On Partial Optimal Transport: Revising the Infeasibility of Sinkhorn and Efficient Gradient Methods

Anh Duc Nguyen[1], **Tuan Dung Nguyen**[2], **Quang Minh Nguyen**[3], **Hoang H. Nguyen**[4], **Lam M. Nguyen**[5], **Kim-Chuan Toh**[1, 6]

[1]Department of Mathematics, National University of Singapore, Singapore
[2]Department of Computer and Information Science, University of Pennsylvania, USA
[3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA
[4]School of Industrial and Systems Engineering, Georgia Institute of Technology, USA
[5]IBM Research, Thomas J. Watson Research Center, USA
[6]Institute of Operations Research and Analytics, National University of Singapore, Singapore
anh_duc@u.nus.edu, joshtn@seas.upenn.edu, nmquang@mit.edu,
hnguyen455@gatech.edu, LamNguyen.MLTD@ibm.com, mattohkc@nus.edu.sg

Abstract

This paper studies the Partial Optimal Transport (POT) problem between two unbalanced measures with at most n supports and its applications in various AI tasks such as color transfer or domain adaptation. There is hence a need for fast approximations of POT with increasingly large problem sizes in arising applications. We first theoretically and experimentally investigate the infeasibility of the state-of-the-art Sinkhorn algorithm for POT, which consequently degrades its qualitative performance in real world applications like point-cloud registration. To this end, we propose a novel rounding algorithm for POT, and then provide a feasible Sinkhorn procedure with a revised computation complexity of $\widetilde{\mathcal{O}}(n^2/\varepsilon^4)$. Our rounding algorithm also permits the development of two first-order methods to approximate the POT problem. The first algorithm, Adaptive Primal-Dual Accelerated Gradient Descent (APDAGD), finds an ε-approximate solution to the POT problem in $\widetilde{\mathcal{O}}(n^{2.5}/\varepsilon)$. The second method, Dual Extrapolation, achieves the computation complexity of $\widetilde{\mathcal{O}}(n^2/\varepsilon)$, thereby being the best in the literature. We further demonstrate the flexibility of POT compared to standard OT as well as the practicality of our algorithms on real applications where two marginal distributions are unbalanced.

Introduction

Optimal Transport (OT) (Villani 2008; Kantorovich 1942), which seeks a minimum-cost coupling between two balanced measures, is a well-studied topic in mathematics and operations research. With the introduction of entropic regularization (Cuturi 2013), the scalability and speed of OT computation have been significantly improved, facilitating its widespread applications in machine learning such as domain adaptation (Courty et al. 2017), and dictionary learning (Rolet, Cuturi, and Peyré 2016). However, OT has a stringent requirement that the input measures must have equal total masses (Chizat et al. 2015), hindering its practicality in various other machine learning applications, which require an optimal matching between two measures with unbalanced masses, such as averaging of neuroimaging data (Gramfort, Peyré, and Cuturi 2015) and image classification (Pele and Werman 2008; Rubner, Tomasi, and Guibas 2000).

In response to such limitations, Partial Optimal Transport (POT), which explicitly constrains the mass to be transported between two unbalanced measures, was proposed. It has been studied from the perspective of partial differential equations by theorists (Figalli 2010; Caffarelli and McCann 2010). Practically, the relaxation of the marginal constraints, which are strictly imposed by the standard OT, and the control over the total mass transported grant POT immense flexibility compared to OT (Chapel, Alaya, and Gasso 2020) and more robustness to outliers (Le et al. 2021). POT has been deployed in various recent AI applications such as color transfer (Bonneel and Coeurjolly 2019), graph neural networks (Sarlin et al. 2019), graph matching (Liu et al. 2020), partial covering (Kawano, Koide, and Otaki 2021), point set registration (Wang et al. 2022), and robust estimation (Nietert, Cummings, and Goldfeld 2023).

Despite its potential applicability, POT still suffers from the computation bottleneck, whereby the more intricate structural constraints imposed on admissible couplings have hindered the direct adaptation of any efficient OT solver in the literature. Currently, the literature (Chapel, Alaya, and Gasso 2020; Le et al. 2021) relies on reformulating POT into an extended OT problem under additional assumptions on the input masses, which can then be solved via existing OT methods, and finally retrieves an admissible POT coupling from the solution to the extended OT problem. This approach has two fundamental drawbacks. First, in the reformulated OT problem, the maximum entry of the extended cost matrix is increased (Chapel, Alaya, and Gasso 2020, Proposition 1), which will always worsen the computational complexity since most efficient algorithms for standard OT (Dvurechensky, Gasnikov, and Kroshnin 2018; Lin, Ho, and Jordan 2019; Guminov et al. 2021) depend on this maximum entry in their complexities. Second, we discover, more details in the Revisiting Sinkhorn section, that although Sinkhorn for POT proposed by (Le et al. 2021) achieves the best known complexity of $\widetilde{\mathcal{O}}(n^2/\varepsilon^2)$, it in fact always

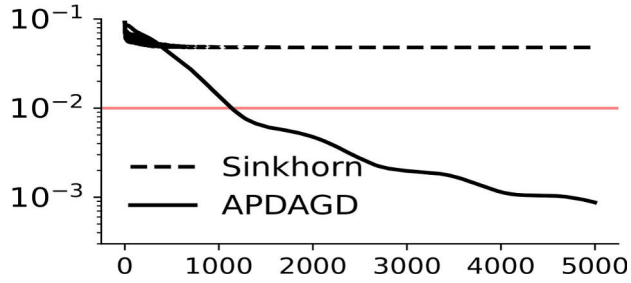

Figure 1: Primal optimality gap against optimization rounds achieved by Sinkhorn (Le et al. 2021) and APDAGD for POT. The marginals are from a color transfer application in our Numerical Experiments section, and the red horizontal line depicts the pre-defined tolerance (ε) for both algorithms.

outputs a *strictly infeasible* solution to the POT problem. In brief, by discarding the last row and column of the reformulated OT solution obtained by Sinkhorn, the POT solution, transportation matrix $\mathbf{X}$, will violate the equality constraint $\mathbf{1}^\top \mathbf{X} \mathbf{1} = s$, which controls the total transported mass. Violating this equality constraint can degrade the results of practical applications in robust regimes such as point cloud registration (Qin et al. 2022) and mini-batch OT (Nguyen et al. 2022) (refer to Remark 4 and our Point Cloud Registration experiment). We theoretically justify the ungroundedness of Sinkhorn for POT in the Revisiting Sinkhorn section and empirically verify this claim in several applications in the Numerical Experiment section. Here, in Figure 1, we specifically investigate Sinkhorn infeasibility in a color transfer example (detailed experimental setup in the later Numerical Experiment section). We show that the optimality gap produced by Sinkhorn is unable to reduce lower than ε, the tolerance of the problem (red line). In other words, Sinkhorn fails to produce an adequate ε-approximate POT solution.

To the best of our knowledge, the invalidity of Sinkhorn means there is currently no efficient method for solving POT in the literature. We attribute this to the fact that while the equivalence between POT and extended OT holds at optimality, all efficient OT solvers instead only output an approximation of the optimum value before projecting it back to the feasible set. However, the well-known rounding algorithm by (Altschuler, Weed, and Rigollet 2017), which is specifically designed for OT, does not guarantee to respect the more intricate structural constraints of POT, resulting in the invalidity of Sinkhorn. Motivated by this challenge and the success of optimization literature for OT, we raise the following central question of this paper:

Can we design a rounding algorithm for POT and then utilize it to develop efficient algorithms for POT that even match the best-known complexities of those for OT?

