Generalized unnormalized optimal transport and its fast algorithms ☆

Wonjun Lee [c], Rongjie Lai [a], Wuchen Li [b,*], Stanley Osher [c]

[a] *Department of Mathematics, Rensselaer Polytechnic Institute, United States of America*
[b] *Department of Mathematics, University of South Carolina, United States of America*
[c] *Department of Mathematics, University of California, Los Angeles, United States of America*

ARTICLE INFO

ABSTRACT

We introduce fast algorithms for generalized unnormalized optimal transport. To handle densities with different total mass, we consider a dynamic model, which mixes the L^p optimal transport with L^p distance. For $p = 1$, we derive the corresponding L^1 generalized unnormalized Kantorovich formula. We further show that the problem becomes a simple L^1 minimization which is solved efficiently by a primal-dual algorithm. For $p = 2$, we derive the L^2 generalized unnormalized Kantorovich formula, a new unnormalized Monge problem and the corresponding Monge-Ampère equation. Furthermore, we introduce a new unconstrained optimization formulation of the problem. The associated gradient flow is essentially related to an elliptic equation which can be solved efficiently. Here the proposed gradient descent procedure together with the Nesterov acceleration involves the Hamilton-Jacobi equation arising from the KKT conditions. Several numerical examples are presented to illustrate the effectiveness of the proposed algorithms.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

Optimal transport describes transport plans and metrics between two densities with equal total mass [28]. It has wide applications in various fields such as physics [14,17], mean field games [10], image processing [23], economics [2], inverse problem [11,29], Kalman filter [13] as well as machine learning [1,19]. In practice, it is also natural to consider transport and metrics between two densities with different total mass. For example, in image processing, it is very common that we need to compare and process images with unequal total intensities [26].

Recently, there has been increasing interests in studying the optimal transport between two densities with different total mass. Based on the linear programming formulation, generalized versions for unnormalized optimal transport have been considered in [25,27]. In this paper, our discussion is based on the fluid-dynamic formulation following [3], which has significantly fewer variables than the linear programming formulation. We consider a source function to provide dynamical behaviors of a source term during transportation. Adding a source term for handling densities with unequal total mass has been considered in [5,7–9,18,20,24]. These methods consider density-dependent source terms and lead to a dynamical mixture of Wasserstein-2 distance and Fisher-Rao distance. The corresponding minimization of the source term is weighted

* Corresponding author.
E-mail addresses: wlee@math.ucla.edu (W. Lee), lair@rpi.edu (R. Lai), wuchen@mailbox.sc.edu (W. Li), sjo@math.ucla.edu (S. Osher).

with the density. More recently, a spatially independent source function was considered in [12] to transport densities with unequal mass. This model results in creating or removing masses in the space uniformly during transportation when moving one density to another. Here, we further extend the model [12] using a spatially dependent source function. As a result, the transportation map between two densities with different masses has the flexibility to create or remove masses locally. In all our models, the source term does not depend on the current density. This property keeps the Hamilton-Jacobi equation arising in the original (normalized) optimal transport problem. We further explore the Kantorovich duality and derive the corresponding unnormalized Monge problems and Monge-Ampère equations. Besides these model derivations, the other main contribution of this paper is to propose fast algorithms for all related dynamical optimal transport problems with source terms.

More specifically, the proposed model is a minimal flux problem mixing both L^p metric and Wasserstein-p metric, following Benamou-Brenier formula [3]:

$$\inf_{\boldsymbol{v},\mu,f} \int_0^1 \int_\Omega \|\boldsymbol{v}(t,x)\|^p \mu(t,x)\mathrm{d}x\mathrm{d}t + \frac{1}{\alpha}\int_0^1 \int_\Omega |f(t,x)|^p\mathrm{d}x\mathrm{d}t,$$

such that

$$\partial_t\mu(t,x) + \nabla\cdot(\mu(t,x)\boldsymbol{v}(t,x)) = f(t,x), \quad \mu(0,x) = \mu_0(x), \quad \mu(1,x) = \mu_1(x).$$

The minimization problem solves for the optimal map between two nonnegative densities μ_0 and μ_1, given a source function f (see the details in Definition 1 from section 2). The optimal map shows how the masses are added or removed by the source function during the transportation. In this paper, in particular, we focus on the cases $p = 1$ and $p = 2$, and design corresponding fast algorithms. For the L^1 case, we propose a primal-dual algorithm [4]. The method updates variables at each iteration with explicit formulas, which only involve low computational cost shrink operators, such as those used in [16]. For the L^2 case, we formulate the minimal flux problem into a novel unconstrained minimization problem as follows

$$\inf_\mu\left\{\int_0^1 \int_\Omega \partial_t\mu(t,x)(-\nabla\cdot(\mu(t,x)\nabla) + \alpha\mathrm{Id})^{-1}\partial_t\mu(t,x)\mathrm{d}x\mathrm{d}t : \right. \tag{1}$$
$$\left. \mu(0,x) = \mu_0(x), \mu(1,x) = \mu_1(x), x \in \Omega\right\},$$

where α is a given positive scalar, Id is the identity operator, and the infimum is taken among all density paths $\mu(t,x)$ with fixed terminal densities μ_0, μ_1. From the associated Euler-Lagrange equation, we derive a Nesterov accelerated gradient descent method to solve the unnormalized optimal transport problem. It turns out that our method only needs to solve an elliptic equation involving the density at each iteration. Thus, fast solvers for elliptic equations can be directly used. Interestingly, the Euler-Lagrange equation of this formulation introduces the Hamilton-Jacobi equation, which characterizes the Lagrange multiplier (see related studies in [15]). We, in fact, construct the gradient descent method in the density path space to solve this equation:

$$\partial_\tau\mu(\tau,t,x) = \partial_t\Phi(\tau,t,x) + \frac{1}{2}\|\nabla\Phi(\tau,t,x)\|^2,$$

with

$$\Phi(\tau,t,x) = (-\nabla\cdot(\mu(\tau,t,x)\nabla) + \alpha Id)^{-1}\partial_t\mu(\tau,t,x).$$

Here τ is an artificial time variable in optimization. The minimizer path $\mu^*(t,x)$ is obtained by solving $\mu^*(t,x) = \lim_{\tau\to\infty}\mu(\tau,t,x)$ numerically.

The outline of this paper is as follows. In section 2, we propose a formulation for the generalized unnormalized optimal transport. We then derive the Kantorovich duality for both cases. We also formulate the generalized unnormalized Monge problem and the corresponding Monge-Ampère equation. In section 3, we propose a fast algorithm for L^1-generalized unnormalized optimal transport using a primal-dual based method. We also propose a new method for L^2-generalized unnormalized optimal transport based on the Nesterov accelerated gradient descent method. In addition, we discuss detailed numerical discretization of the two problems. In section 4, we present several numerical experiments to demonstrate the effectiveness of our algorithms. We conclude the paper in section 5.

2. Generalized unnormalized optimal transport

In this section, we study a formulation of generalized unnormalized optimal transport problem as a natural extension of the exploration studied in [12]. We specifically discuss the L^1 and L^2 versions of the generalized unnormalized optimal transport and their associated Kantorovich dualities. Furthermore, we derive a new generalized unnormalized Monge problem and the corresponding Monge-Ampère equation.

Let $\Omega \subset \mathbb{R}^d$ be a compact convex domain. Denote the space of unnormalized densities $\mathcal{M}(\Omega)$ by

$$\mathcal{M}(\Omega) := \{\mu \in L^1(\Omega) : \mu(x) \geq 0\}.$$

Given two densities $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$, we define the generalized unnormalized optimal transport as follows:

Definition 1 *(Generalized unnormalized optimal transport).* Define the L^p generalized unnormalized Wasserstein distance $UW_p : \mathcal{M}(\Omega) \times \mathcal{M}(\Omega) \to \mathbb{R}$ by

$$UW_p(\mu_0, \mu_1)^p = \inf_{\boldsymbol{v}, \mu, f} \int_0^1 \int_\Omega \|\boldsymbol{v}(t,x)\|^p \mu(t,x) \mathrm{d}x \mathrm{d}t + \frac{1}{\alpha} \int_0^1 \int_\Omega |f(t,x)|^p \mathrm{d}x \mathrm{d}t,$$

such that the dynamical constraint, i.e. the unnormalized continuity equation, holds

$$\partial_t \mu(t,x) + \nabla \cdot (\mu(t,x)\boldsymbol{v}(t,x)) = f(t,x), \quad \mu(0,x) = \mu_0(x), \quad \mu(1,x) = \mu_1(x).$$

The infimum is taken over continuous unnormalized density functions $\mu : [0,1] \times \Omega \to \mathbb{R}$, and Borel vector fields $v : [0,1] \times \Omega \to \mathbb{R}^d$ with zero flux condition on $[0,1] \times \partial\Omega$, and Borel spatially dependent source functions $f : [0,1] \times \Omega \to \mathbb{R}$. A positive constant $\alpha \in (0, \infty)$ is a fixed parameter.

This is a generalized definition of unnormalized optimal transport from [12]. Here, we consider a spatially dependent source function $f(t,x)$. In this paper, we will focus on the cases with $p = 1$ and $p = 2$.

Remark 1. We note that [7] has proposed the model for $p = 2$ without any discussion about numerical methods. In this paper, we mainly study Kantorovich duality and design fast algorithms.

Remark 2. In literature, [8] studied the other dynamical formulations of unbalanced optimal transport problems. In their approach, the optimal source term is expressed as a product of a density function and a scalar field function. In our approach, the optimal source term only depends on a scalar field function. This fact shows that our approach is different from [8] in variational problems and dual (Kantorovich) problems.

2.1. L^1 generalized unnormalized Wasserstein metric

When $p = 1$, the problem (1) becomes

$$UW_1(\mu_0, \mu_1) = \inf_{\boldsymbol{v}, \mu, f} \left\{ \int_0^1 \int_\Omega \|\boldsymbol{v}(t,x)\| \mu(t,x) \mathrm{d}x \mathrm{d}t + \frac{1}{\alpha} \int_0^1 \int_\Omega |f(t,x)| \mathrm{d}x \mathrm{d}t : \right.$$

$$\partial_t \mu(t,x) + \nabla \cdot (\mu(t,x)\boldsymbol{v}(t,x)) = f(t,x) \qquad (2)$$

$$\left. \mu(0,x) = \mu_0(x), \quad \mu(1,x) = \mu_1(x) \right\}.$$

Here $\|\cdot\|$ can be any homogeneous of degree one norm, i.e. l_q norm $\|u\|_q = (\sum_{i=1}^d |u_i|^q)^{\frac{1}{q}}$. In particular, we consider $q = 1, 2$ with

$$\|u\|_1 = |u_1| + \cdots + |u_d| \qquad \text{for } u \in \mathbb{R}^d,$$

or

$$\|u\|_2 = \sqrt{|u_1|^2 + \cdots + |u_d|^2} \qquad \text{for } u \in \mathbb{R}^d.$$

Proposition 2. *The L^1 unnormalized Wasserstein metric is given by*

$$UW_1(\mu_0, \mu_1) = \inf_{\boldsymbol{m}, c} \left\{ \int_\Omega \|\boldsymbol{m}(x)\| \mathrm{d}x + \frac{1}{\alpha} \int_\Omega |c(x)| \mathrm{d}x : \right.$$

$$\left. \mu_1(x) - \mu_0(x) + \nabla \cdot \boldsymbol{m}(x) - c(x) = 0 \right\}. \qquad (3)$$

There exists $\Phi(x)$, such that the minimizer (m, c) for the problem ([3](#)) satisfies

$$\nabla \Phi(x) \in \partial \|\boldsymbol{m}(x)\| \quad and \quad \alpha \Phi(x) \in \partial |c(x)|$$

where $\partial \|\boldsymbol{m}(x)\|$ and $\partial |c(x)|$ denote their sub-differentials.

Proof. Denote

$$\boldsymbol{m}(x) = \int\limits_0^1 \boldsymbol{v}(t, x) \mu(t, x) \mathrm{d}t.$$

Using Jensen's inequality and integration by parts, we can reformulate ([2](#)).

$$\begin{aligned}
&\int\limits_0^1 \int\limits_\Omega \|\boldsymbol{v}(t, x)\| \mu(t, x) \mathrm{d}x \mathrm{d}t + \frac{1}{\alpha} \int\limits_0^1 \int\limits_\Omega |f(t, x)| \mathrm{d}x \mathrm{d}t \\
&\geq \int\limits_\Omega \|\boldsymbol{m}(x)\| \mathrm{d}x + \frac{1}{\alpha} \int\limits_\Omega \left| \int\limits_0^1 f(t, x) \mathrm{d}t \right| \mathrm{d}x.
\end{aligned} \tag{4}$$

Define $c(x) = \int_0^1 f(t, x) \mathrm{d}t$. Integrating on the constraint of problem ([2](#)) with the zero flux condition of v yields,

$$\int\limits_\Omega c(x) \mathrm{d}x = \int\limits_0^1 \int\limits_\Omega f(t, x) \mathrm{d}x \mathrm{d}t = \int\limits_\Omega \mu_1(x) \mathrm{d}x - \int\limits_\Omega \mu_0(x) \mathrm{d}x.$$

Plug $c(x)$ into the equation ([4](#)), we obtain a new formulation.

$$\inf_{\boldsymbol{m}, c} \left\{ \int\limits_\Omega \|\boldsymbol{m}(x)\| \mathrm{d}x + \frac{1}{\alpha} \int\limits_\Omega \|c(x)\| \mathrm{d}x \ : \ \mu_1(x) - \mu_0(x) + \nabla \cdot \boldsymbol{m}(x) - c(x) = 0 \right\}.$$

Note that the minimization path can be attained in the inequality ([4](#)) by choosing $\mu(t, x) = t \mu_0(x) + (1 - t) \mu_1(x)$, $\boldsymbol{m}(x) = \mu(t, x) \boldsymbol{v}(t, x)$ and $f(t, x) = c(x)$. Then $\{\mu(t, x), \boldsymbol{v}(t, x), f(t, x)\}$ is a feasible solution to ([2](#)) and ([3](#)), hence the two minimization problems have the same optimal value.

