

# CONSTRAINED DYNAMICAL OPTIMAL TRANSPORT AND ITS LAGRANGIAN FORMULATION

WUCHEN LI AND STANLEY OSHER

ABSTRACT. We propose dynamical optimal transport (OT) problems constrained in a parameterized probability subset. In applied problems such as deep learning, the probability distribution is often generated by a parameterized mapping function. In this case, we derive a formulation for the constrained dynamical OT.

## 1. INTRODUCTION

Dynamical optimal transport problems play vital roles in fluid dynamics [12] and mean field games [8]. They provide a type of statistical distance and interesting differential structures in the set of probability densities [14]. The full probability set is often intractable, when the dimension of sample space is large. For this reason, parameterized probability subsets have been widely considered, especially in machine learning problems and information geometry [1, 2, 4, 7, 11, 13]. We are interested in studying dynamical OT problems over a parameterized probability subset.

In this note, we follow a series of work found in [6, 9, 10], and introduce general constrained dynamical OT problems in a parametrized probability subset. As in deep learning [3], the probability subset is often constructed by a parameterized mapping. In these cases, we demonstrate that the constrained dynamical OT problems exhibit simple variational structures.

We arrange this note as follows: In section 2, we briefly review the dynamical OT in both Eulerian and Lagrangian coordinates<sup>1</sup>. Using Eulerian coordinates, we propose the constrained dynamical OT over a parameterized probability subset. In section 3, we next derive an equivalent Lagrangian formulation for the constrained problem.

---

*Key words and phrases.* Dynamical optimal transport; Mean field games; Information geometry; Machine learning.

The research is supported by AFOSR MURI proposal number 18RT0073.

<sup>1</sup>In fluid dynamics, the Eulerian coordinates represent the evolution of probability density function of particles, while the Lagrangian coordinates describe the motion of particles. In learning problems, the Eulerian coordinates naturally connect with the minimization problem in term of probability densities, while Lagrangian coordinate refers to the variational problem formulated in samples, whose analog are particles. “In learning, we model problems in Eulerian, and compute them in Lagrangian”. In other words, we often write the objective function in term of densities and compute them via samples.

## 2. CONSTRAINED DYNAMICAL OT

In this section, we briefly review the dynamical OT in a full probability set via both Eulerian and Lagrangian formalisms. Using the Eulerian coordinates, we propose the dynamical OT in a parameterized probability subset.

For the simplicity of exposition, all our derivations assume smoothness. Consider densities  $\rho^0, \rho^1 \in \mathcal{P}_+(\Omega) = \left\{ \rho(x) \in C^\infty(\Omega) : \rho(x) > 0, \int_\Omega \rho(x) dx = 1 \right\}$ , where  $\Omega$  is a  $n$ -dimensional sample space. Here  $\Omega$  can be  $\mathbb{R}^n$ , or a convex compact region in  $\mathbb{R}^n$  with zero flux conditions or periodic boundary conditions.

Dynamical OT studies a variational problem in density space, known as the Benamou-Brenier formula [5]. Given a Lagrangian function  $L: T\Omega \rightarrow [0, +\infty)$  under suitable conditions, consider

$$C(\rho^0, \rho^1) := \inf_{v_t} \int_0^1 \mathbb{E} L(X_t(\omega), v(t, X_t(\omega))) dt, \quad (1a)$$

where  $\mathbb{E}$  is the expectation operator over realizations  $\omega$  in event space and the infimum is taken over all vector fields  $v_t = v(t, \cdot)$ , such that

$$\dot{X}_t(\omega) = v(t, X_t(\omega)), \quad X_0 \sim \rho^0, \quad X_1 \sim \rho^1. \quad (1b)$$

Here  $X_i \sim \rho^i$  represents that  $X_i(\omega)$  satisfies the probability density  $\rho^i(x)$ , for  $i = 0, 1$ . Equivalently, denote the density of particles  $X_t(\omega)$  at space  $x$  and time  $t$  by  $\rho(t, x)$ . Then problem (1) refers to a variational problem in density space:

$$C(\rho^0, \rho^1) := \inf_{v_t} \int_0^1 \int_\Omega L(x, v(t, x)) \rho(t, x) dx dt, \quad (2a)$$

where the infimum is taken over all Borel vector fields  $v_t = v(t, \cdot)$ , such that the density function  $\rho(t, x)$  satisfies the continuity equation:

$$\frac{\partial \rho(t, x)}{\partial t} + \nabla \cdot (\rho(t, x) v(t, x)) = 0, \quad \rho(0, x) = \rho^0(x), \quad \rho(1, x) = \rho^1(x). \quad (2b)$$