We affirmatively answer this question and formally summarize our contributions as follows.

- We theoretically and experimentally show the infeasibility of the state-of-the-art Sinkhorn algorithm for POT due to its incompatible rounding algorithm. We propose a novel POT rounding procedure ROUND-POT (Rounding Algorithm Section), which projects an approximate solution onto the feasible POT set in $\mathcal{O}(n^2)$ time.
- From our theoretical bounds of the Sinkhorn constraint violations and the newly introduced ROUND-POT, we provide a revised procedure for Sinkhorn which will return a feasible POT solution. We also establish the revised complexity of Sinkhorn for POT (Table 1).
- Predicated on our novel dual formulation for entropic regularized POT objective, our proposed Adaptive Primal-Dual Accelerated Gradient Descent (APDAGD) algorithm for POT finds an ε-approximate solution in $\widetilde{\mathcal{O}}(n^{2.5}/\varepsilon)$, which is better in ε than the revised Sinkhorn. Various experiments on synthetic and real datasets and with applications such as point cloud registration, color transfer, and domain adaptation illustrate not only our algorithms' favorable performance against the pre-revised Sinkhorn but also the versatility of POT compared to OT.
- Motivated by our novel rounding algorithm, we further reformulate the POT problem with ℓ_1 penalization as a minimax problem and propose Dual Extrapolation (DE) framework for POT. We prove that DE algorithm can theoretically achieve $\widetilde{\mathcal{O}}(n^2/\varepsilon)$ computational complexity, thereby being the best in the POT literature to the best of our knowledge (Table 1).

Preliminaries

Notation

The set of non-negative real numbers is $\mathbb{R}_+$. We use bold capital font for matrices (e.g., $\mathbf{A}$) and bold lowercase font for vectors (e.g., $\mathbf{x}$). For an $m \times n$ matrix $\mathbf{X}$, vec($\mathbf{X}$) denotes the (mn)-dimensional vector obtained by concatenating the rows of $\mathbf{X}$ and transposing the result. Entrywise multiplication and division for matrices and vectors are respectively denoted by $\odot$ and $\oslash$. For $1 \leq p \leq \infty$, let $\|\cdot\|_p$ be the ℓ_p-norm of matrix or vector. For matrices, $\|\cdot\|_{p\to q}$ is the operator norm: $\|\mathbf{A}\|_{p\to q} = \sup_{\|\mathbf{x}\|_p=1} \|\mathbf{A}\mathbf{x}\|_q$. Three specific cases are considered in this paper: for $q \in \{1, 2, \infty\}$, $\|\mathbf{A}\|_{1\to q}$ is the largest ℓ_q norm of any column of $\mathbf{A}$. We use $\|\mathbf{A}\|_{\max}$ and $\|\mathbf{A}\|_{\min}$ to denote the maximum and minimum entries in absolute value of a matrix $\mathbf{A}$, respectively. The n-vectors of zeros and of ones are respectively denoted by $\mathbf{0}_n$ and $\mathbf{1}_n$. The $(n-1)$-dimensional probability simplex is $\Delta_n = \{\mathbf{v} \in \mathbb{R}^n_+ : \mathbf{v}^\top \mathbf{1}_n = 1\}$.

Partial Optimal Transport

Consider two discrete distributions $\mathbf{r}, \mathbf{c} \in \mathbb{R}^n_+$ with possibly different masses. POT seeks a transport plan $\mathbf{X} \in \mathbb{R}^{n\times n}_+$ which maps $\mathbf{r}$ to $\mathbf{c}$ at the lowest cost. Since the masses at two marginals may differ, only a total mass s such that $0 \leq s \leq \min\{\|\mathbf{r}\|_1, \|\mathbf{c}\|_1\}$ is allowed to be transported (Chapel, Alaya, and Gasso 2020; Le et al. 2021). Formally, the POT problem is written as

$$\mathbf{POT}(\mathbf{r}, \mathbf{c}, s) = \min \langle \mathbf{C}, \mathbf{X} \rangle \;\text{ s.t. }\; \mathbf{X} \in \mathcal{U}(\mathbf{r}, \mathbf{c}, s), \quad (1)$$

where $\mathcal{U}(\mathbf{r}, \mathbf{c}, s)$ is defined as

$$\left\{\mathbf{X} \in \mathbb{R}^{n\times n}_+ : \mathbf{X}\mathbf{1}_n \leq \mathbf{r}, \mathbf{X}^\top \mathbf{1}_n \leq \mathbf{c}, \mathbf{1}_n^\top \mathbf{X} \mathbf{1}_n = s\right\},$$

Algorithm	Regularizer	Cost per iteration	Iteration complexity
Iterative Bregman Projections (Benamou et al. 2015)	Entropic	Unspecified	Unspecified
(Infeasible) Sinkhorn (Le et al. 2021)	Entropic	$\mathcal{O}(n^2)$	$\widetilde{\mathcal{O}}(1/\varepsilon^2)$
(Feasible) Sinkhorn (**This paper**)	Entropic	$\mathcal{O}\left(n^2\right)$	$\widetilde{\mathcal{O}}\left(1/\varepsilon^4\right)$
APDAGD (**This paper**)	Entropic	$\mathcal{O}(n^2)$	$\widetilde{\mathcal{O}}\left(\sqrt{n}/\varepsilon\right)$
Dual Extrapolation (**This paper**)	Area-convex	$\mathcal{O}(n^2)$	$\widetilde{\mathcal{O}}\left(1/\varepsilon\right)$

Table 1: Type of regularizers and orders of complexity for four algorithms for POT approximation.

i.e. the feasible set for the transport map $\mathbf{X}$ is and $\mathbf{C} \in \mathbb{R}_+^{n\times n}$ is a cost matrix. The goal of this paper is to derive efficient algorithms to find an ε-approximate solution to $\mathbf{POT}(\mathbf{r}, \mathbf{c}, s)$, pursuant to the following definition.

Definition 1 (ε-approximation)**.** For $\varepsilon \geq 0$, the matrix $\mathbf{X} \in \mathbb{R}_+^{n\times n}$ is an ε-approximate solution to $\mathbf{POT}(\mathbf{r}, \mathbf{c}, s)$ if $\mathbf{X} \in \mathcal{U}(\mathbf{r}, \mathbf{c}, s)$ and

$$\langle \mathbf{C}, \mathbf{X}\rangle \leq \min \langle \mathbf{C}, \mathbf{X}'\rangle + \varepsilon \;\text{ s.t. }\; \mathbf{X}' \in \mathcal{U}(\mathbf{r}, \mathbf{c}, s).$$

To aid the algorithmic design in the following sections, we introduce two new slack variables $\mathbf{p}, \mathbf{q} \in \mathbb{R}_+^n$ and equivalently express problem (1) as

$$\min_{\mathbf{X}\geq 0, \mathbf{p}\geq 0, \mathbf{q}\geq 0} \langle \mathbf{C}, \mathbf{X}\rangle \tag{2}$$

$$\text{s.t. } \mathbf{X}\mathbf{1}_n + \mathbf{p} = \mathbf{r}, \mathbf{X}^\top \mathbf{1}_n + \mathbf{q} = \mathbf{c}, \mathbf{1}_n^\top \mathbf{X}\mathbf{1}_n = s. \tag{3}$$