Consider the Lagrangian of this minimization problem.

$$\mathcal{L}(\boldsymbol{m}, c, \Phi) = \int\limits_\Omega \|\boldsymbol{m}(x)\| \mathrm{d}x + \frac{1}{\alpha} \int\limits_\Omega |c(x)| \mathrm{d}x + \int\limits_\Omega \Phi(x) \left(\mu_1(x) - \mu_0(x) + \nabla \cdot \boldsymbol{m}(x) - c(x) \right), \tag{5}$$

where $\Phi(x)$ is a Lagrange multiplier. From the Karush–Kuhn–Tucker (KKT) conditions, we derive the following properties of the minimizer

$$0 \in \partial_{\boldsymbol{m}} \mathcal{L} \Rightarrow \nabla \Phi(x) \in \partial \|\boldsymbol{m}(x)\|$$

$$0 \in \partial_c \mathcal{L} \Rightarrow \alpha \Phi(x) \in \partial |c(x)|$$

$$\delta_\Phi \mathcal{L} = 0 \Rightarrow \mu_1(x) - \mu_0(x) + \nabla \cdot \boldsymbol{m}(x) - c(x) = 0. \quad \square$$

Remark 3. In the case that L^1 unnormalized Wasserstein metric with a spatially independent function $f(t)$, c is defined to be $c = \int_0^1 f(t) \mathrm{d}t$, which is a constant. Integrating on a spatial domain for continuity equation,

$$c = \frac{1}{|\Omega|} \left(\int\limits_\Omega \mu_0(x) \mathrm{d}x - \int\limits_\Omega \mu_0(x) \mathrm{d}x \right).$$

As a result, the minimization problem becomes

$$\begin{aligned}
UW_1(\mu_0, \mu_1) = \inf_{\boldsymbol{m}} \Bigg\{ &\int\limits_\Omega \|\boldsymbol{m}(x)\| \mathrm{d}x + \frac{1}{\alpha} \left| \int\limits_\Omega \mu_1(x) \mathrm{d}x - \int\limits_\Omega \mu_0(x) \mathrm{d}x \right| : \\
&\mu_1(x) - \mu_0(x) + \nabla \cdot \boldsymbol{m}(x) = \frac{1}{|\Omega|} \left(\int\limits_\Omega \mu_1(x) \mathrm{d}x - \int\limits_\Omega \mu_0(x) \mathrm{d}x \right) \Bigg\}.
\end{aligned}$$

This is compatible with the result obtained in [12]. In this case, we note that $\boldsymbol{m}(x)$ does not depend on α.

Proposition 3 (L^1 *generalized unnormalized Kantorovich formulation*). *The Kantorovich formulation of L^1 unnormalized Wasserstein metric is the following:*

$$UW_1(\mu_0, \mu_1) = \sup_{\Phi} \left\{ \int_{\Omega} \Phi(x)(\mu_1(x) - \mu_0(x))dx : \|\nabla\Phi\| \leq 1, |\Phi| \leq \frac{1}{\alpha} \right\} \tag{6}$$

Remark 4. The Kantorovich formulation of the generalized unnormalized Wasserstein-1 metric has also been stated in [6] for the $\|\cdot\|_2$ norm.

Proof. From the Lagrangian (5),

$$\inf_{\boldsymbol{m},c} \sup_{\Phi} \mathcal{L}(\boldsymbol{m}, c, \Phi)$$

$$\geq \sup_{\Phi} \inf_{\boldsymbol{m},c} \mathcal{L}(\boldsymbol{m}, c, \Phi)$$

$$= \sup_{\Phi} \inf_{\boldsymbol{m},c} \left\{ \int_{\Omega} \|\boldsymbol{m}(x)\|dx + \frac{1}{\alpha} \int_{\Omega} |c(x)|dx + \int_{\Omega} \Phi(x)(\mu_1(x) - \mu_0(x) + \nabla \cdot \boldsymbol{m}(x) - c(x))dx \right\}$$

$$= \sup_{\Phi} \inf_{\boldsymbol{m},c} \left\{ \int_{\Omega} \|\boldsymbol{m}(x)\|dx + \frac{1}{\alpha} \int_{\Omega} |c(x)|dx + \int_{\Omega} \Phi(x)(\mu_1(x) - \mu_0(x) - c(x))dx \right.$$

$$\left. - \int_{\Omega} \nabla\Phi(x) \cdot \boldsymbol{m}(x)dx + \int_{\partial\Omega} \Phi(x)\boldsymbol{m}(x) \cdot n(x)ds(x) \right\}$$

$$= \sup_{\Phi} \left\{ \int_{\Omega} \Phi(x)(\mu_1(x) - \mu_0(x)) + \inf_{\boldsymbol{m},c} \int_{\Omega} \|\boldsymbol{m}(x)\| - \nabla\Phi(x) \cdot \boldsymbol{m}(x)dx + \int_{\Omega} \frac{1}{\alpha}|c(x)| - \Phi(x)c(x)dx \right\}$$

$$= \sup_{\Phi} \left\{ \int_{\Omega} \Phi(x)(\mu_1(x) - \mu_0(x))dx : \|\nabla\Phi\| \leq 1, |\Phi| \leq \frac{1}{\alpha} \right\}.$$

From the calculation, the optimizer Φ satisfies the following:

$$\nabla\Phi \in \partial\|\boldsymbol{m}(x)\|, \qquad \alpha\Phi \in \partial|c(x)|.$$

We show the duality gap is zero using the Proposition 2.

$$\int_{\Omega} \|\boldsymbol{m}(x)\|dx + \frac{1}{\alpha} \int_{\Omega} |c(x)|dx + \int_{\Omega} \Phi(x)(\mu_1(x) - \mu_0(x) + \nabla \cdot \boldsymbol{m}(x) - c(x))dx$$

$$= \int_{\Omega} \|\boldsymbol{m}(x)\| - \nabla\Phi \cdot \boldsymbol{m}(x)dx + \int_{\Omega} \frac{1}{\alpha}|c(x)| - \Phi(x)c(x)dx + \int_{\Omega} \Phi(x)(\mu_1(x) - \mu_0(x))dx$$

$$= \int_{\Omega} \Phi(x)(\mu_1(x) - \mu_0(x))dx$$

This concludes the proof. □

2.2. L^2 generalized unnormalized Wasserstein metric

Let $p = 2$. From the definition (1), we now consider

$$UW_2(\mu_0, \mu_1)^2 = \inf_{v,\mu,f} \left\{ \int_0^1 \int_{\Omega} \|\boldsymbol{v}(t,x)\|^2 \mu(t,x)dxdt + \frac{1}{\alpha} \int_0^1 \int_{\Omega} \|f(t,x)\|^2 dxdt : \right.$$

$$\partial_t \mu(t,x) + \nabla \cdot (\mu(t,x)\boldsymbol{v}(t,x)) = f(t,x), t \in [0,1], x \in \Omega,$$

$$\left. \mu(0,x) = \mu_0(x), \mu(1,x) = \mu_1(x) \right\}. \tag{7}$$

Proposition 4. *The L^2 generalized unnormalized Wasserstein metric is a well-defined metric function in $M(\Omega)$. In addition, the minimizer $(\boldsymbol{v}(t,x), \mu(t,x), f(t,x))$ for (7) satisfies*

$$\boldsymbol{v}(t,x) = \nabla \Phi(t,x), \qquad f(t,x) = \alpha \Phi(t,x),$$

and

$$\partial_t \mu(t,x) + \nabla \cdot (\mu(t,x)\nabla\Phi(t,x)) = \alpha\Phi(t,x)$$

$$\partial_t \Phi(t,x) + \frac{1}{2}\|\nabla\Phi(t,x)\|^2 \leq 0.$$

In particular, if $\mu(t,x) > 0$, then

$$\partial_t \Phi(t,x) + \frac{1}{2}\|\nabla\Phi(t,x)\|^2 = 0.$$

Proof. Denote $\boldsymbol{m}(t,x) = \mu(t,x)\boldsymbol{v}(t,x)$. Then the problem becomes

$$\frac{1}{2}UW_2(\mu_0,\mu_1)^2 = \inf_{\boldsymbol{m},\mu,f}\left\{ \int_0^1\int_\Omega \frac{\|\boldsymbol{m}(t,x)\|^2}{2\mu(t,x)}\mathrm{d}x\mathrm{d}t + \frac{1}{2\alpha}\int_0^1\int_\Omega |f(t,x)|^2\mathrm{d}x\mathrm{d}t : \right.$$
$$\partial_t\mu(t,x) + \nabla\cdot\boldsymbol{m}(t,x) = f(t,x),$$
$$\left. \mu(0,x) = \mu_0(x), \mu(1,x) = \mu_1(x), x \in \Omega, 0 \leq t \leq 1 \right\}. \tag{8}$$

Denote $\Phi(t,x)$ as a Lagrange multiplier. Consider the Lagrangian

$$\mathcal{L}(\boldsymbol{m},\mu,f,\Phi) = \int_0^1\int_\Omega \frac{\|\boldsymbol{m}(t,x)\|^2}{2\mu(t,x)}\mathrm{d}x\mathrm{d}t + \frac{1}{2\alpha}\int_0^1\int_\Omega |f(t,x)|^2\mathrm{d}x\mathrm{d}t$$
$$+ \int_0^1\int_\Omega \Phi(t,x)\Big(\partial_t\mu(t,x) + \nabla\cdot\boldsymbol{m}(t,x) - f(t,x)\Big)\mathrm{d}x\mathrm{d}t.$$

From KKT condition $\delta_{\boldsymbol{m}}\mathcal{L} = 0, \delta_\mu\mathcal{L} \geq 0, \delta_f\mathcal{L} = 0, \delta_\Phi\mathcal{L} = 0$, the minimizer satisfies the following properties:

$$\frac{\boldsymbol{m}(t,x)}{\mu(t,x)} = \nabla\Phi(t,x) \tag{9}$$

$$-\frac{\|\boldsymbol{m}(t,x)\|^2}{2\mu(t,x)^2} - \partial_t\Phi(t,x) \geq 0 \tag{10}$$

$$f(t,x) = \alpha\Phi(t,x)$$

$$\partial_t\mu(t,x) + \nabla\cdot\boldsymbol{m}(t,x) - f(t,x) = 0.$$

Combining (9) and (10) yields: $\partial_t\Phi(t,x) + \frac{1}{2}\|\nabla\Phi(t,x)\|^2 \leq 0$. □

We next derive the corresponding Monge problem for unnormalized optimal transport with a spatially dependent source function. We note that the following derivations are formal in Eulerian coordinates of fluid dynamics. We are following the proof of Proposition 4 in [12].

Proposition 5 *(Generalized unnormalized Monge problem).*

$$UW_2(\mu_0,\mu_1)^2 = \inf_{M,f(t,x)} \int_\Omega \|M(x)-x\|^2\mu_0(x)\mathrm{d}x + \alpha\int_0^1\int_\Omega |f(t,x)|^2\mathrm{d}x\mathrm{d}t$$

$$+ \int_\Omega\int_0^1\int_0^t f\Big(s, sM(x)+(1-s)x\Big)\|M(x)-x\|^2 Det\Big(s\nabla M(x)+(1-s)\mathbb{I}\Big)\mathrm{d}s\mathrm{d}t\mathrm{d}x \tag{11}$$

where $M : \Omega \to \Omega$ is an invertible mapping function and $f : \Omega \times [0, 1] \to \mathbb{R}$ is a spatially dependent source function. The unnormalized push forward relation holds

$$\mu(1, M(x)) Det(\nabla M(x))$$
$$= \mu(0, x) + \int_0^1 f\left(t, tM(x) + (1-t)\mathbb{I}\right) Det\left(t\nabla M(x) + (1-t)\mathbb{I}\right) dt. \tag{12}$$

Proof. We derive the Lagrange formulation of the unnormalized optimal transport with $p = 2$. Consider a mapping function $X_t(x)$ with vector field $\boldsymbol{v}(t, X_t(x))$ satisfying

$$\frac{d}{dt} X_t(x) = \boldsymbol{v}(t, X_t(x)), \qquad X_0(x) = x. \tag{13}$$

Then

$$\int_\Omega \int_0^1 \|\boldsymbol{v}(t, x)\|^2 \mu(t, x) dt dx = \int_\Omega \int_0^1 \|\boldsymbol{v}(t, X_t(x))\|^2 \mu(t, X_t(x)) Det(\nabla X_t(x)) dx dt$$
$$= \int_\Omega \int_0^1 \|\frac{d}{dt} X_t(x)\|^2 \mu(t, X_t(x)) Det(\nabla X_t(x)) dx dt. \tag{14}$$

Define $J(t, x) := \mu(t, X_t(x)) Det(\nabla X_t(x))$. Differentiate $J(t, x)$ with respect to t,

$$\frac{d}{dt} J(t, x) = \frac{d}{dt} \left\{ \mu(t, X_t(x)) Det(\nabla X_t(x)) \right\}$$
$$= \partial_t \mu(t, X_t(x)) Det(\nabla X_t(x)) + \nabla_X \mu(t, X_t(x)) \cdot \frac{d}{dt} X_t(x) Det(\nabla X_t(x))$$
$$\quad + \mu(t, X_t(x)) \partial_t Det(\nabla X_t(x))$$
$$= \partial_t \mu(t, X_t(x)) Det(\nabla X_t(x)) + \nabla_X \mu(t, X_t(x)) \cdot \frac{d}{dt} X_t(x) Det(\nabla X_t(x))$$
$$\quad + \mu(t, X_t(x)) \nabla \cdot \boldsymbol{v}(t, X_t(x)) Det(\nabla X_t(x))$$
$$= \left(\partial_t \mu + \boldsymbol{v} \cdot \nabla \mu + \mu \nabla \cdot \boldsymbol{v} \right)(t, X_t(x)) Det(\nabla X_t(x))$$
$$= \left(\partial_t \mu + \nabla \cdot (\mu \boldsymbol{v}) \right)(t, X_t(x)) Det(\nabla X_t(x))$$
$$= f\left(t, X_t(x)\right) Det(\nabla X_t(x)).$$

