# A Particle-Evolving Method for Approximating the Optimal Transport Plan


Shu Liu[1(✉)], Haodong Sun[1(✉)], and Hongyuan Zha[2]

[1] Georgia Institute of Technology, Atlanta, USA
{sliu459,hsun310}@gatech.edu
[2] The Chinese University of Hong Kong, Shenzhen, China



**Abstract.** We propose an innovative algorithm that iteratively evolves a particle system to approximate the sample-wised Optimal Transport plan for given continuous probability densities. Our algorithm is proposed via the gradient flow of certain functional derived from the Entropy Transport Problem constrained on probability space, which can be understood as a relaxed Optimal Transport problem. We realize our computation by designing and evolving the corresponding interacting particle system. We present theoretical analysis as well as numerical verifications to our method.



**Keywords:** Optimal Transport · Entropy Transport · Wasserstein gradient flow · Kernel Density Estimation · Interacting particle systems


## 1 Introduction

Optimal transport (OT) provides powerful tools for comparing probability measures in various types. The optimal Transport problem was initially formalized by Gaspard Monge [19] and later generalized by Leonid Kantorovich [12]. Later a series of significant contribution in transportation theory leads to deep connections with more mathematical branches including partial differential equations, geometry, probability theory and statistics [5,12]. Researchers in applied science fields also discover the importance of Optimal Transport. In spite of elegant theoretical results, generally computing Wasserstein distance is not an easy task, especially for the continuous case.

In this paper, instead of solving the standard Optimal Transport (OT) problem, we start with the so-called Entropy Transport (ET) problem, which can be treated as a relaxed OT problem with soft marginal constraints. Recently, the importance of Entropy Transport problem has drawn researchers' attention due to its rich theoretical properties [16]. By restricting ET problem to probability manifold and formulating the gradient flow of the target functional of the Entropy Transport problem, we derive a time-evolution Partial Differential Equation (PDE) that can be then realized by evolving an interacting particle system via Kernel Density Estimation techniques [22].



F. Nielsen and F. Barbaresco (Eds.): GSI 2021, LNCS 12829, pp. 878–887, 2021.
https://doi.org/10.1007/978-3-030-80209-7_94

Our method directly computes for the sample approximation of the optimal coupling to the OT problem between two density functions. This is very different from traditional methods like Linear Programming [20,24,27], Sinkhorn iteration [8], Monge-Ampère Equation [4] or dynamical scheme [3,15,23]; or methods involving neural network optimizations [13,18,25].

Our main contribution is to analyze the theoretical properties of the Entropy Transport problem constrained on probability space and derive its Wasserstein gradient flow. To be specific, we study the existence and uniqueness of the solution to ET problem and further study its $\Gamma$-convergence property to the classical OT problem. Then based on the gradient flow we derive, we propose an innovative particle-evolving algorithm for obtaining the sample approximation of the optimal transport plan. Our method can deal with optimal transport problem between two known densities. As far as we know, despite the classical discretization methods [3,4,15] there is no scalable way to solve this type of problem. We also demonstrate the efficiency of our method by numerical experiments.

## 2 Constrained Entropy Transport as a Regularized Optimal Transport Problem

### 2.1 Optimal Transport Problem and Its Relaxation

In our research, we will mainly focus on Euclidean Space $\mathbb{R}^d$. We denote $\mathcal{P}(E)$ as the probability space defined on the given measurable set $E$. The **Optimal Transport** problem is usually formulated as

$$\inf_{\substack{\gamma\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d),\\ \gamma_1=\mu,\gamma_2=\nu}} \iint c(x,y)\ d\gamma(x,y). \tag{1}$$

Here $\mu,\nu\in\mathcal{P}(\mathbb{R}^d)$, we denote $\gamma_1$ as the marginal distribution of $\gamma$ w.r.t. component $x$ and $\gamma_2$ as the marginal distribution w.r.t. $y$. We call the optimizer of (1)[1] as **Optimal Transport plan** and we denote it as $\gamma_{OT}$.

