Unified (Semi) Unbalanced and Classic Optimal Transport with Equivalent Transformation Mechanism and KKT-Multiplier Regularization

Anonymous Authors[1]

Abstract

Semi-Unbalanced Optimal Transport (SemiUOT) shows great promise in matching two probability measures by relaxing one of the marginal constraints. Previous solvers often incorporate an entropy regularization term, which can result in inaccurate matching solutions. To address this issue, we focus on determining the marginal probability distribution of SemiUOT with KL divergence using the proposed Equivalent Transformation Mechanism (ETM) approach. Furthermore, we extend the ETM-based method into exploiting the marginal probability distribution of Unbalanced Optimal Transport (UOT) with KL divergence for validating its generalization. Once the marginal probabilities of UOT/SemiUOT are determined, they can be transformed into a classical 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 several numerical experiments to demonstrate the effectiveness of our proposed methods in addressing UOT/SemiUOT problems.

1. Introduction

Optimal Transport (OT) technique is a powerful tool for discerning and comparing distinct probability distributions. Nowadays, OT has multiple successful applications in traditional machine learning (Frogner et al., 2015; Feydy et al., 2019; Zhuang et al., 2022; Chuang et al., 2023; Riaz et al., 2023), unsupervised clustering (Asano et al., 2019; Caron et al., 2020), domain adaptation (Damodaran et al., 2018; Courty et al., 2017; Redko et al., 2019), diffusion (Khrulkov et al., 2023; Lipman et al., 2023), generative modeling (Korotin et al., 2023; Onken et al., 2021; Tong et al., 2023) 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 (Cuturi, 2013) for solving OT efficiently, it still suffers from the dilemma of dense and inaccurate solutions (Liu et al., 2023; Lorenz et al., 2021; Dessein et al., 2018). Moreover, classical OT strictly assume that the 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.

Recently, Unbalanced Optimal Transport (UOT) (Benamou, 2003; Chizat, 2017; Séjourné et al., 2023; Scetbon et al., 2023; Séjourné et al., 2022b) and Semi-Unbalanced Optimal Transport (SemiUOT) (Le et al., 2021) have become more attractive in adapting outliers since it allows relaxing marginal constraints for transportation results. UOT adopts several divergences such as Kullback-Leiber (KL) divergence (Pham et al., 2020), ℓ_1 norm (Caffarelli & McCann, 2010) and ℓ_2 norm (Blondel et al., 2018) for the relaxation on OT mass equality constraints by adjusting the corresponding coefficients τ. Meanwhile, KL divergence is the most commonly-used in UOT formulation in real practice (Séjourné et al., 2022a). UOT also provides great applications in transfer learning (Tran et al., 2023; Mukherjee et al., 2021; Pariset et al., 2023), computer vision (Bonneel & Coeurjolly, 2019; De Plaen et al., 2023; Choi et al., 2023; Neklyudov et al., 2023; Chang et al., 2022; Ma et al., 2021; Zhan et al., 2021), structure data exploration (Sato et al., 2020), natural language processing (Arase et al., 2023) and many areas. Previous solvers always involves extra regularization terms including entropy regularization term and proximal point term (Fatras et al., 2021) for tackling UOT problem. While adding additional entropy terms will lead to dense and inaccurate matching solutions. Latest, (Chapel et al., 2021) and (Nguyen et al., 2023) 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 solutions when τ is small, but unsatisfying solutions when τ is much larger. Therefore, it is quite challenging to efficiently exploit accurate solutions for both UOT and SemiUOT problems.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

In this paper, we propose a new method, i.e., ***equivalent transformation mechanism*** (ETM), which relieves the need for extra regularization terms in solving 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. We can further observe that the essence of SemiUOT and UOT lies in correspondingly adjusting the initial weights of different data samples. This provides us with new insights for understanding SemiUOT and UOT problems, i.e., we can transform SemiUOT and UOT problems into classic optimal transport problems based on initial marginal weights. Though we can exactly solve the marginal distributions via conventional iterative methods, e.g., L-BFGS, we further propose ETM-Approx to achieve the approximate results efficiently and convergently. Moreover, ETM-Refine resolves exact solutions via quasi-Newton optimization where the start points are these approximate solutions. Compared with original ETM, ETM-Apporx and ETM-Refine obtain accurate solutions while competitively balancing computational cost. 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 and UOT problems with marginal weights. Therefore we further proposed ***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 MGOT for achieving more accurate results. (3) We conduct extensive experiments on both synthetic and real-world datasets to demonstrate the performance of proposed ETM.

2. Preliminary

We first provide a brief preliminary definition of OT, UOT and SemiUOT. 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. Each data samples has corresponding prior-given mass weights $\boldsymbol{a} \in \mathbb{R}^{M\times 1}$ and $\boldsymbol{b} \in \mathbb{R}^{N\times 1}$. Meanwhile the total masses of these data samples are equal as $\boldsymbol{a}^\top \mathbf{1}_M = \boldsymbol{b}^\top \mathbf{1}_N$. The classical OT problem was defined by (Kantorovich, 1942) with a linear problem to measure the minimum transportation cost among data sample $\boldsymbol{X}$ and $\boldsymbol{Z}$:

$$\min_{\pi_{ij}\geq 0} J_{\mathrm{OT}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle \quad \text{s.t. } \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b},$$

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 $\boldsymbol{Z}$. One can directly solve J_{OT} via utilizing network-flow algorithm (Kennington & Helgason, 1980; Ahuja et al., 1988). To consider more general cases (e.g., filtering out the noise or outliers), one can relax two marginal constraints, i.e., $\boldsymbol{\pi}\mathbf{1}_N \neq \boldsymbol{a}$ and $\boldsymbol{\pi}^\top \mathbf{1}_M \neq \boldsymbol{b}$, to achieve unbalanced optimal transport problem (Pham et al., 2020):

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

where $\mathrm{KL}(\cdot)$ denotes Kullback-Leiber (KL) divergence which has been widely used in dealing with UOT. τ_a and τ_b denote the balanced hyper parameters between the minimizing cost and marginal relaxation. Note that when $\tau_a, \tau_b \to +\infty$ and $\boldsymbol{a}^\top \mathbf{1}_M = \boldsymbol{b}^\top \mathbf{1}_N$, UOT problem will turn into classical OT. Moreover, we can add one marginal constraints to formulate SemiUOT. For instance, we relax the constraint $\boldsymbol{\pi}\mathbf{1}_N \neq \boldsymbol{a}$ and keep the constraint $\boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}$:

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

Previous researches always add entropy regularization term for solving OT, UOT and SemiUOT. Although entropy regularization term can enhance the scalability of solving $\boldsymbol{\pi}^*$, it still suffers from the dense and inaccurate solution dilemma. In the following, we will first investigate the problem of UOT/SemiUOT from the perspective of marginal probability distribution, in order to find out the accurate solution of $\boldsymbol{\pi}^*$ for OT, UOT and SemiUOT.

3. Methodology

In this section, we will provide the calculation details on finding the solutions for commonly-existed UOT and SemiUOT. Previous methods (Pham et al., 2020; Chizat et al., 2018) always directly adopted entropy-based regularization term into tackling UOT and SemiUOT problem. Although such approaches can provide fast computation speed, it will lead to relatively ambiguous and dense solution which does not match most of situations in real practices (Li et al., 2023; Scetbon et al., 2021). Latest, (Chapel et al., 2021) adopted majorization-minimization algorithm or regularization path for solving UOT/SemiUOT problem. However, majorization-minimization algorithm is sensitive to the choice of τ, and still causes inaccurate and dense solutions when $\tau \to +\infty$. Worse still, regularization path could involve heavy matrix computation on inversion, requiring complicated optimization procedure. To solve the above problem, we change the perspective of solving the UOT/SemiUOT problem, i.e., originally exploiting the marginal probability of UOT/SemiUOT via our proposed *Equivalent Transformation Mechanism* (ETM) approach. In this way, we can obtain some interesting insights on understanding the intrinsic characteristics of UOT/SemiUOT. Moreover, we further propose *KKT-Multiplier Regularization* with *Multiplier Regularized Optimal Transpor* (MROT) with theorems and corollaries to achieve more accurate matching solution on SemiUOT and UOT respectively.

3.1. Equivalent Transformation Mechanism

In this section, we will first introduce the proposed equivalent transformation mechanism approach. Specifically, we propose an ETM-based method 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.

Equivalent Transformation Mechanism for SemiUOT. To start with, we first exploit the marginal probability distributions for SemiUOT via the proposed ETM method. Specifically, ETM includes three different approaches, i.e., ETM-Exact, ETM-Approx and ETM-Refine. By utilizing the methods above, one can transform SemiUOT into classic optimal transport problem. In this section, 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:*

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

where $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{s}$ and ζ denotes Lagrange multipliers. Moreover, SemiUOT problem can be transformed into classic optimal transport as follows:

$$\min_{\boldsymbol{\pi}\geq 0}\mathcal{J}_{\mathrm{P}}=\langle \boldsymbol{C},\boldsymbol{\pi}\rangle\quad s.t.\begin{cases}\boldsymbol{\pi}\mathbf{1}_N=\boldsymbol{a}\odot\exp\left(-\dfrac{\boldsymbol{f}^*+\zeta^*}{\tau}\right)=\boldsymbol{\alpha}\\ \boldsymbol{\pi}^\top\mathbf{1}_M=\boldsymbol{b}\end{cases} \tag{2}$$

Note that when $\tau\to\infty$, the source marginal probability can be determined as $\boldsymbol{\pi}\mathbf{1}_N=\omega\boldsymbol{a}$ where $\omega=\langle\boldsymbol{b},\mathbf{1}_N\rangle/\langle\boldsymbol{a},\mathbf{1}_M\rangle$.

