

# Sparse Partial Optimal Transport via Quadratic Regularization

<sup>1</sup>Khang Tran, <sup>2</sup>Khoa Nguyen, <sup>3</sup>Anh Nguyen, <sup>4</sup>Thong Huynh, <sup>5</sup>Son Pham, <sup>4</sup>Sy-Hoang Nguyen-Dang, <sup>6,7</sup>Manh Pham, <sup>1</sup>Bang Vo, <sup>7,8</sup>Mai Ngoc Tran and <sup>9</sup>Dung Luong

<sup>1</sup>Department of Computer Science, Ho Chi Minh University of Science, Ho Chi Minh City, Vietnam

<sup>2</sup>School of Science, Aalto University, Espoo, Finland

<sup>3</sup>Department of Science, Lycée Français Alexandre Yersin de Hanoi, Ha Noi, Vietnam

<sup>4</sup>Department of Math, High School for the Gifted, Ho Chi Minh City, Vietnam

<sup>5</sup>Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Massachusetts, United States

<sup>6</sup>Department of Computer Science, Georgia Institute of Technology, Atlanta, Georgia, United States

<sup>7</sup>Acuitas Education, Ho Chi Minh City, Vietnam

<sup>8</sup>Department of Computer Science, Binh Duong University, Ho Chi Minh City, Vietnam

<sup>9</sup>Department of R&D, VietDynamic, Ho Chi Minh City, Vietnam

## Article history

Received: 19-11-2024

Revised: 31-01-2025

Accepted: 24-02-2025

## Corresponding Author:

Khoa Nguyen

School of Science, Aalto University,

Espoo, Finland

Email: khoa.nguyen@aalto.fi

**Abstract:** Partial Optimal Transport (POT) has recently emerged as a central tool in various Machine Learning (ML) applications. It lifts the stringent assumption of the conventional Optimal Transport (OT) that input measures are of equal masses, which is often not guaranteed in real-world datasets, and thus offers greater flexibility by permitting transport between unbalanced input measures. Nevertheless, existing major solvers for POT commonly rely on entropic regularization for acceleration and thus return dense transport plans, hindering the adoption of POT in various applications that favor sparsity. In this paper, as an alternative approach to the entropic POT formulation in the literature, we propose a novel formulation of POT with quadratic regularization, hence termed quadratic regularized POT (QPOT), which induces sparsity to the transport plan and consequently facilitates the adoption of POT in many applications with sparsity requirements. Extensive experiments on synthetic and CIFAR-10 datasets, as well as real-world applications such as color transfer and domain adaptations, consistently demonstrate the improved sparsity and favorable performance of our proposed QPOT formulation.

**Keywords:** Partial Optimal Transport, Quadratic Regularizer, Optimal Transport

## Introduction

Optimal Transport (OT) (Kantorovich, 1942; Villani, 2008) has long been a well-established mathematical framework for comparing probability distributions by finding the minimum-cost solution for transporting mass between them. However, the high computational cost of OT remains a significant drawback, limiting its practicality in large-scale problems. In recent years, the introduction of regularizers, particularly the entropic regularizer (Cuturi, 2013), emerged as an answer to speeding up OT and improving its scalability, garnering significant attention as a powerful tool in modern Machine Learning (ML) applications. OT applications now span many fields of ML, such as color transfer (Pitié *et al.*, 2007; Pouli and Reinhard, 2011), domain adaptation (Redko *et al.*, 2019), and dictionary learning (Rolet *et al.*, 2016). A primary constraint of OT is its

stringent assumption that both the masses of the source and the target distribution are equal, which does not hold in many real-world datasets. This becomes the primary bottleneck for the application of OT to various ML problems (Rubner *et al.*, 2000; Pele and Werman, 2009; Gramfort *et al.*, 2015; Nguyen *et al.*, 2023b), where such an assumption is not guaranteed.

Partial Optimal Transport (POT) (Chapel *et al.*, 2020) was introduced to address the challenges posed by OT and entropic regularized OT, allowing mass to be transported between unbalanced distributions. POT achieves this by relaxing the marginal constraints that OT strictly imposes (Figalli, 2010; Caffarelli and McCann, 2010), providing greater flexibility (Chapel *et al.*, 2020) and improved robustness to outliers (Le *et al.*, 2021). However, as observed by Nguyen *et al.* (2024a), the current implementation of POT in Chapel *et al.*

(2020) is effectively reformulations of OT with extended cost matrices, resulting in a significant increase in computation cost. Additionally, Nguyen *et al.* (2024a) highlighted that the original Sinkhorn algorithm, when naively used to solve POT, is strictly infeasible as it violates some of the foundational constraints of the problem and proposes a revised feasible Sinkhorn algorithm for POT with worsen complexity.

In addition, the entropic regularizer, along with the Sinkhorn algorithm (Cuturi, 2013; Nguyen *et al.*, 2024a) or relevant solvers (Benamou *et al.*, 2015), has recently become prevalent in the optimization of OT and POT problems thanks to its computational acceleration. However, the usage of entropic regularizers induces strictly dense transport plans (Blondel *et al.*, 2018), which is undesirable for several reasons. First, dense transport plans potentially impose significant memory complexity, making them inefficient when scaling to higher dimensions and larger datasets (Peyré and Cuturi, 2020). Second, the lack of sparsity reduces the interpretability of the solution, obscuring the key connections between the source and target distributions (Solomon *et al.*, 2015). Third, such dense transport plans are often sensitive to noise, which undermines the reliability of the result (Genevay *et al.*, 2019). Finally, dense transport plans enforced by the solvers may nullify the usage of POT in many applications (Pitié *et al.*, 2007; Courty *et al.*, 2017; Muzellec *et al.*, 2017), favoring sparse solutions.

Given the worsening complexity of the Sinkhorn algorithm for POT and the lack of sparsity due to the entropic regularizer in the existing solvers (Nguyen *et al.*, 2024a), there is a strong need for an alternative regularization approach that can accelerate POT while inducing sparsity. This work proposes and benchmarks the quadratic regularizer as an effective solution, offering improved performance and encouraging sparsity for the POT solution. To this end, we propose the novel quadratic regularized POT (QPOT) formulation, which augments the objective of the POT problem with quadratic regularization. We empirically evaluate QPOT across extensive settings of synthetic and CIFAR-10 datasets and real-world applications, such as color transfer and domain adaptation. Using entropic regularized POT (EPOT) as the baseline, our benchmarks demonstrate that QPOT achieves superior sparsity, proving its effectiveness in comparison to EPOT.

