



# Unnormalized optimal transport <sup>☆</sup>

Wilfrid Gangbo, Wuchen Li <sup>\*</sup>, Stanley Osher, Michael Puthawala

Mathematics Department, University of California, Los Angeles, United States of America



## ARTICLE INFO

### Article history:

Received 8 February 2019

Received in revised form 6 August 2019

Accepted 6 September 2019

Available online 18 September 2019

### Keywords:

Optimal transport

Unnormalized density space

Unnormalized Monge-Ampère equation

## ABSTRACT

We propose an extension of the computational fluid mechanics approach to the Monge-Kantorovich mass transfer problem, which was developed by Benamou-Brenier in [4]. Our extension allows optimal transfer of unnormalized and unequal masses. We obtain a one-parameter family of simple modifications of the formulation in [4]. This leads us to a new Monge-Ampère type equation and a new Kantorovich duality formula. These can be solved efficiently by, for example, the Chambolle-Pock primal-dual algorithm [6]. This solution to the extended mass transfer problem gives us a simple metric for computing the distance between two unnormalized densities. The  $L_1$  version of this metric was shown in [25] (which is a precursor of our work here) to have desirable properties.

© 2019 Elsevier Inc. All rights reserved.

## 1. Introduction

Optimal transport (OT) plays important roles in inverse problems [11,29] and machine learning [1,14,21]. It provides a particular distance function, called the Wasserstein metric or Earth Mover's distance, among histograms or density functions [4,28]. In these traditional settings, it assumes that histograms or densities have the same total mass. In real applications, we face a situation where the total mass of each histogram is not equal. For example, when comparing two images, their intensities are not the same. This fact prevents us from applying the classical optimal transport for these problems.

In this paper, we formulate simple and natural extensions of optimal transport in unnormalized density space. In a word, we add a spatial independent source function into the continuity equation and cost functional. There are two benefits of the current approach. On the one hand, the changes of the variational problem are simple. They define a robust  $L^p$  Wasserstein metric in unnormalized density space and do not significantly change the computational complexity of the problem. The proposed model allows us to apply classical algorithms, such as the Chambolle-Pock primal-dual method [6], to solve it. On the other hand, the proposed problem is natural in that it uses the key Hamilton-Jacobi equation as in the original optimal transport problem. These properties allow us to identify new problems corresponding to the Monge problem and Monge-Ampère equation in unnormalized density space.

There have been various extensions of optimal transport for unnormalized or unbalanced densities [2,3,8,5,12,13,20,23,24,26,27]. In particular, [8,9,20,17] propose the Wasserstein-Fisher-Rao or Hellinger-Kantorovich metric.<sup>1</sup> In their studies, a spatially dependent source function is introduced, which contains a ratio function involving the density in the spatial do-

<sup>☆</sup> The research is supported by AFOSR MURI FA9550-18-1-0502.

<sup>\*</sup> Corresponding author.

E-mail addresses: [wgangbo@math.ucla.edu](mailto:wgangbo@math.ucla.edu) (W. Gangbo), [wcli@math.ucla.edu](mailto:wcli@math.ucla.edu) (W. Li), [sjo@math.ucla.edu](mailto:sjo@math.ucla.edu) (S. Osher), [mputhawala@math.ucla.edu](mailto:mputhawala@math.ucla.edu) (M. Puthawala).

<sup>1</sup> In the literature, the Wasserstein-Fisher-Rao metric is called unbalanced OT. To distinguish with their approaches, we call our approach unnormalized OT.

main. Its gradient flow relates to the associated reaction diffusion equations [17]. This leads to rather complicated numerical algorithms and intricate analysis. In addition, [7] and [22] study other spatially dependent source functions. Instead, we propose a density and spatial independent source function, which allow us to keep the original Hamilton-Jacobi equation as in the normalized case. This property allows us to design a simple algorithm and to derive a reasonably simple unnormalized Monge-Ampère equation. The numerical algorithms in [10] involve the entropic regularization of Wasserstein-Fisher-Rao metric. It is based on a related linear programming formulation. This formulation comes from a new Hopf-Lax formula for the solution of a Hamilton-Jacobi equation. The entropic regularization is then added into the relaxed linear programming problem. In contrast to this approach, we apply a simple primal-dual algorithm based on a dynamical formulation of the proposed problem.

The plan of this paper is as follows. In section 2, we propose and study the properties of the unnormalized dynamical optimal transport problem. The unnormalized Monge problem, Monge-Ampère equation and Kantorovich formulations are all derived. In section 3, we present the algorithms and numerical examples for this proposed metric, including the simple unnormalized  $L^1$  case.

### 2. Unnormalized optimal transport

In this section, we introduce unnormalized OT problems and show that the proposed unnormalized metric is well defined. We then derive minimization procedures for unnormalized optimal transport.

Denote  $\Omega \subset \mathbb{R}^d$  as a bounded domain with Lebesgue measure  $|\Omega|$ . Denote the space of normalized densities by

$$\mathcal{P}(\Omega) = \{\mu \in L^1(\Omega) : \mu(x) \geq 0, \int_{\Omega} \mu(x)dx = 1\}.$$

Let the space of unnormalized densities be

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

We note that  $\mathcal{P}(\Omega) \subset \mathcal{M}(\Omega)$ . We next define the optimal transport cost between  $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$ .

**Definition 1** (Unnormalized OT). Define the  $L^p$  unnormalized Wasserstein distance  $UW_p : \mathcal{M}(\Omega) \times \mathcal{M}(\Omega) \rightarrow \mathbb{R}$  by

$$UW_p(\mu_0, \mu_1)^p = \inf_{v, \mu, f} \int_0^1 \int_{\Omega} \|v(t, x)\|^p \mu(t, x) dx dt + \frac{1}{\alpha} \int_0^1 |f(t)|^p dt \cdot |\Omega| \tag{1a}$$

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

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

Here  $\|\cdot\|$  is the Euclidean norm,  $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$ , and the infimum is taken over all continuous unnormalized density functions  $\mu : [0, 1] \times \Omega \rightarrow \mathbb{R}$ , and Borel vector fields  $v : [0, 1] \times \Omega \rightarrow \mathbb{R}^d$  with zero flux condition  $v(t, x) \cdot n(t, x) = 0$  on  $(0, 1) \times \partial\Omega$  with  $n(t, x)$  being the normal vector on the boundary of  $\Omega$ , and Borel spatially independent source functions  $f : [0, 1] \rightarrow \mathbb{R}$ .

The new proposed  $L^p$  Wasserstein metric has an attractive physical interpretation. The above optimization problem can be viewed as a variational fluid dynamics problem in Eulerian coordinates. Definition 1 considers the motion, creation and removal of particles. During this process, the total mass is changing dynamically in a uniform manner, controlled by the positive parameter  $\alpha$  and a spatially independent function  $f(t)$ . We remark that the spatial independence of the source function introduces a very important natural property, which we will repeat. It uses the same Hamilton-Jacobi equation as in the classical optimal transport, which allows us to obtain a new Monge problem, Monge-Ampère equation and Kantorovich duality problem. In addition, this physical analogy follows approaches in [18]. More interestingly, we notice that problem (1) has essentially the same computational complexity as the classical dynamical optimal transport problem. We will present its computational details in section 3.

#### 2.1. $L^1$ unnormalized Wasserstein metric

We first study the  $L^1$  unnormalized Wasserstein metric. When  $p = 1$ , the problem (1a) becomes:

$$UW_1(\mu_0, \mu_1) = \inf_{v(t,x), f(t)} \left\{ \int_0^1 \int_{\Omega} \|v(t, x)\| \mu(t, x) dx dt + \frac{1}{\alpha} \int_0^1 |f(t)| dt \cdot |\Omega| : \partial_t \mu(t, x) + \nabla \cdot (\mu(t, x)v(t, x)) = f(t), \mu(0, x) = \mu_0(x), \mu(1, x) = \mu_1(x) \right\}.$$

Denote

$$m(x) = \int_0^1 v(t, x)\mu(t, x)dt,$$

then by Jensen’s inequality, the above minimization can be obtained into a time independent fashion. Notice the fact that

$$\int_0^1 \int_{\Omega} \|v(t, x)\|\mu(t, x)dxdt \geq \int_{\Omega} \int_0^1 \|v(t, x)\mu(t, x)\|dx = \int_{\Omega} \|m(x)\|dx.$$

By integrating the time variable in the constraint, we observe

$$\begin{aligned} & \left\{ \int_0^1 \int_{\Omega} \|v(t, x)\|\mu(t, x)dxdt + \frac{1}{\alpha} \int_0^1 |f(t)|dt \cdot |\Omega| : \right. \\ & \left. \partial_t \mu(t, x) + \nabla \cdot (\mu(t, x)v(t, x)) = f(t), \mu(0, x) = \mu_0(x), \mu(1, x) = \mu_1(x) \right\} \\ & \geq \left\{ \int_{\Omega} \|m(x)\|dx + \frac{1}{\alpha} \int_0^1 |f(t)|dt \cdot |\Omega| : \mu_1(x) - \mu_0(x) + \nabla \cdot m(x) = \int_0^1 f(t)dt \right\} \\ & \geq \left\{ \int_{\Omega} \|m(x)\|dx + \frac{1}{\alpha} \left| \int_0^1 f(t)dt \right| \cdot |\Omega| : \mu_1(x) - \mu_0(x) + \nabla \cdot m(x) = \int_0^1 f(t)dt \right\}. \end{aligned}$$

Denote  $c = \int_0^1 f(t)dt$ , by integrating on both time and spatial domain for continuity equation (1b), it is clear that

$$c = \frac{1}{|\Omega|} \left( \int_{\Omega} \mu_1(z)dz - \int_{\Omega} \mu_0(z)dz \right).$$

We can show that the minimizer path can be attained in the last inequality, by choosing  $\mu(t, x) = (1 - t)\mu_0(x) + t\mu_1(x)$ . We summarize the above into the following proposition.