The proof of Proposition 1 can be found in Appendix A. We can observe that SemiUOT is set to assign different weights on data samples. 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 (Villani et al., 2009). Therefore, we only need to optimize $\boldsymbol{f}$ and ζ without additional constraints as follows:

$$\min_{\boldsymbol{f},\zeta}L_{\mathrm{P}}=\tau\sum_{i=1}^{M}a_i e^{-\frac{f_i+\zeta}{\tau}}-\sum_{j=1}^{N}\left[\inf_{k\in[M]}[C_{kj}-f_k]-\zeta\right]b_j, \tag{3}$$

We refer to L_{P} as the newly proposed *Exact SemiUOT Equation*. Specifically, we initialize $\zeta=0$ for the optimization. We first fix ζ then adopting L-BFGS method (Berahas et al., 2016; Virtanen et al., 2020) 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 $\nabla_\zeta L_{\mathrm{P}}=0$ and it guarantees $\sum_{i=1}^{M}a_i e^{-(f_i^\ell+\zeta)/\tau}=\sum_{j=1}^{N}b_j$. We iteratively update L_{U} to reach the optimal solution on ζ^*, $\boldsymbol{f}^*$ and $\boldsymbol{g}^*$. Here we refer to the entire optimization procedure as the ETM-Exact approach for addressing the Exact Semi-UOT Equation. Although L_{P} is convex and has unique solutions, the presence of $\inf(\cdot)$ renders it a non-smooth function, leading to inefficient optimization (An et al., 2022). To further accelerate the optimization process, we consider to make a smooth approximation on replacing $\inf(\cdot)$ as $\inf_{k\in[M]}[C_{kj}-f_k]\approx-\epsilon\log[\sum_{k=1}^{M}e^{\frac{f_k-C_{kj}}{\epsilon}}]$. Note that $\epsilon>0$ denotes the balanced hyper parameters among the accuracy and function smoothness. Smaller ϵ (e.g., ϵ approaches to 0) could lead to more accurate while 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 $\widehat{f}$ as below:

$$\min_{\widehat{\boldsymbol{f}},\zeta}\widehat{L}_{\mathrm{P}}=\tau\sum_{i=1}^{M}a_i e^{-\frac{\widehat{f}_i+\zeta}{\tau}}+\sum_{j=1}^{N}b_j\left[\log\left[\sum_{k=1}^{M}e^{\frac{\widehat{f}_k-C_{kj}}{\epsilon}}\right]^{\epsilon}+\zeta\right] \tag{4}$$

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 $\widehat{f}_s$:*

$$\frac{\partial\widehat{L}_{\mathrm{P}}}{\partial\widehat{f}_s}=-a_s e^{-\frac{\widehat{f}_s+\zeta}{\tau}}+e^{\frac{\widehat{f}_s}{\epsilon}}\sum_{j=1}^{N}\left[\frac{b_j\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\sum_{k=1}^{M}\exp\left(\frac{\widehat{f}_k-C_{kj}}{\epsilon}\right)}\right]=0. \tag{5}$$

Specifically, we can adopt fixed-point iteration method for solving Eq.(5) at the ℓ-th iteration as follows:

$$\begin{cases}\widehat{f}_1^{\ell+1}=\nu\left[\log\left(a_1e^{-\frac{\zeta}{\tau}}\right)-\log\left[\sum_{j=1}^{N}\left(\dfrac{b_je^{-C_{1j}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right)\right]\right]\\ \vdots\\ \widehat{f}_M^{\ell+1}=\nu\left[\log\left(a_Me^{-\frac{\zeta}{\tau}}\right)-\log\left[\sum_{j=1}^{N}\left(\dfrac{b_je^{-C_{Mj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right)\right]\right],\end{cases} \tag{6}$$

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

$$\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)=\sum_{k=1}^{M}\exp\left(\frac{\widehat{f}_k^\ell-C_{kj}}{\epsilon}\right). \tag{7}$$

The proposed procedure can be convergence with theoretical guarantee after $\mathcal{T}$-th inner iteration. Finally, updating variable ζ by further considering $\nabla_\zeta\widehat{L}_{\mathrm{P}}=0$ via $\zeta=\tau[\log(\sum_{i=1}^{M}a_i\exp(-\widehat{f}_i^*/\tau))-\log(\sum_{j=1}^{N}b_j)]$. One can achieve the optimal solution on $\widehat{\boldsymbol{f}}$ and ζ accordingly.

The proof of Proposition 2 can be found in Appendix B. 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

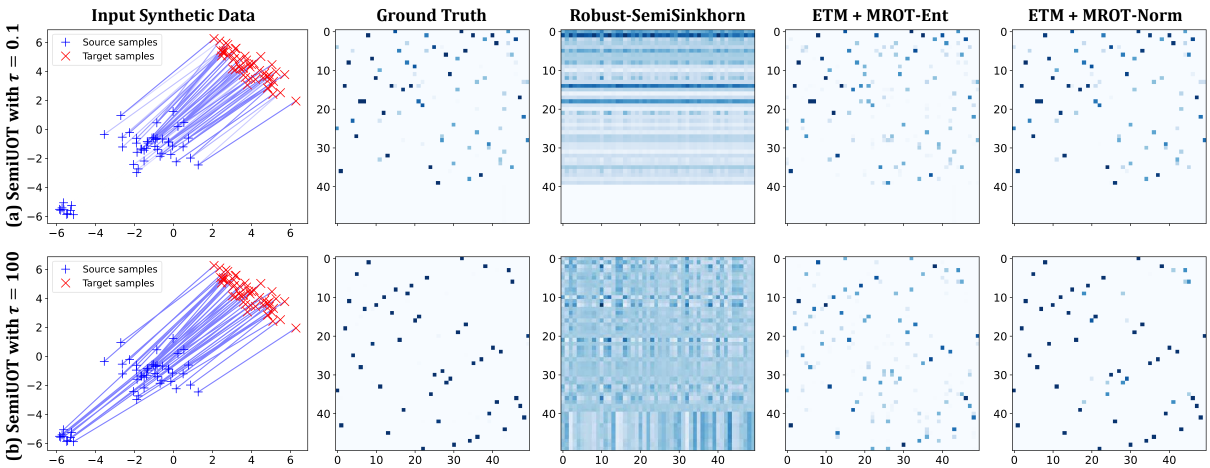


Figure 1. The SemiUOT matching solutions on $\boldsymbol{\pi}^*$ when $\tau = 0.1$ or $\tau = 100$ among the Robust-SemiSinkhorn (Le et al., 2021) and our proposed ETM + MROT-Ent, ETM + MROT-Norm with $\eta_G = 10^2$ and $\epsilon = 10^{-2}$. We set $\eta_{\text{Reg}} = 0.1$ for entropy or L_2-norm regularization term. Our proposed method can avoid ambiguous matching solution and achieve more accurate results.

and implement, while avoiding complex calculations (such as finding the step size and estimating the Hessian matrix) and not requiring a large amount of storage space against previous methods. Therefore, the ETM-Approx approach is an efficient method for determining the optimal 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 a classical optimal transport problem.

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 cost more time on gradient descent 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 $\boldsymbol{f}^*$ in the quasi-Newton optimization procedure on L_{P} (Jin & Mokhtari, 2021; 2023; Rodomanov & Nesterov, 2021) and we regard it as ETM-Refine method. In summary, one can utilize ETM-based approach (e.g., ETM-Exact, ETM-Approx and ETM-Refine) to transform SemiUOT into classic optimal transport problem. We also summarize the optimization details in Appendix C.

Equivalent Transformation Mechanism 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 our methods for solving the marginal probability on UOT which is also a commonly exist optimization problem. That is, we will generalize ETM-based method on solving UOT problem accordingly.

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

$$\min_{\boldsymbol{u},\boldsymbol{v},\zeta} \left[\tau_a \sum_{i=1}^M a_i e^{-\frac{u_i+\zeta}{\tau_a}} + \tau_b \sum_{j=1}^N b_j e^{-\frac{v_j-\zeta}{\tau_b}} \right] \tag{8}$$
$$s.t.\ u_i + v_j + s_{ij} = C_{ij}, \quad s_{ij} \geq 0.$$

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

$$\min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\text{U}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle\ s.t. \begin{cases} \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\dfrac{\boldsymbol{u}^* + \zeta^*}{\tau_a}\right) = \boldsymbol{\alpha} \\ \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \odot \exp\left(-\dfrac{\boldsymbol{v}^* - \zeta^*}{\tau_b}\right) = \boldsymbol{\beta} \end{cases} \tag{9}$$

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

The proof of Proposition 3 can be found in Appendix D. Likewise, we set $v_j = \inf_{k\in[M]} (C_{kj} - u_k)$ according to the c-transform theorem (Villani et al., 2009) to simplify the calculation. Therefore we can obtain *Exact UOT Equation*:

$$\min_{\boldsymbol{u},\zeta} L_{\text{U}} = \tau_a \sum_{i=1}^M a_i e^{-\frac{u_i+\zeta}{\tau_a}} + \tau_b e^{\frac{\zeta}{\tau_b}} \sum_{j=1}^N b_j e^{\frac{\sup_{k\in[M]}(u_k - C_{kj})}{\tau_b}}. \tag{10}$$

We first fix ζ then adopting L-BFGS method to optimize L_{U}. Then we 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 $v_j^\ell = \inf_{k\in[M]}(C_{kj} - u_k^\ell)$ 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 the similar function approximation to replace $\sup(\cdot)$ following Eq.(4), 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.(8) is convex, we can also utilize block gradient descend to optimize the

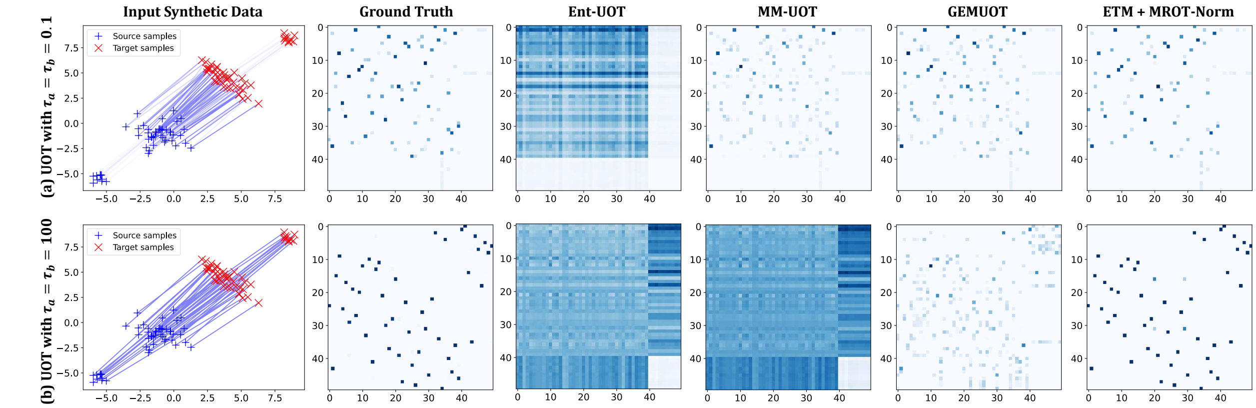


Figure 2. The UOT matching solutions on $\boldsymbol{\pi}^*$ when $\tau_a = \tau_b = 0.1$ or $\tau_a = \tau_b = 100$ among the Ent-UOT (Pham et al., 2020), MM-UOT (Chapel et al., 2021), GEMUOT (Nguyen et al., 2023) and our proposed ETM + MROT-Norm with $\eta_G = 10^2$ and $\eta_{\text{Reg}} = 0.1$.