Here  $\nabla \cdot$  is the divergence operator in  $\Omega$ . In the language of fluid dynamics, problem (1) refers to the Lagrangian formalism, while problem (2) is the associated Eulerian formalism; see details in [14]. The Lagrangian formalism focuses on the motion of each individual particles, while the Eulerian formalism describes the global behavior of all particles. Here (1) and (2) are equivalent since they represent the same variational problem using different coordinate systems. In addition, one often considers  $L(x, v) = \|v\|^p$ ,  $p \geq 1$ , where  $\|\cdot\|$  is the Euclidean norm. In this case, the optimal value of the variational problem defines a distance function in the set of probability space. Denote  $W_p(\rho^0, \rho^1) := C(\rho^0, \rho^1)^{\frac{1}{p}}$ , where  $W_p$  is called the  $L^p$ -Wasserstein distance.

We next study the variational problem (2) constrained on a parameterized probability density set. In other words, consider a parameter space  $\Theta \subset \mathbb{R}^d$  with

$$\mathcal{P}_\Theta = \left\{ \rho(\theta, x) \in C^\infty(\Omega) : \theta \in \Theta, \int_\Omega \rho(\theta, x) dx = 1, \rho(\theta, x) > 0, x \in \Omega \right\}.$$

Here we assume that  $\rho: \Theta \rightarrow \mathcal{P}_+(\Omega)$  is an injective mapping<sup>2</sup>. We introduce the constrained dynamical OT as follows:

$$c(\theta_0, \theta_1) := \inf_{v_t} \int_0^1 \int_{\Omega} L(x, v(t, x)) \rho(\theta_t, x) dx dt, \quad (3a)$$

where  $\theta_t = \theta(t) \in \Theta$ ,  $t \in [0, 1]$ , is a path in parameter space, and the infimum is taken over the Borel vector fields  $v_t = v(t, \cdot)$ , such that the *constrained continuity equation* holds:

$$\frac{\partial}{\partial t} \rho(\theta_t, x) + \nabla \cdot (\rho(\theta_t, x) v(t, x)) = 0, \quad \theta_0, \theta_1 \text{ are fixed.} \quad (3b)$$

We notice that the infimum of problem (3) is taken over density paths lying in the parameterized probability set, i.e.  $\rho(\theta_t, \cdot) \in \rho(\Theta)$ . Here the changing ratio of density is reduced into a finite dimensional direction, i.e.  $\frac{\partial}{\partial t} \rho(\theta_t, x) = (\nabla_{\theta_t} \rho(\theta_t, x), \frac{d\theta_t}{dt})$ , where  $(\cdot, \cdot)$  is an inner product in  $\mathbb{R}^d$ .

A natural question arises. Variational problem (2), together with its constrained problem (3), are written in Eulerian coordinates. They evolve unavoidably with the entire probability density functions. For practical reasons, can we find the Lagrangian coordinates for constrained problem (3)? In other words, what are analogs of (1) in  $\rho(\Theta)$ ? We next demonstrate an answer to this question. We show that there is an expression for the motion of particles, whose density path moves according to the constrained continuity equation (3b).

### 3. LAGRANGIAN FORMULATIONS

In this section, we show the main result of this note, that is the constrained dynamical OT (3) has a simple Lagrangian formulation in Proposition 1.

Consider a parameterized mapping or implicit generative model as follows. Given a input space  $Z \subset \mathbb{R}^{n_1}$ ,  $n_1 \leq n$ , let

$$g_{\theta}: Z \rightarrow \Omega, \quad x = g(\theta, z) \in \Omega, \quad \text{for } z \in Z.$$

Here  $g_{\theta}$  is a mapping function depending on parameters  $\theta \in \Theta$ . Given realizations  $\omega$  in event space, we assume that the random variable  $z(\omega)$  satisfies a density function  $\mu(z) \in \mathcal{P}_+(Z)$ , and denote  $x(\omega) = g(\theta, z(\omega))$  satisfying the density function  $\rho(\theta, x)$ . This means that the map  $g_{\theta}$  pushes forward  $\mu(z)$  to  $\rho(\theta, x)$ , denoted by  $\rho(\theta, x) = g_{\theta\#} \mu(z)$ :

$$\int_Z f(g(\theta, z)) \mu(z) dz = \int_{\Omega} f(x) \rho(\theta, x) dx, \quad \text{for any } f \in C_c^{\infty}(\Omega). \quad (4)$$