We also study a equivalent formulation of this problem:

$$\min_{\mathbf{x}\geq \mathbf{0}} \langle \mathbf{d}, \mathbf{x}\rangle \;\text{ s.t. }\; \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{4}$$

where we perform vectorization with $\mathbf{d}^\top = (\text{vec}(\mathbf{C})^\top, \mathbf{0}_{2n}^\top)$ and $\mathbf{x}^\top = (\text{vec}(\mathbf{X})^\top, \mathbf{p}^\top, \mathbf{q}^\top)$. The constraints in Equation 2 are encoded in $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{(2n+1)\times(n^2+2n)}$ and $\mathbf{b} \in H := \mathbb{R}^{2n+1}$ such that $(\mathbf{A}\mathbf{x})^\top = ((\mathbf{X}\mathbf{1}+\mathbf{p})^\top, (\mathbf{X}^\top\mathbf{1}+\mathbf{q})^\top, \mathbf{1}^\top\mathbf{X}\mathbf{1})$ and $\mathbf{b}^\top = (\mathbf{r}^\top, \mathbf{c}^\top, s)$. In other words, the linear operator $\mathbf{A}$ has the form

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}' & \mathbf{I}_{2n} \\ \mathbf{1}_{n^2}^\top & \mathbf{0}_{2n} \end{pmatrix},$$

where $\mathbf{A}'$ is the edge-incidence matrix of the underlying bipartite graph in OT problems (Dvurechensky, Gasnikov, and Kroshnin 2018; Jambulapati, Sidford, and Tian 2019).

Revisiting Sinkhorn for POT

We can reformulate Problem (1) by adding *dummy points* and extending the cost matrix as

$$\widetilde{\mathbf{C}} = \begin{pmatrix} \mathbf{C} & \mathbf{0}_n \\ \mathbf{0}_n^\top & A \end{pmatrix} \in \mathbb{R}_+^{(n+1)\times(n+1)},$$

where $A > \max(C_{i,j})$ (Chapel, Alaya, and Gasso 2020). Then the two marginals are augmented to $(n+1)$-dimensional vectors as $\widetilde{\mathbf{r}}^\top = (\mathbf{r}^\top, \|\mathbf{c}\|_1 - s)$ and $\widetilde{\mathbf{c}}^\top = (\mathbf{c}^\top, \|\mathbf{r}\|_1 - s)$. (Chapel, Alaya, and Gasso 2020, Proposition 1) show that one can obtain the solution POT by solving this extended OT problem with balanced marginals $\widetilde{\mathbf{r}}, \widetilde{\mathbf{c}}$ and cost matrix $\widetilde{\mathbf{C}}$. In particular, if the OT problem admits an optimal solution of the form

$$\widetilde{\bar{\mathbf{X}}} = \begin{pmatrix} \bar{\mathbf{X}} & \widetilde{\mathbf{p}} \\ \widetilde{\mathbf{q}}^\top & \widetilde{X}_{n+1,n+1} \end{pmatrix} \in \mathbb{R}_+^{(n+1)\times(n+1)},$$

then $\bar{\mathbf{X}} \in \mathbb{R}_+^{n\times n}$ is the solution to the original POT. (Le et al. 2021) seeks an approximate solution to the extended OT problem using the Sinkhorn algorithm (see Algorithm 5 in the Appendix). Then the rounding procedure by (Altschuler, Weed, and Rigollet 2017) is applied to the solution to give a primal feasible matrix. While the two POT inequality constraints are satisfied, we discover in the following Theorem that the equality constraint $\mathbf{1}^\top\bar{\mathbf{X}}\mathbf{1} = s$ is violated. The proof is in Appendix, Revisiting Sinkhorn for POT section.

Theorem 2. *For a POT solution $\bar{\mathbf{X}}$ from (Le et al. 2021), the constraint violation $V := \mathbf{1}^\top\bar{\mathbf{X}}\mathbf{1} - s$ can be bounded as*

$$\tilde{\mathcal{O}}\left(\frac{\|\mathbf{C}\|_{max}^2}{A}\right) \geq V \geq \exp\left(\frac{-12A\log n}{\varepsilon} - \mathcal{O}(\log n)\right).$$

Feasible Sinkhorn Procedure: With these bounds, we deduce that in order for Sinkhorn to be feasible, one needs to **both** utilize our ROUND-POT and choose a sufficiently large A (Theorem 3) as opposed to the common practice of picking A a bit larger than 1 (Le et al. 2021). We derive the revised complexity of Sinkhorn for POT as follows.

Theorem 3 (Revised Complexity for Feasible Sinkhorn with ROUND-POT)**.** *We first derive the sufficient size of A to be $\mathcal{O}\left(\|\mathbf{C}\|_{max}/\varepsilon\right)$. With this large A and* ROUND-POT*, Sinkhorn for POT has a computational complexity of $\tilde{\mathcal{O}}\left(n^2\|\mathbf{C}\|_{max}^2/\varepsilon^4\right)$ as opposed to $\tilde{\mathcal{O}}\left(n^2\|\mathbf{C}\|_{max}^2/\varepsilon^2\right)$ (Le et al. 2021).*

The detailed proof for this theorem is included in Appendix, Revisiting Sinkhorn for POT section. We also empirically verify this worsened complexity in section in Feasible Sinkhorn section in Appendix.

Remark 4. Respecting the equality constraint is crucial for various applications that demand strict adherence to feasible solutions like point cloud registration (Qin et al. 2022) (for avoiding incorrect many-to-many correspondences) and mini-batch OT (Nguyen et al. 2022) (for minimizing misspecification). Hence, it is imperative for POT to transport the exact fraction of mass to achieve an optimal mapping, which is vital for the effective performances of ML models.

Algorithm 1: ROUND-POT

Input: $\mathbf{x} = (\text{vec}(\mathbf{X})^\top, \mathbf{p}^\top, \mathbf{q}^\top)^\top$; marginals $\mathbf{r}$, $\mathbf{c}$; mass s.

1: $\bar{\mathbf{p}} = \texttt{EP}(\mathbf{r}, s, \mathbf{p})$
2: $\bar{\mathbf{q}} = \texttt{EP}(\mathbf{c}, s, \mathbf{q})$
3: $\mathbf{g} = \min\{\mathbf{1}, (\mathbf{r} - \bar{\mathbf{p}}) \oslash \mathbf{X}\mathbf{1}\}$
4: $\mathbf{h} = \min\{\mathbf{1}, (\mathbf{c} - \bar{\mathbf{q}}) \oslash \mathbf{X}^\top\mathbf{1}\}$
5: $\mathbf{X}' = \text{diag}(\mathbf{g})\mathbf{X}\text{diag}(\mathbf{h})$
6: $\mathbf{e}_1 = (\mathbf{r} - \bar{\mathbf{p}}) - \mathbf{X}'\mathbf{1}, \mathbf{e}_2 = (\mathbf{c} - \bar{\mathbf{q}}) - \mathbf{X}'^\top\mathbf{1}$
7: $\bar{\mathbf{X}} = \mathbf{X}' + \mathbf{e}_1\mathbf{e}_2^\top / \|\mathbf{e}_1\|_1$

Output: $\bar{\mathbf{x}} = (\bar{\mathbf{X}}, \bar{\mathbf{p}}, \bar{\mathbf{q}})$