Denote

$$J(t, x) = J(0, x) + \int_0^t \frac{d}{ds} J(s, x) ds.$$

Since $X_0(x) = x$ and $\nabla X_0(x) = \mathbb{I}$, then $J(0, x) = \mu(0, x)$. This yields

$$\mu(t, X_t(x)) Det(\nabla X_t(x)) = \mu(0, x) + \int_0^t f\left(s, X_s(x)\right) Det(\nabla X_s(x)) ds.$$

Since the minimizer in Eulerian coordinates satisfies the Hamilton-Jacobi equation:

$$\partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 = 0,$$

and $\frac{d}{dt} X_t(x) = \nabla \Phi(t, X_t(x))$, then we have $\frac{d^2}{dt^2} X_t(x) = 0$. This implies

$$\frac{d}{dt} X_t(x) = \boldsymbol{v}(t, X_t(x)) = M(x) - x,$$

thus $X_t(x) = (1-t)x + tM(x)$ and $Det(\nabla X_t(x)) = Det((1-t)\mathbb{I} + t\nabla M(x))$. Substitute all the above into (14):

$$
\begin{aligned}
(14) &= \int_0^1 \int_\Omega \|\frac{\mathrm{d}}{\mathrm{d}t}X_t(x)\|^2 J(t,x)\mathrm{d}x\mathrm{d}t \\
&= \int_0^1 \int_\Omega \|M(x) - x\|^2\left(J(0,x) + \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}J(s,x)ds\right)\mathrm{d}x\mathrm{d}t \\
&= \int_0^1 \int_\Omega \|M(x) - x\|^2\mu(0,x)\mathrm{d}x\mathrm{d}t \\
&\quad + \int_0^1 \int_\Omega \|M(x) - x\|^2 \int_0^t f\big(s, X_s(x)\big)Det(\nabla X_s(x))\mathrm{d}s\mathrm{d}x\mathrm{d}t \\
&= \int_\Omega \|M(x) - x\|^2\mu(0,x)dx \\
&\quad + \int_0^1 \int_0^t \int_\Omega \|M(x) - x\|^2 f\left(s, sM(x) + (1-s)x\right)Det\left((1-s)\mathbb{I} + s\nabla M(x)\right)\mathrm{d}x\mathrm{d}s\mathrm{d}t.
\end{aligned}
$$

This concludes the derivation. $\square$

We next find the relation between the spatially dependent source function $f(t,x)$ and the mapping function $M(x)$. For the simplicity of presentation, here we assume the periodic boundary conditions on Ω. We are following the proof of Proposition 5 in [12].

Proposition 6 (*Generalized unnormalized Monge-Ampère equation*). *The optimal mapping function $M(x) = \nabla\Psi(x)$ satisfies the following unnormalized Monge-Ampère equation*

$$
\begin{aligned}
&\mu(1, \nabla\Psi(x))Det(\nabla^2\Psi(x)) - \mu(0,x) \\
&= \alpha \int_0^1 \left(\Psi(x) - \frac{\|x\|^2}{2} + \frac{t\|\nabla\Psi(x) - x\|^2}{2}\right)Det\left(t\nabla^2\Psi(x) + (1-t)\mathbb{I}\right)\mathrm{d}t.
\end{aligned}
\tag{15}
$$

Proof. From the Hopf-Lax formula for the Hamilton-Jacobi equation,

$$
\Phi(1, y) = \sup_x \Phi(0,x) + \frac{\|y - x\|^2}{2}.
$$

Since M is the optimal mapping function, $x = M^{-1}(y)$ is a maximizer of the supremum for each y. Thus, the maximizer satisfies

$$
\nabla\Phi(0,x) + x - M(x) = 0,
$$

and we can rewrite the formula as

$$
\Phi(1, M(x)) = \Phi(0,x) + \frac{\|M(x) - x\|^2}{2}.
$$

We further denote $\Psi(x) = \Phi(0,x) + \frac{\|x\|^2}{2}$, then $M(x) = \nabla\Psi(x)$. From $X_t(x) = (1-t)x + tM(x)$,

$$
\begin{aligned}
\Phi(t, X_t(x)) &= \Phi(0,x) + \frac{\|X_t(x) - x\|^2}{2t} \\
&= \Phi(0,x) + \frac{t\|M(x) - x\|^2}{2} \\
&= \Psi(x) - \frac{\|x\|^2}{2} + \frac{t\|\nabla\Psi(x) - x\|^2}{2}
\end{aligned}
$$

and

$$\nabla X_t(x) = (1-t)\mathbb{I} + t\nabla^2\Psi(x).$$

Substituting $f(t,x) = \alpha\Phi(t,x)$ and $M(x) = \nabla\Psi(x)$ into (12), we get

$$\mu(1,\nabla\Psi(x))Det(\nabla^2\Psi(x)) - \mu(0,x)$$

$$= \int_0^1 \alpha\left(\Psi(x) - \frac{\|x\|^2}{2} + \frac{t\|\nabla\Psi(x) - x\|^2}{2}\right)Det\left(t\nabla^2\Psi(x) + (1-t)\mathbb{I}\right)\mathrm{d}t. \quad \square$$

Now, we show the Kantorovich formulation of the problem (7).

Proposition 7 (*L^2 generalized unnormalized Kantorovich formulation*). *The unnormalized Kantorovich formulation with $f(t,x)$ satisfies*

$$\frac{1}{2}UW_2(\mu_0,\mu_1)^2 = \sup_{\Phi}\left\{\int_\Omega\left(\Phi(1,x)\mu_1(x) - \Phi(0,x)\mu_0(x)\right)\mathrm{d}x - \frac{\alpha}{2}\int_0^1\int_\Omega \Phi(t,x)^2\mathrm{d}x\mathrm{d}t\right\},$$

where the supremum is taken among all $\Phi : [0,1] \times \Omega \to \mathbb{R}$ satisfying

$$\partial_t\Phi(t,x) + \frac{1}{2}\|\nabla\Phi(t,x)\|^2 \le 0.$$

Proof. We introduce a Lagrange multiplier $\Phi(t,x)$ to reformulate the equation (8).

$$\frac{1}{2}UW_2(\mu_0,\mu_1)^2$$

$$= \inf_{\boldsymbol{m},\mu,f}\sup_{\Phi}\left\{\int_0^1\int_\Omega \frac{\|\boldsymbol{m}(t,x)\|^2}{2\mu(t,x)} + \frac{1}{2\alpha}f(x,t)^2 + \Phi(t,x)\left(\partial_t\mu(t,x) + \nabla\cdot\boldsymbol{m}(t,x) - f(t,x)\right)\mathrm{d}x\mathrm{d}t\right\}$$

$$\ge \sup_{\Phi}\inf_{\boldsymbol{m},\mu,f}\left\{\int_0^1\int_\Omega \frac{\|\boldsymbol{m}(t,x)\|^2}{2\mu(t,x)} + \frac{1}{2\alpha}f(x,t)^2 + \Phi(t,x)\left(\partial_t\mu(t,x) + \nabla\cdot\boldsymbol{m}(t,x) - f(t,x)\right)\mathrm{d}x\mathrm{d}t\right\}$$

$$= \sup_{\Phi}\inf_{\boldsymbol{m},\mu,f}\left\{\int_0^1\int_\Omega \frac{\|\boldsymbol{m}(t,x)\|^2}{2\mu(t,x)} - \nabla\Phi(t,x)\cdot\boldsymbol{m}(t,x) + \frac{1}{2\alpha}f(x,t)^2 + \Phi(t,x)\cdot\left(\partial_t\mu(t,x) - f(t,x)\right)\mathrm{d}x\mathrm{d}t\right\}$$

$$= \sup_{\Phi}\inf_{\boldsymbol{m},\mu,f}\left\{\int_0^1\int_\Omega \frac{1}{2}\left\|\frac{\boldsymbol{m}(t,x)}{\mu(t,x)} - \nabla\Phi(t,x)\right\|^2\mu(t,x) - \frac{1}{2}\|\nabla\Phi(t,x)\|^2\mu(t,x)\mathrm{d}x\mathrm{d}t\right.$$

$$+ \int_\Omega \Phi(1,x)\mu_1(x) - \Phi(0,x)\mu_0(x)\mathrm{d}x$$

$$\left. + \int_0^1\int_\Omega -\mu(t,x)\partial_t\Phi(t,x) + \frac{1}{2\alpha}f(t,x)^2 - \Phi(t,x)f(t,x))\mathrm{d}x\mathrm{d}t\right\}.$$

By the Proposition 4, the minimizer $\boldsymbol{m}$ satisfies $\dfrac{\boldsymbol{m}(t,x)}{\mu(t,x)} = \nabla\Phi(t,x)$. Thus,

$$= \sup_{\Phi}\left\{\int_\Omega\left(\Phi(1,x)\mu_1(x) - \Phi(0,x)\mu_0(x)\right)\mathrm{d}x\right.$$

$$\left. + \inf_\mu\int_0^1\int_\Omega -\mu(t,x)\left(\partial_t\Phi(t,x) + \frac{1}{2}\|\nabla\Phi(t,x)\|^2\right)\mathrm{d}x\mathrm{d}t\right.$$

$$+ \inf_{f} \int_0^1 \int_\Omega \frac{1}{2\alpha} f(t,x)^2 - \Phi(t,x) f(t,x) \mathrm{d}x \mathrm{d}t \Bigg\}$$

$$= \sup_{\Phi} \Bigg\{ \int_\Omega \left(\Phi(1,x)\mu_1(x) - \Phi(0,x)\mu_0(x) \right) \mathrm{d}x$$

$$+ \inf_{\mu} \int_0^1 \int_\Omega -\mu(t,x) \left(\partial_t \Phi(t,x) + \frac{1}{2} \|\nabla \Phi(t,x)\|^2 \right) \mathrm{d}x \mathrm{d}t$$

$$+ \inf_{f} \int_0^1 \int_\Omega \frac{1}{2\alpha} \left(f(t,x) - \alpha \Phi(t,x) \right)^2 \mathrm{d}x \mathrm{d}t - \frac{\alpha}{2} \int_0^1 \int_\Omega \Phi(t,x)^2 \mathrm{d}x \mathrm{d}t \Bigg\}.$$

Again from Proposition 4, the minimizer satisfies $f(t,x) = \alpha \Phi(t,x)$. With the assumption $\mu(t,x) \geq 0$ for all $t \in [0,1]$ and $x \in \Omega$, the problem can be written with a constraint.

$$\frac{1}{2} U W_2(\mu_0, \mu_1)^2 = \sup_{\Phi} \Bigg\{ \int_\Omega \left(\Phi(1,x)\mu_1(x) - \Phi(0,x)\mu_0(x) \right) \mathrm{d}x - \frac{\alpha}{2} \int_0^1 \int_\Omega \Phi(t,x)^2 \mathrm{d}x \mathrm{d}t :$$

$$\partial_t \Phi(t,x) + \frac{1}{2} \|\nabla \Phi(t,x)\|^2 \leq 0 \Bigg\}.$$

We next show that the primal-dual gap is zero.

$$\int_0^1 \int_\Omega \frac{\boldsymbol{m}(t,x)^2}{2\mu(t,x)} + \frac{1}{2\alpha} f(t,x)^2 \mathrm{d}x \mathrm{d}t$$

$$= \int_0^1 \int_\Omega \frac{1}{2} \|\nabla \Phi\|^2 \mu(t,x) \mathrm{d}x \mathrm{d}t + \frac{\alpha}{2} \int_0^1 \int_\Omega \Phi(t,x)^2 \mathrm{d}x \mathrm{d}t$$

$$= \int_0^1 \int_\Omega \left(-\frac{1}{2} \|\nabla \Phi(t,x)\|^2 \mu(t,x) + \|\nabla \Phi(t,x)\|^2 \mu(t,x) + \frac{\alpha}{2} \Phi(t,x)^2 \right) \mathrm{d}x \mathrm{d}t$$

$$= \int_0^1 \int_\Omega \partial_t \Phi(t,x) \mu(t,x) + \Phi(t,x) \left(-\nabla \cdot \left(\mu(t,x) \nabla \Phi(t,x) \right) \right) + \frac{\alpha}{2} \Phi(t,x)^2 \mathrm{d}x \mathrm{d}t$$

$$= \int_\Omega \Phi(1,x)\mu_1(x) - \Phi(0,x)\mu_0(x) \mathrm{d}x$$

$$- \int_0^1 \int_\Omega \Phi(t,x) \left(\partial_t \mu(t,x) + \nabla \cdot \left(\mu(t,x) \nabla \Phi(t,x) \right) \right) \mathrm{d}x \mathrm{d}t + \frac{\alpha}{2} \int_0^1 \int_\Omega \Phi(t,x)^2 \mathrm{d}x \mathrm{d}t$$

$$= \int_\Omega \Phi(1,x)\mu_1(x) - \Phi(0,x)\mu_0(x) \mathrm{d}x$$

$$- \int_0^1 \int_\Omega \Phi(t,x) f(t,x) \mathrm{d}x \mathrm{d}t + \frac{\alpha}{2} \int_0^1 \int_\Omega \Phi(t,x)^2 \mathrm{d}x \mathrm{d}t.$$

Using $f(t,x) = \alpha \Phi(t,x)$, we get

$$\frac{1}{2} U W_2(\mu_0, \mu_1)^2 = \int_\Omega \Phi(1,x)\mu(1,x) - \Phi(0,x)\mu(0,x) \mathrm{d}x - \frac{\alpha}{2} \int_0^1 \int_\Omega \Phi(t,x)^2 \mathrm{d}x \mathrm{d}t.$$

This concludes the proof. $\quad \square$

Remark 5. We note that our results and proofs follow directly from the those used in [12]. The major difference between [12] and our paper is that in the case of spatial independent source function, $f(t) = \frac{\alpha}{|\Omega|} \int_\Omega \Phi(t, x) dx$, while in the case of spatial dependent source function, $f(t, x) = \alpha \Phi(t, x)$. This difference remains in the corresponding Monge problem and Kantorvich problem. In particular, we obtain a new spatial dependent unnormalized Monge-Ampère equation (15).

3. Numerical methods

In this section, we propose a Nesterov accelerated gradient descent method to solve L^2 unnormalized OT. In addition, we design a primal-dual hybrid gradient method to solve L^1 unnormalized OT.

3.1. L^2 generalized unnormalized Wasserstein metric

In this section, we present a new numerical implementation for L^2 unnormalized Wasserstein metric. We obtain a unconstrained version of the problem by plugging the PDE constraint into the objective function. Then the accelerated Nesterov gradient descent method is applied to solve the problem. We show that each iteration involves a simple elliptic equation where fast solvers can be applied. This novel numerical method can also be used in normalized optimal transport and unnormalized optimal transport with a spatially independent source function $f(t)$.

Using Proposition 4, we can rewrite the equation (8) as follows:

$$UW_2(\mu_0, \mu_1)^2 = \inf_{\Phi, \mu} \left\{ \int_0^1 \int_\Omega \|\nabla \Phi(t, x)\|_2^2 \mu(t, x) dx dt + \alpha \int_0^1 \int_\Omega |\Phi(t, x)|^2 dx dt : \right.$$
$$\partial_t \mu(t, x) + \nabla \cdot (\mu(t, x) \nabla \Phi(t, x)) = \alpha \Phi(t, x),$$
$$\left. \mu(0, x) = \mu_0(x), \mu(1, x) = \mu_1(x) \right\}.$$

Define an operator $L_\mu = -\nabla \cdot (\mu \nabla)$. The constraint $\partial_t \mu - L_\mu \Phi = \alpha \Phi$ leads to

$$\Phi = (L_\mu + \alpha Id)^{-1} \partial_t \mu. \tag{16}$$

With (16), the minimization problem can be reformulated as

$$UW_2(\mu_0, \mu_1)^2 = \inf_\mu \left\{ \int_0^1 \int_\Omega \mu \|\nabla (L_\mu + \alpha Id)^{-1} \partial_t \mu\|_2^2 dx dt + \alpha \int_0^1 \int_\Omega |(L_\mu + \alpha Id)^{-1} \partial_t \mu|^2 dx dt : \right.$$
$$\left. \mu(0, x) = \mu_0(x), \mu(1, x) = \mu_1(x) \right\}. \tag{17}$$

Using integration by parts,

$$\int_0^1 \int_\Omega \mu \|\nabla (L_\mu + \alpha Id)^{-1} \partial_t \mu\|^2 dx dt + \alpha \int_0^1 \int_\Omega |(L_\mu + \alpha Id)^{-1} \partial_t \mu|^2 dx dt$$

$$= \int_0^1 \int_\Omega - \left(\nabla \mu \nabla (L_\mu + \alpha Id)^{-1} \partial_t \mu \right) \left((L_\mu + \alpha Id)^{-1} \partial_t \mu \right) dx dt$$

$$+ \alpha \int_0^1 \int_\Omega |(L_\mu + \alpha Id)^{-1} \partial_t \mu|^2 dx dt$$

$$= \int_0^1 \int_\Omega \left(L_\mu (L_\mu + \alpha Id)^{-1} \partial_t \mu \right) \left((L_\mu + \alpha Id)^{-1} \partial_t \mu \right) dx dt$$

$$+ \alpha \int_0^1 \int_\Omega |(L_\mu + \alpha Id)^{-1} \partial_t \mu|^2 dx dt$$

$$= \int\limits_0^1 \int\limits_\Omega \left(\left((L_\mu + \alpha Id)(L_\mu + \alpha Id)^{-1} \partial_t \mu \right) \left((L_\mu + \alpha Id)^{-1} \partial_t \mu \right) \right) \mathrm{d}x \mathrm{d}t$$

$$= \int\limits_0^1 \int\limits_\Omega \partial_t \mu(t, x)(L_\mu + \alpha Id)^{-1} \partial_t \mu(t, x) \mathrm{d}x \mathrm{d}t.$$

Thus, the unnormalized Wasserstein-2 distance forms

$$UW_2(\mu_0, \mu_1)^2 = \inf_\mu \left\{ \int\limits_0^1 \int\limits_\Omega \partial_t \mu(t, x)(L_\mu + \alpha Id)^{-1} \partial_t \mu(t, x) \mathrm{d}x \mathrm{d}t : \right. \tag{18}$$

$$\left. \mu(0, x) = \mu_0(x), \mu(1, x) = \mu_1(x) \right\}.$$

Proposition 8. *If $\mu(t, x) > 0$, then the Euler-Lagrange equation of problem (18) satisfies the Hamilton-Jacobi equation, i.e.*

$$\partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 = 0, \quad x \in \Omega, t \in [0, 1]$$

where $\Phi(t, x) = (L_\mu + \alpha Id)^{-1} \partial_t \mu(t, x)$.

Remark 6. For unnormalized optimal transport with a spatially independent source function $f(t)$, the formula uses $(L_\mu + \frac{\alpha}{|\Omega|} \int_\Omega)^{-1}$ instead of $(L_\mu + \alpha Id)^{-1}$, i.e.

$$UW_2(\mu_0, \mu_1)^2 = \inf_\mu \left\{ \int\limits_0^1 \int\limits_\Omega \partial_t \mu(t, x) \left(L_\mu + \frac{\alpha}{|\Omega|} \int\limits_\Omega \right)^{-1} \partial_t \mu(t, x) \mathrm{d}x \mathrm{d}t : \right.$$

$$\left. \mu(0, x) = \mu_0(x), \mu(1, x) = \mu_1(x) \right\}.$$

The Euler-Lagrange equation satisfies the following:

$$\partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 = 0, \quad x \in \Omega, t \in [0, 1]$$

where $\Phi(t, x) = \left(L_\mu + \frac{\alpha}{|\Omega|} \int_\Omega \right)^{-1} \partial_t \mu(t, x)$.

Remark 7. If $\mu(t, x) = 0$, one can show that the Euler-Lagrange equation of problem (18) satisfies

$$\partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 \le 0.$$

Proof. Define

$$\mathcal{I}(\mu) = \int\limits_0^1 \int\limits_\Omega \partial_t \mu(t, x)(L_\mu + \alpha Id)^{-1} \partial_t \mu(t, x) \mathrm{d}x \mathrm{d}t.$$

We now calculate the first variation of $\mathcal{I}(\mu)$ with a perturbation $\eta(t, x) \in C^\infty(\Omega \times [0, 1])$.

$$0 = \lim_{h \to 0} \frac{\mathcal{I}(\mu + h\eta) - \mathcal{I}(\mu)}{h}$$

$$= \lim_{h \to 0} \frac{1}{h} \int\limits_0^1 \int\limits_\Omega \left((\partial_t \mu + h \partial_t \eta)(L_{\mu+h\eta} + \alpha Id)^{-1}(\partial_t \mu + h \partial_t \eta) - \partial_t \mu(t, x)(L_\mu + \alpha Id)^{-1} \partial_t \mu(t, x) \right) \mathrm{d}x \mathrm{d}t$$

$$= \lim_{h \to 0} \left[\int\limits_0^1 \int\limits_\Omega \partial_t \mu \left(\frac{(L_{\mu+h\eta} + \alpha Id)^{-1} - (L_\mu + \alpha Id)^{-1}}{h} \right) \partial_t \mu \right.$$

$$+ 2\partial_t\eta(L_{\mu+h\eta} + \alpha Id)^{-1}\partial_t\mu \mathrm{d}x\mathrm{d}t + O(h)\Bigg]$$

$$= \int_0^1\int_\Omega -\partial_t\mu(L_\mu + \alpha Id)^{-1}L_\eta(L_\mu + \alpha Id)^{-1}\partial_t\mu + 2\partial_t\eta(L_\mu + \alpha Id)^{-1}\partial_t\mu \mathrm{d}x\mathrm{d}t$$

$$= \int_0^1\int_\Omega -\Phi L_\eta\Phi + 2\Phi\partial_t\eta \mathrm{d}x\mathrm{d}t$$

$$= \int_0^1\int_\Omega -\eta\left(\|\nabla\Phi\|^2 + 2\partial_t\Phi\right)\mathrm{d}x\mathrm{d}t.$$

This has to be true for all $\eta \in C^\infty(\Omega \times [0,1])$. Thus, we get

$$\partial_t\Phi + \frac{1}{2}\|\nabla\Phi\|^2 = 0, \quad x \in \Omega, t \in [0,1].$$

This concludes the proof. □

Using Proposition 8, we can formulate a Nesterov accelerated gradient descent method [21] to solve the minimization problem (18).

Algorithm 1 Nesterov Gradient descent method for UW_2 with $f(t,x)$.

---

while not converged **do**

$\quad \mu^{k+\frac{1}{2}} = \mu^k - \tau\nabla\mathcal{I}(\mu^k) = \mu^k + \frac{\tau}{2}\left(\partial_t\Phi^k + \frac{1}{2}\|\nabla\Phi^k\|^2\right)$ where $\Phi^k = (L_{\mu^k} + \alpha Id)^{-1}\partial_t\mu^k$

$\quad \mu^{k+\frac{1}{2}} = \max\{\mu^{k+\frac{1}{2}}, 0\}$

$\quad \mu^{k+1} = (1 - \gamma^k)\mu^{k+\frac{1}{2}} + \gamma^k\mu^k$

end while

Here, τ and γ^k are step sizes of the algorithm.

$$\gamma^k = \frac{1 - \lambda^k}{\lambda^{k+1}}, \quad \lambda^0 = 0, \quad \lambda^k = \frac{1 + \sqrt{1 + 4(\lambda^{k-1})^2}}{2}.$$

Remark 8. The Nesterov accelerated gradient descent method can be used for a spatially independent source function $f(t)$. We simply replace the operator $L_\mu + \alpha Id$ with $L_\mu + \alpha\int_\Omega$ from Algorithm 1.

Remark 9. Here we apply an iterative method, such as conjugate gradient, to solve $(L_{\mu^k} + \alpha Id)^{-1}\partial_t\mu^k$.

Remark 10. We remark that variational problem (18) is convex w.r.t. $\mu(t,x)$. This fact holds following the second variational formula derived in Lemma 2 of [15]. In other words, our gradient descent algorithm is applied to a convex optimization problem (18). For the completeness of this paper, we present the formal derivation here. Denote

$$J(\mu) = \frac{1}{2}\int_0^1\int_\Omega \partial_t\mu(t,x)(L_\mu + \alpha Id)^{-1}\partial_t\mu(t,x)\mathrm{d}x\mathrm{d}t.$$

Consider a test function $h \in C^\infty([0,1] \times \Omega)$, such that $h(0,x) = h(1,x) = 0$. Given $\epsilon \in \mathbb{R}^1$, we claim

$$\frac{d^2}{d\epsilon^2}J(\mu + \epsilon h)|_{\epsilon=0} \geq 0.$$

If the above statement is true, we know that the variational problem (18) is convex w.r.t. $\mu(t,x)$. In fact, by routine computations, we observe that

$$\frac{d^2}{d\epsilon^2}J(\mu + \epsilon h)|_{\epsilon=0}$$
$$= \int_\Omega\left([\partial_t h - L(h)(L_\mu + \alpha Id)^{-1}\partial_t\mu], (L_\mu + \alpha Id)^{-1}[\partial_t h - L(h)(L_\mu + \alpha Id)^{-1}\partial_t\mu]\right)\mathrm{d}x\mathrm{d}t,$$

which finishes the proof.

We next present the discretization of density path in both time and spatial domains, where the spatial domain is given by $1D$ or $2D$. Here we formulate the operator L_μ and derive its inverse into matrix forms; see similar approaches in [15].

3.1.1. 1D discretization

Consider the following one dimensional discretization:

$$\mu = (\mu^0, \cdots, \mu^{N_t}) \in \mathbb{R}^{(N_t+1) \times N_x}$$

$$\mu^n = (\mu_0^n, \cdots, \mu_{N_x-1}^n) \in \mathbb{R}^{N_x} \quad (n = 0, \cdots, N_t)$$

$$\mu_i^n \in \mathbb{R} \quad (i = 0, \cdots, N_x - 1, n = 0, \cdots, N_t)$$

$$\mu_i^0 = \mu_0(i\Delta x), \quad \mu_i^{N_t} = \mu_1(i\Delta x), \quad (i = 0, \cdots, N_x - 1)$$

$$\Delta x = \frac{|\Omega|}{N_x - 1} \quad \Delta t = \frac{1}{N_t}.$$

Using the finite volume method, the weighted Laplacian operator $\tilde{L}_{\mu^n,\alpha} := L_{\mu^n} + \alpha Id$ can be represented as the following matrix:

$$\tilde{L}_{\mu^n,\alpha} = \begin{pmatrix} \frac{\mu_0^n + \mu_1^n}{2\Delta x^2} & -\frac{\mu_0^n + \mu_1^n}{2\Delta x^2} & 0 & \cdots & 0 \\ -\frac{\mu_0^n + \mu_1^n}{2\Delta x^2} & \frac{\mu_0^n + \mu_1^n}{2\Delta x^2} + \frac{\mu_1^n + \mu_2^n}{2\Delta x^2} & -\frac{\mu_1^n + \mu_2^n}{2\Delta x^2} & \cdots & 0 \\ 0 & -\frac{\mu_1^n + \mu_2^n}{2\Delta x^2} & \frac{\mu_1^n + \mu_2^n}{2\Delta x^2} + \frac{\mu_2^n + \mu_3^n}{2\Delta x^2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & -\frac{\mu_{N_x-2}^n + \mu_{N_x-1}^n}{2\Delta x^2} \end{pmatrix} + \alpha Id$$

Further using the forward Euler method in time, formula (18) can be discretized as

$$\int_0^1 \int_\Omega \partial_t \mu(t,x)(L_\mu + \alpha Id)^{-1} \partial_t \mu(t,x) dx dt$$

$$\approx \Delta t \Delta x \sum_{n=0}^{N_t-1} \left\langle \frac{\mu^{n+1} - \mu^n}{\Delta t}, (L_{\mu^n} + \alpha Id)^{-1} \frac{\mu^{n+1} - \mu^n}{\Delta t} \right\rangle_{L^2}$$

$$= \frac{\Delta x}{\Delta t} \sum_{n=0}^{N_t-1} \left\langle \mu^{n+1} - \mu^n, (L_{\mu^n} + \alpha Id)^{-1}(\mu^{n+1} - \mu^n) \right\rangle_{L^2}$$

with μ^0 and μ^{N_t} are given. $\langle \cdot, \cdot \rangle_{L^2}$ is L^2 norm in $\mathbb{R}^{N_x}$ such that

$$\langle a, b \rangle_{L^2} = \sum_{i=0}^{N_x-1} a_i b_i \quad \text{for } a, b \in \mathbb{R}^{N_x}.$$

We are now ready to present the derivative of $E(\mu)$, and formulate the discrete Hamilton-Jacobi equation as in Algorithm 1.