In this case, the parameterized probability set is given as follows:

$$\rho(\Theta) = \left\{ \rho(\theta, x) \in C^{\infty}(\Omega) : \theta \in \Theta, \rho(\theta, x) = g_{\theta\#} \mu(z) \right\}.$$

We next present the constrained dynamical OT in Lagrangian coordinates. We notice a fact that, the vector field in optimal density path of problem (2) or (3) satisfies

$$v(t, x) = D_p H(x, \nabla \Phi(t, x)),$$

---

<sup>2</sup>We abuse the notation of  $\rho$ . Notice that  $\rho(\theta, x)$  is a probability distribution parameterized by  $\theta \in \Theta$ , while  $\rho(x)$  is a probability distribution function in the full probability set.

where

$$H(x, p) = \sup_{v \in T_x \Omega} pv - L(x, v)$$

is the Hamiltonian function associated with  $L$ .

**Proposition 1** (Constrained dynamical OT in Lagrangian formulation). *The constrained dynamical OT has the following formulation:*

$$c(\theta_0, \theta_1) = \inf \left\{ \int_0^1 \mathbb{E}_{z \sim \mu} L(g(\theta_t, z), \frac{d}{dt}g(\theta_t, z)) dt : \right. \\ \left. \frac{d}{dt}g(\theta_t, z) = D_p H(g(\theta_t, z), \nabla_x \Phi(t, g(\theta_t, z))), \theta(0) = \theta_0, \theta(1) = \theta_1 \right\}, \quad (5)$$

where the infimum is taken over all feasible potential functions  $\Phi: [0, 1] \times \Omega \rightarrow \mathbb{R}$  and parameter paths  $\theta: [0, 1] \rightarrow \mathbb{R}^d$ .

*Proof.* Denote

$$\frac{d}{dt}g(\theta_t, z) = v(t, g(\theta_t, z)), \quad \text{with } v(t, g(\theta_t, z)) = D_p H(g(\theta_t, z), \nabla_x \Phi(t, g(\theta_t, z))).$$

We show that the probability density transition equation of  $g(\theta_t, z)$  satisfies the *constrained continuity equation*

$$\frac{\partial}{\partial t} \rho(\theta_t, x) + \nabla \cdot (\rho(\theta_t, x) v(t, x)) = 0, \quad (6)$$

and

$$\mathbb{E}_{z \sim \mu} L(g(\theta_t, z), \frac{d}{dt}g(\theta_t, z)) = \int_{\Omega} L(x, v(t, x)) \rho(\theta_t, x) dx. \quad (7)$$

On the one hand, consider  $f \in C_c^\infty(\Omega)$ , then

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_{z \sim \mu} f(g(\theta_t, z)) &= \frac{d}{dt} \int_{\mathcal{Z}} f(g(\theta_t, z)) \mu(z) dz \\ &= \frac{d}{dt} \int_{\Omega} f(x) \rho(\theta_t, x) dx \\ &= \int_{\Omega} f(x) \frac{\partial}{\partial t} \rho(\theta_t, x) dx, \end{aligned} \quad (8)$$

where the second equality holds from the push forward relation (4).

On the other hand, consider

$$\begin{aligned}
\frac{d}{dt} \mathbb{E}_{z \sim \mu} f(g(\theta_t, z)) &= \lim_{\Delta t \rightarrow 0} \mathbb{E}_{z \sim \mu} \frac{f(g(\theta_{t+\Delta t}, z)) - f(g(\theta_t, z))}{\Delta t} \\
&= \lim_{\Delta t \rightarrow 0} \int_Z \frac{f(g(\theta_{t+\Delta t}, z)) - f(g(\theta_t, z))}{\Delta t} \mu(z) dz \\
&= \int_Z \nabla f(g(\theta_t, z)) \frac{d}{dt} g(\theta_t, z) \mu(z) dz \\
&= \int_Z \nabla f(g(\theta_t, z)) v(t, g(\theta_t, z)) \mu(z) dz \\
&= \int_{\Omega} \nabla f(x) v(t, x) \rho(\theta_t, x) dx \\
&= - \int_{\Omega} f(x) \nabla \cdot (v(t, x) \rho(\theta_t, x)) dx,
\end{aligned} \tag{9}$$

where  $\nabla$ ,  $\nabla \cdot$  are gradient and divergence operators w.r.t.  $x \in \Omega$ . The second to last equality holds from the push forward relation (4), and the last equality holds using the integration by parts w.r.t.  $x$ . Since (8) = (9) for any  $f \in C_c^\infty(\Omega)$ , we have proven (6).