## Related Works

Applications of OT span a wide range of fields. In economics, (Gan *et al.*, 2024) has shown that using OT in resource allocation for Cloud-Edge Collaborative IoT significantly reduces the average energy consumption and delay. In ML, Courty *et al.* (2017) has applied OT for domain adaptation on toys, challenging real visual adaptation examples, and has shown that the method

consistently outperforms state-of-the-art approaches. Bousquet *et al.* (2017), in addition, has studied unsupervised generative modeling in terms of the OT problem and has shown a better understanding of the commonly observed blurriness of images generated by variational auto-encoders. In deep learning, the use of OT to generalize a deep neural network was studied in Zhang *et al.* (2023), while Oh *et al.* (2020) has applied OT in deep learning approaches for accelerated MRI and was able to reconstruct high-resolution MR images. OT is also used in image processing, as (Rabin and Papadakis, 2015; Blondel *et al.*, 2018; Rabin *et al.*, 2014) have applied OT in addition to other methods to gain great results in color transfer tasks. In biology, OT was applied in (Schiebinger *et al.*, 2019) to study developmental time courses to infer ancestor-descendant fates and model the regulatory programs that underlie them.

However, the classical optimal transport problem aims to find a transportation map that keeps the total mass between two probability distributions, requiring their mass to be the same. In certain cases, this condition can be hard to achieve, therefore leading to the formulation of Partial Optimal Transport (POT). The first known study of POT in ML applications was by (Chapel *et al.*, 2020). The work showed that this method is efficient in scenarios where point clouds come from different domains or have different features. Similarly, Qin *et al.* (2022) proposed a point cloud registration algorithm based on partial optimal transport and showed that the proposed method achieves state-of-the-art registration results when dealing with point clouds with significant amounts of outliers and missing points.

To further improve the OT method, recent studies have been about adding a regularizer term to improve certain aspects of the algorithm. The work of Cuturi (2013), for example, has added an entropic regularizer term into the initial problem and shown that the algorithm runs at a speed that is several orders of magnitude faster than that of transportation solvers. Further studies on the efficiency and convergence of the entropic regularized OT can also be found in Lin *et al.* (2022) and Carlier *et al.* (2017).

Even though the entropic regularizer has proven to increase the computing speed significantly, the resulting transportation plan is usually dense. To resolve this problem, many works have been conducted on studying the application of the quadratic regularizer to the OT problem. Blondel *et al.* (2018) shows that the incorporation of the quadratic norm and group lasso regularizations has led to an improvement in the sparsity of the transport plans. Additionally, Lorenz *et al.* (2021) investigated the problem of optimal transport in the so-called Kantorovich form and derived two algorithms to solve the dual problem of the regularized problem. Their experiments have shown that both methods perform well, even for small regularization parameters. Similarly, Nutz

(2025) has proven the existence of the solution of the dual problem for a general square-integrable cost and that the optimal support is indeed sparse for small regularization parameters in a continuous setting with quadratic cost. On the other hand, in Nguyen *et al.* (2023b), the quadratic regularizer was applied simultaneously with the KL divergence to fill the lack of sparse UOT literature. Furthermore, motivated by an application to a sparse mixture of experts, Liu *et al.* (2022) studied OT with explicit cardinality constraints on the transportation plan and showed that this framework is formally equivalent to using squared k-support norm regularization in the primal. In terms of the computational methods, the two main approaches for solving OT are gradient-based methods and the Sinkhorn algorithm. The Sinkhorn algorithm has been studied in Cuturi (2013) and has shown great computational speed. On the other hand, gradient methods have been studied in Nesterov *et al.* (2021) and An *et al.* (2022) and have shown that the proposed method achieves faster convergence and better accuracy with the same parameter compared to the Sinkhorn algorithm. For POT, Nguyen *et al.* (2024a) has shown that Sinkhorn is not feasible and proposes a rounding algorithm to resolve the problem. In UOT, the gradient method has been studied in Nguyen *et al.* (2023b) in addition to the primal-dual theory.

## Problem Formulation

### Optimal Transport

Define the source and target distribution for the transport as  $r, c \in \mathbb{R}_+^n$  with mass  $\|r\|_1 \geq 0$  and  $\|c\|_1 \geq 0$ . The Optimal Transport aims to move the mass of  $r$  to  $c$  and the other way around with minimal cost. Kantorovich (1942) hence reformulate OT as a Linear Programming (LP) problem:

$$OT(r, c) = \min \langle C, X \rangle \text{ s.t. } X \in \mathcal{T}(r, c) \quad (1)$$

where  $C$  is the cost matrix of moving between  $r$  and  $c$ , and  $\mathcal{T}(r, c)$  being the set of solution  $X$  that minimize the overall cost, defined as:

$$\mathcal{T}(r, c) := \{X \in \mathbb{R}_+^{n \times n} : X \mathbf{1}_n = r, X^T \mathbf{1}_n = c\} \quad (2)$$

Revise the formulation of OT, the two equalities of  $\mathcal{T}(r, c)$  enforces a strict constraint where the transported mass must be equal to the mass of both distributions, leading to the requirement where both masses must also be equal for the calculation of OT.

### Partial Optimal Transport

The Partial Optimal Transport (POT), as defined in Chapel *et al.* (2020) and Le *et al.* (2021), relaxes such constraint from OT by introducing a total mass  $0 \leq s \leq \min\{\|r\|_1, \|c\|_1\}$  that is allowed to be transported

between the two distributions. The introduction of  $s$  hence reformulate 1 as:

$$POT(r, c, s) = \min \langle C, X \rangle \quad (3)$$

such that  $X \in \mathcal{U}(r, c, s)$  where

$$\mathcal{U}(r, c, s) := \{X \in \mathbb{R}_+^{n \times n}, X \mathbf{1}_n \leq r, X^T \mathbf{1}_n \leq c, \mathbf{1}_n^T X \mathbf{1}_n = s\}$$

To elaborate, the addition of  $s$  transforms the previous equalities in 2 into inequalities in 3, allowing the mass of each distribution to be transported partially and hence allowing the masses to differ. The loosening of the original constraint also allows POT to be more flexible and encourages sparsity in the transport plan, accelerating computations while maintaining the key connections between the two distributions.

### Regularizers

To be able to modify Linear Programming problems to focus on achieving certain tasks (e.g, lower runtime, higher accuracy, etc), a novel approach is to add a regularizer term to the objective function.

For OT problems, the entropic regularizer has been applied to both OT and POT (Clason *et al.*, 2021; Cuturi, 2013) and has given some optimistic results. The Entropic regularized OT (EOT) problem is OT with the addition of the regularizer Kullback–Leibler divergence (Shlens, 2014) to stabilize the computation and make it solvable using faster algorithms such as the Sinkhorn-Knopp algorithm. Its formula can be written as:

$$EOT(r, c, \varepsilon) = [\min \langle C, X \rangle - \frac{\varepsilon}{2} H(X)] \text{ s.t. } X \in \mathcal{T}(r, c) \quad (4)$$

$$H(X) = \sum_{i,j} X_{i,j} \log X_{i,j} - X_{i,j}$$

Entropic Regularized Partial OT (EPOT) is the combination of POT and EOT, where our objective function would be the same as Eq. (4) but with the additional condition that  $X \in \mathcal{U}(r, c, s)$  which have been denoted in (3). Formally, it can be written as:

$$EPOT(r, c, \varepsilon) = \min [\langle C, X \rangle - \frac{\varepsilon}{2} H(X)] \text{ s.t. } X \in \mathcal{U}(r, c, s) \quad (5)$$

The greatest reason why the entropic regularizer is favored for OT problems is due to its great algorithmic speed (Cuturi, 2013). However, the limitation of the entropic regularizer is that it would result in a dense transport plan, leading to difficulties in interpreting the general pattern, which may lead to suboptimal solutions in applications where sparsity is important. Additionally, when the regularizer coefficient  $\varepsilon$  becomes very small, the Sinkhorn iterations used to solve the entropic OT problem may converge very slowly or face stability issues due to extremely large or small entries in the transport matrix.

To resolve this problem, we propose the Quadratic Regularized POT (QPOT), which is POT with the additional usage of the quadratic regularizer (also known

as the  $\ell_2$  norm). More formally, the formula for QPOT can be written as:

$$QPOT(r, c, s, \varepsilon) = \min [\langle C, X \rangle + \frac{\varepsilon}{2} \|X\|_F^2] \text{ s.t. } X \in \mathcal{X}(r, c, s) \quad (6)$$

where  $\varepsilon$  is the regularizer coefficient and  $\|T\|_F$  is the Frobenius norm of the matrix  $T$  for all matrix  $T$ .

The proposed method can perform very well due to the simplicity in computations of the novel  $\ell_2$  norm as well as its ability to handle outliers. These properties of the  $\ell_2$  norm have also been studied in Zhu *et al.* (2015) and Zhang *et al.* (2019) and have given satisfactory results in terms of sparsity.

## Methods and Implementation

We conducted our experiments on a system powered by an i7-12700K processor with 64GB of DDR5 RAM, comparing the performance of QPOT and EPOT in terms of sparsity through numerical experiments, color transfer, and domain adaptation tasks. Our implementation leveraged Python's cvxpy library, which provides a natural mathematical interface for expressing convex optimization problems. Additionally, the library offers access to multiple open-source solvers, making it well-suited for our experimental setup.

We selected CLARABEL and SCS as the solvers for QPOT and EPOT, respectively. The CLARABEL solver, introduced by Goulart and Chen (2024), based on a primal-dual interior-point method, aligns effectively with the smooth and convex optimization landscape introduced by the quadratic regularizer. CLARABEL is particularly efficient for quadratic objective functions as it does not require epigraphical reformulation of the objective function, leading to faster computation times. It operates with a tight convergence tolerance of approximately  $10^{-8}$  and  $10^{-6}$ , resulting in highly accurate solutions.

Conversely, the SCS solver (introduced by O'Donoghue *et al.* (2016)) proved more suitable for EPOT. SCS employs Douglas-Rachford splitting combined with acceleration techniques, such as Anderson acceleration from Zhang *et al.* (2020), which enhance its flexibility and efficiency when handling smooth and strongly convex regularizers like the entropic regularizer. The SCS solver achieves a convergence tolerance in the range of  $10^{-3}$  and  $10^{-4}$ , providing a balance between computational speed and accuracy, making it a practical choice for large-scale entropic regularized problems.

## Experiments Setup and Results

The performance of QPOT is benchmarked through three experiments, using EPOT as the baseline. The first experiment involves a numerical analysis on the CIFAR-10 dataset and multiple toy distributions. We also conduct additional experiments on real-world machine

learning applications, such as color transfer and domain adaptation. In all settings, we solve QPOT and EPOT using Python's cvxpy package. Furthermore, our experiment setup aims to study the performance of both regularized POTs and observe how they behave under various parameters.

Two noteworthy parameters that are studied throughout all experiments are  $\lambda$  and  $\varepsilon$ .  $\lambda$  is the proportion of mass to be transported in the process. In other words, let  $\lambda = 0.6$  and  $\|r\| = \|c\| = 1$ , then  $\lambda$  is used to simulate the unbalance between the mass of source and target distributions, in this case, the mass to be transported is  $s = \lambda * \min(\|r\|, \|c\|) = 0.6$ . On the other hand,  $\varepsilon$  is the regularizers' strength, where the higher the value of  $\varepsilon$ , the greater the effect of the regularizers, which drives the solver to focus more on optimizing the regularization term rather than the cost of the transportation.

We measured the performance of both QPOT and EPOT based on their sparsities, defined as the number of entries in the transport plan below a certain threshold. For all four experiments, we set the threshold at  $1e-10$ . As discussed, sparsity is beneficial in the Optimal Transport (OT) problem because it improves computation time and promotes key connections between source and target distributions. Moreover, for better visualization, we color the heatmaps of all transport plans white for values smaller than the threshold  $1e-10$  and black for all other values. Since  $\lambda$  being too small would lead to the sparsity of both methods being too close to 1 and hard to differ, we have chosen  $\lambda$  to be 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, and 0.99% which is equivalent to 50, 60, 70, 80, 90, 95 and 99% of the allowed mass to be transported. Furthermore, for each value of  $\lambda$ , we run  $\varepsilon$  from  $10^{-i}$  to  $10^{-j}$ . Where  $i, j$  will be chosen to numerically fit with each experiment.

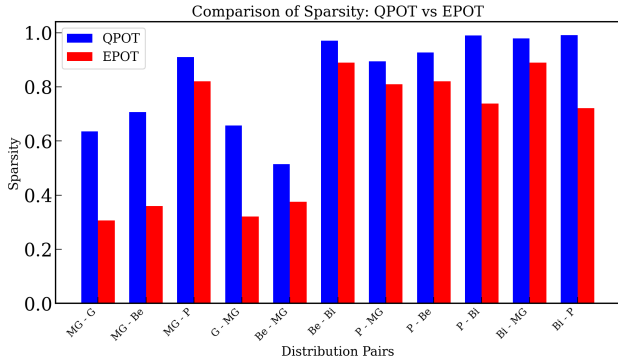
### Numerical Experiments

This section presents the performance of QPOT and EPOT on two datasets. The first dataset is synthetic data sampled based on several of the classic distributions, such as mixed Gaussian, Poisson, Gamma, etc. On the other hand, in the second setup, the masses are drawn from the CIFAR10 dataset.

