Online Learning to Transport via the Minimal Selection Principle

Wenxuan Guo WXGUO@CHICAGOBOOTH.EDU
University of Chicago

YoonHaeng Hur YOONHAENGHUR@UCHICAGO.EDU
University of Chicago

Tengyuan Liang TENGYUAN.LIANG@CHICAGOBOOTH.EDU
University of Chicago

Christopher Thomas Ryan CHRIS.RYAN@SAUDER.UBC.CA
University of British Columbia

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

Motivated by robust dynamic resource allocation in operations research, we study the *Online Learning to Transport* (OLT) problem where the decision variable is a probability measure, an infinite-dimensional object. We draw connections between online learning, optimal transport, and partial differential equations through an insight called the minimal selection principle, originally studied in the Wasserstein gradient flow setting by Ambrosio et al. (2005). This allows us to extend the standard online learning framework to the infinite-dimensional setting seamlessly. Based on our framework, we derive a novel method called the *minimal selection or exploration (MSoE) algorithm* to solve OLT problems using mean-field approximation and discretization techniques. In the displacement convex setting, the main theoretical message underpinning our approach is that minimizing transport cost over time (via the minimal selection principle) ensures optimal cumulative regret upper bounds. On the algorithmic side, our MSoE algorithm applies beyond the displacement convex setting, making the mathematical theory of optimal transport practically relevant to non-convex settings common in dynamic resource allocation.

Keywords: Infinite-dimensional optimization, online learning, optimal transport.

1. Introduction

Online learning and online convex optimization offer an elegant framework for regret minimization under worst-case sequences (Gordon, 1999; Zinkevich, 2003; Shalev-Shwartz, 2012; Orabona, 2021). Principles for these methods have been discovered independently in decision theory, game theory, learning theory, and convex optimization (Cesa-Bianchi and Lugosi, 2006). For problems with a finite-dimensional decision variable, extensive studies have been conducted to understand online algorithms that achieve small regret, either in Euclidean or non-Euclidean settings. Such algorithms usually rely on some notion of (sub-)gradients to iteratively improve the finite-dimensional decision variable. In the simplest Euclidean setting, let $\ell_t \colon \mathbb{R}^d \to \mathbb{R}$ be a smooth convex function and $x_t \in \mathbb{R}^d$ be a finite-dimensional decision variable, both indexed by time $t \in \mathbb{N}$. Online gradient descent with stepsize η satisfies, for any $z \in \mathbb{R}^d$,

$$2\eta\big(\overbrace{\ell_t(x_t) - \ell_t(z)}^{\text{regret term}}\big) - \big(\overbrace{\|x_t - z\|^2 - \|x_{t+1} - z\|^2}^{\text{telescoping term}}\big) \le \overbrace{\|x_t - x_{t+1}\|^2}^{\text{transport cost}} \le \overbrace{\|\eta \nabla \ell_t(x_t)\|^2}^{\text{gradient in Euclidean norm}} . \quad (1)$$

In a simple non-Euclidean setting, let $p_t \in \Delta_d$ be a probability distribution on a d-discrete decision and $\ell_t \in \mathbb{R}^d$ be a sequence of cost vectors where each element $\ell_t[k]$ denotes the cost associated with the k-th action and $\ell_t(p_t) := \sum_{k=1}^d \ell_t[k]p_t[k]$ thus $\nabla \ell_t(\cdot) \equiv \ell_t$. Online mirror descent (specifically, exponentiated gradient) satisfies for any $q \in \Delta_d$

$$\eta\big(\overbrace{\ell_t(p_t) - \ell_t(q)}^{\text{regret term}}\big) - \big(\overbrace{d_{\mathrm{KL}}(q||p_t) - d_{\mathrm{KL}}(q||p_{t+1})}^{\text{telescoping term}}\big) \leq \overbrace{d_{\mathrm{KL}}(p_t||p_{t+1})}^{\text{transport cost}} \leq \overbrace{\|\eta\nabla\ell_t\|_{p_t}^2}^{\text{gradient in local norm}}, \quad (2)$$

where the local norm is defined as $\|\ell_t\|_{p_t}^2 := \sum_{k=1}^d (\ell_t[k])^2 p_t[k]$. It is immediate to spot the resemblance between (1) and (2). The telescoping term marks the progress towards the competing decision (z or q), and the first right-hand side of the expression describes some notion of "transport cost" induced by different geometries.

It is natural to wonder how the theoretical and algorithmic principles behind these finite-dimensional cases extend to the infinite-dimensional case, specifically when the decision variable is a probability distribution over a generic metric space. Such a question is of practical importance. Several dynamic resource allocation problems in operations research involve these infinite-dimensional objects, e.g., routing a network of drones over airspace or allocating drivers across a city topology in ridesharing. Theoretically, the infinite-dimensional object can be viewed as a generalization of both (1) (x from finite to infinite-dimensional) and (2) (p from discrete to continuous measure). As hinted in the previous paragraph where small "transport cost" terms guarantee small regret, optimal transport theory (Villani, 2003; Ambrosio et al., 2005) plays a pivotal role in extending online learning to the infinite-dimensional case. This paper draws connections between online learning, optimal transport, and partial differential equations (PDEs) using an insight called the minimal selection principle. Eventually, we derive new algorithms based on this insight.

Optimal transport (OT) studies how to move in the space of probability measures to minimize a cost metric. This cost metric, credited to Monge in 1781, is associated with the optimal way to move mass between two probability measures. This paper investigates online learning using a toolbox set out by Breiner, who brought perspectives from PDEs, geometry, and functional analysis to the study of OT around 1987 (Brenier, 1987, 1989). Let us first introduce our infinite-dimensional online optimization problem in the language of OT and then lay out the connection between OT and online learning. Consider the following **Online Learning to Transport (OLT)** problem: let $\mathcal{P}(\mathbb{R}^d)$ be the space of probability measures on $\mathbb{R}^d$ where, in each round, $t = 1, \ldots, T$,

- an adversary chooses an energy functional $\mathcal{E}_t \colon \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ without revealing it;
- the player commits a probability measure $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ as the decision variable;
- the adversary reveals the energy functional, and the player suffers the loss $\mathcal{E}_t(\mu_t)$.

The player needs to decide the next μ_{t+1} based on μ_t and all historical information and aims to minimize cumulative regret from facing the adversary. The main conceptual message we discover in OLT is that optimally transporting the probability measures $\mu_1 \to \mu_2 \to \ldots \to \mu_T$ with respect to an appropriate cost enables small cumulative regret.

To provide a glimpse of the angle that inspired us to connect online learning and OT, we follow a viewpoint taken in Benamou and Brenier (1999) and Otto (2001). All concepts introduced in this paragraph have rigorous definitions in Section 2, so we proceed at a high level here. Mimicking the finite-dimensional case, OT theory provides a way to calculate a type of (sub-)gradient of $\mathcal{E}_t$ over $\mathcal{P}(\mathbb{R}^d)$ when equipped with a certain metric. This (sub-)gradient is used to update $\mu_t \to \mu_{t+1}$. To enrich this analogy, we consider a continuous/infinitesimal time analog. In the finite-dimensional cases (1) and (2), the first right-hand side in a continuous time analog would quantify

movement according to the ODE $\frac{\mathrm{d}x_t}{\mathrm{d}t} = -\nabla \ell_t(x_t)$ where $\left\|\frac{\mathrm{d}x_t}{\mathrm{d}t}\right\|^2 = \|\nabla \ell_t(x_t)\|^2$. In the infinite-dimensional case, the density ρ_t (of μ_t w.r.t. the Lebesgue measure) evolves according to the PDE $\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \boldsymbol{\xi}_t)$ with a vector field $\boldsymbol{\xi}_t \colon \mathbb{R}^d \to \mathbb{R}^d$ from the "Fréchet subdifferential" $\partial \mathcal{E}_t(\mu_t)$. The infinitesimal transport cost is[1]

$$\left\|\frac{\partial \rho_t}{\partial t}\right\|_{\rho_t}^2 = \inf_{\boldsymbol{\xi} \in L^2(\mu_t; \mathbb{R}^d)} \left\{ \int \|\boldsymbol{\xi}(x)\|^2 \, \mathrm{d}\mu_t(x) \;\; : \;\; \boldsymbol{\xi} \in \partial \mathcal{E}_t(\mu_t) \right\} . \tag{3}$$

This motivates the *minimal selection principle* of Ambrosio et al. (2005): among all the possible vector fields that belong to the Fréchet subdifferential, we aim to select one whose kinetic energy (captured by the integral $\int \|\boldsymbol{\xi}(x)\|^2 \, \mathrm{d}\mu_t(x)$) is lowest. Assuming a notion of convexity of $\mathcal{E}_t$ along Riemannian geodesics, we formally show that minimizing the transport cost over time using the minimal selection principle yields the smallest cumulative regret upper bounds in the OLT problem.

Equipped with the minimal selection principle, we propose a novel algorithm in Section 3 that we call *minimal selection or exploration (MSoE)* to solve the variational optimization problem in (3) and, in turn, solve OLT using only zeroth-order partial feedback similar to the bandit setting. To make this algorithm numerically tractable, we use a discrete analog of the minimal selection principle using mean-field approximations. It is noteworthy that our MSoE algorithm works beyond the convex setting, which shows how elegant theory from OT is practically relevant to certain non-convex settings common in robust dynamic resource allocation. Simple toy examples demonstrating the algorithm's empirical performance are discussed in Appendix C.

A key feature of this paper is the natural simplicity of its arguments and algorithmic principles that become apparent once the framework connecting online learning and OT is established. Several directions for extending our framework are discussed at the end of the paper.

Related work Our paper is at the intersection of two active research fields: online learning and optimal transport. For the former, due to space limits, we cannot do fair justice to credit all contributions properly; see Orabona (2021) for a comprehensive survey. Here, we give a selective overview. Several improvements of online gradient descent (1) using an adaptive stepsize have been proposed (Streeter and McMahan, 2010; McMahan and Streeter, 2010) with a further modification via scale-freeness (Orabona and Pál, 2015; Orabona and Pál, 2018). Expression (2) can be derived using different principles such as exponentiated gradient (Kivinen and Warmuth, 1997), online mirror descent (OMD), follow-the-regularized-leader (FTRL) (Shalev-Shwartz and Singer, 2007; Abernethy et al., 2008). Modifications of FTRL via adaptive regularization have been introduced (van Erven et al., 2011; De Rooij et al., 2014; Orabona and Pál, 2015). General first-order Riemannian optimization methods have been investigated in Zhang and Sra (2016). Using martingale tools, a theoretical framework for sequential prediction has been studied in Rakhlin and Sridharan (2014). The typical regret bound for online learning with K-discrete actions scales as $\sqrt{T \log K}$ in the full information setting, and as $\sqrt{TK \log K}$ in the bandit/partial feedback setting. Therefore, naive generalizations of these bounds to the infinite-dimensional case ($K \to \infty$) result in diverging bounds. It is thus unclear whether $\sqrt{T}$-regret is achievable in the infinite-dimensional case. We employ tools from optimal transport to resolve this issue and obtain optimal bounds.