In addition, by the definition of the push forward operator (4), we have

$$\begin{aligned}
\mathbb{E}_{z \sim \mu} L(g(\theta_t, z), \frac{d}{dt} g(\theta_t, z)) &= \int_Z L(g(\theta_t, z), v(t, g(\theta_t, z))) \mu(z) dz \\
&= \int_{\Omega} L(x, v(t, x)) \rho(\theta_t, x) dx.
\end{aligned}$$

Thus we prove (7).  $\square$

It is interesting to compare variational problems (1) with (5). We can view  $g(\theta_t, z) \in \Omega$  as “parameterized” particles, whose density function is constrained in the parameterized probability set  $\rho(\Theta)$ . Their motions result at the evolution of probability transition densities in  $\rho(\Theta)$ , satisfying the constrained continuity equation (3b). For this reason, we call (5) the Lagrangian formalism of constrained dynamical OT.

It is also worth noting that each movement of  $g(\theta_t, z)$  results a motion in density path  $\rho(\theta_t, x)$ . The change of density path will identify a potential function  $\Phi(t, x)$  depending on  $\theta_t$ .

In addition, the cost functional in dynamical OT can involve general potential energies, such as linear potential energy:

$$\mathcal{V}(\rho) = \int_{\Omega} V(x) \rho(x) dx,$$

and interaction energy:

$$\mathcal{W}(\rho) = \int_{\Omega} \int_{\Omega} w(x, y) \rho(x) \rho(y) dx dy.$$

Here  $V(x)$  is a linear potential, and  $w(x, y) = w(y, x)$  is a symmetric interaction potential. If  $\rho(\theta, x) \in \rho(\Theta)$ , then

$$\begin{aligned}\mathcal{V}(\rho(\theta, \cdot)) &= \int_{\Omega} V(x)\rho(\theta, x)dx \\ &= \int_Z V(g(\theta, z))\mu(z)dz \\ &= \mathbb{E}_{z \sim \mu} V(g(\theta, z)),\end{aligned}$$

and

$$\begin{aligned}\mathcal{W}(\rho(\theta, \cdot)) &= \int_{\Omega} \int_{\Omega} w(x, y)\rho(\theta, x)\rho(\theta, y)dxdy \\ &= \int_Z \int_Z w(g(\theta, z_1), g(\theta, z_2))\mu(z_1)\mu(z_2)dz_1dz_2 \\ &= \mathbb{E}_{(z_1, z_2) \sim \mu \times \mu} w(g(\theta, z_1), g(\theta, z_2)),\end{aligned}$$

where each second equality in the above two formulas hold because of the constrained mapping relation (4) and  $\mu \circ \mu$  represents an independent joint density function supported on  $Z \times Z$  with marginals  $\mu(z_1)$ ,  $\mu(z_2)$ . Similarly in proposition 1, we have

$$\begin{aligned}c(\theta_0, \theta_1) &= \inf_{v_t, \theta(0)=\theta_0, \theta(1)=\theta_1} \left\{ \int_0^1 \left[ \int_{\Omega} L(x, v(t, x))\rho(\theta_t, x)dx - \mathcal{V}(\rho(\theta_t, \cdot)) - \mathcal{W}(\rho(\theta_t, \cdot)) \right] dt : \right. \\ &\quad \left. \frac{\partial}{\partial t} \rho(\theta_t, x) + \nabla \cdot (\rho(\theta_t, x)v(t, x)) = 0 \right\} \\ &= \inf_{\Phi_t, \theta_t, \theta(0)=\theta_0, \theta(1)=\theta_1} \left\{ \int_0^1 \left[ \int_{\Omega} L(x, D_p H(x, \nabla \Phi(t, x)))\rho(\theta_t, x)dx - \mathcal{V}(\rho(\theta_t, \cdot)) - \mathcal{W}(\rho(\theta_t, \cdot)) \right] dt : \right. \\ &\quad \left. \frac{\partial}{\partial t} \rho(\theta_t, x) + \nabla \cdot (\rho(\theta_t, x)v(t, x)) = 0 \right\} \\ &= \inf_{\Phi_t, \theta_t, \theta(0)=\theta_0, \theta(1)=\theta_1} \left\{ \int_0^1 \left[ \mathbb{E}_{z \sim \mu} L(g(\theta_t, z), \frac{d}{dt}g(\theta_t, z)) - \mathbb{E}_{z \sim \mu} V(g(\theta_t, z)) \right. \right. \\ &\quad \left. \left. - \mathbb{E}_{(z_1, z_2) \sim \mu \times \mu} w(g(\theta_t, z_1), g(\theta_t, z_2)) \right] dt : \right. \\ &\quad \left. \frac{d}{dt}g(\theta_t, z) = D_p H(g(\theta_t, z), \nabla_x \Phi(t, g(\theta_t, z))) \right\}.\end{aligned}$$