Rounding Algorithm

All efficient algorithms for standard OT (Dvurechensky, Gasnikov, and Kroshnin 2018; Lin, Ho, and Jordan 2019; Guminov et al. 2021) only output an infeasible approximation of the optimum value, and leverage the well-known rounding algorithm (Altschuler, Weed, and Rigollet 2017, Algorithm 2) to project it back to the set of admissible couplings. Nevertheless, its ad-hoc design tailored to the OT's marginal constraints makes generalization to the case of POT with more intricate structural constraints non-trivial. In fact, we attribute the rather limited literature on efficient POT solvers to such lack of a rounding algorithm for POT. Specifically, previous works rely on imposing additional assumptions on the input masses to permit reformulation of POT into standard OT with an additional computational burden (Chapel, Alaya, and Gasso 2020; Le et al. 2021). Deviating from the vast literature, we address this fundamental challenge by proposing a novel rounding procedure for POT, termed ROUND-POT (Algorithm 1), to efficiently round any approximate solution to a feasible solution of (2). Given an approximate solution $\mathbf{x} = (\mathbf{X}, \mathbf{p}, \mathbf{q}) \geq 0$ violating the POT constraints of (2) by a predefined error, $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1 \leq \delta$ for some δ, ROUND-POT returns $\bar{\mathbf{x}} = (\bar{\mathbf{X}}, \bar{\mathbf{p}}, \bar{\mathbf{q}}) \geq \mathbf{0}$ strictly in the feasible set, i.e., $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$, and close to $\mathbf{x}$ in ℓ_1 distance.

The Enforcing Procedure (EP) (Algorithm 2) is a novel subroutine to ensure $\mathbf{0} \leq \bar{\mathbf{p}} \leq \mathbf{r}$ and $\|\bar{\mathbf{p}}\|_1 = \|\mathbf{r}\|_1 - s$ (Lemma 5). Equivalently, a similar procedure is applied to $(\mathbf{c}, s, \mathbf{q})$ in step 2 of Algorithm 1 with similar guarantees for $\bar{\mathbf{q}}$. Step 1 transforms $\mathbf{p}$ (or $\mathbf{q}$) to $\mathbf{p}'$ (or $\mathbf{q}'$) so that $\mathbf{0} \leq \mathbf{p}' \leq \mathbf{r}$ (or $\mathbf{0} \leq \mathbf{q}' \leq \mathbf{c}$). The transformation in steps 2 and 3 ensures that $\|\mathbf{p}''\|_1 \leq \|\mathbf{r}\|_1 - s$ (or $\|\mathbf{q}''\|_1 \leq \|\mathbf{c}\|_1 - s$). The rest of the EP steps ensure the other guarantee $\|\bar{\mathbf{p}}\|_1 = \|\mathbf{r}\|_1 - s$. The proof is in Rounding Algorithm section in the Appendix.

Lemma 5 (Guarantees for EP)**.** *We obtain in $\mathcal{O}(n)$ time $\mathbf{0} \leq \bar{\mathbf{p}} \leq \mathbf{r}$ and $\|\bar{\mathbf{p}}\|_1 = \|\mathbf{r}\|_1 - s$.*

For ROUND-POT, steps 3 through 7 check whether the solutions $\mathbf{X}$ violate each of the two equality constraints $\mathbf{X}\mathbf{1} = \mathbf{r} - \bar{\mathbf{p}}$ and $\mathbf{X}^\top\mathbf{1} = \mathbf{c} - \bar{\mathbf{q}}$; if so, the algorithm projects $\mathbf{X}$ into the feasible set. It is noteworthy that these two constraints directly implies the last needed constraint $\mathbf{1}^\top\mathbf{X}\mathbf{1} = s$. Finally, ROUND-POT returns an output that satisfies the required constraints in Equation (2). The following Theorem 6 characterizes the error guarantee of the rounded output $\bar{\mathbf{x}}$. Its detailed proof can be found in Rounding Algorithm section in the Appendix.

Algorithm 2: Enforcing Procedure EP

Input: marginal $\mathbf{r}$ (or $\mathbf{c}$); mass s; slack variable $\mathbf{p}$ (or $\mathbf{q}$).

1: $\mathbf{p}' = \min\{\mathbf{p}, \mathbf{r}\}$
2: **if** $\|\mathbf{p}'\|_1 = 0$ or $\|\mathbf{r}\|_1 = s$ **then**
3: $\quad\alpha = 1$
4: $\quad\mathbf{p}'' = \mathbf{p}'$
5: **else**
6: $\quad\alpha = \min\{1, (\|\mathbf{r}\|_1 - s)/\|\mathbf{p}'\|_1\}$
7: $\quad\mathbf{p}'' = \alpha\mathbf{p}'$
8: **end if**
9: **if** $\|\mathbf{p}'\|_1 > (\|\mathbf{r}\|_1 - s)$ **then**
10: $\quad\bar{\mathbf{p}} = \mathbf{p}''$
11: **else**
12: $\quad i = 0$
13: $\quad$**while** $\|\mathbf{p}''\|_1 \leq \|\mathbf{r}\|_1 - s$ **do**
14: $\quad\quad i = i + 1$
15: $\quad\quad p_i'' = r_i$
16: $\quad$**end while**
17: $\quad p_i'' = p_i'' - (\|\mathbf{p}''\|_1 - \|\mathbf{r}\|_1 + s)$
18: $\quad\bar{\mathbf{p}} = \mathbf{p}''$
19: **end if**

Output: $\bar{\mathbf{p}}$.

Theorem 6 (Guarantees for ROUND-POT)**.** *Let $\mathbf{A}$, $\mathbf{x}$ (consisting of $\mathbf{X}$, $\mathbf{p}$ and $\mathbf{q}$) and $\mathbf{b}$ be defined as in the preliminaries. If $\mathbf{x}$ satisfies that $\mathbf{x} \geq 0$ and $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1 \leq \delta$ for some $\delta \geq 0$, Algorithm 1 outputs $\bar{\mathbf{x}} \geq \mathbf{0}$ (consisting of $\bar{\mathbf{X}}, \bar{\mathbf{p}}$ and $\bar{\mathbf{q}}$) in $\mathcal{O}(n^2)$ time such that $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$ and $\|\mathbf{x} - \bar{\mathbf{x}}\|_1 \leq 23\delta$.*

Adaptive Primal-Dual Accelerated Gradient Descent (APDAGD)

Dual Formulation and Algorithmic Design

Following a similar formulation to (Dvurechensky, Gasnikov, and Kroshnin 2018, Section 3.1), we have the following primal problem with entropic regularization

$$\min_{\mathbf{x} \geq \mathbf{0}} \{f(\mathbf{x}) := \langle \mathbf{d}, \mathbf{x} \rangle + \gamma \langle \mathbf{x}, \log \mathbf{x} \rangle\} \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \quad (5)$$

where $\mathbf{A}\mathbf{x} = \mathbf{b}$ is encoded as explained in Equation (4). Since problem (5) is a linearly constrained convex optimization problem, strong duality holds.

Lemma 7. *With a dual variable $\boldsymbol{\lambda} \in H^* = \mathbb{R}^{2n+1}$, the dual of (5) is given by*

$$\min_{\boldsymbol{\lambda} \in H^*} \left\{ \varphi(\boldsymbol{\lambda}) := \langle \boldsymbol{\lambda}, \mathbf{b} \rangle + \max_{\mathbf{x} \in Q} \left\{ -f(\mathbf{x}) - \langle \mathbf{x}, \mathbf{A}^\top \boldsymbol{\lambda} \rangle \right\} \right\},$$

or equivalently

$$\min_{\mathbf{y}, \mathbf{z}, t} \{-ts - \langle \mathbf{y}, \mathbf{r} \rangle - \langle \mathbf{z}, \mathbf{c} \rangle - \gamma \sum_{i,j=1}^{n} e^{-(C_{i,j} + y_i + z_j + t)/\gamma - 1} + e^{-y_i/\gamma - 1} + e^{-z_j/\gamma - 1} \Bigg\}, \quad (6)$$

where $\mathbf{y}, \mathbf{z}, t$ are dual variables corresponding the POT constraints in (2) as $\boldsymbol{\lambda} = (\mathbf{y}^\top, \mathbf{z}^\top, t)^\top$ (which we simply refer as $(\mathbf{y}, \mathbf{z}, t)$ from now on).

Algorithm 3: Approximating POT by APDAGD

Input: marginals $\mathbf{r}, \mathbf{c}$; cost matrix $\mathbf{C}$.
$\gamma = \varepsilon/(4\log(n)), \widetilde{\varepsilon} = \varepsilon/(8\|\mathbf{C}\|_{\max})$
if $\|\mathbf{r}\|_1 > 1$ **then**
  $\widetilde{\varepsilon} = \min\{\widetilde{\varepsilon}, 8(\|\mathbf{r}\|_1 - s)/(\|\mathbf{r}\|_1 - 1)\}$
end if
if $\|\mathbf{c}\|_1 > 1$ **then**
  $\widetilde{\varepsilon} = \min\{\widetilde{\varepsilon}, 8(\|\mathbf{c}\|_1 - s)/(\|\mathbf{c}\|_1 - 1)\}$
end if
$\widetilde{\mathbf{r}} = (1 - \widetilde{\varepsilon}/8)\,\mathbf{r} + \widetilde{\varepsilon}\mathbf{1}_n/(8n)$
$\widetilde{\mathbf{c}} = (1 - \widetilde{\varepsilon}/8)\,\mathbf{r} + \widetilde{\varepsilon}\mathbf{1}_n/(8n)$
$\widetilde{\mathbf{X}} = \text{APDAGD}(\mathbf{C}, \gamma, \widetilde{\mathbf{r}}, \widetilde{\mathbf{c}}, \widetilde{\varepsilon}/2)$
$\bar{\mathbf{X}} = \text{ROUND-POT}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{r}}, \widetilde{\mathbf{c}}, s)$
Output: $\bar{\mathbf{X}}$.

More details on the dual formulation and properties (strong convexity, smoothness, etc) of the primal and dual objectives are in Appendix. The APDAGD procedure is described in Algorithm 8 in Appendix. So as to approximate POT, we incorporate our novel rounding algorithm with a similar procedure to (Lin, Ho, and Jordan 2019, Algorithm 2), in Algorithm 3.

Computational Complexity

Now, we provide the computational complexity of APDAGD (Theorem 8) and its proof sketch. The detailed proof of this result will be presented in Complexity of APDAGD for POT Detailed Proof subsection in Appendix.

Theorem 8 (Complexity of APDAGD)**.** *The APDAGD algorithm returns ε-approximation POT solution $\widehat{\mathbf{X}} \in \mathcal{U}(\mathbf{r}, \mathbf{c}, s)$ in $\widetilde{\mathcal{O}}\left(n^{5/2}\|\mathbf{C}\|_{\max}/\varepsilon\right)$.*

Proof sketch. **Step 1**: we present the reparameterization $\mathbf{u} = -\mathbf{y}/\gamma - \mathbf{1}, \mathbf{v} = -\mathbf{z}/\gamma - \mathbf{1}$ and $w = -t/\gamma + 1$ for the dual (6), leading to the equivalent dual form

$$\min_{\mathbf{u},\mathbf{v},w} \sum_{i,j=1}^{n} \exp\left(-C_{i,j}/\gamma + u_i + v_j + w\right) + \sum_{i=1}^{n} \exp(u_i) + \sum_{j=1}^{n} \exp(v_j) - \langle \mathbf{u}, \mathbf{r}\rangle - \langle \mathbf{v}, \mathbf{c}\rangle - ws.$$

This transformation will facilitate the bounding of the dual variables in later steps.

Step 2: we proceed to bound the ℓ_∞-norm of the transformed optimal dual variables $\|(\mathbf{u}^*, \mathbf{v}^*, w^*)\|_\infty$. Conventional analyses for OT such as (Lin, Ho, and Jordan 2019) are inapplicable to the case of POT due to the addition of the third dual variable w and more intricate dependencies of the dual variables $\mathbf{u}, \mathbf{v}$. To this end, our novel proof technique establishes the tight bound of $\|(\mathbf{u}^*, \mathbf{v}^*, w^*)\|_\infty = \widetilde{\mathcal{O}}\left(\|\mathbf{C}\|_{\max}\right)$, which consequently translates to the final bound for original dual variables $\|(\mathbf{y}^*, \mathbf{z}^*, t^*)\|_2 = \widetilde{\mathcal{O}}\left(\sqrt{n}\,\|\mathbf{C}\|_{\max}\right)$ in Lemma 18. Bounding the ℓ_2-norm (i.e. bounding R) is crucial because it contributes to the APDAGD guarantees (Theorem 16) and the final complexity.

Step 3: Combining the $\bar{R}$ bound from **Step 2** in view of (Lin, Ho, and Jordan 2019, Proposition 4.10) and the guarantees of ROUND-POT (Theorem 6), we conclude the final computational complexity of APDAGD of $\widetilde{\mathcal{O}}(n^{2.5}/\varepsilon)$. □

Dual Extrapolation (DE)

Our novel POT rounding algorithm permits the development of Dual Extrapolation (DE) for POT. From our analysis, DE is a first-order and parallelizable algorithm that can approximate POT distance up to ε accuracy with $\widetilde{\mathcal{O}}(1/\varepsilon)$ parallel depth and $\widetilde{\mathcal{O}}(n^2/\varepsilon)$ total work.

Setup

For each feasible $\mathbf{x}$, we have $\|\mathbf{x}\|_1 = \|\mathbf{X}\|_1 + \|\mathbf{p}\|_1 + \|\mathbf{q}\|_1 = \|\mathbf{r}\|_1 + \|\mathbf{c}\|_1 - s$. We can normalize $\mathbf{x} = \mathbf{x}/(\|\mathbf{r}\|_1 + \|\mathbf{c}\|_1 - s)$, $\mathbf{b} = \mathbf{b}/(\|\mathbf{r}\|_1 + \|\mathbf{c}\|_1 - s)$. These imply $\mathbf{x} \in \Delta_{n^2+2n}$. The POT problem formulation (4) is now updated as

$$\min_{\mathbf{x}\in\Delta_{n^2+2n}} \langle \mathbf{d}, \mathbf{x}\rangle \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{7}$$

We then consider the ℓ_1 penalization for the problem (7) and show that it has equal optimal value and ε-approximate minimizer to those of the POT formulation (7) (more details in ℓ_1 Penalization subsection in Appendix). Through a primal-dual point of view, the ℓ_1 penalized objective (26) can be rewritten as

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} F(\mathbf{x}, \mathbf{y}) := \mathbf{d}^\top\mathbf{x} + 23\,\|\mathbf{d}\|_\infty \left(\mathbf{y}^\top\mathbf{A}\mathbf{x} - \mathbf{y}^\top\mathbf{b}\right), \tag{8}$$

with $\mathcal{X} = \Delta_{n^2+2n}, \mathcal{Y} = [-1, 1]^{2n+1}$. Note that the term 23 comes from the guarantees of ROUND-POT (Theorem 6, Lemma 20). Let $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ such that $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$. For a bilinear objective $F(\mathbf{x}, \mathbf{y})$ that is convex in $\mathbf{x}$ and concave in $\mathbf{y}$, it is natural to define the gradient operator $g(\mathbf{x}, \mathbf{y}) = (\nabla_{\mathbf{x}} F(\mathbf{x}, \mathbf{y}), -\nabla_{\mathbf{y}} F(\mathbf{x}, \mathbf{y}))$. Specifically for the objective (8), we have $g(\mathbf{x}, \mathbf{y}) = (\mathbf{d} + 23\,\|\mathbf{d}\|_\infty \mathbf{A}^\top\mathbf{y}, -23\,\|\mathbf{d}\|_\infty (\mathbf{A}\mathbf{x} - \mathbf{b}))$. This minimax objective can be solved with the dual extrapolation (Nesterov 2007), which requires strongly convex regularizers. This setup can be relaxed with the notion of area-convexity (Sherman 2017, Definition 1.2), in the following definition.

Definition 9 (Area Convexity)**.** A regularizer r is κ-area-convex w.r.t an operator g if for any x_1, x_2, x_3 in its domain,

$$\kappa \sum_{i}^{3} r(x_i) - 3\kappa r\left(\frac{\sum_i^3 x_i}{3}\right) \geq \langle g(x_2) - g(x_1), x_2 - x_3\rangle.$$

The regularizer chosen for this framework is the Sherman regularizer, introduced in (Sherman 2017)

$$r(\mathbf{x}, \mathbf{y}) = 2\,\|\mathbf{d}\|_\infty \left(10\langle \mathbf{x}, \log\mathbf{x}\rangle + \mathbf{x}^\top\mathbf{A}^\top(\mathbf{y}^2)\right), \tag{9}$$

in which $\mathbf{y}^2$ is entry-wise. While this regularizer has a similar form to that in (Jambulapati, Sidford, and Tian 2019), the POT formulation leads to a different structure of $\mathbf{A}$. For instance, $\|\mathbf{A}\|_1 = 3$ instead of 2. This following lemma shows that chosen r is 9-area-convex (its proof is in the Proof of Lemma 10 section in the Appendix).

Lemma 10. *Regularizer r (9) is 9-area-convex with respect to the gradient operator g, i.e., $\kappa = 9$.*

Algorithm 4: Dual Extrapolation for POT

Input: linearized cost $\mathbf{d}$; linear operator $\mathbf{A}$; constraints $\mathbf{b}$; area-convexity coefficient κ; initial states $\mathbf{s}_{\mathbf{x}}^0 = \mathbf{0}_{n^2+2n}, \mathbf{s}_{\mathbf{y}}^0 = \mathbf{0}_{2n+1}$; iterations M (Theorem 11)

1: $\nabla_{\mathbf{x}} r(\bar{\mathbf{z}}) = 20 \|\mathbf{d}\|_\infty (1 - \log(n^2 + 2n)) \mathbf{1}_{n^2+2n}$
2: $\nabla_{\mathbf{y}} r(\bar{\mathbf{z}}) = \mathbf{0}_{2n+1}$
3: **for** $t = 0, 1, 2, \ldots, T-1$ **do**
4: $\quad \mathbf{v} = \mathbf{s}_{\mathbf{x}}^t - \nabla_{\mathbf{x}} r(\bar{\mathbf{z}})$
5: $\quad \mathbf{u} = \mathbf{s}_{\mathbf{y}}^t - \nabla_{\mathbf{y}} r(\bar{\mathbf{z}})$
6: $\quad (\mathbf{z}_{\mathbf{x}}^t, \mathbf{z}_{\mathbf{y}}^t) = \texttt{AM}(M, \mathbf{v}, \mathbf{u})$
7: $\quad \mathbf{v} = \mathbf{v} + (\mathbf{d} + 23 \|\mathbf{d}\|_\infty \mathbf{A}^\top \mathbf{z}_{\mathbf{y}}^t)/\kappa$
8: $\quad \mathbf{u} = \mathbf{u} - 23 \|\mathbf{d}\|_\infty (\boldsymbol{A}\mathbf{z}_{\mathbf{x}}^t - \mathbf{b})/\kappa$
9: $\quad (\mathbf{w}_{\mathbf{x}}^t, \mathbf{w}_{\mathbf{y}}^t) = \texttt{AM}(M, \mathbf{v}, \mathbf{u})$
10: $\quad \mathbf{s}_{\mathbf{x}}^{t+1} = \mathbf{s}_{\mathbf{x}}^t + (\mathbf{d} + 23 \|\mathbf{d}\|_\infty \mathbf{A}^\top \mathbf{w}_{\mathbf{y}}^t)/(2\kappa)$
11: $\quad \mathbf{s}_{\mathbf{y}}^{t+1} = \mathbf{s}_{\mathbf{y}}^t - 23 \|\mathbf{d}\|_\infty (\boldsymbol{A}\mathbf{w}_{\mathbf{x}}^t - \mathbf{b})/(2\kappa)$
12: **end for**

Output: $\bar{\mathbf{w}}_{\mathbf{x}} = \sum_{t=0}^{T-1} \mathbf{w}_{\mathbf{x}}^t / T$, $\bar{\mathbf{w}}_{\mathbf{y}} = \sum_{t=0}^{T-1} \mathbf{w}_{\mathbf{y}}^t / T$.

Algorithmic Development

The main motivation is the DE Algorithm 7 in Appendix, proposed by (Nesterov 2007). This general DE framework essentially has two proximal steps per iteration, while maintaining a state $\mathbf{s}$ in the dual space. We follow (Jambulapati, Sidford, and Tian 2019) and update $\mathbf{s}$ with $1/2\kappa$ rather than $1/\kappa$ (Nesterov 2007). The proximal steps (steps 2 and 3 in Algorithm 7) needs to minimize:

$$P(\mathbf{x}, \mathbf{y}) := \langle \mathbf{v}, \mathbf{x} \rangle + \langle \mathbf{u}, \mathbf{y} \rangle + r(\mathbf{x}, \mathbf{y}). \tag{10}$$

This can be solved efficiently with an Alternating Minimization (AM) approach (Jambulapati, Sidford, and Tian 2019). Details for AM are included in the Appendix. Combining both algorithms, we have the DE for POT Algorithm 4, where each proximal step is solved by the AM subroutine.

Computational Complexities

Firstly, we bound the regularizer r to satisfy the convergence guarantees (Jambulapati, Sidford, and Tian 2019) in Proof of Lemma 22 subsection in Appendix. In that same subsection, we also derive the required number of iterations T in DE with respect to Θ, the range of the regularizer. Next, we have this essential lemma that bounds the number of iterations to evaluate a proximal step.

Theorem 11 (Complexity of AM)**.** *For $T = \lceil 36\Theta/\varepsilon \rceil$ iterations of DE, AM Algorithm 8 obtains additive error $\varepsilon/2$ in*

$$M = 24 \log \left(\left(840 \|\mathbf{d}\|_\infty / \varepsilon^2 + 6/\varepsilon \right) \Theta + 1336 \|\mathbf{d}\|_\infty / 9 \right)$$

iterations. This is done in wall-clock time $\mathcal{O}(n^2 \log \eta)$ with $\eta = \log n \|\mathbf{d}\|_\infty / \varepsilon$.

The proof of this theorem can be found in Proof of Theorem 11 subsection in Appendix. The main proof idea is to bound the number of iterations required to solve the proximal steps. This explicit bound for the number of AM iterations is novel as in DE for OT (Jambulapati, Sidford, and Tian 2019), the authors runs a while loop and do not analyze its final number of iterations. We can now calculate the computational complexity of the DE algorithm. The proof is in Proof of Theorem 12 subsection in Appendix.

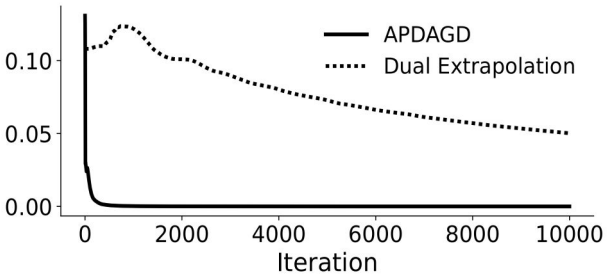

Figure 2: Primal optimality gap for solutions produced by APDAGD and Dual Extrapolation.

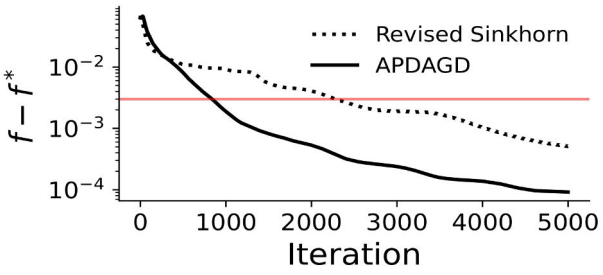

Figure 3: Primal optimality gap against optimization rounds achieved by our revised Sinkhorn and APDAGD for POT. The marginal distributions are taken from the later color transfer application in this section, and the red horizontal line depicts the pre-defined tolerance (ε) for both algorithms.

Theorem 12 (Complexity of DE)**.** *In $\widetilde{\mathcal{O}}(n^2 \|\mathbf{C}\|_{max} / \varepsilon)$ wall-clock time, the DE Algorithm 4 returns $(\bar{\mathbf{w}}_{\mathbf{x}}, \bar{\mathbf{w}}_{\mathbf{y}}) \in \mathcal{Z}$ so that the duality gap (27) is less than ε.*

Numerical Experiments

In this section, we provide numerical results on approximating POT and its applications using the algorithms presented above.[1] In all settings, the optimal solution is found by solving the linear program (1) using the `cvxpy` package. Due to space constraint, we also include many extra experiments such as domain adaptation application, large-scale APDAGD, run time for varying ϵ comparison, and revised Sinkhorn performance in Appendix, further solidifying the efficiency and practicality of the proposed algorithms.

Run Time Comparison

APGAGD vs DE: We are able to implement the DE algorithm for POT while the authors of DE for OT faced numerical overflow and had to use mirror prox instead "for numerical stability considerations" (Jambulapati, Sidford, and Tian 2019). For the setup of Figure 2, we use images in the CIFAR-10 dataset (Krizhevsky and Hinton 2009) as the

[1] Implementation and numerical experiments can be found at https://github.com/joshnguyen99/partialot.

marginals. More details are included in the Further Experiment Setup Details section in Appendix. In Figure 2, despite having a better theoretical complexity, DE has relatively poor practical performance compared to APDAGD. This is not surprising as previous works on this class of algorithms like (Dvinskikh and Tiapkin 2021) reach similar conclusions on DE's practical limitations. This can be partly explained by the large constants that are dismissed by the asymptotic computational complexity. Thus, in our applications in the following subsection, we will use APDAGD.

APGAGD vs Sinkhorn: For the same setting with CIFAR-10 dataset, we report that the ratio between the average per iteration cost of APDAGD and that of Sinkhorn is 0.68. Furthermore, in the Run Time for Varying ε section in Appendix, we reproduce the same result in Figure 6 as **Figure 1 in (Dvurechensky et al., 2018)**, comparing the runtime of APDAGD and Sinkhorn for varying ε.

Revised Sinkhorn

Using the similar setting as the later example of color transfer, we can empirically verify that our revised Sinkhorn can achieve the required tolerance ε of the POT problem in Figure 3 (as opposed to pre-revised Sinkhorn in Figure 1).

Point Cloud Registration

We now present an application of POT in point set registration, a common task in shape analysis. Start with two point clouds in three dimensions $R = \{\mathbf{x}_i \in \mathbb{R}^3 \mid i = 1, \ldots, m\}$ and $Q = \{\mathbf{y}_j \in \mathbb{R}^3 \mid j = 1, \ldots, n\}$. The objective is to find a transformation, consisting of a rotation and a translation, that best aligns the two point clouds. When the initial point clouds contain significant noise or missing data, the registration result is often badly aligned. Here, we consider a scenario where one set has missing values. In Figure 4 (a), the blue point cloud is set to contain the front half of the rabbit, retaining about 45% of the original points. A desirable transformation must align the first halves of the two rabbits correctly. Figure 4 compares the point clouds registration result using these methods. If $\mathbf{T}$ is simply the OT matrix, not subject to a total transported mass constraint, the blue cloud is clearly not well-aligned with the red cloud. This is because the points for which the blue cloud is missing (i.e., the right half of the rabbit) are in the red cloud. If the total mass transported is set to $s = \frac{\min\{m,n\}}{\max\{m,n\}}$, then we end up with a POT matrix $\mathbf{T}$ with all entries summing to s. In Figures 4 (c) and (d), the POT solution leads to a much better result than the OT solution: the left halves of the rabbit are closer together. Importantly, the feasible solution obtained by APDAGD leads to an even better alignment, compared to the pre-revised Sinkhorn, due to its infeasibility.

Color Transfer

For color transfer, a popular application in computer vision, POT offers flexibility in transferring colors between two possibly different-sized images, in contrast to OT which requires two color histograms to be normalized (Bonneel et al. 2015). We follow the setup by (Blondel, Seguy, and Rolet 2018). Implementation details are in the Further Experiment Setup Details section in Appendix. Our results are presented in Figure 5. The third row displays the result produced by APDAGD with different levels of α (or transported mass s). Increasing α makes the wall color closer to the lighter part of the kiwi in the target image but comes at a cost of saturating the window frames' white. We emphasize that α is a tunable parameter, and the user can pick the most suitable transported mass. This can be more flexible than the vanilla OT which requires marginal masses to be normalized. We also emphasize that qualitatively, the solution produced by APDAGD closely matches the exact solution on top right corner, in contrast to pre-revised Sinkhorn's (Le et al. 2021).

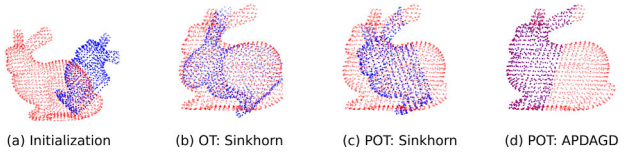

Figure 4: Point cloud registration using POT. (a): Initial point sets with one set (in blue) missing 45% of the points. (b): Registration result obtained after transforming the blue point cloud using OT plan by Sinkhorn. (c): Registration result using POT plan by pre-revised Sinkhorn (Le et al. 2021). (d): Registration result using POT plan by APDAGD.

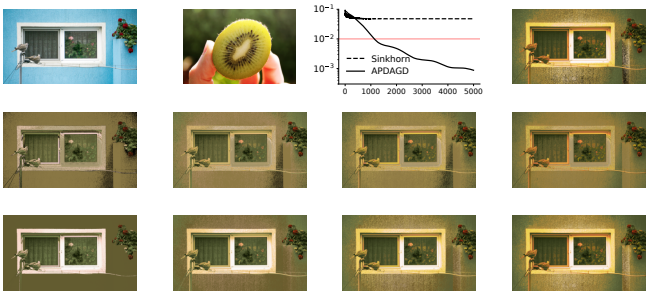

Figure 5: Flexibility in color transfer using POT as opposed to OT. First row (left to right): source and target images (of different sizes), optimality gap at each iteration of Sinkhorn and APDAGD, and image derived from the exact solution by Gurobi when $\alpha = 0.9$. Second row: images derived from the solutions by Sinkhorn at four levels of α: 0.3, 0.5, 0.7 and 0.9. Third row: images derived from the solutions by APDAGD at the same four levels of α.

Conclusion

In this paper, we first examine the infeasibility of Sinkhorn for POT. We then propose a a novel rounding algorithm which facilitates our development of feasible Sinkhorn procedure with guarantees, APDAGD and DE. DE achieves the best theoretical complexity while APDAGD and revised Sinkhorn are practically efficient algorithms, as demonstrated in our extensive experiments. We believe the rigor and applicability of our proposed methods will further facilitate the practical adoptions of POT in AI applications.

Acknowledgements

We would like to thank the reviewers of AAAI 2024 for their detailed and insightful comments and Dr. Darina Dvinskikh for sharing the numerical implementations of their work (Dvinskikh and Tiapkin 2021).

References

Altschuler, J.; Weed, J.; and Rigollet, P. 2017. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Advances in Neural Information Processing Systems*, 1964–1974.

Benamou, J.-D.; Carlier, G.; Cuturi, M.; Nenna, L.; and Peyré, G. 2015. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2): A1111–A1138.

Blondel, M.; Seguy, V.; and Rolet, A. 2018. Smooth and Sparse Optimal Transport. In *AISTATS*, 880–889.

Bonneel, N.; and Coeurjolly, D. 2019. SPOT: Sliced Partial Optimal Transport. *ACM Transactions on Graphics*.

Bonneel, N.; Rabin, J.; Peyré, G.; and Pfister, H. 2015. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1): 22–45.

Caffarelli, L. A.; and McCann, R. J. 2010. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Annals of Mathematics*, 171(2): 673–730.

Chapel, L.; Alaya, M. Z.; and Gasso, G. 2020. Partial Optimal Tranport with applications on Positive-Unlabeled Learning. In *Advances in Neural Information Processing Systems 33*.

Chizat, L.; Peyré, G.; Schmitzer, B.; and Vialard, F.-X. 2015. Unbalanced Optimal Transport: Dynamic and Kantorovich Formulation.

Courty, N.; Flamary, R.; Tuia, D.; and Rakotomamonjy, A. 2017. Optimal Transport for Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9): 1853–1865.

Cuturi, M. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, 2292–2300.

Dvinskikh, D.; and Tiapkin, D. 2021. Improved Complexity Bounds in Wasserstein Barycenter Problem. In Banerjee, A.; and Fukumizu, K., eds., *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, 1738–1746. PMLR.

Dvurechensky, P.; Gasnikov, A.; and Kroshnin, A. 2018. Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn's Algorithm. In *International conference on machine learning*, 1367–1376.

Figalli, A. 2010. The optimal partial transport problem. *Archive for rational mechanics and analysis*, 195(2): 533–560.

Gramfort, A.; Peyré, G.; and Cuturi, M. 2015. Fast Optimal Transport Averaging of Neuroimaging Data. *CoRR*, abs/1503.08596.

Guminov, S.; Dvurechensky, P.; Tupitsa, N.; and Gasnikov, A. 2021. Accelerated Alternating Minimization, Accelerated Sinkhorn's Algorithm and Accelerated Iterative Bregman Projections. arXiv:1906.03622.

Jambulapati, A.; Sidford, A.; and Tian, K. 2019. A Direct $\widetilde{O}(1/\varepsilon)$ Iteration Parallel Algorithm for Optimal Transport. *ArXiv Preprint: 1906.00618*.

Kantorovich, L. V. 1942. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, 199–201.

Kawano, K.; Koide, S.; and Otaki, K. 2021. Partial Wasserstein Covering. *CoRR*, abs/2106.00886.

Krizhevsky, A.; and Hinton, G. 2009. Learning multiple layers of features from tiny images. *Technical Report TR-2009*.

Le, K.; Nguyen, H.; Pham, T.; and Ho, N. 2021. On Multimarginal Partial Optimal Transport: Equivalent Forms and Computational Complexity.

Lin, T.; Ho, N.; and Jordan, M. 2019. On Efficient Optimal Transport: An Analysis of Greedy and Accelerated Mirror Descent Algorithms. In *International Conference on Machine Learning*, 3982–3991.

Liu, W.; Zhang, C.; Xie, J.; Shen, Z.; Qian, H.; and Zheng, N. 2020. Partial Gromov-Wasserstein Learning for Partial Graph Matching. *CoRR*, abs/2012.01252.

Nesterov, Y. 2007. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109(2-3): 319–344.

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

Nietert, S.; Cummings, R.; and Goldfeld, Z. 2023. Robust Estimation under the Wasserstein Distance. *arXiv preprint arXiv:2302.01237*.

Pele, O.; and Werman, M. 2008. A Linear Time Histogram Metric for Improved SIFT Matching. In *European Conference on Computer Vision*.

Qin, H.; Zhang, Y.; Liu, Z.; and Chen, B. 2022. Rigid Registration of Point Clouds Based on Partial Optimal Transport. In *Computer Graphics Forum*, volume 41, 365–378. Wiley Online Library.

Rolet, A.; Cuturi, M.; and Peyré, G. 2016. Fast Dictionary Learning with a Smoothed Wasserstein Loss. In Gretton, A.; and Robert, C. C., eds., *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, 630–638. Cadiz, Spain: PMLR.

Rubner, Y.; Tomasi, C.; and Guibas, L. J. 2000. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2): 99–121.

Sarlin, P.; DeTone, D.; Malisiewicz, T.; and Rabinovich, A. 2019. SuperGlue: Learning Feature Matching with Graph Neural Networks. *CoRR*, abs/1911.11763.

Sherman, J. 2017. Area-convexity, ℓ_∞ regularization, and undirected multicommodity flow. In *STOC*, 452–460. ACM.

Villani, C. 2008. *Optimal transport: Old and New*. Springer.

Wang, Z.; Xue, N.; Lei, L.; and Xia, G.-S. 2022. Partial Wasserstein Adversarial Network for Non-rigid Point Set Registration. In *International Conference on Learning Representations*.