv preprint arXiv:2306.09099*, 2023.

[81] Xingchao Peng, Ben Usman, Neela Kaushik, Dequan Wang, Judy Hoffman, and Kate Saenko. Visda: A synthetic-to-real benchmark for visual domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2021–2026, 2018.

[82] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[83] Khiem Pham, Khang Le, Nhat Ho, Tung Pham, and Hung Bui. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *International Conference on Machine Learning*, pages 7673–7682. PMLR, 2020.

[84] Ievgen Redko, Nicolas Courty, Rémi Flamary, and Devis Tuia. Optimal transport for multi-source domain adaptation under target shift. In *The 22nd International Conference on artificial intelligence and statistics*, pages 849–858. PMLR, 2019.

[85] Bilal Riaz, Yuksel Karahan, and Austin J Brockmeier. Partial optimal transport for support subset selection. *Transactions on Machine Learning Research*, 2023.

[86] Anton Rodomanov and Yurii Nesterov. Rates of superlinear convergence for classical quasi-newton methods. *Mathematical Programming*, pages 1–32, 2021.

[87] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[88] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, pages 213–226. Springer, 2010.

[89] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 153–168, 2018.

[90] Ryoma Sato, Makoto Yamada, and Hisashi Kashima. Fast unbalanced optimal transport on a tree. *Advances in neural information processing systems*, 33:19039–19051, 2020.

[91] Meyer Scetbon, Marco Cuturi, and Gabriel Peyré. Low-rank sinkhorn factorization. In *International Conference on Machine Learning*, pages 9344–9354. PMLR, 2021.

[92] Meyer Scetbon, Michal Klein, Giovanni Palla, and Marco Cuturi. Unbalanced low-rank optimal transport solvers. *NeurIPS*, 2023.

[93] Thibault Séjourné, Clément Bonet, Kilian Fatras, Kimia Nadjahi, and Nicolas Courty. Unbalanced optimal transport meets sliced-wasserstein. *arXiv preprint arXiv:2306.07176*, 2023.

[94] Thibault Séjourné, Jean Feydy, François-Xavier Vialard, Alain Trouvé, and Gabriel Peyré. Sinkhorn divergences for unbalanced optimal transport. 2019.

[95] Thibault Séjourné, Gabriel Peyré, and François-Xavier Vialard. Unbalanced optimal transport, from theory to numerics. *arXiv preprint arXiv:2211.08775*, 2022.

[96] Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. Faster unbalanced optimal transport: Translation invariant sinkhorn and 1-d frank-wolfe. In *International Conference on Artificial Intelligence and Statistics*, pages 4995–5021. PMLR, 2022.

[97] Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, pages 3076–3085. PMLR, 2017.

[98] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[99] Ying Sun, Prabhu Babu, and Daniel P Palomar. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, 65(3):794–816, 2016.

[100] Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian Fatras, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023.

[101] Quang Huy Tran, Hicham Janati, Nicolas Courty, Rémi Flamary, Ievgen Redko, Pinar Demetci, and Ritambhara Singh. Unbalanced co-optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10006–10016, 2023.

[102] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. *CoRR*, abs/1706.07522, 2017.

[103] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[104] Hao Wang, Zhichao Chen, Jiajun Fan, Haoxuan Li, Tianqiao Liu, Weiming Liu, Quanyu Dai, Yichao Wang, Zhenhua Dong, and Ruiming Tang. Optimal transport for treatment effect estimation. *arXiv preprint arXiv:2310.18286*, 2023.

[105] Yujia Xie, Xiangfeng Wang, Ruijia Wang, and Hongyuan Zha. A fast proximal point method for computing exact wasserstein distance. In *Uncertainty in artificial intelligence*, pages 433–453. PMLR, 2020.

[106] Karren D Yang and Caroline Uhler. Scalable unbalanced optimal transport using generative adversarial networks. *arXiv preprint arXiv:1810.11447*, 2018.

[107] Liuyi Yao, Sheng Li, Yaliang Li, Mengdi Huai, Jing Gao, and Aidong Zhang. Representation learning for treatment effect estimation from observational data. *Advances in neural information processing systems*, 31, 2018.

[108] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2720–2729, 2019.

[109] Jin Yu, SVN Vishwanathan, Simon Günter, and Nicol N Schraudolph. A quasi-newton approach to non-smooth convex optimization. In *Proceedings of the 25th international conference on Machine learning*, pages 1216–1223, 2008.

[110] Fangneng Zhan, Yingchen Yu, Kaiwen Cui, Gongjie Zhang, Shijian Lu, Jianxiong Pan, Changgong Zhang, Feiying Ma, Xuansong Xie, and Chunyan Miao. Unbalanced feature transport for exemplar-based image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15028–15038, 2021.

[111] Lei Zhang and Xinbo Gao. Transfer adaptation learning: A decade survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[112] Yubo Zhuang, Xiaohui Chen, and Yun Yang. Wasserstein k-means for clustering probability distributions. *Advances in Neural Information Processing Systems*, 35:11382–11395, 2022.

Appendix

A Notation Table

We provide the important notations and their descriptions for clarification on Table.3.

Symbol	Description
$\boldsymbol{X} \in \mathbb{R}^{M\times D}$	Source domain data matrix
$\boldsymbol{Z} \in \mathbb{R}^{N\times D}$	Target domain data matrix
M	Number of source samples
N	Number of target samples
D	Data dimension
$\boldsymbol{a} \in \mathbb{R}^{M}$	Source marginal probability vector
$\boldsymbol{b} \in \mathbb{R}^{N}$	Target marginal probability vector
$\boldsymbol{\pi} \in \mathbb{R}^{M\times N}$	Coupling matching matrix (transport plan)
$\boldsymbol{C} \in \mathbb{R}^{M\times N}$	Cost (distance) matrix
τ	KL divergence coefficient for SemiUOT
τ_a	KL divergence coefficient for source (UOT)
τ_b	KL divergence coefficient for target (UOT)
$\boldsymbol{\alpha}$	Adjusted source marginal (via ETM)
$\boldsymbol{\beta}$	Adjusted target marginal (via ETM)
$\boldsymbol{f}$	Dual variable (SemiUOT, source)
$\boldsymbol{g}$	Dual variable (SemiUOT, target)
$\boldsymbol{u}$	Dual variable (UOT, source)
$\boldsymbol{v}$	Dual variable (UOT, target)
ζ	Scalar dual variable (shared)
$\boldsymbol{s}$	KKT multiplier variable
ω_{L}	Scaling factor: $\omega_{\mathrm{L}} = \frac{\langle \boldsymbol{b},\mathbf{1}\rangle}{\langle \boldsymbol{a},\mathbf{1}\rangle}$
ϵ	LogSumExp smoothing parameter
ν	Update step: $\nu = \frac{\tau\epsilon}{\tau+\epsilon}$
J_{SemiUOT}	Objective function of SemiUOT
J_{UOT}	Objective function of UOT
L_P	Exact SemiUOT Equation
$\hat{L}_P$	Approximate SemiUOT Equation
L_U	Exact UOT Equation
$\hat{L}_U$	Approximate UOT Equation
$G(\boldsymbol{\pi}, \boldsymbol{s})$	KKT-multiplier regularization term: $\langle \boldsymbol{\pi}, \boldsymbol{s}\rangle$
$L_{\mathrm{Reg}}(\boldsymbol{\pi})$	Regularization term for OT (e.g. entropy or ℓ_2)
η_G	Regularization weight for multiplier term
η_{Reg}	Regularization weight (entropy or ℓ_2)
$\tilde{C}_{ij}$	Adjusted cost: $\tilde{C}_{ij} = C_{ij} + \eta_G s_{ij}$
$\boldsymbol{\psi}$	Dual variable in MROT (source side)
$\boldsymbol{\phi}$	Dual variable in MROT (target side)

Table 3: Important notations

B Proof of Proposition 1

Proposition 1. (Principles of Equivalent Transformation Mechanism for SemiUOT) *Given SemiUOT with KL-Divergence J_{SemiUOT}, one can obtain its Fenchel-Lagrange multipliers form as:*

$$\begin{aligned} &\min_{\boldsymbol{f},\boldsymbol{g},\zeta} \left[\tau \sum_{i=1}^{M} a_i \exp\left(-\frac{f_i+\zeta}{\tau}\right) - \sum_{j=1}^{N} b_j (g_j - \zeta) \right] \\ &s.t.\ f_i + g_j + s_{ij} = C_{ij}, \quad s_{ij} \geq 0. \end{aligned} \tag{16}$$

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

$$\begin{aligned}&\min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\mathrm{P}} = \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle,\\ &s.t.\ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a}\odot \exp\left(-\frac{\boldsymbol{f}^* + \zeta^*}{\tau}\right) = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}.\end{aligned} \tag{17}$$

When $\tau \to \infty$, the source marginal probability is given as $\boldsymbol{\pi}\mathbf{1}_N = \omega_{\mathrm{L}}\boldsymbol{a}$ and $\omega_{\mathrm{L}} = \langle \boldsymbol{b}, \mathbf{1}_N\rangle / \langle \boldsymbol{a}, \mathbf{1}_M\rangle$.

Proof. To start with, we first review the definition of SemiUOT as shown below:

$$\begin{aligned}&\min_{\pi_{ij}\geq 0} J_{\mathrm{SemiUOT}} = \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right)\\ &\text{s.t. } \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}.\end{aligned} \tag{18}$$

Then we can rewrite the optimization problem:

$$\begin{aligned}&\min_{\boldsymbol{\pi}\geq 0} J = \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right)\\ &s.t.\ \begin{cases}(\text{Constraint}):\ \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}\\ (\text{Optional}):\ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{\alpha}\end{cases}.\end{aligned} \tag{19}$$

Note that we do not need to know the exact value of $\boldsymbol{\alpha}$ beforehand. We adopt this optional constraint only for simplifying the following deduction. The Lagrange multipliers of SemiUOT with KL-Divergence is given as:

$$\begin{aligned}\max_{\boldsymbol{s}\geq 0, \boldsymbol{f}, \boldsymbol{g}, \zeta} \min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\mathrm{SemiUOT}} = \tau\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{f} + \zeta, \boldsymbol{\pi}\mathbf{1}_N\rangle + \langle \boldsymbol{g} - \zeta, \boldsymbol{b}\rangle +\\ \langle \boldsymbol{C} - \boldsymbol{f}\otimes \mathbf{1}_N^\top - \mathbf{1}_M \otimes \boldsymbol{g}^\top - \boldsymbol{s}, \boldsymbol{\pi}\rangle,\end{aligned} \tag{20}$$

where $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{s}$ and ζ are dual variables. By taking the differentiation on π_{ij} we have:

$$\begin{aligned}\frac{\partial \mathcal{J}_{\mathrm{SemiUOT}}}{\partial \pi_{ij}} &= \left[\tau \log \frac{\sum\limits_{j=1}^{N} \pi_{ij}}{a_i} + f_i + \zeta\right] + \left(C_{ij} - f_i - g_j - s_{ij}\right)\\ &= C_{ij} + \tau \log \frac{\sum\limits_{j=1}^{N} \pi_{ij}}{a_i} + \zeta - g_j - s_{ij}\\ &= 0.\end{aligned} \tag{21}$$

Therefore, we can obtain the results as:

$$\begin{cases}\sum\limits_{j=1}^{N} \pi_{ij} = a_i \exp\left(-\dfrac{f_i + \zeta}{\tau}\right)\\ \sum\limits_{i=1}^{M} \pi_{ij} = b_j\\ C_{ij} - f_i - g_j - s_{ij} = 0\end{cases}. \tag{22}$$

After that, we can take these back into KL-Divergence to simplify the calculation:

$$\begin{aligned}&\tau\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{f} + \zeta, \boldsymbol{\pi}\mathbf{1}_N\rangle\\ &= \tau\mathrm{KL}\left(\boldsymbol{a}\exp\left(-\frac{\boldsymbol{f}+\zeta}{\tau}\right) \| \boldsymbol{a}\right) + \left\langle \boldsymbol{f} + \zeta, \boldsymbol{a}\exp\left(-\frac{\boldsymbol{f}+\zeta}{\tau}\right)\right\rangle\\ &= \tau \sum_{i=1}^{M}\left[a_i \exp\left(-\frac{f_i+\zeta}{\tau}\right)\log\frac{a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right)}{a_i} - a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right) + a_i\right] + \sum_{i=1}^{M}(f_i+\zeta)a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right)\\ &= \sum_{i=1}^{M}\left[-\tau a_i \exp\left(-\frac{f_i+\zeta}{\tau}\right) + \tau a_i\right].\end{aligned} \tag{23}$$

Therefore we can obtain its Fenchel-Lagrange multipliers form of SemiUOT as:

$$\begin{aligned}\min_{\boldsymbol{f},\boldsymbol{g},\zeta} \mathscr{J}_{\mathrm{SemiUOT}} &= -\tau \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) - \langle \boldsymbol{f}+\zeta, \boldsymbol{\pi}\mathbf{1}_N\rangle - \langle \boldsymbol{g}-\zeta, \boldsymbol{\pi}^\top \mathbf{1}_M\rangle \\ &= \tau \exp\left(-\frac{\zeta}{\tau}\right)\left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{f}}{\tau}\right)\right\rangle - \langle \boldsymbol{g}-\zeta, \boldsymbol{b}\rangle + \mathscr{O}_{\mathrm{Const}} \\ s.t.\ & f_i + g_j \leq C_{ij},\end{aligned} \tag{24}$$

where $\mathscr{O}_{\mathrm{Const}} = -\sum_{i=1}^{M} \tau a_i$ and we can neglect it during the following calculation. Once we obtain the optimal solution on $\boldsymbol{f}^*$, $\boldsymbol{g}^*$ and ζ^*, we will discover that:

$$\tau \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) = \tau \mathrm{KL}\left(\boldsymbol{a}, \exp\left(-\frac{\boldsymbol{f}^* + \zeta^*}{\tau}\right) \| \boldsymbol{a}\right) = \mathrm{Const}. \tag{25}$$

Hence SemiUOT problem can be transformed into classic optimal transport problem accordingly. Finally we can obtain the optimal solution on ζ by considering $\frac{\partial \mathscr{J}_{\mathrm{SemiUOT}}}{\partial \zeta} = 0$ as below:

$$\zeta = \tau \left[\log\left(\sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right)\right) - \log\left(\sum_{j=1}^{N} b_j\right)\right]. \tag{26}$$

Once we set $\tau \to \infty$, the results of the limitation will be shown as:

$$\lim_{\tau \to +\infty} a_i \exp\left(-\frac{f_i + \zeta}{\tau}\right) = \lim_{\tau \to +\infty} a_i \exp\left(-\frac{\zeta}{\tau}\right) = a_i \frac{\langle \boldsymbol{b}, \mathbf{1}_N\rangle}{\langle \boldsymbol{a}, \mathbf{1}_M\rangle} = \omega_{\mathrm{L}} a_i. \tag{27}$$

Therefore we conclude the proof of the Proposition 1. □

C Proof of Proposition 2

Proposition 2. (Calculation for Approximate SemiUOT Equation) *Given Approximate SemiUOT equation $\widehat{L}_{\mathrm{P}}$, it can be optimized via Equivalent Transformation Mechanism with Approximation (ETM-Approx). That is, ETM-Approx aims to solve the following equation for each $\widehat{f}_s$:*

$$\frac{\partial \widehat{L}_{\mathrm{P}}}{\partial \widehat{f}_s} = -a_s \exp\left(-\frac{\widehat{f}_s + \zeta}{\tau}\right) + \exp\left(\frac{\widehat{f}_s}{\epsilon}\right) \sum_{j=1}^{N} \left[\frac{b_j \exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{\widehat{f}_k - C_{kj}}{\epsilon}\right)}\right] = 0. \tag{28}$$

Specifically, we can adopt fixed-point iteration method for solving Eq.(28) *at the ℓ-th iteration as follows:*

$$\begin{cases} \widehat{f}_1^{\ell+1} = \nu \left[\log\left(a_1 \exp\left(-\frac{\zeta}{\tau}\right)\right) - \log\left[\sum_{j=1}^{N}\left(\frac{b_j}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \exp\left(-\frac{C_{1j}}{\epsilon}\right)\right)\right]\right] \\ \vdots \\ \widehat{f}_M^{\ell+1} = \nu \left[\log\left(a_M \exp\left(-\frac{\zeta}{\tau}\right)\right) - \log\left[\sum_{j=1}^{N}\left(\frac{b_j}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \exp\left(-\frac{C_{Mj}}{\epsilon}\right)\right)\right]\right] \end{cases}, \tag{29}$$

where $\nu = \tau\epsilon/(\tau+\epsilon)$ for simplification and $\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})$ denotes the corresponding calculation:

$$\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell}) = \sum_{k=1}^{M} \exp\left(\frac{\widehat{f}_k^{\ell} - C_{kj}}{\epsilon}\right). \tag{30}$$

The proposed procedure can be convergent with a theoretical guarantee. Finally, updating the Lagrange multiplier ζ by further considering $\nabla_\zeta \widehat{L}_{\mathrm{P}} = 0$ via $\zeta = \tau[\log(\sum_{i=1}^{M} a_i \exp(-\widehat{f}_i/\tau)) - \log(\sum_{j=1}^{N} b_j)]$. One can achieve the optimal results on $\widehat{\boldsymbol{f}}^*$ and ζ^* via iterative computing accordingly.

Proof. We first review the proposed Approximate SemiUOT Equation $\widehat{L}_{\mathrm{P}}$ as below:

$$\min_{\widehat{\boldsymbol{f}},\zeta} \widehat{L}_{\mathrm{P}} = \tau \sum_{i=1}^{M} a_i e^{-\frac{\widehat{f}_i+\zeta}{\tau}} + \sum_{j=1}^{N} b_j \left[\epsilon \log \left[\sum_{k=1}^{M} e^{\frac{\widehat{f}_k - C_{kj}}{\epsilon}} \right] + \zeta \right]. \tag{31}$$

Then we consider optimizing $\widehat{f}_s$ as follows:

$$\frac{\partial \widehat{L}_{\mathrm{P}}}{\partial \widehat{f}_s} = 0 \quad \Rightarrow \quad \exp\left(\frac{\tau+\epsilon}{\tau\epsilon} \widehat{f}_s \right) = \frac{a_s e^{-\frac{\zeta}{\tau}}}{\sum\limits_{j=1}^{N} \left(\frac{b_j \exp(-C_{sj}/\epsilon)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}})} \right)}, \tag{32}$$

where $\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}})$ is defined as Eq.(30). At that time we adopt fixed-point iteration method to optimize $\widehat{\boldsymbol{f}}$ accordingly:

$$\begin{cases} \widehat{f}_1^{\ell+1} = \nu \left[\log\left(a_1 e^{-\frac{\zeta}{\tau}} \right) - \log \left[\sum\limits_{j=1}^{N} \left(\dfrac{b_j e^{-C_{1j}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right) \right] \right] = \mathcal{F}_1 \left(\widehat{f}_1^{\ell}, \cdots, \widehat{f}_M^{\ell} \right) \\ \vdots \\ \widehat{f}_s^{\ell+1} = \nu \left[\log\left(a_s e^{-\frac{\zeta}{\tau}} \right) - \log \left[\sum\limits_{j=1}^{N} \left(\dfrac{b_j e^{-C_{sj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right) \right] \right] = \mathcal{F}_s \left(\widehat{f}_1^{\ell}, \cdots, \widehat{f}_M^{\ell} \right) \\ \vdots \\ \widehat{f}_M^{\ell+1} = \nu \left[\log\left(a_M e^{-\frac{\zeta}{\tau}} \right) - \log \left[\sum\limits_{j=1}^{N} \left(\dfrac{b_j e^{-C_{Mj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right) \right] \right] = \mathcal{F}_M \left(\widehat{f}_1^{\ell}, \cdots, \widehat{f}_M^{\ell} \right) \end{cases}, \tag{33}$$

where $\nu = \frac{\tau\epsilon}{\tau+\epsilon}$. By taking the gradient on $\mathcal{F}_s(\widehat{f}_s^{\ell})$ w.r.t $\widehat{f}_s^{\ell}$, we can observe that:

$$\begin{aligned} \frac{\partial \mathcal{F}_s(\widehat{f}_s^{\ell})}{\partial \widehat{f}_s^{\ell}} &= -\frac{\tau\epsilon}{\tau+\epsilon} \frac{1}{\sum\limits_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right] b_j} \frac{\partial}{\partial \widehat{f}_s^{\ell}} \left(\sum_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right] b_j \right) \\ &= \frac{\tau}{\tau+\epsilon} \underbrace{\frac{1}{\sum\limits_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right] b_j} \sum_{j=1}^{N} \left[\frac{b_j \exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \cdot \frac{\exp\left(\frac{\widehat{f}_s^{\ell} - C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right]}_{<1} \\ &< 1. \end{aligned} \tag{34}$$

Likewise we can obtain the result:

$$\begin{aligned} \mathscr{F}_s \left(\widehat{f}_1^{\ell}, \cdots, \widehat{f}_M^{\ell} \right) &= \left| \frac{\partial \mathcal{F}_s(\widehat{f}_1^{\ell})}{\partial \widehat{f}_1^{\ell}} \right| + \cdots + \left| \frac{\partial \mathcal{F}_s(\widehat{f}_s^{\ell})}{\partial \widehat{f}_s^{\ell}} \right| + \cdots + \left| \frac{\partial \mathcal{F}_s(\widehat{f}_M^{\ell})}{\partial \widehat{f}_M^{\ell}} \right| \\ &= \frac{\tau}{\tau+\epsilon} \frac{1}{\sum\limits_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right] b_j} \sum_{j=1}^{N} \left[\frac{b_j \exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \cdot \sum_{u=1}^{M} \left(\frac{\exp\left(\frac{\widehat{f}_u^{\ell} - C_{uj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})} \right) \right] \\ &= \frac{\tau}{\tau+\epsilon} < 1. \end{aligned} \tag{35}$$

We can easily conclude that:

$$\begin{cases} \mathscr{F}_1\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) < 1 \\ \vdots \\ \mathscr{F}_s\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) < 1 \\ \vdots \\ \mathscr{F}_M\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) < 1 \end{cases} . \tag{36}$$

Therefore, we can conclude that the proposed ETM-Approx method guarantees convergence according to Theorem 2.9 in [70]. □

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

Proof. We can formulate the whole optimization process for Proposition 2 as below :

$$\begin{cases} \widehat{f}_1^{\ell+1} = \mathcal{F}_1\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) \\ \vdots \\ \widehat{f}_s^{\ell+1} = \mathcal{F}_s\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) \\ \vdots \\ \widehat{f}_M^{\ell+1} = \mathcal{F}_M\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) \end{cases} \Rightarrow \quad \left(\widehat{f}_1^{\ell+1}, \cdots, \widehat{f}_M^{\ell+1}\right) = \mathcal{F}_{\mathrm{update}}\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right). \tag{37}$$

According to the above discussion, we have the following results:

$$\begin{aligned} \left\|\widehat{\boldsymbol{f}}^{\ell+1} - \widehat{\boldsymbol{f}}^*\right\|_\infty &= \left\|\mathcal{F}_{\mathrm{update}}\left(\widehat{\boldsymbol{f}}^\ell\right) - \mathcal{F}_{\mathrm{update}}\left(\widehat{\boldsymbol{f}}^*\right)\right\|_\infty \\ &\le \frac{\tau}{\tau+\epsilon}\left\|\widehat{\boldsymbol{f}}^\ell - \widehat{\boldsymbol{f}}^*\right\|_\infty \\ &\le \frac{\tau}{\tau+\epsilon}\left\|\widehat{\boldsymbol{f}}^{\ell+1} - \widehat{\boldsymbol{f}}^*\right\|_\infty + \frac{\tau}{\tau+\epsilon}\left\|\widehat{\boldsymbol{f}}^{\ell+1} - \widehat{\boldsymbol{f}}^\ell\right\|_\infty . \end{aligned} \tag{38}$$

Therefore the error between the solution $\widehat{\boldsymbol{f}}^{\ell+1}$ at the $(\ell+1)$ iteration and the optimal results $\widehat{\boldsymbol{f}}^*$ is given as:

$$\left\|\widehat{\boldsymbol{f}}^{\ell+1} - \widehat{\boldsymbol{f}}^*\right\|_\infty \le \frac{\tau+\epsilon}{\epsilon}\left(\frac{\tau}{\tau+\epsilon}\right)^\ell \left\|\widehat{\boldsymbol{f}}^{(1)} - \widehat{\boldsymbol{f}}^{(0)}\right\|_\infty . \tag{39}$$

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

D Algorithm for ETM-Based Method on SemiUOT

We also provide the pseudo algorithm of the proposed ETM-Based approachs (e.g., ETM-Exact, ETM-Approx, and ETM-Refine) for solving SemiUOT in Alg.1 to make a clearer illustration.

E Proof of Proposition 3

Proposition 3. (Principles of Equivalent Transformation Mechanism for UOT) *Given UOT with KL-Divergence J_{UOT}, its Fenchel-Lagrange multipliers form is given:*

$$\begin{aligned} &\min_{\boldsymbol{u},\boldsymbol{v},\zeta} \left[\tau_a \sum_{i=1}^M a_i e^{-\frac{u_i+\zeta}{\tau_a}} + \tau_b \sum_{j=1}^N b_j e^{-\frac{v_j-\zeta}{\tau_b}}\right] \\ &s.t.\ u_i + v_j + s_{ij} = C_{ij}, \quad s_{ij} \ge 0, \end{aligned} \tag{40}$$

where $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{s}$ and ζ denote Lagrange multipliers. Moreover, UOT problem can also be transformed into classic optimal transport as follows:

$$\begin{aligned}
&\min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\mathrm{U}} = \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle \\
&s.t. \begin{cases} \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a}\odot \exp\left(-\dfrac{\boldsymbol{u}^* + \zeta^*}{\tau_a}\right) = \boldsymbol{\alpha} \\ \boldsymbol{\pi}^\top\mathbf{1}_M = \boldsymbol{b}\odot \exp\left(-\dfrac{\boldsymbol{v}^* - \zeta^*}{\tau_b}\right) = \boldsymbol{\beta} \end{cases}.
\end{aligned} \tag{41}$$

Note that when $\tau_a, \tau_b \to \infty$, the source and target marginal probabilities can be determined as $\boldsymbol{\pi}\mathbf{1}_N = \sqrt{\omega_{\mathrm{L}}}\boldsymbol{a}$ and $\boldsymbol{\pi}^\top\mathbf{1}_M = \boldsymbol{b}/\sqrt{\omega_{\mathrm{L}}}$ where $\omega_{\mathrm{L}} = \langle \boldsymbol{b}, \mathbf{1}_N\rangle / \langle \boldsymbol{a}, \mathbf{1}_M\rangle$ respectively.

Proof. To start with, we first rewrite the optimization problem as below:

$$\begin{aligned}
&\min_{\boldsymbol{\pi}\geq 0} J = \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \tau_a \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \tau_b \mathrm{KL}(\boldsymbol{\pi}^\top\mathbf{1}_M \| \boldsymbol{b}) \\
&s.t. \text{ (Optional)}: \ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top\mathbf{1}_M = \boldsymbol{\beta}.
\end{aligned} \tag{42}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote the marginal probabilities for source and target domains respectively. Note that we do not need the true value fo $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ beforehand. That is, the constraints here are optional for the following UOT deduction. The Lagrange multipliers of UOT with KL-Divergence is given as:

$$\max_{\boldsymbol{s}\geq 0, \boldsymbol{u}, \boldsymbol{v}, \zeta} \min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\mathrm{UOT}} = \tau_a \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{u} + \zeta, \boldsymbol{\pi}\mathbf{1}_N\rangle + \tau_b \mathrm{KL}(\boldsymbol{\pi}^\top\mathbf{1}_M \| \boldsymbol{b}) + \langle \boldsymbol{v} - \zeta, \boldsymbol{\pi}^\top\mathbf{1}_M\rangle + \mathscr{C}_{\mathrm{UOT}}, \tag{43}$$

where $\mathscr{C}_{\mathrm{UOT}} = \sum_{i,j}(C_{ij} - u_i - v_j - s_{ij})\pi_{ij} = \langle \boldsymbol{C} - \boldsymbol{u}\otimes \mathbf{1}_N^\top - \mathbf{1}_M \otimes \boldsymbol{v}^\top - \boldsymbol{s}, \boldsymbol{\pi}\rangle$ and $\boldsymbol{u}$, $\boldsymbol{v}$ and ζ are dual variables. By taking the differentiation on π_{ij} we have:

$$\begin{aligned}
\frac{\partial \mathcal{J}_{\mathrm{UOT}}}{\partial \pi_{ij}} &= \left[\tau_a \log \frac{\sum\limits_{j=1}^{N} \pi_{ij}}{a_i} + u_i + \zeta\right] + \left[\tau_b \log \frac{\sum\limits_{i=1}^{M} \pi_{ij}}{b_j} + v_j - \zeta\right] + (C_{ij} - u_i - v_j - s_{ij}) \\
&= C_{ij} + \tau_a \log \frac{\sum\limits_{j=1}^{N} \pi_{ij}}{a_i} + \tau_b \log \frac{\sum\limits_{i=1}^{M} \pi_{ij}}{b_j} - s_{ij} = 0.
\end{aligned} \tag{44}$$

Then we can obtain the results:

$$\begin{cases} \sum\limits_{j=1}^{N} \pi_{ij} = a_i \exp\left(-\dfrac{u_i + \zeta}{\tau_a}\right) \\ \sum\limits_{i=1}^{M} \pi_{ij} = b_j \exp\left(-\dfrac{v_j - \zeta}{\tau_b}\right) \\ C_{ij} - u_i - v_j - s_{ij} = 0 \end{cases}. \tag{45}$$

By taking the above results into KL-Divergence, we can further simplify the results:

$$\begin{cases} \tau_a \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{u} + \zeta, \boldsymbol{\pi}\mathbf{1}_N\rangle = \sum\limits_{i=1}^{M} \left[-\tau_a a_i \exp\left(-\dfrac{f_i + \zeta}{\tau_a}\right) + \tau_a a_i\right] \\ \tau_b \mathrm{KL}\left(\boldsymbol{\pi}^\top\mathbf{1}_M \| \boldsymbol{b}\right) + \langle \boldsymbol{v} - \zeta, \boldsymbol{\pi}^\top\mathbf{1}_M\rangle = \sum\limits_{j=1}^{N} \left[-\tau_b b_j \exp\left(-\dfrac{g_j - \zeta}{\tau_b}\right) + \tau_b b_j\right] \end{cases}. \tag{46}$$

Therefore we can obtain its Fenchel-Lagrange multipliers form of UOT as:

$$\begin{aligned}
\min_{\boldsymbol{u}, \boldsymbol{v}, \zeta} \mathcal{J}_{\mathrm{UOT}} &= -\tau_a \mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) - \langle \boldsymbol{u} + \zeta, \boldsymbol{\pi}\mathbf{1}_N\rangle - \tau_b \mathrm{KL}(\boldsymbol{\pi}^\top\mathbf{1}_M \| \boldsymbol{b}) - \langle \boldsymbol{v} - \zeta, \boldsymbol{\pi}^\top\mathbf{1}_M\rangle \\
&= \tau_a \exp\left(-\frac{\zeta}{\tau_a}\right)\left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right)\left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle + \mathcal{O}_{\mathrm{Const}} \\
s.t.\ & u_i + v_j \leq C_{ij},
\end{aligned} \tag{47}$$

where $\mathcal{O}_{\text{Const}} = -\sum_{i=1}^{M} \tau_a a_i - \sum_{j=1}^{N} \tau_b b_j$, and we can neglect it during the following calculation. Once we obtain the optimal solution on $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and ζ^*, the KL-Divergence will turn out to be constants and therefore the original optimization problem can be transformed into classic optimal transport. Finally we can obtain the optimal solution on ζ by considering $\frac{\partial \mathcal{J}_{\text{UOT}}}{\partial \zeta} = 0$ as below:

$$\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b}\left[\log\left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle - \log\left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle\right]. \tag{48}$$

Once we set $\tau_a \to \infty$ and $\tau_b \to \infty$, the results of the limitation will be shown as:

$$\begin{aligned}
\lim_{\tau_a \to +\infty, \tau_b \to +\infty} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) &= \lim_{\tau_a \to +\infty, \tau_b \to +\infty} a_i \exp\left(-\frac{\zeta}{\tau_a}\right) = a_i \sqrt{\frac{\langle \boldsymbol{b}, \mathbf{1}_N \rangle}{\langle \boldsymbol{a}, \mathbf{1}_M \rangle}} = \sqrt{\omega_{\text{L}}} a_i, \\
\lim_{\tau_a \to +\infty, \tau_b \to +\infty} b_j \exp\left(-\frac{v_j - \zeta}{\tau_b}\right) &= \lim_{\tau_a \to +\infty, \tau_b \to +\infty} b_j \exp\left(-\frac{\zeta}{\tau_b}\right) = b_j \sqrt{\frac{\langle \boldsymbol{a}, \mathbf{1}_M \rangle}{\langle \boldsymbol{b}, \mathbf{1}_N \rangle}} = \frac{1}{\sqrt{\omega_{\text{L}}}} b_j.
\end{aligned} \tag{49}$$

Therefore we conclude the proof of the Proposition 3. □

F Illustrations of Optimization 1

Optimization 1. (Calculation of ETM-Approx approach for UOT) To start with, we first review the Exact UOT Equation is defined as:

$$\begin{aligned}
\min_{\boldsymbol{u}, \zeta} L_{\text{U}} &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(-\frac{v_j}{\tau_b}\right) \\
&= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(\frac{\sup_{k \in [M]}\left(u_k - C_{kj}\right)}{\tau_b}\right),
\end{aligned} \tag{50}$$

where $v_j = -\sup_{k \in [M]}\left(u_k - C_{kj}\right)$ meanwhile the marginal probabilities are set as $\boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-(\boldsymbol{u} + \zeta)/\tau_a\right) = \boldsymbol{\alpha}$ and $\boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \odot \exp\left(-(\boldsymbol{v} - \zeta)/\tau_b\right) = \boldsymbol{\beta}$. Since the optimization problem in Eq.(9) is convex, we can also utilize block gradient descent to optimize the problem. Specifically, we first fix v^l and optimize variable u^l at the l-th iteration by replacing the original marginal probability $\boldsymbol{b}$ in Eq.(6) with $\boldsymbol{\beta}$ accordingly to transform UOT into SemiUOT problem:

$$\begin{aligned}
&\min_{\boldsymbol{\pi} \geq 0} J_{\text{U}}^{u} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau_a \text{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right), \\
&s.t. \begin{cases}
(\text{Constraint}): \ \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \odot \exp\left(-\dfrac{\boldsymbol{v} - \zeta}{\tau_b}\right) = \boldsymbol{\beta} \\
(\text{Optional}): \ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\dfrac{\boldsymbol{u} + \zeta}{\tau_a}\right) = \boldsymbol{\alpha}
\end{cases}.
\end{aligned} \tag{51}$$

At that time, the Fenchel-Lagrange multipliers form of Eq.(51) is given via the Proposition 1:

$$\begin{aligned}
\min_{\boldsymbol{u}} L_{\text{U}}^{u} &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widetilde{u}_i + \zeta}{\tau_a}\right) - \sum_{j=1}^{N} \beta_j (\widetilde{v}_j - \zeta) \\
&= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) - \sum_{j=1}^{N} \left(\inf_{k \in [M]} \left[C_{kj} - u_k\right] - \zeta\right) \beta_j.
\end{aligned} \tag{52}$$

Note that $\widetilde{\boldsymbol{u}}$ and $\widetilde{\boldsymbol{v}}$ denote the Lagrange multiplier for Eq.(51) while we have $\widetilde{v}_j = \inf_{k \in [M]} \left[C_{kj} - u_k\right] = v_j$ and $\widetilde{\boldsymbol{u}} = \boldsymbol{u}$. To further accelerate the optimization process, we consider to make a smooth approximation on replacing $\inf(\cdot)$ as $\inf_{k \in [M]}[C_{kj} - u_k] \approx -\epsilon \log[\sum_{k=1}^{M} e^{\frac{u_k - C_{kj}}{\epsilon}}] = \widehat{v}_j$. Therefore, we first fix $\widehat{v}^l$ and optimize variable $\widehat{u}^l$ at the l-th iteration to solve the following equation on $\widehat{L}^u_{\mathrm{U}}$ accordingly:

$$
\begin{aligned}
\min_{\widehat{\boldsymbol{u}}} \widehat{L}^u_{\mathrm{U}} &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i+\zeta}{\tau_a}\right) + \sum_{j=1}^{N} \beta_j \left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}}\right] + \zeta\right] \\
&= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i+\zeta}{\tau_a}\right) + \sum_{j=1}^{N} b_j \exp\left(-\frac{\widehat{v}_j-\zeta}{\tau_b}\right) \left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}}\right] + \zeta\right].
\end{aligned} \tag{53}
$$

The optimization objective shares a similar formulation as Eq.31. At that time we adopt fixed-point iteration method to optimize $\widehat{\boldsymbol{u}}$ accordingly based on the Proposition 2:

$$
\begin{cases}
\widehat{u}_1^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a+\epsilon}\left[\log\left(a_1 e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N}\left(\frac{\beta_j e^{-C_{1j}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^\ell)}\right)\right]\right] = \mathcal{U}_1\left(\widehat{u}_1^\ell, \cdots, \widehat{u}_M^\ell\right) \\
\vdots \\
\widehat{u}_s^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a+\epsilon}\left[\log\left(a_s e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N}\left(\frac{\beta_j e^{-C_{sj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^\ell)}\right)\right]\right] = \mathcal{U}_s\left(\widehat{u}_1^\ell, \cdots, \widehat{u}_M^\ell\right) \\
\vdots \\
\widehat{u}_M^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a+\epsilon}\left[\log\left(a_M e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N}\left(\frac{\beta_j e^{-C_{Mj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^\ell)}\right)\right]\right] = \mathcal{U}_M\left(\widehat{u}_1^\ell, \cdots, \widehat{u}_M^\ell\right)
\end{cases} . \tag{54}
$$

The iteration process can be shown to converge efficiently based on Proposition 2. After that we fix $\widehat{u}$ and optimize variable $\widehat{v}$ via $\widehat{v}_j = -\epsilon \log[\sum_{k=1}^{M} \exp((\widehat{u}_k - C_{kj})/\epsilon)]$. We can achieve the optimal solution on $\widehat{\boldsymbol{u}}^*$ and $\widehat{\boldsymbol{v}}^*$ via iteratively computing via the above procedure accordingly. Finally, we update ζ via $\zeta = (\tau_a\tau_b/(\tau_a+\tau_b))[\log(\sum_{i=1}^{M} a_i \exp(-\widehat{u}_i^*/\tau_a)) - \log(\sum_{j=1}^{N} b_j \exp(-\widehat{v}_j^*/\tau_b))]$.

G Algorithm for ETM-Based Method on UOT

We also provide the pseudo algorithm of the proposed ETM-Based approachs (e.g., ETM-Exact, ETM-Approx and ETM-Refine) for solving UOT in Alg.2 to make a more clear illustration.

H Proof of Proposition 4

Proposition 4. (The Definition and Usage of KKT-Multiplier Regularization) *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):*

$$
\begin{aligned}
&\min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{G}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle + \eta_{\mathrm{Reg}} \mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}) \\
&s.t.\ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{\beta},
\end{aligned} \tag{55}
$$

where $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi})$ denotes the regularization term on $\boldsymbol{\pi}$. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ denote the final marginal probabilities obtained by ETM-based method, while η_{Reg}, η_G denotes the hyper parameter. Ideally, η_G should be set as a relatively large number. Meanwhile the dual form of MROT is given as:

$$
\max_{\boldsymbol{\psi}, \boldsymbol{\phi}} L_{\mathrm{G}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\mathrm{Reg}} \mathcal{L}^*_{\mathrm{Reg}}\left(\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\mathrm{Reg}}}\right), \tag{56}
$$

where $\widetilde{C}_{ij} = C_{ij} + \eta_G s_{ij}$, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ denote the Lagrange multipliers for MROT. $\mathcal{L}^*_{\mathrm{Reg}}(\cdot)$ denotes the conjugate function of $\mathcal{L}_{\mathrm{Reg}}(\cdot)$ and one can figure out the matching results of $\boldsymbol{\pi}$ via solving the following equation $\nabla_{\pi_{ij}} \mathcal{L}_{\mathrm{Reg}}(\pi_{ij}) = (\psi_i + \phi_j - \widetilde{C}_{ij})/\eta_{\mathrm{Reg}}$.

Algorithm 2 The algorithm of ETM-Based method on UOT

```
Input: $\boldsymbol{C}$: cost matrix; $\boldsymbol{a}, \boldsymbol{b}$: initial marginal probability; $\tau_a, \tau_b, \epsilon$: Hyper parameters.
  Randomly initialize the value of $\boldsymbol{u}^{\text{init}}$.
  Choose ETM-Exact, ETM-Approx or ETM-Refine on UOT for optimization.
(1) Function: ETM-Exact on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \boldsymbol{u}^{t=0} = \boldsymbol{u}^{\text{init}}$)
  Optimize $\boldsymbol{u}$ L-BFGS algorithm to optimize $L_{\text{U}}$ as:
    $\min_{\boldsymbol{u}} L_{\text{U}} = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(\frac{\sup_{k \in [M]} (u_k - C_{kj})}{\tau_b}\right)$
  Optimize $\boldsymbol{v}$ via $v_j = \inf_{k \in [M]} (C_{kj} - u_k)$.
  Optimize $\zeta$ via $\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b} \left[\log \left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle - \log \left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle\right]$ as shown in Eq.(48).
Return: The optimal solutions of $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and $\zeta^*$.
(2) Function: ETM-Approx on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \widehat{\boldsymbol{u}}^{t=0} = \boldsymbol{u}^{\text{init}}$)
Randomly initialize the value of $\widehat{\boldsymbol{v}}^{t'=1}$.
for $t' = 1$ to $T'$ do
  Optimize $\widehat{\boldsymbol{u}}^{t'}$ via Proposition 2 to optimize $\widehat{L}_{\text{U}}^{u}$ as:
    $\min_{\widehat{\boldsymbol{u}}} \widehat{L}_{\text{U}}^{u} = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i + \zeta}{\tau_a}\right) + \sum_{j=1}^{N} b_j \exp\left(-\frac{\widehat{v}_j - \zeta}{\tau_b}\right) \left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}}\right] + \zeta\right]$
  Optimize $\widehat{\boldsymbol{v}}^{t'}$ via $\widehat{v}_j^{t'} = -\epsilon \log[\sum_{k=1}^{M} \exp((\widehat{u}_k^{t'} - C_{kj})/\epsilon)]$.
end for
Optimize $\zeta$ via $\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b} \left[\log \left\langle \boldsymbol{a}, \exp\left(-\frac{\widehat{\boldsymbol{u}}}{\tau_a}\right)\right\rangle - \log \left\langle \boldsymbol{b}, \exp\left(-\frac{\widehat{\boldsymbol{v}}}{\tau_b}\right)\right\rangle\right]$ as shown in Eq.(48).
Return: The optimal solutions of $\widehat{\boldsymbol{u}}^*$, $\widehat{\boldsymbol{v}}^*$ and $\zeta^*$.
(3) Function: ETM-Refine on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \widehat{\boldsymbol{u}}^{t=0} = \boldsymbol{u}^{\text{init}}$)
  Obtain $\widehat{\boldsymbol{u}}^*$ = ETM-Approx on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \widehat{\boldsymbol{u}}^{t=0} = \boldsymbol{u}^{\text{init}}$).
  Obtain $\boldsymbol{u}^*$ = ETM-Exact on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \boldsymbol{u}^{t=0} = \widehat{\boldsymbol{u}}^*$).
Return: The optimal solutions of $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and $\zeta^*$.
```

Proof. We first provide the Lagrange multiplier of MROT as:

$$\begin{aligned}
\max_{\boldsymbol{\psi}, \boldsymbol{\phi}} \min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\text{MROT}} &= \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle + \eta_{\text{Reg}} \mathcal{L}_{\text{Reg}}(\boldsymbol{\pi}) - \langle \boldsymbol{\psi}, \boldsymbol{\pi} \mathbf{1}_N - \boldsymbol{\alpha} \rangle - \langle \boldsymbol{\phi}, \boldsymbol{\pi}^\top \mathbf{1}_M - \boldsymbol{\beta} \rangle \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle + \eta_{\text{Reg}} \inf_{\boldsymbol{\pi}} \left[\sum_{i,j} \left[\frac{C_{ij} + \eta_G s_{ij} - \psi_i - \phi_j}{\eta_{\text{Reg}}} \pi_{ij} + \mathcal{L}_{\text{Reg}}(\pi_{ij})\right]\right] \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\text{Reg}} \sup_{\boldsymbol{\pi}} \left[\sum_{i,j} \left[\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}} \pi_{ij} - \mathcal{L}_{\text{Reg}}(\pi_{ij})\right]\right] \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\text{Reg}} \mathcal{L}_{\text{Reg}}^* \left(\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right).
\end{aligned} \tag{57}$$

At that time we have the following results:

$$\begin{cases} \dfrac{\partial \mathcal{J}_{\text{MROT}}}{\partial \psi_i} = 0 \\ \dfrac{\partial \mathcal{J}_{\text{MROT}}}{\partial \phi_j} = 0 \end{cases} \Rightarrow \begin{cases} \nabla_{\psi_i} \mathcal{L}_{\text{Reg}}^* \left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right) = \alpha_i \\ \nabla_{\phi_j} \mathcal{L}_{\text{Reg}}^* \left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right) = \beta_j \end{cases}. \tag{58}$$

By taking the differentiation on π_{ij} we have:

$$\frac{\partial \mathcal{J}_{\text{MROT}}}{\partial \pi_{ij}} = \widetilde{C}_{ij} + \eta_{\text{Reg}} \nabla_{\pi_{ij}} \mathcal{L}_{\text{Reg}}(\pi_{ij}) - \psi_i - \phi_j = 0. \tag{59}$$

For instance, when $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}) = -\langle \boldsymbol{\pi}, \log(\boldsymbol{\pi}) - 1\rangle$ denotes as the entropy regularization term, the dual form of MROT-Ent is shown as:

$$\begin{cases} \max\limits_{\boldsymbol{\psi},\boldsymbol{\phi}} \mathcal{J}_{\mathrm{MROT-Ent}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle - \eta_{\mathrm{Reg}} \sum\limits_{i,j} \exp\left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\mathrm{Reg}}}\right) \\ \pi_{ij} = \exp\left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\mathrm{Reg}}}\right) \end{cases}. \tag{60}$$

When $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}) = \langle \boldsymbol{\pi}, \boldsymbol{\pi}\rangle/2$ denotes the square-norm regularization term, the dual form of MROT-Norm is shown as:

$$\begin{cases} \max\limits_{\boldsymbol{\psi},\boldsymbol{\phi}} \mathcal{J}_{\mathrm{MROT-Norm}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle - \dfrac{\eta_{\mathrm{Reg}}}{2} \sum\limits_{i,j} \left[\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\mathrm{Reg}}}\right]_+^2 \\ \pi_{ij} = \left[\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\mathrm{Reg}}}\right]_+ \end{cases}. \tag{61}$$

Therefore we conclude the proof of Proposition 4. □

Extensions. MROT can be even extended to solve classic optimal transport problem. That is, the classic optimal transport problem and its dual form can be represented as below:

$$\begin{aligned} & J = \arg\min_{\boldsymbol{\pi}\geq 0} \left\langle \boldsymbol{\pi}, \boldsymbol{C}\right\rangle \\ & s.t.\ \boldsymbol{\pi}\mathbf{1}_{\mathrm{N}} = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_{\mathrm{M}} = \boldsymbol{\beta} \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} & \max_{\boldsymbol{f}^{\triangle}, \boldsymbol{g}^{\triangle}, \boldsymbol{s}} \mathcal{J}_{\mathrm{OT}} = \left\langle \boldsymbol{f}^{\triangle}, \boldsymbol{\alpha}\right\rangle + \left\langle \boldsymbol{g}^{\triangle}, \boldsymbol{\beta}\right\rangle \\ & s.t. \begin{cases} f_i^{\triangle} + g_j^{\triangle} + s_{ij} = C_{ij}, \quad s_{ij} \geq 0 \\ g_j^{\triangle} = \inf\limits_{k\in[M]} \left(C_{kj} - f_k^{\triangle}\right) \end{cases} \end{aligned} \tag{62}$$

where $\boldsymbol{f}^{\triangle}$, $\boldsymbol{g}^{\triangle}$, and $\boldsymbol{s}$ represent the dual variables. To solve the dual form of the classic OT problem, unconstrained optimization techniques, such as L-BFGS or the Sinkhorn algorithm, can be employed to optimize for $\boldsymbol{s}$. Then one can further adopts MROT to solve $\boldsymbol{\pi}$ for classic optimal transport.

In summary, the time complexity of the proposed ETM-Approx+MROT-Ent or ETM-Approx+MROT-Norm method is provided as $\mathcal{O}\left(NM\log\left(1/\varepsilon_{\mathrm{err}}\right) + NMd_{\pi}\right)$ where d_{π} denotes the number of iterations on MROT. Meanwhile, the time complexity of the proposed ETM-Refine+MROT-Ent or ETM-Refine+MROT-Norm method is provided as $\mathcal{O}\left(NM\log\left(1/\varepsilon_{\mathrm{err}}\right) + NM(\log M)d_T + NMd_{\pi}\right)$ where d_T denotes the number of iterations on ETM-Refine.

I Experiments on Domain Adaptations

Datasets. We conduct the unsupervised domain adaptation tasks on *Digits*, *Office-Home*, and *VisDA*. *Digits* is the classic dataset for digit classification which contains three standard digit classification datasets: **MNIST** [53], **USPS**[45] and **SVHN** [76]. Each dataset consists of 10 classes of digits, ranging from 0 to 9. *Office-Home* [102] is a standard benchmark dataset which includes 15,500 images in 65 object classes in office and home settings, forming four dissimilar domains: Artistic images (**Ar**), Clip Art (**Cl**), Product images (**Pr**), and Real-World (**Rw**). *VisDA* [81] is a large-scale computer vision dataset on two domains, i.e., **Synthetic** and **Real** with 280K images in 12 classes.

Performance. We also conduct the UDA domain adaptation tasks on Digits and VisDA and the results are shown in Table.4. We can observe that the proposed ETM-Refine with MROT-Norm on SemiUOT achieves state-of-the-art performance on Digits and VisDA.

J Experiments on Partial Domain Adaptations

Datasets. We further conduct the domain adaptation tasks on new datasets, i.e., *Office-31* [88] and *ImageCLEF* [13]. **Office-31** is the commonly-used computer vision dataset for domain adaptation

Table 4: Classification accuracy (%) on Digits (Source: LeNet) and VisDA dataset (Source:ResNet50) for UDA (unsupervised domain adaptation) task

Method	S→M	M→U	U→M	Avg	VisDA
Source	68.3±0.3	65.3±0.5	66.2±0.2	66.6	52.4
DeepJDOT [25]	95.4±0.1	95.6±0.4	96.4±0.3	95.8	68.0
JUMBOT [34]	98.9±0.1	96.7±0.5	98.2±0.1	97.9	72.5
JUMBOT + UOT(ETM-Refine + MROT-Ent)	99.4±0.1	98.7±0.3	99.2±0.1	99.1	73.6
JUMBOT + UOT(ETM-Refine + MROT-Norm)	99.7±0.1	99.3±0.2	99.6±0.1	**99.5**	**74.2**

Table 5: H-score (%) on *Office-Home* for universal unsupervised domain adaptation

Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
ResNet [43]	44.65	48.04	50.13	46.64	46.91	48.96	47.47	43.17	50.23	48.45	44.76	48.43	47.32
OSBP [89]	39.59	45.09	46.17	45.70	45.24	46.75	45.26	40.54	45.75	45.08	41.64	46.90	44.48
UAN [108]	51.64	51.70	54.30	61.74	57.63	61.86	50.38	47.62	61.46	62.87	52.61	65.19	56.58
CMU [38]	56.02	56.93	59.15	66.95	64.27	67.82	54.72	51.09	66.39	68.24	57.89	69.73	61.60
DCC [54]	57.97	54.05	58.01	74.64	70.62	77.52	64.34	73.60	74.94	80.96	**75.12**	80.38	70.18
TNT [18]	61.90	74.60	80.20	73.50	71.40	79.60	74.20	**69.50**	82.70	77.30	70.10	81.20	74.70
UniOT [16]	67.27	80.54	86.03	73.51	77.33	84.28	75.54	63.33	85.99	77.77	65.37	81.92	76.57
UniOT + UOT(ETM-Refine + MROT-Ent)	68.63	81.72	87.94	75.88	79.03	86.21	77.29	68.77	87.14	78.59	73.62	82.83	78.97
UniOT + UOT(ETM-Refine + MROT-Norm)	**69.02**	**81.95**	**88.36**	**76.12**	**79.36**	**86.49**	**77.03**	69.25	**87.30**	**78.93**	74.18	**82.96**	**79.25**

with 4,652 images from three different domains: *Amazon* (**A**), *Webcam* (**W**) and *DSLR* (**D**). Target domain has the first 10 classes (alphabetical order) following [11]. **ImageCLEF** contains 3 domains with 12 classes, i.e., *Caltech* (**C**), *ImageNet* (**I**) and *Pascal* (**P**). Target domain has the first 6 classes (alphabetical order) following [66].

Baselines. We involve **DeepJDOT** [25], **ROT** [5], **JUMBOT** [34], **ETN** [12], **AR** [42], **m-POT** [77], **MOT** [65], as the model baselines for the domain adaptation task. (1) **DeepJDOT** [25] first adopts optimal transport into solving domain adaptation problem with deep learning framework. (2) **ROT** [5] adopts robust optimal transport into adversarial training for domain adaptation. (3) **JUMBOT** [34] adopts mini-batch unbalanced optimal transport method for domain adaptation. (4) **ETN** [12] utilizes example transfer network to jointly learn domain-invariant representations and the progressive weighting scheme. (5) **AR** [42] adopts adversarial reweighting strategy on source domain data for alignment. (6) **m-POT** [77] adopts partial optimal transport method in the mini-batch settings for domain adaptation. (7) **MOT** [65] adopts masked unbalanced optimal transport technique on considering label information for PDA tasks.

K Experiments on Universal Domain Adaptations

We further conduct the experiments on universal domain adaptations. That is, there are shared labels between the source and target domains. Additionally, there are private labels specific to each domain [33, 111]. We conduct the universal domain adaptations on both *Office-31* and *Office-Home*. Specifically, we set the first 10 classes in alphabetical order as the common label set, the next 10 classes as source private label and the rest 11 classes as target private label for Office-31. Likewise, we set the first 10 classes in alphabetical order as the common label set, the next 5 classes as source private label and the rest 55 classes as target private label for Office-Home. We involve the following models as baselines: (1) **OSBP** [89] adopts domain adversarial learning for open-set domain adaptation, (2) **UAN** [108] utilizes transferability criterion for universal domain adaptation, (3) **CMU** [38] learns to detect open classes with uncertainty estimation, (4) **DCC** [54] adopts domain consensus clustering for adaptation, (5) **TNT** [18] adopts evidential neighborhood contrastive learning for adaptation, (6) **UniOT** [16] adopts unbalanced optimal transport with adaptive filtering for transferring.

We adopt the same experimental settings as UniOT [16]. We utilize the commonly-used H-score [38] to validate the final results as shown in Table 5-6. Note that UniOT + UOT(ETM + MROT) only replaces the entropic UOT in UniOT with our proposed ETM-Refine method with MROT. From that, we can observe that UniOT + UOT(ETM-Refine + MROT-Norm) reaches the best performance, indicating that UOT with ETM + MROT can provide more accurate matching results.

L Experiments on Treatment Effect Estimation

Datasets for Treatment Effect Estimation. We further conduct ETM-Refine on treatment effect estimation with two semi-synthetic datasets IHDP [97] and ACIC [107]. IHDP is set to estimate the effect of specialist home visits on infants' potential cognitive scores and it contains 747 observations and 25 covariates. ACIC includes 4802 observations and 58 covariates, which comes from the collaborative perinatal project.

Table 6: H-Score (%) on *Office-31* for universal unsupervised domain adaptation

Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg
ResNet [43]	49.78	47.92	48.48	54.94	48.96	55.60	50.94
OSBP [89]	51.14	50.23	49.75	55.53	50.16	57.20	52.34
UAN [108]	59.68	58.61	60.11	70.62	60.34	71.42	63.46
CMU [38]	68.11	67.33	71.42	79.32	72.23	80.42	73.14
DCC [54]	88.50	78.54	70.18	79.29	75.87	88.58	80.16
TNT [18]	85.70	80.40	83.80	92.00	79.10	91.20	85.37
UniOT [16]	86.97	88.48	88.35	98.83	87.60	96.57	91.13
UniOT + UOT(ETM-Refine + MROT-Ent)	88.25	89.62	89.47	99.48	89.10	97.94	92.31
UniOT + UOT(ETM-Refine + MROT-Norm)	**88.67**	**90.14**	**90.03**	**99.58**	**89.42**	**98.46**	**92.72**

Table 7: Experimental results on Treatment Effect Estimation tasks.

	ACIC (PEHE)		ACIC (AUUC)		IHDP (PEHE)		IHDP (AUUC)	
	In-Sample	Out-Sample	In-Sample	Out-Sample	In-Sample	Out-Sample	In-Sample	Out-Sample
OLS [2]	3.749	4.340	0.843	0.496	3.856	5.674	0.652	0.492
TARNet [97]	3.236	3.254	**0.886**	0.662	0.749	1.788	0.654	0.711
PSM [87]	5.228	5.094	0.884	0.745	3.219	4.634	0.740	0.681
CFR-WASS [97]	3.128	3.207	0.873	0.669	0.657	1.704	0.656	0.715
ESCFR [104]	2.252	2.316	0.796	0.754	0.502	1.282	0.665	0.719
ESCFR + UOT(ETM-Refine + MROT-Ent)	2.327	2.261	0.839	0.814	0.497	1.275	0.769	0.763
ESCFR + UOT(ETM-Refine + MROT-Norm)	**2.104**	**2.216**	0.883	**0.839**	**0.475**	**1.146**	**0.798**	**0.802**

Results. We involve the following models as baselines: (1) **OLS** [2] utilizes least square regression with treatment as covariates, (2) **TARNet** [97] adopts integral orobability metrics for adaptation, (3) **PSM** [87] adopts propensity score for causal effects, (4) **CFR-WASS** [87] utilizes standard optimal transport for adaptation, (5) **ESCFR** [104] further utilizes unbalanced optimal transport for adaptation. We adopt the same experimental settings as ESCFR [104]. We utilize Precision in Estimation of Heterogeneous Effect (PEHE) [97] and Area Under the Uplift Curve (AUUC) [7] for the evaluation. Note that ESCFR + UOT(ETM-Refine + MROT-Ent) only replaces the entropic UOT in ESCFR with our proposed approximate-to-exact ETM-Refine + MROT-Norm. The experimental results are shown in Table 7. From that, we can observe that ESCFR + UOT(ETM-Refine + MROT-Norm) achieves the best performance, indicating the efficacy of our proposed ETM-Refine method.

M More Experimental Results

Parameter sensitivity. We tune η_{G} on SemiUOT via ETM-Refine with MROT-Norm in range of $\eta_{\mathrm{G}} \in \{0, 1, 100\}$ using the same data samples shown in Fig.1 and show the results in Fig.5. We can observe that when η_{G} is smaller (e.g., $\eta_{\mathrm{G}} = 0$ or $\eta_{\mathrm{G}} = 1$), the proposed KKT-multiplier regularization term $\mathcal{G}(\boldsymbol{\pi}, \boldsymbol{s}) = \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle$ may struggle to play a significant role during the optimization process. Meanwhile when $\eta_{\mathrm{G}} = 100$, ETM-Refine with MROT-Norm can achieve more accurate matching results. 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. Moreover, we conduct the experiments for the absolute error when $\tau = 1$ with $N = 500$ synthetic data samples on both SemiUOT and UOT and report the results in Fig.6(a)-(b). Larger value on η_G can provide more useful KKT-multiplier information and boost the model performance and therefore we set $\eta_{\mathrm{G}} = 100$ empirically. Furthermore, we conduct the hyper parameter experiments by varying $\epsilon = \{0.01, 0.05, 0.1, 0.5, 1\}$ on UDA task in Office-Home and report the results in Fig.6(c). We can observe that smaller value of ϵ can provide a more accurate approximation with higher accuracy and thus we set $\epsilon = 0.01$.

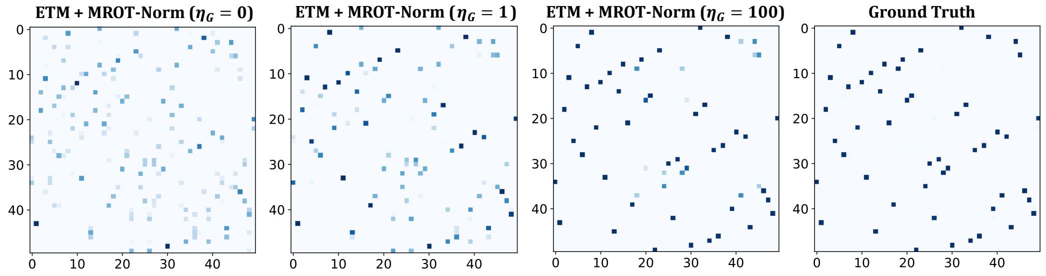
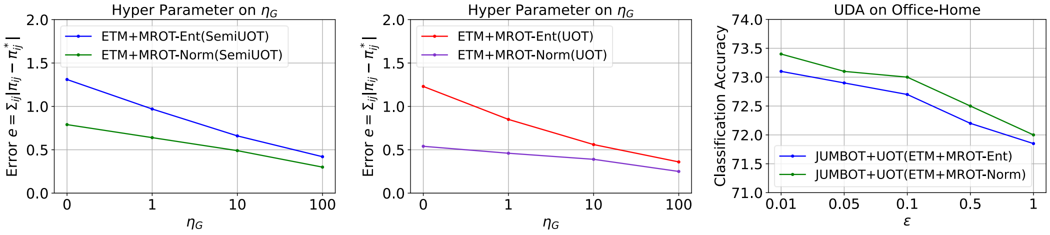
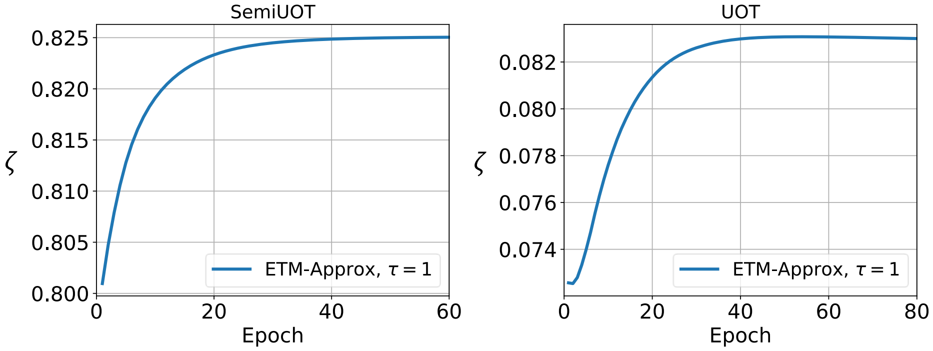


Figure 5: The matching results on ETM + MROT-Norm on SemiUOT with different values of $\eta_G = \{0, 1, 100\}$.



(a) η_G on SemiUOT (Synthetic Dataset) **(b) η_G on UOT (Synthetic Dataset)** **(c) ϵ on UOT (UDA on Office-Home)**

Figure 6: The hyper parameters on η_G and ϵ.



(a) The result of ζon SemiUOT **(b) The result of ζon UOT**

Figure 7: The results of ζ and L_{P} on UOT and SemiUOT.

N Miscellaneous Discussions

N.1 The role of ζ in ETM-based method

We first discuss why we should involve translation invariant ζ in both L_{U} and L_{P}. Specifically, we first analyze the case of SemiUOT. The Fenchel-Lagrange conjugate form of SemiUOT without translation invariant mechanism is given as:

$$\min_{\boldsymbol{f},\boldsymbol{g},\zeta} \left[\tau \sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right) - \sum_{j=1}^{N} b_j g_j \right] \quad s.t.\ f_i + g_j \leq C_{ij}. \tag{63}$$

We can adopt c-transform on Eq.(63) to obtain the unconstrained optimization problem as:

$$\min_{\boldsymbol{f}} \widetilde{L}_{\mathrm{P}} = \tau \sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right) - \sum_{j=1}^{N} \inf_{k\in[M]} [C_{kj} - f_k] b_j, \tag{64}$$

We adopt L-BFGS to optimize $\widetilde{L}_{\mathrm{P}}$ using the same data samples as shown in Fig.1 with $\tau = 1$. Meanwhile, the translation invariant term ζ in SemiUOT should be calculated as follows:

$$\zeta = \tau \log\left(\sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right)\right) - \tau \log\left(\sum_{j=1}^{N} b_j\right). \tag{65}$$

Ideally, ζ should equals to 0 since $\sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right) = \sum_{j=1}^{N} b_j$. However, we can observe that $\zeta > 0$ during the iteration epoch on optimizing $\widetilde{L}_{\mathrm{P}}$ as shown in Fig.7(a). Therefore we can conclude that ζ is indispensable during the calculation on SemiUOT. Likewise, the Fenchel-Lagrange conjugate form of UOT without translation invariant mechanism is given as:

$$\min_{\boldsymbol{v},\boldsymbol{u}} \left[\tau_a \left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right) \right\rangle + \tau_b \left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right) \right\rangle \right] \quad s.t.\ u_i + v_j \leq C_{ij}. \tag{66}$$

Here we can adopt c-transform on Eq.(66) to obtain the unconstrained optimization problem as:

$$\min_{\boldsymbol{u}} \widetilde{L}_{\mathrm{U}} = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i}{\tau_a}\right) + \tau_b \sum_{j=1}^{N} b_j \exp\left(\frac{\sup_{k=1}^{M}\left(u_k - C_{kj}\right)}{\tau_b}\right). \tag{67}$$

We also adopt L-BFGS to optimize $\widetilde{L}_{\mathrm{U}}$ using the same data samples as shown in Fig.2 with $\tau_a = \tau_b = 1$. Meanwhile, the translation invariant term ζ in UOT should be calculated as follows:

$$\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b}\left[\log\left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle - \log\left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle\right]. \tag{68}$$

Ideally, ζ should equals to 0 since $\left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle = \left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle$. However, we can observe that $\zeta > 0$ during the iteration epoch on optimizing $\widetilde{L}_{\mathrm{U}}$ as shown in Fig.7(b). Therefore we can conclude that ζ is indispensable during the calculation on UOT. In conclusion, the concept of translation invariant was first proposed in [96]. However, [96] only utilizes translation invariant for entropic UOT. **We highlight that, in this paper, we further extend translation invariant for standard UOT/SemiUOT scenario**. We illustrate that translation invariant is essential in solving UOT and SemiUOT problems.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction sections reflect the paper's contributions and scope in the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper focuses exclusively on discrete Semi-UOT and UOT problems. Other types of optimal transport problems, such as GW-based Semi-UOT and UOT, are beyond the scope of this work and remain promising directions for future research.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All theoretical claims are correctly induced with reasonable assumptions and correct proofs.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide the details of experimental implementation and algorithms in Section 5 main paper and Appendix D, G, I, J, K, for reproducibility.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide the code in appendix.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We provide full experimental details with the code, in the main paper and appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We report our results with five times of repetitions and show the significant results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide both computation complexity and computation time in this paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Yes, we do.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: This work has no societal impact.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper poses no such risks

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.