#### Sparsity Experiment on Toy Distributions

The first setup compares the sparsities of QPOT and EPOT on multiple toy distributions in a one-to-one setting. We selected the pairs of distributions on several combinations of six different classic distributions and generated their mass by sampling them using Python's numpy library. For each type of distribution, we generated  $10^5$  samples, which are divided into 100 bins. Our sampled distributions include: the Gamma distribution (G):  $\Gamma(\alpha = 7, \beta = 1)$ , the Poisson distribution

(P):  $\text{Pois}(\lambda = 5)$ , the Binomial distribution  $\beta$  ( $n = 10, p = 0.4$ ) (Bi); the Beta distribution (Be):  $\text{Beta}(\alpha = 2, \beta = 2)$ , and the Mixed Gaussian distribution (MG):  $\mathcal{N}_1(\mu_1 = 1, \sigma_1 = 2), \mathcal{N}_2(\mu_2 = 10, \sigma_2 = 1.5)$ . The results are plotted out in Figure 1.

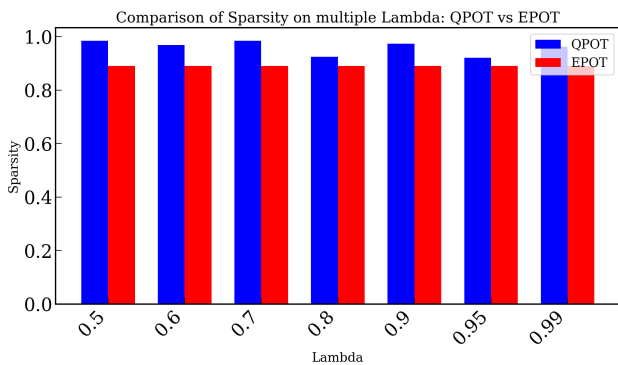


**Fig. 1:** Comparing pairs of toy distributions

It can be observed from Figure 1 that for all experimented pairs, QPOT has a more sparse transport plan compared to EPOT. Furthermore, for some pairs of toy distributions like MG-G or MG-Be, the sparsity of QPOT even doubles that of EPOT, which again highlights the dominance of our proposed method.

### Sparsity Experiment on Multiple Mass

The setting of  $\lambda$  for this experiment is similar to what was mentioned at the beginning of this section. Additionally, we have chosen the regularizer coefficient  $\varepsilon$  to be  $1e - 6$  and run our experiment on two toy distributions which are the Binomial Distribution  $\beta$  ( $n = 10, p = 0.4$ ) and the Mixed Gaussian distribution  $\mathcal{N}_1(\mu_1 = 1, \sigma_1 = 2), \mathcal{N}_2(\mu_2 = 10, \sigma_2 = 1.5)$ . Figure 2 below shows the sparsity of both methods for each  $\lambda$ .



**Fig. 2:** Sparsity of QPOT and EPOT for multiple transport masses

Similar to before, in all conducted experiments, QPOT shows better performance in terms of sparsity compared to EPOT. Furthermore, overall values of  $\lambda$ , QPOT sparsity always stay higher than 0.9, while the same is not true for EPOT. This has confirmed the effectiveness of QPOT in finding a sparse solution even on different masses.

### Sparsity Experiment on Multiple Regularizer Coefficients

The next setup studies the effect of the regularizer strength on the performance of both EPOT and QPOT by varying the value  $\varepsilon$ . As shown in Figure 3, we benchmarked both POT methods on  $\varepsilon$  running from  $10^{-0.5}$  to  $10^{-6}$  with the pair of  $\text{Pois}(\lambda = 5)$  and  $\text{Beta}(\alpha = 2, \beta = 2)$  are selected as the source and target distributions. We also studied this effect on two  $\lambda$  values 0.99 and 0.7, as the first value simulates the setup of almost equal distribution, while it is studied that the sparsities of the two methods are the highest when  $\lambda = 0.7$ . Furthermore, similar to the previous experiment, for each type of distribution, we used  $10^5$  samples to generate and divided them into 100 bins.

In both cases of  $\lambda$ , it can be seen that QPOT surpasses EPOT for all experimented  $\varepsilon$ . Additionally, it can also be observed that for most values of epsilon, the sparsity of EPOT is usually below 0.9, while the sparsity of QPOT mostly stays above 0.95. This, again, shows how dominant QPOT is when it comes to sparsity.



**Fig. 3:** Sparsity on Numerical Experiment on Poisson and Beta with: (a)  $\lambda = 0.7$  (b)  $\lambda = 0.99$

### Sparsity Experiment on CIFAR10 Dataset

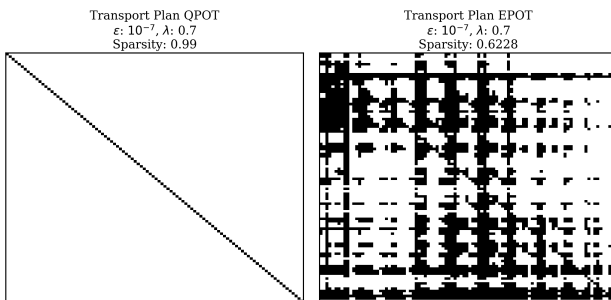
Numerical experimentation on CIFAR-10 allows us to evaluate the optimal transport methods and on standard benchmark datasets where the masses are drawn from images

The CIFAR-10 dataset is a well-established dataset that provides a diverse collection of 60,000 color images

in RGB format, each sized  $32 \times 32$  pixels, divided into 10 object classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. It includes 50,000 training images and 10,000 test images, making it a benchmark for many machine learning tasks that involve images. Despite its low resolution, CIFAR-10 remains challenging due to its diverse and complex visual content. The diversity in object classes, and consequently the differences in pixel value distribution images, makes CIFAR-10 an excellent candidate for this experiment. In this experiment, we used a pair of colored images from the CIFAR-10. We have decided to convert RGB images to grayscale images and to downscale the dimensions of the images from  $32 \times 32$  to  $10 \times 10$ . The grayscale-reduced images are flattened into histograms with 100 bins.

We smooth and normalize both histograms by adding a small value of  $10^{-6}$  to each bin. We calculate the cost matrix  $C$  using the squared Euclidean distance between pixel positions and normalize it so that the maximum value is 1 ( $\|c\|_{max} = 1$ ).

For the experiment,  $\lambda = 0.7$  is chosen as it yielded the best sparsity in the previous experiment. Moreover, we select the regularizer coefficient  $\varepsilon \approx 2 \times 10^{-7}$ . The heat map of the final result is illustrated in Figure 4.



**Fig. 4:** Transport map for QPOT and EPOT on CIFAR10 with  $\lambda = 0.7$  and  $\varepsilon = 10^{-7}$

Just like the aforementioned sections, this experiment has again highlighted the dominance of QPOT in terms of getting a sparse transport plan compared to EPOT. While the EPOT heat map is very dense and barely shows any pattern, the QPOT heat map, on the other hand, has given us a much more recognizable solution.