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

$$\begin{aligned} UW_1(\mu_0, \mu_1) = \inf_m \left\{ \int_{\Omega} \|m(x)\|dx + \frac{1}{\alpha} \left| \int_{\Omega} (\mu_0(x) - \mu_1(x))dx \right| : \right. \\ \left. \mu_1(x) - \mu_0(x) + \nabla \cdot m(x) = \frac{1}{|\Omega|} \int_{\Omega} (\mu_1(z) - \mu_0(z))dz \right\}. \end{aligned}$$

In addition, in one space dimension on the interval  $\Omega = [0, 1]$ , the  $L^1$  unnormalized Wasserstein metric has the following explicit solution:

$$\begin{aligned} UW_1(\mu_0, \mu_1) = \int_{\Omega} \left| \int_0^x (\mu_1(y) - \mu_0(y))dy - x \int_{\Omega} (\mu_1(z) - \mu_0(z))dz \right| dx \\ + \frac{1}{\alpha} \left| \int_{\Omega} (\mu_1(z) - \mu_0(z))dz \right|. \end{aligned}$$

The formulation in Proposition 2 has been proposed in [25] for inverse problems. It is one of the prime motivations for this paper. We also note the minimizer satisfies the following form [19]:

$$\begin{cases} \frac{m(x)}{\|m(x)\|} = \nabla \Phi(x), & \text{if } \|m(x)\| \neq 0 \\ \nabla \cdot m(x) = \left( \mu_0(x) - \frac{1}{|\Omega|} \int_{\Omega} \mu_0(z)dz \right) - \left( \mu_1(x) - \frac{1}{|\Omega|} \int_{\Omega} \mu_1(z)dz \right). \end{cases}$$

**Proposition 3.** The  $L^1$  unnormalized Wasserstein metric is given in terms of the original Wasserstein-1 metric by:

$$UW_1(\mu_0, \mu_1) = W_1(\tilde{\mu}_0, \tilde{\mu}_1) + \frac{|\Omega|}{\alpha} |c|,$$

where  $W_1$  is the classical Wasserstein-1 metric between two unnormalized densities and

$$c = \frac{1}{|\Omega|} \int_{\Omega} (\mu_1(z) - \mu_0(z)) dz$$

$$\tilde{\mu}_0 = (\mu_1 - \mu_0 - c)_-$$

$$\tilde{\mu}_1 = (\mu_1 - \mu_0 - c)_+.$$

Notice  $\tilde{\mu}_1 - \tilde{\mu}_0 = \mu_1 - \mu_0 - c$ . Then from Proposition 2, we prove the result. See details in [25].

2.2.  $L^2$  unnormalized Wasserstein metric

We next present the result when  $p = 2$ . Similar derivations can also be established for  $p \in (1, \infty)$ . For the simplicity of presentation, we now assume  $|\Omega| = 1$ .

**Proposition 4.** The  $L^2$  unnormalized Wasserstein metric (1) is a well-defined metric function in  $\mathcal{M}(\Omega)$ . In addition, the minimizer  $(v(t, x), \mu(t, x), f(t))$  for problem (1) satisfies

$$v(t, x) = \nabla \Phi(t, x), \quad f(t) = \alpha \int_{\Omega} \Phi(t, x) dx,$$

and

$$\begin{cases} \partial_t \mu(t, x) + \nabla \cdot (\mu(t, x) \nabla \Phi(t, x)) = \alpha \int_{\Omega} \Phi(t, x) dx \\ \partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 \leq 0 \\ \mu(0, x) = \mu_0(x), \quad \mu(1, x) = \mu_1(x). \end{cases} \tag{2}$$

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

$$\partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 = 0. \tag{3}$$

**Remark 5.** We note that equation (2) implies

$$\alpha \int_0^1 \int_{\Omega} \Phi(t, x) dx dt = \int_{\Omega} \mu_1(y) dy - \int_{\Omega} \mu_0(y) dy.$$

This means that unlike the classical OT, we are not only solving for the unique  $\nabla \Phi$ , but also for **the unique**  $\Phi$ .

**Remark 6.** Our  $UW_2(\mu_0, \mu_1)$  involving  $f(t)$  as a source term does have the property that the distance between measures depends linearly on the area of the domain. We will address this issue in a future work.

**Proof.** Denote  $m(t, x) = \mu(t, x)v(t, x)$  and

$$F(m, \mu) = \begin{cases} \frac{\|m\|^2}{\mu} & \text{if } \mu > 0; \\ 0 & \text{if } \mu = 0, m = 0; \\ +\infty & \text{otherwise,} \end{cases}$$

then variational problem (1) can be reformulated as

$$\begin{aligned}
 UW_2(\mu_0, \mu_1)^2 = \inf_{m, \mu, f} \left\{ \int_0^1 \int_{\Omega} F(m(t, x), \mu(t, x)) dx dt + \frac{1}{\alpha} \int_0^1 |f(t)|^2 dt : \right. \\
 \left. \begin{aligned}
 \partial_t \mu(t, x) + \nabla \cdot (m(t, x)) &= f(t), \\
 \mu(0, x) &= \mu_0(x), \quad \mu(1, x) = \mu_1(x) \} \right.
 \end{aligned}
 \tag{4}$$

It is clear that (4) is the reformulation of (1). We first prove that the variational problem (4) is well defined. In other words, there exists a feasible path for the dynamical constraint. We construct a feasible path  $\mu_t$  connecting any  $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$ . The proof is divided into three steps:

Step 1. Construct a density path  $t \in [0, \frac{1}{3}]$ , there exists a feasible path connecting  $\mu_0$  and a uniform measure with total mass  $\int_{\Omega} \mu_0 dx$ . In this case, the density path is a normalized (classical) OT between two densities. We set  $f(t) = 0$  when  $t \in [0, 1/3]$ , there always exists such a path.

Step 2. Construct a density path  $t \in [\frac{1}{3}, \frac{2}{3}]$ , there exists a feasible path connecting a uniform measure with total mass  $\int_{\Omega} \mu_0 dx$  and a uniform measure with total mass  $\int_{\Omega} \mu_1 dx$ . In this case, we let the transport flux  $m(t, x) = 0$ , and choose  $f(t) = 3(\int_{\Omega} \mu^1(x) dx - \int_{\Omega} \mu^0(x) dx)$ .

Step 3. Construct a density path  $t \in [\frac{2}{3}, 1]$ , there exists a feasible path connecting a uniform measure with total mass  $\int_{\Omega} \mu_1 dx$  and  $\mu_1$ . In this case, we set  $f(t) = 0$ . Following the classical OT, we find a feasible path.

Combining steps 1,2,3, the proposed path is feasible with finite cost functional. We next show that the problem has a minimizer. Since the constraint set is not empty, then it is classical to show that the cost functional  $F(m, \mu) + \frac{1}{\alpha} f(t)^2$  is convex and is lower semicontinuous, while the constraint is linear. So the variational problem (2) has a minimizer.

We next apply a Lagrange multiplier to find the minimizer. Denote  $\Phi(t, x)$  as the multiplier with

$$\mathcal{L}(m, \mu, \Phi) = \int_0^1 \int_{\Omega} \frac{\|m(t, x)\|^2}{2\mu(t, x)} + \Phi(t, x) \left( \partial_t \mu(t, x) + \nabla \cdot m(t, x) - f(t) \right) dx dt + \frac{1}{2\alpha} \int_0^1 f(t)^2 dt.$$

Assuming  $\delta_m \mathcal{L} = 0, \delta_{\mu} \mathcal{L} \geq 0, \delta_f \mathcal{L} = 0$ , we derive the property of minimizer as follows:

$$\begin{cases} \frac{m(t, x)}{\mu(t, x)} = \nabla \Phi(t, x) \\ -\frac{m(t, x)^2}{2\mu(t, x)^2} - \partial_t \Phi(t, x) \geq 0 \\ f(t) = \alpha \int_{\Omega} \Phi(t, x) dx. \end{cases}$$

Here if  $\mu > 0$ , we obtain  $\delta_{\mu} \mathcal{L} = 0$ , which gives equality in the second formula of the above system. Using the fact  $\frac{m(t, x)}{\mu(t, x)} = \nabla \Phi(t, x)$ , we prove the result. In this case, the non-negativity, symmetry, triangle inequality of the metric follow directly from the definition.  $\square$

We next derive our new Monge problem for unnormalized OT. This approach uses the Lagrange coordinates arising in problem (1).

**Proposition 7** (Unnormalized Monge problem).

$$\begin{aligned}
 UW_2(\mu_0, \mu_1)^2 = \inf_{M, f(t)} \int_{\Omega} \|M(x) - x\|^2 \mu_0(x) dx + \frac{1}{\alpha} \int_0^1 f(t)^2 dt \\
 + \int_0^1 \int_0^t f(s) \int_{\Omega} \|M(x) - x\|^2 \text{Det}(s \nabla M(x) + (1-s)\mathbb{I}) ds dx dt,
 \end{aligned}
 \tag{5a}$$

where the infimum is among all one to one, invertible mapping functions  $M: \Omega \rightarrow \Omega$  and a source function  $f: [0, 1] \rightarrow \mathbb{R}$ , such that the unnormalized push forward relation holds

$$\mu(1, M(x)) \text{Det}(\nabla M(x)) = \mu(0, x) + \int_0^1 f(t) \text{Det}(t \nabla M(x) + (1-t)\mathbb{I}) dt.
 \tag{5b}$$

**Proof.** We now derive the Lagrange formulation of the unnormalized OT (1). Consider any mapping function  $X_t(x)$  with vector field  $v(t, X_t(x))$ , i.e.

$$\frac{d}{dt} X_t(x) = v(t, X_t(x)), \quad X_0(x) = x.$$

Then

$$\begin{aligned} \int_0^1 \int_{\Omega} \|v(t, x)\|^2 \mu(t, x) dx dt &= \int_0^1 \int_{\Omega} \|v(t, X_t(x))\|^2 \mu(t, X_t(x)) dX_t(x) dt \\ &= \int_0^1 \int_{\Omega} \left\| \frac{d}{dt} X_t(x) \right\|^2 \mu(t, X_t(x)) \text{Det}(\nabla X_t(x)) dx dt. \end{aligned} \quad (6)$$

We next derive the differential equation for  $J(t, x) := \mu(t, X_t(x)) \text{Det}(\nabla X_t(x))$ . Later on, we use the notation  $J(t) = J(t, x)$  and  $\frac{d}{dt} J(t) = \frac{\partial}{\partial t} J(t, x)$ . Since

$$\begin{aligned} \frac{d}{dt} J(t, x) &= \frac{d}{dt} \left\{ \mu(t, X_t(x)) \text{Det}(\nabla X_t(x)) \right\} \\ &= \partial_t \mu(t, X_t(x)) \text{Det}(\nabla X_t(x)) + \nabla_X \mu(t, X_t(x)) \frac{d}{dt} X_t(x) \text{Det}(\nabla X_t(x)) \\ &\quad + \mu(t, X_t(x)) \partial_t \text{Det}(\nabla X_t(x)) \\ &= \left\{ \partial_t \mu + \nabla \mu \cdot v + \nabla \cdot v \mu \right\} (t, X_t(x)) \text{Det}(\nabla X_t(x)) \\ &= \left\{ \partial_t \mu + \nabla \cdot (\mu v) \right\} (t, X_t(x)) \text{Det}(\nabla X_t(x)) \\ &= f(t) \text{Det}(\nabla X_t(x)), \end{aligned}$$

where the third equality is derived by the Jacobi identity, i.e.

$$\partial_t \text{Det}(\nabla X_t(x)) = \nabla \cdot v(t, X_t(x)) \text{Det}(\nabla X_t(x)),$$

and the last equality holds following our proposed continuity equation with spatial independent source function (1b).

Notice

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

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

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

Since the minimizer in Eulerian coordinates satisfies the Hamilton-Jacobi equation in (3):

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

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

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

thus  $X_t(x) = (1-t)x + tM(x)$  and  $\text{Det}(\nabla X_t(x)) = \text{Det}((1-t)\mathbb{I} + t\nabla M(x))$ .

Substituting all the above relations into (6):

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

Thus we prove the results.  $\square$

We next find the relation between the spatial independent source function  $f(t)$  and the mapping function  $M(x)$ . For the simplicity of presentation, we assume periodic boundary conditions on  $\Omega$ .

**Proposition 8** (Unnormalized Monge-Ampère equation). *The optimal mapping function  $M(x) = \nabla \Psi(x)$  satisfies the following unnormalized Monge-Ampère equation*

$$\begin{aligned}
 &\mu(1, \nabla \Psi(x)) \text{Det}(\nabla^2 \Psi(x)) - \mu(0, x) \\
 &= \alpha \int_0^1 \text{Det}(t\nabla^2 \Psi(x) + (1-t)\mathbb{I}) \int_{\Omega} \left( \Psi(y) - \frac{\|y\|^2}{2} + \frac{t\|\nabla \Psi(y) - y\|^2}{2} \right) \text{Det}(t\nabla^2 \Psi(y) + (1-t)\mathbb{I}) dy dt.
 \end{aligned}$$

**Proof.** Let us rewrite the minimizer (2) into a time independent formulation. From the Hopf-Lax formula for the Hamilton-Jacobi equation,

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

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

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

and

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

From (2) and the above two formulas, then

$$\begin{aligned}
 f(t) &= \alpha \int_{\Omega} \Phi(t, x) dx = \alpha \int_{\Omega} \Phi(t, X_t(x)) dX_t(x) \\
 &= \alpha \int_{\Omega} \Phi(t, X_t(x)) \text{Det}(\nabla X_t(x)) dx
 \end{aligned}$$

$$= \alpha \int_{\Omega} \left\{ \Psi(x) - \frac{\|x\|^2}{2} + \frac{t \|\nabla \Psi(x) - x\|^2}{2} \right\} \text{Det} \left( (1-t)\mathbb{I} + t \nabla^2 \Psi(x) \right) dx.$$

Substituting  $f(t)$ 's formula and  $M(x) = \nabla \Psi(x)$  into (5b), we derive the result.  $\square$

We now present the Kantorovich duality formulation of the problem (1).

**Proposition 9** (Unnormalized Kantorovich formulation).

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

where the supremum is taken among all  $\Phi: [0, 1] \rightarrow \Omega$  satisfying

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

**Proof.** As in [15,16], we derive the duality formula by integration by parts as follows. Notice the fact that

$$\begin{aligned} & \frac{1}{2} UW_2(\mu_0, \mu_1)^2 \\ &= \inf_{m, \mu, f} \left\{ \int_0^1 \int_{\Omega} \frac{m(t, x)^2}{2\mu(t, x)} dx dt + \frac{1}{2\alpha} \int_0^1 f(t)^2 dt : \partial_t \mu + \nabla \cdot m = 0, \mu(0, x) = \mu_0(x), \mu(1, x) = \mu_1(x) \right\} \\ &= \inf_{m, \mu, f} \sup_{\Phi} \left\{ \int_0^1 \int_{\Omega} \frac{m(t, x)^2}{2\mu(t, x)} + \frac{1}{2\alpha} f(t)^2 + \Phi(t, x) \left( \partial_t \mu(t, x) + \nabla \cdot m(t, x) - f(t) \right) dx dt \right\} \\ &\geq \sup_{\Phi} \inf_{m, \mu, f} \left\{ \int_0^1 \int_{\Omega} \frac{m(t, x)^2}{2\mu(t, x)} + \frac{1}{2\alpha} f(t)^2 + \Phi(t, x) \left( \partial_t \mu(t, x) + \nabla \cdot m(t, x) - f(t) \right) dx dt \right\} \\ &= \sup_{\Phi} \inf_{m, \mu, f} \left\{ \int_0^1 \int_{\Omega} \frac{m(t, x)^2}{2\mu(t, x)} - \nabla \Phi(t, x) \cdot m(t, x) + \frac{1}{2\alpha} f(t)^2 + \Phi(t, x) \cdot \left( \partial_t \mu(t, x) - f(t) \right) dx dt \right\} \\ &= \sup_{\Phi} \inf_{m, \mu, f} \left\{ \int_0^1 \int_{\Omega} \frac{1}{2} \left( \frac{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) dx dt \right. \\ &\quad \left. + \int_{\Omega} \left( \Phi(1, x) \mu(1, x) - \Phi(0, x) \mu(0, x) \right) dx \right. \\ &\quad \left. + \int_0^1 \int_{\Omega} \left( -\mu(t, x) \partial_t \Phi(t, x) + \frac{1}{2\alpha} f(t)^2 - \Phi(t, x) f(t) \right) dx dt \right\} \\ &= \sup_{\Phi} \left\{ \int_{\Omega} \left( \Phi(1, x) \mu(1, x) - \Phi(0, x) \mu(0, x) \right) dx \right. \\ &\quad \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) dx dt \right. \\ &\quad \left. + \inf_f \int_0^1 \left( \int_{\Omega} \frac{1}{2\alpha} f(t)^2 - \Phi(t, x) f(t) \right) dx dt \right\} \\ &= \sup_{\Phi} \left\{ \int_{\Omega} \left( \Phi(1, x) \mu(1, x) - \Phi(0, x) \mu(0, x) \right) dx - \frac{1}{2\alpha} \int_0^1 \left( \alpha \int_{\Omega} \Phi(t, x) dx \right)^2 dt \right\} \end{aligned}$$



$$\begin{aligned}
 & + \inf_{\mu} \left\{ - \int_0^1 \int_{\Omega} \mu(t, x) \left( \partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 \right) dx dt \right\} \\
 & + \frac{1}{2\alpha} \inf_f \int_0^1 \left( f(t) - \alpha \int_{\Omega} \Phi(t, x) dx \right)^2 dt \Big\} \\
 = & \sup_{\Phi} \left\{ \int_{\Omega} \left( \Phi(1, x) \mu(1, x) - \Phi(0, x) \mu(0, x) \right) dx - \frac{\alpha}{2} \int_0^1 \left( \int_{\Omega} \Phi(t, x) dx \right)^2 dt : \right. \\
 & \left. \partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 \leq 0 \right\}.
 \end{aligned}$$