Beyond well-established analytic results of OT, computational (Cuturi, 2013; Genevay et al., 2016; Altschuler et al., 2017) and statistical (Weed and Bach, 2019; Liang, 2021; Weed and Berthet,

1. The curious reader may identify the above resemblance to the local norm in (2): here the local Riemannian geometry quantifies the transport cost.

2019; Liang, 2019; Hütter and Rigollet, 2021) aspects have been emerging research areas at the intersection of OT, probability distribution estimation, and numerical sampling. At the same time, OT has been a versatile tool for various applied tasks such as image retrieval (Rubner et al., 2000), computational linguistics (Kusner et al., 2015), and domain adaptation (Courty et al., 2017). One of the successful applications of OT is generative sampling, such as GANs (Goodfellow et al., 2014). OT has improved generative sampling in several ways: by modifying GANs using OT-based probability metrics (Arjovsky et al., 2017; Genevay et al., 2018) or by utilizing dual formulations of OT (Seguy et al., 2018; Makkuva et al., 2020). More recently, further improvements (Bunne et al., 2019; Hur et al., 2021) have been made by considering the Gromov-Wasserstein (Mémoli, 2011), a generalization of OT ideas for identifying isomorphism in metric measure spaces.

Notation Let $\mathcal{P}(\mathbb{R}^d)$ denote the set of all Borel probability measures on $\mathbb{R}^d$. Let $\mathcal{P}^r(\mathbb{R}^d)$ denote the subset of $\mathcal{P}(\mathbb{R}^d)$ of Borel probability measures that are absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^d$ on $\mathbb{R}^d$. For $\mu \in \mathcal{P}^r(\mathbb{R}^d)$, we write $\mu = \rho \cdot \mathcal{L}^d$ to signify that ρ is a density function of μ with respect to $\mathcal{L}^d$. Let $\Pi(\mu, \nu)$ denote the collection of all couplings of $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, that is, $\gamma \in \Pi(\mu, \nu)$ is a Borel probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\gamma(A \times \mathbb{R}^d) = \mu(A)$ and $\gamma(\mathbb{R}^d \times B) = \nu(B)$ for all Borel subsets $A, B \subset \mathbb{R}^d$. For a measurable map $\mathbf{t}\colon \mathbb{R}^d \to \mathbb{R}^p$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$, we define the pushforward measure $(\mathbf{t}_{\#}\mu)(B) = \mu\{x \in \mathbb{R}^d : \mathbf{t}(x) \in B\}$ for any Borel subset $B \subset \mathbb{R}^p$; hence $\mathbf{t}_{\#}\mu \in \mathcal{P}(\mathbb{R}^p)$. We let δ_x denote the Dirac measure for $x \in \mathbb{R}^d$ and $\mathbf{i}\colon \mathbb{R}^d \to \mathbb{R}^d$ the identity map $\mathbf{i}(x) = x$ for all $x \in \mathbb{R}^d$. Lastly, $\|\cdot\|$ denotes the standard Euclidean norm on $\mathbb{R}^d$, $[x]_+ = \max(x, 0)$ for $x \in \mathbb{R}$, and let $[n]$ denote the unordered set $\{1, \dots, n\}$ for $n \in \mathbb{N}$.

2. Optimal Transport, Minimal Selection Principle, and Regret Bound

In this section, we present our main theoretical results for OLT. As noted earlier, the key to our analysis is a connection with optimal transport theory. We start by introducing the notion of Wasserstein space to structure the decision space of OLT. Recall from (1) and (2) that a notion of difference (transport cost) between two consecutive decisions plays a role in deriving a regret bound. To apply this idea to OLT, we utilize the Wasserstein distance between probability measures (decisions in OLT). Next, we discuss differential calculus over Wasserstein space, which serves as a key building block for deriving a regret bound for OLT. In Wasserstein space, we first define a subdifferential and then select an element associated with the smallest size, measured by the local Riemannian geometry of the Wasserstein space. Such an element, which we call a minimal selection, is defined by a variational problem and serves as a functional gradient. Based on this, we make transparent a strategy to obtain $\sqrt{T}$-regret that generalizes seamlessly to the infinite-dimensional setting. Lastly, we mention a connection between OLT and Wasserstein gradient flow.

2.1. Wasserstein Space

One of the most important consequences of optimal transport theory is that we can define a distance between $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ by finding a coupling that gives the smallest transport cost. Concretely, we define the Wasserstein distance W_2 on $\mathcal{P}(\mathbb{R}^d)$ as $W_2^2(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, \mathrm{d}\gamma(x, y)$. The Wasserstein distance W_2 is indeed a distance over $\mathcal{P}_2(\mathbb{R}^d)$, where $\mathcal{P}_2(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|x\|^2 \, \mathrm{d}\mu(x) < \infty\}$. From this, we obtain a metric space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ of probability measures called the Wasserstein space.

One important property of W_2 is that we can find a unique optimal coupling under certain regularity conditions. Here, an optimal coupling is any coupling $\gamma \in \Pi(\mu, \nu)$ associated with the smallest transport cost, that is, $W_2^2(\mu, \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, \mathrm{d}\gamma(x, y)$. Proposition 2.1 of Villani (2003) tells that $\Pi_o(\mu, \nu)$, the set of all optimal couplings, is always nonempty. The following result states that $\Pi_o(\mu, \nu)$ is a singleton if μ is absolutely continuous with respect to the Lebesgue measure; moreover, such an optimal coupling is concentrated on the graph of some unique map.

Lemma 1 *Let $\mathcal{P}_2^r(\mathbb{R}^d) = \mathcal{P}_2(\mathbb{R}^d) \cap \mathcal{P}^r(\mathbb{R}^d)$. Given $\mu \in \mathcal{P}_2^r(\mathbb{R}^d)$ and $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, there exists a unique optimal coupling $\gamma \in \Pi(\mu, \nu)$, namely $\Pi_o(\mu, \nu) = \{\gamma\}$. Also, there exists a unique map $\mathbf{t}_\mu^\nu \colon \mathbb{R}^d \to \mathbb{R}^d$ such that γ is concentrated on the graph of $\mathbf{t}_\mu^\nu$, that is,*

$$\gamma(\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : y = \mathbf{t}_\mu^\nu(x)\}) = 1 \, .$$

This implies $(\mathbf{t}_\mu^\nu)_{\#}\mu = \nu$ and $W_2^2(\mu, \nu) = \int_{\mathbb{R}^d} \|x - \mathbf{t}_\mu^\nu(x)\|^2 \, \mathrm{d}\mu(x)$.

Remark 2 *This is a part of Brenier's theorem (Brenier, 1991), which contains further properties of $\mathbf{t}_\mu^\nu$ such as monotonicity. See Theorem 2.12 of Villani (2003) for details.*

2.2. Displacement Convexity and Subdifferential Calculus

Having defined the decision space for OLT, we explore how to minimize a loss function over this space. As noted earlier, the key is finding an object analogous to the gradient in the finite-dimensional setting. Ambrosio et al. (2005) achieves this by rigorously defining notions of convexity and differential calculus on Wasserstein space. We reproduce a few relevant results from Ambrosio et al. (2005). Throughout, we consider a functional $\mathcal{E} \colon \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ with a nonempty domain $D(\mathcal{E}) := \{\mu \in \mathcal{P}_2^r(\mathbb{R}^d) : \mathcal{E}(\mu) < \infty\} \neq \emptyset$; for a concise summary, we restrict the domain to $\mathcal{P}_2^r(\mathbb{R}^d)$, see Section 10.1 of Ambrosio et al. (2005) for more details. First, we discuss convexity of $\mathcal{E}$. To import convexity into $\mathcal{P}_2^r(\mathbb{R}^d)$, we need a concept that replaces the concept of 'line segment' in a vector space. McCann (1997) introduces the displacement interpolation for connecting two elements of $\mathcal{P}_2(\mathbb{R}^d)$ and defines the convexity based on it.

Definition 3 (Displacement interpolation and Convexity) *We define the displacement interpolation between $\mu, \nu \in \mathcal{P}_2^r(\mathbb{R}^d)$ as a curve $(\pi_\eta^{\mu \to \nu})_{\eta \in [0,1]}$ in $\mathcal{P}_2^r(\mathbb{R}^d)$ such that*

$$\pi_\eta^{\mu \to \nu} := ((1 - \eta)\mathbf{i} + \eta \mathbf{t}_\mu^\nu)_{\#}\mu \, .$$

We say $\mathcal{E}$ is displacement convex if $\mathcal{E}(\pi_\eta^{\mu \to \nu}) \leq (1 - \eta)\mathcal{E}(\mu) + \eta \mathcal{E}(\nu)$ for all $\mu, \nu \in D(\mathcal{E})$ and $\eta \in [0, 1]$.

Remark 4 *To see that the displacement interpolation $(\pi_\eta^{\mu \to \nu})_{\eta \in [0,1]}$ serves as a segment connecting μ and ν one can verify that $\pi_0^{\mu \to \nu} = \mu$, $\pi_1^{\mu \to \nu} = \nu$, and $\pi_\eta^{\mu \to \nu} \in \mathcal{P}_2^r(\mathbb{R}^d)$ for all $\eta \in (0, 1)$. See Chapter 7 of Ambrosio et al. (2005) for details.*

Next, we introduce the Fréchet subdifferential, which delineates the differentiable structure on the Wasserstein space. Recall that derivatives or subgradients of functionals defined on a Hilbert space are linear functionals over that space. To mimic their features, we define the subdifferential at $\mu \in \mathcal{P}_2^r(\mathbb{R}^d)$ by means of a Hilbert space $L^2(\mu; \mathbb{R}^d)$ as follows.

Definition 5 (Fréchet Subdifferential) *For $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, let $L^2(\mu;\mathbb{R}^d)$ be the collection of vector fields $\boldsymbol{\xi}\colon \mathbb{R}^d \to \mathbb{R}^d$ satisfying $\|\boldsymbol{\xi}\|^2_{L^2(\mu;\mathbb{R}^d)} := \int_{\mathbb{R}^d} \|\boldsymbol{\xi}(x)\|^2 \,\mathrm{d}\mu(x) < \infty$. For a lower semi-continuous functional $\mathcal{E}$ and $\mu \in D(\mathcal{E})$, we say that $\boldsymbol{\xi} \in L^2(\mu;\mathbb{R}^d)$ belongs to the Fréchet subdifferential $\partial\mathcal{E}(\mu)$ if*

$$\liminf_{\nu\to\mu} \frac{\mathcal{E}(\nu) - \mathcal{E}(\mu) - \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}(x), \mathbf{t}^\nu_\mu(x) - x\rangle \,\mathrm{d}\mu(x)}{W_2(\nu,\mu)} \geq 0 \,, \tag{4}$$

where $\liminf_{\nu\to\mu}$ is based on the convergence under W_2.

Definition 5 is a local definition as one can see from (4). For a displacement convex functional, however, Fréchet subdifferentials admit a global characterization. Moreover, as the next result shows, we can find a unique member of the Fréchet subdifferential with the smallest norm; see Section 10.1 of Ambrosio et al. (2005).