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.(4) with $\boldsymbol{b} \odot \exp(-(\widehat{\boldsymbol{v}} - \zeta)/\tau_b) = \boldsymbol{\beta}$ accordingly to transform UOT into SemiUOT problem:

$$\min_{\widehat{\boldsymbol{u}}} \widehat{L}_{\text{U}}^u = \tau_a \sum_{i=1}^{M} a_i e^{-\frac{\widehat{u}_i + \zeta}{\tau_a}} + \sum_{j=1}^{N} \beta_j \left[\log \left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}} \right]^{\epsilon} + \zeta \right]. \tag{11}$$

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

$$\frac{\partial \widehat{L}_{\text{U}}^u}{\partial \widehat{u}_s} = -a_s e^{-\frac{\widehat{u}_s + \zeta}{\tau_a}} + e^{\frac{\widehat{u}_s}{\epsilon}} \sum_{j=1}^{N} \left[\frac{\beta_j \exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{\widehat{u}_k - C_{kj}}{\epsilon}\right)} \right] = 0. \tag{12}$$

Obviously, it is equivalent to replace $\boldsymbol{b}$ with $\boldsymbol{\beta}$ in Eq.(5) for solving Eq.(12). Then we can utilize the iteration step shown in Eq.(6) 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 E.

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.(10) as ETM-Refine. Hence, UOT can be transformed into classic optimal transport using the ETM-based method.

3.2. 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 UOT and SemiUOT with commonly used KL Divergence via proposed ETM-based method. 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 classical optimal transport via adopting proposed ETM approach flexibly. We can further utilize existing OT solver for solving $\boldsymbol{\pi}^*$ as:*

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

This observation provides us with entirely new unified insights into solving the matching results of $\boldsymbol{\pi}^*$ for UOT and SemiUOT. 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. Specifically, one can further adopt Sinkhorn (Cuturi, 2013; Carlier, 2022), ℓ_2-norm term (Blondel et al., 2018) or some other sparsification OT solver (Liu et al., 2023; Genevay et al., 2016) with different regularization terms to achieve the transportation solution on $\boldsymbol{\pi}^*$.

Although previous efficient OT solvers (e.g., Sinkhorn (Cuturi, 2013)) could figure out $\boldsymbol{\pi}^*$ efficiently, they often produce ambiguous matching results that may deviate significantly from the correct solutions (Montesuma et al., 2023; Liu et al., 2023). Therefore, finding an accurate matching solution for $\boldsymbol{\pi}^*$ efficiently remains a challenge. Recalling the whole process of ETM-based approach, we not only obtain the marginal probabilities for each data sample, but also derive the multipliers $\boldsymbol{s}$ which can be further utilized.

Corollary 2. *Given the optimal $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$ in UOT via ETM-based approach, 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 approach as $s_{ij} = \max(0, C_{ij} - f_i^* - g_j^*)$. Multipliers $\boldsymbol{s}$ indicates 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 calculating $\boldsymbol{\pi}^*$.

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):*

$$\begin{aligned} &\min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{G}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle + \eta_{\mathrm{Reg}} \mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}) \\ &s.t. \ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{\beta}, \end{aligned} \tag{13}$$

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 approach and 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:

$$\max_{\boldsymbol{\psi}, \boldsymbol{\phi}} L_{\mathrm{G}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\mathrm{Reg}} \mathcal{L}^*_{\mathrm{Reg}} \left(\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\mathrm{Reg}}} \right) \tag{14}$$

where $\widetilde{C}_{ij} = C_{ij} + \eta_G s_{ij}$ and $\boldsymbol{\phi}$ 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 $\nabla_{\pi_{ij}} \mathcal{L}_{\mathrm{Reg}}(\pi_{ij}) = (\psi_i + \phi_j - \widetilde{C}_{ij}) / \eta_{\mathrm{Reg}}$.

The deduction process of MROT can be found in Appendix G. That is, minimizing $s_{ij}\pi_{ij}$ to 0 could result in $s_{ij}\pi_{ij} = 0$ which is consistent with the KKT complementary condition. Generally, we can adopt different kinds of regularization term $\mathcal{L}_{\mathrm{Reg}}(\cdot)$ on MROT for 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). The matching results of $\boldsymbol{\pi}$ in MROT-Ent can be obtained as:

$$\pi_{ij} = \exp\left(-\frac{\eta_G s_{ij}}{\eta_{\mathrm{Reg}}} \right) \exp\left(\frac{\psi_i + \phi_j - C_{ij}}{\eta_{\mathrm{Reg}}} \right) \tag{15}$$

Obviously, the multipliers information $\boldsymbol{s}$ has been involved for achieving 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. More discussions on MROT-Norm can be found in Appendix. In conclusion, we can integrate the ETM-based approach with MROT method to solve UOT and SemiUOT, achieving accurate results for both marginal probabilities and the matching solution π_{ij}.

4. 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 (Peyré et al., 2019; De Plaen et al., 2023; Séjourné et al., 2019; Le et al., 2022). 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 (Sinkhorn & Knopp, 1967; Balaji et al., 2020). Latest, (Chapel et al., 2021) further considers UOT without entropy terms by Majorization-Minimization (MM) (Chizat et al., 2018; Sun et al., 2016) or regularization path methods (Mairal & Yu, 2012; Massias et al., 2018; Liu & Nocedal, 1989). However, the nature of MM algorithm inherits inexact proximal point of KL term (Xie et al., 2020), 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. (2) *Related works on SemiUOT*: SemiUOT with KL divergence only relaxes one of the marginal constraints comparing with UOT. (Le et al., 2021) first fully investigated the corresponding problem and proposed Robust-SemiSinkhorm algorithm. Nevertheless, it still suffers from inaccurate matching solutions with entropy regularization term. Currently, there only exists extremely few works for solving SemiUOT (Montesuma et al., 2024). Therefore, how to efficiently provide accurate solution on both UOT and SemiUOT is still a challenging problem.

5. Experiments

In this section, we conduct experiments on both synthetic and real-world datasets to evaluate our proposed methods.

5.1. Experimental setup

Synthetic Datasets. We first conduct the experiments on the synthetic datasets. That is, we set 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 (Flamary et al., 2021; Chapel et al., 2021). We will sample a number of source and target data via $\boldsymbol{x} \sim \mathbb{P}_X$ and $\boldsymbol{z} \sim \mathbb{P}_Z$ respectively to establish the synthetic datasets.

Datasets for Domain Adaptation. We conduct the unsupervised domain adaptation tasks on *Digits* (Lecun et al., 1998; Hull, 2002; Netzer et al., 2011), *Office-Home* (Venkateswara et al., 2017), and *VisDA* (Peng et al., 2018). More details on these datasets are provided in Appendix H.

Baselines. We compare ETM-Refine with MGOT method with the following state-of-the-art UOT/SemiUOT solvers on the synthetic datasets. (1) **Ent-UOT** (Pham et al., 2020)

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

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 (He et al., 2016)	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 (Cao et al., 2019)	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 (Fatras et al., 2021)	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 (Gu et al., 2021)	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 (Nguyen et al., 2022)	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 (Luo & Ren, 2023)	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 + 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 + 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**
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 (He et al., 2016)	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 (Damodaran et al., 2018)	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 (Balaji et al., 2020)	47.2	71.8	76.4	58.6	68.1	70.2	56.5	45.0	75.8	69.4	52.1	80.6	64.3
JUMBOT (Fatras et al., 2021)	55.2	75.5	80.8	65.5	74.4	74.9	65.2	52.7	79.2	73.0	59.9	83.4	70.0
JUMBOT + UOT(MMUOT)	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 + 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 + 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**

Table 2. Classification accuracy (%) on Digits (Source: LeNet) and VisDA dataset (Source:ResNet50) for UDA 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 (Damodaran et al., 2018)	95.4±0.1	95.6±0.4	96.4±0.3	95.8	68.0
JUMBOT (Fatras et al., 2021)	98.9±0.1	96.7±0.5	98.2±0.1	97.9	72.5
JUMBOT + UOT(ETM + MROT-Ent)	99.4±0.1	98.7±0.3	99.2±0.1	99.1	73.6
JUMBOT + UOT(ETM + MROT-Norm)	99.7±0.1	99.3±0.2	99.6±0.1	**99.5**	**74.2**

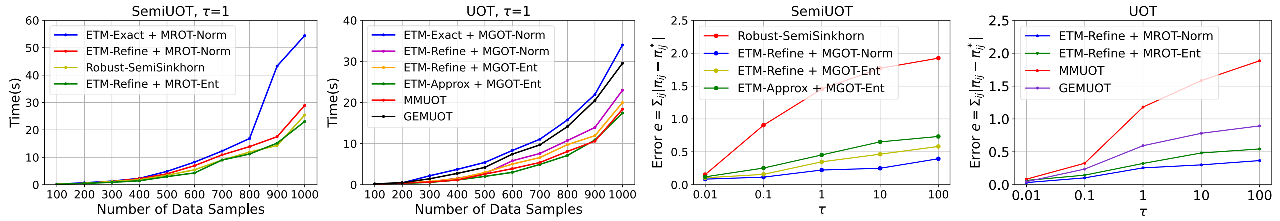
utilizes the entropy regularization term on tackling UOT problem. (2) **MMUOT** (Chapel et al., 2021) adopts majority maximization algorithm for solving UOT. (3) **GEMUOT** (Nguyen et al., 2023) adopts ℓ_2-norm term for reaching transport solutions on UOT which is the state-of-the-art approach. (4) **Robust-SemiSinkhorn** (Le et al., 2021) adopts the entropy regularization term for solving SemiUOT problem. We also involve **DeepJDOT** (Damodaran et al., 2018), **ROT** (Balaji et al., 2020), **JUMBOT** (Fatras et al., 2021), **ETN** (Cao et al., 2019), **AR** (Gu et al., 2021), **m-POT** (Nguyen et al., 2022), **MOT** (Luo & Ren, 2023) as the model baselines for the domain adaptation task. The model details will be provided in Appendix.I.

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., $a_i = \frac{1}{M}$ and $b_j = \frac{1}{N}$. And we adopt square Euclidean distance for the cost C_{ij}. Besides, we follow the same framework and experimental settings as UDA model **JUMBOT** (Fatras et al., 2021) for domain adaptation. Meanwhile, we adopt the same framework and experimental settings as partial UDA model **MOT** (Luo & Ren, 2023) for partial domain adaptation. For all the experiments, we perform five random experiments and report the average results.

5.2. Performance on Synthetic and Real-World Datasets

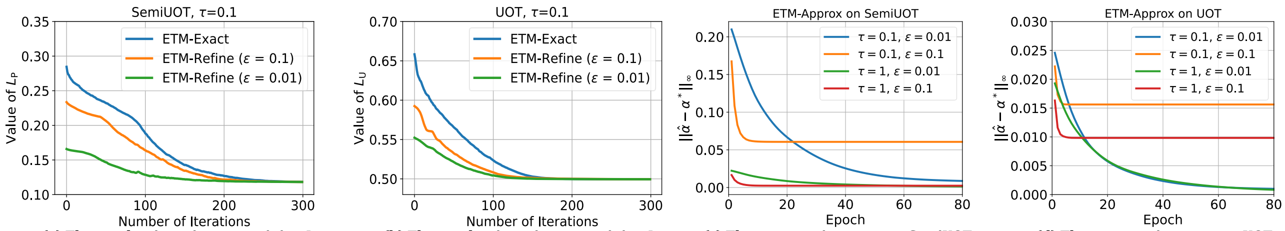
Performance on Synthetic Datasets. We sample 50 data samples on both source and target distributions for finding $\boldsymbol{\pi}^*$ on UOT/SemiUOT. We first set $\tau = \{0.1, 100\}$ on SemiUOT and the matching solutions are shown in Fig.1(a)-(b). Note that we randomly sample 20% of noise in the source datasets. We can observe that previous method **Robust-SemiSinkhorn** could lead to ambiguious results. Our proposed ETM-Refine with MROT+Ent can reach relatively clear 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 transport solutions which are inaccurate. (2) **MMUOT** obtains relatively accurate solutions when τ is small. However, **MMUOT** 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 **MMUOT** 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) ETM-Refine with MGOT-Norm can reach more accurate results with a smaller error compared to the standard UOT solutions.

Performance on Real-World Datasets. We further conduct the experiments on the real-world datasets to validate the our proposed method. The experimental UDA task results on *Office-Home*, *Digits* and *VisDA* are shown in Table.1 and Table.2. We also directly adopt ℓ_2-norm (Blondel et al., 2018) and sparse solver (Liu et al., 2023) on solving $\mathcal{J}_{\mathrm{U}}$. We can observe that replacing entropy-based UOT with other regularization term (e.g., **GEMUOT** or $\mathcal{J}_{\mathrm{U}}$ with ℓ_2-norm) could lead to better results. Moreover, our proposed ETM-Refine with MROT obtains the best performance, which indicates the method efficacy for finding more accurate results. Then we adopt the same experimental protocol as **MOT** to establish the partial UDA task where target label space is a subset of source label space and it is more challenging than classic UDA task (Cao et al., 2018; Luo et al., 2020). The



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

Figure 3. The time consumption and computation error analysis $e = \sum_{ij} ||\pi_{ij} - \pi^*_{ij}||_1$ on UOT and SemiUOT. (Best viewed in color)



(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 values of $\epsilon = \{0.01, 0.1\}$ with the loss descent curve and computation error $e_\alpha = ||\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*||_\infty$ for the proposed ETM-Approx methods on solving SemiUOT and UOT problems.

partial UDA results on *Office-Home* are also shown in Table.1. We can easily observe that MOT + SemiUOT (ETM + MROT-Norm) achieves the best performance on the partial UDA task. UOT relaxes the dual transportation constraints, thus resulting in that some target samples cannot be transported to the source domain. Meanwhile, SemiUOT could overcome the mentioned issue while avoid negative transfer in partial UDA, which boosts the model performance. We also conduct more experiments with applications and please kindly refer to Appendix.I, J, K for more details.

5.3. Analysis

Solver Comparison. To further analyses the proposed ETM-based method with MROT, we conduct the solver comparison on the aspects of *computation time*, *computation error*, and *model performance on real-world datasets*. We first sample the same number of source/target data samples from $\mathbb{P}_X$ and $\mathbb{P}_Z$ respectively. Then we conduct the experiments on both UOT and SemiUOT with $\tau_a = \tau_b = \tau = 1$ and the results are shown in Fig.3(a)-(b). We can conclude that ETM-Exact with MROT-Norm is most time-consuming. Meanwhile ETM-Refine reaches similar computation time with ETM-Approx suggests that utilizing $\widehat{\boldsymbol{u}}^*$ or $\widehat{\boldsymbol{f}}^*$ could accelerate the process for finding $\boldsymbol{u}^*$ or $\boldsymbol{f}^*$ via L-BFGS algorithm. Moreover, we calculate the computation error e between matching solution $\boldsymbol{\pi}$ learned by ETM with MROT and the standard UOT/SemiUOT solution with CVXPY as $\boldsymbol{\pi}^*$ via absolute error $e = \sum_{i,j} ||\boldsymbol{\pi}_{ij} - \boldsymbol{\pi}^*_{ij}||_1$. 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. We also collect the computation error on UOT using **MMUOT** and **GEMUOT**. We can observe that our propose 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. 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 effect 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 less accurate. 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.L, M.

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 UOT/SemiUOT is reweighting data samples accordingly and thus UOT/SemiUOT 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 experiments to demonstrate the superior performance of our proposed ETM with MROT, on both synthetic and real-world datasets of different tasks and applications.

7. Potential Broader Impact

This paper provides a new insight on solving (semi) unbalanced optimal transport problem. Moreover, we first utilize multipliers information into solving transportation solutions. The extensive experiments on both real-world and synthetic datasets with diverse domain adaptation problems show the efficacy of the proposed ETM method with MROT.

References

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. Network flows. 1988.

An, D., Lei, N., Xu, X., and Gu, X. Efficient optimal transport algorithm by accelerated gradient descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 10119–10128, 2022.

Angrist, J. D. and Imbens, G. W. 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.

Arase, Y., Bao, H., and Yokoi, S. Unbalanced optimal transport for unbalanced word alignment. *ACL*, 2023.

Asano, Y. M., Rupprecht, C., and Vedaldi, A. Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*, 2019.

Balaji, Y., Chellappa, R., and Feizi, S. Robust optimal transport with applications in generative modeling and domain adaptation. *Advances in Neural Information Processing Systems*, 33:12934–12944, 2020.

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

Berahas, A. S., Nocedal, J., and Takác, M. A multi-batch l-bfgs method for machine learning. *Advances in Neural Information Processing Systems*, 29, 2016.

Betlei, A., Diemert, E., and Amini, M.-R. Uplift modeling with generalization guarantees. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 55–65, 2021.

Blondel, M., Seguy, V., and Rolet, A. Smooth and sparse optimal transport. In *International conference on artificial intelligence and statistics*, pp. 880–889. PMLR, 2018.

Bonneel, N. and Coeurjolly, D. Spot: sliced partial optimal transport. *ACM Transactions on Graphics (TOG)*, 38(4): 1–13, 2019.

Caffarelli, L. A. and McCann, R. J. Free boundaries in optimal transport and monge-ampere obstacle problems. *Annals of mathematics*, pp. 673–730, 2010.

Cao, Z., Ma, L., Long, M., and Wang, J. Partial adversarial domain adaptation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 135–150, 2018.

Cao, Z., You, K., Long, M., Wang, J., and Yang, Q. Learning to transfer examples for partial domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2985–2994, 2019.

Caputo, B., Müller, H., Martinez-Gomez, J., Villegas, M., Acar, B., Patricia, N., Marvasti, N., Üsküdarlı, S., Paredes, R., Cazorla, M., 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*, pp. 192–211. Springer, 2014.

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

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., and Joulin, A. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020.

Chang, W., Shi, Y., Tuan, H., and Wang, J. Unified optimal transport framework for universal domain adaptation. *Advances in Neural Information Processing Systems*, 35: 29512–29524, 2022.

Chapel, L., Flamary, R., Wu, H., Févotte, C., and Gasso, G. Unbalanced optimal transport through non-negative penalized linear regression. *Advances in Neural Information Processing Systems*, 34:23270–23282, 2021.

Chen, L., Lou, Y., He, J., Bai, T., and Deng, M. Evidential neighborhood contrastive learning for universal domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 6258–6267, 2022.

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

Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609, 2018.

Choi, J., Choi, J., and Kang, M. Generative modeling through the semi-dual formulation of unbalanced optimal transport. *NeurIPS*, 2023.

Chuang, C.-Y., Jegelka, S., and Alvarez-Melis, D. Infoot: Information maximizing optimal transport. In *International Conference on Machine Learning*, pp. 6228–6242. PMLR, 2023.

Courty, N., Flamary, R., Habrard, A., and Rakotomamonjy, A. Joint distribution optimal transportation for domain adaptation. *Advances in neural information processing systems*, 30, 2017.

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

Damodaran, B. B., Kellenberger, B., Flamary, R., Tuia, D., and Courty, N. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 447–463, 2018.

De Plaen, H., De Plaen, P.-F., Suykens, J. A., Proesmans, M., Tuytelaars, T., and Van Gool, L. Unbalanced optimal transport: A unified framework for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3198–3207, 2023.

Dessein, A., Papadakis, N., and Rouas, J.-L. Regularized optimal transport and the rot mover's distance. *The Journal of Machine Learning Research*, 19(1):590–642, 2018.

Farahani, A., Voghoei, S., Rasheed, K., and Arabnia, H. R. A brief review of domain adaptation. *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020*, pp. 877–894, 2021.

Fatras, K., Séjourné, T., Flamary, R., and Courty, N. Unbalanced minibatch optimal transport; applications to domain adaptation. In *International Conference on Machine Learning*, pp. 3186–3197. PMLR, 2021.

Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-i., Trouvé, A., and Peyré, G. Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2681–2690. PMLR, 2019.

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., et al. Pot: Python optimal transport. *The Journal of Machine Learning Research*, 22(1):3571–3578, 2021.

Frogner, C., Zhang, C., Mobahi, H., Araya, M., and Poggio, T. A. Learning with a wasserstein loss. *Advances in neural information processing systems*, 28, 2015.

Fu, B., Cao, Z., Long, M., and Wang, J. 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*, pp. 567–583. Springer, 2020.

Genevay, A., Cuturi, M., Peyré, G., and Bach, F. Stochastic optimization for large-scale optimal transport. *Advances in neural information processing systems*, 29, 2016.

Gu, X., Yu, X., Sun, J., Xu, Z., et al. Adversarial reweighting for partial domain adaptation. *Advances in Neural Information Processing Systems*, 34:14860–14872, 2021.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

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

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

Jin, Q. and Mokhtari, A. Non-asymptotic superlinear convergence of standard quasi-newton methods. *Mathematical Programming*, 200(1):425–473, 2023.

Kantorovich, L. *On the transfer of masses*. 1942.

Kennington, J. L. and Helgason, R. V. *Algorithms for network programming*. John Wiley & Sons, Inc., 1980.

Khrulkov, V., Ryzhakov, G., Chertkov, A., and Oseledets, I. Understanding ddpm latent codes through optimal transport. *ICLR*, 2023.

Korotin, A., Selikhanovych, D., and Burnaev, E. Neural optimal transport. *ICLR*, 2023.

Le, K., Nguyen, H., Nguyen, Q. M., Pham, T., Bui, H., and Ho, N. On robust optimal transport: Computational complexity and barycenter computation. *Advances in Neural Information Processing Systems*, 34:21947–21959, 2021.

Le, K., Nguyen, H., Nguyen, K., Pham, T., and Ho, N. On multimarginal partial optimal transport: Equivalent forms and computational complexity. In *International Conference on Artificial Intelligence and Statistics*, pp. 4397–4413. PMLR, 2022.

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

Li, G., Kang, G., Zhu, Y., Wei, Y., and Yang, Y. Domain consensus clustering for universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9757–9766, 2021.

Li, M., Yu, J., Li, T., and Meng, C. Importance sparsification for sinkhorn algorithm. *arXiv preprint arXiv:2306.06581*, 2023.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. *ICLR*, 2023.