### Color Transfer



**Fig. 5:** Dataset for Color Transfer

Color transfer is an image editing process where the color palette of one image is applied to another. It is often used to match the mood, tone, or aesthetic of two images by transferring color characteristics (hue, saturation, and brightness) from a reference image to a target image and has applications in robotics and ML (He *et al.*, 2018; Lee and Han, 2005). The key to color transfer is that the new image would have the same color palette as the target image without losing its contextual information.

Define an RGB image as  $x \in \mathbb{R}^{h \times w \times 3}$  with each color channel is a matrix of  $\mathbb{R}^{h \times w}$  corresponding to Red, Green, or Blue. A pixel  $x_{ij}$  in an image at coordinate  $(i, j)$  hence consists of three values  $x_{ij} = \{\text{red, green, blue}\}$  where each value range from 0 to 255, where the higher the value, the stronger the color component for that channel. Hence, each image can be interpreted as an array of pixels, with each color channel forming a color distribution. Therefore, the goal of the color transfer task is equivalent to transporting the color distribution of the source image to the target image.

The data for this experiment is a pair of images (Figure 5) representing the source and the target distribution, each having the dimension of  $256 \times 256$ . This pair of images is chosen because of their distinct color palettes. However, instead of conducting the experiments on the RGB images, we convert the images to the LUV color space for two main reasons. Firstly, the LUV color space primarily encodes the color information as the U and V channels, consequently reducing the dimensionality of the color space from three (Red-Green-Blue) to two (UV-color chromaticity). Moreover, LUV is more accurate in representing the colors, which is in contrast to RGB, where two colors can be close in the color space but appear very different to the human eye. The conversion of RGB to LUV is as follows:  $L = R + G + B$ ,  $U = G/L$ , and  $V = B/L$ .

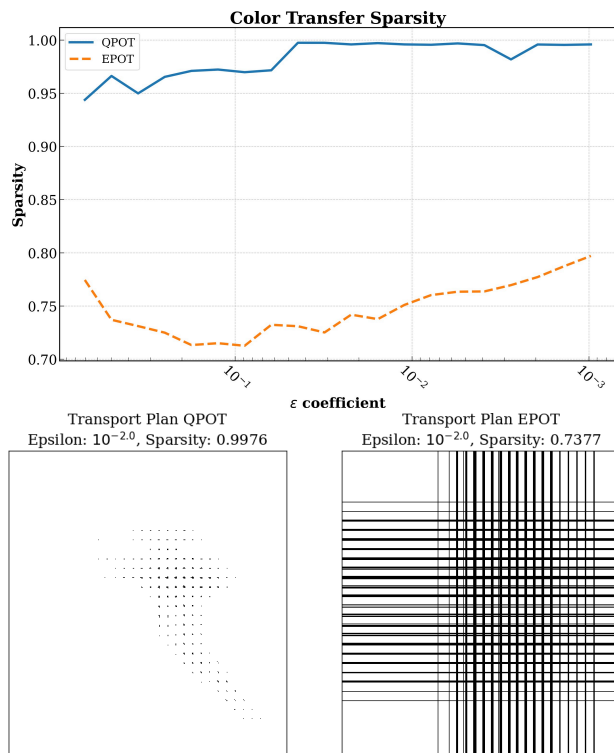
After converting the color space, we extract the color histogram with  $n$  bins for each image by taking the frequency of each U and V channel. These color histograms then undergo smoothing and normalization, yielding the total mass of each histogram to  $\|r\|_1 = 1$  and  $\|c\|_1 = 1$ . From the histogram, we then defined the cost matrix  $C$  as  $C_{i,j} = \|a_i - b_j\|_2^2$  where  $a_i \in a_1, \dots, a_n$  are the bin values of each histogram. Since we are using  $256 \times 256$  pictures, setting  $\varepsilon$  too small would lead to numerical problems while calculating. Thus, we have chosen to set  $i, j$  to  $-0.2$  and  $-6$  respectively, and use the step size of  $10^{-0.2}$ .

The resulting sparsity is seemingly the same for all experimented  $\lambda$ . We have chosen to plot out the case of  $\lambda = 0.7$  in Figure 6. It can be observed that QPOT completely dominates EPOT in terms of sparsity, ranging over all the values of  $\varepsilon$ . This greatly highlights the effectiveness of QPOT in getting a sparse transportation

plan. In Figure 6 we have also drawn the heatmaps of the results for  $\epsilon = 10^{-2}$  of the two methods. These heat maps are drawn by first setting all the values of the transported plan from 0 to  $1e-10$  and the rest to 1. The zero-values, then, are plotted white, and the one-values are plotted black. It can be seen that QPOT has a much more sparse transport plan, which clearly shows the pattern of the solution. The heat maps of EPOT, on the other hand, are very dense and have no pattern whatsoever. This firmly confirms the effectiveness of QPOT in terms of sparsity.

### Domain Adaptation

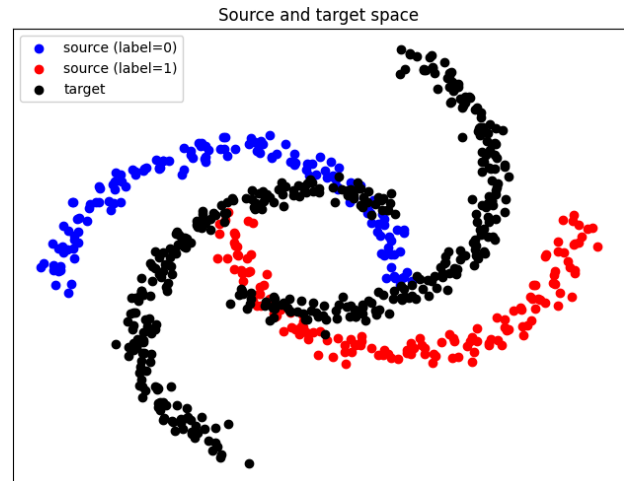
In many real-world applications, the process of labeling a newly collected and unlabeled data set (or target domain) requires the usage of models trained on largely annotated data sets (or source domains). However, problems such as the differences in the probability distribution can impede the effectiveness of directly applying the learned models to the latter. In recent years, new ML research has been conducted to resolve this problem. As a result, Domain Adaptation (DA), a family of techniques that handles cases where source and target samples follow different probabilities, was invented to adapt models trained from the source data onto the target data.



**Fig. 6:** (a) Sparsity graph  $\lambda = 0$ . (a) Sparsity graph  $\lambda = 0$ ; (b) Transport plans of QPOT and EPOT  $\lambda = 0$ . (b) Transport plans of QPOT and EPOT  $\lambda = 0$

In our experiment's setting, the "moon" dataset, a widely-used scikit-learn (Pedregosa *et al.*, 2011) synthetic dataset for benchmarking classification

algorithms on linearly non-separable data, is used for both the source dataset and the target dataset. This dataset consists of two interleaving half circles representing two distinct classes, commonly used to test algorithms capable of handling complex, non-linear classification boundaries. The experiment requires a source and a target dataset where the target distribution is created by rotating the source distribution by 50 degrees, simulating a domain covariate shift.