Lemma 6 (Minimal Selection Principle) *Let $\mathcal{E}$ be a lower semi-continuous and displacement convex functional, then a vector field $\boldsymbol{\xi}$ belongs to the Fréchet subdifferential $\partial\mathcal{E}(\mu)$ if and only if*

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \geq \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}(x), \mathbf{t}^\nu_\mu(x) - x\rangle \,\mathrm{d}\mu(x) \quad \forall \nu \in \mathcal{P}^r_2(\mathbb{R}^d) \,.$$

In this case, $\partial\mathcal{E}(\mu)$ has a unique element $\partial^o\mathcal{E}(\mu)$ called the minimal selection *with the smallest norm in the following sense:*

$$\partial^o\mathcal{E}(\mu) = \arg\min\{\|\boldsymbol{\xi}\|^2_{L^2(\mu;\mathbb{R}^d)} : \boldsymbol{\xi} \in \partial\mathcal{E}(\mu)\} \,.$$

As we shall see in Theorem 8, the minimal selection $\partial^o\mathcal{E}(\mu)$ plays a role analogous to a gradient in providing a regret bound for OLT comparable to (1) and (2). We conclude this subsection with two canonical examples of the minimal selection principle.

Example 1 (Potential Functional) *Given a lower semi-continuous function $V\colon \mathbb{R}^d \to (-\infty,\infty]$, we call $\mathcal{V}\colon \mathcal{P}_2(\mathbb{R}^d) \to (-\infty,\infty]$ a potential functional associated with V if*

$$\mathcal{V}(\mu) := \int_{\mathbb{R}^d} V(x) \,\mathrm{d}\mu(x)$$

and $\partial^o\mathcal{V}(\mu) = \nabla V$ if V is convex and satisfies some regularity conditions; see Section 10.4 of Ambrosio et al. (2005).

Example 2 (Interaction Functional) *Given a lower semi-continuous function $W\colon \mathbb{R}^d \to [0,\infty)$, we call $\mathcal{W}\colon \mathcal{P}_2(\mathbb{R}^d) \to [0,\infty]$ an interaction functional associated with W if*

$$\mathcal{W}(\mu) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} W(x-y) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y)$$

and $\partial^o\mathcal{W}(\mu) = \nabla W * \rho$, where $\mu = \rho\cdot\mathcal{L}^d$, if W is convex and satisfies some regularity conditions; see Section 10.4 of Ambrosio et al. (2005). Here, $*$ denotes a convolution of vector field ∇W and density ρ.

2.3. Regret Bound for OLT

We are ready to derive a regret bound for the OLT problem. The following result, often referred to as Evolution Variational Inequality (EVI) in the PDE literature, clarifies why the minimal selection principle is key to obtaining a regret bound.

Lemma 7 (EVI) *Let $\mathcal{E}$ be a lower semi-continuous and displacement convex functional. Let $\mu \in D(\mathcal{E})$ and $\boldsymbol{\xi} \in \partial\mathcal{E}(\mu)$. Fix $\eta > 0$ and define $\mu_\eta := (\mathbf{i} - \eta\boldsymbol{\xi})_{\#}\mu$. Then, for any $\nu \in \mathcal{P}_2^r(\mathbb{R}^d)$,*

$$\mathcal{E}(\mu) - \mathcal{E}(\nu) \leq \frac{W_2^2(\mu,\nu) - W_2^2(\mu_\eta,\nu)}{2\eta} + \frac{\eta}{2}\int_{\mathbb{R}^d} \|\boldsymbol{\xi}(x)\|^2 \,\mathrm{d}\mu(x) \ . \tag{5}$$

Proof Recall from Lemma 1 that $W_2^2(\mu,\nu) = \int_{\mathbb{R}^d} \|x - \mathbf{t}_\mu^\nu(x)\|^2 \,\mathrm{d}\mu(x)$. Let $(\mathbf{i} - \eta\boldsymbol{\xi}, \mathbf{t}_\mu^\nu)$ be the map from $\mathbb{R}^d$ to $\mathbb{R}^{2d}$ that maps $x \mapsto (x - \eta\boldsymbol{\xi}(x), \mathbf{t}_\mu^\nu(x))$, then $(\mathbf{i} - \eta\boldsymbol{\xi}, \mathbf{t}_\mu^\nu)_{\#}\mu \in \Pi(\mu_\eta,\nu)$ and thus

$$W_2^2(\mu_\eta,\nu) = \inf_{\gamma\in\Pi(\mu_\eta,\nu)} \int_{\mathbb{R}^d} \|x-y\|^2 \,\mathrm{d}\gamma(x,y) \leq \int_{\mathbb{R}^d} \|(x - \eta\boldsymbol{\xi}(x)) - \mathbf{t}_\mu^\nu(x)\|^2 \,\mathrm{d}\mu(x) \ .$$

Hence,

$$\begin{aligned}
\frac{W_2^2(\mu,\nu) - W_2^2(\mu_\eta,\nu)}{2\eta} &\geq \frac{1}{2\eta}\int_{\mathbb{R}^d} \left(\|x - \mathbf{t}_\mu^\nu(x)\|^2 - \|x - \eta\boldsymbol{\xi}(x) - \mathbf{t}_\mu^\nu(x)\|^2\right) \mathrm{d}\mu(x) \\
&= \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}(x), x - \mathbf{t}_\mu^\nu(x)\rangle \,\mathrm{d}\mu(x) - \frac{\eta}{2}\int_{\mathbb{R}^d} \|\boldsymbol{\xi}(x)\|^2 \,\mathrm{d}\mu(x) \\
&\geq \mathcal{E}(\mu) - \mathcal{E}(\nu) - \frac{\eta}{2}\int_{\mathbb{R}^d} \|\boldsymbol{\xi}(x)\|^2 \,\mathrm{d}\mu(x) \quad (\because \text{Lemma 6}) \ .
\end{aligned}$$

■

From (5), it is now obvious that the minimal selection $\boldsymbol{\xi} = \partial^o\mathcal{E}(\mu)$ gives the smallest upper bound. Also, notice the similarity of (1), (2), and (5); the last term on the right-hand side of (5) corresponds to the norm of a gradient in (1) and (2). The minimal selection enables us to import bounding techniques in (1) and (2) into the OLT problem. Based on this connection, we propose a strategy to tackle the OLT: for each time t, the player finds the minimal selection $\boldsymbol{\xi}_t = \partial^o\mathcal{E}_t(\mu_t)$ and updates $\mu_{t+1} = (\mathbf{i} - \eta\boldsymbol{\xi}_t)_{\#}\mu_t$. Using (5), we obtain the following regret bound.

Theorem 8 (Regret Bound: Discrete-Time) *Assume $\mathcal{E}_t$ of the OLT is displacement convex for all $t \in [T]$ and $T \in \mathbb{N}$. The player selects $\boldsymbol{\xi}_t = \partial^o\mathcal{E}_t(\mu_t)$ and updates $\mu_{t+1} = (\mathbf{i} - \eta\boldsymbol{\xi}_t)_{\#}\mu_t$ for each round t, where $\eta > 0$ is a fixed stepsize. Then, for any $\nu \in \mathcal{P}_2^r(\mathbb{R}^d)$,*

$$\sum_{t=1}^T \mathcal{E}_t(\mu_t) - \sum_{t=1}^T \mathcal{E}_t(\nu) \leq \frac{W_2^2(\mu_1,\nu) - W_2^2(\mu_{T+1},\nu)}{2\eta} + \frac{\eta}{2}\sum_{t=1}^T \int_{\mathbb{R}^d} \|\boldsymbol{\xi}_t(x)\|^2 \,\mathrm{d}\mu_t(x) \ . \tag{6}$$

Remark 9 *Under mild conditions such as μ_t and ν are supported on a bounded domain and $\boldsymbol{\xi}_t$ are uniformly bounded, we have $W_2(\mu_t,\nu) \leq D$ and $\|\boldsymbol{\xi}_t\|_{L^2(\mu_t;\mathbb{R}^d)} \leq L$ for all t for some constants $D, L > 0$. Then, the right-hand side of (6) is upper bounded by $\frac{D^2}{2\eta} + \frac{\eta L^2 T}{2}$, which becomes $DL\sqrt{T}$ for $\eta = \frac{D}{L\sqrt{T}}$ and yields $\sqrt{T}$-regret.*

Remark 10 *We clarify that the EVI is an existing technique in the PDE literature and has already been used for different purposes, for instance, see Dieuleveut et al. (2017) and Salim et al. (2020). We emphasize, however, that Theorem 8 utilizes such a technique for the first time in the context of the online learning given a sequence of adversarial functionals $(\mathcal{E}_t)_{t=1}^T$, and an arbitrary reference measure ν.*

2.4. Connections to Wasserstein Gradient Flow

We conclude this section with a connection between online learning and Wasserstein gradient flow through a lens of continuous-time OLT, further highlighting why it is natural to employ insights from PDEs. In the infinitesimal limit as $\eta \to 0$, the algorithm solving OLT amounts to a Wasserstein gradient flow. To see this, consider the steepest descent on Wasserstein space according to the energy functional $\mathcal{E}_t$ at time t, $\mu_{t+\eta} := \arg\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{E}_t(\mu) + \frac{1}{2\eta} W_2^2(\mu, \mu_t)$. In the infinitesimal limit, the steepest descent defines continuous-time evolution of probability measures $(\mu_t)_{t\in[0,T]}$, naturally described by PDEs. The following continuity equation is the natural counterpart of discrete update of the steepest descent: letting $\mu_t = \rho_t \cdot \mathcal{L}^d$, solve

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \boldsymbol{\xi}_t) \quad \text{where } \boldsymbol{\xi}_t \in \partial \mathcal{E}_t(\mu_t) \,. \tag{7}$$

Therefore, it is clear that our online OLT is a discrete analogue of the online Wasserstein gradient flow, with a time-varying energy functional $\mathcal{E}_t$. A solution $(\mu_t)_{t\in[0,T]}$ enjoys the following regret bound; see Appendix B.1 for the proof.

Theorem 11 (Regret Bound: Continuous-Time) *Assume $\mathcal{E}_t$ is displacement convex for all $t \in [0,T]$. Under mild regularity assumptions, a solution $(\mu_t)_{t\in[0,T]}$ to (7) satisfies*

$$\int_0^T \mathcal{E}_t(\mu_t)\,\mathrm{d}t - \int_0^T \mathcal{E}_t(\nu)\,\mathrm{d}t \le \frac{W_2^2(\mu_0,\nu) - W_2^2(\mu_T,\nu)}{2} \quad \forall \nu \in \mathcal{P}_2^r(\mathbb{R}^d) \,. \tag{8}$$

Remark 12 *If $\mathcal{E}_t$ is time-invariant, say $\mathcal{E}_t = \mathcal{E}$ for all $t \in [0,T]$, then (7) amounts to finding a solution to the standard Wasserstein gradient flow given $\mathcal{E}$ (Chapter 11 of Ambrosio et al. (2005)); minimizing the regret is exactly minimizing a functional $\mathcal{E}$ by solving a Wasserstein gradient flow. Note that the RHS is time-independent, namely, we have bounded total regret and fast rate as $1/T$.*

3. Minimal Selection Algorithm with Zeroth-Order Information

In this section, we propose a more practical framework for OLT and methods to solve it based on only zeroth-order, partial feedback. First, let us briefly explain our motivation for studying this practical framework. Recall that the strategy we studied in the previous section to solve OLT was to find the minimal selection $\boldsymbol{\xi}_t = \partial^o \mathcal{E}_t(\mu_t)$ and update μ_t to $(\mathbf{i} - \eta \boldsymbol{\xi}_t)_{\#}\mu_t$ for all t. The EVI gave us a regret bound as in Theorem 8. However, such a strategy is difficult to implement in practice. Finding $\partial^o \mathcal{E}_t(\mu_t)$ requires the player to access the Fréchet subdifferential $\partial \mathcal{E}_t(\mu_t)$ for each t. Such information is not directly available as nature typically only reveals the loss $\mathcal{E}_t(\mu_t)$ and not the entire collection of vector fields in $\partial^o \mathcal{E}_t(\mu_t)$. For instance, suppose $\mathcal{E}_t$ is a potential functional associated with $V_t \colon \mathbb{R}^d \to \mathbb{R}$, then $\nabla V_t = \partial^o \mathcal{E}_t(\mu_t)$ due to Example 1. It is not obvious how the player can determine the vector field ∇V_t based only on the zeroth-order information V_t.

As a result, we need a more reasonable setting to implement a strategy based on the minimal selection principle. We achieve this goal by querying only through zeroth-order information, drawing a parallel to the bandit/partial feedback setting with finite arms. To make our discussion concrete, we focus on the case where $\mathcal{E}_t = \mathcal{V}_t$ (potential functional) that occurs frequently in applications; the loss functional is a potential functional associated with some $V_t \colon \mathbb{R}^d \to \mathbb{R}$ for all $t \in \mathbb{N}$. An extension to the interaction functional, another common setting in applications, will be discussed later in this section. Under this framework, nature reveals $\mathcal{V}_t(\mu_t)$ by telling the values of V_t **only** on the support of μ_t and some fixed set of grid points. As we will see shortly, such a setting enables the player to execute a strategy based on the minimal selection principle, thereby obtaining a regret bound based on the EVI. In other words, instead of observing the gradient ∇V_t (minimal selection), the player approximates it using only zeroth-order information. As such, our formulation bridges the gap between the sophisticated theory in the previous section and its practical realization.

3.1. OLT with Zeroth-Order Information and Minimal Selection Algorithm

We formally define our online learning problem. Fix a domain $\Omega \subset \mathbb{R}^d$ and a set $Z \subset \Omega$ of grid points or hubs, say $Z = \{z_1, \dots, z_n\}$, known to the player.[2] At round $t \in \mathbb{N}$,

- the player chooses a discrete measure $\mu_t := \frac{1}{m}\sum_{j=1}^m \delta_{x_j^t}$, where we call $x_1^t, \dots, x_m^t \in \Omega$ decision points,
- nature reveals a lower semi-continuous function $V_t \colon \mathbb{R}^d \to \mathbb{R}$ only on $\{x_j^t\}_{j\in[m]} \cup Z$, namely the zeroth-order information on the player's decision points and the fixed grid points, and the player suffers a loss given as the potential functional associated with V_t, that is, $\mathcal{V}_t(\mu_t) = \frac{1}{m}\sum_{j=1}^m V_t(x_j^t)$ (Example 1).

Here, we view the grid points as the canonical locations of the domain Ω. Meanwhile, decision points x_j^t are any elements of Ω, not necessarily the grid points. By using the zeroth-information $V_t(z_i)$, the player competes against the grid points, meaning that she aims to choose her decision points that would incur smaller losses than the grid points would do. Accordingly, we consider a regret $\sum_{t=1}^T \mathcal{V}_t(\mu_t) - \sum_{t=1}^T \mathcal{V}_t(\nu)$ for any $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ supported on Z.

Inspired by the minimal selection principle (Lemma 6), we propose an algorithm for this learning problem.

Definition 13 (Minimal Selection Algorithm) *After round $t \in \mathbb{N}$, the player solves*

$$\min_{\xi_1,\dots,\xi_m \in \mathbb{R}^d} \quad \frac{1}{m}\sum_{j=1}^m \|\xi_j\|^2 \tag{9}$$

$$\textit{subject to} \quad V_t(x_j^t) - V_t(z_i) \le \langle \xi_j, x_j^t - z_i \rangle \ \ \forall (i,j) \in [n] \times [m] \ . \tag{10}$$

After obtaining a minimizer $(\xi_1^t, \dots, \xi_m^t)$, the player updates $x_j^{t+1} = x_j^t - \eta \xi_j^t$, where $\eta > 0$ is a suitable stepsize.

Remark 14 *Note that this is a convex program. Also, there exists a unique minimizer $(\xi_1^t, \dots, \xi_m^t)$ provided the constraint (10) is feasible.*

2. For instance, z_i's could be hub points (in ridesharing applications), be stochastically sampled based on some reference measure, or be grid points for discretization.

Not surprisingly, this algorithm amounts to the minimal section principle in a discrete setting. Recall that $\|\boldsymbol{\xi}\|^2_{L^2(\mu_t;\mathbb{R}^d)} = \frac{1}{m}\sum_{j=1}^m \|\boldsymbol{\xi}(x_j^t)\|^2$ for any $\boldsymbol{\xi} \in L^2(\mu_t;\mathbb{R}^d)$, hence the objective function (9) corresponds to the norm of an element of the Fréchet subdifferential $\partial\mathcal{V}_t(\mu_t)$. Meanwhile, constraint (10) amounts to finding a certified subgradient of V_t at x_j^t. Recall from Example 1 that $\partial^o\mathcal{V}_t(\mu_t) = \nabla V_t$. Essentially, this algorithm aims to find a minimizer $\xi_j^t = \nabla V_t(x_j^t)$. Also, the update rule $x_j^{t+1} = x_j^t - \eta\xi_j^t$ corresponds to $\mu_{t+1} = (\mathbf{i} - \eta\boldsymbol{\xi}_t)_\#\mu_t$ in Theorem 8, where $\boldsymbol{\xi}_t(x_j^t) = \xi_j^t$.

Now, we can utilize the EVI to obtain a regret bound. By adapting the proof of Lemma 7, we obtain the following result similar to Theorem 8; see Appendix A.1 for the proof.

Theorem 15 *Assume $\Omega = \mathbb{R}^d$. Let μ_t and ξ_j^t be the output of the minimal selection algorithm. If V_t is convex for all $t \in \mathbb{N}$, then for any measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ supported on Z and any $\eta > 0$,*

$$\sum_{t=1}^T \mathcal{V}_t(\mu_t) - \sum_{t=1}^T \mathcal{V}_t(\nu) \le \frac{W_2^2(\mu_1,\nu) - W_2^2(\mu_{T+1},\nu)}{2\eta} + \frac{\eta}{2}\sum_{t=1}^T \left(\frac{1}{m}\sum_{j=1}^m \|\xi_j^t\|^2\right) . \tag{11}$$

Remark 16 *Remark that the RHS of the above equation is precisely the minimal objective value in (9). Additionally, if $\Omega \subsetneq \mathbb{R}^d$, the update rule $x_j^{t+1} = x_j^t - \eta\xi_j^t$ may cause $x_j^{t+1} \notin \Omega$. We can prevent this via a projection step. See Subsection 3.3 for details.*

Remark 17 *We emphasize that the key of the aforementioned zeroth-order framework is that the decision points can take any values in Ω, allowing the support of μ_t to change continuously on Ω. On the contrary, if we had restricted the support of μ_t to be contained in Z, the decision points would have moved discontinuously over Z. In applications such as real-time dynamic resource allocation, such a discontinuous movement is impractical as one needs to allocate the decision points (resources such as drones or drivers) within a short period of time. In conclusion, our framework is more suitable in practice due to the continuously varying support of μ_t. At the same time, our framework preserves the infinite-dimensional nature.*

3.2. Beyond Convex Case: Exploration and Regret Bound

If V_t is non-convex, constraint set (10) might be infeasible. We propose a simple modification of the algorithm using exploration and provide a regret bound that extends beyond the convex case.

Note that we may consider the previous learning problem at each decision point separately. The minimal selection algorithm amounts to solving, for each $j \in [m]$,

$$\min_{\xi_j \in \mathbb{R}^d} \quad \|\xi_j\|^2 \quad \text{subject to } V_t(x_j^t) - V_t(z_i) \le \langle \xi_j, x_j^t - z_i\rangle \ \ \forall i \in [n] . \tag{12}$$

If the above problem is not feasible, we sample a stochastic, isotropic Gaussian vector with a suitable scaling and update x_j^t along this exploration direction.

Definition 18 (Minimal Selection or Exploration (MSoE) Algorithm) *After round $t \in \mathbb{N}$, let S^t be a subset of $[m]$ such that $j \in S^t$ if there is $\xi_j \in \mathbb{R}^d$ satisfying $V_t(x_j) - V_t(z_i) \le \langle \xi_j, x_j^t - z_i\rangle$ for all $i \in [n]$. For $j \in S^t$, the player solves (12) and updates $x_j^{t+1} = x_j^t - \eta\xi_j^t$ using the unique solution ξ_j^t to (12). For $j \notin S^t$, the player samples a Gaussian vector $g_j^t \sim N(0, I_d)$ and updates*

$$x_j^{t+1} = x_j^t - \sqrt{\eta \frac{\max_{i\in[n]}[V_t(x_j^t) - V_t(z_i)]_+}{d}}\, g_j^t .$$

We assume all the Gaussian vectors are independent of each other.

In short, we modify the minimal selection algorithm to let the feasible decision points move along minimal selection directions as before, while infeasible decision points fully explore in a random direction. The fact that the discretization stepsize scales with $\sqrt{\eta}$ is motivated by Euler discretization of Langevin dynamics. The adaptive stepsize factor $\max_{i\in[n]}[V_t(x_j^t) - V_t(z_i)]_+$ follows the relative potential difference between infeasible decision points and grid points. For the MSoE algorithm, an expected regret bound can be derived that depends on the fraction of infeasible points; see Appendix A.2 for the proof.

Theorem 19 *Assume $\Omega = \mathbb{R}^d$. Let μ_t, ξ_j^t, and g_j^t be the output of the MSoE algorithm in Definition 18. For any V_t and any measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ supported on Z,*

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathcal{V}_t(\mu_t) - \sum_{t=1}^{T}\mathcal{V}_t(\nu)\right] \le \frac{W_2^2(\mu_1,\nu)}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{m}\sum_{j\in S^t}\|\xi_j^t\|^2\right] + \frac{3}{2}\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{m}\sum_{j\notin S^t}\max_{i\in[n]}[V_t(x_j^t) - V_t(z_i)]_+\right] . \tag{13}$$

From (13), we can expect a moderate regret bound provided that the proportion $\frac{|[m]\setminus S^t|}{m}$ of infeasible decision points is small. We prove that this is indeed the case under some assumptions on the sequence of functions V_t. To see this, we first define a feasible region.

Definition 20 *For $t \in \mathbb{N}$, let*

$$R^t := \{x \in \Omega : \exists \xi \text{ such that } V_t(x) - V_t(z_i) \le \langle \xi, x - z_i\rangle \ \ \forall i \in [n]\}$$

and for each $x \in R^t$ define

$$\xi_x = \operatorname*{arg\,min}_{\xi\in\mathbb{R}^d} \|\xi\|^2 \quad \text{subject to } V_t(x) - V_t(z_i) \le \langle \xi, x - z_i\rangle \ \ \forall i \in [n] .$$

Lastly, for $\eta > 0$ let $R_\eta^t := \{x - \eta\xi_x : x \in R^t\}$.