We have shown that the minimizer over  $m$  is obtained at  $\frac{m}{\mu} = \nabla \Phi$ , and  $f(t) = \alpha \int_{\Omega} \Phi(t, x) dx$ . The last equality holds because  $\mu(t, x) \geq 0$ , thus  $\partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 \leq 0$ .

We next show that the primal-dual gap is zero. From Proposition 4, the minimizer  $(\mu, \Phi)$  satisfies (2). Thus

$$\begin{aligned}
 & \int_0^1 \int_{\Omega} \frac{m(t, x)^2}{2\mu(t, x)} dx dt + \frac{1}{2\alpha} \int_0^1 f(t)^2 dt \\
 = & \int_0^1 \int_{\Omega} \frac{1}{2} \|\nabla \Phi(t, x)\|^2 \mu(t, x) dx dt + \frac{\alpha}{2} \int_0^1 \left( \int_{\Omega} \Phi(t, x) dx \right)^2 dt \\
 = & \int_0^1 \int_{\Omega} \left( -\frac{1}{2} \|\nabla \Phi(t, x)\|^2 \mu(t, x) + \|\nabla \Phi(t, x)\|^2 \mu(t, x) \right) dx dt + \frac{\alpha}{2} \int_0^1 \left( \int_{\Omega} \Phi(t, x) dx \right)^2 dt \\
 = & \int_0^1 \int_{\Omega} \partial_t \Phi(t, x) \mu(t, x) + \Phi(t, x) \left( -\nabla \cdot (\mu(t, x) \nabla \Phi(t, x)) \right) dx dt + \frac{\alpha}{2} \int_0^1 \left( \int_{\Omega} \Phi(t, x) dx \right)^2 dt \\
 = & \int_{\Omega} \Phi(1, x) \mu(1, x) dx - \int_{\Omega} \Phi(0, x) \mu(0, x) dx \\
 & - \int_0^1 \int_{\Omega} \Phi(t, x) \left( \partial_t \mu(t, x) + \nabla \cdot (\mu(t, x) \nabla \Phi(t, x)) \right) dx dt + \frac{\alpha}{2} \int_0^1 \left( \int_{\Omega} \Phi(t, x) dx \right)^2 dt \\
 = & \int_{\Omega} \Phi(1, x) \mu(1, x) dx - \int_{\Omega} \Phi(0, x) \mu(0, x) dx \\
 & - \int_0^1 \int_{\Omega} \Phi(t, x) f(t) dx dt + \frac{\alpha}{2} \int_0^1 \left( \int_{\Omega} \Phi(t, x) dx \right)^2 dt \\
 = & \int_{\Omega} \Phi(1, x) \mu(1, x) dx - \int_{\Omega} \Phi(0, x) \mu(0, x) dx + \left( -\alpha + \frac{\alpha}{2} \right) \int_0^1 \left( \int_{\Omega} \Phi(t, x) dx \right)^2 dt \\
 = & \int_{\Omega} \Phi(1, x) \mu(1, x) dx - \int_{\Omega} \Phi(0, x) \mu(0, x) dx - \frac{\alpha}{2} \int_0^1 \left( \int_{\Omega} \Phi(t, x) dx \right)^2 dt.
 \end{aligned}$$

This concludes the proof.  $\square$

### 3. The numerical method

In this section, we propose to apply a primal-dual algorithm to solve unnormalized OT numerically. We then provide several numerical examples to demonstrate the effectiveness of this procedure.

3.1. Algorithm

We present a primal-dual algorithm for problem (1). In particular, our method is based on its reformulation (4), named the minimal flux problem. Define the Lagrangian of (4):

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

where  $\Phi(t, x)$  is the Lagrange multiplier of the unnormalized continuity equation (1b).

Convex analysis shows that  $(m^*(t, x), \mu^*(t, x), f^*(t))$  is a solution to (4) if and only if there is a  $\Phi^*$  such that  $(m^*, \Phi^*)$  is a saddle point of  $\mathcal{L}(m, \mu, f, \Phi)$ . In other words, we can compute minimization (4) by solving the following minimax problem

$$\inf_{m, \mu, f} \sup_{\Phi} \mathcal{L}(m, \mu, f, \Phi).$$

It is clear that  $\mathcal{L}$  is convex in  $m, \mu, f$  and concave in  $\Phi$ , and the interaction term is a linear operator. This property allows us to apply the Chambolle-Pock first order primal-dual algorithm [6]. We choose to use the blockwise update as follows:

$$\left\{ \begin{aligned} m^{k+1}(t, x) &= \arg \inf_m \mathcal{L}(m, \mu^k, f^k, \Phi^k) + \frac{1}{2\tau_1} \int_0^1 \int_{\Omega} \|m(t, x) - m^k(t, x)\|^2 dx dt \\ \mu^{k+1}(t, x) &= \arg \inf_{\mu \geq 0} \mathcal{L}(m^k, \mu, f^k, \Phi^k) + \frac{1}{2\tau_1} \int_0^1 \int_{\Omega} \|\mu(t, x) - \mu^k(t, x)\|^2 dx dt \\ f^{k+1}(t) &= \arg \inf_f \mathcal{L}(m^k, \mu^k, f, \Phi^k) + \frac{1}{2\tau_1} \int_0^1 \|f(t) - f^k(t)\|^2 dt \\ \tilde{\Phi}^{k+1}(t, x) &= \arg \sup_{\Phi} \mathcal{L}(\tilde{m}, \tilde{\mu}, \tilde{f}, \Phi) - \frac{1}{2\tau_2} \int_0^1 \int_{\Omega} \|\Phi(t, x) - \Phi^k(t, x)\|^2 dx dt \\ (\tilde{m}, \tilde{\mu}, \tilde{f}) &= 2(m^{k+1}, \mu^{k+1}, f^{k+1}) - (m^k, \mu^k, f^k) \end{aligned} \right. \tag{8}$$

where  $\tau_1, \tau_2$  are given step sizes for primal, dual variables. These steps can be interpreted as a gradient descent in the primal variable  $(m, \mu, f)$  and a gradient ascent in the dual variable  $\Phi$ .

It turns out that the optimizations in above update (8) have explicit formulas. The first line becomes

$$m^{k+1}(t, x) = \arg \inf_m \left\{ \frac{\|m(t, x)\|^2}{2\mu^k(t, x)} - m(t, x) \cdot \nabla \Phi^k(t, x) + \frac{1}{2\tau_1} \|m(t, x) - m^k(t, x)\|^2 \right\} = \frac{\mu^k(t, x)}{\mu^k(t, x) + \tau_1} (\tau_1 \nabla \Phi^k(t, x) + m^k(t, x)).$$

The second line of (8) simplifies to

$$\mu^{k+1}(t, x) = \arg \inf_{\mu \geq 0} \frac{\|m^k(t, x)\|^2}{2\mu(t, x)} - \partial_t \Phi^k(t, x) \cdot \mu(t, x) + \frac{1}{2\tau_1} |\mu(t, x) - \mu^k(t, x)|^2.$$

The above problem has an analytical solution by solving a cubic equation. The third line of (8) gives

$$f^{k+1}(t) = \arg \inf_f \left\{ \frac{1}{2\alpha} f(t)^2 - f(t) \int_{\Omega} \Phi^k(t, x) dx + \frac{1}{2\tau_1} \|f(t) - f^k(t)\|^2 \right\} = \frac{\alpha}{\alpha + \tau_1} \left( \tau_1 \int_{\Omega} \Phi^k(t, x) dx + f^k(t) \right).$$

The fourth line of (8) gives

$$\begin{aligned} \Phi^{k+1}(t, x) &= \arg \sup_{\Phi} \left\{ \Phi(t, x) \cdot (\partial_t \tilde{\mu}(t, x) + \nabla \cdot \tilde{m}(t, x) - \tilde{f}(t)) - \frac{1}{2\tau_2} \|\Phi(t, x) - \Phi^k(t, x)\|^2 \right\} \\ &= \Phi^k(t, x) + \tau_2 \left( \partial_t \tilde{\mu}^{k+1}(t, x) + \nabla \cdot \tilde{m}(t, x) - \tilde{f}(t) \right). \end{aligned}$$

Combining all above formulas, we are now ready to state the algorithm.