We can also reformulate (1) as $\min_{\gamma\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)}\left\{\mathcal{E}_\iota(\gamma|\mu,\nu)\right\}$ where

$$\mathcal{E}_\iota(\gamma|\mu,\nu)=\iint c(x,y)d\gamma(x,y)+\int\iota\left(\frac{d\gamma_1}{d\mu}\right)d\mu+\int\iota\left(\frac{d\gamma_2}{d\nu}\right)d\nu \tag{2}$$

Here $\iota$ is defined as $\iota(1)=0$ and $\iota(s)=+\infty$ when $s\neq 1$. We now derive a relaxed version of (1) by replacing $\iota(\cdot)$ with $\Lambda F(\cdot)$, where $\Lambda>0$ is a tunable positive parameter and $F$ is a smooth convex function with $F(1)=0$ and 1 is the unique minimizer. In our research, we mainly focus on $F(s)=s\log s-s+1$ [16]. It is worth mentioning that after such relaxation, the constraint term $\int F(\frac{d\gamma_1}{d\mu})\ d\mu$ is usually called Kullback-Leibler (KL) divergence [14] and is denoted as $D_{\mathrm{KL}}(\gamma_1\|\mu)$.

[1] When $\mu,\nu$ are absolute continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, the optimizer of (1) is guaranteed to be unique.

From now on, we should focus on the following functional involving cost $c(x,y) = h(x-y)$ with $h$ as a strictly convex function, and enforcing the marginal constraints by using KL-divergence:

$$\mathcal{E}_{\Lambda,\mathrm{KL}}(\gamma|\mu,\nu) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\ d\gamma(x,y) + \Lambda D_{\mathrm{KL}}(\gamma_1\|\mu) + \Lambda D_{\mathrm{KL}}(\gamma_2\|\nu). \tag{3}$$

## 2.2 Constrained Entropy Transport Problem and Its Properties

For the following discussions, we always assume that $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ are **absolute continuous** with respect to the Lebesgue measure on $\mathbb{R}^d$. We now focus on solving the following problem

$$\min_{\gamma\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)} \{\mathcal{E}_{\Lambda,\mathrm{KL}}(\gamma|\mu,\nu)\}. \tag{4}$$

A similar problem

$$\min_{\gamma\in\mathcal{M}(\mathbb{R}^d\times\mathbb{R}^d)} \{\mathcal{E}_{\Lambda,\mathrm{KL}}(\gamma|\mu,\nu)\} \tag{5}$$

has been studied in [16] with $\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ being replaced by the space of positive measures $\mathcal{M}(\mathbb{R}^d\times\mathbb{R}^d)$ and is named as **Entropy Transport** problem therein. In our case, since we restrict $\gamma$ to probability space, we call (4) **constrained Entropy Transport problem** and call $\mathcal{E}_{\Lambda,\mathrm{KL}}$ the Entropy Transport functional.

Let us denote $\mathcal{E}_{\min} = \inf_{\gamma\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)}\{\mathcal{E}_{\Lambda,\mathrm{KL}}(\gamma|\mu,\nu)\}$. The following theorem shows the existence of the optimal solution to problem (4). It also describes the relationship between the solution to the constrained ET problem (4) and the solution to the general ET problem (5):

**Theorem 1.** *Suppose $\tilde{\gamma}$ is the solution to original Entropy Transport problem* (5)*. Then we have $\tilde{\gamma} = Z\gamma$, here $Z = e^{-\frac{\mathcal{E}_{\min}}{2\Lambda}}$ and $\gamma \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ is the solution to constrained Entropy Transport problem* (4)*.*

The proof is provided in [17]. The following corollary guarantees the uniqueness of optimal solution to (4):