The set R^t contains the points for which the minimal selection algorithm is feasible, hence $j \in S^t$ if and only if $x_j^t \in R^t$. Also, ξ_x is well-defined for $x \in R^t$ as discussed in Remark 14. Lastly, R_η^t denotes the region obtained by updating points in R^t according to the minimal selection algorithm. Now, we introduce an assumption to control the proportion of infeasible decision points.

Assumption 1 (Non-expansion condition) *There exists $\eta > 0$ so that $R_\eta^t \subseteq R^{t+1}$ for all $t \in \mathbb{N}$.*

In other words, this assumption guarantees that any feasible point at some round is still feasible after the update; $j \in S^t$ implies $j \in S^{t+1}$. Hence, the proportion $\frac{|[m]\setminus S^t|}{m}$ of infeasible points does not grow. We impose a further assumption that guarantees this proportion decreases.

Assumption 2 (Shrinking condition) *There exist $\eta > 0$ and $\gamma \in (0,1)$ such that for all $t \in \mathbb{N}$,*

$$\inf_{x\in\Omega\setminus R^t} \mathbb{P}_{g\sim N(0,I_d)}\left(x - \sqrt{\eta\frac{\max_{i\in[n]}[V_t(x) - V_t(z_i)]_+}{d}}\, g \in R^{t+1}\right) \ge \gamma . \tag{14}$$

In short, each infeasible point can be updated to a feasible point by a random direction with probability at least γ.[3] Combining this assumption with Assumption 1, we can show that the proportion $\frac{|[m]\setminus S^t|}{m}$ of infeasible decision points shrinks by a factor $1-\gamma$ in each iteration. This observation yields the following regret bound. For simplicity, we assume each V_t is uniformly bounded; see Appendix A.3 for the proof.

Theorem 21 (Shrinking fraction of infeasible decision points) *Assume $\Omega = \mathbb{R}^d$. Let μ_t, ξ_j^t, and g_j^t be the output of the MSoE algorithm. If Assumptions 1 and 2 are satisfied with parameters $\eta, \gamma > 0$, then for any measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ supported on Z,*

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathcal{V}_t(\mu_t) - \sum_{t=1}^{T} \mathcal{V}_t(\nu)\right] \leq \frac{W_2^2(\mu_1,\nu)}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T} \mathbb{E}\left[\frac{1}{m}\sum_{j\in S^t} \|\xi_j^t\|^2\right] + 3B\gamma^{-1}\frac{|[m]\setminus S^1|}{m}\,, \quad (15)$$

where $B = \max_{t\in[T]} \sup_{x\in\Omega} |V_t(x)|$.

3.3. Further Extensions

We briefly discuss a few extensions of our learning problem and the minimal selection algorithm.

Relaxed minimal selection algorithm We introduce another modification of the minimal selection algorithm using slack variables. After round $t \in \mathbb{N}$, for each $j \in [m]$, the player solves

$$\min_{\xi_j\in\mathbb{R}^d, s_j\geq 0} \quad \|\xi_j\|^2 + \frac{2}{\eta}s_j \quad (16)$$

$$\text{subject to} \quad -s_j + V_t(x_j^t) - V_t(z_i) \leq \langle \xi_j, x_j^t - z_i\rangle \;\; \forall i \in [n]\,. \quad (17)$$

After finding minimizers (ξ_j^t, s_j^t), the player updates $x_j^{t+1} = x_j^t - \eta\xi_j^t$, where $\eta > 0$ is a suitable stepsize. Now that variables are ξ_j and s_j, the constraint (17) is always feasible. The resulting optimization problem is still convex and admits a unique solution (ξ_j^t, s_j^t). Also, for $j \in S^t$, that is, for a decision point for which the minimal selection algorithm is feasible, one can easily verify that the relaxed minimal selection algorithm boils down to the minimal selection algorithm. Moreover, we can obtain regret bounds similar to (11) or (13); see (25) and (27) in Appendix B.3.

Interaction functional We consider an extension of our problem to a different class of loss functional. Suppose we replace the potential functional with the interaction functional. Nature reveals a lower semi-continuous function $W^t\colon \mathbb{R}^d \to [0,\infty)$ only on $\{x_j^t\}_{j\in[m]} \cup Z$, that is, the player knows $W^t(x_j^t - z_i)$ for all $(i,j) \in [n]\times[m]$. The player suffers a loss given as the interaction functional associated with W^t (Example 2):

$$\mathcal{W}_t(\mu_t) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} W^t(x-y)\,\mathrm{d}\mu_t(x)\,\mathrm{d}\mu_t(y) = \frac{1}{m^2}\sum_{j,k=1}^{m} W^t(x_j^t - x_k^t)\,. \quad (18)$$

3. Assumptions 1 and 2 are about relations between two adjacent functions V_t and V_{t+1}. We present a simple example of V_t satisfying these assumptions in Appendix B.2.

In this case, we define the minimal selection algorithm as follows: after finishing round $t \in \mathbb{N}$, the player solves

$$\min_{\xi_1,\dots,\xi_m \in \mathbb{R}^d} \quad \frac{1}{m}\sum_{j=1}^{m} \|\xi_j\|^2 \tag{19}$$

$$\text{subject to} \quad \frac{1}{m}\sum_{k=1}^{m} W^t(x_j^t - x_k^t) - \min_{k\in[n]} W^t(z_i - z_k) \le \langle \xi_j, x_j^t - z_i \rangle \ \ \forall (i,j) \in [n]\times[m] \ . \tag{20}$$

Again, we have a convex program that admits a unique solution $(\xi_1^t, \dots, \xi_m^t)$ provided the constraint (20) is feasible. Using the EVI, we can derive a regret bound similar to (11); see Theorem 24 in Appendix B.3.

Projection We briefly mention a projection step to ensure $x_j^t \in \Omega$. The idea is to change the update rule from $x_j^{t+1} = x_j^t - \eta\xi_j^t$ to $x_j^{t+1} = P_\Omega(x_j^t - \eta\xi_j^t)$, where $P_\Omega \colon \mathbb{R}^d \to \Omega$ is defined as

$$P_\Omega(x) = \operatorname*{arg\,min}_{\omega\in\Omega} \|x - \omega\| \ .$$

Recall that the projection P_Ω is well-defined provided Ω is closed and convex. Using this projection step, we can always ensure decision points are contained in Ω. Moreover, this modification does not affect the regret bound (11); see Theorem 25 in Appendix B.3.

4. Discussion

In this paper, we have introduced and studied an infinite-dimensional online learning problem called the Online Learning to Transport (OLT) problem, where the decision variable is a probability measure on a fully nonparametric space, the Wasserstein space. Leveraging tools from optimal transport theory, we have established several theoretical results regarding the OLT problem; equipping the decision space with the Wasserstein distance, we have utilized the evolution variational inequality and the minimal selection principle to derive $\sqrt{T}$-regret bound for the OLT problem. Also, we have studied a practical framework for the OLT problem using zeroth-order information, where we develop a concrete algorithm called the Minimal Selection or Exploration (MSoE) that is numerically tractable and works beyond the convex setting.

Lastly, we mention a few directions for future research. First, theoretical results in Section 2.3 may be improved with a stronger notion of convexity. For instance, one might study if $\log(T)$-regret is achievable based on λ-convexity (Definition 9.1.1 of Ambrosio et al. (2005)) of $\mathcal{E}_t$. Another important direction is to modify the minimal selection strategy with adaptive stepsize η_t. In our work, $\sqrt{T}$-regret is achievable for a fixed stepsize depending on some universal constants as mentioned Remark 9. Hence, to remedy this limitation, we might need to design an adaptive stepsize η_t possibly in terms of $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{t-1}$ (or their norms). Meanwhile, on the practical side, it would be interesting to see how the zeroth-order information framework and the MSoE algorithm work in real-life examples; for instance, we may compare our method with existing methods for real driver allocation datasets, which will shed light on developing even more practical framework and algorithms.

Acknowledgments

Liang acknowledges the generous support from the NSF Career Award (DMS-2042473), and the William S. Fishman Faculty Research Fund at the University of Chicago Booth School of Business.

References

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, 2008.

Jason Altschuler, Jonathan Weed, and Philippe Rigollet. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1961–1971, 2017.

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser Verlag, Basel, 2005.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 214–223, 2017.

Jean-David Benamou and Yann Brenier. A numerical method for the optimal time-continuous mass transport problem and related problems. *Contemporary Mathematics*, 226:1–12, 1999.

Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *omptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique*, 305(19):805–808, 1987.

Yann Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

Charlotte Bunne, David Alvarez-Melis, Andreas Krause, and Stefanie Jegelka. Learning Generative Models across Incomparable Spaces. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 851–861, 2019.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2017.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NIPS)*, volume 26, 2013.

Steven De Rooij, Tim Van Erven, Peter D. Grünwald, and Wouter M. Koolen. Follow the leader if you can, hedge if you must. *J. Mach. Learn. Res.*, 15(1):1281–1316, 2014.

Aymeric Dieuleveut, Alain Durmus, and Francis Bach. Bridging the gap between constant step size stochastic gradient descent and markov chains, 2017. URL https://arxiv.org/abs/1707.06386.

Aude Genevay, Marco Cuturi, Gabriel Peyré, and Francis Bach. Stochastic Optimization for Large-scale Optimal Transport. In *Advances in Neural Information Processing Systems (NIPS)*, volume 29, 2016.

Aude Genevay, Gabriel Peyre, and Marco Cuturi. Learning generative models with Sinkhorn divergences. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 84, pages 1608–1617, 2018.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NIPS)*, volume 27, 2014.

Geoffrey J Gordon. Regret bounds for prediction problems. In *Proceedings of the 12th Annual Conference on Computational Learning Theory (COLT)*, pages 29–40, 1999.

YoonHaeng Hur, Wenxuan Guo, and Tengyuan Liang. Reversible Gromov-Monge sampler for simulation-based inference. *arXiv:2109.14090*, 2021.

Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166 – 1194, 2021.

Jyrki Kivinen and Manfred K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger. From word embeddings to document distances. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 957–966, 2015.

Tengyuan Liang. Estimating certain integral probability metric (IPM) is as hard as estimating under the IPM. *arXiv:1911.00730*, 2019.

Tengyuan Liang. How well generative adversarial networks learn distributions. *Journal of Machine Learning Research*, 22(228):1–41, 2021.

Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal Transport Mapping via Input Convex Neural Networks. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020.

Robert J. McCann. A convexity principle for interacting gases. *Advances in Mathematics*, 128(1): 153–179, 1997.

H Brendan McMahan and Matthew Streeter. Adaptive bound optimization for online convex optimization. *arXiv:1002.4908*, 2010.

Facundo Mémoli. Gromov–Wasserstein Distances and the Metric Approach to Object Matching. *Foundations of Computational Mathematics*, 11(4):417–487, August 2011.

Francesco Orabona. A Modern Introduction to Online Learning. *arXiv:1912.13213*, 2021.

Francesco Orabona and Dávid Pál. Scale-free algorithms for online linear optimization. In *Proceedings of the 26th International Conference on Algorithmic Learning Theory (COLT)*, pages 287–301, 2015.

Francesco Orabona and Dávid Pál. Scale-free online learning. *Theoretical Computer Science*, 716: 50–69, 2018.

Felix Otto. The geometry of dissipative evolution equations: The porous medium equation. *Communications in Partial Differential Equations*, 26(1&2):101–174, 2001.

Alexander Rakhlin and Karthik Sridharan. *Statistical Learning and Sequential Prediction*. 2014.

Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, 40(2):99–121, November 2000.

Adil Salim, Anna Korba, and Giulia Luise. The wasserstein proximal gradient algorithm. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12356–12366. Curran Associates, Inc., 2020.

Vivien Seguy, Bharath Bhushan Damodaran, Rémi Flamary, Nicolas Courty, Antoine Rolet, and Mathieu Blondel. Large-Scale Optimal Transport and Mapping Estimation. In *International Conference on Learning Representations (ICLR)*, 2018.

Shai Shalev-Shwartz. Online Learning and Online Convex Optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Shai Shalev-Shwartz and Yoram Singer. A primal-dual perspective of online learning algorithms. *Machine Learning*, 69(2):115–142, December 2007.

Matthew Streeter and H Brendan McMahan. Less regret via online conditioning. *arXiv:1002.4862*, 2010.

Tim van Erven, Peter Grunwald, Wouter M. Koolen, and Steven de Rooij. Adaptive hedge. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1656–1664, 2011.

Cédric Villani. *Topics in Optimal Transportation*. American Mathematical Society, Providence, RI, 2003.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620 – 2648, 2019.

Jonathan Weed and Quentin Berthet. Estimation of smooth densities in Wasserstein distance. In *Proceedings of the 32nd Conference on Learning Theory (COLT)*, pages 3118–3119, 2019.

Hongyi Zhang and Suvrit Sra. First-order methods for geodesically convex optimization. In *Proceedings of the 29th Annual Conference on Learning Theory (COLT)*, volume 49, pages 1617–1638, 2016.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, pages 928–936, 2003.

Appendix A. Proofs of the Regret Bounds in Section 3

A.1. Proof of Theorem 15

By convexity of V_t, the constraint set (10) is always feasible, hence ξ_j^t is well-defined. Let $\nu = \sum_i a_i \delta_{z_i}$ for some $a_1, \dots, a_n \geq 0$ such that $\sum_{i=1}^n a_i = 1$. We can find an optimal coupling between μ_t and ν, which takes the following form: $\sum_{i=1}^n \sum_{j=1}^m \pi_{ji} \delta_{(x_j^t, z_i)}$, where $\sum_{i=1}^n \pi_{ji} = \frac{1}{m}$ and $\sum_{j=1}^m \pi_{ji} = a_i$. Note that $\sum_{i=1}^n \sum_{j=1}^m \pi_{ji} \delta_{(x_j^{t+1}, z_i)}$ is a coupling between ν and μ_{t+1} as well. Hence,

$$\begin{aligned} W_2^2(\mu_{t+1}, \nu) - W_2^2(\mu_t, \nu) &\leq \sum_{i=1}^n \sum_{j=1}^m \|x_j^{t+1} - z_i\|^2 \pi_{ji} - \sum_{i=1}^n \sum_{j=1}^m \|x_j^t - z_i\|^2 \pi_{ji} \\ &= \sum_{i=1}^n \sum_{j=1}^m \left(\eta^2 \|\xi_j^t\|^2 - 2\eta \langle \xi_j^t, x_j^t - z_i \rangle \right) \pi_{ji} \\ &\leq \sum_{i=1}^n \sum_{j=1}^m \left(\eta^2 \|\xi_j^t\|^2 - 2\eta (V_t(x_j^t) - V_t(z_i)) \right) \pi_{ji} \\ &= \eta^2 \left(\frac{1}{m} \sum_{j=1}^m \|\xi_j^t\|^2 \right) - 2\eta \left(\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu) \right) . \end{aligned}$$

Therefore, we have

$$\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu) \leq \frac{W_2^2(\mu_t, \nu) - W_2^2(\mu_{t+1}, \nu)}{2\eta} + \frac{\eta}{2} \cdot \frac{1}{m} \sum_{j=1}^m \|\xi_j^t\|^2 , \tag{21}$$

and (11) follows by summing (21) iteratively.

A.2. Proof of Theorem 19

We first prove the following:

$$\begin{aligned} \mathbb{E}\left[\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu) \mid \mathcal{F}_t \right] \leq & \frac{\mathbb{E}\left[W_2^2(\mu_t, \nu) - W_2^2(\mu_{t+1}, \nu) \mid \mathcal{F}_t \right]}{2\eta} \\ & + \frac{\eta}{2} \cdot \frac{1}{m} \sum_{j \in S^t} \|\xi_j^t\|^2 + \frac{3}{2} \cdot \frac{1}{m} \sum_{j \notin S^t} \max_{i \in [n]} [V_t(x_j^t) - V_t(z_i)]_+ , \end{aligned} \tag{22}$$

where $\mathcal{F}_t$ denotes the σ-field generated by $\{g_j^s : \forall s < t \text{ and } \forall j \in S^s\}$ for $t > 1$ and $\mathcal{F}_1$ is the trivial σ-field. We write $x_j^{t+1} = x_j^t - v_j^t$, where

$$v_j^t = \sqrt{\eta \frac{\max_{i \in [n]} [V_t(x_j^t) - V_t(z_i)]_+}{d}} g_j^t \quad \text{if } j \notin S^t$$

and $v_j^t = \eta \xi_j^t$ otherwise. As in the proof of Theorem 15, let $\sum_{i=1}^n \sum_{j=1}^m \pi_{ji} \delta_{(x_j^t, z_i)}$ be an optimal coupling between μ_t and $\nu = \sum_i a_i \delta_{z_i}$. Again, it is a coupling between ν and μ_{t+1}, hence

$$\begin{aligned} W_2^2(\mu_{t+1}, \nu) - W_2^2(\mu_t, \nu) &\le \sum_{i=1}^n \sum_{j=1}^m \|x_j^{t+1} - z_i\|^2 \pi_{ji} - \sum_{i=1}^n \sum_{j=1}^m \|x_j^t - z_i\|^2 \pi_{ji} \\ &= \sum_{i=1}^n \sum_{j=1}^m \left(\|v_j^t\|^2 - 2\langle v_j^t, x_j^t - z_i \rangle \right) \pi_{ji} \\ &= \frac{1}{m} \sum_{j=1}^m \|v_j^t\|^2 - 2 \sum_{i=1}^n \sum_{j=1}^m \langle v_j^t, x_j^t - z_i \rangle \pi_{ji} \ . \end{aligned}$$

Now, taking the conditional expectation $\mathbb{E}[\cdot \mid \mathcal{F}_t]$, we have

$$\mathbb{E}[W_2^2(\mu_{t+1}, \nu) - W_2^2(\mu_t, \nu) \mid \mathcal{F}_t] \le \frac{1}{m} \sum_{j=1}^m \mathbb{E}\left[\|v_j^t\|^2 \mid \mathcal{F}_t \right] - 2 \sum_{i=1}^n \sum_{j=1}^m \mathbb{E}\left[\langle v_j^t, x_j^t - z_i \rangle \pi_{ji} \mid \mathcal{F}_t \right] \ .$$

Now, we upper bound the last two terms. For $j \in S^t$, we have $\mathbb{E}[\|v_j^t\|^2 \mid \mathcal{F}_t] = \eta^2 \|\xi_j^t\|^2$ because $v_j^t = \xi_j^t$ is measurable with respect to $\mathcal{F}_t$. Meanwhile, for $j \notin S^t$, since $\mathbb{E}\, \|g_j^t\|^2 = d$,

$$\mathbb{E}\left[\|v_j^t\|^2 \mid \mathcal{F}_t \right] = \eta \max_{i \in [n]} [V_t(x_j^t) - V_t(z_i)]_+ \ .$$

Hence,

$$\frac{1}{m} \sum_{j=1}^m \mathbb{E}\left[\|v_j^t\|^2 \mid \mathcal{F}_t \right] = \frac{\eta^2}{m} \sum_{j \in S^t} \|\xi_j^t\|^2 + \frac{\eta}{m} \sum_{j \notin S^t} \max_{i \in [n]} [V_t(x_j^t) - V_t(z_i)]_+ \ . \tag{23}$$

Next, for $j \in S^t$, recall that

$$\langle v_j^t, x_j^t - z_i \rangle \ge \eta \big(V_t(x_j^t) - V_t(z_i) \big) \ .$$

For $j \notin S^t$, since $\mathbb{E}\, g_j^t = 0$,

$$\mathbb{E}\left[\langle v_j^t, x_j^t - z_i \rangle \mid \mathcal{F}_t \right] = 0 \ge \eta \big(V_t(x_j^t) - V_t(z_i) \big) - \eta \max_{i \in [n]} [V_t(x_j^t) - V_t(z_i)]_+ \ .$$

Hence,

$$\begin{aligned} &- 2 \sum_{i=1}^n \sum_{j=1}^m \mathbb{E}\left[\langle v_j^t, x_j^t - z_i \rangle \pi_{ji} \mid \mathcal{F}_t \right] \\ &\le -2\eta \sum_{i=1}^n \sum_{j \in S^t} \big(V_t(x_j^t) - V_t(z_i) \big) \pi_{ji} \\ &\quad - 2\eta \sum_{i=1}^n \sum_{j \notin S^t} \big(V_t(x_j^t) - V_t(z_i) \big) \pi_{ji} + 2\eta \sum_{i=1}^n \sum_{j \notin S^t} \max_{i \in [n]} [V_t(x_j^t) - V_t(z_i)]_+ \pi_{ji} \\ &= -2\eta \sum_{i=1}^n \sum_{j=1}^m \big(V_t(x_j^t) - V_t(z_i) \big) \pi_{ji} + \frac{2\eta}{m} \sum_{j \notin S^t} \max_{i \in [n]} [V_t(x_j^t) - V_t(z_i)]_+ \\ &= -2\eta \left(\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu) \right) + \frac{2\eta}{m} \sum_{j \notin S^t} \max_{i \in [n]} [V_t(x_j^t) - V_t(z_i)]_+ \ . \end{aligned}$$

Combining this result with (23) and using $\mathbb{E}\left[\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu) \mid \mathcal{F}_t\right] = \mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu)$, we obtain (22). Now, taking the expectation to the both sides of (22) and sum iteratively to obtain (13).

A.3. Proof of Theorem 21

First, using uniform boundedness of V_t, we have

$$\sum_{j \notin S^t} \mathbb{E}\left[\max_{i\in[n]}[V_t(x_j^t) - V_t(z_i)]_+\right] \le 2B \sum_{j=1}^m 1_{j\notin S^t} \ .$$

Combining this with (22) and taking the expectation, we have

$$\mathbb{E}\left[\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu)\right] \le \frac{\mathbb{E}\left[W_2^2(\mu_t,\nu) - W_2^2(\mu_{t+1},\nu)\right]}{2\eta} + \frac{\eta}{2}\mathbb{E}\left[\frac{1}{m}\sum_{j\in S^t}\|\xi_j^t\|^2\right] + \frac{3B}{m}\sum_{j=1}^m \mathbb{E}[1_{j\notin S^t}] \ .$$

Summing this over $t \in [T]$, we have

$$\mathbb{E}\left[\sum_{t=1}^T \mathcal{V}_t(\mu_t) - \sum_{t=1}^T \mathcal{V}_t(\nu)\right] \le \frac{W_2^2(\mu_1,\nu)}{2\eta} + \frac{\eta}{2}\sum_{t=1}^T \mathbb{E}\left[\frac{1}{m}\sum_{j\in S^t}\|\xi_j^t\|^2\right] + \frac{3B}{m}\sum_{t=1}^T\sum_{j=1}^m \mathbb{E}[1_{j\notin S^t}] \ . \tag{24}$$

Now, we upper bound the last term on the right-hand side. Since $j \notin S^t$ implies $j \notin S^{t-1}$, we have

$$\begin{aligned}
&\mathbb{E}\left[1_{j\notin S^t} \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[1_{j\notin S^{t-1} \text{ and } j\notin S^t} \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[1_{x_j^t \notin R^t} \mid \mathcal{F}_{t-1}\right] 1_{x_j^{t-1}\in\Omega\setminus R^{t-1}} \\
&= \mathop{\mathbb{P}}_{g\sim N(0,I)}\left(x_j^{t-1} - \sqrt{\eta\tfrac{\max_{i\in[n]}[V_{t-1}(x)-V_{t-1}(z_i)]_+}{d}}\, g \notin R^t\right) 1_{x_j^{t-1}\in\Omega\setminus R^{t-1}} \\
&\le (1-\gamma) 1_{x_j^{t-1}\in\Omega\setminus R^{t-1}} \\
&= (1-\gamma) 1_{j\notin S^{t-1}} \ ,
\end{aligned}$$

where the inequality is due to Assumption 2. Therefore, taking the conditional expectation recursively according to the filtration, we have $\mathbb{E}[1_{j\notin S^t}] \le (1-\gamma)^{t-1} 1_{j\notin S^1}$. Hence,

$$\sum_{j=1}^m\sum_{t=1}^T \mathbb{E}[1_{j\notin S^t}] \le (1 + (1-\gamma) + \cdots + (1-\gamma)^{T-1}) \sum_{j=1}^m 1_{j\notin S^1} \le \gamma^{-1}|[m]\setminus S^1| \ .$$

Combining this with (24), we obtain (15).

Appendix B. Supplementary Explanations

B.1. Proof of Theorem 11

The proof uses differentiability of Wasserstein distance (Theorem 8.13 of Villani (2003)). If $\boldsymbol{\xi}_t(x)$ is a C^1 function of x and t that is globally bounded, for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, one can calculate that

$$\left.\frac{\mathrm{d}W_2^2(\mu,\mu_t)}{\mathrm{d}t}\right|_{t=s} = 2\int \left\langle x - \mathbf{t}_{\mu_s}^{\mu}(x), \boldsymbol{\xi}_s(x)\right\rangle \mathrm{d}\mu_s(x) \le 2[\mathcal{E}_s(\mu) - \mathcal{E}_s(\mu_s)] \ .$$

The inequality is due to the characterization of Fréchet subdifferential in Lemma 6. Integrating over $s \in [0, T]$, we obtain the regret bound.

B.2. Assumptions 1 and 2

First, we illustrate the notions of feasible and infeasible regions for some loss functions in one dimension. In Figure1, we plot a w-shape loss function V. For simplicity, we set two global minimizers z_1, z_2 as grid points, and x_1, x_2 as player's decision points. One can see that x_1 is feasible where the minimal selection subdifferential ξ_1 is the slope of the blue line l_1, passing through $(x_1, V(x_1))$, $(z_1, V(z_1))$. The decision point x_2, lying in the barrier between two global minimizers, is infeasible: for any line passing through $(x_2, V(x_2))$, it is impossible to have smaller values than the loss function at both grid points z_1, z_2. The red lines in the figure are two attempts. With this intuition, one can verify that the feasible region is $(-\infty, z_1] \cup [z_2, +\infty)$, whereas the infeasible region is (z_1, z_2).

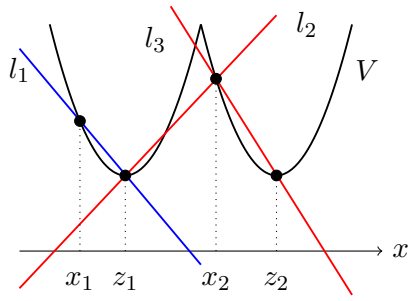


Figure 1: An illustration of a w-shape non-convex loss $V : \mathbb{R} \to \mathbb{R}$. Two global minimizers z_1, z_2 serve as grid points. x_1, x_2 are player's decision points.

Next, we showcase a series of w-shape non-convex functions that satisfy Assumptions 1, 2.

Lemma 22 *Consider loss functions*

$$V_t(x) = \begin{cases} a^t(x+1)^2, & \text{if } x < 0, \\ a^t(x-1)^2, & \text{if } x \geq 0, \end{cases}$$

and grid points $-1 = z_1 < z_2 < \cdots < z_n = 1$ *for some arbitrary* n. *If there exists* $\epsilon > 0$ *such that* $a^t \geq \epsilon$, $a^t\eta < \frac{1}{2}$ *for all* t, *then Assumptions 1, 2 hold.*

Proof The feasible region is $R^t = (-\infty, -1] \cup [1, \infty)$ for all t. We first verify Assumption 1. Without loss of generality, we consider positive feasible points $x^t = 1 + \delta^t \in R^t$ for some $\delta^t \geq 0$. The minimal selection of subdifferential returns $\xi_x = 2a^t\delta^t$. Therefore $x^{t+1} = x^t - 2\eta a^t\delta^t = 1 + (1 - 2a^t\eta)\delta$. Under the condition $a^t\eta < \frac{1}{2}$, we have $x^{t+1} \geq 1$, i.e., $R^t_\eta \subset R^{t+1}$.

Now, we verify Assumption 2. For any infeasible $x^t \in [0, 1)$, update by random exploration

$$x^{t+1} = x_t - \sqrt{\eta V_t(x^t)}g \ ,$$

where $g \sim \mathcal{N}(0,1)$. Then,

$$\begin{aligned}\mathbb{P}(x^{t+1} \in R^{t+1}) &\geq \mathbb{P}(x^t - \sqrt{\eta V_t(x^t)}g \geq 1) \\ &= \mathbb{P}(-\sqrt{\eta a^t}(1-x^t)g \geq 1 - x^t) \\ &= \mathbb{P}\left(g \leq -\frac{1}{\sqrt{a^t \eta}}\right) \\ &\geq \mathbb{P}\left(g \leq -\frac{1}{\sqrt{\epsilon\eta}}\right) ,\end{aligned}$$

where the last inequality is due to the condition $a^t \geq \epsilon$. Note that the lower bound above holds for any $x \in [0,1)$. By a symmetric argument, we conclude that Assumption 2 is satisfied with $\gamma = \mathbb{P}(g \leq -1/\sqrt{\epsilon\eta})$. ■

B.3. Results in Subsection 3.3

Here, we provide detailed explanations on the results presented in Subsection 3.3. First, using a similar EVI technique, we derive the following regret bound for the relaxed minimal selection algorithm.

Theorem 23 *Assume $\Omega = \mathbb{R}^d$. Let μ_t, ξ_j^t, and s_j^t be the output of the relaxed minimal selection algorithm. For any V_t and any measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ supported on Z,*

$$\sum_{t=1}^{T} \mathcal{V}_t(\mu_t) - \sum_{t=1}^{T} \mathcal{V}_t(\nu) \leq \frac{W_2^2(\mu_1,\nu) - W_2^2(\mu_{T+1},\nu)}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\frac{1}{m}\sum_{j=1}^{m}\left(\|\xi_j^t\|^2 + \frac{2s_j^t}{\eta}\right) . \quad (25)$$

Proof As in the proof of Theorem 15, let $\sum_{i=1}^{n}\sum_{j=1}^{m}\pi_{ji}\delta_{(x_j^t, z_i)}$ be an optimal coupling between μ_t and $\nu = \sum_i a_i \delta_{z_i}$. Again, it is a coupling between ν and μ_{t+1}, hence

$$\begin{aligned}W_2^2(\mu_{t+1},\nu) - W_2^2(\mu_t,\nu) &\leq \sum_{i=1}^{n}\sum_{j=1}^{m}\|x_j^{t+1} - z_i\|^2\pi_{ji} - \sum_{i=1}^{n}\sum_{j=1}^{m}\|x_j^t - z_i\|^2\pi_{ji} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\eta^2\|\xi_j^t\|^2 - 2\eta\langle\xi_j^t, x_j^t - z_i\rangle\right)\pi_{ji} \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\eta^2\|\xi_j^t\|^2 + 2\eta s_j^t - 2\eta(V_t(x_j^t) - V_t(z_i))\right)\pi_{ji} \\ &= \frac{\eta^2}{m}\sum_{j=1}^{m}\left(\|\xi_j^t\|^2 + \frac{2s_j^t}{\eta}\right) - 2\eta\left(\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu)\right) .\end{aligned}$$

Hence,

$$\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu) \leq \frac{W_2^2(\mu_t,\nu) - W_2^2(\mu_{t+1},\nu)}{2\eta} + \frac{\eta}{2}\cdot\frac{1}{m}\sum_{j=1}^{m}\left(\|\xi_j^t\|^2 + \frac{2s_j^t}{\eta}\right) \quad (26)$$

(25) follows by summing (26) iteratively. ■

In fact, we can further upper bound (26). Note that $\xi_j = 0$ and $s_j = \max_{i\in[n]}[V_t(x_j^t) - V_t(z_i)]_+$ satisfy the constraint (17), hence the minimizer (ξ_j^t, s_j^t) satisfies

$$\|\xi_j^t\|^2 + \frac{2s_j^t}{\eta} \le \frac{2\max_{i\in[n]}[V_t(x_j^t) - V_t(z_i)]_+}{\eta} .$$

Using this upper bound for $j \notin S^t$, we have

$$\begin{aligned}\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu) \le \frac{W_2^2(\mu_t,\nu) - W_2^2(\mu_{t+1},\nu)}{2\eta} \\ + \frac{\eta}{2}\cdot\frac{1}{m}\sum_{j\in S^t}\|\xi_j^t\|^2 + \frac{1}{m}\sum_{j\notin S^t}\max_{i\in[n]}[V_t(x_j^t) - V_t(z_i)]_+ .\end{aligned} \tag{27}$$

Notice that this upper bound is comparable to (22). Therefore, the relaxed minimal selection algorithm and the minimal selection or exploration algorithm admit essentially the same upper bound.

Next, we present a regret bound for the case where the loss is given according to the interaction functional.

Theorem 24 *Assume $\Omega = \mathbb{R}^d$ and consider the game with the loss (18) discussed in Subsection 3.3. Let μ_t and ξ_j^t be the output of the minimal selection algorithm with the constraint (20). If W^t is convex, for any measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ supported on Z,*

$$\sum_{t=1}^T \mathcal{W}_t(\mu_t) - \sum_{t=1}^T \mathcal{W}_t(\nu) \le \frac{W_2^2(\mu_1,\nu) - W_2^2(\mu_{T+1},\nu)}{2\eta} + \frac{\eta}{2}\sum_{t=1}^T\left(\frac{1}{m}\sum_{j=1}^m\|\xi_j^t\|^2\right) . \tag{28}$$

Proof As in the proof of Theorem 15, let $\sum_{i=1}^n\sum_{j=1}^m \pi_{ji}\delta_{(x_j^t,z_i)}$ be an optimal coupling between μ_t and $\nu = \sum_i a_i\delta_{z_i}$. Again, it is a coupling between ν and μ_{t+1}, hence

$$\begin{aligned}W_2^2(\mu_{t+1},\nu) - W_2^2(\mu_t,\nu) &\le \sum_{i=1}^n\sum_{j=1}^m\|x_j^{t+1} - z_i\|^2\pi_{ji} - \sum_{i=1}^n\sum_{j=1}^m\|x_j^t - z_i\|^2\pi_{ji} \\ &= \sum_{i=1}^n\sum_{j=1}^m\left(\eta^2\|\xi_j^t\|^2 - 2\eta\langle\xi_j^t, x_j^t - z_i\rangle\right)\pi_{ji} .\end{aligned}$$

Using (20),

$$\begin{aligned}\sum_{i=1}^n\sum_{j=1}^m -2\eta\langle\xi_j^t, x_j^t - z_i\rangle\pi_{ji} &\le \sum_{i=1}^n\sum_{j=1}^m -2\eta\left(\frac{1}{m}\sum_{k=1}^m W^t(x_j^t - x_k^t) - \min_{k\in[n]} W^t(z_i - z_k)\right)\pi_{ji} \\ &= -2\eta\left(\frac{1}{m^2}\sum_{j,k=1}^m W^t(x_j^t - x_k^t) - \sum_{i=1}^n a_i\cdot\min_{k\in[n]} W^t(z_i - z_k)\right) \\ &\le -2\eta\left(\frac{1}{m^2}\sum_{j,k=1}^m W^t(x_j^t - x_k^t) - \sum_{i=1}^n a_i\sum_{k=1}^n a_k W^t(z_i - z_k)\right) \\ &= -2\eta\left(\mathcal{W}_t(\mu_t) - \mathcal{W}_t(\nu)\right) .\end{aligned}$$

Therefore, we have

$$\mathcal{W}_t(\mu_t) - \mathcal{W}_t(\nu) \leq \frac{W_2^2(\mu_t, \nu) - W_2^2(\mu_{t+1}, \nu)}{2\eta} + \frac{\eta}{2} \cdot \frac{1}{m} \sum_{j=1}^{m} \|\xi_j^t\|^2 \,, \tag{29}$$

and (28) follows by summing (29) iteratively. ■

Lastly, we prove that the projection step does not affect the regret bound.

Theorem 25 *Assume Ω is closed and convex. Let μ_t and ξ_j^t be the output of the minimal selection algorithm with a modified update rule $x_j^{t+1} = P_\Omega(x_j^t - \eta\xi_j^t)$. If V_t is convex, for any measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ supported on Z,*

$$\sum_{t=1}^{T} \mathcal{V}_t(\mu_t) - \sum_{t=1}^{T} \mathcal{V}_t(\nu) \leq \frac{W_2^2(\mu_1, \nu) - W_2^2(\mu_{T+1}, \nu)}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \left(\frac{1}{m} \sum_{j=1}^{m} \|\xi_j^t\|^2 \right) \,.$$

Proof As in the proof of Theorem 15, let $\sum_{i=1}^{n}\sum_{j=1}^{m} \pi_{ji}\delta_{(x_j^t, z_i)}$ be an optimal coupling between μ_t and $\nu = \sum_i a_i \delta_{z_i}$. Again, it is a coupling between ν and μ_{t+1}, hence

$$W_2^2(\mu_{t+1}, \nu) - W_2^2(\mu_t, \nu) \leq \sum_{i=1}^{n}\sum_{j=1}^{m} \|x_j^{t+1} - z_i\|^2 \pi_{ji} - \sum_{i=1}^{n}\sum_{j=1}^{m} \|x_j^t - z_i\|^2 \pi_{ji} \,.$$

One property of the projection P_Ω is that $\|x_j^{t+1} - \omega\| \leq \|x_j^t - \eta\xi_j^t - \omega\|$ for all $\omega \in \Omega$. Thus,

$$\begin{aligned}
W_2^2(\mu_{t+1}, \nu) - W_2^2(\mu_t, \nu) &\leq \sum_{i=1}^{n}\sum_{j=1}^{m} \|x_j^{t+1} - z_i\|^2 \pi_{ji} - \sum_{i=1}^{n}\sum_{j=1}^{m} \|x_j^t - z_i\|^2 \pi_{ji} \\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{m} \|x_j^t - \eta\xi_j^t - z_i\|^2 \pi_{ji} - \sum_{i=1}^{n}\sum_{j=1}^{m} \|x_j^t - z_i\|^2 \pi_{ji} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \left(\eta^2 \|\xi_j^t\|^2 - 2\eta \langle \xi_j^t, x_j^t - z_i \rangle \right) \pi_{ji} \\
&\leq \eta^2 \left(\frac{1}{m} \sum_{j=1}^{m} \|\xi_j^t\|^2 \right) - 2\eta \left(\mathcal{V}_t(\mu_t) - \mathcal{V}_t(\nu) \right) \,,
\end{aligned}$$

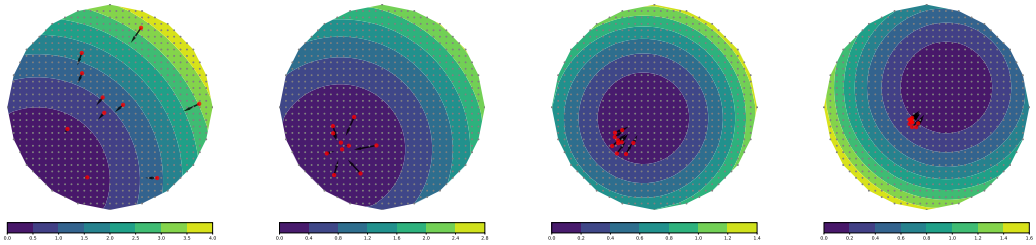
where the last inequality directly follows as in the proof of Theorem 15. Therefore, the regret bounds in Theorem 15 still hold. ■

Remark 26 *We can apply this projection step to the MSoE algorithm and the relaxed minimal selection algorithm as well; we modify their update rules by combining them with P_Ω. Then, the regret bounds in Theorems 19 and 23 still hold.*

Appendix C. Simulations

In this section, we examine the empirical performance of the minimal selection and the MSoE algorithms. Here, we consider a toy example where $\Omega = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$, $m = 10$, and Z consists of uniform grid points of Ω with $n = 797$, obtained by forming uniform grid points of $[-1, 1]^2$ and choose a subset of included in Ω. Code for reproducing all the results below is provided in the supplementary material.

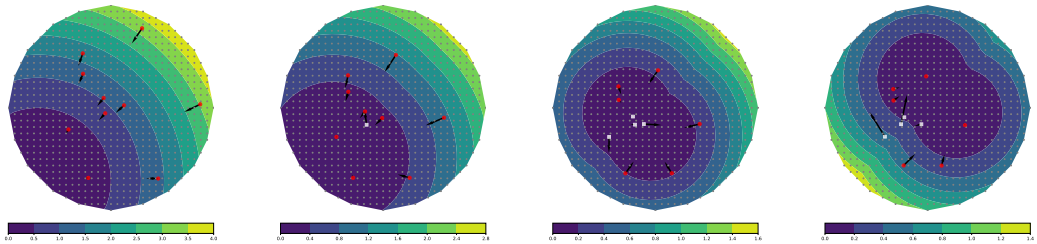
Convex case First, we consider a simple quadratic function $V_t(x) = \|x - u_t\|^2$, where $u_1 = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ and $u_t = u_1 + (0.15t, 0.15t)$. For better understanding, we may imagine a situation, where we deploy m drones to track a moving target u_t using the signals V_t (distances to the target) captured at Z (fixed stations with sensors) and $\{x_j^t\}_{j\in[m]}$ (drones with mobile sensors). Figure 2 shows how the minimal selection algorithm moves the decision points. At $t = 1$ (Figure 2(a)), the decision points, which are randomly initialized, start moving towards the darker region based on ξ_j^t's from the minimal selection algorithm. In this simple toy example, we can see that the decision points quickly gather around the minimum of V_t (the target) and follow it.



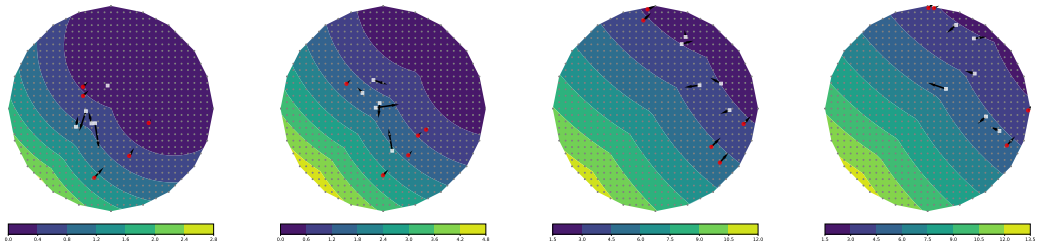
(*a*) $t = 1$ (*b*) $t = 3$ (*c*) $t = 5$ (*d*) $t = 7$

Figure 2: Minimal selection algorithm for convex V_t with stepsize $\eta = 0.2$. The gray dots and the red circles denote Z and $\{x_j^t\}_{j\in[m]}$, respectively. The black solid arrows show ξ_j^t's. The contour regions represent the level of V_t (darker = smaller as shown in the horizontal colorbars).

Non-convex case As a simple non-convex example, consider $V_t(x) = \min\{\|x-u_t\|^2, \|x-v_t\|^2\}$, where $u_1 = v_1 = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$, $u_t = u_1 + (0.165t, 0.11t)$, and $v_t = v_1 + (0.11t, 0.165t)$. As in the convex case, we may interpret u_t and v_t as moving targets, while the signal is the distance to a closer target. Figure 3 shows how the MSoE algorithm works. As opposed to the convex case, we can check that there are infeasible decision points ($j \notin S^t$ as in Definition 18) after $t = 3$. The MSoE algorithm let such points move along random directions, thereby continuing the tracking situation; the plain minimal selection would have ended up stopping all the decision points, losing the targets. Although infeasible points might get far away from the targets as in Figure 3(e) and Figure 3(f), once they get into a feasible region, they start moving again towards the darker area based on ξ_j^t's from the minimal selection. Figure 3(g) and Figure 3(h) show that all the decision points somehow get closer to the darker area after repeating the minimal selection and exploration.



(a) $t = 1$ (b) $t = 3$ (c) $t = 5$ (d) $t = 7$



(e) $t = 9$ (f) $t = 12$ (g) $t = 18$ (h) $t = 19$

Figure 3: MSoE algorithm for non-convex case with stepsize $\eta = 0.05$. The red circles denote the feasible decision points ($j \in S^t$) and the white squares denote the infeasible decision points ($j \notin S^t$).