Journal of Computer Science

Original Research Paper

Sparse Partial Optimal Transport via Quadratic Regularization

Khang Tran[1]*, Khoa Nguyen[2]*, Anh Nguyen[3], Thong Huynh[4], Son Pham[5], Sy-Hoang Nguyen-Dang[4], Manh Pham[6, 7], Bang Vo[1], Mai Ngoc Tran[8, 7], Dung Luong[9]

[1]*Ho Chi Minh University of Science, Ho Chi Minh City, Vietnam;*
[2]*Aalto University, Espoo, Finland;*
[3]*Lycée Français Alexandre Yersin de Hanoi, Ha Noi, Vietnam;*
[4]*High School for the Gifted, Ho Chi Minh City, Vietnam;*
[5]*University of Massachusetts Amherst, Massachusetts, USA;*
[6]*Georgia Institute of Technology, Atlanta, Georgia, USA;*
[7]*Acuitas Education, Ho Chi Minh City, Vietnam;*
[8]*Binh Duong University, Ho Chi Minh City, Vietnam;*
[9]*VietDynamic, Ho Chi Minh City, Vietnam;*

Article history
Received:
Revised:
Accepted:

Corresponding Authors:
Khoa Nguyen
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 favoring 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), 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

*These authors contributed equally to this work.

Science Publications

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., 2023a), the current implementations of POT in (Chapel et al., 2020) and (Le et al., 2021) are effectively reformulations of OT with extended cost matrices, resulting in a significant increase in computation cost. Additionally, (Nguyen et al., 2023a) 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., 2023a) 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., 2018a), 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., 2017a; Muzellec et al., 2016) favoring sparse solutions.

Contribution: Given the worsen complexity of the Sinkhorn algorithm for POT and the lack of sparsity due to entropic regularizor in the existing solvers (Nguyen et al., 2023a), 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., 2017b) has applied OT for domain adaptation on toys, and 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., 2018b; 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; 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. The work in (Blondel et al., 2018a) shows that the incorporation of the quadratic norm and group lasso regularizations have led to the 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, 2024) 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., 2024c), 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; 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., 2024c) in addition to the primal-dual theory.

Problem Formulation

Optimal Transport

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

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

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

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

. Revise the formulation of OT, the two equalities of $\mathscr{U}(\mathbf{r}, \mathbf{c})$ 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 $0 \leq s \leq \min\{\|\mathbf{r}\|_1, \|\mathbf{c}\|_1\}$ that is allowed to be transported between the two distribution. The introduction of s hence reformulate 1 as:

$$\begin{aligned} &\mathbf{POT}(\mathbf{r}, \mathbf{c}, s) = \min\langle \mathbf{C}, \mathbf{X}\rangle \\ &\text{s.t.} \quad \mathbf{X} \in \mathscr{U}(\mathbf{r}, \mathbf{c}, s) \end{aligned} \tag{3}$$

where

$$\begin{aligned} \mathscr{U}(\mathbf{r}, \mathbf{c}, s) := \{&\mathbf{X} \in \mathbb{R}_+^{n\times n}, \mathbf{X}\mathbf{1}_n \leq \mathbf{r}, \mathbf{X}^T\mathbf{1}_n \leq \mathbf{c}, \\ &\mathbf{1}_n^T\mathbf{X}\mathbf{1}_\mathbf{n} = s\}. \end{aligned}$$

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 encourage 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 an addition of the regularizer Kullback–Leibler divergence (Kullback, 1951) to stabilize the computation and make it solvable using faster algorithms such as the Sinkhorn-Knopp algorithm. Its formula can be written as:

$$\begin{aligned} &\mathbf{EOT}(\mathbf{r}, \mathbf{c}, \varepsilon) = \min\left[\langle \mathbf{C}, \mathbf{X}\rangle - \frac{\varepsilon}{2}H(\mathbf{X})\right] \\ &\text{s.t.} \quad \mathbf{X} \in \mathscr{T}(\mathbf{r}, \mathbf{c}) \end{aligned} \tag{4}$$

where:

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

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 $\mathbf{X} \in \mathscr{U}(\mathbf{r}, \mathbf{c}, s)$ which have been denoted in (3). Formally, it can be written as:

$$\begin{aligned} &\mathbf{EPOT}(\mathbf{r}, \mathbf{c}, \varepsilon) = \min \left[\langle \mathbf{C}, \mathbf{X} \rangle - \frac{\varepsilon}{2} H(\mathbf{X}) \right] \\ &\text{s.t.} \quad \mathbf{X} \in \mathscr{U}(\mathbf{r}, \mathbf{c}, s) \end{aligned} \tag{5}$$

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 ε 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 ℓ_2 norm). More formally, the formula for QPOT can be written as:

$$\begin{aligned} &\mathbf{QPOT}(r, c, s, \varepsilon) = \min \left[\langle \mathbf{C}, \mathbf{X} \rangle + \frac{\varepsilon}{2} ||\mathbf{X}||_F^2 \right] \\ &\text{s.t.} \quad \mathbf{X} \in \mathscr{U}(\mathbf{r}, \mathbf{c}, s) \end{aligned} \tag{6}$$

where ε is the regularizer coefficient and $||\mathbf{T}||_F$ is the Frobenius norm of the matrix $\mathbf{T}$ for all matrix $\mathbf{T}$.

The proposed method can perform very well due to the simplicity in computations of the novel ℓ_2 norm as well as its ability to handle outliers. These properties of the ℓ_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.

Experiments

The performance of QPOT is benchmarked on three experiments with EPOT as the baseline. The first experiment is the numerical experiment on the CIFAR10 dataset and multiple toy distributions. Additional experiments on real-world machine-learning applications, such as color transfer and domain adaptation, are also utilized. In all settings, QPOT and EPOT are solved using Python's *cvxpy* package. Moreover, the experiment's setup is to study the performance of both regularized POTs and how they behave on various parameters.

Two noteworthy parameters that are studied throughout all experiments are λ and ε. λ is the proportion of mass to be transported in the process. In other words, let $\lambda = 0.6$ and $\|\mathbf{r}\|_1 = \|\mathbf{c}\|_1 = 1$, then λ is used to simulate the unbalance between the mass of source and target distributions, in this case, the mass to be transported is $s = \lambda * \min(\|\mathbf{r}\|_1, \|\mathbf{c}\|_1) = 0.6$. On the other hand, ε is the regularizers' strength, where the higher the value of ε, 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.

The performance of both QPOT and EPOT are measured by their sparsities, which is the number of below-threshold entries in the transport plan. The threshold for all four experiments is set at $1e-10$. As discussed, sparsity is desired in the problem of OT since it allows for better computation time and encourages key connections between source and target distributions. Additionally, for better visualization, the heat maps of all transport plans will be colored white at values smaller than the threshold $1e-10$ and black at others.

Since λ being too small would lead to the sparsity of both methods being too close to 1 and hard to differ, we have chosen λ 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 λ, we run ε 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

Sparsity experiment on multiple pair of distributions

The first setup compares the sparsities of QPOT and EPOT on multiple toy distributions in a one-to-one setting. The pairs of distribution are selected on several combinations of six different classic distributions and were sampled using Python's *numpy* library. Our sampled distributions include: the Gamma distribution (G): $\Gamma(\alpha = 7, \beta = 1)$, the Poisson distribution (P): $Pois(\lambda = 5)$, the Binomial distribution $\beta(n = 10, p = 0.4)$ (Bi):, the Beta distribution (Be): $Beta(\alpha = 2, \beta = 2)$, 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.

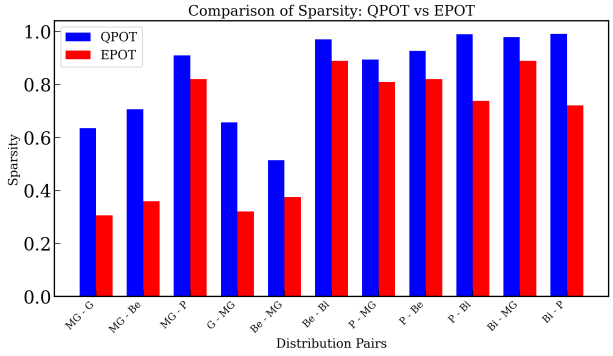


Fig. 1. Comparing on pairs of toy distributions

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 λ for this experiment is similar to what was mentioned at the beginning of this section. Additionally, we have chosen the regularizer coefficient ε 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 λ.

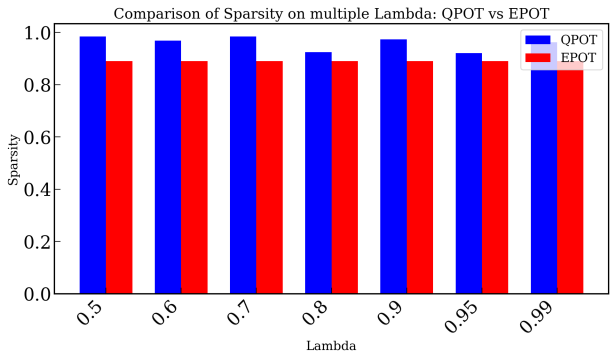


Fig. 2. Sparsity of QPOT and QPOT for multiple transport masses

Similar to before, in all conducted experiments, QPOT shows better performance in terms of sparsity compared to EPOT. Furthermore, overall values of λ, 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 coefficient

The next setup studies the effect of the regularizer strength on the performance of both EPOT and QPOT by varying the value ε. With that in mind, both POT methods are benchmarked on ε running from $10^{-0.5}$ to 10^{-6} with the pair of $Pois(\lambda = 5)$ and $Beta(\alpha = 2, \beta = 2)$ are selected as the source and target distributions. This effect is also studied on two λ 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 $\lambda = 0.7$.

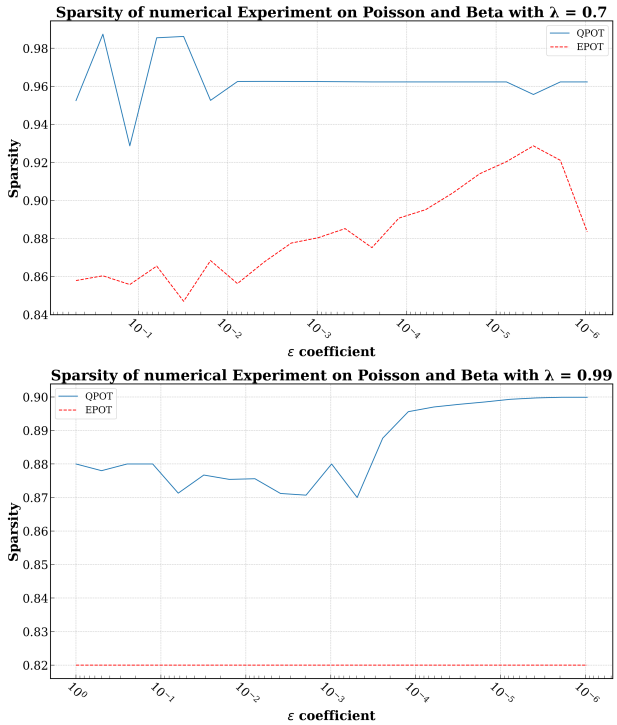


Fig. 3. Sparsity on Numerical Experiment on Poisson and Beta with: (a) $\lambda = 0.7$ (b) $\lambda = 0.99$

In both cases of λ, it can be seen that QPOT surpasses EPOT for all experimented ε. 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.

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 is a well-established dataset that provides a diverse collection of images across ten distinct classes, with 6000 images per class. In this experience, we used a pair of colored images from the CIFAR-10, each having a dimension of 32 x 32 in RGB format. We have decided to convert RGB images to grayscale images and to downscale the dimensions of the images from 32 x 32 to 10 x 10. The grayscale-reduced images are flattened into histograms with 100 bins.

We smoothen 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.

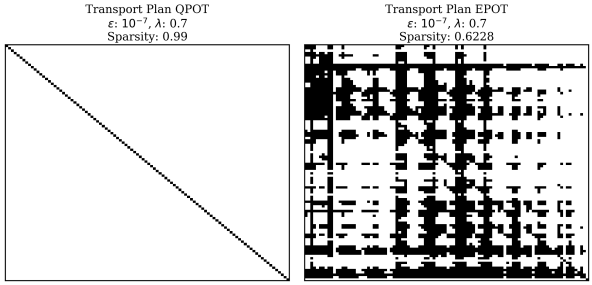


Fig. 4. Transport map for QPOT and EPOT on CIFAR10 with $\lambda = 0.7$ and $\varepsilon = 10^{-7}$

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. 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.

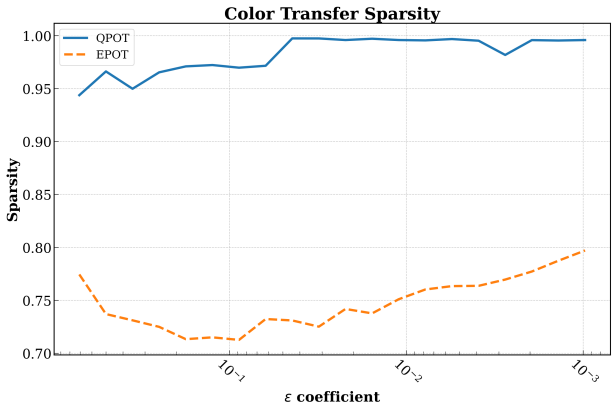
Define an RGB image as $\mathbf{x} \in \mathbb{R}^{h \times w \times 3}$ with each color channel is a matrix of $\mathbb{R}^{h \times w}$ corresponding to Red, Green, or Blue. A pixel x_{ij} in an image at coordinate (i, j) hence consists of three values $x_{ij} = \{\text{red, green, blue}\}$ where each 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 256×256. This pair of images is chosen because of their distinct color palettes. However, instead of carrying out the experiments on the RGB images, the images are instead converted to the LUV color space for two 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 eyes. The conversion of RGB to LUV is as follows: $L = R + G + B$, $U = G/L$, and $V = B/L$.

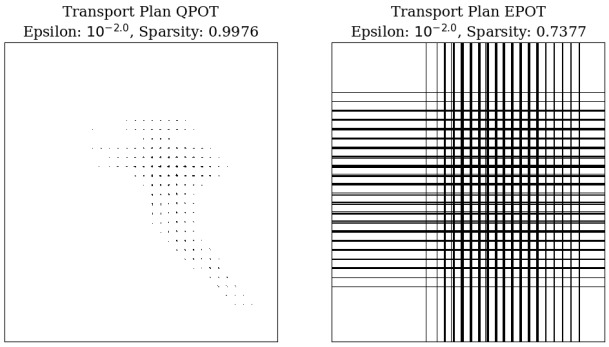
After the color space conversion, the color histogram of n bins of each image is extracted by taking the frequency of each U and V. These color histograms then undergo smoothing and normalization, yielding the total mass of each histogram to $\|\mathbf{r}\|_1 = 1$ and $\|\mathbf{c}\|_1 = 1$. From the histogram, the cost matrix $\mathbf{C}$ is then defined as $C_{i,j} = \|a_i - b_j\|_2^2$ where $a_i \in \{a_1, ..., a_n\}$ and $b_j \in \{b_1, ..., b_n\}$ are the bin values of each histogram. Since we are using 256×256 pictures, setting ε 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 λ. 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 ε. 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 $\varepsilon = 10^{-2}$ of the two methods. These heat maps

are drawn by first setting all the values of the transported plan from 0 to 1e-10 to 0 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 transporting 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.



(a) Sparsity graph $\lambda = 0.7$



(b) Transport plans of QPOT and EPOT $\lambda = 0.7$

Fig. 6. (a) Sparsity, and (b) Transport plan comparison of QPOT and EPOT for Color Transfer with $\lambda = 0.7$

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 Adaption (DA), a family of techniques that handles cases where source and target samples follow different probabilities, had been invented to adapt models trained from the source data onto the target data.

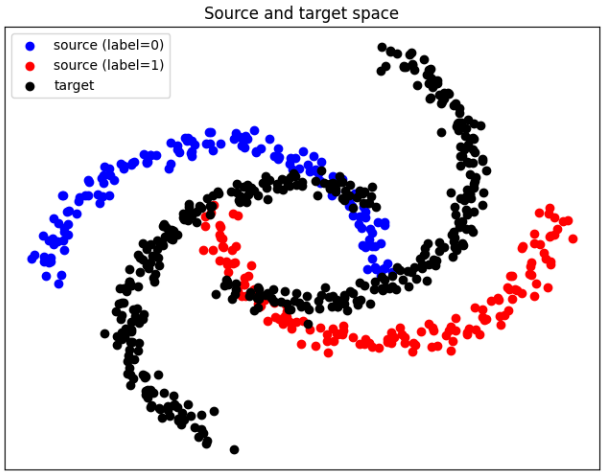


Fig. 7. Moon Dataset for domain adaptation

In our experiment setting, the ”moon” dataset , a widely-used scikit-learn (Pedregosa, 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.

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 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 amount of data points in the target dataset down to 300. The cost matrix is then calculated based on the Euclidian 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 supporting vector machine (SVM) model on the target dataset predicted by our domain adaption 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 ε 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

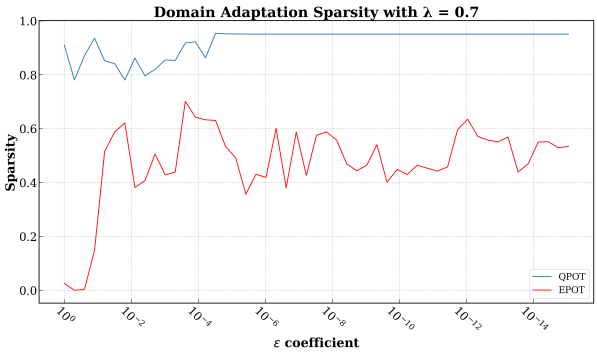
around 0.9, dramatically surpasses that of EPOT, which typically hovers around 0.5 and peaks at around 0.7. This performance of sparsity applies to all λ values in the experiment. The accuracy of the SVM classifier also shows that there is no tradeoff between sparsity and accuracy from QPOT, where the method provides a good accuracy consistently over the ε range. Specifically, the accuracies of both solvers are illustrated in Figure 8 in the case of $\varepsilon = 10^{-4}$ 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.

Conclusions and Future Works

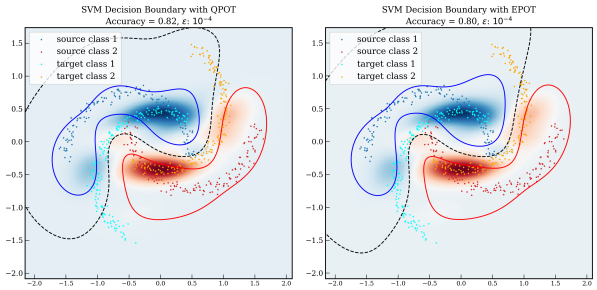
In this paper, we proposed and examined the Quadratic-regularized Partial Optimal Transport formulation (QPOT) and implemented a variety of experiments to show its superiority compared to the classical Entropic-regularized Partial Optimal Transport (EPOT) formulation. In particular, we experimented with both synthetic data, 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.

Overall, QPOT outperforms EPOT in terms of sparsity for all the experiments. For the experiments on toy distributions, QPOT has been shown to induce sparsity in transport plans, even for different regularizer coefficients and transport masses. For the Color Transfer task, the sparsity of EPOT never exceed 0.8, while the the sparsity of QPOT mostly stayed above the 0.95 threshold. Moreover, the heat maps of the experiment also highlight the ability of QPOT to show the general patterns of the solution, which is much better than EPOT. Lastly, on the task of Domain Adaption, QPOT still shows its dominance in terms of sparsity, especially with values of ε lower than 10^{-6}, while still keeping the accuracy roughly the same as EPOT.

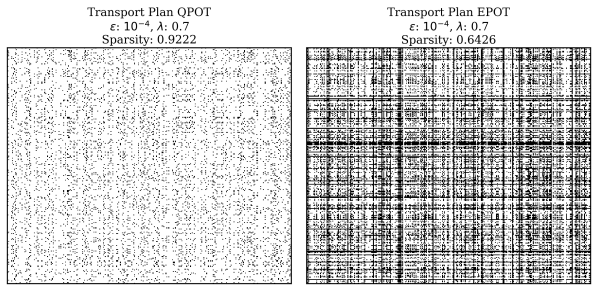
We believe that the new QPOT formulation will lay the foundation for many interesting future works. For example, the sparsity and numerical stability of the QPOT could further enhance the performances in many ML applications. Another approach for future improvements would be to develop accelerated computational methods such as APDAGD (Dvurechensky et al., 2018) to solve QPOT. Additionally, one can adapt Stochastic or Constrained Decentralized Optimization methods (Nguyen et al., 2024b, 2018; Nguyen and Maguluri, 2024; 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, 2023d,c; Nguyen-Vinh et al., 2024) arising in modern distributed settings (Nguyen et al., 2021, 2023b).



(a) Sparsity $\lambda = 0.7$ and $\varepsilon = 10^{-4}$



(b) Accuracy $\lambda = 0.7$



(c) Transport Plans QPOT and EPOT $\lambda = 0.7$ and $\varepsilon = 10^{-4}$

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

Acknowledgment

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

References

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

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. (2018a). Smooth and sparse optimal transport. In Storkey, A. and Perez-Cruz, F., editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 880–889. PMLR.

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

Bousquet, O., Gelly, S., Tolstikhin, I., Simon-Gabriel, C.-J., and Schoelkopf, B. (2017). From optimal transport to generative modeling: the vegan cookbook. *arXiv preprint arXiv:1705.07642*.

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

Carlier, G., Duval, V., Peyré, G., and Schmitzer, B. (2017). Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418.

Chapel, L., Alaya, M. Z., and Gasso, G. (2020). Partial optimal tranport with applications on positive-unlabeled learning. *Advances in Neural Information Processing Systems*, 33:2903–2913.

Clason, C., Lorenz, D. A., Mahler, H., and Wirth, B. (2021). Entropic regularization of continuous optimal transport problems. *Journal of Mathematical Analysis and Applications*, 494(1):124432.

Courty, N., Flamary, R., Tuia, D., and Rakotomamonjy, A. (2017a). Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865.

Courty, N., Flamary, R., Tuia, D., and Rakotomamonjy, A. (2017b). 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. *Advances in neural information processing systems*, 26.

Dinh, M. N. and Nguyen, Q. M. (2020). Measurements of errors in large-scale computational simulations at runtime. In *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 1–7.

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*, pages 1367–1376. PMLR.

Figalli, A. (2010). The optimal partial transport problem. *Archive for Rational Mechanics and Analysis*, 195(2):533–560.

Gan, D., Ge, X., and 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.

Genevay, A., Chizat, L., Bach, F., Cuturi, M., and Peyré, G. (2019). Sample complexity of sinkhorn divergences.

Gramfort, A., Peyré, G., and Cuturi, M. (2015). Fast optimal transport averaging of neuroimaging data. volume 24.

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

Kullback, S. (1951). Kullback-leibler divergence.

Le, K., Nguyen, H., Nguyen, Q., Ho, N., Pham, T., and Bui, H. (2021). On robust optimal transport: Computational complexity, low-rank approximation, and barycenter computation. *CoRR*, abs/2102.06857.

Lin, T., Ho, N., and 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., and Blondel, M. (2022). Sparsity-constrained optimal transport. *arXiv preprint arXiv:2209.15466*.

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

Muzellec, B., Nock, R., Patrini, G., and Nielsen, F. (2016). Tsallis regularized optimal transport and ecological inference. In *AAAI Conference on Artificial Intelligence*.

Nesterov, Y., Gasnikov, A., Guminov, S., and Dvurechensky, P. (2021). Primal–dual accelerated gradient methods with small-dimensional relaxation oracle. *Optimization Methods and Software*, 36(4):773–810.

Nguyen, A. D., Nguyen, T. D., Nguyen, Q. M., Nguyen, H. H., Nguyen, L. M., and Toh, K.-C. (2023a). On partial optimal transport: Revising the infeasibility of sinkhorn and efficient gradient methods.

Nguyen, A. D., Nguyen, T. D., Nguyen, Q. M., Nguyen, H. H., Nguyen, L. M., and Toh, K.-C. (2024a). On partial optimal transport: Revising the infeasibility of sinkhorn and efficient gradient methods. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38,

pages 8090–8098.

Nguyen, H. H., Li, Y., and Zhao, T. (2024b). Stochastic constrained decentralized optimization for machine learning with fewer data oracles: a gradient sliding approach.

Nguyen, H. H. and Maguluri, S. T. (2024). Stochastic approximation for nonlinear discrete stochastic control: Finite-sample bounds.

Nguyen, L. M., Nguyen, P. H., van Dijk, M., Richtárik, P., Scheinberg, K., and Takáč, M. (2018). Sgd and hogwild! convergence without the bounded gradients assumption.

Nguyen, M., Loghin, D., and Dinh, T. T. A. (2021). Understanding the scalability of hyperledger fabric. *ArXiv*, abs/2107.09886.

Nguyen, Q. M., Jeong, H., and Grover, P. (2020). Coded qr decomposition. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 191–196.

Nguyen, Q. M., Khanh Le, N., and Nguyen, L. M. (2023b). Scalable and secure federated xgboost. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.

Nguyen, Q. M., Nguyen, H. H., Zhou, Y., and Nguyen, L. M. (2024c). On unbalanced optimal transport: Gradient methods, sparsity and approximation error.

Nguyen, Q. M., Nguyen, L. M., and Das, S. (2023c). Correlated attention in transformers for multivariate time series.

Nguyen, Q. M., Weissburg, I., and Jeong, H. (2023d). Coded computing for fault-tolerant parallel qr decomposition.

Nguyen-Vinh, K., Vo-Huynh, Q.-N., Nguyen-Minh, K., Hoang, M., and 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*, pages 53–67. Springer Nature Singapore, Singapore.

Nutz, M. (2024). Quadratically regularized optimal transport: Existence and multiplicity of potentials. *arXiv preprint arXiv:2404.06847*.

Oh, G., Sim, B., Chung, H., Sunwoo, L., and Ye, J. C. (2020). Unpaired deep learning for accelerated mri using optimal transport driven cyclegan. *IEEE Transactions on Computational Imaging*, 6:1285–1296.

Pedregosa, F. (2011). Scikit-learn: Machine learning in python fabian. *Journal of machine learning research*, 12:2825.

Pele, O. and Werman, M. (2009). Fast and robust earth mover's distances. *2009 IEEE 12th International Conference on Computer Vision*, pages 460–467.

Peyré, G. and Cuturi, M. (2020). Computational optimal transport.

Pitié, F., Kokaram, A. C., and Dahyot, R. (2007). Automated colour grading using colour distribution transfer. *Comput. Vis. Image Underst.*, 107:123–137.

Pitié, F., Kokaram, A. C., and Dahyot, R. (2007). Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1):123–137. Special issue on color image processing.

Pouli, T. and Reinhard, E. (2011). Extended papers from npar 2010: Progressive color transfer for images of arbitrary dynamic range. *Computers and Graphics*, 35:67–80.

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, pages 365–378. Wiley Online Library.

Rabin, J., Ferradans, S., and Papadakis, N. (2014). Adaptive color transfer with relaxed optimal transport. In *2014 IEEE international conference on image processing (ICIP)*, pages 4852–4856. IEEE.

Rabin, J. and Papadakis, N. (2015). Non-convex relaxation of optimal transport for color transfer between images. In *Geometric Science of Information: Second International Conference, GSI 2015, Palaiseau, France, October 28-30, 2015, Proceedings 2*, pages 87–95. Springer.

Redko, I., Courty, N., Flamary, R., and Tuia, D. (2019). Optimal Transport for Multi-source Domain Adaptation under Target Shift. In *22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019*, volume 89, Naha, Japan.

Rolet, A., Cuturi, M., and Peyré, G. (2016). Fast dictionary learning with a smoothed wasserstein loss. In *Artificial intelligence and statistics*, pages 630–638. PMLR.

Rubner, Y., Tomasi, C., and Guibas, L. (2000). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40:99–121.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.

Solomon, J., De Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., and Guibas, L. (2015). Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (ToG)*, 34(4):1–11.

Villani, C. (2008). Stability of a 4th-order curvature condition arising in optimal transport theory. *Journal of Functional Analysis*, 255(9):2683–2708.

Zhang, G., Yuan, Y., and Sun, D. (2022). An efficient hpr algorithm for the wasserstein barycenter problem with $o(Dim(P)/\varepsilon)$ computational complexity.

Zhang, J., Liu, T., and Tao, D. (2023). An optimal

transport analysis on generalization in deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(6):2842–2853.

Zhang, J., Wan, Y., Chen, Z., and Meng, X. (2019). Non-negative and local sparse coding based on l2-norm and hessian regularization. *Information Sciences*, 486:88–100.

Zhu, Q., Zhang, D., Sun, H., and Li, Z. (2015). Combining l1-norm and l2-norm based sparse representations for face recognition. *Optik*, 126(7-8):719–724.