



HAL
open science

Mean Field Optimization problems: stability results and Lagrangian discretization

Kang Liu, Laurent Pfeiffer

► **To cite this version:**

Kang Liu, Laurent Pfeiffer. Mean Field Optimization problems: stability results and Lagrangian discretization. 2023. hal-04269378

HAL Id: hal-04269378

<https://inria.hal.science/hal-04269378>

Preprint submitted on 3 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

MEAN FIELD OPTIMIZATION PROBLEMS: STABILITY RESULTS AND LAGRANGIAN DISCRETIZATION

KANG LIU^{1,2} AND LAURENT PFEIFFER¹

ABSTRACT. We formulate and investigate a mean field optimization (MFO) problem over a set of probability distributions μ with a prescribed marginal m . The cost function depends on an aggregate term, which is the expectation of μ with respect to a contribution function. This problem is of particular interest in the context of Lagrangian potential mean field games (MFGs) and their discretization. We provide a first-order optimality condition and prove strong duality. We investigate stability properties of the MFO problem with respect to the prescribed marginal, from both primal and dual perspectives. In our stability analysis, we propose a method for recovering an approximate solution to an MFO problem with the help of an approximate solution to an MFO with a different marginal m , typically an empirical distribution. We combine this method with the stochastic Frank-Wolfe algorithm of [6] to derive a complete resolution method.

Keywords: optimization with probability measures, potential mean field games, non-atomic games, stability analysis, Frank-Wolfe algorithm.

1. INTRODUCTION

This article is dedicated to a general class of optimization problems involving probability measures with a prescribed marginal m . We will refer to them as Mean Field Optimization (MFO) problems. These typically arise in multi-agent optimization problems, for which a mean-field formulation of the problem, involving the probability distribution of the decisions of the agents (rather than an enumeration of them), is not only meaningful but also provides us with convexity properties of great numerical interest.

The first ambition of our work is to provide a general framework for the formulation of such situations. For the sake of clarity, we introduce here the MFO problems investigated in this work. We refer the reader to Sec. 2.3 for a complete description of the required assumptions. Let X and Y be two complete and separable metric spaces and let \mathcal{H} be a separable Hilbert space. Let Z be a closed subset of $X \times Y$ and let m be a probability measure on X . We consider the following problem, parametrized by m :

$$(P_m) \quad \inf_{\mu \in \mathcal{P}_m(Z)} f \left(\int_Z g(x, y) d\mu(x, y) \right),$$

where $g: Z \rightarrow \mathcal{H}$ is a Borel measurable function and $f: \mathcal{H} \rightarrow \mathbb{R}$ is a convex function. The admissible set $\mathcal{P}_m(Z)$ is the set of all probability measures on Z whose marginal distribution on X is m . Our model allows for heterogeneity within the agents, which is modeled by some parameter $x \in X$. The decision variables of the agents are generically denoted by $y \in Y$ and the probability measure μ represents the distribution of the parameter-decision pairs (x, y) of our agents. The distribution of the parameters of the agents is given by m , that is why we impose that μ has its first marginal

¹UNIVERSITÉ PARIS-SACLAY, CNRS, CENTRALESUPÉLEC, INRIA, LABORATOIRE DES SIGNAUX ET SYSTÈMES, 91190, GIF-SUR-YVETTE, FRANCE.

²INSTITUT POLYTECHNIQUE DE PARIS, CNRS, ECOLE POLYTECHNIQUE, CMAP, 91120 PALAISEAU, FRANCE.

E-mail address: kang.liu@polytechnique.edu, laurent.pfeiffer@inria.fr.

Date: November 1, 2023.

equal to m . At an abstract level, we can interpret the term $\int_Z g d\mu$ as a common good, obtained by aggregating the contributions of all the agents.

Motivation. Our original interest for MFO problems comes from non-atomic games with a potential structure, for which finding a Nash equilibrium is equivalent to solving an MFO problem. Among these games, we have a special interest for Lagrangian Mean Field Games (MFGs), in which the agents each optimize the trajectory of some dynamical system and are parametrized by their initial condition. Problem (P_m) also arises in energy management problems, more specifically, in problems involving many small consumption (or production) units, for example electrical cars. For such problems, the common good is the total energy consumption, the parameter x could model any relevant characteristic of the cars (such as their charging capacity) while the variable y describes their charging profiles. Finally, let us mention that MFO problems find applications in supervised learning, more specifically in the training of neuron networks with one hidden layer: in the mean field approximation of such problems, the variable μ simply describes the probability distribution of the weights of the neurons [13, 24, 12]. For other applications of MFO problems in learning, we refer to [7, 13, 35] and the references therein.

Numerical approach. The numerical resolution of Problem (P_m) poses two main difficulties: the numerical manipulation of probability measures and the treatment of the marginal constraint. Let us focus on the first difficulty by supposing momentarily that m is a Dirac measure located at some point x^* , so that the problem (P_m) can simply be written as an optimization problem on $\mathcal{P}(Y)$. Unless Y is finite, $\mathcal{P}(Y)$ is an infinite dimensional set. A first approach would consist in discretizing the set Y , which would preserve the convexity of the problem. This approach suffers from the curse of dimensionality, since it requires an exponential number of points with respect to the dimension of Y . A second approach would consist in representing the probability measure as the empirical mean of a set of N points to be optimized. The clear drawback of this approach is the loss of convexity of the discretized problem; yet we mention that it proves efficient in the context of supervised learning problem [13]. In the general context of MFO problems, the Frank-Wolfe (FW) algorithm is a particularly advantageous algorithm. It produces a finitely supported approximate solution and leverages the convexity of the original problem (without using a coarse discretization of Y). More specifically, it generates after k iterations an $\mathcal{O}(1/k)$ -optimal solution supported by at most k points.

Coming back to the case of a general marginal, we propose to discretize m with an empirical measure m_N associated with N points and thus to solve:

$$(P_{m_N}) \quad \inf_{\mu \in \mathcal{P}_{m_N}(Z)} f \left(\int_Z g d\mu \right).$$

This idea was already proposed in [30], in an MFG context. By the disintegration theorem, (P_{m_N}) is equivalent to a problem involving N probability measures. A direct implementation of the FW algorithm would lead to an approximate solution possibly involving kN points after k iterations of the algorithm. We will see that the Stochastic Frank-Wolfe algorithm, which we introduced and analyzed in [6], allows to obtain an approximate solution of (P_{m_N}) relying on only N support points, to the price of an additional error term of order $\mathcal{O}(1/N)$ in the main convergence result. Let us mention that the Frank-Wolfe algorithm (also called conditional gradient method) was already applied to potential MFGs, see for example [16, 22, 23]. It can be seen as a generalization of the fictitious play, investigated in particular in [10, 20].

Theoretical results. At a theoretical level, we first establish a first-order necessary and sufficient optimality condition for (P_m) and an existence result, both relying on rather standard arguments. Then we perform a stability analysis of the problem with respect to the marginal m . It is of course motivated by the need to understand the effect of the discretization of m in the numerical approach

described above. We provide a constructive method, which we call *bridging method*. It allows to construct an approximate solution to (P_m) , given an approximate solution to the problem with a different (but close) marginal. This allows to prove that the value of problem (P_m) is Lipschitz continuous with respect to m , for the Kantorovich-Rubinstein distance. Finally, we introduce a dual problem to (P_m) and prove that strong duality holds. We prove that the unique solution to the dual problem has a Hölder dependence with respect to m .

Organization. In Section 2, we present some notations and results in measure theory and set-valued functions, as well as the rigorous description of the data of problem (P_m) . Section 3 is dedicated to the primal problem: We provide a first-order optimality condition and an existence result. We perform in Section 4 a stability analysis for the primal problem, based on our bridging method. In Section 5, we formulate the dual problem of (P_m) , we prove strong duality, and we prove the stability of the dual solution. We provide our numerical method in Section 6. We perform in Section 7 some numerical simulations for a Lagrangian MFG model taken from [18] and for a congestion problem.

2. PRELIMINARIES

2.1. Results in measure theory. A metric space is called a *Polish space* if it is complete and separable. Let X be a Polish space equipped with a metric d_X , and let \mathcal{X} be a σ -algebra on X . The Borel σ -algebra on X is denoted by \mathcal{B}^X . Given any measure m on \mathcal{X} , we refer to the triplet (X, \mathcal{X}, m) as a measure space. Measure spaces are said to be complete if for any $A \in \mathcal{X}$ with $m(A) = 0$ and for any subset B of A , we have $B \in \mathcal{X}$. We define

$$\mathcal{P}(X) := \{m \text{ is a positive Borel measure on } X, \text{ and } m(X) = 1\};$$

$$\mathcal{P}^1(X) := \left\{ m \in \mathcal{P}(X) \mid \exists x_0 \in X \text{ such that } \int_X d_X(x, x_0) dm < +\infty \right\}.$$

Let δ_x denote the Dirac measure at point x . We denote by $\mathcal{P}_\delta(\Omega)$ the set of finitely supported probability measures, defined by

$$\mathcal{P}_\delta(X) := \left\{ \sum_{k=1}^K \omega_k \delta_{x_k} \mid K \in \mathbb{N}, (\omega_k)_{k=1}^K \in (\mathbb{R}_+)^K, (x_k)_{k=1}^K \in X^K, \sum_{k=1}^K \omega_k = 1 \right\}.$$

In particular, we call $m \in \mathcal{P}_\delta(X)$ an *empirical distribution* if $\lambda_k = 1/K$ for $k = 1, \dots, K$.

The set $\mathcal{P}(X)$ is endowed with the narrow topology. We say that a sequence $(m_n)_{n \geq 1}$ in $\mathcal{P}(Z)$ narrowly converges to some $m \in \mathcal{P}(X)$ if for any bounded and continuous function $F: X \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} \int_X F dm_n = \int_X F dm.$$

The space $\mathcal{P}^1(X)$ is endowed with the *Kantorovich–Rubinstein Distance*,

$$d_1(m_0, m_1) := \sup_{F \in \text{Lip}_1(X)} \int_\Omega F d(m_0 - m_1),$$

where $\text{Lip}_1(X)$ is the set of all 1-Lipschitz continuous functions on X . For any $m \in \mathcal{P}(X)$, the support of m is defined by

$$(2.1) \quad \text{supp}(m) := \{x \in X \mid m(V) > 0 \text{ for all open set } V \text{ such that } x \in V\}.$$

Lemma 2.1. *Let $m \in \mathcal{P}(X)$. Let $F: X \rightarrow \mathbb{R}_+$ be a Borel measurable function. Assume that*

$$\int_X F dm = 0.$$

Then $F = 0$, m -a.e. Moreover, if $F^{-1}(\{0\})$ is closed, then $\text{supp}(m) \subseteq F^{-1}(\{0\})$.

Proof. The fact that $F = 0$, m -a.e., is from [28, Thm. 1.39(a)]. Now, let $F^{-1}(\{0\})$ be closed. Suppose that there exists $x \in \text{supp}(m)$ such that $x \notin F^{-1}(\{0\})$. Since $F^{-1}(\{0\})$ is closed, there exists an open neighborhood V of x such that $F(x) > 0$, for all $x \in V$. By the definition of the support of a probability measure, we have $m(V) > 0$. Therefore, $\int_X F dm \geq \int_V F dm > 0$, contradiction. ■

2.2. Results about set-valued functions. In this subsection, we consider a metric space X equipped with a metric d_X , a σ -algebra \mathcal{X} on X , and a measure m on \mathcal{X} . Additionally, we fix a Polish space Y with a metric d_Y , and we denote the Borel σ -algebra on Y by \mathcal{B}^Y . We call F a set-valued function from X to Y if $F(x) \subseteq Y$ for all $x \in X$, denoted by $X \rightsquigarrow Y$ for short. The graph of F is defined by

$$\text{Graph}(F) := \{(x, y) \in X \times Y \mid y \in F(x)\}.$$

We say that F has closed (non-empty) images, if for any $x \in X$, $F(x)$ is closed (non-empty) in Y .

Let us give some definitions concerning regularity properties of set-valued functions, which are from [2, Def. 1.4.1, Def. 1.4.2, Def. 1.4.5, and Def. 8.1.1].

Definition 2.2. Let $F: X \rightsquigarrow Y$ be a set-valued function with non-empty images.

- (1) (Lower semi-continuity). The set-valued function F is *lower semi-continuous* at point $x \in X$ if for any $y \in F(x)$ and any sequence $(x_n \in X)_{n \geq 1}$ converging to x , there exists $y_n \in F(x_n)$ converging to y . The set-valued function F is said to be lower semi-continuous if it is lower semi-continuous at each point $x \in X$.
- (2) (Upper semi-continuity). The set-valued function F is *upper semi-continuous* at point $x \in X$ if for any neighborhood \mathcal{U} of $F(x)$, there exists $\eta > 0$ such that for any $x' \in B_X(x, \eta)$, we have

$$F(x') \subseteq \mathcal{U}.$$

The set-valued function F is said to be upper semi-continuous if it is upper semi-continuous at each point $x \in X$.

- (3) (Lipschitz continuity). When X and Y are normed vector spaces, we say that F is L -Lipschitz continuous on X , for some $L > 0$, if for any $x_1, x_2 \in X$,

$$F(x_1) \subseteq F(x_2) + B_Y(0, Ld_X(x_1, x_2)).$$

Here $B_Y(0, r)$ denotes the closed ball in Y centered at 0 with radius $r > 0$.

- (4) (Measurability). The set-valued function F is *measurable* if the inverse image of any open subset \mathcal{O} of Y is measurable, i.e.,

$$F^{-1}(\mathcal{O}) := \{x \in X \mid F(x) \cap \mathcal{O} \neq \emptyset\} \in \mathcal{X}.$$

An important property of measurable set-valued functions is the existence of measurable selections.

Theorem 2.3 (Measurable selection). *Let $F: X \rightsquigarrow Y$ be a measurable set-valued function with non-empty images. Then F has a measurable selection f , i.e., $f: X \rightarrow Y$ is $(\mathcal{X}, \mathcal{B}^Y)$ -measurable and $f(x) \in F(x)$ for any $x \in X$.*

Proof. See [2, Thm. 8.1.3]. ■

The following two lemmas will allow us to prove the measurability of some set-valued functions.

Lemma 2.4. *If $F: X \rightsquigarrow Y$ is a set-valued function such that $F^{-1}(\mathcal{C}) \in \mathcal{X}$ for any closed subset \mathcal{C} of Y , then F is measurable.*

Proof. See [11, Prop. III.11]. ■

Lemma 2.5. *Let (X, \mathcal{X}, m) be a complete measure space, with m a positive measure such that $m(X) = 1$. Then any set-valued mapping $F: X \rightsquigarrow Y$ is measurable if and only if $\text{Graph}(F)$ belongs to $\mathcal{X} \otimes \mathcal{B}^Y$.*

Proof. See [2, Thm. 8.1.4]. ■

2.3. Data setting and technical lemmas. Recall the MFO problem (P_m) . We consider the following setting:

- Two Polish spaces and their Borel σ -algebras: (X, \mathcal{B}^X) and (Y, \mathcal{B}^Y) .
- A probability distribution on X : $m \in \mathcal{P}(X)$.
- A set-valued function $F: X \rightsquigarrow Y$ with a closed graph and non-empty images. Let

$$Z := \text{Graph}(F), \quad Z_x := F(x), \quad \forall x \in X.$$

- The admissible set of probability measures:

$$\mathcal{P}_m(Z) := \{\mu \in \mathcal{P}(Z) \mid \pi_1 \# \mu = m\},$$

where $\pi_1: Z \rightarrow X, (x, y) \mapsto x$.

- A separable Hilbert space: \mathcal{H} .
- Two Borel measurable functions: $g: Z \rightarrow \mathcal{H}$ and $f: \mathcal{H} \rightarrow \mathbb{R}$.

The integral $\int_Z g d\mu$ in (P_m) should be interpreted in the Bochner integration sense. We refer to [14, Appx. E] for Bochner integrable functions.

Lemma 2.6. *If there exists a constant $M > 0$ such that $\|g(z)\| \leq M$ for any $z \in Z$, then the function g is Bochner integrable with respect to any $\mu \in \mathcal{P}(Z)$, i.e., $\int_Z g d\mu$ exists. Moreover, for any $\lambda \in \mathcal{H}$, we have*

$$\left\langle \lambda, \int_Z g d\mu \right\rangle = \int_Z \langle \lambda, g \rangle d\mu.$$

As a consequence, for any $\mu_1, \mu_2 \in \mathcal{P}(Z)$, we have

$$\left\langle \int_Z g d\mu_1, \int_Z g d\mu_2 \right\rangle = \int_Z \int_Z \langle g(x), g(y) \rangle d\mu_1(x) d\mu_2(y).$$

Proof. As \mathcal{H} is separable, the function g is strongly measurable. Moreover, as the constant function M is Bochner integrable with respect to any $\mu \in \mathcal{P}(Z)$, and $|g(z)| \leq M$ for any $z \in Z$, it follows from [14, Prop. E.2, Thm. E.6] that g is Bochner integrable with respect to any $\mu \in \mathcal{P}(Z)$. Therefore, we can apply [14, Prop. E.11] to obtain the first equality of this lemma. The second equality is obtained by applying twice the first one. ■

Theorem 2.7 (Disintegration theorem). *For any $\mu \in \mathcal{P}_m(Z)$, there exists a family of probability measures $\{\mu_x \in \mathcal{P}(Y)\}_x$ such that for any Borel measurable function $f: Z \rightarrow \mathbb{R}_+$, we have*

$$\int_Z f d\mu = \int_X \int_{Z_x} f(x, y) d\mu_x(y) dm(x).$$

Moreover, for a.e. $x \in X$, μ_x is uniquely determined.

Proof. See [1, Thm. 5.3.1]. ■

Remark 2.8. It is not difficult to generalize Theorems 2.7 to functions f bounded from below, by adding to f a sufficient large positive constant.

3. OPTIMALITY CONDITION

3.1. Assumptions and constants. To simplify the presentation of the assumptions and the results of the article, we introduce the following (set-valued) functions, parameterized by $\lambda \in \mathcal{H}$:

- $g_\lambda: Z \rightarrow \mathbb{R}$ and $u_\lambda: X \rightarrow \mathbb{R}$,

$$g_\lambda(x, y) = \langle \lambda, g(x, y) \rangle, \quad u_\lambda(x) = \inf_{y \in Z_x} g_\lambda(x, y);$$

- $G_\lambda: X \rightsquigarrow \mathbb{R}$ and $\mathbf{BR}_\lambda: X \rightsquigarrow Y$,

$$G_\lambda(x) = \{g_\lambda(x, y) \mid y \in Z_x\}, \quad \mathbf{BR}_\lambda(x) = \operatorname{argmin}_{y \in Z_x} g_\lambda(x, y).$$

Assumption A. The following holds:

- (1) The function g is bounded. The function f is convex and differentiable, and ∇f is Lipschitz continuous with modulus L .
- (2) Let $\mathcal{H}_f := \nabla f(\mathcal{H})$. Fixing any $\lambda \in \mathcal{H}_f$, we have:
 - the function g_λ is lower semi-continuous;
 - the set-valued function G_λ is lower semi-continuous;
 - the set-valued function \mathbf{BR}_λ has non-empty images.