---

**Algorithm: Primal-Dual method for Unnormalized OT**

**Input:** Unnormalized densities  $\mu_0, \mu_1$ ;  
 Initial guess of  $m^0, \mu^0, \Phi^0, f^0$ , step size  $\tau_1, \tau_2$ .  
**Output:** Minimizer  $\mu(t, x)$ ; Dual variable  $\Phi(t, x)$ ; Value  $UW_2(\mu_0, \mu_1)$ .

---

1. **For**  $k = 1, 2, \dots$  Iterate until convergence
  2.  $m^{k+1}(t, x) = \frac{\mu^k(t, x)}{\mu^k(t, x) + \tau_1} \left( \tau_1 \nabla \Phi^k(t, x) + m^k(t, x) \right)$ ;
  3. Solve  $\mu^{k+1}(t, x) = \arg \inf_{\mu \geq 0} \frac{\|m^k(t, x)\|^2}{2\mu(t, x)} - \partial_t \Phi^k(t, x) \cdot \mu(t, x) + \frac{1}{2\tau_1} |\mu(t, x) - \mu^k(t, x)|^2$ ;
  4.  $f^{k+1}(t) = \frac{\alpha}{\alpha + \tau_1} \left( \tau_1 \int_{\Omega} \Phi^k(t, x) dx + f^k(t) \right)$ ;
  5.  $\Phi^{k+1}(t, x) = \Phi^k(t, x) + \tau_2 \left( \partial_t \tilde{\mu}^{k+1}(t, x) + \nabla \cdot \tilde{m}(t, x) - \tilde{f}(t) \right)$ ;
  6.  $(\tilde{m}, \tilde{\mu}, \tilde{f}) = 2(m^{k+1}, \mu^{k+1}, f^{k+1}) - (m^k, \mu^k, f^k)$ ;
  7. **end**
- 

**Remark 10.** We note that the above approach is not the classical primal-dual method. The classical method requires us to minimize  $(m, \mu)$  simultaneously, in which a cubic equation in term of  $\frac{m}{\mu}$  needs to be solved. Here we mainly implement a block relaxation version and perform a sequential minimization by updating the values of  $(m, \mu)$ . In each iteration, we make sure that  $(m, \mu)$  lies in the domain of minimization problem (4).

3.2. Numerical grid

To apply the algorithm, we first define our numerical grid. For simplicity we consider the case where the space of interest is  $\Omega = [0, 1]^d$  and time  $\mathcal{T} = [0, 1]$ . Further, for the following explanations we consider the problem when  $d = 2$ , however, our grid construction can be constructed on any dimension by extending it in the obvious way. We will use the same symbol to represent both the continuous  $\mu, m, \Phi, f$  and their respective discretized counterparts, as the difference between the two should be clear from context alone.

Let  $n_t, n_x$ , and  $n_y$  be given then notate  $\Delta t = \frac{1}{n_t - 1}$ ,  $\Delta x = \frac{1}{n_x - 1}$ , and  $\Delta y = \frac{1}{n_y - 1}$ . Using this notation we define the following sets:

$$\begin{aligned} \Omega_{(i,j)} &= [i\Delta x, (i+1)\Delta x] \times [j\Delta y, (j+1)\Delta y] \\ \mathcal{T}_{(k)} &= [k\Delta t, (k+1)\Delta t] \end{aligned}$$

$$\Omega_{(i-1/2,j)} = [(i-1/2)\Delta x, (i+1/2)\Delta x] \times [j\Delta y, (j+1)\Delta y] \text{ for } i = 0, \dots, n_x$$

$$\Omega_{(i,j-1/2)} = [i\Delta x, (i+1)\Delta x] \times [(j-1/2)\Delta y, (j+1/2)\Delta y] \text{ for } j = 0, \dots, n_y$$

where  $i = 0, \dots, n_x - 1, j = 0, \dots, n_y - 1$ , and  $k = 0 \dots, n_t - 1$  unless otherwise specified.

For the discretized problem we consider a  $f_{(k)}$  that is constant along each  $\mathcal{T}_{(k)}$ , and consider  $\mu_{(k,i,j)}$  and  $\Phi_{(k,i,j)}$  that are constant along each  $\mathcal{T}_{(k)} \times \Omega_{(i,j)}$ . The vector  $m_{(k,i,j)}$  has two components  $m_{x,(k,i-1/2,j)}$  and  $m_{y,(k,i,j-1/2)}$ , that are constant along  $\mathcal{T}_{(k)} \times \Omega_{(i-1/2,j)}$  and  $\mathcal{T}_{(k)} \times \Omega_{(i,j-1/2)}$  respectively. Numerically  $m$  quantifies the movement of density between each of the  $\Omega_{(i,j)}$  and its spacial neighbors (i.e.  $\Omega_{(i-1,j)}, \Omega_{(i,j-1)}, \Omega_{(i+1,j)}$ , and  $\Omega_{(i,j+1)}$ ) and so it is natural to define the components of  $m$  not on  $\Omega_{(i,j)}$  but rather on  $\Omega_{(i-1/2,j)}, \Omega_{(i+1/2,j)}, \Omega_{(i,j-1/2)}$  and  $\Omega_{(i,j+1/2)}$ .

Using the above notation, we write the steps of the algorithm as:

$$\begin{aligned} m_{x,(k,i-1/2,j)} &= \begin{cases} \frac{\mu_{(k,i-1,j)} + \mu_{(k,i-1,j)}}{\mu_{(k,i,j)} + \mu_{(k,i-1,j)} + 2\tau_1} \left( \tau_1 + \nabla_x \Phi_{(k,i-1/2,j)} + m_{x,(k,i-1/2,j)} \right) & \text{if } i = 1, \dots, n_x - 1 \\ 0 & \text{if } i = 0, n_x \end{cases} \\ m_{y,(k,i,j-1/2)} &= \begin{cases} \frac{\mu_{(k,i,j)} + \mu_{(k,i,j-1)}}{\mu_{(k,i,j)} + \mu_{(k,i,j-1)} + 2\tau_1} \left( \tau_1 + \nabla_y \Phi_{(k,i,j-1/2)} + m_{y,(k,i,j-1/2)} \right) & \text{if } j = 1, \dots, n_y - 1 \\ 0 & \text{if } j = 0, n_y \end{cases} \end{aligned}$$

**Table 1**  
Numerical parameters for our experiments. Note that for our one dimensional experiments,  $n_y$  has no value.

Parameter	Value	Parameter	Value
Discretization		Optimization	
$n_t$	15	Iterations	200,000
$n_x$	35	$\tau_1$	$10^{-3}$
$n_y$	35	$\tau_2$	$10^{-1}$
		$\alpha$	100

$$\begin{aligned} \mu_{(k,i,j)} &= \text{root}^+(1, -(\tau_1 * \partial_t \Phi_{(k,i,j)} + \mu_{(k,i,j)}), 0, \\ &\quad - \frac{\tau_1}{8} \left( (m_{(k,i+1/2,j)} + m_{(k,i-1/2,j)})^2 + (m_{(k,i,j+1/2)} + m_{(k,i,j-1/2)})^2 \right)) \\ f^{(k)} &= \frac{\alpha}{\alpha + \tau_1} \left( \tau_1 + \sum_i \sum_j \Phi_{(k,i,j)} \Delta x \Delta y + f^{(k)} \right) \\ \Phi_{(k,i,j)} &= \tau_2 * \left( \partial_t \tilde{\mu}_{(k,i,j)} + \nabla \cdot \tilde{m}_{(k,i,j)} - \tilde{f}^{(k)} \right) + \Phi_{(k,i,j)} \end{aligned}$$