Proposition 9. Denote $\tilde{L}_{\mu^n,\alpha} := L_{\mu^n} + \alpha Id$. Let

$$E(\mu) := \frac{\Delta x}{\Delta t} \sum_{n=0}^{N_t-1} \left\langle \mu^{n+1} - \mu^n, \tilde{L}_{\mu^n,\alpha}^{-1}(\mu^{n+1} - \mu^n) \right\rangle_{L^2}.$$

Suppose $x \in \Omega$. The derivative of $E(\mu)$ with respect to μ^n ($n = 1, \cdots, N_t - 1$) is

$$\frac{\delta E(\boldsymbol{\mu})}{\delta \boldsymbol{\mu}^n} = \frac{\Delta x}{\Delta t}\left(-2\tilde{L}_{\boldsymbol{\mu}^n,\alpha}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n) + 2\tilde{L}_{\boldsymbol{\mu}^{n-1},\alpha}(\boldsymbol{\mu}^n-\boldsymbol{\mu}^{n-1})\right.$$

$$\left. - \left(\left\langle \tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n), L_{\boldsymbol{e}_i}\tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n)\right\rangle_{L^2}\right)_{i=0}^{N_x-1}\right)$$

where $\boldsymbol{e}_i \in \mathbb{R}^{N_x}$ is an index vector defined as

$$\boldsymbol{e}_i = \begin{cases} 1 & i^{th}\ index \\ 0 & else. \end{cases}$$

Proof. Differentiating $E(\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}^n$ for $n = 1, \cdots, N_t - 1$, we get

$$\frac{\delta E(\boldsymbol{\mu})}{\delta \boldsymbol{\mu}^n} = \frac{\delta}{\delta \boldsymbol{\mu}^n}\left(\Delta t \Delta x \sum_{m=0}^{N_t-1}(\boldsymbol{\mu}^{m+1}-\boldsymbol{\mu}^m)\tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}(\boldsymbol{\mu}^{m+1}-\boldsymbol{\mu}^m)\right)$$

$$= \Delta t \Delta x\left(-2\tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n) + 2\tilde{L}_{\boldsymbol{\mu}^{n-1},\alpha}^{-1}(\boldsymbol{\mu}^n-\boldsymbol{\mu}^{n-1})\right.$$

$$\left. + (\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n)\frac{\partial \tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}}{\partial \boldsymbol{\mu}^n}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n)\right),$$

and

$$(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n)\frac{\delta \tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}}{\delta \boldsymbol{\mu}_n^n}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n) = -\left\langle \boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n, \tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}L_{\boldsymbol{e}_i}\tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n)\right\rangle_{L^2}$$

$$= -\left\langle \tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n), L_{\boldsymbol{e}_i}\tilde{L}_{\boldsymbol{\mu}^n,\alpha}^{-1}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n)\right\rangle_{L^2}.$$

This concludes the proof. □

Consider $\boldsymbol{u} = (u_0, \cdots, u_{N_x-1})^T \in \mathbb{R}^{N_x}$, then $\langle \boldsymbol{u}, L_{\boldsymbol{e}_i}\boldsymbol{u}\rangle_{L_2}$ forms the R.H.S. of the discrete Hamilton-Jacobi equation as follows

$$\langle \boldsymbol{u}, L_{\boldsymbol{e}_i}\boldsymbol{u}\rangle_{L^2} = \begin{cases} \frac{1}{2}\left(\frac{u_{i+1}-u_i}{\Delta x}\right)^2 + \frac{1}{2}\left(\frac{u_i-u_{i-1}}{\Delta x}\right)^2, & i = 1, \cdots, N_x - 2 \\ \frac{1}{2}\left(\frac{u_{i+1}-u_i}{\Delta x}\right)^2, & i = 0 \\ \frac{1}{2}\left(\frac{u_i-u_{i-1}}{\Delta x}\right)^2, & i = N_x - 1. \end{cases}$$

3.1.2. 2D discretization

Now, consider the two dimensional discretization. Assume $\Omega = [0, 1] \times [0, 1]$ and $t \in [0, 1]$.

$$\boldsymbol{\mu} = (\boldsymbol{\mu}^0, \cdots, \boldsymbol{\mu}^{N_t}) \in \mathbb{R}^{(N_t+1)\times N_x \times N_y}$$

$$\boldsymbol{\mu}^n = (\mu_{ij}^n)_{i=0\ \ j=0}^{N_x-1\ N_y-1} \in \mathbb{R}^{N_x \times N_y} \quad (n = 0, \cdots, N_t)$$

$$\mu_{ij}^n \in \mathbb{R} \quad (i = 0, \cdots, N_x - 1, j = 0, \cdots, N_y - 1, n = 0, \cdots, N_t)$$

$$\mu_{ij}^0 = \mu_0(i\Delta x, j\Delta y), \quad \mu_{ij}^{N_t} = \mu_1(i\Delta x, j\Delta y), \quad (i = 0, \cdots, N_x - 1, j = 0, \cdots, N_y - 1)$$

$$\Delta x = \frac{1}{N_x - 1}, \quad \Delta y = \frac{1}{N_y - 1}, \quad \Delta t = \frac{1}{N_t}.$$

Similar to 1D case, using the finite volume method, formula (18) can be discretized as

$$\int_0^1 \int_0^1 \int_0^1 \partial_t \mu(t, x, y)(L_\mu + \alpha Id)^{-1}\partial_t \mu(t, x, y)\mathrm{d}x\mathrm{d}y\mathrm{d}t$$

$$\approx \frac{\Delta x \Delta y}{\Delta t}\sum_{n=0}^{N_t-1}\left\langle \boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n, (L_{\boldsymbol{\mu}^n}+\alpha Id)^{-1}(\boldsymbol{\mu}^{n+1}-\boldsymbol{\mu}^n)\right\rangle_{L^2}$$

with $\boldsymbol{\mu}^0$ and $\boldsymbol{\mu}^{N_t}$ are given and $\langle \cdot, \cdot \rangle_{L^2}$ is defined as

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle_{L^2} = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} a_{ij} b_{ij} \qquad \text{for } \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{N_x \times N_y}.$$

The major difference between 1D discretization and 2D discretization arises from the weighted Laplacian operator $\tilde{L}_{\boldsymbol{\mu}^n, \alpha}$. Consider $\boldsymbol{w} = (w_{i,j})_{i=0 \ \ j=0}^{N_x-1 \ N_y-1} \in \mathbb{R}^{N_x \times N_y}$. For $i = 0, \cdots, N_x - 1$ and $j = 0, \cdots, N_y - 1$, the operator can be described as follows:

$$
\begin{aligned}
&(\tilde{L}_{\boldsymbol{\mu}^n, \alpha} \boldsymbol{w})_{ij} \\
&= -\frac{1}{\Delta x^2} \left(\frac{\mu_{i+1,j}^n + \mu_{i,j}^n}{2} w_{i+1,j} - 2\left(\frac{\mu_{i+1,j}^n + \mu_{i,j}^n}{2} + \frac{\mu_{i,j}^n + \mu_{i-1,j}^n}{2} \right) w_{i,j} + \frac{\mu_{i,j}^n + \mu_{i-1,j}^n}{2} w_{i-1,j} \right) \\
&\quad - \frac{1}{\Delta y^2} \left(\frac{\mu_{i,j+1}^n + \mu_{i,j}^n}{2} w_{i,j+1} - 2\left(\frac{\mu_{i,j+1}^n + \mu_{i,j}^n}{2} + \frac{\mu_{i,j}^n + \mu_{i,j-1}^n}{2} \right) w_{i,j} + \frac{\mu_{i,j}^n + \mu_{i,j-1}^n}{2} w_{i,j-1} \right) \\
&\quad + \alpha w_{i,j}.
\end{aligned}
$$

Here, we assume the Neumann boundary on the spatial domain Ω. Thus,

$$
\begin{aligned}
w_{-1,j} &= w_{0,j}, & w_{N_x,j} &= w_{N_x-1,j}, & j &= 0, \cdots, N_y - 1 \\
w_{i,-1} &= w_{i,0}, & w_{i,N_y} &= w_{i,N_y-1}, & i &= 0, \cdots, N_x - 1 \\
\mu_{-1,j}^n &= \mu_{0,j}^n, & \mu_{N_x,j}^n &= \mu_{N_x-1,j}^n, & j &= 0, \cdots, N_y - 1 \\
\mu_{i,-1}^n &= \mu_{i,0}^n, & \mu_{i,N_y}^n &= \mu_{i,N_y-1}^n, & i &= 0, \cdots, N_x - 1.
\end{aligned}
$$

Proposition 10. *Denote* $\tilde{L}_{\boldsymbol{\mu}^n, \alpha} := L_{\boldsymbol{\mu}^n} + \alpha \, Id$. *Let*

$$E(\boldsymbol{\mu}) := \frac{\Delta x \Delta y}{\Delta t} \sum_{n=0}^{N_t-1} \left\langle \boldsymbol{\mu}^{n+1} - \boldsymbol{\mu}^n, \tilde{L}_{\boldsymbol{\mu}^n, \alpha}^{-1} (\boldsymbol{\mu}^{n+1} - \boldsymbol{\mu}^n) \right\rangle_{L^2}.$$

Suppose $x \in \Omega = [0, 1] \times [0, 1]$. *The derivative of* $E(\boldsymbol{\mu})$ *with respect to* $\boldsymbol{\mu}^n$ ($n = 1, \cdots, N_t - 1$) *is*

$$
\begin{aligned}
\frac{\delta E(\boldsymbol{\mu})}{\delta \boldsymbol{\mu}^n} = \frac{\Delta x \Delta y}{\Delta t} \Bigg(&-2\tilde{L}_{\boldsymbol{\mu}^n, \alpha} (\boldsymbol{\mu}^{n+1} - \boldsymbol{\mu}^n) + 2\tilde{L}_{\boldsymbol{\mu}^{n-1}, \alpha} (\boldsymbol{\mu}^n - \boldsymbol{\mu}^{n-1}) \\
&- \left(\left\langle \tilde{L}_{\boldsymbol{\mu}^n, \alpha}^{-1} (\boldsymbol{\mu}^{n+1} - \boldsymbol{\mu}^n), L_{\boldsymbol{e}_{ij}} \tilde{L}_{\boldsymbol{\mu}^n, \alpha}^{-1} (\boldsymbol{\mu}^{n+1} - \boldsymbol{\mu}^n) \right\rangle_{L^2} \right)_{i=0, j=0}^{N_x-1, N_y-1} \Bigg),
\end{aligned}
$$

where $\boldsymbol{e}_{ij}$ *is an index vector such that* $\boldsymbol{e}_{k,l} = 1$ *if* $k = i$ *and* $l = j$ *and* 0 *otherwise.*

Proof. The proof follows exactly the one in Proposition 9. □

Consider a vector $\boldsymbol{u} = (u_{ij})_{i=0 \ \ j=0}^{N_x-1 \ N_y-1} \in \mathbb{R}^{N_x \times N_y}$ that satisfies the Neumann boundary condition. Similar to 1D case, $\langle \boldsymbol{u}, L_{\boldsymbol{e}_{i,j}} \boldsymbol{u} \rangle_{L^2}$ can be computed easily based on the operator and it forms the R.H.S. of the discrete Hamilton-Jacobi equation. For $i = 0, \cdots, N_x - 1$ and $j = 0, \cdots, N_y - 1$,

$$
\begin{aligned}
\langle \boldsymbol{u}, L_{\boldsymbol{e}_{i,j}} \boldsymbol{u} \rangle_{L^2} = &\frac{1}{2} \left(\frac{u_{i+1,j} - u_{i,j}}{\Delta x} \right)^2 + \frac{1}{2} \left(\frac{u_{i,j} - u_{i-1,j}}{\Delta x} \right)^2 \\
&+ \frac{1}{2} \left(\frac{u_{i,j+1} - u_{i,j}}{\Delta y} \right)^2 + \frac{1}{2} \left(\frac{u_{i,j} - u_{i,j-1}}{\Delta y} \right)^2.
\end{aligned}
$$

3.2. L^1 generalized unnormalized Wasserstein metric

Our discussion here mainly focuses on $\|u\|_1 = \sum_i |u_i|$. The algorithm can be simply extended to $\|u\|_2 = \sqrt{\sum_i u_i^2}$ using the corresponding shrinkage operator. With the Lagrangian (5), we consider a saddle point problem.

$$\inf_{\boldsymbol{m},c} \sup_{\Phi} \mathcal{L}(m, c, \Phi).$$

We can use PDHG [4] to solve the saddle point problem by minimizing $\mathcal{L}(\boldsymbol{m}, c, \Phi)$ over m and c and maximizing over Φ.

$$\boldsymbol{m}^{k+1} = \arg\min_{\boldsymbol{m}} \left(\|\boldsymbol{m}\|_1 + \frac{\epsilon}{2}\|\boldsymbol{m}\|_2^2 + \left\langle \Phi^k, \nabla \cdot \boldsymbol{m} \right\rangle_{L^2} + \frac{1}{2\lambda}\|\boldsymbol{m} - \boldsymbol{m}^k\|_2^2 \right) \tag{19}$$

$$c^{k+1} = \arg\min_{c} \left(\frac{1}{\alpha}\|c\|_1 + \frac{\epsilon}{2}\|c\|_2^2 - \left\langle \Phi^k, c \right\rangle_{L^2} + \frac{1}{2\lambda}\|c - c^k\|_2^2 \right) \tag{20}$$

$$\Phi^{k+1} = \arg\max_{\Phi} \left(\left\langle \Phi, \nabla \cdot (2\boldsymbol{m}^{k+1} - \boldsymbol{m}^k) - (2c^{k+1} - c^k) + \mu_1 - \mu_0 \right\rangle_{L^2} - \frac{1}{2\tau}\|\Phi - \Phi^k\|_2^2 \right) \tag{21}$$

where λ and τ are step sizes of the algorithm. Note that we add a small $\|\cdot\|_2^2$ perturbation in (19) and (20) to strictly convexify the problem. This adjustment can overcome the possible non-uniqueness of the optimal transport problem. This trick is also related to so called the elastic net regularization [22], whose proximal operator is essentially the same as the proximal operator of L^1 norm shrink operator.