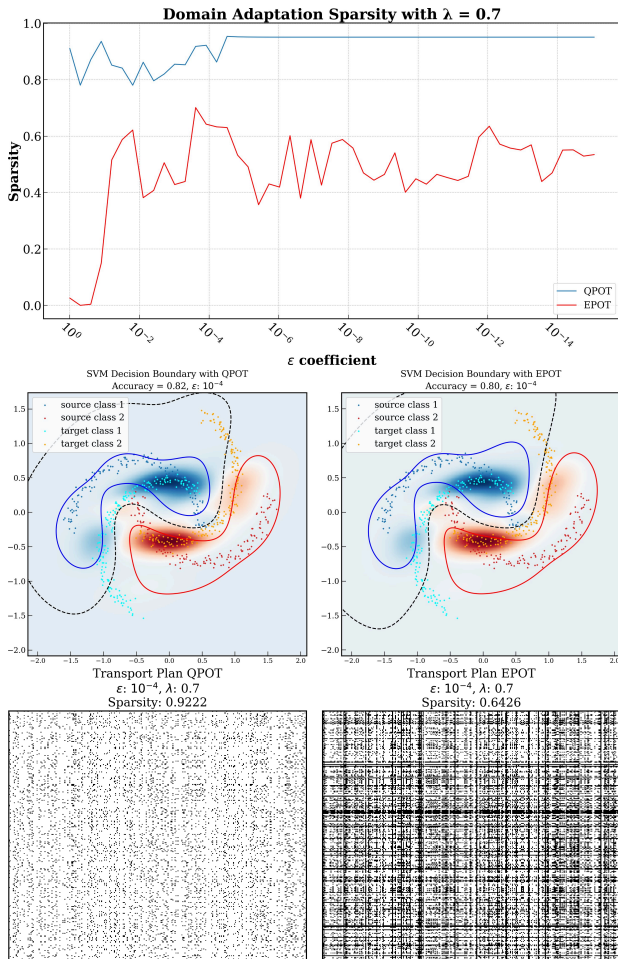
**Fig. 7:** Moon Dataset for domain adaptation

Figure 7 illustrates the dataset where the source dataset has a total of 300 data points, with 150 data points for each class. The target dataset, on the other hand, has a total of 400 data points that are yet to be labeled. Since the number of data points in the target dataset is larger than the amount in the source datasets, we first use K-means clustering to reduce the number of data points in the target dataset down to 300. The cost matrix is then calculated based on the Euclidean distance and is normalized by dividing all numbers by the largest number.

The metric for measuring QPOT and EPOT performance is typically the sparsity. However, in this experiment, to measure how accurately our domain adaptation performed, we train a support vector machine (SVM) model on the target dataset predicted by our domain adaptation and then test it on the source dataset to find the accuracy. This accuracy shows how similar the classification of the target dataset is compared to the classification of the source dataset.

The results of benchmarking QPOT and EPOT on domain adaptation with  $\epsilon$  varying from  $10^{-0.3}$  to  $10^{-15}$  and  $\lambda = 0.7$  are illustrated in Figure 8. The Figure suggests that the sparsity QPOT achieves, averaging around 0.9, dramatically surpasses that of EPOT, which typically hovers around 0.5 and peaks at around 0.7. This performance of sparsity applies to all  $\lambda$  values in the experiment. The accuracy of the SVM classifier also shows that there is no tradeoff between sparsity and accuracy from QPOT, where the method provides a good

accuracy consistently over the  $\epsilon$  range. Specifically, the accuracies of both solvers are illustrated in Figure 8 in the case of  $\epsilon = 10^{-4}$ , showing an improvement in accuracy from QPOT over EPOT while maintaining high sparsity. Moreover, the heat maps of both methods have also been plotted in Figure 8 to further highlight the dominant sparsity of the transport plan created by QPOT compared to that created by EPOT.



**Fig. 8:** (a) Sparsity (b) Accuracy with decision boundary of SVM, and (c) Transport plan comparison of QPOT and EPOT for Domain Adaptation with  $\lambda = 0.7$

## Discussion

As shown in Figure 3, Figure 4, Figure 6, and Figure 8(a, c), introducing quadratic terms into the formulation of Partial Optimal Transport (POT) significantly improves sparsity compared to other variants for various tasks. Moreover, QPOT demonstrates strong robustness across different distribution pairs (Figure 1) and varying transport mass levels (Figure 2). Notably, in real-world tasks such as color transfer and domain adaptation, QPOT achieves over a 20% improvement in performance compared to its counterpart, EPOT. In particular, for domain adaptation, where accuracy serves as an important metric, QPOT also outperforms EPOT (Figure

8(b)), further demonstrating its effectiveness. These results underscore the potential of our approach as a strong foundation for future advancements in sparse OT frameworks. We believe this formulation opens up new directions for both theoretical exploration and practical applications, especially in areas requiring sparse and scalable solutions.

## Conclusion

In this paper, we proposed and examined the Quadratic-regularized Partial Optimal Transport formulation (QPOT) and implemented a variety of experiments to show its superiority compared to the classical Entropic-regularized Partial Optimal Transport (EPOT) formulation. In particular, we experimented with both synthetic data, the CIFAR-10 dataset, and then applied our formulation to the Domain Adaptation and Color Transfer applications. For each experiment, we tested different values of the transport mass and the regularizer coefficient to show the proficiency and robustness of QPOT.

While entropic regularization is widely favored for its compatibility with the celebrated Sinkhorn algorithm, recent advancements in gradient-based methods suggest that quadratic regularization offers a more seamless integration. Notably, works such as Blondel *et al.*, (2018) and Nguyen *et al.* (2023b) demonstrated the superior efficiency of quadratic regularization in practice. Consistent with these findings, our experiments show that QPOT consistently outperforms EPOT across all settings in terms of sparsity, producing transport plans that make it significantly easier to infer general patterns.

We believe that the new QPOT formulation will lay the foundation for many interesting future works. For example, the sparsity and numerical stability of the QPOT could further enhance the performance and thus facilitate its adoption in many ML and generative AI applications (Nguyen *et al.*, 2025; Vu *et al.*, 2025; Li *et al.*, 2024). Another approach for future improvements would be to develop accelerated computational methods, such as APDAGD (Dvurechensky *et al.*, 2018), to solve QPOT. Additionally, one can adapt Stochastic or Constrained Decentralized Optimization methods (Wai *et al.*, 2017; Nguyen *et al.*, 2024b, 2018; Zhang *et al.*, 2022) to develop sample-efficient computational methods for noisy, dynamic and multi-agent applications (Dinh and Nguyen, 2020; Nguyen *et al.*, 2020, 2023c; Nguyen-Vinh *et al.*, 2024) arising in modern distributed settings (Nguyen *et al.*, 2021, 2023a).

## Acknowledgment

We thank the Journal of Computer Science for providing the resources and platform that made this publication possible. The journal offered us not only a place to share our findings, but also a space where our interest in the topic had room to grow and be expressed.



We appreciate the editorial team's support throughout the review process.

We also extend our sincere thanks to VietDynamic and Binh Duong University for their valuable technical insight and support in shaping the problem formulation. Their contributions were instrumental in guiding the experiments toward meaningful real-world applications and in scaling the method to larger settings.

## Funding Information

We want to express our gratitude to VietDynamic and Binh Duong University for their support of this research.

## Authors' Contributions

**Khang Tran and Khoa Nguyen:** Co-authored the manuscript and conducted experiments on synthetic data, color transfer, and domain adaptation applications.

**Anh Nguyen and Mai Tran:** Focused on the CIFAR-10 experiments.

**Thong Huynh:** Executed and supported the domain adaptation experiments.

**Son Pham, Manh Pham, and Mai Ngoc Tran:** Contributed to the QPOT formulation, manuscript writing, and identification of future research directions.

**Sy-Hoang Nguyen-Dang:** Assisted with domain adaptation experiments.

**Bang Vo:** Optimized experiments for large-scale settings.

**Dung Luong:** Helped run large-scale color transfer experiments and provided technical and industrial application guidance.

## Ethics

This manuscript is the authors' original work and has not been previously published elsewhere.

## References

- An, D., Lei, N., Xu, X., & Gu, X. (2022). Efficient Optimal Transport Algorithm by Accelerated Gradient Descent. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(9), 10119-10128. <https://doi.org/10.1609/aaai.v36i9.21251>
- Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., & Peyré, G. (2015). Iterative Bregman Projections for Regularized Transportation Problems. *SIAM Journal on Scientific Computing*, 37(2), A1111-A1138. <https://doi.org/10.1137/141000439>
- Blondel, M., Seguy, V., & Rolet, A. (2018). Smooth and sparse optimal transport. *International Conference on Artificial Intelligence and Statistics*, 880-889.
- Bousquet, O., Gelly, S., Tolstikhin, I., Simon-Gabriel, C.-J., & Schoelkopf, B. (2017). From optimal transport to generative modeling: the vegan cookbook. *ArXiv:1705.07642*.

- Caffarelli, L., & McCann, R. (2010). Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Annals of Mathematics*, 171(2), 673-730. <https://doi.org/10.4007/annals.2010.171.673>
- Carlier, G., Duval, V., Peyré, G., & Schmitzer, B. (2017). Convergence of Entropic Schemes for Optimal Transport and Gradient Flows. *SIAM Journal on Mathematical Analysis*, 49(2), 1385-1418. <https://doi.org/10.1137/15m1050264>
- Clason, C., Lorenz, D. A., Mahler, H., & Wirth, B. (2021). Entropic regularization of continuous optimal transport problems. *Journal of Mathematical Analysis and Applications*, 494(1), 124432. <https://doi.org/10.1016/j.jmaa.2020.124432>
- Courty, N., Flamary, R., Tuia, D., & Rakotomamonjy, A. (2017). Optimal Transport for Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9), 1853-1865. <https://doi.org/10.1109/tpami.2016.2615921>
- Dinh, M. N., & Nguyen, Q. M. (2020). Measurements of errors in large-scale computational simulations at runtime. *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*. 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh, Vietnam. <https://doi.org/10.1109/rivf48685.2020.9140785>
- Dvurechensky, P., Gasnikov, A., & Koshnina, A. (2018). Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn's Algorithm. *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. <https://doi.org/10.1007/s00205-008-0212-7>
- Gan, D., Ge, X., & Li, Q. (2024). An Optimal Transport-Based Federated Reinforcement Learning Approach for Resource Allocation in Cloud-Edge Collaborative IoT. *IEEE Internet of Things Journal*, 11(2), 2407-2419. <https://doi.org/10.1109/jiot.2023.3292368>
- Genevay, A., Chizat, L., Bach, F., Cuturi, M., & Peyré, G. (2019). Sample complexity of sinkhorn divergences. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, 1574-1583.
- Gramfort, A., Peyré, G., & Cuturi, M. (2015). Fast optimal transport averaging of neuroimaging data. *International Conference on Information Processing in Medical Imaging*, 261-272.
- He, K., Wang, R., Tao, D., Cheng, J., & Liu, W. (2018). Color Transfer Pulse-Coupled Neural Networks for Underwater Robotic Visual Systems. *IEEE Access*, 6, 32850-32860. <https://doi.org/10.1109/access.2018.2845855>
- Kantorovich, L. V. (1942). On the translocation of masses. *English Translation in Journal of Mathematical Sciences*, 37(4), 227-229.

- Le, K., Nguyen, H., Nguyen, Q., Ho, N., Pham, T., & Bui, H. (2006). On robust optimal transport: Computational complexity, low-rank approximation, and barycenter computation. *ArXiv:2102.06857*.
- Lee, H.-S., & Han, D. (2005). Implementation of real time color gamut mapping using neural network. *Proceedings of the 2005 IEEE Midnight-Summer Workshop on Soft Computing in Industrial Applications, 2005. SMCia/05.*, 138-141. <https://doi.org/10.1109/smcia.2005.1466962>
- Li, X., Chen, J., Chai, Y., & Xiong, H. (2024). GiLOT: Interpreting generative language models via optimal transport. *Proceedings of the 41st International Conference on Machine Learning*, 27515-27530.
- Lin, T., Ho, N., & Jordan, M. I. (2022). On the efficiency of entropic regularized algorithms for optimal transport. *Journal of Machine Learning Research*, 23(137), 1-42.
- Liu, T., Puigcerver, J., & Blondel, M. (2022). Sparsity-constrained optimal transport. *ArXiv:2209.15466*.
- Lorenz, D. A., Manns, P., & Meyer, C. (2021). Quadratically Regularized Optimal Transport. *Applied Mathematics & Optimization*, 83(3), 1919-1949. <https://doi.org/10.1007/s00245-019-09614-w>
- Muzellec, B., Nock, R., Patrini, G., & Nielsen, F. (2017). Tsallis Regularized Optimal Transport and Ecological Inference. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10854>
- Nesterov, Y., Gasnikov, A., Guminov, S., & Dvurechensky, P. (2021). Primal-dual accelerated gradient methods with small-dimensional relaxation oracle. *Optimization Methods and Software*, 36(4), 773-810. <https://doi.org/10.1080/10556788.2020.1731747>
- Nguyen, A. D., Nguyen, T. D., Nguyen, Q. M., Nguyen, H. H., Nguyen, L. M., & Toh, K.-C. (2024a). On Partial Optimal Transport: Revising the Infeasibility of Sinkhorn and Efficient Gradient Methods. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8), 8090-8098. <https://doi.org/10.1609/aaai.v38i8.28648>
- Nguyen, H. H., Li, Y., & Zhao, T. (2024b). Stochastic constrained decentralized optimization for machine learning with fewer data oracles: a gradient sliding approach. *ArXiv:2404.02511*.
- Nguyen, L. M., Nguyen, P. H., van Dijk, M., Richtárik, P., Scheinberg, K., & Takác, M. (2018). Sgd and hogwild! convergence without the bounded gradients assumption. *International Conference on Machine Learning*, 3750-3758.
- Nguyen, M. Q., Loghin, D., & Dinh, T. T. A. (2021). Understanding the scalability of Hyperledger Fabric. *ArXiv:2107.09886*.
- Nguyen, Q. M., Jeong, H., & Grover, P. (2020). Coded QR Decomposition. *2020 IEEE International Symposium on Information Theory (ISIT)*. 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA. <https://doi.org/10.1109/isit44484.2020.9173985>
- Nguyen, Q. M., Khanh Le, N., & Nguyen, L. M. (2023a). Scalable and Secure Federated XGBoost. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece. <https://doi.org/10.1109/icassp49357.2023.10097233>
- Nguyen, Q. M., Nguyen, H. H., Zhou, Y., & Nguyen, L. M. (2023b). On unbalanced optimal transport: Gradient methods, sparsity and approximation error. *Journal of Machine Learning Research*, 24(1), 18430-18430.
- Nguyen, Q. M., Nguyen, L. M., & Das, S. (2023). Correlated Attention in Transformers for Multivariate Time Series. *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1-5. <https://doi.org/10.1109/ICASSP49660.2025.10889540>
- Nguyen, Q. M., Weissburg, I., & Jeong, H. (2023c). Coded computing for fault-tolerant parallel qr decomposition. *ArXiv:2311.11943*.
- Nguyen-Vinh, K., Vo-Huynh, Q.-N., Nguyen-Minh, K., Hoang, M., & Rangaraju, S. (2024). Case Study: Utilising of Deep Learning Models for Fault Detection and Diagnosis of Photovoltaic Modules to Improve Solar Energy Constructions' O&M Activities Quality. 53-67. [https://doi.org/10.1007/978-981-99-4792-8\\_5](https://doi.org/10.1007/978-981-99-4792-8_5)
- Nutz, M. (2025). Quadratically Regularized Optimal Transport: Existence and Multiplicity of Potentials. *SIAM Journal on Mathematical Analysis*, 57(3), 2622-2649. <https://doi.org/10.1137/24m1638136>
- O'Donoghue, B., Chu, E., Parikh, N., & Boyd, S. (2016). Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding. *Journal of Optimization Theory and Applications*, 169(3), 1042-1068. <https://doi.org/10.1007/s10957-016-0892-3>
- Oh, G., Sim, B., Chung, H., Sunwoo, L., & Ye, J. C. (2020). Unpaired Deep Learning for Accelerated MRI Using Optimal Transport Driven CycleGAN. *IEEE Transactions on Computational Imaging*, 6, 1285-1296. <https://doi.org/10.1109/tci.2020.3018562>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825-2830.

- Pele, O., & Werman, M. (2009). Fast and robust Earth Mover's Distances. *2009 IEEE 12th International Conference on Computer Vision*. 2009 IEEE 12th International Conference on Computer Vision (ICCV), Kyoto.  
<https://doi.org/10.1109/iccv.2009.5459199>
- Peyré, G., & Cuturi, M. (2020). Computational optimal transport. *ArXiv:1803.00567*.
- Pitié, F., Kokaram, A. C., & Dahyot, R. (2007). Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1-2), 123-137.  
<https://doi.org/10.1016/j.cviu.2006.11.011>
- Pouli, T., & Reinhard, E. (2011). Extended papers from npar 2010: Progressive color transfer for images of arbitrary dynamic range. *Computers and Graphics*, 35(1), 67-80.
- Qin, H., Zhang, Y., Liu, Z., & Chen, B. (2022). Rigid Registration of Point Clouds Based on Partial Optimal Transport. *Computer Graphics Forum*, 41(6), 365-378.  
<https://doi.org/10.1111/cgf.14614>
- Rabin, J., Ferradans, S., & Papadakis, N. (2014). Adaptive color transfer with relaxed optimal transport. *2014 IEEE International Conference on Image Processing (ICIP)*. 2014 IEEE International Conference on Image Processing (ICIP), Paris, France.  
<https://doi.org/10.1109/icip.2014.7025983>
- Rabin, J., & Papadakis, N. (2015). *Non-convex Relaxation of Optimal Transport for Color Transfer Between Images*. 87-95.  
[https://doi.org/10.1007/978-3-319-25040-3\\_10](https://doi.org/10.1007/978-3-319-25040-3_10)
- Rakotomamonjy, A., Flamary, R., Gasso, G., Alaya, M. E., Berar, M., & Courty, N. (2022). Optimal transport for conditional domain matching and label shift. *Machine Learning*, 111(5), 1651-1670.  
<https://doi.org/10.1007/s10994-021-06088-2>
- Redko, I., Courty, N., Flamary, R., & Tuia, D. (2019). Optimal Transport for Multi-source Domain Adaptation under Target Shift. *22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 849-858.
- Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A., & Lander, E. S. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell*, 176(4), 928-943.e22.  
<https://doi.org/10.1016/j.cell.2019.01.006>
- Solomon, J., Goes, F. de, Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., & Guibas, L. (2015). Convolutional Wasserstein Distances: Efficient Optimal Transportation on Geometric Domains. *ACM Transactions on Graphics (TOG)*, 34(4), 1-11. <https://doi.org/10.1145/2766963>
- Vu, B., Nguyen, H., Nguyen, Q. M., Le, D., Pham, H., & Nguyen, P. L. (2025). *Foundation Model and Temporal Priors-guided Transductive Few-shot Action Recognition*. ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India.  
<https://doi.org/10.1109/ICASSP49660.2025.10888090>
- Wai, H.-T., Lafond, J., Scaglione, A., & Moulines, E. (2017). Decentralized Frank-Wolfe Algorithm for Convex and Nonconvex Problems. *IEEE Transactions on Automatic Control*, 62(11), 5522-5537. <https://doi.org/10.1109/TAC.2017.2685559>
- Zhang, J., Liu, T., & Tao, D. (2023). An Optimal Transport Analysis on Generalization in Deep Learning. In *IEEE Transactions on Neural Networks and Learning Systems* (Vol. 34, Issue 6, pp. 2842-2853).  
<https://doi.org/10.1109/tnnls.2021.3109942>