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

Liu, T., Puigcerver, J., and Blondel, M. Sparsity-constrained optimal transport. *ICLR*, 2023.

Lorenz, D. A., Manns, P., and Meyer, C. Quadratically regularized optimal transport. *Applied Mathematics & Optimization*, 83(3):1919–1949, 2021.

Luo, Y.-W. and Ren, C.-X. Mot: Masked optimal transport for partial domain adaptation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3531–3540. IEEE, 2023.

Luo, Y.-W., Ren, C.-X., Dai, D.-Q., and Yan, H. Unsupervised domain adaptation via discriminative manifold propagation. *IEEE transactions on pattern analysis and machine intelligence*, 44(3):1653–1669, 2020.

Ma, Z., Wei, X., Hong, X., Lin, H., Qiu, Y., and Gong, Y. Learning to count via unbalanced optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 2319–2327, 2021.

Mairal, J. and Yu, B. Complexity analysis of the lasso regularization path. *ICML*, 2012.

Massias, M., Gramfort, A., and Salmon, J. Celer: a fast solver for the lasso with dual extrapolation. In *International Conference on Machine Learning*, pp. 3315–3324. PMLR, 2018.

Mathews, J. Numerical methods using matlab, 2004.

Montesuma, E. F., Mboula, F. N., and Souloumiac, A. Recent advances in optimal transport for machine learning. *arXiv preprint arXiv:2306.16156*, 2023.

Montesuma, E. F., Mboula, F. M. N., and Souloumiac, A. Recent advances in optimal transport for machine learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Mukherjee, D., Guha, A., Solomon, J. M., Sun, Y., and Yurochkin, M. Outlier-robust optimal transport. In *International Conference on Machine Learning*, pp. 7850–7860. PMLR, 2021.

Neklyudov, K., Brekelmans, R., Tong, A., Atanackovic, L., Liu, Q., and Makhzani, A. A computational framework for solving wasserstein lagrangian flows. *arXiv preprint arXiv:2310.10649*, 2023.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., and Ng, A. Y. Reading digits in natural images with unsupervised feature learning. *Nips Workshop on Deep Learning & Unsupervised Feature Learning*, 2011.

Nguyen, K., Nguyen, D., Pham, T., Ho, N., et al. Improving mini-batch optimal transport via partial transportation. In *International Conference on Machine Learning*, pp. 16656–16690. PMLR, 2022.

Nguyen, Q. M., Nguyen, H. H., Zhou, Y., and Nguyen, L. M. On unbalanced optimal transport: Gradient methods, sparsity and approximation error. *JMLR*, 2023.

Onken, D., Fung, S. W., Li, X., and Ruthotto, L. Ot-flow: Fast and accurate continuous normalizing flows via optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 9223–9232, 2021.

Pariset, M., Hsieh, Y.-P., Bunne, C., Krause, A., and De Bortoli, V. Unbalanced diffusion schr\" odinger bridge. *arXiv preprint arXiv:2306.09099*, 2023.

Peng, X., Usman, B., Kaushik, N., Wang, D., Hoffman, J., and Saenko, K. Visda: A synthetic-to-real benchmark for visual domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2021–2026, 2018.

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

Pham, K., Le, K., Ho, N., Pham, T., and Bui, H. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *International Conference on Machine Learning*, pp. 7673–7682. PMLR, 2020.

Redko, I., Courty, N., Flamary, R., and Tuia, D. Optimal transport for multi-source domain adaptation under target shift. In *The 22nd International Conference on artificial intelligence and statistics*, pp. 849–858. PMLR, 2019.

Riaz, B., Karahan, Y., and Brockmeier, A. J. Partial optimal transport for support subset selection. *Transactions on Machine Learning Research*, 2023.

Rodomanov, A. and Nesterov, Y. Rates of superlinear convergence for classical quasi-newton methods. *Mathematical Programming*, pp. 1–32, 2021.

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

Saenko, K., Kulis, B., Fritz, M., and Darrell, T. 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*, pp. 213–226. Springer, 2010.

Saito, K., Yamamoto, S., Ushiku, Y., and Harada, T. Open set domain adaptation by backpropagation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 153–168, 2018.

Sato, R., Yamada, M., and Kashima, H. Fast unbalanced optimal transport on a tree. *Advances in neural information processing systems*, 33:19039–19051, 2020.

Scetbon, M., Cuturi, M., and Peyré, G. Low-rank sinkhorn factorization. In *International Conference on Machine Learning*, pp. 9344–9354. PMLR, 2021.

Scetbon, M., Klein, M., Palla, G., and Cuturi, M. Unbalanced low-rank optimal transport solvers. *NeurIPS*, 2023.

Séjourné, T., Feydy, J., Vialard, F.-X., Trouvé, A., and Peyré, G. Sinkhorn divergences for unbalanced optimal transport. 2019.

Séjourné, T., Peyré, G., and Vialard, F.-X. Unbalanced optimal transport, from theory to numerics. *arXiv preprint arXiv:2211.08775*, 2022a.

Séjourné, T., Vialard, F.-X., and Peyré, G. Faster unbalanced optimal transport: Translation invariant sinkhorn and 1-d frank-wolfe. In *International Conference on Artificial Intelligence and Statistics*, pp. 4995–5021. PMLR, 2022b.

Séjourné, T., Bonet, C., Fatras, K., Nadjahi, K., and Courty, N. Unbalanced optimal transport meets sliced-wasserstein. *arXiv preprint arXiv:2306.07176*, 2023.

Shalit, U., Johansson, F. D., and Sontag, D. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, pp. 3076–3085. PMLR, 2017.

Sinkhorn, R. and Knopp, P. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

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

Tong, A., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Fatras, K., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023.

Tran, Q. H., Janati, H., Courty, N., Flamary, R., Redko, I., Demetci, P., and Singh, R. Unbalanced co-optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 10006–10016, 2023.

Van der Maaten, L. and Hinton, G. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Venkateswara, H., Eusebio, J., Chakraborty, S., and Panchanathan, S. Deep hashing network for unsupervised domain adaptation. *CoRR*, abs/1706.07522, 2017.

Villani, C. et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

Wang, H., Chen, Z., Fan, J., Li, H., Liu, T., Liu, W., Dai, Q., Wang, Y., Dong, Z., and Tang, R. Optimal transport for treatment effect estimation. *arXiv preprint arXiv:2310.18286*, 2023.

Xie, Y., Wang, X., Wang, R., and Zha, H. A fast proximal point method for computing exact wasserstein distance. In *Uncertainty in artificial intelligence*, pp. 433–453. PMLR, 2020.

Yao, L., Li, S., Li, Y., Huai, M., Gao, J., and Zhang, A. Representation learning for treatment effect estimation from observational data. *Advances in neural information processing systems*, 31, 2018.

You, K., Long, M., Cao, Z., Wang, J., and Jordan, M. I. Universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2720–2729, 2019.

Zhan, F., Yu, Y., Cui, K., Zhang, G., Lu, S., Pan, J., Zhang, C., Ma, F., Xie, X., and Miao, C. Unbalanced feature transport for exemplar-based image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15028–15038, 2021.

Zhang, L. and Gao, X. Transfer adaptation learning: A decade survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

Zhuang, Y., Chen, X., and Yang, Y. Wasserstein k-means for clustering probability distributions. *Advances in Neural Information Processing Systems*, 35:11382–11395, 2022.

Appendix

A. 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:*

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

where $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{s}$ and ζ denotes Lagrange multipliers. Moreover, SemiUOT problem can be transformed into classic optimal transport as follows:

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

Note that when $\tau\to\infty$, the source marginal probability can be determined as $\boldsymbol{\pi}\mathbf{1}_N=\omega\boldsymbol{a}$ where $\omega=\langle\boldsymbol{b},\mathbf{1}_N\rangle/\langle\boldsymbol{a},\mathbf{1}_M\rangle$.

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

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

Then we can rewrite the optimization problem:

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

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:

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

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

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

Therefore we can obtain the results as:

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

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

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

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

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

where $\mathcal{O}_{\mathrm{Const}} = -\sum_{i=1}^{M}\tau a_i$ 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:

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

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

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

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

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

Therefore we conclude the proof of the Proposition 1. □

B. 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 $\widehat{f}_s$:*

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

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

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

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

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

The proposed procedure can be convergence with theoretical guarantee after $\mathcal{T}$-th inner iteration. Finally, updating the Lagrange multiplier ζ by 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 solution 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:

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

At that time we adopt fixed-point iteration method to optimize $\widehat{\boldsymbol{f}}$ accordingly:

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

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

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

Likewise we can obtain the result:

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

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 method guarantees convergence according to Theorem 2.9 in (Mathews, 2004). □

C. 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 more clear illustration.

Algorithm 1 The algorithm of ETM-Based method on SemiUOT

Input: $\boldsymbol{C}$: cost matrix; $\boldsymbol{a}, \boldsymbol{b}$: initial marginal probability; τ, ϵ: 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_{P} as:

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

Optimize $\boldsymbol{g}$ via $g_j = \inf\limits_{k\in[M]} (C_{kj} - f_k^t)$.

Optimize ζ via $\zeta = \tau[\log(\sum_{i=1}^{M} a_i \exp(-f_i/\tau)) - \log(\sum_{j=1}^{N} b_j)]$ as shown in Eq.(26).

Return: The optimal solutions of $\boldsymbol{f}^*$, $\boldsymbol{g}^*$ and ζ^*.

(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 $\widehat{L}_{\text{P}}$ as:

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

Optimize $\widehat{\boldsymbol{g}}$ via $\widehat{g}_j = -\epsilon \log[\sum_{k=1}^{M} \exp((\widehat{f}_k - C_{kj})/\epsilon)]$.

Optimize ζ via $\zeta = \tau[\log(\sum_{i=1}^{M} a_i \exp(-\widehat{f}_i/\tau)) - \log(\sum_{j=1}^{N} b_j)]$ as shown in Eq.(26).

Return: The optimal solutions of $\widehat{\boldsymbol{f}}^*$, $\widehat{\boldsymbol{g}}^*$ and ζ^*.

(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 ζ^*.