where

$$\begin{aligned} \nabla_x \Phi_{(k,i-1/2,j)} &= \frac{\Phi_{(k,i,j)} - \Phi_{(k,i-1,j)}}{\Delta x} \\ \nabla_y \Phi_{(k,i,j-1/2)} &= \frac{\Phi_{(k,i,j)} - \Phi_{(k,i,j-1)}}{\Delta y} \\ \partial_t \Phi_{(k,i,j)} &= \begin{cases} \frac{1}{\Delta t} \left( \frac{\Phi_{(1,i,j)}}{2} + \Phi_{(0,i,j)} \right) & \text{if } k = 0 \\ \frac{1}{\Delta t} \left( \frac{\Phi_{(2,i,j)}}{2} - \Phi_{(0,i,j)} \right) & \text{if } k = 1 \\ \frac{1}{2\Delta t} \left( \Phi_{(k+1,i,j)} - \Phi_{(k-1,i,j)} \right) & \text{if } 1 < k < n_t - 2 \\ \frac{1}{\Delta t} \left( \Phi_{(n_t-1,i,j)} - \frac{\Phi_{(n_t-3,i,j)}}{2} \right) & \text{if } k = n_t - 2 \\ \frac{1}{\Delta t} \left( -\Phi_{(n_t-1,i,j)} - \frac{\Phi_{(n_t-2,i,j)}}{2} \right) & \text{if } k = n_t - 1 \end{cases} \\ \text{root}^+(a, b, c, d) &= \text{the largest real solution to } ax^3 + bx^2 + cx + d = 0 \\ \partial_t \mu_{(k,i,j)} &= \begin{cases} \frac{1}{\Delta t} \left( \mu_{(1,i,j)} - \mu_{(0,i,j)} \right) & \text{if } k = 0 \\ \frac{1}{2\Delta t} \left( \mu_{(k+1,i,j)} - \mu_{(k-1,i,j)} \right) & \text{if } 0 < k < n_t - 1 \\ \frac{1}{\Delta t} \left( \mu_{(n_t-1,i,j)} - \mu_{(n_t-2,i,j)} \right) & \text{if } k = n_t - 1 \end{cases} \\ \nabla \cdot m_{(k,i,j)} &= \frac{m_{x,(k,i+1/2,j)} - m_{x,(k,i-1/2,j)}}{\Delta x} + \frac{m_{y,(k,i,j+1/2)} - m_{y,(k,i,j-1/2)}}{\Delta y} \end{aligned}$$

Note that the unusual boundary conditions of  $\partial_t \Phi$  arise from the need to satisfy

$$\sum_k \Phi_{(k,i,j)} \partial_t \mu_{(k,i,j)} \Delta t = - \sum_k \partial_t \Phi_{(k,i,j)} \mu_{(k,i,j)} \Delta t \quad \forall i, j.$$

### 3.3. Numerical experiments

Now we present our numerical results. The first two experiments are in one dimension, and the rest are in two dimension. The numerical parameters for our experiments are given in Table 1.

#### 3.4. Experiment 1

Here we consider the problem where  $\mu_0$  and  $\mu_1$  are both one dimensional Gaussians of equal integral,  $\Omega = [0, 1]$  and

$$\begin{aligned} \mu_0 &= N \left( x; \frac{1}{3}, 0.1 \right) \\ \mu_1 &= N \left( x; \frac{2}{3}, 0.1 \right) \end{aligned}$$

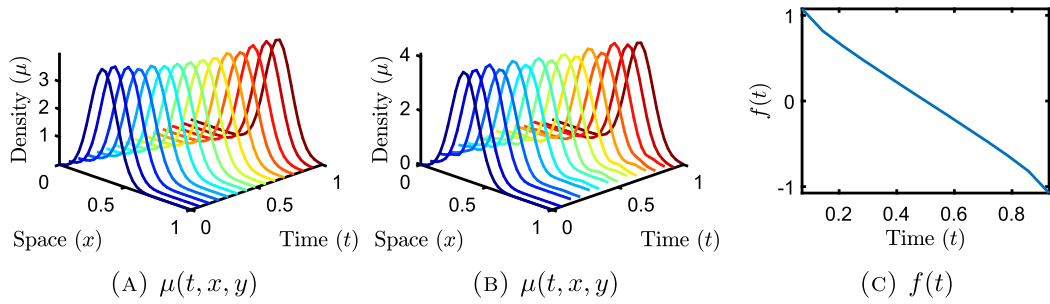


Fig. 1. A plot of (A)  $W_2(\mu_0, \mu_1)$ , (B)  $UW_2(\mu_0, \mu_1)$  and (C)  $f(t)$  in the unbalanced case.

$$N(x; \mu_x, \sigma^2) = Ce^{-\frac{(x-\mu_x)^2}{2\sigma^2}} \text{ where } C \text{ is such that } \int_{\Omega} N(x; \mu_x, \sigma^2) dx = 1$$

We plot the results in Fig. 1. In this case the input densities are balanced and so  $W_2(\mu_0, \mu_1)$  and  $UW_2(\mu_0, \mu_1)$  appear similar. Indeed  $UW_2(\mu_0, \mu_1) = 0.055$  and  $W_2(\mu_0, \mu_1) = 0.056$ .

Note that even in this simple case the behavior of  $f(t)$  is nuanced. In this case,  $\mu_0$  and  $\mu_1$  are smooth, of equal integral and  $W_2(\mu_0, \mu_1)$  is given by a simple analytical formula, and  $f(t)$  is not identically zero. Integrating Equation 1b in space and time yields  $|\Omega| \int_{[0,1]} f(t) dt = \int_{\Omega} \mu_1 dx - \int_{\Omega} \mu_0 dx$ , and so for balanced inputs  $\int_{[0,1]} f(t) dt = 0$ , but experiment 1 shows that  $f \neq 0$ .

### 3.5. Experiment 2

Again consider  $\Omega = [0, 1]$ , however in this experiment we analyze the asymptotic behavior of  $UW_2(\mu_0, \mu_1)$  as a function of  $\alpha$  and  $\alpha \rightarrow 0$  and  $\alpha \rightarrow \infty$ . Here

$$\begin{aligned} \mu_0 &= N(x; 0, 0.1) + N\left(x; \frac{1}{3}, 0.1\right) \\ \mu'_0 &= \frac{1}{2} \left( N(x; 0, 0.1) + N\left(x; \frac{1}{3}, 0.1\right) \right) \\ \mu_1 &= N\left(x; \frac{2}{3}, 0.1\right). \end{aligned}$$

The balanced case refers to  $UW_2(\mu'_0, \mu_1)$ , and the unbalanced refers to  $UW_2(\mu_0, \mu_1)$ . In both cases we compute the unnormalized Wasserstein distance. The results are given in Fig. 2.

Figs. 2a - 2c show that (at least numerically)  $UW_2(\mu_0, \mu_1; \alpha)$ ,  $f(t, \alpha)$  and  $\Phi(t, x; \alpha)$  converge as  $\alpha \rightarrow 0^+$ ,  $\alpha \rightarrow \infty$  when  $\int_{\Omega} \mu_0 dx = \int_{\Omega} \mu_1 dx$ . Further it seems plausible that for balanced inputs  $UW_2(\mu_0, \mu_1; \alpha) \rightarrow W_2(\mu_0, \mu_1)$  as  $\alpha \rightarrow 0^+$ . For any  $\alpha$  the  $\mu, m$  and  $\Phi$  from  $W_2(\mu_0, \mu_1)$  along with  $f(t) \equiv 0$  satisfy the constraint of Equation 1b. Formally sending  $\alpha \rightarrow 0^+$  causes  $f(t)$  to approach 0.

Figs. 2d - 2f illustrate the asymptotic behavior of  $UW_2(\mu_0, \mu_1; \alpha)$  w.r.t.  $\alpha$  when the inputs are unbalanced. In that case we (numerically) see that as  $\alpha \rightarrow 0$ ,  $f(t; \alpha)$  converges to a non-zero value, and both  $UW_2(\mu_0, \mu_1; \alpha)$  and  $\Phi(t, x; \alpha)$  diverge. This too is consistent with the formal argument that  $UW_2(\mu_0, \mu_1; \alpha) \rightarrow W_2(\mu_0, \mu_1)$  as  $\alpha \rightarrow 0^+$ .

In a predecessor of this work [4] the authors solve for  $W_2(\mu_0, \mu_1)$  using Lagrange multipliers in a similar formulation to equations (1a), (1b). In their work the Lagrange multiplier  $\Phi(t, x)$  is given up to an additive constant. If indeed  $UW_2(\mu_0, \mu_1; \alpha) \rightarrow W_2(\mu_0, \mu_1)$  as  $\alpha \rightarrow 0^+$  and  $\Phi(t, x; \alpha)$  does converge then  $\Phi(t, x; 0^+)$  is given uniquely (as a limit) and there is no issue of undetermined constants.

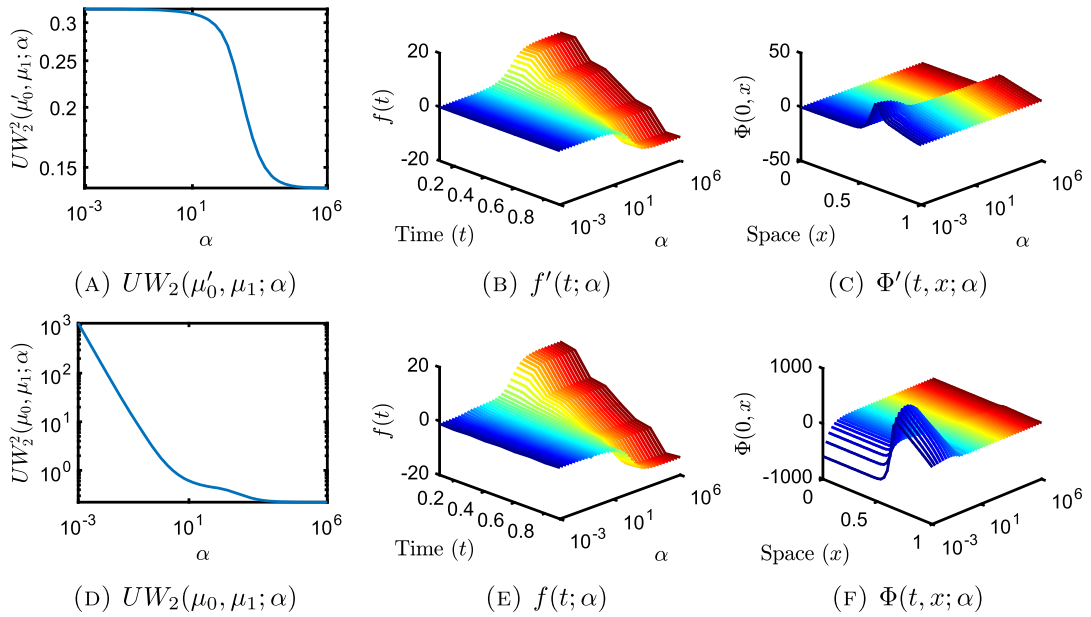
### 3.6. Experiment 3

Now consider the two dimensional problem where  $\Omega = [0, 1]^2$ . In this case

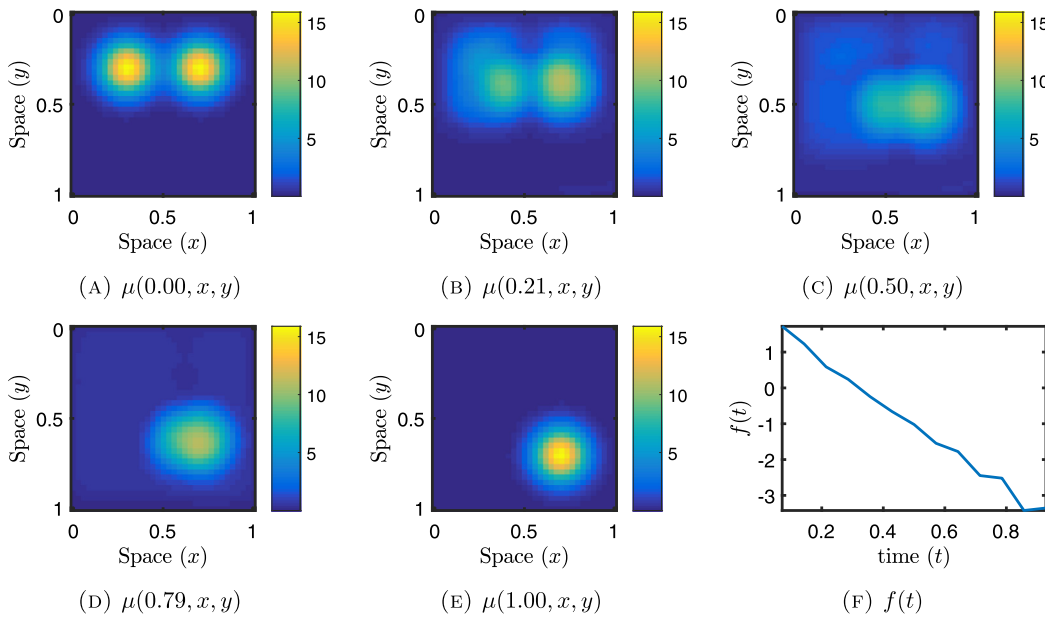
$$\begin{aligned} \mu_0(x, y) &= N(x, y; 0.3, 0.3, 0.1, 0.1) + N(x, y; 0.7, 0.3, 0.1, 0.1) \\ \mu_1(x, y) &= N(x, y; 0.7, 0.7, 0.1, 0.1) \end{aligned}$$

$$N(x, y; \mu_x, \mu_y, \sigma_1^2, \sigma_2^2) = Ce^{-\frac{(x-\mu_x)^2}{2\sigma_1^2} + \frac{(y-\mu_y)^2}{2\sigma_2^2}},$$

where  $C$  is a normalization constant such that  $\int_{\Omega} N(x, y; \mu_x, \mu_y, \sigma_1^2, \sigma_2^2) dx dy = 1$ . The results from our experiments are shown in Fig. 3. Note that although the mass of  $\mu_0$  is twice that of  $\mu_1$ , the optimal  $f(t)$  is not non-positive. Indeed from



**Fig. 2.** A plot of the asymptotic behavior of  $UW_2$  in  $\alpha$  with balanced and unbalanced inputs. Balanced: (A)  $UW_2(\mu'_0, \mu_1; \alpha)$ , (B)  $f'(t; \alpha)$ , (C)  $\Phi'(t, x; \alpha)$ , and unbalanced: (D)  $UW_2(\mu_0, \mu_1; \alpha)$ , (E)  $f(t; \alpha)$ , (F)  $\Phi(t, x; \alpha)$ .



**Fig. 3.** Plots of the  $\mu(t, x, y)$  and  $f(t)$  for  $UW_2(\mu_0, \mu_1)$ . (A)  $\mu(0.00, x, y)$ , (B)  $\mu(0.21, x, y)$ , (C)  $\mu(0.50, x, y)$ , (D)  $\mu(0.79, x, y)$ , (E)  $\mu(1.00, x, y)$ , (F)  $f(t)$ .

$t = 0$  to  $t \approx \frac{1}{4}$ ,  $f(t)$  is positive, before staying non-positive for the rest of the interval. This again illustrates that even in the case of Gaussian movement the behavior of  $f(t)$  is nuanced, and violates naive basic intuition.

### 3.7. Experiment 4

Consider again the two dimensional problem, however this time we choose  $\mu_0$  and  $\mu_1$  to be the cats in [19]. Our results are summarized in Fig. 4. This illustrates that our new method can be used as a general purpose OT solver for unbalanced inputs, and so can be used to interpolate between two functions.

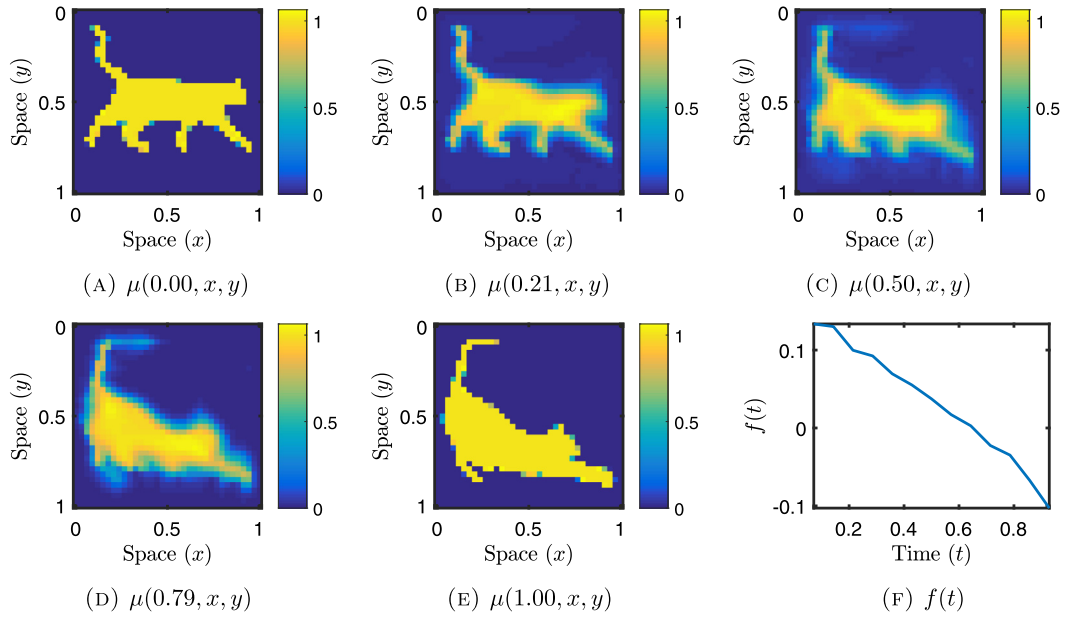


Fig. 4. Plots of the  $\mu(t, x, y)$  and  $f(t)$  for  $UW_2(\mu_0, \mu_1)$ . (A)  $\mu(0.00, x, y)$ , (B)  $\mu(0.21, x, y)$ , (C)  $\mu(0.50, x, y)$ , (D)  $\mu(0.79, x, y)$ , (E)  $\mu(1.00, x, y)$ , (F)  $f(t)$ .

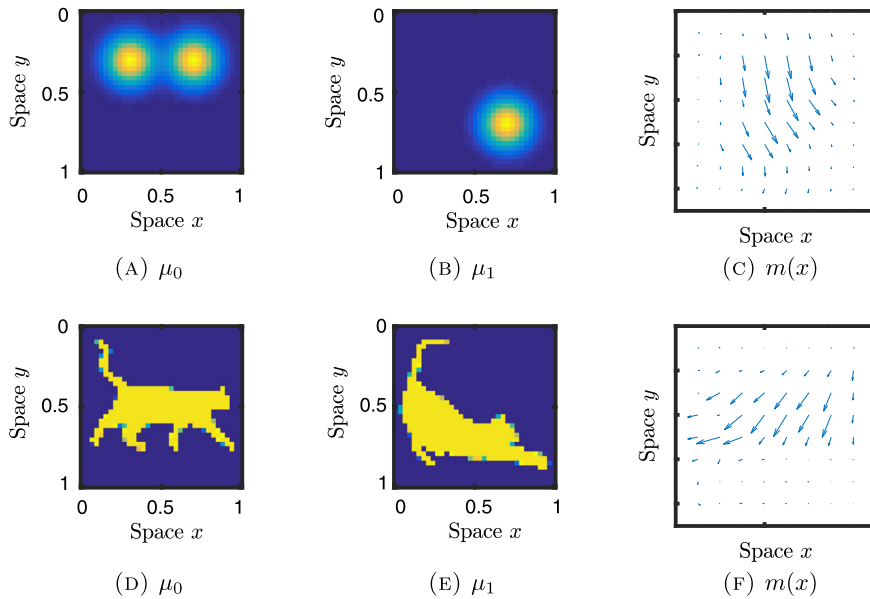


Fig. 5. Plots of the  $\mu_0, \mu_1$  and  $m(x)$  for  $UW_1(\mu_0, \mu_1)$  for the two gaussian movement (A)  $\mu_0$ , (B)  $\mu_1$ , (C)  $m(x)$  and two (D)  $\mu_0$ , (E)  $\mu_1$ , (F)  $m(x)$ .

### 3.8. $L^1$ unnormalized Wasserstein metric

In this subsection, we also present several numerical results for  $UW_1$  in Fig. 5. For the completeness of this paper, we present the details here. Let  $\Omega$  be a bounded domain. First  $\mu_0, \mu_1$  are given probability densities in  $\mathcal{P}(\Omega)$  with equal total mass, i.e.  $\int \mu_0 dx = \int \mu_1 dx = 1$ . The classical Wasserstein-1 distance has the minimal flux formulation as follows

$$W_1(\mu_0, \mu_1) = \inf_m \left\{ \int_{\Omega} \|m(x)\| dx : \nabla \cdot m(x) = \mu^0 - \mu^1 \right\},$$

where  $m$  is a flux function satisfying the zero flux condition on the boundary of  $\Omega$ .

As shown in Proposition 2, if  $\mu_0, \mu_1$  do not have the equal total mass, i.e.  $\int_{\Omega} \mu_0 dx \neq \int_{\Omega} \mu_1 dx$ , we modify the problem to

$$UW_1(\mu_0, \mu_1) = \inf_m \left\{ \int_{\Omega} \|m(x)\| dx + \frac{|c|}{\alpha} : \nabla \cdot m(x) = \mu_0 - \mu_1 + c \right\}, \tag{9}$$

where  $c = \frac{1}{|\Omega|} \left( \int_{\Omega} \mu_1(z) dz - \int_{\Omega} \mu_0(z) dz \right)$ . The only difference between above two minimizations is that we add a constant  $c$  into the constraint of the variational problem. The constant  $c$  just slightly modifies the algorithm.

Here we apply a primal-dual algorithm for computing the unnormalized Wasserstein-1 distance [19]. Define the Lagrangian

$$L(m, \Phi) = \int \|m(x)\| dx + \int \Phi(x) (\nabla \cdot m(x) + \mu^1(x) - \mu^0(x) - c) dx,$$

where  $\Phi: \Omega \rightarrow \mathbb{R}$  is the Lagrange multiplier corresponding to the equality constraint. We apply the first-order primal-dual method of Chambolle and Pock [6]. As in [19], the update for  $m, \Phi$  are as follows.

### Primal-Dual for Unnormalized Wasserstein-1 metric

**Input:** Discrete probabilities  $\mu^0, \mu^1$   
 Initial guess of  $m^0$ , step size  $\tau_1, \tau_2$   
**Output:** Optimal flux  $m$  and unnormalized Wasserstein-1 metric

- 1 Denote  $c = \frac{1}{|\Omega|} (\int \mu^1(x) dx - \int \mu^0(x) dx)$
- 2 **For**  $k = 1, 2, \dots$  (Iterate until convergence)
- 3  $m^{k+1}(x) = \text{shrink}(m^k(x) + \tau_1(\nabla \Phi^k)(x), \tau_1)$  for  $x \in \Omega$
- 4  $\Phi^{k+1}(x) = \Phi^k(x) + \tau_2((\nabla \cdot (2m^{k+1} - m^k))(x) + \mu^1(x) - \mu^0(x) - c)$  for  $x \in \Omega$
- 5 **end**

Here the shrink operator in line 3 yields

$$\text{shrink}(v, \tau) = \begin{cases} (1 - \tau/\|v\|)v & \text{for } \|v\| \geq \tau \\ 0 & \text{for } \|v\| < \tau. \end{cases}$$

### 4. Discussion

In this paper, we propose and solve an unnormalized optimal transport problem. We show that the proposed distance is well defined, and we obtain the minimizer using the same key Hamilton-Jacobi equation (3). More importantly, computing the  $L^p$  unnormalized Wasserstein metric has essentially the same computational complexity as the normalized one. In the future, we intend to study these related geometric properties and applications in inverse problems, machine learning and mean field games.

### 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.

### Acknowledgement

We would like to thank the reviewers for many suggestions to improve the paper. We thank Wonjun Lee for his help with several numerical experiments.

### References

- [1] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, arXiv:1701.07875 [cs, stat], 2017, cs, stat.
- [2] J. Barrett, L. Prigozhin, Partial L1 Monge–Kantorovich problem: variational formulation and numerical approximation, *Interfaces Free Bound.* (2009) 201–238.
- [3] J.-D. Benamou, Numerical resolution of an “unbalanced” mass transport problem, *ESAIM: Math. Model. Numer. Anal.* 37 (5) (2003) 851–868.
- [4] J.-D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, *Numer. Math.* 84 (3) (2000) 375–393.
- [5] L. Caffarelli, R. McCann, Free boundaries in optimal transport and Monge–Ampère obstacle problems, *Ann. Math.* 171 (2) (2010) 673–730.
- [6] 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.
- [7] L. Chayes, H.K. Lei, Transport and equilibrium in non-conservative systems, *Adv. Differ. Equ.* 23 (1/2) (2018) 1–64.



- [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] L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard, Scaling algorithms for unbalanced optimal transport problems, *Math. Comput.* (2018).
- [11] B. Engquist, Y. Yang, Seismic inversion and the data normalization for optimal transport, arXiv:1810.08686 [math], 2018.
- [12] A. Figalli, The optimal partial transport problem, *Arch. Ration. Mech. Anal.* 195 (2) (2010) 533–560.
- [13] A. Figalli, N. Gigli, A new transportation distance between non-negative measures, with applications to gradient flows with Dirichlet boundary conditions, *J. Math. Pures Appl.* 94 (2) (2010) 107–130.
- [14] C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo, T. Poggio, Learning with a Wasserstein Loss, arXiv:1506.05439, cs, stat, 2015.
- [15] W. Gangbo, An elementary proof of the polar factorization of vector-valued functions, *Arch. Ration. Mech. Anal.* 128 (4) (Dec 1994) 381–399.
- [16] W. Gangbo, R.J. McCann, The geometry of optimal transportation, *Acta Math.* 177 (2) (1996) 113–161.
- [17] S. Kondratyev, L. Monsaingeon, D. Vorotnikov, A new optimal transport distance on the space of finite radon measures, *Adv. Differ. Equ.* 21 (11–12) (2016) 1117–1164, 11.
- [18] W. Li, Geometry of probability simplex via optimal transport, arXiv:1803.06360 [math], 2018.
- [19] 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.
- [20] 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.
- [21] A.T. Lin, W. Li, S. Osher, G. Montufar, Wasserstein proximal of GANs, 2018.
- [22] 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.
- [23] B. Piccoli, F. Rossi, Generalized Wasserstein distance and its application to transport equations with source, *Arch. Ration. Mech. Anal.* 211 (1) (2014) 335–358.
- [24] B. Piccoli, F. Rossi, On properties of the generalized Wasserstein distance, *Arch. Ration. Mech. Anal.* 222 (3) (2016) 1339–1365.
- [25] M.A. Puthawala, C.D. Hauck, S.J. Osher, Diagnosing forward operator error using optimal transport, 2018.
- [26] E.K. Ryu, W. Li, P. Yin, S. Osher, Unbalanced and partial  $L^1$  Monge–Kantorovich problem: a scalable parallel first-order method, *J. Sci. Comput.* 75 (3) (2018) 1596–1613.
- [27] B. Schmitzer, B. Wirth, Dynamic models of Wasserstein-1-type unbalanced transport, arXiv:1705.04535 [math], 2017.
- [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.