**Corollary 1.** *The constrained ET problem* (4) *admits a unique optimal solution.*

Despite the discussions on the constrained problem (4) with fixed $\Lambda$, we also establish asymptotic results for (4) with quadratic cost $c(x,y) = |x-y|^2$ as $\Lambda \to +\infty$. For the rest of this section, we define:

$$\mathcal{P}_2(E) = \left\{\gamma\middle|\ \gamma\in\mathcal{P}(E), \gamma \ll \mathscr{L}^d,\ \int_E |x|^2 d\gamma(x) < +\infty\right\} \quad E \ \text{ measurable.}$$

Here we denote $\mathscr{L}^d$ as the Lebesgue measure on $\mathbb{R}^d$.

Let us now consider $\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d)$ and assume it is equipped with the topology of weak convergence. We are able to establish the following $\Gamma$-convergence result.

**Theorem 2.** *Suppose* $c(x, y) = |x-y|^2$*, let us assume* $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $\mu, \nu \ll \mathscr{L}^d$ *and both* $\mu, \nu$ *satisfy the Logarithmic Sobolev inequality with constants* $K_\mu, K_\nu > 0$. *Let* $\{\Lambda_n\}$ *be a positive increasing sequence with* $\lim_{n\to\infty} \Lambda_n = +\infty$.

*We consider the sequence of functionals* $\{\mathcal{E}_{\Lambda_n,\mathrm{KL}}(\cdot|\mu,\nu)\}$. *Recall the functional* $\mathcal{E}_\iota(\cdot|\mu,\nu)$ *defined in* (2). *Then* $\{\mathcal{E}_{\Lambda_n,\mathrm{KL}}(\cdot|\mu,\nu)\}$ $\Gamma$*- converges to* $\mathcal{E}_\iota(\cdot|\mu,\nu)$ *on* $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$.

*Furthermore,* (4) *with functional* $\mathcal{E}_{\Lambda_n,\mathrm{KL}}(\cdot|\mu,\nu)$ *admits a unique optimal solution, let us denote it as* $\gamma_n$. *At the same time, the Optimal Transport problem* (1) *also admits a unique optimal solution, we denote it as* $\gamma_{OT}$. *Then* $\lim_{n\to\infty} \gamma_n = \gamma_{OT}$ *in* $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$.

*Remark 1.* We say a distribution $\mu$ satisfies the Log-Sobolev inequality with $K > 0$ if for any $\tilde{\mu} \ll \mu$, $D_{\mathrm{KL}}(\tilde{\mu}|\mu) \leq \frac{1}{2K}\mathcal{I}(\tilde{\mu}|\mu)$ always holds. Here $\mathcal{I}(\tilde{\mu}|\mu) = \int |\nabla \log\left(\frac{d\tilde{\mu}}{d\mu}\right)|^2 \, d\tilde{\mu}$.

Theorem 2 justifies the asymptotic convergence property of the approximation solutions $\{\gamma_n\}$ to the desired Optimal Transport plan $\gamma_{OT}$. The proof of Theorem 2 is provided in [17].

## 3 Wasserstein Gradient Flow Approach for Solving the Regularized Problem

### 3.1 Wasserstein Gradient Flow

There are already numerous references [2,11,21] regarding Wasserstein gradient flows of functionals defined on the Wasserstein manifold-like structure $(\mathcal{P}_2(\mathbb{R}^d), g^W)$. The Wasserstein manifold-like structure is the manifold $\mathcal{P}_2(\mathbb{R}^d)$ equipped with a special metric $g^W$ compatible to the 2-Wasserstein distance [2,26]. Under this setting, the Wasserstein gradient flow of a certain functional $\mathcal{F}$ can thus be formulated as:

$$\frac{\partial \gamma_t}{\partial t} = -\mathrm{grad}_W \mathcal{F}(\gamma_t) \tag{6}$$

### 3.2 Wasserstein Gradient Flow of Entropy Transport Functional

We now come back to our constrained Entropy Transport problem (4). There are mainly two reasons why we choose to compute the Wasserstein gradient flow of functional $\mathcal{E}_{\Lambda,\mathrm{KL}}(\cdot|\mu,\nu)$:

- Computing the Wasserstein gradient flow is equivalent to applying gradient descent method to determine the minimizer of the ET functional (3);
- In most of the cases, Wasserstein gradient flows can be viewed as a time evolution PDE describing the density evolution of a stochastic process. As a result, once we derived the gradient flow, there will be a natural particle version associated to the gradient flow, which makes the computation of gradient flow tractable.

Now let us compute the Wasserstein gradient flow of $\mathcal{E}_{\Lambda,\mathrm{KL}}(\cdot|\mu,\nu)$:

$$\frac{\partial \gamma_t}{\partial t} = -\mathrm{grad}_W \mathcal{E}_{\Lambda,\mathrm{KL}}(\gamma_t|\mu,\nu), \quad \gamma_t|_{t=0} = \gamma_0 \tag{7}$$

To keep our notations concise, we denote $\rho(\cdot,t) = \frac{d\gamma_t}{d\mathscr{L}^{2d}}$, $\varrho_1 = \frac{d\mu}{d\mathscr{L}^{d}}$, $\varrho_2 = \frac{d\nu}{d\mathscr{L}^{d}}$, we can show that the previous Eq. (7) can be written as:

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\rho\ \nabla(c(x,y) + \Lambda \log(\frac{\rho_1(x,t)}{\varrho_1(x)}) + \Lambda \log(\frac{\rho_2(y,t)}{\varrho_2(y)}))) \tag{8}$$

Here $\rho_1(\cdot,t) = \int \rho(\cdot,y,t)\ dy$ and $\rho_2(\cdot,t) = \int \rho(x,\cdot,t)\ dx$ are density functions of marginals of $\gamma_t$.

### 3.3 Relating the Wasserstein Gradient Flow to a Particle System

Let us treat (8) as certain kind of continuity equation, i.e. we treat $\rho(\cdot,t)$ as the density of the time-evolving random particles. Then the vector field that drives the random particles at time $t$ should be $-\nabla(c(x,y) + \Lambda \log\left(\frac{\rho_1(x,t)}{\varrho_1(x)}\right) + \Lambda \log\left(\frac{\rho_2(y,t)}{\varrho_2(y)}\right))$. This helps us design the following dynamics $\{(X_t,Y_t)\}_{t\geq 0}$: (here $\dot{X}_t$ denotes the time derivative $\frac{dX_t}{dt}$)

$$\begin{cases} \dot{X}_t = -\nabla_x c(X_t,Y_t) + \Lambda(\nabla \log \varrho_1(X_t) - \nabla \log \rho_1(X_t,t)); \\ \dot{Y}_t = -\nabla_y c(X_t,Y_t) + \Lambda(\nabla \log \varrho_2(Y_t) - \nabla \log \rho_2(Y_t,t)); \end{cases} \tag{9}$$

Here $\rho_1(\cdot,t)$ denotes the probability density of random variable $X_t$ and $\rho_2(\cdot,t)$ denotes the density of $Y_t$. If we assume (9) is well-posed, then the density $\rho_t(x,y)$ of $(X_t,Y_t)$ solves the PDE (8).

When we take a closer look at (9), the movement of the particle $(X_t,Y_t)$ at certain time $t$ depends on the probability density $\rho(X_t,Y_t,t)$, which can be approximated by the distribution of the surrounding particles near $(X_t,Y_t)$. Generally speaking, we plan to evolve (9) as a particle aggregation model in order to converge to a sample-wised approximation of the Optimal Transport plan $\gamma_{OT}$ for OT problem (1).

## 4 Algorithmic Development

To simulate the stochastic process (9) with the Euler scheme, we apply the Kernel Density Estimation [22] here to approximate the gradient log function $\nabla \log \rho(x)$ by convolving it with kernel $K(x,\xi)$[2]:

$$\nabla \log \rho(x) \approx \nabla \log(K * \rho)(x) = \frac{(\nabla_x K) * \rho(x)}{K * \rho(x)} \tag{10}$$

[2] In this paper, we choose the Radial Basis Function (RBF) as the kernel: $K(x,\xi) = \exp(-\frac{|x-\xi|^2}{2\tau^2})$.

**Algorithm 1.** Random Batch Particle Evolution Algorithm

**Input:** The density functions of the marginals $\varrho_1, \varrho_2$, timestep $\Delta t$, total number of iterations $T$, parameters of the chosen kernel $K$
**Initialize:** The initial locations of all particles $X_i(0)$ and $Y_i(0)$ where $i = 1, 2, \cdots, n$,
**for** t = 1,2,···,T **do**
  Shuffle the particles and divide them into $m$ batches: $\mathcal{C}_1, \cdots, \mathcal{C}_m$
  **for** each batch $\mathcal{C}_q$ **do**
    Update the location of each particle $(X_i, Y_i)$ $(i \in \mathcal{C}_q)$ according to (11)
  **end for**
**end for**
**Output:** A sample approximation of the optimal coupling: $X_i(T), Y_i(T)$ for $i = 1, 2, \cdots, n$

Here $K * \rho(x) = \int K(x,\xi)\rho(\xi)d\xi$, $(\nabla_x K) * \rho(x) = \int \nabla_x K(x,\xi)\rho(\xi)d\xi$[3]. Such technique is also known as blobing method, which was first studied in [6] and has already been applied to Bayesian sampling [7]. With the blobing method, $\nabla \log \rho(x)$ is evaluated based on the locations of the particles:

$$\frac{\mathbb{E}_{\xi\sim\rho}\nabla_x K(x,\xi)}{\mathbb{E}_{\xi\sim\rho}K(x,\xi)} \approx \frac{\sum_{k=1}^{N}\nabla_x K(x,\xi_k)}{\sum_{k=1}^{N}K(x,\xi_k)} \quad \xi_1, ..., \xi_N, \text{i.i.d.} \sim \rho$$

Now we are able to simulate (9) with the following interacting particle system involving $N$ particles $\{(X_i, Y_i)\}_{i=1,...,N}$. For the $i$-th particle, we have:

$$\begin{cases} \dot{X}_i(t) = -\nabla_x c(X_i(t), Y_i(t)) - \Lambda\left(\nabla V_1(X_i(t)) + \dfrac{\sum_{k=1}^{N}\nabla_x K(X_i(t), X_k(t))}{\sum_{k=1}^{N}K(X_i(t), X_k(t))}\right) \\ \dot{Y}_i(t) = -\nabla_y c(X_i(t), Y_i(t)) - \Lambda\left(\nabla V_2(Y_i(t)) + \dfrac{\sum_{k=1}^{N}\nabla_x K(Y_i(t), Y_k(t))}{\sum_{k=1}^{N}K(Y_i(t), Y_k(t))}\right) \end{cases} \tag{11}$$

Here we denote $V_1 = -\log \varrho_1$, $V_2 = -\log \varrho_2$. Since we only need the gradients of $V_1, V_2$ , our algorithm can deal with unnormalized probability measures. We numerically verify that when $t \to \infty$, the empirical distribution $\frac{1}{N}\sum_{i=1}^{N}\delta_{(X_i(t),Y_i(t))}$ will converge to the optimal distribution $\gamma_{cET}$ that solves (4) with sufficient large $N$ and $\Lambda$, while the rigorous proof is reserved for our future work. The algorithm scheme is summarized in the algorithm 1.

*Remark 2.* Inspired by [10], we apply the Random Batch Methods (RBM) here to reduce the computational effort required to approximate $\nabla \log \rho(x)$ with blobing method: We divide all $N$ particles into $m$ batches equally, and we only consider the particles in the same batch as the particle $X_i$ when we evaluate the $\nabla \log \rho(X_i)$. Now in each time step, the computational cost is reduced from $\mathcal{O}(n^2)$ to $\mathcal{O}(n^2/m)$.

[3] Notice that we always use $\nabla_x K$ to denote the partial derivative of $K$ with respect to the first components.

## 5 Numerical Experiments

In this section, we test our algorithm on several toy examples.

**Gaussian.** We first apply the algorithm to compute the sample approximation of the Optimal Transport plan between two 1D Gaussian distributions $\mathcal{N}(-4,1), \mathcal{N}(6,1)$[4]. We set $\lambda = 200, \Delta t = 0.001, c(x,y) = |x-y|^2$ and run it with 1000 particles $(X_i, Y_i)$'s for 1000 iterations. We initialize the particles by drawing 1000 i.i.d. sample points from $\mathcal{N}(-20,4)$ as $X_i$'s and 1000 i.i.d. sample points from $\mathcal{N}(20,2)$ as $Y_i$'s. The empirical results are shown in Fig. 1 and Fig. 2.

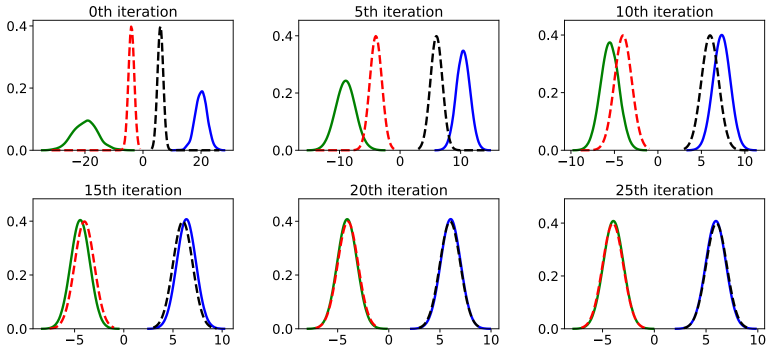


**Fig. 1.** Marginal plot for 1D Gaussian example. The red and black dashed lines correspond to two marginal distribution respectively and the solid blue and green lines are the kernel estimated density functions of particles at certain iterations. After first 25 iterations, the particles have matched the marginal distributions very well. (Color figure online)

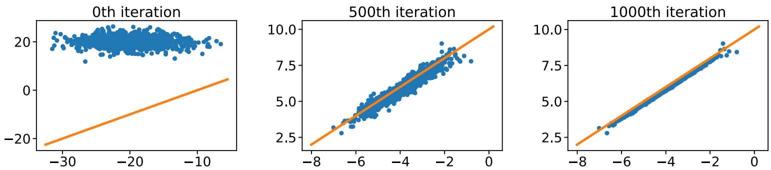


**Fig. 2.** The sample approximation for 1D Gaussian example. The orange dash line corresponds to the Optimal Transport map $T(x) = x + 10$. (Color figure online)

**Gaussian Mixture.** Then we apply the algorithm to two 1D Gaussian mixture $\varrho_1 = \frac{1}{2}\mathcal{N}(-1,1) + \frac{1}{2}\mathcal{N}(1,1), \varrho_2 = \frac{1}{2}\mathcal{N}(-2,1) + \frac{1}{2}\mathcal{N}(2,1)$. For experiment, we set $\lambda = 60, \Delta t = 0.0004, c(x,y) = |x-y|^2$ and run it with 1000 particles $(X_i, Y_i)$'s for 5000 iterations. We initialize the particles by drawing 2000 i.i.d. sample points from $\mathcal{N}(0,2)$ as $X_i$'s and $Y_i$'s. Figure 3 gives a visualization of the marginal distributions and the Optimal Transport map.

[4] Here $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean value $\mu$ and variace $\sigma^2$.

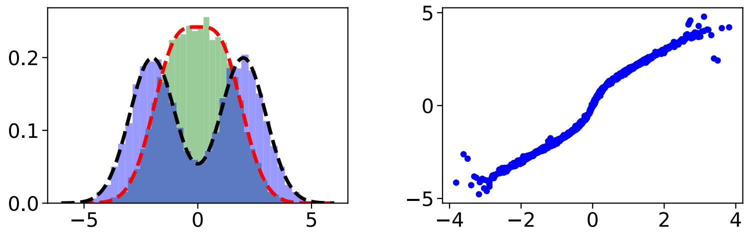


**Fig. 3.** 1D Gaussian mixture. Left. Marginal plot. The dash lines correspond to two marginal distributions. The histogram indicates the distribution of particles after 5000 iterations. Right. Sample approximation for the optimal coupling.

**Wasserstein Barycenters.** Our framework can be easily extended to solve the Wasserstein barycenter problem [1,9]

$$\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \sum_{i=1}^{m} \lambda_i W_2^2(\mu, \mu_i) \tag{12}$$

where $\lambda_i > 0$ are the weights. Similar to our previous treatment on OT problem, we can relax the marginal constraints in (12) and consider

$$\min_{\gamma \in \mathcal{P}(\mathbb{R}^{(m+1)d})} \int_{\mathbb{R}^{(m+1)d}} \sum_{j=1}^{m} \lambda_j |x - x_j|^2 \, d\gamma(x, x_1, ..., x_m) + \sum_{j=1}^{m} \Lambda_j D_{\mathrm{KL}}(\gamma_j \| \mu_j) \tag{13}$$

Then we can apply similar particle-evolving method to compute for problem (13), which can be treated as an approximation of the original barycenter problem (12). Here is an illustrative example: Given two Gaussian distributions $\rho_1 = \mathcal{N}(-10, 1), \rho_2 = \mathcal{N}(10, 1)$, and cost function $c(x, x_1, x_2) = w_1|x - x_1|^2 + w_2|x - x_2|^2$, we can compute sample approximation of the barycenter $\bar{\rho}$ of $\rho_1, \rho_2$. We try different weights $[w_1, w_2] = [0.25, 0.75], [0.5, 0.5], [0.75, 0.25]$ to test our algorithm. The experimental results are shown in Fig. 4. The distribution of the particles corresponding to the barycenter random variable $X_0$ converges to $\mathcal{N}(5, 1), \mathcal{N}(0, 1), \mathcal{N}(-5, 1)$ successfully after 2000 iterations.

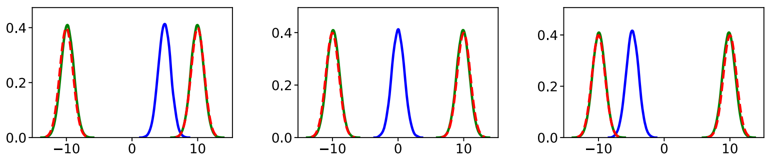


**Fig. 4.** Density plots for 1D Wasserstein barycenter example. The red dashed lines correspond to two marginal distributions respectively and the solid green lines are the kernel estimated density functions of the particles $X_1$'s and $X_2$'s. The solid blue line represents the kernel estimated density function of the particles corresponding to the barycenter. $[w1, w2] = [0.25, 0.75], [0.5, 0.5], [0.75, 0.25]$ from left to right. (Color figure online)

## 6 Conclusion

We propose the constrained Entropy Transport problem (4) and study some of its theoretical properties. We discover that the optimal distribution of (4) can be treated as an approximation to the optimal plan of the Optimal Transport problem (1) in the sense of $\Gamma$-convergence. We also construct the Wasserstein gradient flow of the Entropy Transport functional. Based on that, we propose a novel algorithm that computes for the sample-wised optimal transport plan by evolving an interacting particle system. We demonstrate the effectiveness of our method by several illustrative examples. More theoretical analysis and numerical experiments on higher dimensional cases including comparisons with other methods will be included in our future work.

## References

1. Agueh, M., Carlier, G.: Barycenters in the Wasserstein space. SIAM J. Math. Anal. **43**(2), 904–924 (2011)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows: In Metric Spaces and in the Space of Probability Measures. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-7643-8722-8
3. Benamou, J.D., Brenier, Y.: A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000)
4. Benamou, J.D., Froese, B.D., Oberman, A.M.: Numerical solution of the optimal transportation problem using the Monge-Ampère equation. J. Comput. Phys. **260**, 107–126 (2014)
5. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math. **44**(4), 375–417 (1991)
6. Carrillo, J.A., Craig, K., Patacchini, F.S.: A blob method for diffusion. Calc. Var. Partial. Differ. Equ. **58**(2), 53 (2019)
7. Chen, C., Zhang, R., Wang, W., Li, B., Chen, L.: A unified particle-optimization framework for scalable Bayesian sampling. arXiv preprint arXiv:1805.11659 (2018)
8. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems, pp. 2292–2300 (2013)
9. Cuturi, M., Doucet, A.: Fast computation of Wasserstein barycenters. In: International Conference on Machine Learning, pp. 685–693. PMLR (2014)
10. Jin, S., Li, L., Liu, J.G.: Random batch methods (RBM) for interacting particle systems. J. Comput. Phys. **400**, 108877 (2020)
11. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
12. Kantorovich, L.: On translation of mass (in Russian), c r. Doklady. Acad. Sci. USSR **37**, 199–201 (1942)
13. Korotin, A., Egiazarian, V., Asadulaev, A., Safin, A., Burnaev, E.: Wasserstein-2 generative networks. arXiv preprint arXiv:1909.13082 (2019)
14. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**(1), 79–86 (1951)
15. Li, W., Ryu, E.K., Osher, S., Yin, W., Gangbo, W.: A parallel method for earth mover's distance. J. Sci. Comput. **75**(1), 182–197 (2018)

16. Liero, M., Mielke, A., Savaré, G.: Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. Invent. Math. **211**(3), 969–1117 (2018)
17. Liu, S., Sun, H., Zha, H.: Approximating the optimal transport plan via particle-evolving method. arXiv preprint arXiv:2105.06088 (2021)
18. Makkuva, A., Taghvaei, A., Oh, S., Lee, J.: Optimal transport mapping via input convex neural networks. In: International Conference on Machine Learning, pp. 6672–6681. PMLR (2020)
19. Monge, G.: Mémoire sur la théorie des déblais et des remblais. Histoire de l'Académie Royale des Sciences de Paris (1781)
20. Oberman, A.M., Ruan, Y.: An efficient linear programming method for optimal transportation. arXiv preprint arXiv:1509.03668 (2015)
21. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Comm. Partial Differ. Equ. **26**(1–2), 101–174 (2001)
22. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. **33**(3), 1065–1076 (1962)
23. Ruthotto, L., Osher, S.J., Li, W., Nurbekyan, L., Fung, S.W.: A machine learning framework for solving high-dimensional mean field game and mean field control problems. Proc. Natl. Acad. Sci. **117**(17), 9183–9193 (2020)
24. Schmitzer, B.: A sparse multiscale algorithm for dense optimal transport. J. Math. Imaging Vis. **56**(2), 238–259 (2016)
25. Seguy, V., Damodaran, B.B., Flamary, R., Courty, N., Rolet, A., Blondel, M.: Large-scale optimal transport and mapping estimation. arXiv preprint arXiv:1711.02283 (2017)
26. Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-71050-9
27. Walsh III, J.D., Dieci, L.: General auction method for real-valued optimal transport. arXiv preprint arXiv:1705.06379 (2017)