Three useful constants below are defined, following Assumption A:

$$M := \sup_{z \in Z} \|g(z)\|, \quad D := \sup_{z_1, z_2 \in Z} \|g(z_1) - g(z_2)\|^2, \quad C := \sup_{\mu \in \mathcal{P}(Z)} \left\| \nabla f \left(\int_Z g d\mu \right) \right\|.$$

We present here a lemma following Assumption A. A similar result for the Lagrangian MFG is presented in [9, Lem. 3.4].

Lemma 3.1. *Under Assumption A, for any $\lambda \in \mathcal{H}_f$, the set-valued function \mathbf{BR}_λ has a closed graph.*

Proof. Let $x_k \in X$ converge to some $\bar{x} \in X$, and let $y_k \in \mathbf{BR}_\lambda(x_k)$ converge to some $\bar{y} \in Y$. We have to prove that $\bar{y} \in \mathbf{BR}_\lambda(\bar{x})$. First, we have $\bar{y} \in Z_{\bar{x}}$, since Z is closed. Fix any $y \in Z_{\bar{x}}$. Since G_λ is lower semi-continuous, there exists a sequence $(\hat{y}_k)_{k \in \mathbb{N}}$ in Z_{x_k} such that

$$g_\lambda(\bar{x}, y) = \lim_{k \rightarrow \infty} g_\lambda(x_k, \hat{y}_k).$$

By the lower semi-continuity of g_λ , we have

$$g_\lambda(\bar{x}, \bar{y}) \leq \liminf_{k \rightarrow \infty} g_\lambda(x_k, y_k).$$

Since $y_k \in \mathbf{BR}_\lambda(x_k)$ and $\hat{y}_k \in Z_{x_k}$, we have $g_\lambda(x_k, y_k) \leq g_\lambda(x_k, \hat{y}_k)$ for any k . Passing to the limit in this inequality (using the above inequalities), we deduce that $g_\lambda(\bar{x}, \bar{y}) \leq g_\lambda(\bar{x}, y)$. Thus, \mathbf{BR}_λ has a closed graph. ■

In Section 5, we will consider the dual problem of (\mathbf{P}_m) . For the analysis of Section 5, Assumption A needs to be strengthened as follows:

Assumption A*. Assumption A(1) holds true and Assumption A(2) holds true for all $\lambda \in \operatorname{dom}(f^*)$, where f^* is the Fenchel conjugate of f .

Remark 3.2. Assumption A* is indeed stronger than Assumption A since $\mathcal{H}_f = \nabla f(\mathcal{H}) \subseteq \operatorname{dom}(f^*)$. This inclusion is deduced from Fenchel's relation: $y = \nabla f(x) \Leftrightarrow f^*(y) = \langle x, y \rangle - f(x)$.

3.2. First-order-optimality condition. The following lemma plays a key role in proving the first-order optimality condition for (\mathbf{P}_m) .

Lemma 3.3. *Let Assumption A hold true. For any $\lambda \in \mathcal{H}_f$, we have*

$$\inf_{\mu \in \mathcal{P}_m(Z)} \int_Z g_\lambda d\mu = \int_X u_\lambda dm.$$

Here we present a proof of Lemma 3.3 for the case where m has finite support, that is, $m \in \mathcal{P}_\delta(X)$. This particular case provides us with insight into the general proof, and proves beneficial for resolving the discretized problem introduced in Section 6.

Proof of Lemma 3.3 when $m \in \mathcal{P}_\delta(X)$. Fix any $\mu \in \mathcal{P}_m(Z)$. Since g is bounded over Z , the function g_λ is bounded from below. By Lemma 2.7 and Remark 2.8, we have

$$\int_Z g_\lambda d\mu = \int_X \int_{Z_x} g_\lambda(x, y) d\mu_x(y) dm(x) \geq \int_X u_\lambda dm,$$

where the second inequality follows from the definition of u_λ .

Let us prove the converse inequality. Let us fix $m \in \mathcal{P}_\delta(X)$. Let $K \in \mathbb{N}$, let $(x_k)_{k=1, \dots, K} \in X^K$ and let $(\omega_k)_{k=1, \dots, K} \in \mathbb{R}_+^K$ be such that $\sum_{k=1}^K \omega_k = 1$ and $m = \sum_{k=1}^K \omega_k \delta_{x_k}$. For any $k = 1, \dots, K$, let $y_k \in \mathbf{BR}_\lambda(x_k)$. Let us define $\tilde{\mu} = \sum_{k=1}^K \omega_k \delta_{(x_k, y_k)}$. Clearly $\tilde{\mu} \in \mathcal{P}_m(Z)$. Moreover,

$$\int_Z g_\lambda d\tilde{\mu} = \sum_{k=1}^K \omega_k g_\lambda(x_k, y_k) = \sum_{k=1}^K \omega_k u_\lambda(x_k) = \int_X u_\lambda dm.$$

The conclusion follows, moreover, $\tilde{\mu}$ minimizes $\int_Z g_\lambda d\mu$ over $\mathcal{P}_m(Z)$. ■

In the general case, one has to find a measurable selection of \mathbf{BR}_λ , which requires us to prove the measurability of \mathbf{BR}_λ , which cannot be done in a direct fashion. The complete proof is given in Appendix A.

Theorem 3.4 (First-order optimality condition). *Let Assumption A(1) hold true. Let $\bar{\mu} \in \mathcal{P}_m(Z)$ and $\bar{\lambda} = \nabla f(\int_Z g d\bar{\mu})$. Consider the following three assertions:*

- (1) *The measure $\bar{\mu}$ is a solution of problem (\mathbf{P}_m) ;*
- (2) *$\int_Z g_{\bar{\lambda}} d\bar{\mu} = \inf_{\mu \in \mathcal{P}_m(Z)} \int_Z g_{\bar{\lambda}} d\mu$;*
- (3) *$\text{supp}(\bar{\mu}_x) \subseteq \mathbf{BR}_{\bar{\lambda}}(x)$, m -a.e., where $\bar{\mu}_x$ is defined by the disintegration theorem.*

Then, assertions (1) and (2) are equivalent. Moreover, under Assumption A(2), assertions (1), (2), and (3) are equivalent.

Proof. Step 1. (Equivalence between (1) and (2)). We first prove that (1) \Rightarrow (2). Suppose that $\bar{\mu}$ is a solution of problem (\mathbf{P}_m) . Take an arbitrary $\mu \in \mathcal{P}_m(Z)$. Then, for any $\alpha \in [0, 1]$, we have

$$\begin{aligned} f\left(\int_Z g d\bar{\mu}\right) &\leq f\left(\int_Z g d(\bar{\mu} + \alpha(\mu - \bar{\mu}))\right) \\ &\leq f\left(\int_Z g d\bar{\mu}\right) + \alpha \left\langle \bar{\lambda}, \int_Z g d(\mu - \bar{\mu}) \right\rangle + \frac{\alpha^2 LD}{2}, \end{aligned}$$

where the second inequality follows from the Lipschitz-continuity of ∇f and the definition of D . Therefore

$$0 \leq \left\langle \bar{\lambda}, \int_Z g d(\mu - \bar{\mu}) \right\rangle + \frac{\alpha LD}{2}$$

Let α go to 0. We obtain that

$$(3.1) \quad \left\langle \bar{\lambda}, \int_Z g d\bar{\mu} \right\rangle = \inf_{\mu \in \mathcal{P}_m(Z)} \left\langle \bar{\lambda}, \int_Z g d\mu \right\rangle.$$

This implies (2) by the definition of $g_{\bar{\lambda}}$.

We now prove (2) \Rightarrow (1). Let (2) hold true. We obtain (3.1) by the definition of $g_{\bar{\lambda}}$. The convexity of f implies that for any $\mu \in \mathcal{P}_m(Z)$,

$$f\left(\int_Z g d\mu\right) \geq f\left(\int_Z g d\bar{\mu}\right) + \left\langle \bar{\lambda}, \int_Z g d\mu - \int_Z g d\bar{\mu} \right\rangle \geq f\left(\int_Z g d\bar{\mu}\right).$$

Therefore, $\bar{\mu}$ is a solution of problem (\mathbf{P}_m) .

Step 2. (Equivalence between (2) and (3)). By Theorem 2.7, we have

$$\int_Z g_{\bar{\lambda}} d\bar{\mu} = \int_X \int_{Z_x} g_{\bar{\lambda}}(x, y) d\bar{\mu}_x(y) dm(x).$$

By Lemma 3.3, we have

$$\inf_{\mu \in \mathcal{P}_m(Z)} \int_Z g_{\bar{\lambda}} d\mu = \int_X u_{\bar{\lambda}} dm.$$

Therefore, assertion (2) is equivalent to

$$(3.2) \quad \int_X \int_{Z_x} g_{\bar{\lambda}}(x, y) d\bar{\mu}_x(y) dm(x) = \int_X u_{\bar{\lambda}} dm.$$

Let (3) hold true. It follows that $\int_{Z_x} g_{\bar{\lambda}}(x, y) d\bar{\mu}_x(y) = u_{\bar{\lambda}}(x)$, m -a.e., which implies (3.2).

Let (2) hold true. We obtain (3.2). The function $x \mapsto \left(\int_{Z_x} g_{\bar{\lambda}}(x, y) d\bar{\mu}_x(y)\right) - u_{\bar{\lambda}}(x)$ is nonnegative, for a.e. $x \in X$, by the definition of $u_{\bar{\lambda}}$. By (3.2), its integral is null, thus, as a consequence of Lemma 2.1, we have

$$(3.3) \quad \int_{Z_x} g_{\bar{\lambda}}(x, y) d\bar{\mu}_x(y) = u_{\bar{\lambda}}(x) = \inf_{y \in Z_x} g_{\bar{\lambda}}(x, y), \quad m\text{-a.e.}$$

Fix $x \in X$ such that equality holds in (3.3). Consider the map $y \in Z_x \mapsto g_{\bar{\lambda}}(x, y) - u_{\bar{\lambda}}(x)$. It is nonnegative, with a null integral, and $\mathbf{BR}_{\bar{\lambda}}(x)$ is non-empty and closed. Then assertion (3) follows with Lemma 2.1. \blacksquare

Corollary 3.5. *Under Assumption A, $\bar{\mu}$ is a solution of (\mathbf{P}_m) if and only if the following equilibrium equation is satisfied:*

$$(3.4) \quad \begin{cases} \bar{\lambda} = \nabla f\left(\int_Z g d\bar{\mu}\right), \\ \text{supp}(\bar{\mu}_x) \subseteq \mathbf{BR}_{\bar{\lambda}}(x), \quad m\text{-a.e.} \end{cases}$$

Proof. This is a consequence of Theorem 3.4. \blacksquare

The conditions in (3.4) can be interpreted as the conditions for a Nash equilibrium in a non-atomic game, in which the agents through the variable $\bar{\gamma}$. The relation $\bar{\lambda} = \nabla f\left(\int_Z g d\bar{\mu}\right)$ shows how $\bar{\lambda}$ results from the collective behavior of the agents, while the relation $\text{supp}(\bar{\mu}_x) \subseteq \mathbf{BR}_{\bar{\lambda}}(x)$ shows that the agents behave optimally, for some criterion that depends on $\bar{\lambda}$. We will discuss some more concrete examples in Section 7.

3.3. Existence of a solution under tightness assumptions. We denote by $\mathbf{val}(\mathbf{P}_m)$ the value of problem (\mathbf{P}_m) . We can easily deduce from Assumption A that $\mathbf{val}(\mathbf{P}_m) > -\infty$. The following proposition demonstrates the existence of a solution to problem (\mathbf{P}_m) under some additional assumptions.

Proposition 3.6 (Existence). *Let Assumption A hold true. Let $(\mu_n)_{n \geq 1}$ be a minimizing sequence for problem (\mathbf{P}_m) . Suppose that $\{\mu_n\}_{n \geq 1}$ is tight in $\mathcal{P}(Z)$, i.e. for any $\epsilon > 0$, there exists a compact subset K_ϵ of Z such that*

$$\mu_n(K_\epsilon) \geq 1 - \epsilon, \quad \forall n \geq 1.$$

Then every accumulation point of $\{\mu_n\}_{n \geq 1}$ for the narrow topology (there exists at least one) is a solution of (\mathbf{P}_m) .

Proof. By Prokhorov's theorem [33, p. 43], the set $\{\mu_n\}_{n \geq 1}$ is relatively compact with respect to the narrow topology. Without loss of generality, suppose that μ_n narrowly converges to some $\bar{\mu} \in \mathcal{P}(Z)$. The set $\mathcal{P}_m(Z)$ is closed with respect to narrow topology by [29, Prop. 2.4]. This implies that $\bar{\mu} \in \mathcal{P}_m(Z)$. Let $\bar{\lambda} = \nabla f(\int_Z g d\bar{\mu})$. Since f is convex, we have

$$(3.5) \quad f\left(\int_Z g d\mu_n\right) \geq f\left(\int_Z g d\bar{\mu}\right) + \int_Z g_{\bar{\lambda}} d(\mu_n - \bar{\mu}).$$

Since $g_{\bar{\lambda}}: Z \rightarrow \mathbb{R}$ is lower semi-continuous and bounded from below by Assumption A, we deduce the following inequality from [33, Lem. 4.3]:

$$\liminf_{n \rightarrow +\infty} \int_Z g_{\bar{\lambda}} d(\mu_n - \bar{\mu}) \geq 0.$$

In inequality (3.5), letting n go to infinity, by the definition of μ_n , we have

$$\mathbf{val}(\mathbf{P}_m) = \liminf_{n \rightarrow +\infty} f\left(\int_Z g d\mu_n\right) \geq f\left(\int_Z g d\bar{\mu}\right) \geq \mathbf{val}(\mathbf{P}_m).$$

Therefore, $\bar{\mu}$ is a solution of problem (\mathbf{P}_m) . ■

4. STABILITY ANALYSIS AND BRIDGING METHOD

In this section, we study the stability of the primal problem (\mathbf{P}_m) with respect to its parameter m . We need the following assumptions (recall the data setting introduced in Sec. 2.3).

Assumption B. The following holds:

- (1) The space X is a closed subset of a separable Banach space;
- (2) The function $g: Z \rightarrow \mathcal{H}$ is continuous;
- (3) The set Z_x is compact for any $x \in X$ and the set-valued function $F: X \rightsquigarrow Y$ is upper semi-continuous;
- (4) There exists $L_g \geq 0$ such that the set-valued function

$$(4.1) \quad \mathcal{Z}: X \rightsquigarrow \mathcal{H}, \quad x \mapsto \{g(x, y) \mid y \in Z_x\}$$

is L_g -Lipschitz on X .

Let m_0 and m_1 lie in $\mathcal{P}(X)$. We consider the following two instances of (\mathbf{P}_m) with $m = m_0$ and $m = m_1$ respectively:

$$(\mathbf{P}_{m_0}) \quad \inf_{\mu \in \mathcal{P}_{m_0}(Z)} f\left(\int_Z g d\mu\right);$$

$$(\mathbf{P}_{m_1}) \quad \inf_{\mu \in \mathcal{P}_{m_1}(Z)} f\left(\int_Z g d\mu\right).$$

Suppose that we have an (approximate) solution of problem (\mathbf{P}_{m_0}) , denoted by $\bar{\mu}_0$. Our goal is to propose a feasible approach for recovering an approximate solution of problem (\mathbf{P}_{m_1}) from $\bar{\mu}_0$ and to study the performance of this approximation. We call it the bridging method. It relies on $\bar{\mu}_0$ and the solution of the optimal transport problem (OT1) stated later. To introduce (OT1), we need to define some projection operators

Recall that $\pi_1: Z \rightarrow X, (x, y) \mapsto x$, and $\pi_2: Z \rightarrow X, (x, y) \mapsto y$. The other projection operators used in this subsection are defined as:

$$\begin{aligned} \tilde{\pi}_1: X \times X &\rightarrow X, (x, x') \mapsto x, & \tilde{\pi}_2: X \times X &\rightarrow X, (x, x') \mapsto x', & \pi_3: Z \times X &\rightarrow X, (x, y, x') \mapsto x', \\ \pi_{12}: Z \times X &\rightarrow Z, (x, y, x') \mapsto (x, y), & \pi_{13}: Z \times X &\rightarrow Z, (x, y, x') \mapsto (x, x'). \end{aligned}$$

It directly follows from the above definitions that

$$\tilde{\pi}_1 \circ \pi_{13} = \pi_1 \circ \pi_{12} \quad \text{and} \quad \tilde{\pi}_2 \circ \pi_{13} = \pi_3.$$

Now we consider the following optimal transport problem:

$$(OT1) \quad \inf_{\rho \in \Pi(m_0, m_1)} \int_{X \times X} d_X(x, x') d\rho(x, x'),$$

where $\Pi(m_0, m_1) = \{\rho \in \mathcal{P}(X \times X) \mid \tilde{\pi}_1 \# \rho = m_0, \tilde{\pi}_2 \# \rho = m_1\}$. It follows from [33, Rem. 6.5] that if m_0 and m_1 lie in $\mathcal{P}^1(X)$, then $d_1(m_0, m_1) = \mathbf{val}(OT1)$.

The following particular example will provide an intuitive understanding of our bridging method.

A particular case. Let us assume that the distributions m_0 , m_1 , and $\bar{\mu}_0$ are empirical distributions with supports of size N , i.e., there exists $(x_i)_{i=1}^N, (\tilde{x}_i)_{i=1}^N \in X^N$ and $(y_i)_{i=1}^N \in \prod_{i=1}^N Z_{x_i}$ such that

$$(4.2) \quad m_0 = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \quad m_1 = \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{x}_i}, \quad \bar{\mu}_0 = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)}.$$

Lemma 4.1. *Let m_0 and m_1 be defined by (4.2). Then problem (OT1) has a solution*

$$(4.3) \quad \rho = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, x'_i)},$$

where $\{x'_1, \dots, x'_N\}$ is a permutation of $\{\tilde{x}_1, \dots, \tilde{x}_N\}$.

Proof. This is a consequence of [26, Prop. 2.1]. ■

Let ρ be given by Lemma 4.1. By Assumption B(4), for any i , there exists $y'_i \in Z_{x'_i}$ such that

$$(4.4) \quad \|g(x'_i, y'_i) - g(x_i, y_i)\| \leq L_g d_X(x_i, x'_i).$$

In our bridging method, each x_i is transformed to x'_i while simultaneously moving y_i to the point $y'_i \in Z_{x'_i}$ for $i = 1, \dots, N$. This can be expressed as follows:

$$(4.5) \quad \bar{\mu}_0 = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)} \longrightarrow \mu_1 = \frac{1}{N} \sum_{i=1}^N \delta_{(x'_i, y'_i)}.$$

To provide a clearer formula of the construction of μ_1 , we introduce the empirical distribution $\nu_N \in \mathcal{P}(Z \times X)$ and the mapping $s_N: \{(x_i, y_i, x'_i)\}_{i=1}^N \rightarrow \{(x_i, y'_i)\}_{i=1}^N$, defined as:

$$\begin{aligned} \nu_N &= \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i, x'_i)}, \\ s_N(x_i, y_i, x'_i) &= (x'_i, y'_i), \quad \forall i = 1, 2, \dots, N. \end{aligned}$$

It can be observed that $\pi_{12} \# \nu_N = \bar{\mu}_0$ and $\pi_3 \# \nu_N = m_1$. Furthermore, we will demonstrate later in Lemma 4.2 that ν_N is a solution of another optimal transport problem (OT2). Then the approximate solution μ_1 of problem (P $_{m_1}$) can be written as:

$$\mu_1 = s_N \# \nu_N.$$

The distribution μ_1 belongs to $\mathcal{P}_{m_1}(Z)$, and furthermore,

$$\begin{aligned} \left\| \int_Z g d\bar{\mu}_0 - \int_Z g d\mu_1 \right\| &= \left\| \frac{1}{N} \sum_{i=1}^N (g(x_i, y_i) - g(x'_i, y'_i)) \right\| \\ &\leq \frac{L_g}{N} \sum_{i=1}^N \|g(x_i, y_i) - g(x'_i, y'_i)\| \\ &\leq \frac{L_g}{N} \sum_{i=1}^N d_X(x_i, x'_i) = L_g d_1(m_0, m_1), \end{aligned}$$

where the second line follows from the triangle inequality and the third line follows from (4.4) and Lemma 4.1. The above inequality demonstrates that the distance between the aggregates associated with $\bar{\mu}_0$ and μ_1 is controlled by the d_1 -distance of m_0 and m_1 .

The general case. To investigate the stability and present the bridging method in the general case, we draw inspiration from the constructions of ν_N and s_N in the previous particular case and introduce the following:

- the auxiliary optimal transport problem:

$$(OT2) \quad \inf_{\nu \in \Pi(\bar{\mu}_0, m_1)} \int_{Z \times X} d_X(x, x') d\nu(x, y, x'),$$

where $\Pi(\bar{\mu}_0, m_1) := \{\nu \in \mathcal{P}(Z \times X) \mid \pi_{12}\#\nu = \bar{\mu}_0, \pi_3\#\nu = m_1\}$;

- the set-valued function $S: Z \times X \rightsquigarrow Z$,

$$(4.6) \quad S(x, y, x') = \{(x', y') \in Z \mid \|g(x', y') - g(x, y)\| \leq L_g d_X(x, x')\}.$$

Note that problems (OT2) and (OT1) are similar in so far as the integrand of the cost function is the same, moreover, the second marginal of ν in (OT2) (resp. ρ in (OT1)) must be equal to m_1 . The following lemma shows the equivalence between problems (OT1) and (OT2). We will see that the solution of (OT2) will play the role of ν_N in the particular case mentioned earlier.

Lemma 4.2. *If m_0 and m_1 lie in $\mathcal{P}^1(X)$, then both problems (OT1) and (OT2) have solutions, moreover,*

$$\mathbf{val}(OT1) = \mathbf{val}(OT2) = d_1(m_0, m_1).$$

Proof. Since $m_0, m_1 \in \mathcal{P}^1(X)$, by [33, Rem. 6.5], we have $d_1(m_0, m_1) = \mathbf{val}(OT1)$. The existence of solutions of problems (OT2) and (OT1) is from [33, Thm. 4.1].

Let ν be a solution to (OT2) and let $\rho = \pi_{13}\#\nu$, which is clearly an element of $\mathcal{P}(X \times X)$. By the basic properties of push-forward measures, we have that $\tilde{\pi}_1\#\rho = \tilde{\pi}_1\#(\pi_{13}\#\nu) = (\tilde{\pi}_1 \circ \pi_{13})\#\nu$. Using the relation $\tilde{\pi}_1 \circ \pi_{13} = \pi_1 \circ \pi_{12}$, we obtain that $(\tilde{\pi}_1 \circ \pi_{13})\#\rho = (\pi_1 \circ \pi_{12})\#\rho = \pi_1\#(\pi_{12}\#\nu) = \pi_1\#\bar{\mu}_0 = m_0$. It follows that $\tilde{\pi}_1\#\rho = m_0$. By similar arguments, we deduce that $\tilde{\pi}_2\#\rho = m_1$ from the relation $\tilde{\pi}_2 \circ \pi_{13} = \pi_3$. Therefore, $\mu \in \Pi(m_0, m_1)$, moreover,

$$\int_{X \times X} d_X(x, x') d\pi_{13}\#\nu(x, x') = \int_{Z \times X} d_X(\pi_{13}(x, y, x')) d\nu(x, y, x') = \mathbf{val}(OT2).$$

It follows that $d_1(m_0, m_1) \leq \mathbf{val}(OT2)$.

On the other hand, let ρ be a solution of (OT1). Since $\bar{\mu}_0$ and ρ have the same marginal distribution m_0 with respect to their first variable, by the Gluing lemma [33, p. 11], there exists a probability measure $\nu \in \mathcal{P}(X \times Y \times X)$ such that

$$\pi_{12}\#\nu = \bar{\mu}_0, \quad \pi_{13}\#\nu = \rho.$$

Since $\bar{\mu}_0 \in \mathcal{P}(Z)$, we have $\nu \in \mathcal{P}(Z \times X)$. From the relation $\pi_3 = \tilde{\pi}_2 \circ \pi_{13}$, we deduce that $\pi_3 \# \nu = \tilde{\pi}_2 \# (\pi_{13} \# \nu) = \tilde{\pi}_2 \# \rho = m_1$. Thus, $\nu \in \Pi(\bar{\mu}_0, m_1)$, moreover,

$$\begin{aligned} \int_{Z \times X} d_X(x, x') d\nu(x, y, x') &= \int_{Z \times X} d_X(\pi_{13}(x, y, x')) d\nu(x, y, x') = \int_{X \times X} d_X(x, x') d\pi_{13} \# \nu(x, x') \\ &= \int_{X \times X} d_X(x, x') d\rho(x, x') = d_1(m_0, m_1). \end{aligned}$$

It follows that $d_1(m_0, m_1) \geq \mathbf{val}(\text{OT2})$. \blacksquare

The following two lemmas demonstrate that the set-valued function S has a measurable selection, which will be denoted by s . We will see that the function s will fulfill the role of s_N in the particular case discussed earlier.

Lemma 4.3. *Under Assumption B(3), let $(x_n)_{n \geq 1}$ be a sequence in X converging to some $x_0 \in X$. Then any sequence $(y_n \in Z_{x_n})_{n \geq 1}$ has a convergent sub-sequence with its limit in Z_{x_0} .*

Proof. Since $F: X \rightsquigarrow Y$ is upper semi-continuous, for any $k \geq 1$, there exists $\eta_k > 0$ such that for any $x \in B_X(x_0, \eta_k)$, we have $Z_x \subseteq B_Y(Z_{x_0}, 1/k) := \cup_{y \in Z_{x_0}} B_Y(y, 1/k)$. For $k = 1$, there exists $\varphi(1) \in \mathbb{N}_+$ such that $Z_{x_{\varphi(1)}} \subseteq B_Y(Z_{x_0}, 1)$, i.e., there exists $\bar{y}_1 \in Z_{x_0}$ such that

$$d_Y(y_{\varphi(1)}, \bar{y}_1) \leq 1.$$

Assume now we have $\varphi(k) \in \mathbb{N}_+$ and $\bar{y}_k \in Z_{x_0}$ for $k = 1, \dots, K$ such that

$$d_Y(y_{\varphi(k)}, \bar{y}_k) \leq \frac{1}{k}, \quad k = 1, \dots, K.$$

Since $x_n \rightarrow x_0$, we can find $\varphi(K+1) > \varphi(K)$ such that $x_{\varphi(K+1)} \in B_X(x_0, \eta_{K+1})$. As a consequence, there exists $\bar{y}_{K+1} \in Z_{x_0}$ such that

$$d_Y(y_{\varphi(K+1)}, \bar{y}_{K+1}) \leq \frac{1}{K+1}.$$

Since Z_{x_0} is compact, $(\bar{y}_k)_{k \geq 1}$ has a convergent sub-sequence $(\bar{y}_{\phi(k)})_{k \geq 1}$ with a limit $\bar{y} \in Z_{x_0}$. By the triangle inequality,

$$d_Y(y_{\phi(\varphi(k))}, \bar{y}) \leq d_Y(y_{\phi(\varphi(k))}, \bar{y}_{\phi(\varphi(k))}) + d_Y(\bar{y}_{\phi(\varphi(k))}, \bar{y}) \leq \frac{1}{\phi(\varphi(k))} + d_Y(\bar{y}_{\phi(\varphi(k))}, \bar{y}).$$

Since ϕ and φ are strictly increasing functions going to $+\infty$, we have $\lim_{k \rightarrow +\infty} d_Y(y_{\phi(\varphi(k))}, \bar{y}) = 0$. Therefore, $(y_{\phi(\varphi(k))})_{k \geq 1}$ is a convergent sub-sequence of $(y_n)_{n \geq 1}$ with its limit $\bar{y} \in Z_{x_0}$. \blacksquare

Lemma 4.4. *Under Assumption B, the set-valued function S has a Borel measurable selection function $s: Z \times X \rightarrow Z$. Furthermore, we have $\|g(s(x, y, x')) - g(x, y)\| \leq L_g d_X(x, x')$.*

Proof. We will apply Theorem 2.3 and Lemma 2.4 to prove the result. The images of S are non-empty since the set-valued mapping \mathcal{Z} (defined in (4.1)) is supposed to be L_g -Lipschitz. Let us first verify that S has non-empty closed images. Fix any $(x, y, x') \in Z \times X$, and assume that $(x', z_n) \in S(x, y, x')$ converges to some $(x', z) \in Z$. It suffices to prove that $(x', z) \in S(x, y, x')$, i.e., $\|g(x', z) - g(x, y)\| \leq L_g d_X(x, x')$. This is true since g is continuous and $z_n \rightarrow z$.

Then, let us show that $S^{-1}(\mathcal{C})$ is closed for any closed subset \mathcal{C} in Z . By (4.6), we have

$$\begin{aligned} S^{-1}(\mathcal{C}) &= \{(x, y, x') \in Z \times X \mid S(x, y, x') \cap \mathcal{C} \neq \emptyset\} \\ &= \left\{ (x, y, x') \in Z \times X \mid \exists y' \in Z_{x'} \text{ such that } \left\{ \begin{array}{l} (x', y') \in \mathcal{C}, \\ \|g(x', y') - g(x, y)\| \leq L_g d_X(x, x') \end{array} \right. \right\}. \end{aligned}$$

If $S^{-1}(\mathcal{C}) = \emptyset$, then the conclusion is obvious. Assume that $S^{-1}(\mathcal{C}) \neq \emptyset$ and let $(x_n, y_n, x'_n)_{n \geq 1} \in S^{-1}(\mathcal{C})$ be a convergent sequence with its limit point $(x_0, y_0, x'_0) \in Z \times X$. Then, it suffices to prove that $(x_0, y_0, x'_0) \in S^{-1}(\mathcal{C})$. Since $(x_n, y_n, x'_n) \in S^{-1}(\mathcal{C})$, there exists $y'_n \in Z_{x'_n}$, for any n , such that

$$(x'_n, y'_n) \in \mathcal{C}, \quad \|g(x'_n, y'_n) - g(x_n, y_n)\| \leq L_g d_X(x_n, x'_n).$$

By Lemma 4.3, the sequence $(y'_n)_{n \geq 1}$ has a convergent sub-sequence $(y'_{\varphi(n)})_{n \geq 1}$ with its limit $y'_0 \in Z_{x'_0}$. Hence, $\lim_{n \rightarrow \infty} (x'_{\varphi(n)}, y'_{\varphi(n)}) = (x'_0, y'_0)$. Since \mathcal{C} is closed, we have $(x'_0, y'_0) \in \mathcal{C}$. By the triangle inequality,

$$\begin{aligned} \|g(x'_0, y'_0) - g(x_0, y_0)\| &\leq \|g(x'_0, y'_0) - g(x'_{\varphi(n)}, y'_{\varphi(n)})\| + \|g(x'_{\varphi(n)}, y'_{\varphi(n)}) - g(x_{\varphi(n)}, y_{\varphi(n)})\| \\ &\quad + \|g(x_{\varphi(n)}, y_{\varphi(n)}) - g(x_0, y_0)\|. \end{aligned}$$

By the continuity of g , we have

$$\begin{aligned} \|g(x'_0, y'_0) - g(x_0, y_0)\| &\leq \limsup_{n \rightarrow \infty} \|g(x'_{\varphi(n)}, y'_{\varphi(n)}) - g(x_{\varphi(n)}, y_{\varphi(n)})\| \\ &\leq \limsup_{n \rightarrow \infty} L_g d_X(x_{\varphi(n)}, x'_{\varphi(n)}) = L_g d_X(x_0, x'_0). \end{aligned}$$

Therefore, $y'_0 \in Z_{x'_0}$, $(x'_0, y'_0) \in \mathcal{C}$ and $\|g(x'_0, y'_0) - g(x_0, y_0)\| \leq L_g d_X(x_0, x'_0)$. It follows that $(x_0, y_0, x'_0) \in S^{-1}(\mathcal{C})$, which implies that $S^{-1}(\mathcal{C})$ is closed, thus a Borel set.

Lemma 2.4 shows that the set-valued function S is Borel measurable, and Theorem 2.3 shows the existence of a Borel measurable selection s of S . Since $s(x, y, x') \in S(x, y, x')$, the inequality $\|g(s(x, y, x')) - g(x, y)\| \leq L_g d_X(x, x')$ is a direct consequence of (4.6). \blacksquare

Lemma 4.5. *Let Assumptions A-B hold true. Let $\nu \in \Pi(\bar{\mu}_0, m_1)$ and s be the Borel measurable selection of S obtained Lemma 4.4. Let $\mu_1 = s\#\nu$. Then $\mu_1 \in \mathcal{P}_{m_1}(Z)$ and*

$$f\left(\int_Z g d\mu_1\right) - f\left(\int_Z g d\bar{\mu}_0\right) \leq L_g(C + LM) \int_{Z \times X} d_X(x, x') d\nu(x, y, x').$$

Proof. Step 1. (Properties of μ_1). Since $\nu \in \mathcal{P}(Z \times S)$ and $s: Z \times S \rightarrow Z$ is a Borel measurable function, we have $\mu = s\#\nu \in \mathcal{P}(Z)$. Observing that $\pi_1 \circ s = \pi_3$, it follows that $\pi_1\#\mu = (\pi_1 \circ s)\#\nu = \pi_3\#\nu = m_1$. Thus, $\mu_1 \in \mathcal{P}_{m_1}(Z)$. For any $\lambda \in \mathcal{H}$, $g_\lambda: Z \rightarrow \mathbb{R}$ is bounded. Then,

$$(4.7) \quad \int_Z g_\lambda d\mu_1 = \int_{Z \times X} g_\lambda \circ s d\nu.$$

Step 2. (Quadratic upper bound). By the Lipschitz continuity of ∇f , we have

$$(4.8) \quad f\left(\int_Z g d\mu_1\right) - f\left(\int_Z g d\bar{\mu}_0\right) \leq \int_Z g_{\lambda_0} (d\mu_1 - d\bar{\mu}_0) + \frac{L}{2} \left\| \int_Z g d\mu_1 - \int_Z g d\bar{\mu}_0 \right\|^2,$$

where $\lambda_0 = \nabla f(\int_Z g d\bar{\mu}_0)$.

Step 3. (First-order estimate). Let us study the first-order term in (4.8). By (4.7), we have

$$\int_Z g_{\lambda_0} d\mu_1 = \int_{Z \times X} g_{\lambda_0} \circ s d\nu = \int_{Z \times X} \langle \lambda_0, g(s(x, y, x')) \rangle d\nu(x, y, x').$$

From the relation $\pi_{12}\#\nu = \bar{\mu}_0$, we deduce that

$$\int_Z g_{\lambda_0} d\bar{\mu}_0 = \int_{Z \times X} \langle \lambda_0, g(x, y) \rangle d\nu(x, y, x').$$

Using the previous two equalities, Lemma 4.4, and the Cauchy–Schwarz inequality, we obtain that

$$\int_Z g_{\lambda_0} (d\mu_1 - d\bar{\mu}_0) \leq L_g \|\lambda_0\| \int_{Z \times X} d_X(x, x') d\nu(x, y, x').$$

Step 4. (Second-order estimate). Let us study the second-order term in (4.8). Developing it and using Lemma 2.6, we obtain that

$$(4.9) \quad \left\| \int_Z g d\mu_1 - \int_Z g d\bar{\mu}_0 \right\|^2 = \gamma_1 + \gamma_2,$$

where

$$\begin{aligned} \gamma_1 &= \int_Z \int_Z \langle g(z_1), g(z_2) \rangle d(\mu_1 - \bar{\mu}_0)(z_1) d\mu_1(z_2); \\ \gamma_2 &= \int_Z \int_Z \langle g(z_1), g(z_2) \rangle d(\bar{\mu}_0 - \mu_1)(z_1) d\bar{\mu}_0(z_2). \end{aligned}$$

Fix any $z_2 \in Z$. Following the same argument as in step 3, we have

$$\left\| \int_Z \langle g(z_1), g(z_2) \rangle d(\mu_1 - \bar{\mu}_0)(z_1) \right\| \leq L_g \|g(z_2)\| \int_{Z \times X} d_X(x, x') d\nu(x, y, x').$$

It follows that

$$\gamma_1 + \gamma_2 \leq 2L_g M \int_{Z \times X} d_X(x, x') d\nu(x, y, x').$$

Step 5. As a consequence of Steps 2-4, we deduce that

$$f \left(\int_Z g d\mu_1 \right) - f \left(\int_Z g d\bar{\mu}_0 \right) \leq L_g (\|\lambda_0\| + LM) \int_{Z \times X} d_X(x, x') d\nu(x, y, x').$$

By the definition of the constant C , we have $C \geq \|\lambda_0\|$. The conclusion follows. \blacksquare

We can now state the bridging algorithm that enables us to obtain an approximate solution of (\mathbf{P}_{m_1}) , given an approximate solution $\bar{\mu}_0$ of (\mathbf{P}_{m_0}) .

Algorithm 1: Bridging method

Input: $m_0, m_1 \in \mathcal{P}^1(Z)$, and $\bar{\mu}_0 \in \mathcal{P}_{m_0}(Z)$.

Step 1. Find a solution ρ of the optimal transport problem (OT1).

Step 2. Find $\nu \in \mathcal{P}(Z \times X)$ such that $\pi_{12}\#\nu = \bar{\mu}_0$ and $\pi_{13}\#\nu = \rho$.

Step 3. Set $\mu_1 = s\#\nu \in \mathcal{P}_{m_1}(Z)$, where s is constructed in Lemma 4.4.

Output: μ_1 .

Remark 4.6. We have already discussed the case where m_0, m_1 , and $\bar{\mu}_0$ are empirical distributions. We discuss now the slightly more general case where only m_0 and $\bar{\mu}_0$ are empirical distributions:

$$m_0 = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \quad \bar{\mu}_0 = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)}.$$

This situation corresponds to the algorithm presented in Section 6. Since $\rho \in \mathcal{P}_{m_0}(X \times X)$, by Lemma 6.5, we have $\rho = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \rho_{x_i}$, where ρ_{x_i} is defined in Theorem 2.7. Then the probability distribution μ , obtained in general with the Gluing lemma, is given here in an explicit form:

$$\nu = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)} \otimes \rho_{x_i}.$$

Theorem 4.7. *Let Assumptions A-B hold true. Assume that $m_0, m_1 \in \mathcal{P}^1(Z)$ and that $\bar{\mu}_0$ is an ϵ_0 -minimizer of problem (\mathbf{P}_{m_0}) for some $\epsilon_0 \geq 0$. The following holds true.*

(1) $|\mathbf{val}(\mathbf{P}_{m_0}) - \mathbf{val}(\mathbf{P}_{m_1})| \leq L_g(C + LM)d_1(m_0, m_1)$;

(2) If $\mu_1 \in \mathcal{P}_{m_1}(Z)$ is the output of Algorithm 1, then μ_1 is an η -minimizer of problem (\mathbf{P}_{m_1}) , where

$$\eta = \epsilon_0 + 2L_g(C + LM)d_1(m_0, m_1).$$

Proof. We prove (1). Fix any $\epsilon > 0$. Let μ_0^ϵ be an ϵ -minimizer of problem (\mathbf{P}_{m_0}) . By Lemma 4.2, there exists $\nu^\epsilon \in \Pi(\mu_0^\epsilon, m_1)$ such that

$$\int_{Z \times X} d_X(x, x') d\nu^\epsilon(x, y, x') = d_1(m_0, m_1).$$

We deduce from Lemma 4.5 that there exists $\mu_1^\epsilon \in \mathcal{P}_{m_1}(Z)$ associated with ν^ϵ such that

$$f\left(\int_Z g\mu_1^\epsilon\right) - f\left(\int_Z g d\mu_0^\epsilon\right) \leq L_g(C + LM)d_1(m_0, m_1).$$

Since $\mu_1^\epsilon \in \mathcal{P}_{m_1}(Z)$, $\mathbf{val}(\mathbf{P}_{m_1}) \leq f\left(\int_Z g\mu_1^\epsilon\right)$. Combining this with the fact $\mathbf{val}(\mathbf{P}_{m_0}) \geq f\left(\int_Z g d\mu_0^\epsilon\right) - \epsilon$, we obtain that

$$\mathbf{val}(\mathbf{P}_{m_1}) - \mathbf{val}(\mathbf{P}_{m_0}) \leq f\left(\int_Z g\mu_1^\epsilon\right) - f\left(\int_Z g d\mu_0^\epsilon\right) + \epsilon.$$

Therefore, $\mathbf{val}(\mathbf{P}_{m_1}) - \mathbf{val}(\mathbf{P}_{m_0}) \leq L_g(C + LM)d_1(m_0, m_1)$ by the arbitrariness of ϵ . We conclude the first part of the proof by exchanging the positions of m_0 and m_1 .

Let μ_1 be the output of Algorithm 1. To prove (2), we do the following decomposition:

$$f\left(\int g d\mu_1\right) - \mathbf{val}(\mathbf{P}_{m_1}) = \gamma_1 + \gamma_2 + \gamma_3,$$

where

$$\gamma_1 = f\left(\int g d\mu_1\right) - f\left(\int g d\bar{\mu}_0\right), \quad \gamma_2 = f\left(\int g d\bar{\mu}_0\right) - \mathbf{val}(\mathbf{P}_{m_0}), \quad \gamma_3 = \mathbf{val}(\mathbf{P}_{m_0}) - \mathbf{val}(\mathbf{P}_{m_1}).$$

From the proof of Lemma 4.2, we know that ν (the result of step 2 in Algorithm 1) is a solution of (OT2). Then, Lemma 4.5 shows that $\gamma_1 \leq L_g(C + LM)d_1(m_0, m_1)$. Since $\bar{\mu}_0$ is an ϵ -minimizer, $\gamma_2 \leq \epsilon$. By point (1), $\gamma_3 \leq L_g(C + LM)d_1(m_0, m_1)$. The conclusion follows. \blacksquare

5. DUALITY ANALYSIS

5.1. The dual problem. This section is dedicated to the duality analysis of the primal problem (\mathbf{P}_m) . In the sequel of this section, let Assumptions \mathbf{A}^* and \mathbf{B} hold true. Consider the equivalent formulation of problem of (\mathbf{P}_m) ,

$$(\tilde{\mathbf{P}}_m) \quad \inf_{\mu \in \mathcal{P}_m(Z), \beta \in \mathcal{H}} f(\beta), \quad \text{s.t. } \beta = \int_Z g d\mu.$$

The Lagrangian $\mathcal{L}: \mathcal{H}^2 \times \mathcal{P}_m(Z) \rightarrow \mathbb{R}$ associated with $(\tilde{\mathbf{P}}_m)$ writes,

$$\mathcal{L}(\lambda, \beta, \mu) = f(\beta) + \left\langle \lambda, \int_Z g d\mu - \beta \right\rangle.$$

Then, the dual problem of $(\tilde{\mathbf{P}}_m)$ is,

$$(5.1) \quad \sup_{\lambda \in \mathcal{H}} \inf_{\beta \in \mathcal{H}, \mu \in \mathcal{P}_m(Z)} \mathcal{L}(\lambda, \beta, \mu) = \sup_{\lambda \in \mathcal{H}} \left(-f^*(\lambda) + \inf_{\mu \in \mathcal{P}_m(Z)} \int_Z \langle \lambda, g(z) \rangle d\mu(z) \right),$$

where f^* is the Fenchel conjugate of f . For any $\lambda \in \mathcal{H}$, since g is bounded over Z , the second term $\inf_{\mu \in \mathcal{P}_m(Z)} \int_Z \langle \lambda, g(z) \rangle d\mu(z)$ is finite. Therefore, it suffices to study (5.1) for $\lambda \in \text{dom}(f^*)$, i.e.,

$$\sup_{\lambda \in \text{dom}(f^*)} \left(-f^*(\lambda) + \inf_{\mu \in \mathcal{P}_m(Z)} \int_Z g_\lambda d\mu \right).$$

The result of Lemma 3.3 holds true for all $\lambda \in \text{dom}(f^*)$ under Assumption \mathbf{A}^* . Applying it to the previous problem, we obtain the following equivalent dual problem:

$$(D_m) \quad - \inf_{\lambda \in \text{dom}(f^*)} \mathcal{D}_m(\lambda) := f^*(\lambda) - \int_X u_\lambda dm.$$

Lemma 5.1. *The function \mathcal{D}_m is strongly convex with modulus $1/L$. As a consequence, problem (D_m) has a unique solution, denoted by $\lambda^*(m)$. Moreover, there exists a constant C^* independent of m such that*

$$\|\lambda^*(m)\| \leq C^*.$$

Proof. Since ∇f is L -Lipschitz continuous, we know that f^* is strongly convex with modulus $1/(2L)$ (i.e. $f^* - 1/L\|\cdot\|^2$ is convex) (see [3, Thm. 18.15]). Let us consider $u_\lambda(x)$ as a function of λ while fixing any $x \in X$. By definition, $\lambda \mapsto u_\lambda(x)$ is the infimum of a family of affine functions (with respect to λ), thus it is concave with respect to λ . Consequently, $-\int_X u_\lambda dm$ is convex with respect to λ . Therefore, \mathcal{D}_m is $1/L$ -strongly convex. Additionally, $\text{dom}(f^*)$ is both convex and closed. These properties guarantee the existence and uniqueness of the minimizer $\lambda^*(m)$.

Since M is an upper bound of $\|g(z)\|$, it follows that for all $\lambda \in \mathcal{H}$:

$$-M\|\lambda\| \leq \inf_{y \in Z_x} -\|\lambda\| \|g(x, y)\| \leq u_\lambda(x) \leq \sup_{y \in Z_x} \|\lambda\| \|g(x, y)\| \leq M\|\lambda\|.$$

Let $\lambda_0 \in \text{dom}(f^*)$. As $\mathcal{D}_m(\lambda^*(m)) \leq \mathcal{D}_m(\lambda_0)$, we can derive the following inequalities:

$$f^*(\lambda_0) + M\|\lambda_0\| \geq \mathcal{D}_m(\lambda_0) \geq \mathcal{D}_m(\lambda^*(m)) \geq f^*(\lambda^*(m)) - M\|\lambda^*(m)\|.$$

The strong convexity of f^* yields that

$$\frac{1}{2L} \|\lambda^*(m) - \lambda_0\|^2 + \langle p_0, \lambda^*(m) - \lambda_0 \rangle \leq f^*(\lambda^*(m)) - f^*(\lambda_0),$$

where $p_0 \in \partial f^*(\lambda_0)$. Combining the two above inequalities, we obtain:

$$\frac{1}{2L} \|\lambda^*(m) - \lambda_0\|^2 + \langle p_0, \lambda^*(m) - \lambda_0 \rangle \leq M(\|\lambda^*(m)\| + \|\lambda_0\|).$$

where $p_0 \in \partial f^*(\lambda_0)$. The announced result follows, with $C^* = 3\|\lambda_0\| + 2L(M + \|p_0\|)$. \blacksquare

5.2. Strong duality. Let us now prove the strong duality principle between (P_m) and (D_m), i.e., $\text{val}(\mathbf{P}_m) = \text{val}(\mathbf{D}_m)$. We will apply the Fenchel-Rockafellar theorem [27] to prove this relation.

Proposition 5.2. *Assume that the set $G_m := \{\int_\mu g d\mu \mid \mu \in \mathcal{P}_m(Z)\} \subseteq \mathcal{H}$ is closed. Then,*

- (1) $\text{val}(\mathbf{P}_m) = \text{val}(\mathbf{D}_m)$;
- (2) the primal problem (P_m) has a solution;
- (3) let $\lambda^*(m)$ be the solution of (D_m) and let μ be a solution of (P_m), then

$$\lambda^*(m) = \nabla f \left(\int_Z g d\mu \right).$$

Proof. Let us consider the following optimization problem with variable in \mathcal{H} :

$$(5.2) \quad \inf_{z \in \mathcal{H}} f(z) + \chi_{G_m}(z).$$

It is obvious that $\text{val}(\mathbf{P}_m) = \text{val}(\tilde{\mathbf{P}}_m) = \text{val}(5.2)$. The dual problem of (5.2) writes

$$(5.3) \quad \sup_{\lambda \in \mathcal{H}} -f^*(\lambda) - \chi_{G_m}^*(-\lambda).$$

By the definition of the Fenchel conjugate and the definition of G_m , we have

$$-\chi_{G_m}^*(-\lambda) = \inf_{z \in G_m} \langle \lambda, z \rangle = \inf_{\mu \in \mathcal{P}_m(Z)} \left\langle \lambda, \int_Z g d\mu \right\rangle.$$

Therefore, $\mathbf{val}(\mathbf{D}_m) = \mathbf{val}(5.1) = \mathbf{val}(5.3)$. Let us apply the Fenchel-Rockafellar theorem to (5.2). The function f is convex and continuous. The function χ_{G_m} is convex and lower semi-continuous from the fact that G_m is convex and closed. It is obvious that G_m is non-empty. Therefore, $0 \in \text{int}(\mathcal{H} - G_m) = \text{int}(\text{dom}f - \text{dom}\chi_{G_m})$. By the Fenchel-Rockafellar theorem [27], $\mathbf{val}(5.3) = \mathbf{val}(5.2)$, thus, $\mathbf{val}(\mathbf{P}_m) = \mathbf{val}(\mathbf{D}_m)$.

Since G_m is non-empty, bounded, convex, and closed, and since f is continuous and convex, we deduce from [8, Cor. 3.23] that problem (5.2) has a solution. Therefore, (\mathbf{P}_m) has a solution, denoted by μ . Since $\lambda^*(m)$ is the solution of (\mathbf{D}_m) , by the strong duality,

$$-f^*(\lambda^*(m)) + \inf_{z \in G_m} \langle \lambda^*(m), z \rangle = f \left(\int_Z g d\mu \right).$$

On the other hand, by the definition of Fenchel's conjugate,

$$f \left(\int_Z g d\mu \right) + f^*(\lambda^*(m)) \leq \left\langle \lambda^*(m), \int_Z g d\mu \right\rangle.$$

Combining the previous two inequalities and the fact that $\int_Z g d\mu \in G_m$, we deduce that

$$f \left(\int_Z g d\mu \right) + f^*(\lambda^*(m)) = \left\langle \lambda^*(m), \int_Z g d\mu \right\rangle.$$

We obtain that $\lambda^*(m) = \nabla f(\int_Z g d\mu)$ from Fenchel's relation. ■

5.3. Stability of the dual solution.

Lemma 5.3. *For any $\lambda_1, \lambda_2 \in \mathcal{H}$ and $x_1, x_2 \in X$, it holds that*

$$|u_{\lambda_1}(x_1) - u_{\lambda_2}(x_2)| \leq L_g \|\lambda_1\| d_X(x_1, x_2) + M \|\lambda_1 - \lambda_2\|.$$

Proof. By the triangle inequality,

$$|u_{\lambda_1}(x_1) - u_{\lambda_2}(x_2)| \leq |u_{\lambda_1}(x_1) - u_{\lambda_1}(x_2)| + |u_{\lambda_1}(x_2) - u_{\lambda_2}(x_2)|.$$

By the definition of u_λ , we have

$$u_{\lambda_1}(x_1) - u_{\lambda_1}(x_2) = \inf_{y_1 \in Z_{x_1}} \langle \lambda_1, g(x_1, y_1) \rangle - \inf_{y_2 \in Z_{x_2}} \langle \lambda_1, g(x_2, y_2) \rangle.$$

Let \tilde{y}_2^ϵ be an ϵ -minimizer of $\inf_{y_2 \in Z_{x_2}} \langle \lambda_1, g(x_2, y_2) \rangle$, with $\epsilon > 0$. By the Lipschitz continuity of \mathcal{Z} , there exists $\tilde{y}_1^\epsilon \in Z_{x_1}$ such that

$$\|g(x_1, \tilde{y}_1^\epsilon) - g(x_2, \tilde{y}_2^\epsilon)\| \leq L_g d_X(x_1, x_2).$$

By the Cauchy-Schwarz inequality, we have

$$u_{\lambda_1}(x_1) - u_{\lambda_1}(x_2) \leq \langle \lambda_1, g(x_1, \tilde{y}_1^\epsilon) - g(x_2, \tilde{y}_2^\epsilon) \rangle + \epsilon \leq \|\lambda_1\| L_g d_X(x_1, x_2) + \epsilon.$$

By the arbitrariness of ϵ , we have $|u_{\lambda_1}(x_1) - u_{\lambda_1}(x_2)| \leq \|\lambda_1\| L_g d_X(x_1, x_2)$.

On the other hand,

$$u_{\lambda_1}(x_2) - u_{\lambda_2}(x_2) = \inf_{y \in Z_{x_2}} \langle \lambda_1, g(x_2, y) \rangle - \inf_{y \in Z_{x_2}} \langle \lambda_2, g(x_2, y) \rangle \leq \sup_{y \in Z_{x_2}} \langle \lambda_1 - \lambda_2, g(x_2, y) \rangle.$$

By the Cauchy-Schwarz inequality and the definition of M , we have that

$$u_{\lambda_1}(x_2) - u_{\lambda_2}(x_2) \leq M \|\lambda_1 - \lambda_2\|.$$

The conclusion follows. ■

Lemma 5.4 (Stability of the dual problem). *For any $m_0, m_1 \in \mathcal{P}(\Omega)$, we have*

$$(5.4) \quad |\mathcal{D}_{m_0}(\lambda^*(m_0)) - \mathcal{D}_{m_1}(\lambda^*(m_1))| \leq C^* L_g d_1(m_0, m_1),$$

$$(5.5) \quad \|\lambda^*(m_0) - \lambda^*(m_1)\|^2 \leq 2C^* L_g L d_1(m_0, m_1),$$

where C^* is the a priori bound of $\|\lambda^*(\cdot)\|$ obtained in Lemma 5.1

Proof. According to Lemma 5.1, we know that $\|\lambda^*(m_0)\|$ and $\|\lambda^*(m_1)\|$ are smaller than C^* . Then, by Lemma 5.3, $u_{\lambda^*(m_0)}(x)$ and $u_{\lambda^*(m_1)}(x)$ are (C^*L_g) -Lipschitz continuous with respect to x . Hence,

$$\begin{aligned} \mathcal{D}_{m_0}(\lambda^*(m_0)) &= f^*(\lambda^*(m_0)) - \int_X u_{\lambda^*(m_0)}(x) dm_0(x) \\ (5.6) \quad &= f^*(\lambda^*(m_0)) - \int_X u_{\lambda^*(m_0)}(x) dm_1(x) + \int_X u_{\lambda^*(m_0)}(x) d(m_1 - m_0)(x) \\ &\geq \mathcal{D}_{m_1}(\lambda^*(m_0)) - C^*L_g d_1(m_0, m_1), \end{aligned}$$

where the third line is by the definition of the Kantorovich–Rubinstein distance. Since $\lambda^*(m_1)$ minimizes \mathcal{D}_{m_1} and since \mathcal{D}_{m_1} is $1/L$ -strongly convex, we have

$$(5.7) \quad \mathcal{D}_{m_1}(\lambda^*(m_0)) \geq \mathcal{D}_{m_1}(\lambda^*(m_1)) + \frac{1}{2L} \|\lambda^*(m_0) - \lambda^*(m_1)\|^2.$$

Combining (5.6) and (5.7), we obtain that

$$\mathcal{D}_{m_0}(\lambda^*(m_0)) \geq \mathcal{D}_{m_1}(\lambda^*(m_1)) + \frac{1}{2L} \|\lambda^*(m_0) - \lambda^*(m_1)\|^2 - C^*L_g d_1(m_0, m_1).$$

In particular, we have $\mathcal{D}_{m_1}(\lambda^*(m_1)) - \mathcal{D}_{m_0}(\lambda^*(m_0)) \leq C^*L_g d_1(m_0, m_1)$. Exchanging the positions of m_0 and m_1 in (5.6), we obtain

$$(5.8) \quad \mathcal{D}_{m_1}(\lambda^*(m_1)) \geq \mathcal{D}_{m_0}(\lambda^*(m_0)) + \frac{1}{2L} \|\lambda^*(m_0) - \lambda^*(m_1)\|^2 - C^*L_g d_1(m_0, m_1).$$

Inequality (5.4) follows immediately and (5.5) is deduced by summing (5.6)-(5.8). \blacksquare

5.4. Directional derivative of the value function. The value function of problem (P_m) is defined by

$$V: \mathcal{P}^1(X) \rightarrow \mathbb{R}, \quad m \mapsto \mathbf{val}(P_m).$$

Our goal is to characterize the directional derivative of V . Define the following function:

$$v: \mathcal{P}^1(X) \times X \rightarrow \mathbb{R}, \quad (m, x) \mapsto u_{\lambda^*(m)}(x).$$

Proposition 5.5. *Assume that G_m is closed for any $m \in \mathcal{P}^1(X)$. Then for any $m_0, m_1 \in \mathcal{P}^1(X)$, we have*

$$\lim_{t \rightarrow 0^+} \frac{V(m_0 + t(m_1 - m_0)) - V(m_0)}{t} = \int_X v(m_0, x) d(m_1 - m_0)(x).$$

As a consequence, v is the directional derivative of V , i.e.,

$$V(m_1) - V(m_0) = \int_{t=0}^1 \int_X v(m_0 + t(m_1 - m_0), x) d(m_1 - m_0)(x) dt.$$

Proof. For any $t \in [0, 1]$, let $m_t = m_0 + t(m_1 - m_0)$. By the strong duality, we have

$$V(m_t) - V(m_0) = \mathcal{D}_{m_0}(\lambda^*(m_0)) - \mathcal{D}_{m_t}(\lambda^*(m_t)).$$

From (5.6), we deduce that

$$(5.9) \quad \mathcal{D}_{m_0}(\lambda^*(m_0)) - \mathcal{D}_{m_t}(\lambda^*(m_t)) \geq \int_X v(m_0, x) d(m_t - m_0)(x) = t \int_X v(m_0, x) d(m_1 - m_0)(x).$$

On the other hand, let μ_0 and μ_1 be solutions of (P_m) with $m = m_0$ and m_1 respectively. Let $\mu_t = \mu_0 + t(\mu_1 - \mu_0)$. It is obvious that $\pi_1 \# \mu_t = m_t$. Therefore,

$$V(m_t) - V(m_0) \leq f\left(\int_Z g d\mu_t\right) - f\left(\int_Z g d\mu_0\right).$$

By Proposition 5.2, $\lambda^*(m_0) = \nabla f(\int_Z g d\mu_0)$. Since ∇f is L -Lipschitz, it follows that

$$f\left(\int_Z g d\mu_t\right) - f\left(\int_Z g d\mu_0\right) \leq t \int_X v(m_0, x) d(m_1 - m_0)(x) + \frac{Lt^2}{2} \left\| \int_Z g d(\mu_1 - \mu_0) \right\|^2.$$

Recall the definition of D . Combining the two inequalities above, we have

$$(5.10) \quad V(m_t) - V(m_0) \leq t \int_X v(m_0, x) d(m_1 - m_0)(x) + \frac{LDt^2}{2}.$$

We have $V(m_0) = -\mathcal{D}_{m_0}(\lambda^*(m_0))$ and $V(m_t) = -\mathcal{D}_{m_t}(\lambda^*(m_t))$. Using (5.9)-(5.10) and letting t go to 0^+ , we obtain that

$$\lim_{t \rightarrow 0^+} \frac{V(m_t) - V(m_0)}{t} = \int_X v(m_0, x) d(m_1 - m_0)(x).$$

From Lemmas 5.3-5.4, we deduce that the function $v(m, x)$ is continuous in $\mathcal{P}^1(X) \times X$ with respect to the distance (d_1, d_X) . Let us define two functions from $[0, 1]$ to \mathbb{R} ,

$$\begin{aligned} \bar{V} &: [0, 1] \rightarrow \mathbb{R}, t \mapsto V(m_t); \\ \bar{v} &: [0, 1] \rightarrow \mathbb{R}, t \mapsto \int_X v(m_t, x) d(m_1 - m_0)(x). \end{aligned}$$

For any $0 \leq t \leq T < 1$, observe that $m_T = m_t + \frac{T-t}{1-t}(m_1 - m_t)$ and $m_1 - m_t = (1-t)(m_1 - m_0)$. By using the same arguments as in (5.9)-(5.10), we have

$$(T-t)\bar{v}(t) \leq \bar{V}(T) - \bar{V}(t) \leq (T-t)\bar{v}(t) + \frac{LD(T-t)^2}{2(1-t)^2}.$$

We deduce that $\bar{v}(t)$ is the right derivative of \bar{V} at t for any $t \in [0, 1]$. By exchanging positions of m_0 and m_1 , we can prove that $\bar{v}(t)$ is the left derivative of \bar{V} at t for any $t \in (0, 1]$. Therefore, \bar{V} is differentiable at each point on $[0, 1]$ and \bar{v} is its derivative. Since \bar{v} is continuous, by the fundamental theorem of calculus [28, Thm. 7.21], we have that $\bar{V}(1) - \bar{V}(0) = \int_{t=0}^1 \bar{v}(t) dt$. ■

6. NUMERICAL APPROACH

We present in this section our numerical method for solving (P_m) . The first step of resolution consists in discretizing m . We replace it by an empirical distribution m_N and focus next on the resolution of (P_{m_N}) . By Theorem 4.7(1), we have

$$(6.1) \quad |\mathbf{val}(P_m) - \mathbf{val}(P_{m_N})| \leq L_g(C + LM)d_1(m, m_N).$$

We give theoretical bounds for the minimal value of $d_1(m, m_N)$ in Subsection 6.1. Then we discuss the resolution of (P_{m_N}) with the Frank-Wolfe algorithm in Subsection 6.2. Finally in Subsection 6.3 we propose to use a variant of the Frank-Wolfe algorithm, called Stochastic Frank-Wolfe (SFW) algorithm, introduced in [6]. This method generates a solution to (P_{m_N}) which is an empirical distribution.

6.1. Discretization. In view of (6.1), one should look for an empirical distribution $m_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ that is as close as possible to m for the d_1 -distance. This problem is commonly known as the optimal quantization problem, and for detailed information on this topic, we refer to [17]. Here, we present a slightly modified version of an optimal quantization result obtained in [25, Prop. 12]. For any subset A of X , we denote by $r_N(A)$ the minimum radius r required to cover A with N closed balls of radius r . It is defined by

$$r_N(A) := \inf_{x \in A^N} \min \left\{ r \geq 0 \mid A \subseteq \bigcup_{i=1}^N B_X(x_i, r) \right\}.$$

The upper box-counting dimension (or the upper Minkowski dimension) of A [15, p. 41] is defined as follows:

$$\bar{D}(A) := \inf \left\{ \bar{D} > 0 \mid \exists \bar{C} > 0 \text{ such that } r_N(A) \leq \bar{C}N^{-1/\bar{D}}, \forall N \in \mathbb{N}_+ \right\}.$$

Lemma 6.1. Let $m \in \mathcal{P}^1(X)$, and let $A \subseteq X$ be the support of m . There exists a sequence $(m_N)_{N \geq 1}$ of empirical distributions on X such that the following holds:

(1) If $\bar{D}(A) > 1$, then there exists a constant \tilde{C}_1 such that for any $N \geq 1$,

$$d_1(m, m_N) \leq \tilde{C}_1 N^{-\frac{1}{\bar{D}(A)}}.$$

(2) If $\bar{D}(A) = 1$, then there exists a constant \tilde{C}_2 such that for any $N \geq 1$,

$$d_1(m, m_N) \leq \tilde{C}_2 N^{-1} \log N.$$

(3) If $\bar{D}(A) < 1$, then there exists a constant \tilde{C}_3 such that for any $N \geq 1$,

$$d_1(m, m_N) \leq \tilde{C}_3 N^{-1}.$$

Proof. This follows from the proof presented in [25, Prop. 12], with the only difference being that in the final inequality, we employ the triangle inequality for the d_1 -distance instead of the Minkowski inequality for the Wasserstein-2 distance. \blacksquare

Remark 6.2. If A is a subset of a smooth d -dimensional submanifold of a Euclidean space, then $\bar{D}(A) \leq d$. This estimate is deduced from [15, p. 48 (i)-(ii)].

6.2. Frank-Wolfe algorithm. For general convex optimization problems, the Frank-Wolfe algorithm relies on the resolution of a sequence of linearized problems, obtained by replacing the cost function of the problem by a first-order Taylor approximation of it. In the context of problem (\mathbf{P}_{m_N}) , the linearized problem is of the general form:

$$(6.2) \quad \inf_{\mu \in \mathcal{P}_{m_N}(Z)} \left\langle \lambda, \int_Z g d\mu \right\rangle,$$

for some $\lambda \in \nabla f(\mathcal{H})$.

A key observation from Lemma 3.3 is that a solution of the linearized problem, denoted by μ_λ , can be obtained as in the proof of Lemma 3.3, in the simple case where m is a finitely-supported probability measure: for all $i = 1, \dots, N$, find $y_i \in \mathbf{BR}_\lambda(x_i)$ and set $\mu_\lambda = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)}$. Therefore, one can consider applying the Frank-Wolfe algorithm to solve (\mathbf{P}_{m_N}) , in which the main task is to solve (6.2).

Algorithm 2: Frank-Wolfe Algorithm

Initialization: $\mu^0 \in \mathcal{P}_{m_N}(Z)$. Set $K \geq 1$.
for $k = 0, 1, 2, \dots, K - 1$ **do**
 Compute $\lambda^k = \nabla f \left(\int_Z g d\mu^k \right)$.
 Solve (6.2) for $\lambda = \lambda^k$, the solution is denoted by μ_{λ^k} .
 Choose $\omega_k \in [0, 1]$.
 Set $\mu^{k+1} = (1 - \omega_k)\mu^k + \omega_k\mu_{\lambda^k}$.

end for

Remark 6.3. If we take $\omega_k = 1/(k+1)$ for all k , then it is easy to see that $\mu^K = \frac{1}{K} \sum_{k=0}^{K-1} \mu_{\lambda^k}$. We recover the fictitious play of [10], applied to the Lagrangian discretization of first-order MFGs.

Lemma 6.4. Let Assumption A hold true. In Algorithm 2, we set $\omega_k = 2/(k+2)$ for all k . Then for any $K \geq 1$,

$$f \left(\int_Z g d\mu^K \right) - \mathbf{val}(\mathbf{P}_{m_N}) \leq \frac{2LD}{K}.$$

Proof. This is a consequence of [6, Prop. 3.4]. \blacksquare

6.3. Stochastic Frank-Wolfe algorithm. In Algorithm 2, at each iteration, we generate the output by taking a convex combination of the previous iteration's result and the solution of (6.2). This process requires us to add N points from $\mathbf{BR}_{\lambda^k}(x_i)$, for $i = 1, \dots, N$, to stock the support of solution at each iteration. As a consequence, this approach can lead to a memory overflow issue, as K going to infinity. The large support of μ^K will also raise the difficulty of Step 2 in Algorithm 1, in which we will take $\bar{\mu}_0 = \mu^K$. To address this issue, we will use the stochastic Frank-Wolfe algorithm [6] to (\mathbf{P}_{m_N}) . This approach will enable us to obtain an approximate empirical solution (\mathbf{P}_{m_N}) , and can effectively handle the large support of μ^K .

Lemma 6.5. *Let $\mu \in \mathcal{P}(Z)$. Then μ lies in $\mathcal{P}_{m_N}(Z)$ if and only if there exists $\mu_i \in \mathcal{P}(Z_{x_i})$ for any $i = 1, \dots, N$ such that $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \mu_i$.*

Proof. If $\mu_i \in \mathcal{P}(Z_{x_i})$, then $\pi_1 \#(\delta_{x_i} \otimes \mu_i) = \delta_{x_i}$. Since the push-forward operator $\#$ is linear, we have that $\pi_1 \#(\frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \mu_i) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} = m_N$.

Conversely, let us assume that $\mu \in \mathcal{P}_{m_N}(Z)$. By Theorem 2.7 and its remark, we can conclude that there exists $\mu_{x_i} \in \mathcal{P}(Z_{x_i})$ for $i = 1, \dots, N$ such that for any bounded and continuous function $h: Z \rightarrow \mathbb{R}$, we have

$$\int_Z h d\mu = \frac{1}{N} \sum_{i=1}^N \int_{Z_{x_i}} h(x_i, y_i) d\mu_{x_i}(y_i).$$

Applying Fubini's theorem to the equality above, we have

$$\int_Z h d\mu = \frac{1}{N} \sum_{i=1}^N \int_Z h d(\delta_{x_i} \otimes \mu_{x_i}) = \int_Z h d\left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \mu_{x_i}\right).$$

This implies that $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \mu_{x_i}$. ■

According to Lemma 6.5 and Fubini's theorem, the discretized problem (\mathbf{P}_{m_N}) is equivalent to

$$(6.3) \quad \inf_{\mu_i \in \mathcal{P}(Z_{x_i})} f\left(\frac{1}{N} \sum_{i=1}^N \int_{Z_{x_i}} g(x_i, y_i) d\mu_i(y_i)\right).$$

Problem (6.3) is the randomized relaxation of an N -agent optimization problem as investigated in [6],

$$(6.4) \quad \inf_{y \in \prod_{i=1}^N Z_{x_i}} f\left(\frac{1}{N} \sum_{i=1}^N g(x_i, y_i)\right).$$

Problem (6.4) is equivalent to a version of problem (6.3) in which the probability measures μ_i are restricted to be Dirac measures. In particular, we can associate with each feasible element $y = (y_i)_{i=1, \dots, N} \in \prod_{i=1}^N Z_{x_i}$ (for problem (6.4)) the tuple $(\delta_{x_i})_{i=1, \dots, N}$, which is feasible for (6.3), and the probability distribution $\frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)}$, which is feasible for (\mathbf{P}_{m_N}) .

We apply the following Stochastic Frank-Wolfe algorithm, investigated in [6], to solve problems (6.3) and (6.4). Let $\text{Bern}(\omega)$ be the Bernoulli distribution with a parameter $\omega \in [0, 1]$.

Algorithm 3: Stochastic Frank-Wolfe Algorithm

Initialization: $y^0 \in \prod_{i=1}^N Z_{x_i}$. Set $K \geq 1$.
for $k = 0, 1, 2, \dots, K - 1$ **do**
 Compute $\lambda^k = \nabla f(\frac{1}{N} \sum_{i=1}^N g(x_i, y_i^k))$.
 for $i = 1, 2, \dots, N$ **do**
 Find $\hat{y}_i^k \in \mathbf{BR}_{\lambda^k}(x_i)$.
 end for
 Choose $n_k \in \mathbb{N}^*$. Set $\omega_k = 2/(k + 2)$.
 for $j = 1, 2, \dots, n_k$ **do**
 for $i = 1, 2, \dots, N$ **do**
 Simulate $P_i^{k,j} \sim \text{Bern}(\omega_k)$, independently of all previously defined random variables.
 Set $\hat{y}_i^{k,j} = (1 - P_i^{k,j})y_i^k + P_i^{k,j}\hat{y}_i^k$.
 end for
 Define $\hat{y}^{k,j} = (\hat{y}_i^{k,j})_{i=1, \dots, N}$.
 end for
 Find $y^{k+1} \in \text{argmin} \{f(\frac{1}{N} \sum_{i=1}^N g(x_i, y_i)) \mid y \in \{\hat{y}^{k,j}, j = 1, 2, \dots, n_k\}\}$.
end for

The interest of Algorithm 3 is that it provides an approximate solution to (6.4), and the associated empirical distribution serves as a reliable approximate solution of the problem (6.3), as demonstrated in the following lemma. Additionally, this empirical distribution has a fixed support size N , which does not increase with the iteration number, making the algorithm memory-efficient.

Lemma 6.6. *in Algorithm 3, whatever the numbers $(n_k)_{k \in \mathbb{N}}$, we have for any $K = 1, 2, \dots, 2N$ that*

$$\mathbb{E} \left[f \left(\frac{1}{N} \sum_{i=1}^N g(x_i, y_i^K) \right) \right] - \mathbf{val}(6.3) \leq \frac{4LD}{K}.$$

Proof. This is from [6, Thm. 3.7]. ■

Remark 6.7. Lemma 6.6 provides a convergence result for Algorithm 3 in terms of expectation. An estimate of the following quantity can be found in [6, Thm. 3.7]:

$$\mathbb{P} \left[f \left(\frac{1}{N} \sum_{i=1}^N g(x_i, y_i^K) \right) \geq \mathbf{val}(6.3) + \epsilon + \frac{4LD}{K} \right],$$

for a given $\epsilon > 0$. In particular, this probability can be made arbitrarily small, provided that the numbers n_k are large enough.

In order to obtain an approximate solution of (\mathbf{P}_m) , we combine Algorithm 3 with Algorithm 1. Let us consider the outcome y^K of Algorithm 3 after K iterations, for $1 \leq K \leq 2N$ and for arbitrary numbers $n_k \geq 1$ of simulations. Let $\mu_N^K = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i^K)}$. Moving on to Algorithm 1, we utilize the following inputs: $m_0 = m_N$, $m_1 = m$, and $\tilde{\mu}_0 = \mu_N^K$. The output of Algorithm 1 is denoted as $\tilde{\mu}^K$, which is an element of the set $\mathcal{P}_m(Z)$. We have the following convergence result for the combination of Algorithm 1 and 3.

Theorem 6.8. *Let Assumptions A-B hold true, and let $m \in \mathcal{P}^1(Z)$. Then,*

$$\mathbb{E} \left[f \left(\int_Z g d\tilde{\mu}^K \right) \right] - \mathbf{val}(\mathbf{P}_m) \leq \frac{4LD}{K} + 2L_g(C + LM)d_1(m_N, m).$$

Proof. Since $f\left(\int_Z g d\mu_N^K\right) = f\left(\frac{1}{N} \sum_{i=1}^N g(x_i, y_i^K)\right)$, by Lemma 4.5, we have

$$f\left(\int_Z g \tilde{\mu}^K\right) - f\left(\frac{1}{N} \sum_{i=1}^N g(x_i, y_i^K)\right) \leq L_g(C + LM)d_1(m_N, m), \quad \text{almost surely.}$$

Taking expectation on both sides of the previous inequality, and applying Lemma 6.6 and the relation $\mathbf{val}(6.3) = \mathbf{val}(\mathbf{P}_{m_N})$, we have

$$\mathbb{E}\left[f\left(\int_Z g d\tilde{\mu}^K\right)\right] - \mathbf{val}(\mathbf{P}_{m_N}) \leq \frac{2LD}{K} + L_g(C + LM)d_1(m_N, m).$$

Combining with Theorem 4.7(1), the proof is complete. \blacksquare

Remark 6.9. The realization of Algorithm 1 can be simplified in Theorem 6.8 thanks to the empirical structure of m_0 and $\bar{\mu}_0$, as noted in Remark 4.6.

7. EXAMPLES AND NUMERICAL RESULTS

7.1. The traffic assignment problem. The traffic assignment problem is a non-atomic game whose potential formulation takes the form of problem (\mathbf{P}_m) . We describe it briefly in this subsection. Consider a finite set of nodes \mathcal{N} and a finite set of edges $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$. The model is a static model that describes how the agents move on the network, taking into account their origins and destinations as well as the congestion on each arc.

We fix a subset X of $\mathcal{N} \times \mathcal{N}$. Each parameter $x = (x_1, x_2) \in X$ represents an origin-destination pair. Next, we denote by Y the set of subsets of \mathcal{E} . For each x , we consider a set of possible paths connecting x_1 and x_2 , denoted Z_x . Mathematically, we simply describe a path as a subset of \mathcal{E} , so $Z_x \subseteq Y$.

For the definition of the potential problem, we define $\mathcal{H} = \mathbb{R}^{\mathcal{E}}$. The function g is defined by $g: (x, y) \in Z \mapsto (g(x, y)_e)_{e \in \mathcal{E}} \in \mathcal{H}$, where

$$g(x, y)_e = \begin{cases} 1, & \text{if } e \in y, \\ 0, & \text{otherwise.} \end{cases}$$

In words, $g(x, y)_e = 1$ is the edge e belongs to the path y , 0 otherwise. Next, we fix a family of functions $\phi_e: [0, \infty) \rightarrow [0, \infty)$, parametrized by $e \in \mathcal{E}$. We assume that these functions are non-decreasing and we fix a primitive Φ_e for each of them. The functions Φ_e are convex. Finally, we define $f: \mathcal{H} \rightarrow \mathbb{R}$ by

$$f(q) = \sum_{e \in \mathcal{E}} \Phi_e(q_e).$$

With these definitions at hand, it remains to interpret the optimality conditions for the associated MFO problem. Let us consider $\bar{\mu} \in \mathcal{P}_m(Z)$ and let $\bar{\lambda} = \nabla f(\int_Z g d\bar{\mu}) \in \mathcal{H}$. A direct calculation shows that

$$\bar{\lambda}_e = \phi_e(q_e), \quad \text{where: } q_e = \sum_{\substack{(x,y) \in Z \\ e \in y}} \bar{\mu}(x, y).$$

We can interpret q_e as the proportion of agents using the edge e . We interpret ϕ_e as a function that gives the travelling time on the edge e in function of the congestion q_e . So here the dual variable $\bar{\lambda}$ has a natural interpretation as a vector containing all the travelling times of the network. Finally, for any $x \in X$, we have

$$\inf_{y \in Z_x} \langle \bar{\lambda}, g(x, y) \rangle = \inf_{y \in Z_x} \sum_{e \in y} \bar{\lambda}_e.$$

Here $\sum_{e \in y} \bar{\lambda}_e$ describes the total duration of the path y . As a consequence, solving the MFO problem is equivalent to find $\bar{\mu}$ such that for any x , $\bar{\mu}_x$ is supported by the optimal paths (among

those connecting x_1 to x_2), the travel time of a path y begin defined by the above relations. This notion of equilibrium is known as Wardrop equilibrium in the literature. Using the MFO setting, we recover the well-known equivalence between Wardrop equilibria and their potential formulation, see [31, Chapter 3]. We mention here that the modelling is different (but equivalent) to the standard one in which one rather describes the distribution of the agents with respect to the edges, instead of using the distribution with respect to the paths. We also note that the Frank-Wolfe algorithm is a very standard algorithm for solving those problems, see [31, Section 5.2].

7.2. Lagrangian MFGs. We propose here a class of potential Lagrangian MFGs. As before, we first formulate the potential problem, in the form of an MFO problem and interpret next the optimality conditions as a non-atomic game. There is a large amount of literature on Lagrangian MFGs, we refer the reader to [4, 5, 9, 29, 30] and the references therein. Our intention here is only to formulate a model that fits with the framework of MFO problems, we do not check the corresponding assumptions, which must be done on a case-by-case basis.

Let us fix a domain $\Omega \subseteq \mathbb{R}^d$ and a final time $T > 0$. Let $\text{AC}([0, T], \mathbb{R}^d)$ be the set of all absolutely continuous functions from $[0, 1]$ to \mathbb{R}^d . For any $x \in \Omega$, we denote,

$$Y := \{y \in \text{AC}([0, T], \mathbb{R}^d) \mid y(t) \in \Omega, \forall t \in [0, T]\}, \quad Z_x := \{y \in Y \mid y(0) = x\}.$$

Let $Z = \{(x, y) \mid x \in \Omega, y \in Z_x\}$. Let $m \in \mathcal{P}(\Omega)$ be the distribution of the initial states of the players. We fix three functions $L: \mathbb{R}^d \rightarrow \mathbb{R}$, $h: \mathbb{R}^d \rightarrow \mathbb{R}^k$, and $\Phi: \mathbb{R}^k \rightarrow \mathbb{R}$. We define $g: (x, y) \in Z \rightarrow (g_1(x, y), g_2(x, y)) \in \mathbb{R} \times L^2(0, T; \mathbb{R}^k)$ by

$$g_1(x, y) = \int_0^T L(\dot{y}(t)) dt \quad \text{and} \quad g_2(x, y) = h \circ y$$

and we define $f: \mathbb{R} \times L^2(0, T; \mathbb{R}^k) \rightarrow \mathbb{R}$ as

$$f(q_1, q_2) = q_1 + \int_0^T \Phi(q_2(t)) dt.$$

With these definitions of X , Y , Z_x , g , and f , we have a full description of an MFO problem. Let us write the corresponding optimality conditions. Let $\bar{\mu} \in \mathcal{P}_m(Z)$ and let $\bar{\lambda} = \nabla f(\int_Z g d\bar{\mu})$. Then, $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2)$, with $\bar{\lambda}_1 = 1$ and

$$\bar{\lambda}_2(t) = \nabla \Phi \left(\int_Z h(y(t)) d\bar{\mu}(x, y) \right).$$

For any $t \in [0, T]$, denote by $e_t: Z \rightarrow \mathbb{R}^d$ the mapping defined by $e_t(x, y) = y(t)$. Then $\bar{\lambda}_2$ is equivalently defined by

$$\bar{\lambda}_2(t) = \nabla \Phi \left(\int_{\Omega} h(y') d\bar{m}_t(y') \right), \quad \text{where: } m_t = e_t \# \bar{\mu}.$$

In this context, the minimization problem in (3.4) is equivalent to the following optimal control problem:

$$\inf_{y \in Z_x} \int_0^T \left[L(\dot{y}(t)) + \left\langle \nabla \Phi \left(\int_{\Omega} h d\bar{m}_t \right), h(y(t)) \right\rangle \right] dt.$$

Let us note that the above problem is not convex in general. It could be solved by dynamic programming in the situation where the dimension of y is moderate.

7.3. Numerical results for a competition problem with a non-renewable resource.

Model. We consider a Lagrangian MFG in which the agents exploit their own stock of an exhaustible resource. The model is taken from [18]. We fix a time horizon $[0, T]$ where $T \in [0, +\infty)$ (the case $T = \infty$ investigated in [18] is not considered here). The state variable of a representative agent is the level of the stock of resource at any time, denoted $(X_t^q)_{t \in [0, T]}$ and the control is the speed of extraction at any time, denoted q . The dynamic of a given producer with an initial position $x_0 \geq 0$ is described as follows:

$$X_t^q := x_0 - \int_0^t q_\tau d\tau, \quad t \in [0, T],$$

where $q_t \geq 0$, for any $t \in [0, T]$. We impose that $X_T^q \geq 0$, which implies that $X_t^q \geq 0$ at any time.

We define the set of aggregate production, denoted as \mathcal{G} , by

$$\mathcal{G} := \left\{ Q \in \mathbb{L}^2([0, T], \mathbb{R}) \mid 0 \leq Q(t) \leq \frac{1}{2}, \forall t \in [0, T] \right\}.$$

The price of the resource for this representative producer depends on its extracting speed and an aggregate production $Q \in \mathcal{G}$,

$$p_t := 1 - q_t - \epsilon Q_t, \quad t \in [0, T],$$

where $\epsilon \in (0, 1)$ is a constant. The gain of this representative producer writes,

$$\int_0^T e^{-rt} q_t (1 - q_t - \epsilon Q_t) dt.$$

where $r \geq 0$ is a discount rate. Therefore, given an aggregate production $Q \in \mathcal{G}$ and an initial position $x_0 \geq 0$, we can formulate an optimal control problem associated with this representative producer,

$$(7.1) \quad \begin{cases} \inf_{q \in \mathcal{G}} J^Q(q) := \int_0^T e^{-rt} q_t (q_t - 1 + \epsilon Q_t) dt; \\ \text{s.t.} \quad \int_0^T q_t dt \leq x_0. \end{cases}$$

Lemma 7.1. *Problem (7.1) has a unique solution $q^Q(x_0)$. Moreover, $0 \leq q^Q(x_0)(t) \leq \frac{1}{2}$, for a.e. $t \in (0, T)$.*

Proof. It is easy to see that \mathcal{G} is a non-empty and convex subset of $\mathbb{L}^2([0, T], \mathbb{R})$. Following [28, Thm. 3.12], if $(f_n)_{n \geq 1}$ converges to f in \mathbb{L}^2 sense, then there exists a subsequence of $(f_n)_{n \geq 1}$ converges to f a.e. As a consequence, f lies in \mathcal{G} . Therefore, \mathcal{G} is closed. Furthermore, by Hölder's inequality, we obtain that $\{q \in \mathbb{L}^2([0, T], \mathbb{R}) \mid \int_0^T q_t dt \leq x_0\}$ is non-empty, convex and closed in $\mathbb{L}^2([0, T], \mathbb{R})$. It follows that the admissible set of problem (7.1) is non-empty, closed and convex in Hilbert space $\mathbb{L}^2([0, T], \mathbb{R})$. On the other hand, the cost function $J^Q(\cdot)$ is strongly convex. Then the existence of the solution of (7.1) comes from [8, Cor. 3.23] and the uniqueness is by the strong convexity of J^Q .

Let q be the solution to (7.1). Define $q'(t) = \min\{q(t), \frac{1}{2}\}$, for a.e. $t \in (0, T)$. Since $q' \leq q$, q' is also feasible for problem (7.1). Moreover, the running cost $q_0 \mapsto q_0(q_0 - 1 + \epsilon Q_t)$ is increasing for $q_0 \geq \frac{1}{2}$. As a consequence, $J^Q(q') \leq J^Q(q)$. Therefore, q' is optimal, and since the solution is unique, we have $q = q'$, which proves that $q \leq \frac{1}{2}$. ■

Let $m \in \mathcal{P}([0, +\infty))$ denote the distribution of the initial conditions of the producers. The aggregate production rate corresponding to q^Q is given by

$$Q_t^Q := \int_0^\infty q_t^Q(x_0) dm(x_0), \quad \forall t \in [0, T].$$

Following [18], we call Nash equilibrium a solution Q^* to the fix-point problem:

$$(7.2) \quad Q^* = Q^{Q^*}, \quad Q^* \in \mathcal{G}.$$

Potential problem. In this paragraph, we find an optimization problem associated with the fixed point problem (7.2), which is a particular case of problem (P_m). Let us specify metric spaces and admissible sets in (P_m) associated with (7.2):

$$X = [0, \infty), \quad Y = \mathcal{G}, \quad F(x) = \left\{ q \in \mathcal{G} \mid \int_0^T q_t dt \leq x \right\}, \quad Z = \text{Graph}(F), \quad Z_x = F(x).$$

Let us define the separable Hilbert space $\mathbb{L}_{e^{-rt}}^2([0, T])$ [28, Example. 4.5(b)]:

$$\mathbb{L}_{e^{-rt}}^2([0, T]) := \left\{ \zeta: [0, T] \rightarrow \mathbb{R} \text{ is Lebesgue measurable} \mid \int_0^T e^{-rt} |\zeta(t)|^2 dt < +\infty \right\},$$

with a scalar product,

$$\langle f_1, f_2 \rangle_{\mathbb{L}_{e^{-rt}}^2([0, T])} = \int_0^T e^{-rt} f_1(t) f_2(t) dt.$$

It is easy to check that $Y = \mathcal{G} \subseteq \mathbb{L}_{e^{-rt}}^2([0, T])$. Then, in (P_m), we set $\mathcal{H} = \mathbb{R} \times \mathbb{L}_{e^{-rt}}^2([0, T])$,

$$g: Z \rightarrow \mathcal{H}, (x, q) \mapsto \left(\int_0^T e^{-rt} (q_t^2 - q_t) dt, q \right),$$

$$f: \mathcal{H} \rightarrow \mathbb{R}, (y_1, y_2) \mapsto y_1 + \frac{\epsilon}{2} \|y_2\|_{\mathbb{L}_{e^{-rt}}^2([0, T])}^2.$$

Therefore, problem (P_m) associated with (7.2) writes:

$$(7.3) \quad \inf_{\mu \in \mathcal{P}_m(Z)} \int_Z \int_0^T e^{-rt} (q_t^2 - q_t) dt d\mu(x, q) + \frac{\epsilon}{2} \int_0^T e^{-rt} \left(\int_Z q_t d\mu(x, q) \right)^2 dt.$$

Proposition 7.2. *If $\bar{\mu}$ is a solution of problem (7.3), then $Q^* = \int_Z q d\bar{\mu}(x, q)$ is a Nash equilibrium of the optimal exploitation of exhaustible resources problem, i.e., Q^* is a solution of (7.2).*

Proof. Let us first check that Assumption A holds true for problem (7.3). It is easy to see that Assumption A(1) and the first and the third points in Assumption A(2) are true by the continuity of g and Lemma 7.1. Let us prove that G_λ is lower semi-continuous for any $\lambda \in \mathcal{H}_f$. This is a consequence of the claim that the set-valued function $\mathcal{Z}: X \rightsquigarrow \mathcal{H}, x \mapsto \{g(x, y) \mid y \in Z_x\}$ is locally Lipschitz, i.e. Lipschitz in any compact set of X . To see the local Lipschitz continuity, we fix any $x_1 < x_2$ in X . If $q \in Z_{x_1}$, then we have immediately that $q \in Z_{x_2}$. This implies that $Z_{x_1} \subseteq Z_{x_2}$. On the other hand, let $q \in Z_{x_2}$. We construct $q' \in Z_{x_1}$ by the following method:

$$q'_t = \begin{cases} q_t, & \text{if } \int_0^t q_\tau d\tau \leq x_1; \\ 0, & \text{otherwise.} \end{cases}$$

As a consequence, we have that $\|q' - q\|_{\mathbb{L}^1([0, T])} \leq x_2 - x_1$. Therefore, by Hölder's inequality,

$$\|q' - q\|_{\mathbb{L}_{e^{-rt}}^2([0, T])}^2 \leq \|e^{-rt}(q' - q)\|_{\mathbb{L}^\infty([0, T])} \|q' - q\|_{\mathbb{L}^1([0, T])} \leq x_2(x_2 - x_1).$$

This implies that $Z_{x_2} \subseteq Z_{x_1} + \mathcal{B}_Y(0, \sqrt{x_2(x_2 - x_1)})$. Therefore, Assumption A follows.

Let $\bar{\mu}$ be a solution of problem (7.3), $\bar{\lambda} = \nabla f(\int_Z g d\bar{\mu})$ and $Q^* = \int_Z q d\bar{\mu}(x, q)$. By the definitions of f and g , we obtain that $\bar{\lambda} = (1, \epsilon Q^*)$, moreover,

$$g_{\bar{\lambda}}(x, q) = \int_0^T e^{-rt} q_t (q_t - 1 + \epsilon Q_t^*) dt.$$

By Lemma 7.1, $\mathbf{BR}_{\bar{\lambda}}(x_0) = \{q^{Q^*}(x_0)\}$ for any $x_0 \in X$. By Corollary 3.5, we have that $(\bar{\lambda}, \bar{\mu})$ satisfies the following equilibrium equation:

$$\begin{cases} \bar{\lambda} = (1, \epsilon \int_Z q d\bar{\mu}) \\ \bar{\mu}_x = \delta_{q^{Q^*}(x)}, \quad m\text{-a.e.} \end{cases}$$

Combining with Theorem 2.7, we obtain that $\int_Z q d\bar{\mu} = \int_X q^{Q^*}(x) dm(x)$. Recall that $Q^* = \int_Z q d\bar{\mu}$, then (7.2) follows. \blacksquare

Numerical simulations. Let the initial measure m be an exponential distribution with parameter $a \geq 0$, i.e., $dm(x) = ae^{-ax}dx$ for all $x \geq 0$. Let us independently sample the distribution m for N times, denoting the samples by x_1, x_2, \dots, x_N , and $m_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$. The time space $[0, T]$ is discretized with a step size $\Delta t = T/M$ for some $M \geq 1$. Then, a totally discretized problem associated with (7.3) writes:

$$(7.4) \quad \begin{cases} \inf_{q \in \mathbb{R}^{N \otimes M}} & J_N(q) := \frac{\Delta t}{N} \sum_{i=1}^N \sum_{t=0}^{M-1} e^{-rt\Delta t} (q_{i,t}^2 - q_{i,t}) + \frac{\epsilon \Delta t}{2} \sum_{t=0}^{M-1} e^{-rt} \left(\frac{1}{N} \sum_{i=1}^N q_{i,t} \right)^2, \\ \text{such that} & q_i \in S^M(x_i) := \{q \in [0, 1/2]^M \mid \Delta t \sum_{t=0}^{M-1} q_t \leq x_i\}, \quad i = 1, 2, \dots, N. \end{cases}$$

We apply Algorithm 3 to solve (7.4). At each iteration, the evaluation of a best-response, for each producer i amounts to solve a problem of the following form:

$$(7.5) \quad \begin{cases} \inf_{q_i \in \mathbb{R}^M} & \Delta t \sum_{t=0}^{M-1} e^{-rt\Delta t} q_{i,t} (q_{i,t} - 1 + \epsilon Q_t), \\ \text{such that} & q_i \in S^M(x_i), \end{cases}$$

for a given $Q \in [0, 1/2]^M$. This problem is a convex quadratic programming problem in \mathbb{R}^M that can be dealt with by some solvers, such as GUROBI [19].

For the resolution of the problem, we chose the following parameters: $T = 10$, $\epsilon = r = a = 1$, $N = 100$, $M = 100$, $K = 100$, $n_k = 10$, for all k . Figure 1 shows the extracting speeds and the stocks of three producers with initial stocks: 0.9, 1.2, and 3.1. From Figure 1, we see that the producers with the higher initial stock have the same extracting speed as those with a lower initial stock, at the beginning. However, as the smaller agents exhaust their resource, the larger ones progressively raise their extraction speed. Once the extraction speed reaches its maximum value, it rapidly decreases to zero. These observations are consistent with the findings of [18, Sec. 3.3].

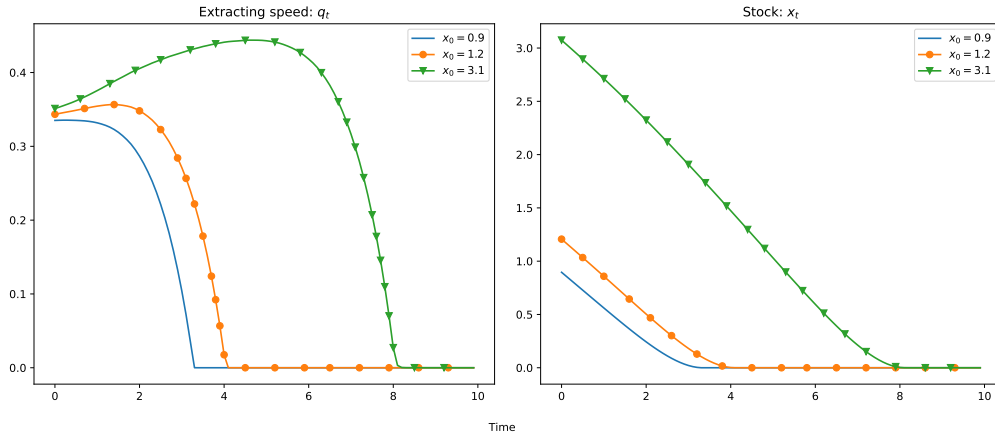


FIGURE 1. Extracting speeds and stocks of three producers with initial stocks: 0.9, 1.2, and 3.1.

To study the error caused by sampling, we independently sample the exponential distribution m for $100 * N$ times, and group them into batches of N . The empirical distribution corresponding to each batch is set as the initial distribution. Then we apply Algorithm 3 to compute Q^* corresponding to each initial distribution. In Figure 2, we show the mean and standard deviation of the results of the 100 simulations.

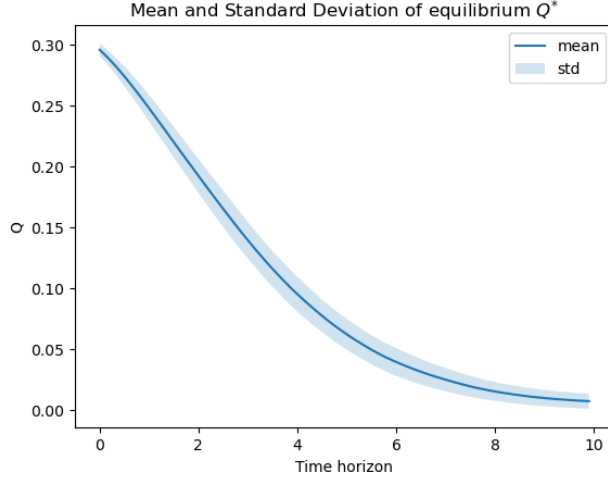


FIGURE 2. Mean and standard deviation of the equilibria of 100 batches

7.4. Numerical results for a congestion game.

Model. Consider a second numerical example within the context of the minimal-time deterministic MFG. We set the following parameters: the state space is fixed as $[0, 1]$, a maximum duration is denoted by $T > 0$, and an upper bound for the speed is given by $\bar{V} > 0$. In this particular example, the dynamics governing each player are characterized by the set Z , defined as:

$$Z = \{(x, \gamma) \in [0, 1] \times \text{AC}([0, T]) \mid \dot{\gamma}_t = v_t, \gamma_0 = x, 0 \leq v_t \leq \bar{V}\}.$$

The objective for the players in this example is to reach the target point 1 as soon as possible, while simultaneously ensuring that the density at each point does not become excessively high. To quantify this, we introduce the congestion function $\tilde{\mathcal{F}}: \mathcal{P}([0, 1]) \rightarrow \mathbb{R}$, which is defined as follows:

$$\tilde{\mathcal{F}}(m) = \begin{cases} \int_0^1 m(x)^2 dx, & \text{if } m \ll dx, \\ +\infty, & \text{otherwise.} \end{cases}$$

Given an initial distribution $m_0 \in \mathcal{P}([0, 1])$, the resulting deterministic MFG problem can be expressed as follows:

$$(7.6) \quad \inf_{\mu \in \mathcal{P}_{m_0}(Z)} J(\mu) := \int_Z \int_0^T \mathbb{I}_{[0,1]}(\gamma_t) dt d\mu(x, \gamma) + \alpha \int_0^T \tilde{\mathcal{F}}(e_t \# \pi_2 \# \mu) dt,$$

where $\alpha > 0$ is a penalty parameter.

Regularization. Note that the congestion function $\tilde{\mathcal{F}}$ does not fit to the framework studied in this article. To address this, we begin by approximating $\tilde{\mathcal{F}}$ with a function that aligns with our framework. We achieve this by partitioning the interval $[0, 1]$ into $J \in \mathbb{N}_+$ small, uniform subintervals: I_1, \dots, I_J , where $I_j = [(j-1)\Delta x, j\Delta x]$, and $\Delta x = 1/J$. Subsequently, we approximate $\tilde{\mathcal{F}}(m)$ as follows:

$$\tilde{\mathcal{F}}(m) \approx \Delta x \sum_{j=1}^J \left(\frac{1}{\Delta x} \int_0^1 \mathbb{I}_{I_j}(x) dm(x) \right)^2.$$

To facilitate the execution of the numerical experiments, we replace the indicator function \mathbb{I}_{I_j} by some smooth functions. Let $k \geq J$ be a positive integer. We introduce two smooth functions,

denoted as $\varphi_k \in \mathcal{C}^\infty(\mathbb{R})$ and $\phi_{k,\Delta x} \in \mathcal{C}^\infty(\mathbb{R})$. These functions are parametrized by the variables k and Δx , and are defined as follows:

$$\varphi_k(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ \frac{1}{1+e^{1/kx-1/(1-kx)}}, & \text{if } 0 < x < \frac{1}{k}, \\ 1 & \text{otherwise,} \end{cases}$$

$$\phi_{k,\Delta x}(x) = \begin{cases} 0, & \text{if } x \leq -\frac{1}{k}, \\ \varphi_k(x + 1/k), & \text{if } -\frac{1}{k} < x < 0, \\ 1 & \text{if } 0 \leq x \leq \Delta x - \frac{1}{k}, \\ 1 - \varphi_k(x - \Delta x + \frac{1}{k}), & \text{otherwise.} \end{cases}$$

Then, we approximate \mathbb{I}_{I_j} by $h_j: \mathbb{R}_+ \rightarrow [0, 1], x \mapsto \phi_{k,\Delta x}(x - (j-1)\Delta x)$ for $j = 1, \dots, J$, and $\mathbb{I}_{[0,1]}$ by $h_0: \mathbb{R}_+ \rightarrow [0, 1]$,

$$h_0(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1 - \frac{1}{k}, \\ 1 - \varphi_k(x - 1 + \frac{1}{k}), & \text{otherwise.} \end{cases}$$

An important property of $\{h_j\}_{j=1,\dots,J}$ is that $\sum_{j=1}^J h_j(x) = h_0(x)$, which is 1 for any $x \in [0, 1 - 1/k]$, see Figure 3. The resulting approximated MFO problem associated with (7.6) is,

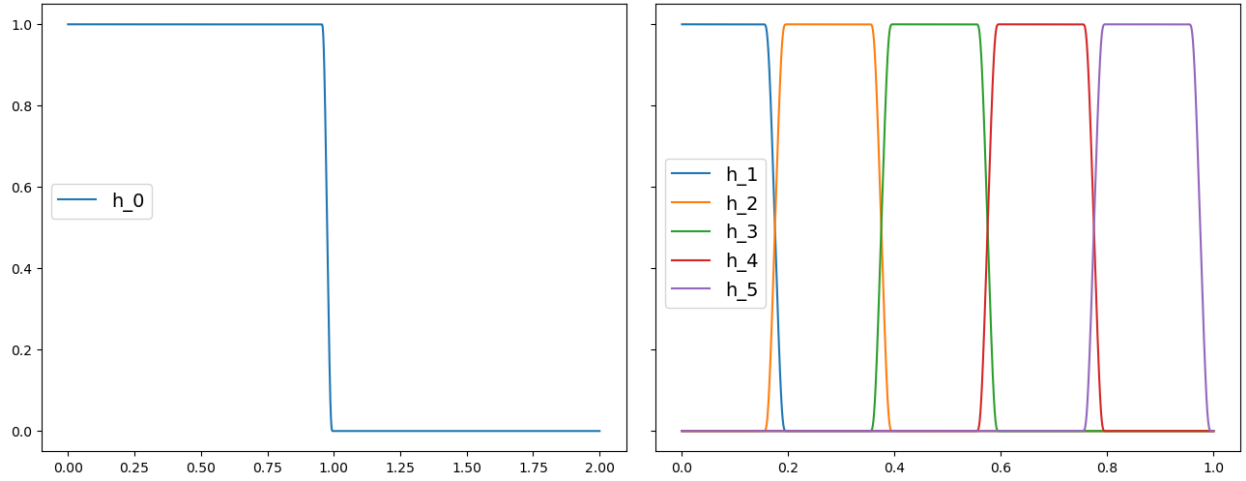


FIGURE 3. Example of h_0 and $\{h_j\}_{j=1,\dots,J}$ with $J = 5$ and $k = 20$.

$$(7.7) \quad \inf_{\mu \in \mathcal{P}_{m_0}(Z)} \int_Z \int_0^T h_0(\gamma_t) dt d\mu(x, \gamma) + \frac{\alpha}{\Delta x} \sum_{j=1}^J \int_0^T \left(\int_Z h_j(\gamma_t) d\mu(x, \gamma) \right)^2 dt.$$

Discretization and numerical results. To proceed with our numerical experiments, we discretize the time horizon $[0, T]$ into M steps, each of duration $\Delta t = T/M$. Additionally, we discretize the initial distribution m_0 by $m_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$. We formulate the fully discretized problem associated with (7.7) as follows:

$$(7.8) \quad \begin{cases} \inf_{(\gamma^i)_{i=1,\dots,N}} & \frac{\Delta t}{N} \sum_{i=1}^N \sum_{t=0}^{M-1} h_0(\gamma_t^i) + \frac{\alpha \Delta t}{\Delta x} \sum_{j=1}^J \sum_{t=0}^{M-1} \left(\frac{1}{N} \sum_{i=1}^N h_j(\gamma_t^i) \right)^2, \\ \text{such that} & 0 \leq \gamma_{t+1}^i - \gamma_t^i \leq \bar{V} \Delta t, \gamma_0^i = x_i, \text{ for } t = 0, \dots, M, \text{ and } i = 1, \dots, N. \end{cases}$$

Therefore, given some $(\bar{\gamma}^i)_{i=1,\dots,N}$ satisfying the constraint in (7.8), the sub-problem for player i is

$$(7.9) \quad \begin{cases} \inf_{\gamma^i} & \Delta t \sum_{t=0}^{M-1} h_0(\gamma_t^i) + \frac{2\alpha\Delta t}{\Delta x} \sum_{j=1}^J \sum_{t=0}^{M-1} \bar{y}_{j,t} h_j(\gamma_t^i), \\ \text{such that} & 0 \leq \gamma_{t+1}^i - \gamma_t^i \leq \bar{V}\Delta t, \gamma_0^i = x_i, \text{ for } t = 0, \dots, M, \end{cases}$$

where $\bar{y}_{j,t} = \frac{1}{N} \sum_{i=1}^N h_j(\bar{\gamma}_t^i)$. The sub-problem (7.9) is a finite-dimensional non-convex optimization problem, which is addressed by the open-source solver “scipy.optimize.minimize” [34].

Let us specify the parameters used in the numerical simulation of problem (7.7) as follows:

$$J = 5, k = 20, \bar{V} = 3, T = 1, \alpha = 1, M = 20, N = 200, m_0 = \text{Uni}_{[0,0.2]},$$

where “Uni” represents the uniform distribution, and the points x_i are drawn from samples of m_0 .

We first present in Figure 4 convergence results of Algorithm 3 for the discretized problem (7.8) in 100 iterations, utilizing parameter settings of $n_k = 1$ and 5. We see that in both choices, the algorithm converges to a local minimum very fast (fewer than 20 iterations). In Figure 5, we compare the optimal trajectories of γ^i for two cases: $\alpha = 0$ and $\alpha = 1$. It is evident that, in the case of $\alpha = 0$, the optimal strategy for each player is to move at the maximum speed, $\bar{V} = 3$, as there is no penalty for density. This is depicted in the left part of Figure 5. However, when $\alpha = 1$, players starting from greater initial positions choose to run at the maximum speed, whereas those with lower initial positions prefer to wait briefly to avoid congestion in density with those starting farther ahead. In Figure 6a, we draw the agents’ state distributions at each time for both $\alpha = 0$ and $\alpha = 1$, along with a regularized version obtained through interpolation in Figure 6b. For a more detailed view of the density evolution before reaching the target, we further depict the restricted distributions within the spatial interval $[0, 0.8]$ in Figure 6c, using a distinct color scale.

Remark 7.3. Let us underline that for this example, the optimization problems involved in the evaluation of the best-response mapping are non-convex. As mentioned above, we address them with the open-source solver `scipy.optimize.minimize` whose default method for tackling constrained non-linear optimization problems is the SLSQP (Sequential Least Squares Programming) algorithm, a quasi-Newton-type algorithm. Consequently, the quality of the initial guess plays a crucial role in the resolution of sub-problems. In the context of Algorithm 3, our experience shows that at iteration k , it is more efficient to initialise the evaluation of $\bar{y}_i^k \in \mathbf{BR}_{\lambda^k}(x_i)$ with y_i^k (rather than \bar{y}_i^{k-1}). We conjecture that the chance for the solver to generate a local solution is higher when initializing with \bar{y}_i^{k-1} .

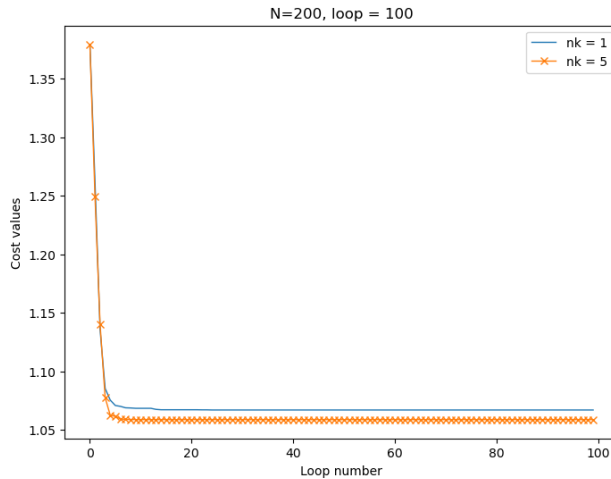


FIGURE 4. Convergence results of Algorithm 3.

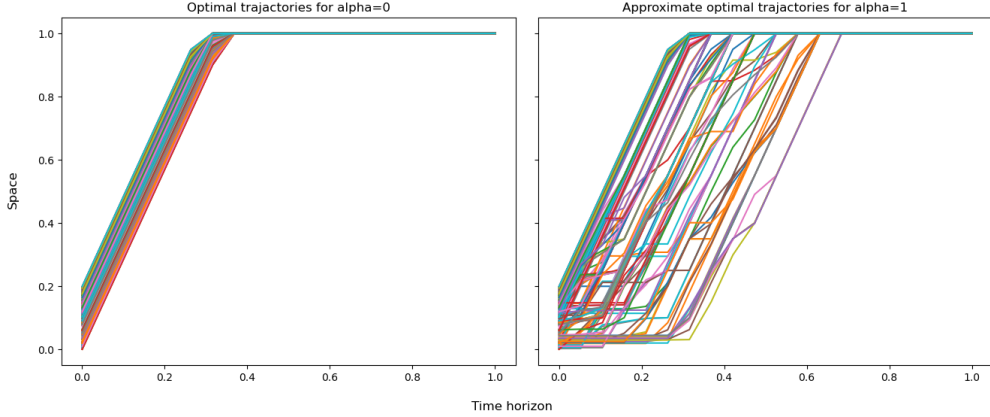


FIGURE 5. Optimal trajectory of each player: the case $\alpha = 0$ (left), the case $\alpha = 1$ (right)

8. CONCLUSION

We have provided a general framework for analyzing Mean Field Optimization problems. We have proposed a general method, based on an extension of the Frank-Wolfe algorithm for solving MFO problems, with a convergence guarantee, assuming that some best-response function can be efficiently computed (with a solver or with specific methods). Numerous extensions of the current setting could be considered. For example, one could formulate a stochastic setting with a random variable impacting all agents. In this setting the evaluation of λ (in the SFW algorithm) may require to use Monte-Carlo approximations, adding a new source of error in the general algorithm. One may also realize a general convergence analysis that would take into account the need to discretize the sets Z_x (in particular in the case of MFGs, where Z_x is an infinite dimension set). Finally, at a purely numerical level, we could investigate variants of the proposed method in which the distribution m is discretized progressively. This would reduce the number of subproblems to solve in the early iterations of the SFW algorithm. We also mention that the SFW is robust in the following sense: at the end of iteration k , if y^{k+1} is replaced by any other point yielding a reduction of the cost function, then the general convergence properties of the SFW algorithm are preserved. This fact could motivate the design of heuristic improvements on a case-by-case basis.

APPENDIX A. PROOF OF LEMMA 3.3

Before proving Lemma 3.3, let us recall the definitions of the restriction of a measure and the completion of a probability space, taken from [28, Thm. 1.36].

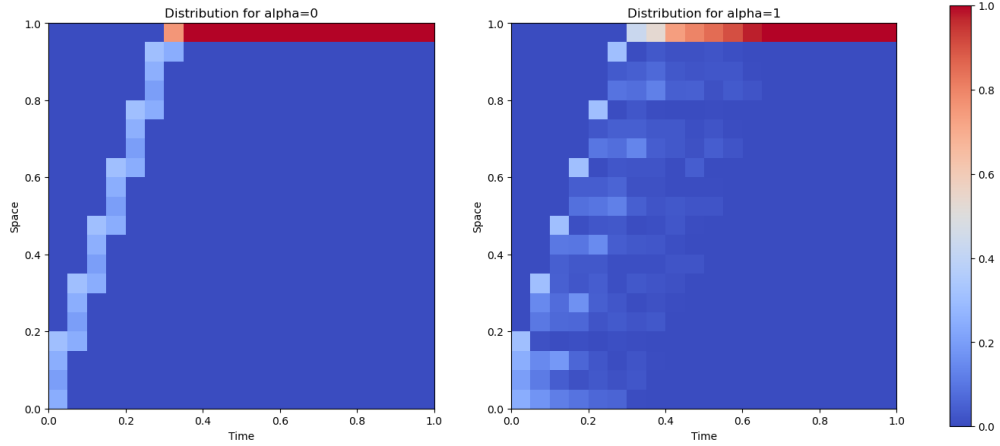
Definition A.1 (Restriction). Let X_1 be a Polish space, let \mathcal{X} and \mathcal{X}' be two σ -algebras on X_1 such that $\mathcal{X}' \subseteq \mathcal{X}$, and let ν be a measure on \mathcal{X} . The restriction measure of ν on \mathcal{X}' is defined as follows:

$$\nu|_{\mathcal{X}'}(A) := \nu(A), \quad \text{for any } A \in \mathcal{X}'.$$

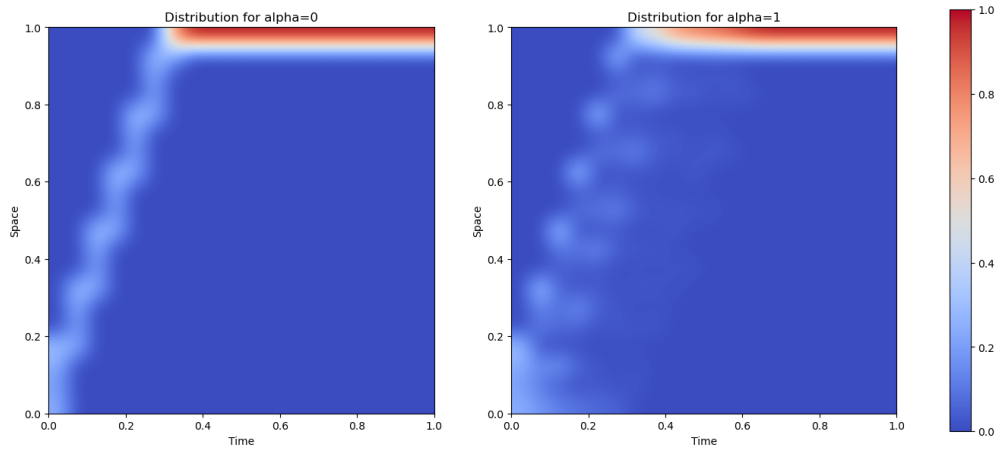
Definition A.2 (Completion). Let $(X_1, \mathcal{B}^{X_1}, \nu)$ be a probability space. Let \mathcal{B}_ν be the collection of all $E \subseteq X_1$ such that there exists A and B in \mathcal{B}^{X_1} , $A \subseteq E \subseteq B$, and $\nu(B - A) = 0$. For such an E , we define a function $\hat{\nu}(E)$ as

$$\hat{\nu}(E) = \nu(A).$$

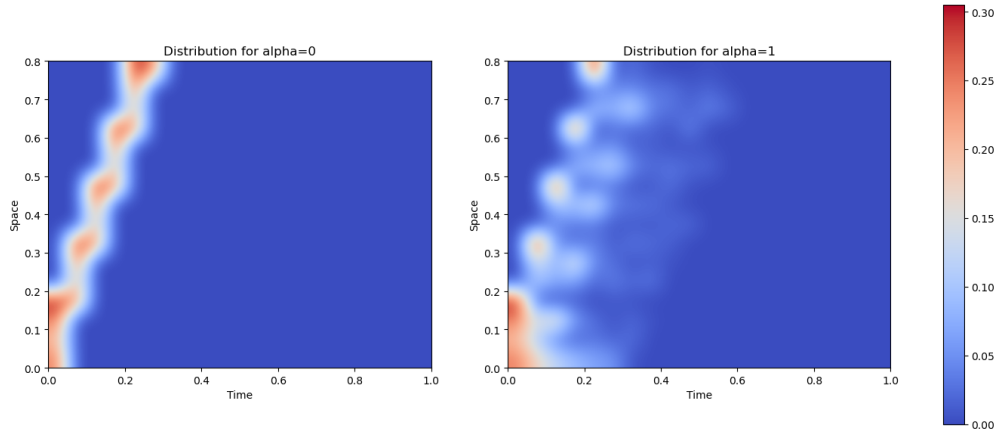
Then $(X_1, \mathcal{B}_\nu, \hat{\nu})$ is a complete measure space. We say that $(X_1, \mathcal{B}_\nu, \hat{\nu})$ is the completion of $(X_1, \mathcal{B}^{X_1}, \nu)$.



(A) Comparison of distributions of positions at each time: the case $\alpha = 0$ (left), the case $\alpha = 1$ (right).



(B) Comparison of regularized distributions of positions at each time: the case $\alpha = 0$ (left), the case $\alpha = 1$ (right).



(C) Comparison of regularized distributions of positions in horizon $[0, 0.8]$ at each time: the case $\alpha = 0$ (left), the case $\alpha = 1$ (right).

FIGURE 6. Distributions

Sketch of the proof of Lemma 3.3. The proof of the direction that the left-hand-side of (3.2) is greater than the right-hand-side is the same as the proof for the case that $m \in \mathcal{P}_\delta(X)$.

Let us prove the converse inequality. Let $(X, \mathcal{B}_m, \hat{m})$ be the completion of the probability space (X, \mathcal{B}^X, m) . Fix any $\lambda \in \mathcal{H}_f$. By Assumption A, the set-valued function $\mathbf{BR}_\lambda: X \rightsquigarrow Y$ has non-empty closed images. By Lemma 3.1, $\text{Graph}(\mathbf{BR}_\lambda)$ is closed in $X \times Y$, thus is a $\mathcal{B}_m \otimes \mathcal{B}^Y$ -measurable set. By Lemma 2.5 and Theorem 2.3, the set-valued function $\mathbf{BR}_\lambda: X \rightsquigarrow Y$ is $(\mathcal{B}_m, \mathcal{B}^Y)$ -measurable, and there exists a $(\mathcal{B}_m, \mathcal{B}^Y)$ -measurable function $\mathbf{br}_\lambda: X \rightarrow Y$ such that for any $x \in X$,

$$\mathbf{br}_\lambda(x) \in \mathbf{BR}_\lambda(x).$$

We define $\mathcal{A}: X \rightarrow Z$, $x \mapsto (x, \mathbf{br}_\lambda(x))$. Since \mathbf{br}_λ is $(\mathcal{B}_m, \mathcal{B}^Y)$ -measurable, we have that \mathcal{A} is $(\mathcal{B}_m, \mathcal{B}_m \otimes \mathcal{B}^Y)$ -measurable, see [21, Lem. 1.8]. Let \mathcal{B}^Z be the Borel σ -algebra on Z . It is obvious that $\mathcal{B}^Z \subseteq \mathcal{B}^X \otimes \mathcal{B}^Y \subseteq \mathcal{B}_m \otimes \mathcal{B}^Y$. Let us take

$$\tilde{\mu} = \mathcal{A} \# \hat{m}|_{\mathcal{B}^Z}.$$

Then $\tilde{\mu}$ is a positive Borel measure on Z . Moreover, we deduce from Definitions A.1-A.2 that

$$\tilde{\mu}(Z) = \mathcal{A} \# \hat{m}(Z) = \hat{m}(X) = m(X) = 1.$$

Therefore, $\tilde{\mu} \in \mathcal{P}(Z)$. Assume that the following two equalities hold true:

$$(A.1) \quad \pi_1 \# \tilde{\mu} = m,$$

$$(A.2) \quad \int_Z g_\lambda d\tilde{\mu} = \int_X g_\lambda \circ \mathcal{A} dm.$$

By the definitions of u_λ and \mathcal{A} ,

$$(A.3) \quad g_\lambda \circ \mathcal{A}(x) = g_\lambda(x, \mathbf{br}_\lambda(x)) = \inf_{y \in \mathbf{BR}_\lambda(x)} g_\lambda(x, y) = u_\lambda(x), \quad \forall x \in X.$$

Combining (A.1)-(A.3), we obtain that

$$\inf_{\mu \in \mathcal{P}_m(Z)} \int_Z g_\lambda d\mu \leq \int_Z g_\lambda d\tilde{\mu} = \int_X g_\lambda \circ \mathcal{A} dm = \int_X u_\lambda dm.$$

The conclusion follows. ■

For completing the proof of Lemma 3.3, it remains to prove equalities (A.1)-(A.2). They are deduced from Lemmas A.3-A.4:

- To prove (A.1), we take $\tilde{X} = X$ and $h = \pi_1$ in Lemma A.3;
- To prove (A.2), we take $\tilde{X} = [-M\|\lambda\|, +\infty)$ and $h = g_\lambda$ in Lemma A.3, and Lemma A.4 implies that $g_\lambda \circ \mathcal{A} = u_\lambda$ is Borel measurable.

Recall the definition of $\mathcal{A}: X \rightarrow Z$ in the previous proof and recall that $\tilde{\mu} = \mathcal{A} \# \hat{m}|_{\mathcal{B}^Z}$.

Lemma A.3. *Let \tilde{X} be a Polish space. Let $h: Z \rightarrow \tilde{X}$ be a Borel measurable function. Assume that $h \circ \mathcal{A}: X \rightarrow \tilde{X}$ is Borel measurable. Then $h \# \tilde{\mu} = (h \circ \mathcal{A}) \# m$. As a consequence, if $\tilde{X} = [c, +\infty)$ for some $c \in \mathbb{R}$, then*

$$\int_Z h d\tilde{\mu} = \int_X h \circ \mathcal{A} dm.$$

Proof. Let B be any Borel set in \tilde{X} . By the property of push-forward measure, $h \# \tilde{\mu}(B) = \tilde{\mu}(h^{-1}(B))$. Since h is Borel measurable, $h^{-1}(B) \in \mathcal{B}^Z$. Thus $h \# \tilde{\mu}(B) = \mathcal{A} \# \hat{m}(h^{-1}(B))$. Next, by the property of the push-forward measure,

$$\mathcal{A} \# \hat{m}(h^{-1}(B)) = \hat{m}(\mathcal{A}^{-1}h^{-1}(B)) = \hat{m}((h \circ \mathcal{A})^{-1}(B)).$$

Since $h \circ \mathcal{A}$ is Borel measurable, we have that $(h \circ \mathcal{A})^{-1}(B) \in \mathcal{B}^X$. As a consequence,

$$\hat{m}((h \circ \mathcal{A})^{-1}(B)) = m((h \circ \mathcal{A})^{-1}(B)) = (h \circ \mathcal{A}) \# m(B).$$

Therefore, $h\#\tilde{\mu}(B) = (h \circ \mathcal{A})\#m(B)$ for any Borel set $B \subseteq \tilde{X}$. This concludes the first part of the proof. In the case where $X_3 = [c, +\infty)$ for some $c \in \mathbb{R}$, since $c = \int_{X_2} c d\tilde{\mu} = \int_{X_1} c \circ \mathcal{A} dm$, it suffices to prove the conclusion for $h - c$ instead of h . Therefore, we can assume that $X_3 = \mathbb{R}_+$. By the change-of-variable formula for push-forward measures,

$$\int_Z h d\tilde{\mu} = \int_{\mathbb{R}_+} x d(h\#\tilde{\mu}(x)).$$

Next, it follows from the equality $h\#\tilde{\mu} = (h \circ \mathcal{A})\#m$ that

$$\int_{\mathbb{R}_+} x d(h\#\tilde{\mu}(x)) = \int_{\mathbb{R}_+} x d((h \circ \mathcal{A})\#m(x)).$$

Again, by the change-of-variable formula, we obtain that

$$\int_{\mathbb{R}_+} x d((h \circ \mathcal{A})\#m(x)) = \int_X h \circ \mathcal{A} dm.$$

The conclusion follows. ■

Lemma A.4. *Under Assumption A, the function u_λ is upper semi-continuous for any $\lambda \in \mathcal{H}_f$, thus Borel measurable.*

Proof. Let $\lambda \in \mathcal{H}_f$. Since g is bounded over Z , we have that $u_\lambda(x) > -\infty$ for any $x \in X$. Fix any $x \in X$. Let $y \in \mathbf{BR}_\lambda(x)$. Let $(x_n \in X)_{n \geq 1}$ be a sequence converging to x . By the lower semi-continuity of G_λ , there exists $y_n \in Z_{x_n}$ such that $g_\lambda(x, y) = \lim_{n \rightarrow \infty} g_\lambda(x_n, y_n)$. Therefore,

$$u_\lambda(x) = g_\lambda(x, y) = \lim_{n \rightarrow \infty} g_\lambda(x_n, y_n) \geq \limsup_{n \rightarrow \infty} u_\lambda(x_n).$$

We obtain the upper semi-continuity of u_λ for any $\lambda \in \mathcal{H}_f$. Since any upper semi-continuous function defined on a metric space is the limit of a monotonically decreasing sequence of continuous functions [32, Thm. 3], we deduce that u_λ is Borel measurable. ■

REFERENCES

- [1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.
- [2] J.-P. Aubin and Hélène Frankowska. *Set-valued analysis*. Springer Science & Business Media, 2009.
- [3] H.H. Bauschke and P.L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.
- [4] J.-D. Benamou, G. Carlier, and F. Santambrogio. Variational mean field games. *Active Particles, Volume 1: Advances in Theory, Models, and Applications*, pages 141–171, 2017.
- [5] J.F. Bonnans, J. Gianatti, and L. Pfeiffer. A lagrangian approach for aggregative mean field games of controls with mixed and final constraints. *SIAM Journal on Control and Optimization*, 61(1):105–134, 2023.
- [6] J.F. Bonnans, K. Liu, N. Oudjane, L. Pfeiffer, and C. Wan. Large-scale nonconvex optimization: randomization, gap estimation, and numerical resolution. *arXiv preprint*, 2022.
- [7] N. Boyd, G. Schiebinger, and B. Recht. The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization*, 27(2):616–639, 2017.
- [8] H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 2. Springer, 2011.
- [9] P. Cannarsa and R. Capuani. Existence and uniqueness for mean field games with state constraints. *PDE models for multi-agent phenomena*, pages 49–71, 2018.
- [10] P. Cardaliaguet and S. Hadikhannoo. Learning in mean field games: the fictitious play. *ESAIM: Control, Optimisation and Calculus of Variations*, 23(2):569–591, 2017.
- [11] C. Castaing and M. Valadier. *Convex analysis and measurable multifunctions*, volume 580. Springer, 2006.
- [12] F. Chen, Z. Ren, and S. Wang. Entropic fictitious play for mean field optimization problem. *Journal of Machine Learning Research*, 24(211):1–36, 2023.
- [13] L. Chizat and F. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. *Advances in neural information processing systems*, 31, 2018.
- [14] D.L. Cohn. *Measure theory*, volume 1. Springer, 2013.
- [15] K. Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.

- [16] M. Geist, J. Pérolat, M. Laurière, R. Elie, S. Perrin, O. Bachem, R. Munos, and O. Pietquin. Concave utility reinforcement learning: The mean-field game viewpoint. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pages 489–497, 2022.
- [17] A. Gersho and R.M. Gray. *Vector quantization and signal compression*, volume 159. Springer Science & Business Media, 2012.
- [18] P. Graewe, U. Horst, and R. Sircar. A maximum principle approach to a deterministic mean field game of control with absorption. *SIAM Journal on Control and Optimization*, 60(5):3173–3190, 2022.
- [19] LLC Gurobi Optimization. Gurobi optimizer reference manual, 2018.
- [20] S. Hadikhanloo and F.J. Silva. Finite mean field games: fictitious play and convergence to a first order continuous mean field game. *Journal de Mathématiques Pures et Appliquées*, 132:369–397, 2019.
- [21] O. Kallenberg. *Foundations of modern probability*, volume 2. Springer, 1997.
- [22] P. Lavigne and L. Pfeiffer. Generalized conditional gradient and learning in potential mean field games. *Applied Mathematics and Optimization*, 88, 2023.
- [23] K. Liu and L. Pfeiffer. A mesh-independent method for second-order potential mean field games. *ArXiv preprint*, 2023.
- [24] S. Mei, A. Montanari, and P.-M. Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33):E7665–E7671, 2018.
- [25] Q. Mérigot and J.-M. Mirebeau. Minimal geodesics along volume-preserving maps, through semidiscrete optimal transport. *SIAM Journal on Numerical Analysis*, 54(6):3465–3492, 2016.
- [26] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
- [27] R.T. Rockafellar. *Convex analysis*, volume 11. Princeton university press, 1997.
- [28] W. Rudin. *Real and complex analysis*, volume 3. McGraw-Hill, 1987.
- [29] F. Santambrogio and W. Shim. A Cucker–Smale inspired deterministic mean field game with velocity interactions. *SIAM Journal on Control and Optimization*, 59(6):4155–4187, 2021.
- [30] C. Sarrazin. Lagrangian discretization of variational mean field games. *SIAM Journal on Control and Optimization*, 60(3):1365–1392, 2022.
- [31] Yosef Sheffi. *Urban transportation networks*, volume 6. Prentice-Hall, Englewood Cliffs, NJ, 1985.
- [32] H. Tong. Some characterizations of normal and perfectly normal spaces. *Duke Math. J.*, pages 289–292, 1952.
- [33] C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [34] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.
- [35] M. Wang. Vanishing price of decentralization in large coordinative nonconvex optimization. *SIAM Journal on Optimization*, 27(3):1977–2009, 2017.