D. Proof of Proposition 3

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

$$\min_{\boldsymbol{u},\boldsymbol{v},\zeta}\left[\tau_a\sum_{i=1}^{M}a_i e^{-\frac{u_i+\zeta}{\tau_a}}+\tau_b\sum_{j=1}^{N}b_j e^{-\frac{v_j-\zeta}{\tau_b}}\right] \quad s.t.\ u_i+v_j+s_{ij}=C_{ij},\quad s_{ij}\geq 0. \tag{37}$$

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

$$\begin{aligned}&\min_{\boldsymbol{\pi}\geq 0}\mathcal{J}_{\mathrm{U}}=\langle\boldsymbol{C},\boldsymbol{\pi}\rangle\\ &s.t.\begin{cases}\boldsymbol{\pi}\mathbf{1}_N=\boldsymbol{a}\odot\exp\left(-\dfrac{\boldsymbol{u}^*+\zeta^*}{\tau_a}\right)=\boldsymbol{\alpha}\\ \boldsymbol{\pi}^\top\mathbf{1}_M=\boldsymbol{b}\odot\exp\left(-\dfrac{\boldsymbol{v}^*-\zeta^*}{\tau_b}\right)=\boldsymbol{\beta}\end{cases}\end{aligned} \tag{38}$$

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

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

$$\begin{aligned}&\min_{\boldsymbol{\pi}\geq 0}J=\langle\boldsymbol{C},\boldsymbol{\pi}\rangle+\tau_a\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N\|\boldsymbol{a}\right)+\tau_b\mathrm{KL}(\boldsymbol{\pi}^\top\mathbf{1}_M\|\boldsymbol{b})\\ &s.t.\ (\text{Optional}):\ \ \boldsymbol{\pi}\mathbf{1}_N=\boldsymbol{\alpha},\quad \boldsymbol{\pi}^\top\mathbf{1}_M=\boldsymbol{\beta}\end{aligned} \tag{39}$$

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:

$$\max_{\boldsymbol{s}\geq 0,\boldsymbol{u},\boldsymbol{v},\zeta}\min_{\boldsymbol{\pi}\geq 0}\mathcal{J}_{\mathrm{UOT}}=\tau_a\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N\|\boldsymbol{a}\right)+\langle\boldsymbol{u}+\zeta,\boldsymbol{\pi}\mathbf{1}_N\rangle+\tau_b\mathrm{KL}(\boldsymbol{\pi}^\top\mathbf{1}_M\|\boldsymbol{b})+\langle\boldsymbol{v}-\zeta,\boldsymbol{\pi}^\top\mathbf{1}_M\rangle+\mathscr{C}_{\mathrm{UOT}} \tag{40}$$

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:

$$\begin{aligned}\frac{\partial\mathcal{J}_{\mathrm{UOT}}}{\partial\pi_{ij}}&=\left[\tau_a\log\frac{\sum_{j=1}^{N}\pi_{ij}}{a_i}+u_i+\zeta\right]+\left[\tau_b\log\frac{\sum_{i=1}^{M}\pi_{ij}}{b_j}+v_j-\zeta\right]+(C_{ij}-u_i-v_j-s_{ij})\\ &=C_{ij}+\tau_a\log\frac{\sum_{j=1}^{N}\pi_{ij}}{a_i}+\tau_b\log\frac{\sum_{i=1}^{M}\pi_{ij}}{b_j}-s_{ij}=0\end{aligned} \tag{41}$$

Then we can obtain the results:

$$\begin{cases}\displaystyle\sum_{j=1}^{N}\pi_{ij}=a_i\exp\left(-\frac{u_i+\zeta}{\tau_a}\right)\\ \displaystyle\sum_{i=1}^{M}\pi_{ij}=b_j\exp\left(-\frac{v_j-\zeta}{\tau_b}\right)\\ C_{ij}-u_i-v_j-s_{ij}=0\end{cases} \tag{42}$$

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

$$\begin{cases}\tau_a\mathrm{KL}\left(\boldsymbol{\pi}\mathbf{1}_N\|\boldsymbol{a}\right)+\langle\boldsymbol{u}+\zeta,\boldsymbol{\pi}\mathbf{1}_N\rangle=\displaystyle\sum_{i=1}^{M}\left[-\tau_a a_i\exp\left(-\frac{f_i+\zeta}{\tau_a}\right)+\tau_a a_i\right]\\ \tau_b\mathrm{KL}\left(\boldsymbol{\pi}^\top\mathbf{1}_M\|\boldsymbol{b}\right)+\langle\boldsymbol{v}-\zeta,\boldsymbol{\pi}^\top\mathbf{1}_M\rangle=\displaystyle\sum_{j=1}^{N}\left[-\tau_b b_j\exp\left(-\frac{g_j-\zeta}{\tau_b}\right)+\tau_b b_j\right]\end{cases} \tag{43}$$

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

$$\begin{aligned}
\min_{\boldsymbol{u},\boldsymbol{v},\zeta} \mathcal{J}_{\text{UOT}} &= -\tau_a \text{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) - \langle \boldsymbol{u}+\zeta, \boldsymbol{\pi}\mathbf{1}_N \rangle - \tau_b \text{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}) - \langle \boldsymbol{v}-\zeta, \boldsymbol{\pi}^\top \mathbf{1}_M \rangle \\
&= \tau_a \exp\left(-\frac{\zeta}{\tau_a}\right) \left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle + \mathcal{O}_{\text{Const}} \\
s.t.\ & u_i + v_j \leq C_{ij}
\end{aligned} \tag{44}$$

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:

$$\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b}\left[\log\left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle - \log\left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle\right]. \tag{45}$$

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

$$\begin{aligned}
\lim_{\tau_a \to +\infty, \tau_b \to +\infty} a_i \exp\left(-\frac{u_i+\zeta}{\tau_a}\right) &= \lim_{\tau_a \to +\infty, \tau_b \to +\infty} a_i \exp\left(-\frac{\zeta}{\tau_a}\right) = a_i \sqrt{\frac{\langle \boldsymbol{b}, \mathbf{1}_N \rangle}{\langle \boldsymbol{a}, \mathbf{1}_M \rangle}} = \sqrt{\omega} a_i \\
\lim_{\tau_a \to +\infty, \tau_b \to +\infty} b_j \exp\left(-\frac{v_j-\zeta}{\tau_b}\right) &= \lim_{\tau_a \to +\infty, \tau_b \to +\infty} b_j \exp\left(-\frac{\zeta}{\tau_b}\right) = b_j \sqrt{\frac{\langle \boldsymbol{a}, \mathbf{1}_M \rangle}{\langle \boldsymbol{b}, \mathbf{1}_N \rangle}} = \frac{1}{\sqrt{\omega}} b_j
\end{aligned} \tag{46}$$

Therefore we conclude the proof of the Proposition 3. □

E. 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:

$$\begin{aligned}
\min_{\boldsymbol{u},\zeta} L_{\text{U}} &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i+\zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(-\frac{v_j}{\tau_b}\right) \\
&= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i+\zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(\frac{\sup_{k\in[M]}(u_k - C_{kj})}{\tau_b}\right)
\end{aligned} \tag{47}$$

where $v_j = -\sup_{k\in[M]}(u_k - C_{kj})$ 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.(8) is convex, we can also utilize block gradient descend 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.(4) with $\boldsymbol{\beta}$ accordingly to transform UOT into SemiUOT problem:

$$\begin{aligned}
&\min_{\boldsymbol{\pi}\geq 0} J_{\text{U}}^{u} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau_a \text{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) \\
&s.t. \begin{cases} (\text{Constraint}): \ \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \odot \exp\left(-\dfrac{\boldsymbol{v}-\zeta}{\tau_b}\right) = \boldsymbol{\beta} \\ (\text{Optional}): \ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\dfrac{\boldsymbol{u}+\zeta}{\tau_a}\right) = \boldsymbol{\alpha} \end{cases}
\end{aligned} \tag{48}$$

At that time, the Fenchel-Lagrange multipliers form of Eq.(48) is given via the Proposition 1:

$$\begin{aligned}
\min_{\boldsymbol{u}} L_{\text{U}}^{u} &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widetilde{u}_i+\zeta}{\tau_a}\right) - \sum_{j=1}^{N} \beta_j(\widetilde{v}_j - \zeta) \\
&= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i+\zeta}{\tau_a}\right) - \sum_{j=1}^{N} \left(\inf_{k\in[M]}[C_{kj} - u_k] - \zeta\right) \beta_j
\end{aligned} \tag{49}$$

Note that $\widetilde{\boldsymbol{u}}$ and $\widetilde{\boldsymbol{v}}$ denote the Lagrange multiplier for Eq.(48) while we have $\widetilde{v}_j = \inf_{k\in[M]}[C_{kj} - u_k] = v_j$ 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:

$$\begin{aligned}\min_{\widehat{\boldsymbol{u}}} \widehat{L}^u_{\mathrm{U}} &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i + \zeta}{\tau_a}\right) + \sum_{j=1}^{N} \beta_j \left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}}\right] + \zeta\right] \\ &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i + \zeta}{\tau_a}\right) + \sum_{j=1}^{N} b_j \exp\left(-\frac{\widehat{v}_j - \zeta}{\tau_b}\right)\left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}}\right] + \zeta\right]\end{aligned} \tag{50}$$

At that time we adopt fixed-point iteration method to optimize $\widehat{\boldsymbol{u}}$ accordingly based on the Proposition 2:

$$\begin{cases} \widehat{u}_1^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a + \epsilon}\left[\log\left(a_1 e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N}\left(\frac{\beta_j e^{-C_{1j}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^{\ell})}\right)\right]\right] = \mathcal{U}_1\left(\widehat{u}_1^{\ell}, \cdots, \widehat{u}_M^{\ell}\right) \\ \vdots \\ \widehat{u}_s^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a + \epsilon}\left[\log\left(a_s e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N}\left(\frac{\beta_j e^{-C_{sj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^{\ell})}\right)\right]\right] = \mathcal{U}_s\left(\widehat{u}_1^{\ell}, \cdots, \widehat{u}_M^{\ell}\right) \\ \vdots \\ \widehat{u}_M^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a + \epsilon}\left[\log\left(a_M e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N}\left(\frac{\beta_j e^{-C_{Mj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^{\ell})}\right)\right]\right] = \mathcal{U}_M\left(\widehat{u}_1^{\ell}, \cdots, \widehat{u}_M^{\ell}\right), \end{cases} \tag{51}$$

The iteration process can be shown to converge 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 $\zeta = (\tau_a\tau_b/(\tau_a + \tau_b))[\log(\sum_{i=1}^{M} a_i \exp(-\widehat{u}_i^*/\tau_a)) - \log(\sum_{j=1}^{N} b_j \exp(-\widehat{v}_j^*/\tau_b))]$.

F. 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.

G. 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):*

$$\begin{aligned}&\min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{G}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle + \eta_{\mathrm{Reg}} \mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}) \\ &s.t.\ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{\beta},\end{aligned} \tag{52}$$

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 approach and 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:

$$\max_{\boldsymbol{\psi}, \boldsymbol{\phi}} L_{\mathrm{G}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\mathrm{Reg}} \mathcal{L}^*_{\mathrm{Reg}}\left(\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\mathrm{Reg}}}\right) \tag{53}$$

where $\widetilde{C}_{ij} = C_{ij} + \eta_G s_{ij}$ and $\boldsymbol{\phi}$ 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 $\nabla_{\pi_{ij}} \mathcal{L}_{\mathrm{Reg}}(\pi_{ij}) = (\psi_i + \phi_j - \widetilde{C}_{ij})/\eta_{\mathrm{Reg}}$.

Algorithm 2 The algorithm of ETM-Based method on UOT

Input: $\boldsymbol{C}$: cost matrix; $\boldsymbol{a}, \boldsymbol{b}$: initial marginal probability; τ_a, τ_b, ϵ: 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_{U} as:

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

Optimize $\boldsymbol{v}$ via $v_j = \inf_{k\in[M]} (C_{kj} - u_k)$.
Optimize ζ 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.(45).
Return: The optimal solutions of $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and ζ^*.
(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 ζ 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.(45).
Return: The optimal solutions of $\widehat{\boldsymbol{u}}^*$, $\widehat{\boldsymbol{v}}^*$ and ζ^*.
(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 ζ^*.

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

$$\begin{aligned}
\max_{\boldsymbol{\psi},\boldsymbol{\phi}} \min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\text{MROT}} &= \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s}\rangle + \eta_{\text{Reg}} \mathcal{L}_{\text{Reg}}(\boldsymbol{\pi}) - \langle \boldsymbol{\psi}, \boldsymbol{\pi}\mathbf{1}_N - \boldsymbol{\alpha}\rangle - \langle \boldsymbol{\phi}, \boldsymbol{\pi}^\top \mathbf{1}_M - \boldsymbol{\beta}\rangle \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle + \eta_{\text{Reg}} \inf_{\boldsymbol{\pi}} \left[\sum_{i,j}\left[\frac{C_{ij} + \eta_G s_{ij} - f_i - g_j}{\eta_{\text{Reg}}}\pi_{ij} + \mathcal{L}_{\text{Reg}}(\pi_{ij})\right]\right] \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle - \eta_{\text{Reg}} \sup_{\boldsymbol{\pi}} \left[\sum_{i,j}\left[\frac{f_i + g_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\pi_{ij} - \mathcal{L}_{\text{Reg}}(\pi_{ij})\right]\right] \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle - \eta_{\text{Reg}} \mathcal{L}_{\text{Reg}}^{*}\left(\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right)
\end{aligned} \tag{54}$$

At that time we have the following results:

$$\begin{cases} \dfrac{\partial \mathcal{J}_{\text{MROT}}}{\partial \psi_i} = 0 \\ \dfrac{\partial \mathcal{J}_{\text{MROT}}}{\partial \phi_j} = 0 \end{cases} \Rightarrow \begin{cases} \nabla_{\psi_i} \mathcal{L}_{\text{Reg}}^{*}\left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right) = \alpha_i \\ \nabla_{\phi_j} \mathcal{L}_{\text{Reg}}^{*}\left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right) = \beta_j \end{cases} \tag{55}$$

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

$$\frac{\partial \mathcal{J}_{\text{MROT}}}{\partial \pi_{ij}} = \widetilde{C}_{ij} + \eta_{\text{Reg}} \nabla_{\pi_{ij}} \mathcal{L}_{\text{Reg}}(\pi_{ij}) - \psi_i - \phi_j = 0 \tag{56}$$

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

$$\begin{cases} \max_{\boldsymbol{\psi},\boldsymbol{\phi}} \mathcal{J}_{\text{MROT-Ent}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle - \eta_{\text{Reg}} \sum_{i,j} \exp\left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right) \\ \pi_{ij} = \exp\left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right) \end{cases} \tag{57}$$

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

$$\begin{cases} \max_{\boldsymbol{\psi},\boldsymbol{\phi}} \mathcal{J}_{\text{MROT-Norm}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle - \dfrac{\eta_{\text{Reg}}}{2} \sum_{i,j} \left[\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right]_+^2 \\ \pi_{ij} = \left[\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right]_+ \end{cases} \tag{58}$$

Therefore we conclude the proof of the Proposition 4. □

H. Datasets on Domain Adaptations

Datasets. We conduct the unsupervised domain adaptation tasks on *Digits*, *Office-Home*, and *VisDA*. *Digits* is the classical dataset for digit classification which contains three standard digit classification datasets: **MNIST** (Lecun et al., 1998), **USPS**(Hull, 2002) and **SVHN** (Netzer et al., 2011). Each dataset consists of 10 classes of digits, ranging from 0 to 9. *Office-Home* (Venkateswara et al., 2017) 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* (Peng et al., 2018) is the large-scale cross-domain dataset in computer vision on two domains, i.e., **Synthetic** and **Real** with 280K images in 12 classes.

I. Experiments on Partial Domain Adaptations

Datasets. We further conduct the domain adaptation tasks on new datasets, i.e., *Office-31* (Saenko et al., 2010) and *ImageCLEF* (Caputo et al., 2014). **Office-31** is the commonly-used computer vision dataset for domain adaptation 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 (Cao et al., 2018). **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 (Luo et al., 2020).

Baselines. We involve **DeepJDOT** (Damodaran et al., 2018), **ROT** (Balaji et al., 2020), **JUMBOT** (Fatras et al., 2021), **ETN** (Cao et al., 2019), **AR** (Gu et al., 2021), **m-POT** (Nguyen et al., 2022), **MOT** (Luo & Ren, 2023), **DMP** (Luo et al., 2020) as the model baselines for the domain adaptation task.

- **DeepJDOT** (Damodaran et al., 2018) first adopts optimal transport into solving domain adaptation problem with deep learning framework.
- **ROT** (Balaji et al., 2020) adopts robust optimal transport into adversarial training for domain adaptation.
- **JUMBOT** (Fatras et al., 2021) adopts mini-batch unbalanced optimal transport method for domain adaptation.
- **ETN** (Cao et al., 2019) utilizes example transfer network to jointly learn domain-invariant representations and the progressive weighting scheme.
- **AR** (Gu et al., 2021) adopts adversarial reweighting strategy on source domain data for alignment.
- **m-POT** (Nguyen et al., 2022) adopts partial optimal transport method in the mini-batch settings for domain adaptation.
- **DMP** (Luo et al., 2020) adopt discriminative manifold propagation for domain adaptation.
- **MOT** (Luo & Ren, 2023) adopts masked unbalanced optimal transport technique on considering label information for PDA tasks.

Table 3. Classification accuracy (%) on *Office-31* for partial unsupervised domain adaptation

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg
ResNet (He et al., 2016)	75.6	96.3	98.1	83.4	83.9	85.0	87.1
ETN (Cao et al., 2019)	84.7	97.4	99.2	91.3	90.2	92.8	92.6
JUMBOT (Fatras et al., 2021)	90.2	98.9	99.3	94.5	93.8	93.4	95.0
AR (Gu et al., 2021)	93.5	**100.0**	99.7	96.8	95.5	96.0	96.9
m-POT (Nguyen et al., 2022)	96.2	99.5	**100.0**	97.6	94.4	95.3	97.2
MOT (Luo & Ren, 2023)	99.3	**100.0**	**100.0**	98.7	96.1	96.4	98.4
MOT + UOT(ETM + MROT-Ent)	99.4	**100.0**	**100.0**	98.9	96.8	97.3	98.7
MOT + UOT(ETM + MROT-Norm)	99.6	**100.0**	**100.0**	99.2	97.3	97.7	99.0
MOT + SemiUOT(ETM + MROT-Ent)	**99.8**	**100.0**	**100.0**	99.4	97.8	98.4	99.2
MOT + SemiUOT(ETM + MROT-Norm)	99.7	**100.0**	**100.0**	**99.7**	**98.4**	**98.6**	**99.4**

Table 4. Classification accuracy (%) on *ImageCLEF* for partial unsupervised domain adaptation

Method	I→P	P→I	I→C	C→I	C→P	P→C	Avg
ResNet (He et al., 2016)	78.3	86.9	91.0	84.3	72.5	91.5	84.1
ETN (Cao et al., 2019)	79.6	88.5	92.9	87.2	74.1	93.4	86.0
JUMBOT (Fatras et al., 2021)	80.1	91.3	93.6	90.9	75.7	94.2	87.6
AR (Gu et al., 2021)	83.1	92.8	94.5	92.4	76.3	95.0	89.0
DMP (Luo et al., 2020)	82.4	94.5	96.7	94.3	78.7	96.4	90.5
MOT (Luo & Ren, 2023)	87.7	95.0	98.0	95.0	87.0	98.7	93.6
MOT + UOT(ETM + MROT-Ent)	88.3	95.6	98.4	95.3	87.6	99.0	94.0
MOT + UOT(ETM + MROT-Norm)	88.7	95.9	98.7	95.8	88.0	99.1	94.4
MOT + SemiUOT(ETM + MROT-Ent)	89.1	96.2	99.2	96.1	88.5	99.4	94.8
MOT + SemiUOT(ETM + MROT-Norm)	**89.6**	**96.7**	**99.4**	**96.5**	**89.1**	**99.6**	**95.2**

Performance. We also conduct the partial domain adaptation tasks on Office-31 and ImageCLEF and the results are shown in Table.4-5. We can observe that the proposed ETM-Refine with MROT-Norm on SemiUOT achieves state-of-the-art performance on Office-31 and ImageCLEF.

J. 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 (Farahani et al., 2021; Zhang & Gao, 2022). 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** (Saito et al., 2018) adopts domain adversarial learning for open-set domain adaptation, (2) **UAN** (You et al., 2019) utilizes transferability criterion for universal domain adaptation, (3) **CMU** (Fu et al., 2020) learns to detect open classes with uncertainty estimation, (4) **DCC** (Li et al., 2021) adopts domain consensus clustering for adaptation, (5) **TNT** (Chen et al., 2022) adopts evidential neighborhood contrastive learning for adaptation, (6) **UniOT** (Chang et al., 2022) adopts unbalanced optimal transport with adaptive filtering for transferring.

We adopt the same experimental settings as UniOT (Chang et al., 2022). We utilize the commonly-used H-score (Fu et al., 2020) to validate the final results as shown in Table.6-7. 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 + MROT) reaches the best performance, indicating that UOT with ETM + MROT can provide more accurate matching results. Moreover, we adopt T-SNE method (Van der Maaten & Hinton, 2008) to plot the source and target data features in the latent space as shown in Fig.8(a)-(b). We can find that: (1) original UniOT could lead to rather scattered features in the latent space. That is because UniOT with entropic UOT could lead to dense and inaccurate matching solutions which limits the model potentials. (2) Our proposed UniOT + UOT(ETM + MROT-Norm) can provide more compact features since it provides more accurate solutions. Thus it further illustrates the efficacy of our proposed ETM + MROT-Norm method.

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 (He et al., 2016)	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 (Saito et al., 2018)	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 (You et al., 2019)	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 (Fu et al., 2020)	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 (Li et al., 2021)	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 (Chen et al., 2022)	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 (Chang et al., 2022)	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 + 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 + 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**

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 (He et al., 2016)	49.78	47.92	48.48	54.94	48.96	55.60	50.94
OSBP (Saito et al., 2018)	51.14	50.23	49.75	55.53	50.16	57.20	52.34
UAN (You et al., 2019)	59.68	58.61	60.11	70.62	60.34	71.42	63.46
CMU (Fu et al., 2020)	68.11	67.33	71.42	79.32	72.23	80.42	73.14
DCC (Li et al., 2021)	88.50	78.54	70.18	79.29	75.87	88.58	80.16
TNT (Chen et al., 2022)	85.70	80.40	83.80	92.00	79.10	91.20	85.37
UniOT (Chang et al., 2022)	86.97	88.48	88.35	98.83	87.60	96.57	91.13
UniOT + UOT(ETM + MROT-Ent)	88.25	89.62	89.47	99.48	89.10	97.94	92.31
UniOT + UOT(ETM + MROT-Norm)	**88.67**	**90.14**	**90.03**	**99.58**	**89.42**	**98.46**	**92.72**

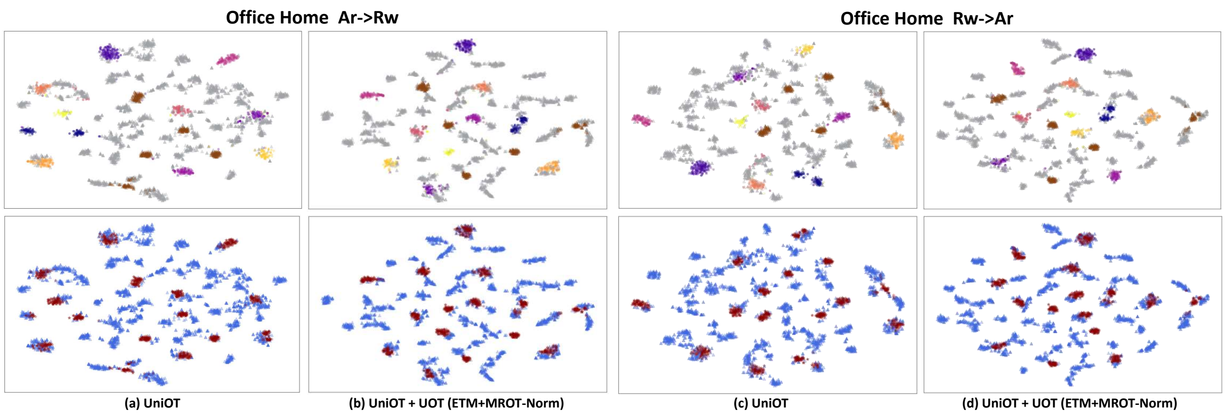


Figure 5. The T-SNE of data features on Ar → Rw (Office-Home) and Rw → Ar (Office-Home). The first row shows the original data sample distribution: The brown and gray colors denote the source and target private classes respectively. The rest are common label set with different colors. The second row indicates the mapping between source and target domain: The red and blue points denote the source and target samples respectively.

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 (Angrist & Imbens, 1995)	3.749	4.340	0.843	0.496	3.856	5.674	0.652	0.492
TARNet (Shalit et al., 2017)	3.236	3.254	**0.886**	0.662	0.749	1.788	0.654	0.711
PSM (Rosenbaum & Rubin, 1983)	5.228	5.094	0.884	0.745	3.219	4.634	0.740	0.681
CFR-WASS (Shalit et al., 2017)	3.128	3.207	0.873	0.669	0.657	1.704	0.656	0.715
ESCFR (Wang et al., 2023)	2.252	2.316	0.796	0.754	0.502	1.282	0.665	0.719
ESCFR + UOT(ETM + MROT-Ent)	2.327	2.261	0.839	0.814	0.497	1.275	0.769	0.763
ESCFR + UOT(ETM + MROT-Norm)	**2.104**	**2.216**	0.883	**0.839**	**0.475**	**1.146**	**0.798**	**0.802**

K. Experiments on Treatment Effect Estimation

Datasets for Treatment Effect Estimation. We further conduct ETM on treatment effect estimation with two semi-synthetic datasets IHDP (Shalit et al., 2017) and ACIC (Yao et al., 2018). 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.

Results. We involve the following models as baselines: (1) **OLS** (Angrist & Imbens, 1995) utilizes least square regression with treatment as covariates, (2) **TARNet** (Shalit et al., 2017) adopts integral orobability metrics for adaptation, (3) **PSM** (Rosenbaum & Rubin, 1983) adopts propensity score for causal effects, (4) **CFR-WASS** (Rosenbaum & Rubin, 1983)

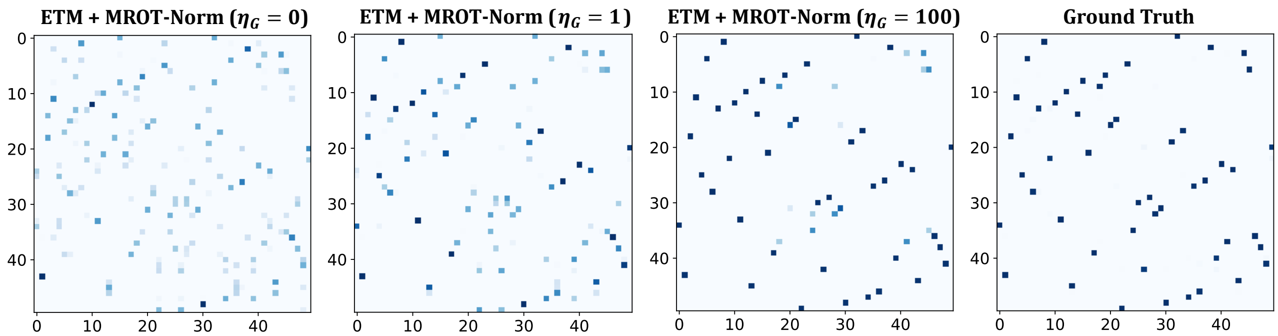


Figure 6. The matching results on ETM + MROT-Norm on SemiUOT with different values of $\eta_G = \{0, 1, 100\}$.

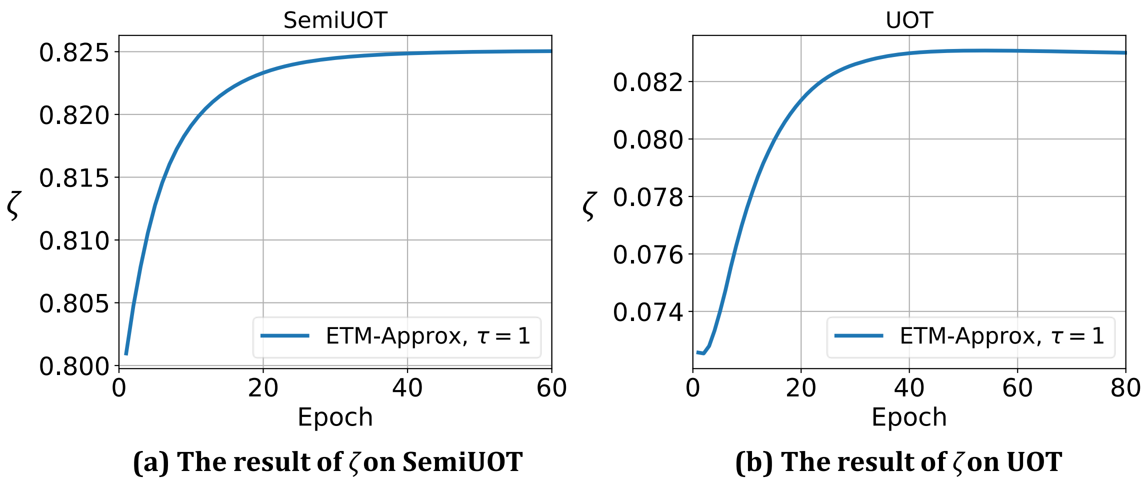


Figure 7. The results of ζ and L_{P} on UOT and SemiUOT.

utilizes standard optimal transport for adaptation, (5) **ESCFR** (Wang et al., 2023) further utilizes unbalanced optimal transport for adaptation. We adopt the same experimental settings as ESCFR (Wang et al., 2023). We utilize Precision in Estimation of Heterogeneous Effect (PEHE) (Shalit et al., 2017) and Area Under the Uplift Curve (AUUC) (Betlei et al., 2021) for the evaluation. Note that ESCFR + UOT(ETM + MROT) only replaces the entropic UOT in ESCFR with our proposed approximate-to-exact ETM +MROT. The experimental results are shown in Table 8. From that we can observe that ESCFR + UOT(ETM + MROT) achieves the best performance, indicating the efficacy of our proposed ETM method.

L. 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.6. 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 comparing with the ground truth result. We can conclude that choosing larger value of η_{G} can fully utilize the knowledge provided by KKT multiplier and enhance the final results. Therefore we set $\eta_{\mathrm{G}} = 100$ empirically.

M. Miscellaneous Discussions

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 analyse the case of SemiUOT. The Fenchel-Lagrange conjugate form of SemiUOT without translation invariant mechanism is given as:

$$\begin{aligned} &\min_{\boldsymbol{f},\boldsymbol{g},\zeta} [\tau \sum_{i=1}^{M} a_i e^{-\frac{f_i}{\tau}} - \sum_{j=1}^{N} b_j g_j] \\ &s.t.\ f_i + g_j \le C_{ij} \end{aligned} \tag{59}$$

We can adopt c-transform on Eq.(59) to obtain the unconstrained optimization problem as:

$$\min_{\boldsymbol{f}} \widetilde{L}_{\mathrm{P}} = \tau \sum_{i=1}^{M} a_i e^{-\frac{f_i}{\tau}} - \sum_{j=1}^{N} \inf_{k\in[M]} [C_{kj} - f_k] b_j, \tag{60}$$

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:

$$\zeta = \tau \log \left(\sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right)\right) - \tau \log \left(\sum_{j=1}^{N} b_j\right) \tag{61}$$

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 imdispenable during the calculation on SemiUOT. Likewise, the Fenchel-Lagrange conjugate form of UOT without translation invariant mechanism is given as:

$$\begin{aligned} &\min_{\boldsymbol{v},\boldsymbol{u}} \left[\tau_a \langle \boldsymbol{a}, e^{-\frac{\boldsymbol{u}}{\tau_a}}\rangle + \tau_b \langle \boldsymbol{b}, e^{-\frac{\boldsymbol{v}}{\tau_b}}\rangle\right] \\ &s.t. \ u_i + v_j \leq C_{ij}. \end{aligned} \tag{62}$$

Here we can adopt c-transform on Eq.(62) to obtain the unconstrained optimization problem as:

$$\min_{\boldsymbol{u}} \widetilde{L}_{\mathrm{U}} = \tau_a \sum_{i=1}^{M} a_i e^{-\frac{u_i}{\tau_a}} + \tau_b \sum_{j=1}^{N} b_j e^{\frac{\sup_{k=1}^{M}\left(u_k - C_{kj}\right)}{\tau_b}} \tag{63}$$

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:

$$\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b} \left[\log \left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle - \log \left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle\right] \tag{64}$$

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 imdispenable during the calculation on UOT. In conclusion, the concept of translation invariant was first proposed in (Séjourné et al., 2022b). However, (Séjourné et al., 2022b) 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.