Algorithm 2 PDHG for UW_1 with $f(t, x)$.

$$\boldsymbol{m}^{k+1} = 1/(1 + \epsilon\lambda) shrink\left(\boldsymbol{m}^k + \lambda\nabla\Phi^k, \lambda \right)$$

$$c^{k+1} = 1/(1 + \epsilon\lambda) shrink\left(c^k + \lambda\Phi^k, \frac{\lambda}{\alpha} \right)$$

$$\Phi^{k+1} = \Phi^k + \tau\left(\nabla \cdot (2\boldsymbol{m}^{k+1} - \boldsymbol{m}^k) - (2c^{k+1} - c^k) + \mu_1 - \mu_0 \right)$$

---

where the *shrink* operator is defined as following:

$$(shrink(u, t))_i = \begin{cases} (1 - t/|u_i|)u_i, & \text{for } \|u_i\|_1 \geq t; \\ 0, & \text{for } \|u_i\|_1 < t. \end{cases} \quad i = 1, \cdots, d.$$

Remark 11. This algorithm can also be extended to $\|\cdot\|_2$ by simply replacing the above shrink operator as

$$shrink(u, t) = \begin{cases} (1 - t/\|u\|_2)u, & \text{for } \|u\|_2 \geq t; \\ 0, & \text{for } \|u\|_2 < t. \end{cases}$$

3.2.1. Discretization

Consider the following two dimensional discretization on a domain $\Omega = [0, 1] \times [0, 1]$ based on the finite volume method.

$$\Delta x = \frac{1}{N_x}, \Delta y = \frac{1}{N_y}$$

$$\mu_{ij}^0 = \mu_0(i\Delta x, j\Delta y), \quad \mu_{ij}^1 = \mu_1(i\Delta x, j\Delta y)$$

$$V = \{(i, j) : i = 0, \cdots, N_x, j = 0, \cdots, N_y\}$$

$$E_x = \{(i \pm \frac{1}{2}, j) : i = 1, \cdots, N_x - 1, j = 0, \cdots, N_y)\}$$

$$E_y = \{(i, j \pm \frac{1}{2}) : i = 0, \cdots, N_x, j = 1, \cdots, N_y - 1)\}$$

$$\boldsymbol{\Phi} = (\Phi_{ij})_{ij \in V} \in \mathbb{R}^{(N_x+1) \times (N_y+1)}, \quad \boldsymbol{c} = (c_{ij})_{ij \in V} \in \mathbb{R}^{(N_x+1) \times (N_y+1)}$$

$$\boldsymbol{mx} = (mx_e)_{e \in E_x} \in \mathbb{R}^{N_x \times (N_y+1)}, \quad \boldsymbol{my} = (my_e)_{e \in E_y} \in \mathbb{R}^{(N_x+1) \times N_y}$$

$$mx_{i+\frac{1}{2}, j} \approx \int\limits_{i\Delta x}^{(i+1)\Delta x} \int\limits_{(j-1/2)\Delta y}^{(j+1/2)\Delta y} m_x(x, y)\mathrm{d}y\mathrm{d}x$$

$$my_{i, j+\frac{1}{2}} \approx \int\limits_{(i-1/2)\Delta x}^{(i+1/2)\Delta x} \int\limits_{j\Delta y}^{(j+1)\Delta y} m_y(x, y)\mathrm{d}y\mathrm{d}x.$$

Here m satisfies the zero flux condition. Thus, $\boldsymbol{mx}$ and $\boldsymbol{my}$ satisfy the following boundary conditions on m:

$$mx_{-\frac{1}{2},j} = mx_{N_x+\frac{1}{2},j} = 0, \quad j = 0, \cdots, N_y$$
$$my_{i,-\frac{1}{2}} = my_{i,N_y+\frac{1}{2}} = 0, \quad i = 0, \cdots, N_x.$$

The discretization of Algorithm 2 can be written as the following:

$$mx_{i+\frac{1}{2},j}^{k+\frac{1}{2}} = \frac{1}{1+\epsilon\lambda}\left(mx_{i+\frac{1}{2},j}^{k} + \frac{\lambda}{\Delta x}(\Phi_{i+1,j} - \Phi_{i,j})\right)$$

$$my_{i,j+\frac{1}{2}}^{k+\frac{1}{2}} = \frac{1}{1+\epsilon\lambda}\left(my_{i,j+\frac{1}{2}}^{k} + \frac{\lambda}{\Delta y}(\Phi_{i,j+1} - \Phi_{i,j})\right)$$

$$c_{ij}^{k+\frac{1}{2}} = \frac{1}{1+\epsilon\lambda}shrink\left(c^k + \lambda\Phi_{ij}^k, \frac{\lambda}{\alpha}\right)$$

$$\boldsymbol{mx}^{k+1} = 2\boldsymbol{mx}^{k+\frac{1}{2}} - \boldsymbol{mx}^k$$

$$\boldsymbol{my}^{k+1} = 2\boldsymbol{my}^{k+\frac{1}{2}} - \boldsymbol{my}^k$$

$$\boldsymbol{c}^{k+1} = 2\boldsymbol{c}^{k+\frac{1}{2}} - \boldsymbol{c}^k$$

$$\Phi_{ij}^{k+1} = \Phi_{ij}^k + \tau\left(\frac{1}{\Delta x}(mx_{i+\frac{1}{2},j}^{k+1} - mx_{i-\frac{1}{2},j}^{k+1}) + \frac{1}{\Delta y}(my_{i,j+\frac{1}{2}}^{k+1} - my_{i,j-\frac{1}{2}}^{k+1}) - c_{ij}^{k+1} + \mu_{ij}^1 - \mu_{ij}^0\right).$$

4. Numerical experiments

In this section, we show the numerical results with various examples for L^1 and L^2 unnormalized optimal transport with the spatially dependent source function. The computations were conducted on 2019 MacBook Pro with 2.6 GHz 6-Core and 16 GB RAM.

4.1. Nesterov accelerated gradient descent for UW_2

We present four numerical experiments with different μ_0 and μ_1 using Algorithm 1.

4.1.1. Experiment 1
Consider a one dimensional problem on $\Omega = [0,1]$ with μ_0 and μ_1 in $\mathcal{M}(\Omega)$ as

$$\mu_0 = N(x; \frac{1}{5}, 0.0001)$$
$$\mu_1 = N(x; \frac{4}{5}, 0.0001) \cdot 1.4$$

Here we choose $N(x, \mu, \sigma^2) = C\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ with an appropriate choice of C satisfying $\int_\Omega N(x; \mu, \sigma^2)dx = 1$. Note that $\int_\Omega \mu_0 dx = 1$ and $\int_\Omega \mu_1 dx = 1.4$. We use the Algorithm 1 to compute the minimizer $\mu(t,x)$ of $UW_2(\mu_0, \mu_1)$. The parameters chosen for the experiment are

$$N_x = 40, N_t = 30, \tau = 0.1, \text{maximum iterations} = 200,000.$$

Fig. 1 shows the L^2 unnormalized optimal transport with a spatially dependent source function $f(t,x)$ with different α values. The parameter α determines the ratio between transportation and linear interpolation for μ_0 and μ_1. If α is small, the geodesic of generalized unnormalized optimal transport is similar to the normalized (classical) optimal transport geodesics. As the parameter α increases, the generalized unnormalized optimal transport geodesic behaves closer to the Euclidean geodesics.

Fig. 2 shows the transportation with a spatially independent source function $f(t)$. It is clear to see that the masses are created or removed locally for the transportation with $f(t,x)$, while they are created or removed globally for the transportation with $f(t)$.

For each 1-dimensional numerical experiment, the computation took about 5 seconds for 200,000 iterations.

4.1.2. Experiment 2
In this experiment, we can see how the size of the domain affects the unnormalized Wasserstein distances for both a spatially dependent source function $f(t,x)$ and a spatially independent source function $f(t)$. Consider a one dimensional problem between two densities with different total masses. Fig. 3 shows plots for the size of the domain $|\Omega|$ vs. the unnormalized Wasserstein distance UW_2. As expected, for the spatially independent source function, the distance increases as $|\Omega|$ increases since the source function affects the transportation globally. Thus, more masses are created or removed as $|\Omega|$ increases. However, the unnormalized Wasserstein distance with the spatially dependent source function does not
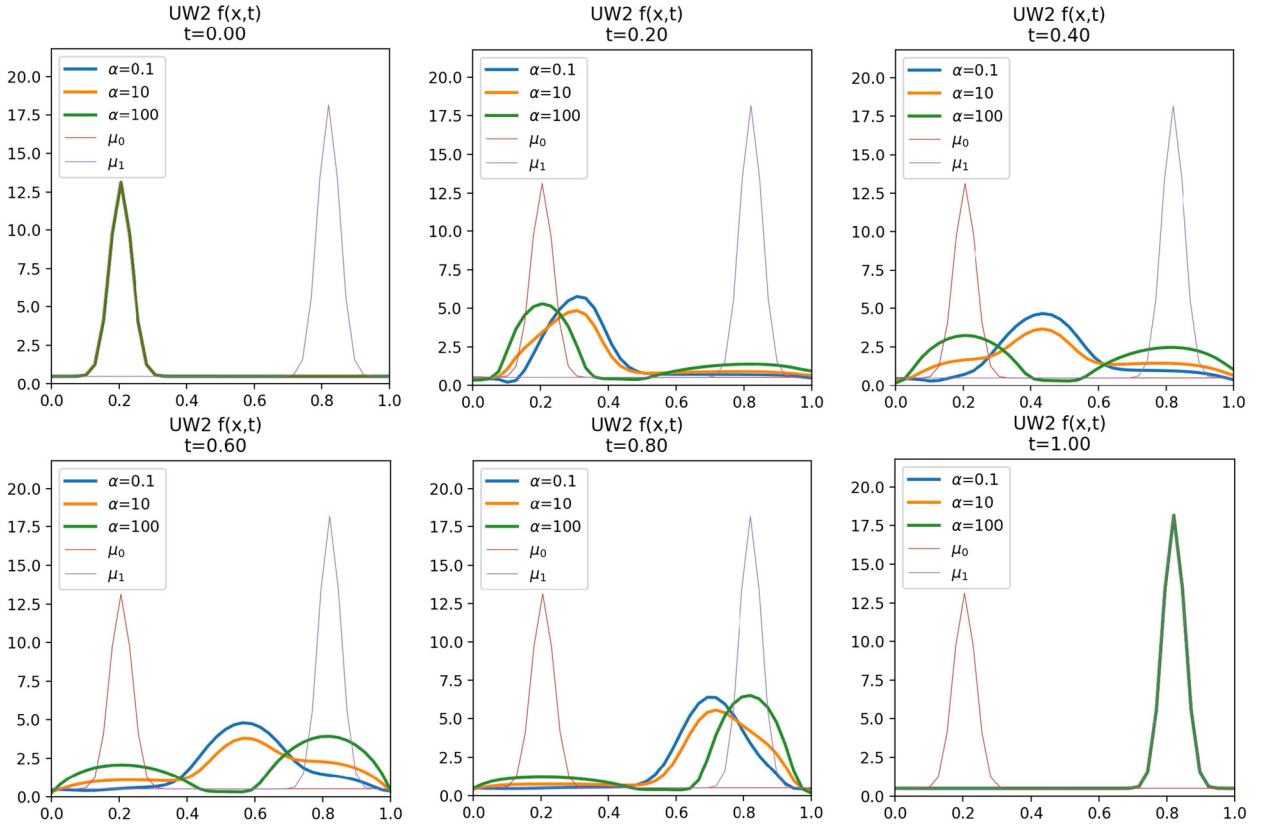
Fig. 1. *Experiment 1.* L^2 Unnormalized optimal transportation with a spatially dependent source function $f(t,x)$. The figures show the transportation of the densities from $t = 0$ (top left) to $t = 1$ (bottom right). Blue line shows $\alpha = 0.1$, orange line shows $\alpha = 10$, and green line shows $\alpha = 100$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

depend on $|\Omega|$. This actually provides an advantage of using the spatially dependent source function over the spatially independent source function when we need a consistent Wasserstein distance for any size of the domain.

4.1.3. Experiment 3

Consider a two dimensional problem with the following input values:

$$\mu_0 = N\left((x,y), (\frac{1}{3}, \frac{1}{3}), (\frac{\sqrt{2}}{20}, \frac{\sqrt{2}}{20})\right) + N\left((x,y), (\frac{2}{3}, \frac{1}{3}), (\frac{\sqrt{2}}{20}, \frac{\sqrt{2}}{20})\right)$$

$$\mu_1 = N\left((x,y), (\frac{2}{3}, \frac{2}{3}), (\frac{\sqrt{2}}{20}, \frac{\sqrt{2}}{20})\right)$$

where $N\left((x,y); (\mu_x, \mu_y), (\sigma_x^2, \sigma_y^2)\right) = C \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}\right)$ and C is a constant such that $\int_\Omega N((x,y); (\mu_x, \mu_y), (\sigma_x^2, \sigma_y^2))\mathrm{d}x\mathrm{d}y = 1$. Using the Algorithm 1, we calculate the minimizers of $UW_2(\mu_0, \mu_1)$ with a spatially dependent source function $f(t,x)$. The parameters are chosen as

$$N_x = 35, N_y = 35, N_t = 15, \tau = 0.1, \text{maximum iterations} = 6,000.$$

Fig. 4 shows the transportation with $\alpha = 1$ and $\alpha = 1000$, respectively. The same phenomena can be observed as in 1D case from *Experiment 1*. In other words, the geodesic with the spatially dependent source function with small α in Fig. 4 behaves closer to the normalized (classical) optimal transport geodesic and the geodesic with large α behaves closer to the Euclidean geodesic. The computation took 402.54 seconds for $\alpha = 1$ and 77.82 seconds for $\alpha = 1000$. Note that when α is small, the condition number of the Laplacian operator gets larger. This results in slower convergence rate of conjugate gradient method for inverting the Laplacian operator.
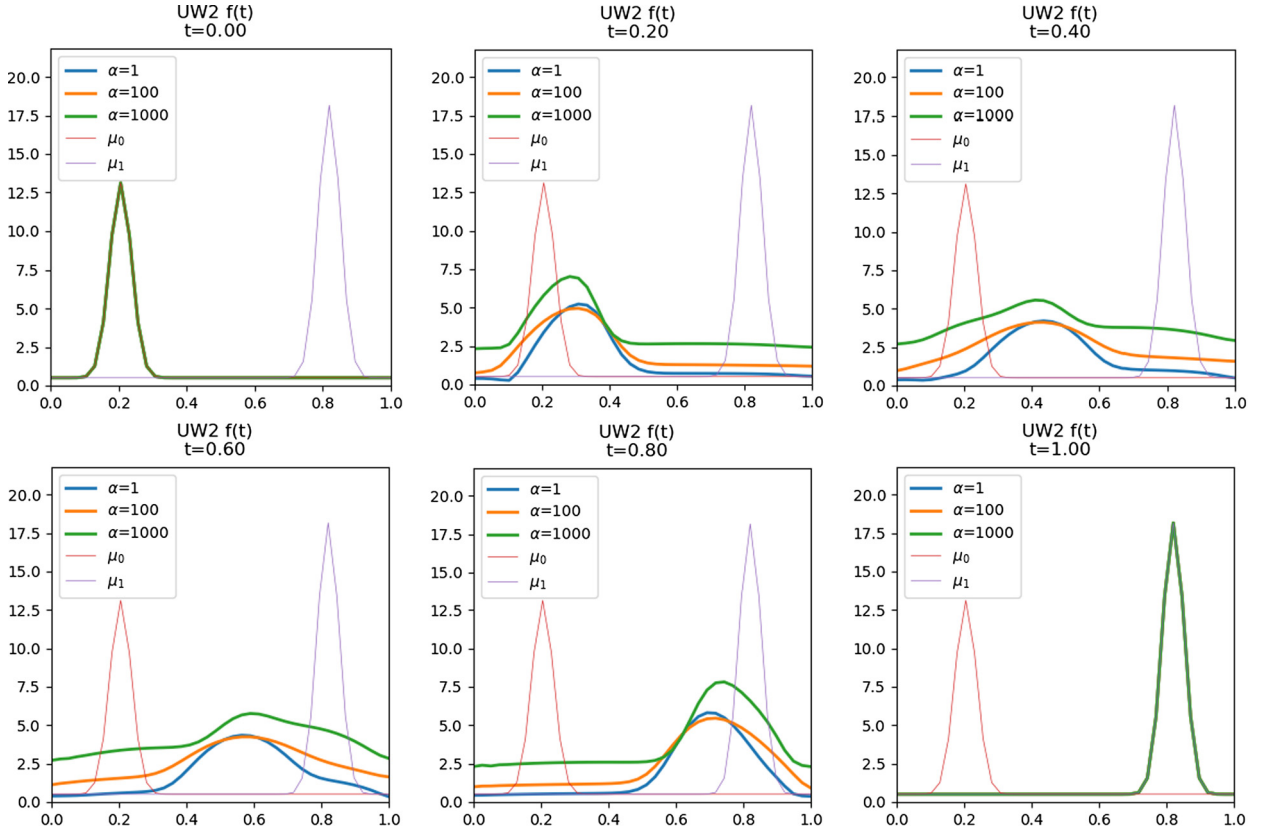
Fig. 2. *Experiment 1. L^2 Unnormalized optimal transportation with a spatially independent source function $f(t)$. The figures show the transportation of the densities from $t = 0$ (top left) to $t = 1$ (bottom right). Blue lines show $\alpha = 1$, orange lines show $\alpha = 100$, and green lines show $\alpha = 1000$.*
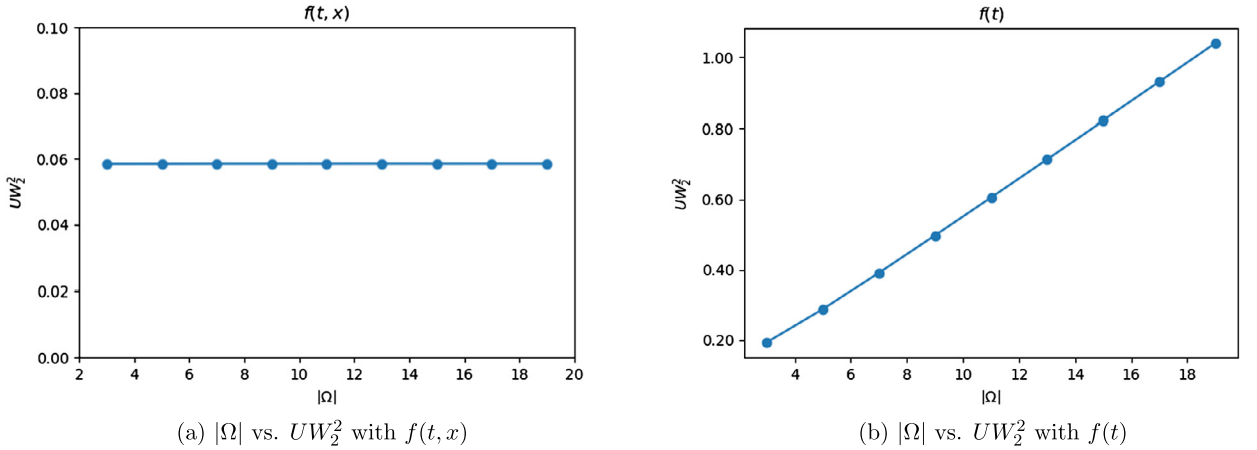

Fig. 3. *Experiment 2. The size of the domain $|\Omega|$ vs. L^2 unnormalized Wasserstein metrics for $f(t, x)$ and $f(t)$. x-axis represents $|\Omega|$ and y-axis represents $UW_2(\mu_0, \mu_1)^2$. Both $f(t, x)$ and $f(t)$ use $\alpha = 100$.*
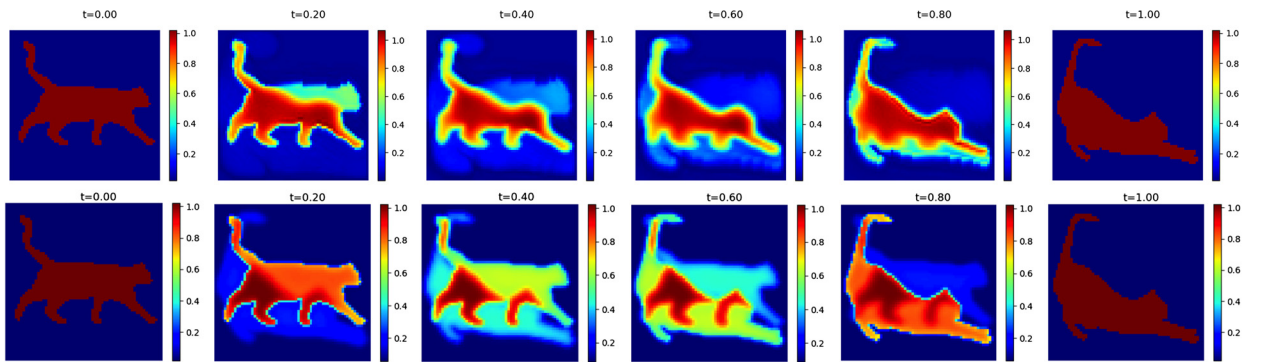
4.1.4. Experiment 4

In this experiment, we are interested in calculating L^2 unnormalized Wasserstein distance between two images. We show two sets of experiments with different initial and terminal densities. First, consider images of two cats with different total masses defined on the domain $\Omega = [0, 1] \times [0, 1]$. We use Algorithm 1 with the following parameters:

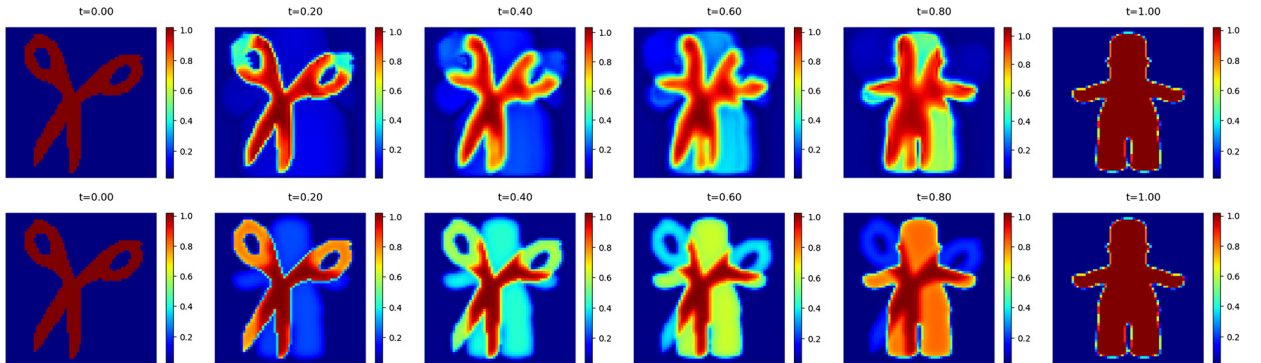$$N_x = 64, \ N_y = 64, \ N_t = 15, \text{ maximum iterations} = 4,000.$$

Fig. 5 shows transportation between two cats images with $\alpha = 10$ and $\alpha = 1000$, respectively. The computation took 353.85 seconds for $\alpha = 10$ and 93.67 seconds for $\alpha = 1000$.
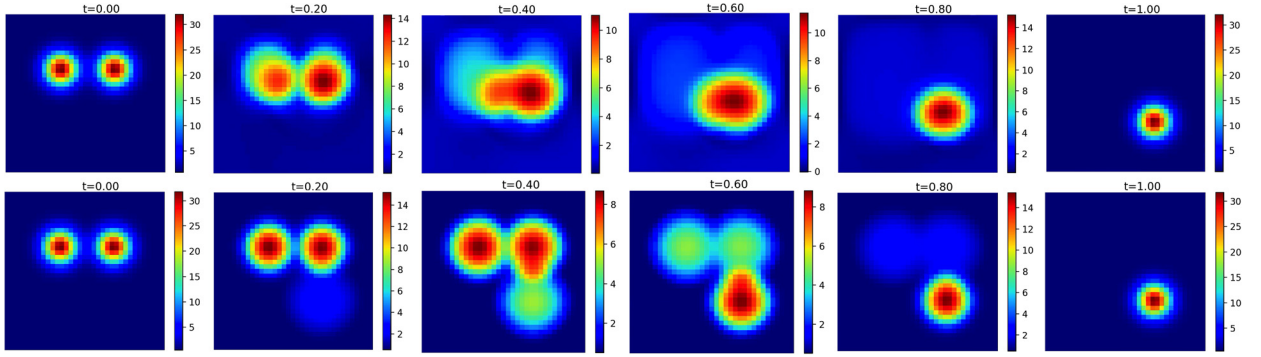
Fig. 4. *Experiment 3.* L^2 generalized unnormalized optimal transportation: 2D example with a spatially dependent source function $f(t,x)$. The first row is with $\alpha = 1$. The second row is with $\alpha = 1000$.

Fig. 5. *Experiment 4.* L^2 generalized unnormalized optimal transportation between two cats with a spatially dependent source function $f(t,x)$. The first row is with $\alpha = 10$. The second row is with $\alpha = 1000$.


Fig. 6. *Experiment 4.* L^2 generalized unnormalized optimal transportation between a pair of scissors and Homer Simpson with a spatially dependent source function $f(t,x)$. The first row is with $\alpha = 10$. The second row is with $\alpha = 1000$.

Additionally, we consider images of a pair of scissors and Homer Simpson. We again use Algorithm 1 with the same set of parameters as above. Fig. 6 shows transportation between two images with $\alpha = 10$ and $\alpha = 1000$, respectively. The computation took 353.00 seconds for $\alpha = 10$ and 93.96 seconds for $\alpha = 1000$.

4.2. Primal dual algorithm for UW_1

We conduct two numerical examples of L^1 unnormalized optimal transport using Algorithm 2. For UW_1 experiments, we use maximum iterations for the stopping condition.
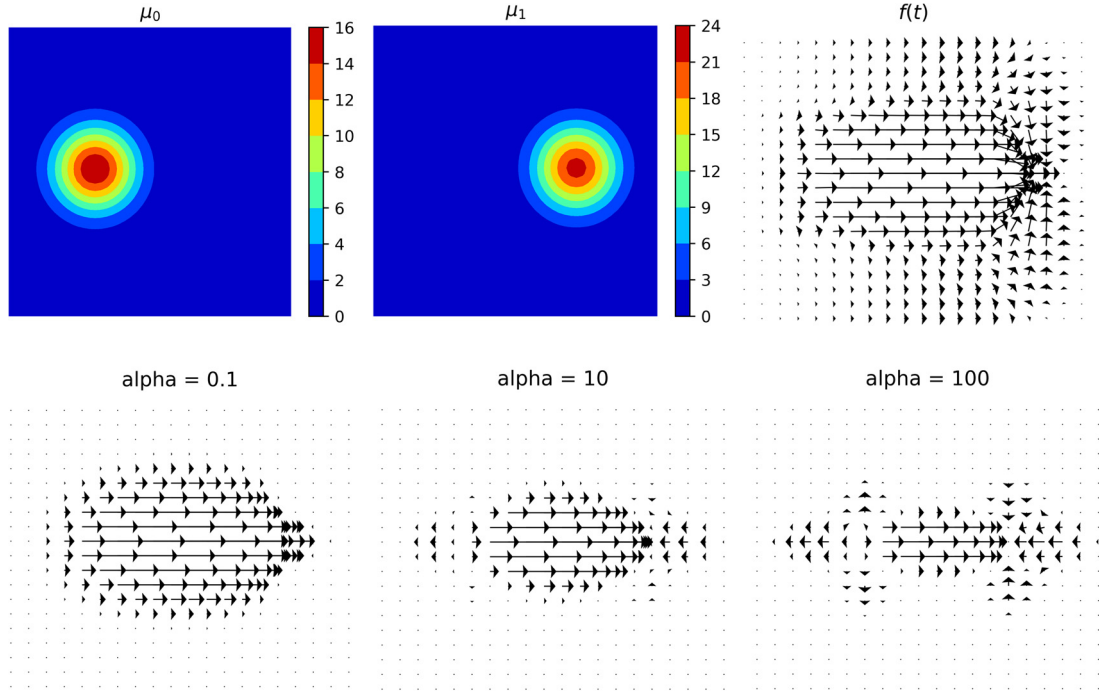
Fig. 7. *Experiment 5.* Top left: initial density μ_0. Top middle: the terminal density μ_1. Top right: the solution $\boldsymbol{m}$ of L^1 unnormalized optimal transportation with a spatially independent source $f(t)$. The bottom images show the solution of L^1 unnormalized optimal transportation with $f(t, x)$ using different α values. Bottom left: $\alpha = 0.1$, bottom middle: $\alpha = 10$, bottom right: $\alpha = 100$.

4.2.1. Experiment 5

Assume $\Omega = [0, 1] \times [0, 1]$. Consider the two dimensional problem with the following initial densities:

$$\mu_0 = N\left((x, y), (\frac{1}{3}, \frac{1}{2}), (\frac{1}{10}, \frac{1}{10})\right)$$

$$\mu_1 = N\left((x, y), (\frac{2}{3}, \frac{1}{2}), (\frac{1}{10}, \frac{1}{10})\right) \cdot 1.4$$

N is the same as the one used in Experiment 3. We chose the parameters as:

$$N_x = N_y = 40, \epsilon = 0.001, \lambda = 0.001, \tau = 0.1, \text{maximum iterations} = 30,000.$$

In Fig. 7, the initial densities μ_0 and μ_1 are shown on the top two plots and the minimizers $\boldsymbol{m}$'s are plotted for three different α values in the second row. As a comparison, the top right picture in Fig. 7 shows the result from L^1 transportation with a spatially independent source function $f(t)$. This experiment shows the clear difference between L^1 unnormalized optimal transport with $f(t, x)$ and with $f(t)$. While the minimizer $\boldsymbol{m}$ from the unnormalized optimal transport with $f(t, x)$ is nonzero only on the area between two densities, the minimizer from the unnormalized optimal transport with $f(t)$ is nonzero everywhere. This is because the spatially dependent source function $f(t, x)$ affects the minimizer locally but the spatially independent source function $f(t)$ affects the minimizer globally. The computation took 2.02 seconds for $\alpha = 1$, 2.74 seconds for $\alpha = 10$, and 2.67 seconds for $\alpha = 100$.
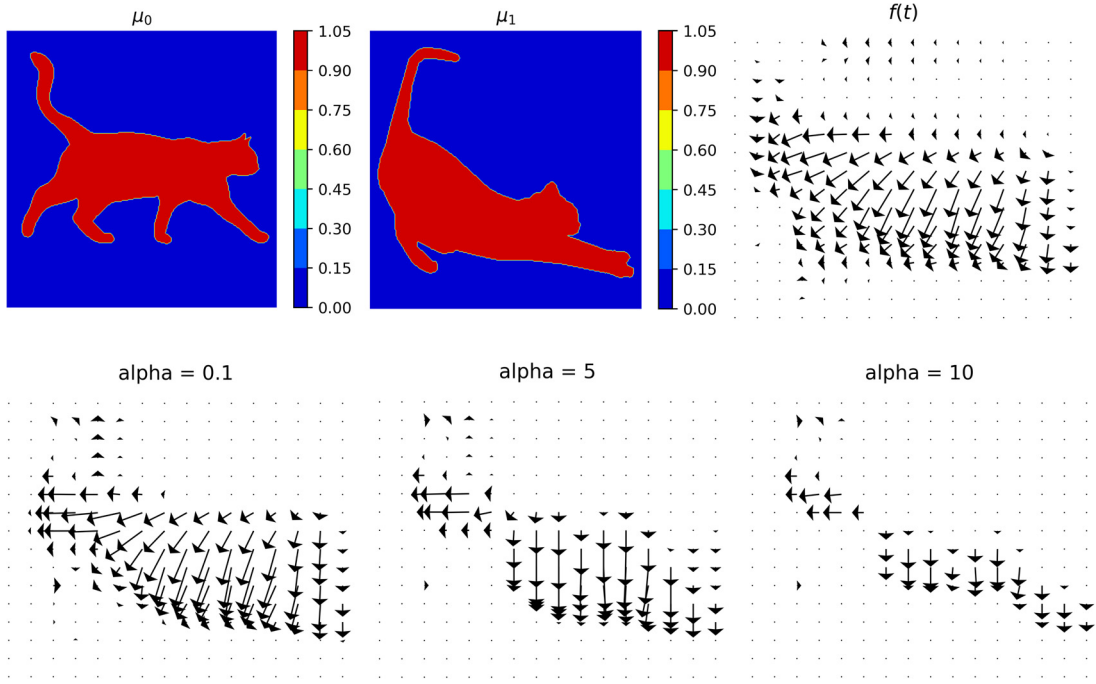
4.2.2. Experiment 6

In this experiment, we are interested in UW_1 distance between two images. Consider the same 2D example as in the *Experiment 4*. We use the Algorithm 2 with the following parameters:

$$N_x = N_y = 256, \epsilon = 0.001, \lambda = 0.0001, \tau = 0.01, \text{maximum iterations} = 40,000.$$

Fig. 8 plots the results of L^1 unnormalized optimal transport with a spatially dependent source function $f(t, x)$ with different α values 0.1, 5, and 10 in the second row. As a comparison, the top right picture in Fig. 8 shows the result from L^1 transportation with a spatially independent source function $f(t)$. The result is similar to the *Experiment 5*. The minimizer $\boldsymbol{m}$ from L^1 unnormalized optimal transport with $f(t)$ has nonzero values on the surrounding area of the two densities, but the minimizers from unnormalized optimal transport with $f(t, x)$ are zero on that surrounding area. The computation took 153.07 seconds for $\alpha = 1$, 192.82 seconds for $\alpha = 10$, and 186.27 seconds for $\alpha = 100$.

Fig. 8. *Experiment 6.* Top left: initial density μ_0. Top middle: the terminal density μ_1. Top right: the solution **m** of L^1 unnormalized optimal transportation with a spatially independent source $f(t)$. The bottom images show the solution of L^1 unnormalized optimal transportation with $f(t, x)$ using different α values. Bottom left: $\alpha = 0.1$, bottom middle: $\alpha = 5$, bottom right: $\alpha = 10$.
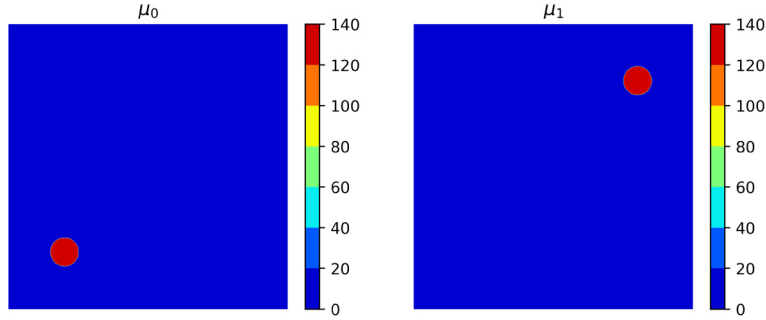
Fig. 9. Initial and terminal densities for Experiment 7.

4.2.3. Experiment 7

In this last experiment, we demonstrate the spatial convergence of Algorithm 2. Assume $\Omega = [0, 1] \times [0, 1]$. Consider an initial density to be a circle of radius 0.05 at $(0.2, 0.2)$ with a mass 1 and a terminal density to be a circle of radius 0.05 at $(0.8, 0.8)$ with a mass 1 (Fig. 9). UW_1 distance between these two densities is 0.8 which equals L^1 distance. We use the following parameters:

$$\alpha = 0.001, \epsilon = 0.01, \lambda = 0.001, \tau = 0.01, \text{maximum iterations} = 50,000.$$

We repeat the experiment with 4 different space discretizations ($N_x = N_y = 16, 32, 64, 128$). The table below summarizes the result of the experiment which shows the algorithm is accurate to order 1 in space.

N_x	N_y	Time (s)	Error
16	16	0.62	5.4×10^{-2}
32	32	2.39	3.8×10^{-2}
64	64	9.87	1.4×10^{-2}
128	128	44.14	9.4×10^{-3}

5. Discussion

In this paper, we introduced a new class of L^p generalized unnormalized optimal transport distance with a spatially dependent source function. We presented new fast algorithms for L^1 and L^2 generalized unnormalized optimal transport. For L^1 case, we derived the Kantorovich duality and used a primal-dual algorithm which has explicit formulas with low computational costs. For L^2 case, we derived the duality formula, the generalized unnormalized Monge problem and corresponding Monge-Ampère equation. We applied a weighted Laplacian operator L_μ to formulate the problem into an unconstrained optimization. The gradient operator of this unconstrained optimization is precisely the Hamilton-Jacobi equation. We apply the Nesterov accelerated gradient descent method to solve this minimization problem.

Our algorithm can be applied to general unnormalized/unbalanced optimal transport problems. It is also suitable for considering general variational mean-field games. In future works, we will derive new formulations for all related L^p unbalanced or unnormalized mean-field games and design fast numerical algorithms to solve them.

CRediT authorship contribution statement

Wonjun Lee: Software, Writing – original draft. **Rongjie Lai:** Conceptualization, Methodology. **Wuchen Li:** Conceptualization, Methodology. **Stanley Osher:** Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, arXiv:1701.07875 [cs, stat], 2017.
[2] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[3] J.-D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. 84 (3) (2000) 375–393.
[4] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vis. 40 (1) (2011) 120–145.
[5] L. Chayes, H.K. Lei, Transport and equilibrium in non-conservative systems, Adv. Differ. Equ. 23 (1/2) (2018) 1–64.
[6] Y. Chen, T. Georgiou, L. Ning, A. Tannenbaum, Matricial Wasserstein-1 distance, IEEE Control Syst. Lett. 1 (1) (04 2017).
[7] Y. Chen, T.T. Georgiou, A. Tannenbaum, Interpolation of matrices and matrix-valued densities: the unbalanced case, Eur. J. Appl. Math. 30 (3) (2019) 458–480.
[8] L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard, Unbalanced optimal transport: geometry and Kantorovich formulation, arXiv:1508.05216 [math], 2015.
[9] L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard, An interpolating distance between optimal transport and Fisher–Rao metrics, Found. Comput. Math. 18 (1) (2018) 1–44.
[10] Y.T. Chow, W. Li, S. Osher, W. Yin, Algorithm for Hamilton-Jacobi equations in density space via a generalized Hopf formula, arXiv:1805.01636 [math], 2018.
[11] B. Engquist, Y. Yang, Seismic inversion and the data normalization for optimal transport, arXiv:1810.08686 [math], 2018.
[12] W. Gangbo, W. Li, S. Osher, M. Puthawala, Unnormalized optimal transport, arXiv preprint arXiv:1902.03367, 2019.
[13] A. Garbuno-Inigo, F. Hoffmann, W. Li, A.M. Stuart, Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler, arXiv:1903.08866 [math], 2019.
[14] F. Léger, W. Li, Hopf-Cole transformation via generalized Schrödinger bridge problem, arXiv:1901.09051 [math], 2019.
[15] W. Li, Geometry of probability simplex via optimal transport, arXiv:1803.06360 [math], 2018.
[16] W. Li, E.K. Ryu, S. Osher, W. Yin, W. Gangbo, A parallel method for Earth mover's distance, J. Sci. Comput. 75 (1) (2018) 182–197.
[17] W. Li, P. Yin, S. Osher, Computations of optimal transport distance with Fisher information regularization, J. Sci. Comput. 75 (3) (2018) 1581–1595.
[18] M. Liero, A. Mielke, G. Savaré, Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures, Invent. Math. 211 (3) (2018) 969–1117.
[19] A.T. Lin, W. Li, S. Osher, G. Montufar, Wasserstein Proximal of GANs, 2018.
[20] J. Maas, M. Rumpf, C. Schönlieb, S. Simon, A generalized model for optimal transport of images including dissipation and density modulation, arXiv:1504.01988 [math], 2015.
[21] Y.E. Nesterov, A method for solving the convex programming problem with convergence rate o (1/kˆ 2), Dokl. Akad. Nauk SSSR 269 (1983) 543–547.
[22] N. Parikh, S. Boyd, Proximal algorithms, Found. Trends Optim. 1 (3) (2014) 127–239.
[23] G. Peyré, M. Cuturi, Computational optimal transport, arXiv:1803.00567 [stat], 2018.
[24] B. Piccoli, F. Rossi, Generalized Wasserstein distance and its application to transport equations with source, Arch. Ration. Mech. Anal. 211 (1) (2014) 335–358.
[25] B. Piccoli, F. Rossi, On properties of the generalized Wasserstein distance, Arch. Ration. Mech. Anal. 222 (3) (2016) 1339–1365.
[26] E.K. Ryu, W. Li, P. Yin, S. Osher, Unbalanced and partial L1 Monge–Kantorovich problem: a scalable parallel first-order method, J. Sci. Comput. (2017) 1–18.
[27] M. Thorpe, S. Park, S. Kolouri, G.K. Rohde, D. Slepuažev, A transportation L^p distance for signal analysis, J. Math. Imaging Vis. 59 (2) (Oct. 2017) 187–210.
[28] C. Villani, Optimal Transport: Old and New, Grundlehren der mathematischen Wissenschaften., vol. 338, Springer, Berlin, 2009.
[29] Y. Yang, B. Engquist, J. Sun, B.F. Hamfeldt, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion, Geophysics 83 (1) (2018) R43–R62.