We next demonstrate an example of constrained dynamical OT problems.

**Example 1** (Constrained  $L^2$ -Wasserstein distance). *Let  $L(x, v) = \|v\|^2$  and denote  $d_{W_2}(\theta_0, \theta_1) = c(\theta_0, \theta_1)^{\frac{1}{2}}$ , then*

$$d_{W_2}(\theta_0, \theta_1)^2 = \inf \left\{ \int_0^1 \mathbb{E}_{z \sim \mu} \left\| \frac{d}{dt}g(\theta_t, z) \right\|^2 dt : \frac{d}{dt}g(\theta_t, z) = \nabla \Phi(t, g(\theta_t, z)), \theta(0) = \theta_0, \theta(1) = \theta_1 \right\}.$$

*Observe that (5) forms a geometric action energy function in parameter space  $\Theta$ , in which the metric tensor can be extracted explicitly. In other words, denote  $G(\theta) \in \mathbb{R}^{d \times d}$  by*

$$\dot{\theta}^\top G(\theta) \dot{\theta} = \dot{\theta}^\top \mathbb{E}_{\mu} (\nabla_{\theta} g(\theta, z) \nabla_{\theta} g(\theta, z)^\top) \dot{\theta},$$

*with the constraint*

$$(\dot{\theta}, \nabla_{\theta} g(\theta, z)) = \nabla_x \Phi(g(\theta, z)).$$

Here  $\nabla_{\theta}g(\theta, z) \in \mathbb{R}^{d \times n}$ ,  $\Phi$  is a potential function satisfying

$$-\nabla \cdot (\rho(\theta, x)\nabla\Phi(x)) = (\nabla_{\theta}\rho(\theta, x), \dot{\theta}),$$

and  $G(\theta) = \mathbb{E}_{z \sim \mu}(\nabla_{\theta}g(\theta, z)\nabla_{\theta}g(\theta, z)^{\top}) \in \mathbb{R}^{d \times d}$  is a semi-positive definite matrix.

## REFERENCES

- [1] S.-i. A. Natural Gradient Works Efficiently in Learning. *Neural Computation*, 10(2):251–276, 1998.
- [2] S. Amari. *Information Geometry and Its Applications*. Number volume 194 in Applied mathematical sciences. Springer, Japan, 2016.
- [3] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *arXiv:1701.07875 [cs, stat]*, 2017.
- [4] N. Ay, J. Jost, H. V. Lê, and L. J. Schwachhöfer. *Information Geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete A @series of modern surveys in mathematics\$13. Folge, volume 64. Springer, Cham, 2017.
- [5] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [6] Y. Chen and W. Li. Natural gradient in Wasserstein statistical manifold. *arXiv:1805.08380 [cs, math]*, 2018.
- [7] A. Halder and T. T. Georgiou. Gradient Flows in Uncertainty Propagation and Filtering of Linear Gaussian Systems. *arXiv:1704.00102 [cs, math]*, 2017.
- [8] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.
- [9] W. Li. Geometry of probability simplex via optimal transport. *arXiv:1803.06360 [math]*, 2018.
- [10] W. Li and G. Montufar. Natural gradient via optimal transport I. *arXiv:1803.07033 [cs, math]*, 2018.
- [11] K. Modin. Geometry of matrix decompositions seen through optimal transport and information geometry. *Journal of Geometric Mechanics*, 9(3):335–390, 2017.
- [12] E. Nelson. Derivation of the Schrödinger Equation from Newtonian Mechanics. *Physical Review*, 150(4):1079–1085, 1966.
- [13] G. Pistone. Lagrangian Function on the Finite State Space Statistical Bundle. *Entropy*, 20(2):139, 2018.
- [14] C. Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.

*E-mail address:* `wcli@math.ucla.edu`

*E-mail address:* `sjo@math.ucla.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES.