

Dianetti, Jodi; Dumitrescu, Roxana; Ferrari, Giorgio; Xu, Renyuan

Working Paper

Entropy regularization in mean-field games of optimal stopping

Center for Mathematical Economics Working Papers, No. 755

Provided in Cooperation with:

Center for Mathematical Economics (IMW), Bielefeld University

Suggested Citation: Dianetti, Jodi; Dumitrescu, Roxana; Ferrari, Giorgio; Xu, Renyuan (2025) : Entropy regularization in mean-field games of optimal stopping, Center for Mathematical Economics Working Papers, No. 755, Bielefeld University, Center for Mathematical Economics (IMW), Bielefeld, <https://nbn-resolving.de/urn:nbn:de:0070-pub-30076673>

This Version is available at:

<https://hdl.handle.net/10419/333540>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

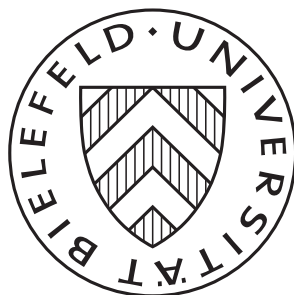
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Entropy Regularization in Mean-Field Games of Optimal Stopping

Jodi Dianetti, Roxana Dumitrescu, Giorgio Ferrari and Renyuan Xu



ENTROPY REGULARIZATION IN MEAN-FIELD GAMES OF OPTIMAL STOPPING

JODI DIANETTI, ROXANA DUMITRESCU, GIORGIO FERRARI, AND RENYUAN XU

ABSTRACT. We study mean-field games of optimal stopping (OS-MFGs) and introduce an entropy-regularized framework to enable learning-based solution methods. By utilizing randomized stopping times, we reformulate the OS-MFG as a mean-field game of singular stochastic controls (SC-MFG) with entropy regularization. We establish the existence of equilibria and prove their stability as the entropy parameter vanishes. Fictitious play algorithms tailored for the regularized setting are introduced, and we show their convergence under both Lasry–Lions monotonicity and supermodular assumptions on the reward functional. Our work lays the theoretical foundation for model-free learning approaches to OS-MFGs.

Keywords: Mean-field game of optimal stopping, singular stochastic control, entropy regularization, randomized stopping times, fictitious play algorithm.

AMS subject classification: 91A16, 60G40, 93E20, 68T01

1. INTRODUCTION

The study of mean-field games (MFGs) has become central to the analysis of large-population stochastic control systems, where individual agents interact through the empirical distribution of states and/or controls. For methodologies, techniques, and applications, we refer to the two-volume book [12]. In this paper, we focus on a specific class of such problems: mean-field games of optimal stopping (OS-MFGs). Optimal stopping has a wide range of applications in Economics, Finance, Operations Research, and other applied fields. Relevant examples include pricing American derivatives, entry-exit problems (real options models), and optimal timing for buying or selling an asset, among others. On the other hand, OS-MFGs have received attention only very recently. Among the fundamental contributions on this topic (without learning considered), we first mention [52], where the authors consider a specific game in which the interaction occurs through the number of players who have already stopped. In [13], the authors study an optimal stopping game with Brownian common noise, inspired by a model of bank runs. They prove the strong existence of mean-field equilibria in a setting with strategic complementarity, and weak existence in a setting with continuity. In [7], a purely analytical approach is adopted to solve the problem – namely, the study of the system of coupled Hamilton-Jacobi-Bellman and Fokker-Planck equations – and the existence of mixed solutions is proved. More recently, a compactification approach based on the linear programming formulation is developed in a series of papers [8], [33] and [34], in which the existence of relaxed solutions, interpreted via occupational measures, is established. These techniques have been then applied to solve entry-exit optimal stopping games in electricity markets in [1] and [5], and in the case with common noise in [35]. In [32], a rigorous connection between the occupation measures arising in the linear programming approach and randomized stopping is established. In

Date: October 20, 2025.

J. Dianetti: Department of Economics and Finance, University of Rome Tor Vergata. Email: jodi.dianetti@uniroma2.it.

R. Dumitrescu: ENSAE-CREST, Institut Polytechnique de Paris. Email: roxana.dumitrescu@ensae.fr.

G. Ferrari: Center for Mathematical Economics, Bielefeld University. Email: giorgio.ferrari@uni-bielefeld.de.

R. Xu: Department of Finance and Risk Engineering, NYU. Email: rx2364@nyu.edu.

addition, [53] studies an optimal stopping game through the lens of the master equation. Another recent contribution is [46], where the authors develop a new approach to solving various types of MFGs (including optimal stopping) based on a mean-field version of the Bank-El Karoui representation theorem for stochastic processes. In the context of optimal stopping MFGs with partial information, [9, 10, 59] study the existence of the mean-field solution for a one-dimensional filtering problem under different model set-ups. Finally, related works on McKean-Vlasov optimal stopping are [61] and [21].

Despite their importance, OS-MFGs pose significant theoretical and computational challenges. First, they require more subtle techniques to be solved than MFGs with standard control because of the irregularity of the flow measure generated by the simultaneous exit of a significant number of players. This class of games becomes particularly challenging when the goal is to develop model-free, learning-based solution methods.

The primary motivation of our work is to establish the foundations of a suitable OS-MFG framework that supports model-free solution methods. A central challenge in this direction is to rigorously formalize and analyze randomized and explorative strategies in games, which play a crucial role in the design of reinforcement learning (RL) algorithms. Establishing the existence and uniqueness of (mean-field) equilibria that accommodate such strategies is therefore essential—not only for theoretical completeness but also for enabling practical learning algorithms with provable guarantees. In the absence of such a foundation, rigorous analysis of multi-agent learning remains elusive. Moreover, the convergence of RL methods often depends on the properties of the underlying iterative schemes (e.g., fictitious play), whose well-posedness similarly hinges on the existence and uniqueness of equilibria in games that allow for randomized behavior (see, e.g., [40, 41]). This need becomes even more pronounced in environments characterized by irregular or sparse decision-making, such as optimal stopping problems, where agents make a single, irreversible decision based on limited information. Unlike standard Markov Decision Processes (MDPs) [54, 60], where frequent actions and rewards can reduce the reliance on explicit exploration, optimal stopping games require more deliberate and carefully designed exploration mechanisms to facilitate information acquisition [26]. On the technical side, however, a major challenge arises in applying RL to OS-MFGs. Standard equilibrium-finding algorithms in MFGs typically rely on the stability of optimal controls with respect to variations in the distribution of agents. This stability, however, is particularly difficult to ensure in the context of optimal stopping, where understanding how equilibrium stopping policies respond to small perturbations in the measure flow is highly nontrivial. To illustrate this, consider that in a Markovian setting, equilibrium stopping rules are often given by hitting times of the underlying state process at a free boundary. Determining the stability properties (e.g., Lipschitz continuity) of these free boundaries is already a technically intricate problem, even in single-agent optimal stopping settings (see, e.g., [23] and the discussion therein).

Our contributions. To address these challenges, we develop a new theoretical framework that enables the use of model-free algorithms for OS-MFGs. The core idea is to reformulate the game as a MFG of singular stochastic controls (SC-MFG) by introducing randomized stopping times (see [30, 62]). A randomized stopping time can be interpreted as the (conditional) probability of stopping before a certain time t . This interpretation leads naturally to an adapted, nonnegative, nondecreasing process with right-continuous sample paths, bounded by 1 – an object that, following terminology from control theory, we refer to as a singular control (see Chapter VIII in [37] for an introduction to singular stochastic controls). To promote exploration, we introduce an entropy regularization term into the stopping functional, as previously established in the single-agent setting of [30]. This regularization term is governed by a temperature parameter $\lambda \geq 0$ and induces strict concavity in the performance criterion. This change enables optimizers to deviate from pure 0-1 strategies (of not stopping or stopping) and instead favor reflecting-type controls. In the context of MFG, the equilibrium is characterized via a two-step procedure. First, given a fixed flow of measures

and a joint distribution, we determine the singular control that maximizes the regularized reward functional. Second, we impose the so-called consistency condition, which requires that the fixed input measures coincide with the distribution of the state process before stopping and the joint distribution of the (randomized) stopping time and the stopped state. Naturally, in this framework, the consistency conditions must be reformulated in terms of randomized stopping times – namely, in terms of the random Borel measure on the time axis induced by the singular control. The particular structure of these consistency conditions prevents us from applying existing existence results for mean-field equilibria in singular control games (see, e.g., [11, 25, 38, 39]). We therefore establish a novel existence result for equilibria in our SC-MFG framework, which we believe is of independent theoretical interest (see Theorem 3.6).

Having proved the existence of equilibria in randomized stopping times, we then move on to establishing λ -stability as the temperature parameter $\lambda \downarrow 0$. We show that any equilibrium of the SC-MFG approximates, as $\lambda \downarrow 0$ (up to a subsequence), an equilibrium of the original MFG of optimal stopping without entropy regularization. Furthermore, under a Lasry-Lions monotonicity condition, we can prove a stronger result: the entire sequence of mean-field equilibria of the singular control game (parametrized by λ) converges to an equilibrium of the original optimal stopping game (see Theorem 3.10). This result ensures that the solutions of the regularized problems converge to equilibria of the original OS-MFG, providing a crucial theoretical justification for using entropy-based approximations in RL methods.

We then design novel fictitious-play algorithms tailored to the entropy regularized SC-MFG framework. We establish their convergence under both the Lasry–Lions monotonicity condition (see Theorem 4.5) and in the case of supermodular reward structures (see Theorem 4.9). It is important to note that the presence of the entropy regularization term introduces nonlinearity into the performance criterion, placing our model outside the scope of recent approaches based on linear programming for MFGs [8, 33, 34]. This nonlinearity necessitates the development of new analytical and numerical tools. The fictitious-play algorithm under the Lasry–Lions condition therefore extends previous algorithms developed under this monotonicity assumption to cases where the payoff functional is non-linear with respect to the control. Under the supermodularity condition, the fictitious-play algorithm is new in the literature. The theoretical foundation presented here will be complemented by a companion paper focused on the algorithmic design, related theoretical convergence analysis, and numerical validation of the proposed framework, which is currently in progress [28].

Related literature. Reinforcement learning applied to optimal stopping problems is closely tied to the broader challenge of learning in environments with sparse rewards [26, 45]. In these settings, a reward is only granted at the stopping time, resulting in extreme sparsity and significant learning challenges compared to more conventional control tasks, whether formulated in continuous time or as classical MDPs.

When the model is fully known, recent studies such as [56] and [58] have introduced deep learning techniques to approximate optimal stopping boundaries. In a related vein, [6] investigated randomized formulations of optimal stopping and provided convergence results for both forward and backward Monte Carlo-based optimization schemes. Similarly, [3] showcased the effectiveness of deep learning methods for solving high-dimensional singular control problems.

In the realm of continuous-time RL, a notable advancement comes from [24], who proposed a general policy gradient framework applicable to a wide array of stochastic control problems, including optimal stopping, impulse control, and switching. Their approach leverages connections between stochastic control and its randomized counterparts. However, their framework currently lacks theoretical guarantees for convergence. The work [31] studies a regularized version of the one-dimensional American Put option under a Shannon entropy framework. By employing an intensity-based control formulation, they introduce exploration and prove convergence of the policy iteration algorithm (PIA) for fixed temperature parameters. Nevertheless, the convergence of the optimal

policy as the temperature parameter $\lambda \downarrow 0$ remains an open question. An alternative perspective was recently offered by [22], who reformulated the optimal stopping problem as a regular control problem with binary (0 or 1) actions and applied entropy regularization. This transformed the problem into a classical entropy-regularized control setting, enabling the use of standard RL algorithms. The very recent paper [51] proposes a continuous-time reinforcement learning framework for a class of singular stochastic control problems without entropy regularization and generalizes the existing policy evaluation theories with regular controls to learn optimal singular control law and develop a policy improvement theorem.

Another important research direction leverages non-parametric statistical methods for stochastic processes to design learning-based control algorithms [16, 17, 18, 19]. In particular, [17] and [18] addressed one-dimensional singular and impulse control problems by learning critical thresholds using a non-parametric diffusion framework. This approach was extended to multivariate reflection problems in [19], while [16] contributed theoretical guarantees, including regret bounds and non-asymptotic PAC estimates.

Entropy regularized MFGs and their reinforcement learning (RL) formulations have received increasing attention over the past year. We do not attempt to provide an exhaustive literature review here, but instead highlight a few key references: [15], which presents an RL approach to model-free MFGs for MDPs; [2], which discusses Q-learning in regularized MFGs for MDPs; [14], which provides a convergence analysis of machine learning algorithms for mean-field control and games over a finite time horizon in continuous time and space; and [36, 42], which addresses entropy regularized MFGs in diffusion settings.

As far as it concerns the more specific class of entropy-regularized MFGs of optimal stopping – as those considered in this paper – the only paper brought to our attention is [65]. This paper investigates a discrete-time major-minor MFG where the major player can choose to either control or stop. The goal is to find a relaxed (randomized) stopping equilibrium, formulated as a fixed point of a set-valued map, whose direct analysis is difficult due to the major player’s influence. To address this, the authors introduce entropy regularization for the major player’s problem and reformulate the minor players’ stopping problems using linear programming over occupation measures, in the spirit of [8]. They prove the existence of regularized equilibria via a fixed-point theorem and show that, as regularization vanishes, these equilibria converge to a solution of the original problem, thus establishing the existence of a relaxed equilibrium. Although there are clear connections between [65] and the present paper, the mathematical analysis is entirely different, as the problem in [65] involves a Stackelberg game (which is not present in our case) on top of the mean-field interaction, within a discrete-time and discrete-space setting – unlike our continuous-time and continuous-space framework.

Organization of the paper. The rest of the paper is organized as follows. In Section 2 we present the OS-MFG problem and introduce its entropy-regularized formulation through randomized stopping times. In Section 3 we prove existence and stability of equilibria, while in Section 4 we present the fictitious play algorithms and their convergence. Finally, Appendix A collects results on a connection between the studied MFG of singular control and (another) auxiliary MFG of optimal stopping.

1.1. General notation. Let $p \geq 1$ and (E, d) a non-empty Polish space. Define the set $\mathcal{P}_p^{sub}(E)$ of subprobability measures ν on the Borel subsets of E such that $\int_E d(z, x_0)^p \nu(dz) < \infty$, for some (and thus all) $x_0 \in E$. The space $\mathcal{P}_p(E)$ represents the set of elements of $\mathcal{P}_p^{sub}(E)$ which are probability measures, and it is endowed with the Wasserstein distance d_p defined as

$$(1.1) \quad d_p(\bar{\nu}, \nu) := \inf_{\pi \in \Pi(\bar{\nu}, \nu)} \int_{E \times E} d(\bar{z}, z)^p \pi(d\bar{z}, dz),$$

where $\Pi(\bar{\nu}, \nu)$ is the set of $\pi \in \mathcal{P}_p(E \times E)$ which has $\bar{\nu}$ as first marginal and ν as second marginal. Following [20, Section B], to define the Wasserstein distance d'_p on $\mathcal{P}_p^{sub}(E)$, we introduce a cemetery point ∂ and obtain the enlarged space $\bar{E} := E \cup \partial$. By defining $d(x, \partial) := d(x, x_0) + 1$, we extend the definition of d on \bar{E} , in such a way that (\bar{E}, d) is still Polish. We define the classical Wasserstein distance d_p on \bar{E} using the same expression (1.1). The Wasserstein type distance d'_p on $\mathcal{P}_p^{sub}(\bar{E})$ is given by

$$d'_p(\mu, \nu) := d_p(\bar{\mu}, \bar{\nu}),$$

where

$$\bar{\mu}(\cdot) := \mu(\cdot \cap E) + (1 - \mu(E))\delta_\partial, \quad \bar{\nu}(\cdot) := \nu(\cdot \cap E) + (1 - \nu(E))\delta_\partial.$$

Fix $T > 0$. We introduce the set $M_p(E)$ of measurable functions $m : [0, T] \rightarrow \mathcal{P}_p^{sub}(E)$ identified dt -a.e. on $[0, T]$, endowed with the convergence in measure which is induced by the metric

$$(1.2) \quad d_p^M(m, \nu) := \int_0^T 1 \wedge d'_p(m_t, \nu_t) dt, \quad m, \nu \in M_p(E).$$

A sequence $(m^n)_n$ converges to m ($m^n \rightarrow m$, in short) if for any $\varepsilon > 0$ we have

$$\int_0^T \mathbb{1}_{\{t : d'_p(m_t^n, m_t) > \varepsilon\}} dt \rightarrow 0.$$

Recall that, for any convergent sequence $m^n \rightarrow m$, there exists a subsequence $(m^{n_k})_k$ such that

$$(1.3) \quad d'_p(m_t^n, m_t) \rightarrow 0, \quad dt\text{-a.e.}$$

We refer to Appendix B in [34] for further details.

Throughout the paper, $C > 0$ will denote a generic constant that may change from line to line.

2. MFGs OF OPTIMAL STOPPING AND ENTROPY REGULARIZATION

This section formally introduces the OS-MFG problem with randomization and entropy regularization, leading to a singular control formulation. More specifically, Section 2.1 introduces preliminaries and model assumptions; Section 2.2 describes the OS-MFG problem; and Section 2.3 develops the framework for randomized stopping times, entropy regularization, and the resulting singular control problem.

2.1. Preliminaries and assumptions. Let $T \in (0, \infty)$ be given and fixed throughout the rest of the paper. For $d, d_1 \in \mathbb{N} \setminus \{0\}$, consider the continuous functions $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d_1}$, $f : [0, T] \times \mathbb{R}^d \times \mathcal{P}^{sub}(\mathbb{R}^d) \rightarrow \mathbb{R}$, $g : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$.

Define

$$M_p := M_p(\mathbb{R}^d) \text{ and } \mathcal{M} := M_p \times \mathcal{P}_p([0, T] \times \mathbb{R}^d),$$

endowed with the product topology. Elements of \mathcal{M} will typically be denoted by couples (m, μ) . Moreover, given a sequence $(m^n, \mu^n)_n \subset \mathcal{M}$ and a limit point $(m, \mu) \in \mathcal{M}$, we will write

$$(m^n, \mu^n)_n \rightarrow (m, \mu)$$

to denote the convergence in the product topology.

On a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider a d_1 -dimensional Brownian motion $W := (W_t)_t$ and a \mathbb{R}^d -valued square integrable \mathcal{F}_0 -measurable random variable x_0 independent of W . Denote by $\mathbb{F}^{x_0, W} := (\mathcal{F}_t^{x_0, W})_t$ the right-continuous extension of the filtration generated by W and x_0 . For a time-horizon $T \in (0, \infty)$, let \mathcal{T} be the set of $\mathbb{F}^{x_0, W}$ -stopping times τ such that $\tau \leq T$ a.s..

Assumption 2.1. *The data of the problem (x_0, b, σ, f, g) verify:*

$$(1) \quad \mathbb{E}[|x_0|^p] < \infty.$$

(2) There exists a constant $L > 0$ such that

$$\begin{aligned} |b(t, x)| + |\sigma(t, x)| &\leq L(1 + |x|), \quad \forall t \in [0, T], x \in \mathbb{R}^d, \\ |b(t, \bar{x}) - b(t, x)| + |\sigma(t, \bar{x}) - \sigma(t, x)| &\leq L|\bar{x} - x|, \quad \forall t \in [0, T], x, \bar{x} \in \mathbb{R}^d. \end{aligned}$$

(3) There exists a constant $K > 0$ such that

$$\begin{aligned} |f(t, x, m)| &\leq K \left(1 + |x|^p + \int_{\mathbb{R}^d} |z|^p m(dz) \right), \quad \forall (t, x, m) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_p^{sub}(\mathbb{R}^d), \\ |g(x, \mu)| &\leq K \left(1 + |x|^p + \int_{[0, T] \times \mathbb{R}^d} |z|^p \mu(ds, dz) \right), \quad \forall (x, \mu) \in \mathbb{R}^d \times \mathcal{P}_p([0, T] \times \mathbb{R}^d). \end{aligned}$$

(4) The functions f and g are continuous, and g is continuous in μ , locally uniformly with respect to x . That is, there exist a constant $K > 0$ and a function $w_g : \mathcal{P}_p([0, T] \times \mathbb{R}^d) \times \mathcal{P}_p([0, T] \times \mathbb{R}^d) \rightarrow [0, \infty)$ with $\lim_{d_p(\mu, \bar{\mu}) \rightarrow 0} w_g(\bar{\mu}, \mu) = 0$ such that $|g(x, \bar{\mu}) - g(x, \mu)| \leq K(1 + |x|^p)w_g(\bar{\mu}, \mu)$ for any $x \in \mathbb{R}^d, \mu, \bar{\mu} \in \mathcal{P}_p([0, T] \times \mathbb{R}^d)$.

2.2. The MFG of optimal stopping. Consider the mean-field game of optimal stopping (OS-MFG, in short) in which, for a given behavior $(m, \mu) \in \mathcal{M}$ of the population of players, the representative player maximizes, over the choice of stopping times $\tau \in \mathcal{T}$, the profit functional

$$(2.1) \quad \begin{aligned} J(\tau, m, \mu) &:= \mathbb{E} \left[\int_0^\tau f(t, X_t, m_t) dt + g(X_\tau, \mu) \right], \\ \text{subject to } dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad t \in [0, T], \quad X_0 = x_0. \end{aligned}$$

Thanks to Assumption 2.1, there exists a unique strong solution to the stochastic differential equation (SDE) for X with continuous paths a.s., and such that

$$(2.2) \quad \mathbb{E} \left[\sup_{t \in [0, T]} |X_t|^p \right] < \infty.$$

Moreover, the profit functional $J(\tau, m, \mu)$ is well defined for any $\tau \in \mathcal{T}$, the maximization problem has finite value (i.e., $\sup_\tau J(\tau, m, \mu) < \infty$), and there exists an optimal stopping (i.e., $\tau^* \in \arg \max_\tau J(\tau, m, \mu)$). We refer the interested reader to Appendix D in [49] for further details.

We consider the following notion of equilibrium.

Definition 2.2 (OS-MFG equilibrium with strict stopping). *An optimal stopping MFG equilibrium (with strict stopping) is a triple $(\hat{\tau}, \hat{m}, \hat{\mu}) \in \mathcal{T} \times \mathcal{M}$ such that*

$$(2.3) \quad \begin{cases} \hat{\tau} \in \arg \max_{\tau \in \mathcal{T}} J(\tau, \hat{m}, \hat{\mu}), \\ \hat{m}_t(A) = \mathbb{P}(X_t \in A, t \leq \hat{\tau}), \quad \forall A \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T], \\ \hat{\mu}(B \times [0, t]) = \mathbb{P}(X_{\hat{\tau}} \in B, \hat{\tau} \leq t), \quad \forall B \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T]. \end{cases}$$

The first condition is the optimality condition, while the second and the third condition correspond to the so-called consistency conditions.

Under the equilibrium notion in Definition 2.2, Problem 2.1 introduces a MFG where agents interact through the running reward function f with instantaneous mean-field measure m_t and through the terminal cost g with measure μ quantifying the stopping time and stopping position.

2.3. Singular control formulation and entropy regularization. In this subsection, we reformulate the OS-MFG problem with the aim of providing a theoretical framework that enables agents to learn equilibria using RL algorithms. To this end, we modify the original OS-MFG framework by incorporating the exploration–exploitation trade-off, which is empirically known to enhance the convergence of learning methods. Fictitious play algorithms will be discussed later in Section 4.

2.3.1. *Randomized stopping times and singular controls.* We first introduce the concept of randomized stopping times. We borrow ideas from the literature in game theory (see, e.g., [62]) and from [30], where the entropy regularized version of an optimal stopping problem is considered.

A randomized stopping time can be intrinsically connected to singular control when interpreted as the conditional probability of stopping before a given time t . To build this connection formally, define the set of stochastic processes

$$\mathcal{A} := \{ \xi : \Omega \times [0, T] \rightarrow [0, 1], \mathbb{F}^{x_0, W}\text{-adapted, nondecreasing, càdlàg, with } \xi_{0-} = 0 \text{ and } \xi_T = 1 \}.$$

In the rest of this paper, with reference to the terminology of stochastic control theory, we shall refer to an element of \mathcal{A} as to a singular control (see, e.g., Chapter VIII in [37]).

Consider a random variable $U : \Omega \rightarrow [0, 1]$ which is uniformly distributed and independent from W and x_0 . Unless to enlarge the original probability space, we can assume U to be defined on $(\Omega, \mathcal{F}, \mathbb{P})$ as well, and to be \mathcal{F} -measurable. Given $\xi \in \mathcal{A}$, a randomized stopping time is defined as

$$\tau^\xi := \inf \{ t \in [0, T] \mid \xi_t > U \},$$

with the convention $\inf \emptyset = T$.

Notice that τ^ξ is a stopping time with respect to the enlarged filtration $\mathbb{F}^{x_0, W, U}$, generated by x_0, W and U . Hence τ^ξ is not necessarily an $\mathbb{F}^{x_0, W}$ -stopping time. However, if $\tau \in \mathcal{T}$, the process $\xi^\tau \in \mathcal{A}$ defined by $\xi^\tau := (\mathbb{1}_{\{t \geq \tau\}})_t$ is such that $\tau = \tau^{\xi^\tau}$. This gives a natural inclusion of $\{(\xi_t^\tau)_t, \xi_t^\tau = \mathbb{1}_{\{t \geq \tau\}}\}$ into \mathcal{A} . We finally observe that, thanks to the definition of τ^ξ and the independence of U with respect to W and x_0 , we have

$$(2.4) \quad \mathbb{P}(\tau^\xi \leq t \mid \mathcal{F}_t^{x_0, W}) = \mathbb{P}(U \leq \xi_t \mid \mathcal{F}_t^{x_0, W}) = \int_0^{\xi_t} du = \xi_t,$$

so that ξ_t can be interpreted as the (conditional) probability of stopping before time t .

With slight abuse of notation, we will evaluate the payoff J as in (2.1) on randomized stopping times as well. In particular, using (2.4), an application of the tower property and of the independence of U with respect to (W, x_0) allows to rewrite the payoff functional for any $(m, \mu) \in \mathcal{M}$ as

$$(2.5) \quad J(\tau^\xi, m, \mu) = \mathbb{E} \left[\int_0^T f(s, X_s, m_s) (1 - \xi_s) ds + \int_{[0, T]} g(X_s, \mu) d\xi_s \right].$$

Here, $\xi \in \mathcal{A}$ is the singular control associated to the randomized stopping time τ^ξ , and for a generic continuous process $Y := (Y_t)_t$ we have set $\int_{[0, T]} Y_s d\xi_s := Y_0 \xi_0 + \int_0^T Y_s d\xi_s$.

2.3.2. *Consistency conditions for randomized strategies.* Equation (2.5) provides the expression of the representative player's payoff functional when she is allowed to play a randomized stopping rule. We now derive a convenient form of the consistency conditions in (2.3) for randomized stopping times τ^ξ in terms of the related singular control $\xi \in \mathcal{A}$.

For any $\xi \in \mathcal{A}$, by using (2.4), we rewrite the first consistency condition in (2.3) as

$$(2.6) \quad \begin{aligned} \mathbb{P}(X_t \in A, t < \tau^\xi) &= \mathbb{P}(X_t \in A, \xi_t \leq U) \\ &= \mathbb{E} \left[\mathbb{E} \left[\mathbb{1}_{\{X_t \in A, U \geq \xi_t\}} \mid \mathcal{F}_t^{x_0, W} \right] \right] \\ &= \mathbb{E} \left[\int_0^1 \mathbb{1}_{\{X_t \in A, \xi_t \leq u\}} du \right] \\ &= \mathbb{E} \left[\int_{\xi_t}^1 \mathbb{1}_{\{X_t \in A\}} dq \right] \\ &= \mathbb{E} \left[\mathbb{1}_{\{X_t \in A\}} (1 - \xi_t) \right], \quad \forall A \in \mathcal{B}(\mathbb{R}^d), t \in [0, T], \end{aligned}$$

where, again, independence of $U \sim \text{Un}(0, 1)$ with respect to (W, x_0) has been employed.

Similarly, for the second consistency condition in (2.3), we find

$$\begin{aligned}\mathbb{P}\left(X_{\tau^\xi} \in B, \tau^\xi \leq t\right) &= \mathbb{E}\left[\mathbb{1}_{\{X_{\tau^\xi} \in B, \tau^\xi \leq t\}}\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{X_{\tau^\xi} \in B, \xi_t \geq U\}}\right], \quad \forall B \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T].\end{aligned}$$

Furthermore, letting

$$\tau^\xi(u) := \inf\{t \mid \xi_t > u\}, \quad u \in [0, 1],$$

we have $\tau^\xi = \tau^\xi(U)$. Hence

$$\begin{aligned}\mathbb{P}\left(X_{\tau^\xi} \in B, \tau^\xi \leq t\right) &= \mathbb{E}\left[\mathbb{1}_{\{X_{\tau^\xi} \in B, \xi_t \geq U\}}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{X_{\tau^\xi(U)} \in B, \xi_t \geq U\}} \mid \mathcal{F}_t^{x_0, W}\right]\right] \\ &= \mathbb{E}\left[\int_0^1 \mathbb{1}_{\{X_{\tau^\xi(u)} \in B, \xi_t \geq u\}} du\right] \\ &= \mathbb{E}\left[\int_0^1 \mathbb{1}_{\{X_{\tau^\xi(u)} \in B, \tau^\xi(u) \leq t\}} du\right].\end{aligned}$$

By using the change of variables formula (see Proposition 4.9 in Chapter 0 in [57], among others), we finally obtain

$$(2.7) \quad \mathbb{E}\left[\mathbb{1}_{\{X_{\tau^\xi} \in B, \xi_t \geq U\}}\right] = \mathbb{E}\left[\int_{[0, T]} \mathbb{1}_{\{X_s \in B, s \leq t\}} d\xi_s\right] = \mathbb{E}\left[\int_{[0, t]} \mathbb{1}_{\{X_s \in B\}} d\xi_s\right].$$

Thus, if for $(\hat{\xi}, \hat{m}, \hat{\mu}) \in \mathcal{A} \times \mathcal{M}$ the triple $(\tau^{\hat{\xi}}, \hat{m}, \hat{\mu})$ satisfy the consistency conditions in (2.3), then by (2.6) and (2.7) it follows that

$$(2.8) \quad \begin{aligned}\hat{m}_t(A) &= \mathbb{E}\left[\mathbb{1}_{\{X_t \in A\}}(1 - \hat{\xi}_t)\right], \quad \forall A \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T], \\ \hat{\mu}(B \times [0, t]) &= \mathbb{E}\left[\int_{[0, t]} \mathbb{1}_{\{X_s \in B\}} d\hat{\xi}_s\right], \quad \forall B \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T],\end{aligned}$$

which is the desired reformulation in terms of $\hat{\xi}$.

2.3.3. Entropy regularization and new equilibrium concept. It is clear from (2.5) that $\mathcal{A} \ni \xi \mapsto J(\xi, m, \mu) \in \mathbb{R}$ is linear, and it admits optimizers of the "bang-bang" form $\xi_t = \mathbb{1}_{t \geq \tau}$, for some strict optimal stopping $\tau \in \mathcal{T}$. Hence, there is no guarantee that equilibria are of randomized type – a phenomenon already noted in [30] in a single-agent setting without mean-field interaction.

To encourage randomized equilibrium strategies, we introduce an entropy-regularized version of the profit functional J . Specifically, we consider a general entropy function \mathcal{E} that satisfies the following conditions.

Assumption 2.3. *The entropy function $\mathcal{E} : [0, 1] \rightarrow [0, \infty)$ is bounded, strictly concave, and has maximum in $(0, 1)$.*

A natural choice for the entropy function \mathcal{E} is the *cumulative residual entropy* (see [55]) which is given by $\mathcal{E}(z) = -z \log z$, employed in [30]. We also refer to [44] for examples related to Choquet integrals.

Given a temperature parameter $\lambda \geq 0$ that controls the level of exploration, for any $(m, \mu) \in \mathcal{M}$ we define the entropy regularized profit functional

$$(2.9) \quad J^\lambda(\xi, m, \mu) := \mathbb{E}\left[\int_0^T (f(t, X_t, m_t)(1 - \xi_t) + \lambda \mathcal{E}(\xi_t)) dt + \int_{[0, T]} g(X_t, \mu) d\xi_t\right], \quad \xi \in \mathcal{A}.$$

A control $\xi^* \in \mathcal{A}$ is optimal for (m, μ) if $J^\lambda(\xi^*, m, \mu) \geq J^\lambda(\xi, m, \mu)$ for any other $\xi \in \mathcal{A}$. The existence of an optimal control follows by classical arguments (see e.g. [50]), after noticing that the function $\xi \mapsto J^\lambda(\xi, m, \mu)$ is concave. In the case $\lambda > 0$, such a concavity is strict, and the optimizer is unique.

The entropy \mathcal{E} is nonnegative and achieves its highest values when the probability ξ_t is near the maximum point of \mathcal{E} . Since the maximum point of \mathcal{E} is an internal point of $[0, 1]$, this incentivizes the use of randomized stopping times. Indeed, in the case $\lambda > 0$, it can be seen that the unique optimal control is not of bang-bang type (see also [30]).

In light of the consistency conditions derived in (2.8), we introduce the following notion of equilibrium to the entropy regularized MFG with singular controls (λ -SC-MFG, in short).

Definition 2.4 (Entropy regularized MFG equilibrium with singular controls). *An equilibrium to the λ -SC-MFG problem is a triple $(\hat{\xi}, \hat{m}, \hat{\mu})$ such that*

$$(2.10) \quad \begin{cases} \hat{\xi} \in \arg \max_{\xi \in \mathcal{A}} J^\lambda(\xi, \hat{m}, \hat{\mu}), \\ \hat{m}_t(A) = \mathbb{E} \left[\mathbf{1}_{\{X_t \in A\}} (1 - \hat{\xi}_t) \right], \quad \forall A \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T], \\ \hat{\mu}(B \times [0, t]) = \mathbb{E} \left[\int_{[0, t]} \mathbf{1}_{\{X_s \in B\}} d\hat{\xi}_s \right], \quad \forall B \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T]. \end{cases}$$

Remark 2.5. *From the mathematical point of view, we notice that:*

- (1) *When $\lambda = 0$, we have $J^0(\xi, m, \mu) = J(\tau^\xi, m, \mu)$, so that the problem (2.10) is just the definition of equilibrium for the OS-MFG with mixed stopping strategies. Such a game is very closely related to those studied in [8, 33, 34], using a linear programming approach. A rigorous relation between mixed stopping strategies and the occupation measures arising in the linear-programming approach is provided in [32], where mixed strategies are expressed in terms of probability kernels.*
- (2) *When $\lambda > 0$, the problem (2.10) is a new type of MFG with singular controls which has not been considered in the literature. In particular, such a model cannot be embedded into the (general) results obtained in [11, 25, 38, 39] because of the structure of our consistency condition (2.8) involving the integral with respect to the $d\xi$ -random measure.*

Remark 2.6. *For any fixed $(m, \mu) \in \mathcal{M}$, the optimal value and the corresponding optimal strategy of the control problem remain stable as $\lambda \rightarrow 0$ (see [30] for details). However, due to the well-known sensitivity of Nash equilibria to model perturbations, it is a nontrivial question whether the equilibria themselves remain stable in the limit $\lambda \rightarrow 0$.*

From an applied and learning perspective, this question is particularly important, as it provides theoretical guarantees on the closeness between the equilibria of the entropy-regularized problem (2.10) – which we aim to learn – and those of the original problem (2.3). We address this issue and establish a positive result in Theorem 3.10.

3. ANALYSIS OF ENTROPY REGULARIZED OS-MFGs

The aim of this section is to study in detail the entropy regularized λ -SC-MFG (2.10) and its relations, as the temperature parameter $\lambda \rightarrow 0$, with our original OS-MFG (2.3).

3.1. Preliminary results. We first endow the set of singular controls \mathcal{A} with a topology. Notice that \mathcal{A} is a closed convex subset of the Hilbert space $\mathbb{L}^2(\Omega \times [0, T])$, i.e. the set of $\mathcal{F} \otimes \mathcal{B}([0, T])$ -measurable functions ℓ such that $\mathbb{E}[\int_0^T \ell_t^2 dt] < \infty$: indeed, if a sequence $(\xi^n)_n \subset \mathcal{A}$ converges to some $\xi \in \mathbb{L}^2(\Omega \times [0, T])$, it is always possible (e.g., via Lemmas 4.5 and 4.6 in [48]) to find a càdlàg monotone nondecreasing modification of ξ , so that $\xi \in \mathcal{A}$. Thus, consider on \mathcal{A} the topology of weak convergence (in the Hilbert sense) of $\mathbb{L}^2(\Omega \times [0, T])$, and denote the convergence with $\xi^n \rightarrow \xi$. Since

the set \mathcal{A} is convex and closed for the strong topology, it is closed for the weak topology. Therefore, since \mathcal{A} is also bounded, it is compact when endowed with the weak topology.

We define the consistency map $\Gamma : \mathcal{A} \rightarrow \mathcal{M}$ as the map $\Gamma(\xi) := (m^\xi, \mu^\xi)$ with

$$(3.1) \quad \begin{aligned} m_t^\xi(A) &:= \mathbb{E} [\mathbb{1}_{\{X_t \in A\}} (1 - \xi_t)], \quad \forall A \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T], \\ \mu^\xi(B \times [0, t]) &:= \mathbb{E} \left[\int_{[0, t]} \mathbb{1}_{\{X_t \in B\}} d\xi_t \right], \quad \forall B \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T]. \end{aligned}$$

For our subsequent analysis, in the next three lemmas we discuss boundedness, closedness and compactness of the set $\Gamma(\mathcal{A})$.

Lemma 3.1. *The consistency map Γ is bounded; that is, there exists a constant $C_p > 0$ such that*

$$\int_{\mathbb{R}^d} |z|^p m_t^\xi(dz) \leq C_p, \quad \forall t \in [0, T], \quad \text{and} \quad \int_{[0, T] \times \mathbb{R}^d} (t^p + |z|^p) \mu^\xi(dt, dz) \leq C_p,$$

for any $\xi \in \mathcal{A}$.

Proof. By using the boundedness condition $0 \leq \xi_t \leq 1$, we obtain

$$\int_{\mathbb{R}^d} |z|^p m_t^\xi(dz) = \mathbb{E}[|X_t|^p (1 - \xi_t)] \leq \mathbb{E} \left[\sup_{t \in [0, T]} |X_t|^p \right] \leq C_p,$$

for a suitable constant C_p (see (2.2)). This shows the first estimate. For the second estimate, again by boundedness of ξ , we have

$$\int_{[0, T] \times \mathbb{R}^d} (t^p + |z|^p) \mu^\xi(dt, dz) = \mathbb{E} \left[\int_{[0, T]} (t^p + |X_t|^p) d\xi_t \right] \leq T^p + \mathbb{E} \left[\sup_{t \in [0, T]} |X_t|^p \right] \leq C_p,$$

unless to take a larger C_p . □

Lemma 3.2. *The consistency map $\Gamma : \mathcal{A} \rightarrow \mathcal{M}$ has closed graph.*

Proof. We have to show that, if $(m^n, \mu^n) \in \Gamma(\mathcal{A})$ converges to (m, μ) , then $(m, \mu) \in \Gamma(\mathcal{A})$. Let $\xi^n \in \mathcal{A}$ such that $(m^n, \mu^n) = (m^{\xi^n}, \mu^{\xi^n})$. By compactness of the set \mathcal{A} , there exists a subsequence ξ^{n_k} which converges weakly to ξ^* . Since $\sup_{t \geq 0} \xi_t^{n_k} \leq 1$, we can employ Lemma 3.5 in [47] in order to find a subsequence (not relabelled) $(n_k)_k$ and a limit point $\xi \in \mathcal{A}$ such that, setting $\zeta_t^j := \frac{1}{j} \sum_{k=1}^j \xi_t^{n_k}$, we have

$$\int_{[0, T]} \phi_t d\zeta_t^j \rightarrow \int_{[0, T]} \phi_t d\xi_t, \quad \mathbb{P}\text{-a.s.}, \quad \text{as } j \rightarrow \infty, \quad \text{for any } \phi \in C_b([0, T]).$$

One can easily check that $\xi = \xi^*$. Therefore, it holds

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi(x) m_t(dx) &= \lim_n \frac{1}{N} \sum_{k=1}^N \int_{\mathbb{R}^d} \varphi(x) m_t^{n_k}(dx) \\ &= \lim_n \frac{1}{N} \sum_{k=1}^N \mathbb{E} [\varphi(X_t) (1 - \xi_t^{n_k})] \\ &= \lim_n \mathbb{E} [\varphi(X_t) (1 - \zeta_t^n)] \\ &= \mathbb{E} [\varphi(X_t) (1 - \xi_t)], \quad \forall t \in [0, T], \quad \varphi \in C_b(\mathbb{R}^d), \end{aligned}$$

and

$$\begin{aligned}
\int_{[0,T] \times \mathbb{R}^d} \varphi(t, x) \mu(dt, dx) &= \lim_n \frac{1}{N} \sum_{k=1}^N \mathbb{E} \left[\int_{[0,T]} \varphi(t, X_t) d\xi_t^{n_k} \right] \\
&= \lim_n \mathbb{E} \left[\int_{[0,T]} \varphi(t, X_t) d\xi_t^N \right] \\
&= \mathbb{E} \left[\int_{[0,T]} \varphi(t, X_t) d\xi_t \right], \quad \forall \varphi \in C_b([0, T] \times \mathbb{R}^d),
\end{aligned}$$

from which we conclude that $(m, \mu) = \Gamma(\xi)$, thus completing the proof. \square

Lemma 3.3. *The set $\Gamma(\mathcal{A}) \subset \mathcal{M}$ is compact.*

Proof. Let us first introduce the set \mathcal{M}^* of pairs $(\mu, m) \in \mathcal{P}_p([0, T] \times \mathbb{R}^d) \times M_p$, such that for all $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$,

$$\int_{[0,T] \times \mathbb{R}^d} u(t, x) \mu(dt, dx) = \int_{\mathbb{R}^d} u(0, x) m_0(dx) + \int_0^T \int_{\mathbb{R}^d} (\partial_t u + \mathcal{L}u)(t, x) m_t(dx) dt,$$

where $\mathcal{L}u(t, x) := \nabla_x u(t, x)^\top b(t, x) + \frac{1}{2} \text{Tr}[\sigma^\top (H_x u) \sigma]$, with $\nabla_x u := (\partial_{x_1} u, \dots, \partial_{x_d} u)$, $H_x u$ the Hessian matrix of u with respect to x and Tr the trace operator. We also denote by $m_0 := \mathcal{L}(x_0)$. For a given $(\xi_t) \in \mathcal{A}$, by applying Ito's formula, it can be shown that the associated pair $(m^\xi, \mu^\xi) = \Gamma(\xi)$ belongs to \mathcal{M}^* . Therefore, $\Gamma(\mathcal{A}) \subset \mathcal{M}^*$. By Theorem 2.10 in [34], the set \mathcal{M}^* is compact for the topology of the convergence in measure.

Furthermore, thanks to Lemma 3.2, the set $\Gamma(\mathcal{A})$ is a closed subset of \mathcal{M}^* . It thus follows that it is compact. \square

Remark 3.4. *Under some additional assumptions (e.g. uniform ellipticity) it can be actually shown that $\Gamma(\mathcal{A}) = \mathcal{M}^*$ (see [32]).*

Next, for $(m, \mu, \lambda) \in \mathcal{M} \times [0, \infty)$, define the best response map $R_\lambda(m, \mu) \subset \mathcal{A}$ as

$$(3.2) \quad R_\lambda(m, \mu) := \arg \max_{\xi \in \mathcal{A}} J^\lambda(\xi, m, \mu).$$

For any $(m, \mu, \lambda) \in \mathcal{M} \times [0, \infty)$, via classical concavity arguments (see e.g. [50]), we have that $R_\lambda(m, \mu)$ is nonempty. Notice that, when $\lambda > 0$, the strict concavity of $\xi \mapsto J^\lambda(\xi, m, \mu)$ implies that $R_\lambda(m, \mu)$ is actually a singleton. In this case, with abuse of notation we will indicate with $R_\lambda(m, \mu) = (R_\lambda(m, \mu)_t)_t$ both the singleton and its only element.

The best response map has closed graph, and enjoys the following continuity properties.

Lemma 3.5. *Under Assumptions 2.1, the following statements hold true:*

- (1) *The best response map $(m, \mu, \lambda) \mapsto R_\lambda(m, \mu)$ is continuous in $\Gamma(\mathcal{A}) \times (0, \infty)$.*
- (2) *For any sequence $(m^n, \mu^n, \lambda^n)_n \subset \Gamma(\mathcal{A}) \times [0, \infty)$ converging to $(m, \mu, 0)$, and any $\xi^n \in R_{\lambda^n}(m^n, \mu^n)$, there exists a subsequence $(n_k)_k$ and $\xi^* \in R_0(m, \mu)$ such that $\xi^{n_k} \rightarrow \xi^*$.*

Proof. Consider a sequence $(m^n, \mu^n, \lambda^n)_n$ converging to (m, μ, λ) and take $\xi^n \in R_{\lambda^n}(m^n, \mu^n)$. By compactness of \mathcal{A} in the weak topology, for any subsequence of indexes $(n_k)_k$, we can find a further subsequence (not relabeled) and a limit point ξ^* such that $\xi^{n_k} \rightarrow \xi^*$ weakly in $\mathbb{L}^2(\Omega \times [0, T])$ as $k \rightarrow \infty$. We need to show that ξ^* is optimal for $J^\lambda(\cdot, m, \mu)$. To this end, notice that, by optimality of ξ^{n_k} for $J^{\lambda^{n_k}}(\cdot, m^{n_k}, \mu^{n_k})$, we have

$$(3.3) \quad J^{\lambda^{n_k}}(\xi^{n_k}, m^{n_k}, \mu^{n_k}) \geq J^{\lambda^{n_k}}(\xi, m^{n_k}, \mu^{n_k}), \quad \forall \xi \in \mathcal{A}.$$

In order to discuss limits as $n_k \rightarrow \infty$ in the previous inequality, we first show that

$$(3.4) \quad \limsup_k \sup_{\xi \in \mathcal{A}} |J^{\lambda^{n_k}}(\xi, m^{n_k}, \mu^{n_k}) - J^\lambda(\xi, m, \mu)| = 0.$$

Using the boundedness of \mathcal{E} and of the controls ξ , we first find

$$\begin{aligned} & \sup_{\xi \in \mathcal{A}} |J^{\lambda^{n_k}}(\xi, m^{n_k}, \mu^{n_k}) - J^\lambda(\xi, m, \mu)| \\ & \leq \sup_{\xi \in \mathcal{A}} \mathbb{E} \left[\int_0^T |f(t, X_t, m_t^{n_k}) - f(t, X_t, m_t)| (1 - \xi_t) dt + \int_{[0, T]} |g(X_t, \mu^{n_k}) - g(X_t, \mu)| d\xi_t \right] \\ & \quad + |\lambda^{n_k} - \lambda| \sup_{\xi \in \mathcal{A}} \mathbb{E} \left[\int_0^T \mathcal{E}(\xi_t) dt \right] \\ & \leq \mathbb{E} \left[\int_0^T |f(t, X_t, m_t^{n_k}) - f(t, X_t, m_t)| dt \right] + \mathbb{E} \left[\sup_{t \in [0, T]} |g(X_t, \mu^{n_k}) - g(X_t, \mu)| \right] + C|\lambda^{n_k} - \lambda|, \end{aligned}$$

so that it remains to study the convergence of the two terms under expectation. For the term involving f , we show that for any subsequence of $(n_k)_k$ (not relabelled), there exists a further subsequence (again, not relabelled) such that

$$(3.5) \quad \lim_k \mathbb{E} \left[\int_0^T |f(t, X_t, m_t^{n_k}) - f(t, X_t, m_t)| dt \right] = 0,$$

so that the whole sequence converges. Indeed, given a subsequence, we can find a further subsequence such that $d'_p(m_t^{n_k}, m_t) \rightarrow 0$ dt -a.e. (see (1.3)), and thanks to the continuity of f , we have that $f(t, X_t, m_t^{n_k}) \rightarrow f(t, X_t, m_t)$ $\mathbb{P} \otimes dt$ -a.e. Moreover, the growth conditions of f give

$$\begin{aligned} \sup_k |f(t, X_t, m_t^{n_k}) - f(t, X_t, m_t)| & \leq C \left(1 + \sup_{t \in [0, T]} |X_t|^p + \sup_k \int_{\mathbb{R}^d} |z|^p m_t^{n_k}(dz) + \int_{\mathbb{R}^d} |z|^p m_t(dz) \right) \\ & \leq C \left(1 + \sup_{t \in [0, T]} |X_t|^p + 2C_p \right), \end{aligned}$$

for a constant C_p as in Lemma 3.1 (notice indeed that the estimate $\int_{\mathbb{R}^d} |z|^p m_t(dz)$ easily follows by convergence). Since $\mathbb{E} \left[\sup_{t \in [0, T]} |X_t|^p \right]$ is finite (see (2.2)), the limits in (3.5) follows by the dominated convergence theorem. We next show the limit for the term involving g . Using the uniform continuity of g , for w_g as in Assumption 2.1 we have

$$\lim_k \mathbb{E} \left[\sup_{t \in [0, T]} |g(X_t, \mu^{n_k}) - g(X_t, \mu)| \right] \leq \lim_k C \left(1 + \mathbb{E} \left[\sup_{t \in [0, T]} |X_t|^p \right] \right) w_g(\mu^{n_k}, \mu) = 0.$$

This limit, together with (3.5), allows us to obtain (3.4).

As a consequence of (3.4), we have

$$(3.6) \quad \lim_k J^{\lambda^{n_k}}(\xi, m^{n_k}, \mu^{n_k}) = J^\lambda(\xi, m, \mu),$$

as well as

$$(3.7) \quad \lim_k |J^{\lambda^{n_k}}(\xi^{n_k}, m^{n_k}, \mu^{n_k}) - J^\lambda(\xi^{n_k}, m, \mu)| = 0.$$

We next want to show that

$$(3.8) \quad V^* := \limsup_k J^\lambda(\xi^{n_k}, m, \mu) \leq J^\lambda(\xi^*, m, \mu).$$

Fix a subsequence (not relabelled) $(n_k)_k$ such that $V^* := \lim_k J^\lambda(\xi^{n_k}, m, \mu)$. Since $\sup_{t \geq 0} \xi_t^{n_k} \leq 1$, we can again employ Lemma 3.5 in [47] in order to find a subsequence (not relabelled) $(\bar{n}_k)_k$ and a limit point $\bar{\xi} \in \mathcal{A}$ such that, setting $\zeta_t^j := \frac{1}{j} \sum_{k=1}^j \xi_t^{n_k}$, we have

$$\int_{[0,T]} \phi_t d\zeta_t^j \rightarrow \int_{[0,T]} \phi_t d\bar{\xi}_t, \quad \mathbb{P}\text{-a.s.}, \text{ as } j \rightarrow \infty, \text{ for any } \phi \in C_b([0, T]).$$

The processes $\bar{\xi}$ and ξ^* coincide. Furthermore, by the convergence of ζ^j and the concavity of $J^\lambda(\cdot, m, \mu)$ we find

$$J^\lambda(\xi^*, m, \mu) = \lim_j J^\lambda(\zeta^j, m, \mu) \geq \lim_j \frac{1}{j} \sum_{k=1}^j J^\lambda(\xi^{n_k}, m, \mu) = V^*,$$

which proves (3.8).

Finally, by using (3.6), (3.7) and (3.8) to take limits as $k \rightarrow \infty$ in (3.3), we obtain

$$J^\lambda(\xi^*, m, \mu) \geq J^\lambda(\xi, m, \mu), \quad \forall \xi \in \mathcal{A},$$

showing that $\xi^* \in R_\lambda(m, \mu)$. When $\lambda = 0$, this completes the proof of Claim (2).

To show Claim (1), we notice that, when $\lambda > 0$, by uniqueness of the optimal control we have that $\xi^* = R_\lambda(m, \mu)$. In this case, we have shown that for any subsequence $(n_k)_k$ we can find a further subsequence (not relabelled) such that $(\xi^{n_k})_k$ converges to $R_\lambda(m, \mu)$. Hence, the whole sequence $(\xi^n)_n$ converges to $R_\lambda(m, \mu)$, thus completing the proof. \square

3.2. Existence of equilibria of the SC-MFG. We are now ready to discuss the following existence result.

Theorem 3.6. *Under Assumption 2.1, for any $\lambda \geq 0$ there exists an equilibrium to the λ -SC-MFG (2.10).*

Proof. The proof consists in employing Kakutani–Glicksberg–Ky Fan fixed point theorem to the map $\mathcal{R}_\lambda : \mathcal{A} \rightarrow \mathcal{A}$, given by the composition $\mathcal{R}_\lambda := R_\lambda \circ \Gamma$. Noticing that \mathcal{A} is a nonempty compact convex subset of the locally convex Hausdorff topological vector space $\mathbb{L}^2(\Omega \times [0, T])$, it only remains to show that the map \mathcal{R}_λ has closed graph.

To this end, for any sequence $\xi^n \rightarrow \xi$, with $\mathcal{R}_\lambda(\xi^n) \rightarrow \xi^\lambda \in \mathcal{A}$, we need to show that $\xi^\lambda \in \mathcal{R}_\lambda(\xi)$. Set $(m^n, \mu^n) := (m^{\xi^n}, \mu^{\xi^n}) := \Gamma(\xi^n)$ and notice that, due to the compactness of $\Gamma(\mathcal{A})$ (see Lemma 3.3), there exists a subsequence (not relabeled) and a limit point $(m, \mu) \in \mathcal{M}$ such that $(m^n, \mu^n) \rightarrow (m, \mu)$. Moreover, by Lemma 3.2, we have

$$(3.9) \quad (m, \mu) = \Gamma(\xi).$$

Now, by assumption, $\mathcal{R}_\lambda(\xi^n) = R_\lambda(\Gamma(\xi^n))$ converges to ξ^λ and, using (3.9), we have $\Gamma(\xi^{n_k}) \rightarrow \Gamma(\xi)$ on a subsequence n_k . Thanks to Lemma 3.5, the map $R_\lambda : \mathcal{M} \rightarrow \mathcal{A}$ has closed graph, so that

$$\xi^\lambda \in R_\lambda(\Gamma(\xi)) = \mathcal{R}_\lambda(\xi).$$

This shows that \mathcal{R}_λ has closed graph, and hence completes the proof. \square

3.3. λ -stability of equilibria. A crucial point when studying the limit, as $\lambda \rightarrow 0$, of the equilibria to the entropy regularized λ -SC-MFG (2.10) is whether the equilibrium to the 0-SC-MFG (2.10) is unique or not.

To this end, we introduce the following requirement, which is a version of the so called Lasry–Lions monotonicity conditions.

Assumption 3.7 (Monotonicity). *For any $\lambda \geq 0$, the following hold true:*

- (1) *For any (m, μ) , $R_\lambda(m, \mu)$ is a singleton.*

$$(2) \quad J^\lambda(\xi, m^\xi, \mu^\xi) - J^\lambda(\bar{\xi}, m^\xi, \mu^\xi) - (J^\lambda(\xi, m^{\bar{\xi}}, \mu^{\bar{\xi}}) - J^\lambda(\bar{\xi}, m^{\bar{\xi}}, \mu^{\bar{\xi}})) \leq 0.$$

Notice that Condition 1 in Assumption 3.7 is always satisfied when $\lambda > 0$.

Remark 3.8. We provide here an example in which Condition (2) in Assumption (3.7) is satisfied. For example, one could consider f and g of the form

$$f(t, x, m) = \bar{k}(x) \bar{f}\left(t, \int_{\mathbb{R}^d} \bar{k}(x) m_t(dx)\right), \quad g(x, \mu) = \bar{\ell}(x) \bar{h}\left(\int_{[0, T] \times \mathbb{R}^d} \bar{\ell}(x) \mu(dt, dx)\right),$$

with \bar{f} non-increasing in the second argument and \bar{h} non-decreasing.

Theorem 3.9. Under Assumptions 2.1 and 3.7, for any $\lambda \geq 0$ there exists a unique equilibrium to λ -SC-MFG (2.10).

Proof. The proof of this result this theorem is standard and slightly adapted from [8], we include it for completeness.

Arguing by contradiction, assume that there exists two distinct mean-field equilibria (ξ, m, μ) , $(\bar{\xi}, \bar{m}, \bar{\mu})$. By uniqueness of the optimal control and the definition of equilibrium, we have

$$J^\lambda(\xi, m^\xi, \mu^\xi) - J^\lambda(\bar{\xi}, m^\xi, \mu^\xi) > 0 \quad \text{and} \quad J^\lambda(\bar{\xi}, m^{\bar{\xi}}, \mu^{\bar{\xi}}) - J^\lambda(\xi, m^{\bar{\xi}}, \mu^{\bar{\xi}}) > 0.$$

Summing up these two inequalities, we obtain

$$J^\lambda(\xi, m^\xi, \mu^\xi) - J^\lambda(\bar{\xi}, m^\xi, \mu^\xi) - (J^\lambda(\xi, m^{\bar{\xi}}, \mu^{\bar{\xi}}) - J^\lambda(\bar{\xi}, m^{\bar{\xi}}, \mu^{\bar{\xi}})) > 0,$$

which contradicts the second condition in Assumption 3.7. Therefore, the equilibrium is unique. \square

Entropy regularization and the corresponding temperature parameter λ are introduced to encourage randomization in a learning environment. It is important to understand the closeness between the equilibria of the entropy-regularized problem (2.10) – which we aim to learn – and those of the original problem (2.3). We discuss this convergence in the following theorem.

Theorem 3.10. Under Assumption 2.1, for any $\lambda > 0$, let $(\hat{\xi}^\lambda, \hat{m}^\lambda, \hat{\mu}^\lambda)$ be an equilibrium to the λ -SC-MFG. Then

- (1) $(\hat{\xi}^\lambda, \hat{m}^\lambda, \hat{\mu}^\lambda) \rightarrow (\hat{\xi}^0, \hat{m}^0, \hat{\mu}^0)$ up to subsequence as $\lambda \rightarrow 0$, where $(\hat{\xi}^0, \hat{m}^0, \hat{\mu}^0)$ is an equilibrium to the 0-SC-MFG;
- (2) Under the additional Assumption 3.7, we have $(\hat{\xi}^\lambda, \hat{m}^\lambda, \hat{\mu}^\lambda) \rightarrow (\hat{\xi}^0, \hat{m}^0, \hat{\mu}^0)$ as $\lambda \rightarrow 0$.

Proof. To simplify the notation, we simply write $(\xi^\lambda, m^\lambda, \mu^\lambda)$ instead of $(\hat{\xi}^\lambda, \hat{m}^\lambda, \hat{\mu}^\lambda)$, for any $\lambda \geq 0$.

Fix any sequence $(\lambda_n)_n \subset (0, \infty)$ with $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$ and set

$$(\xi^n, m^n, \mu^n) := (\xi^{\lambda_n}, m^{\lambda_n}, \mu^{\lambda_n}).$$

Since (ξ^n, m^n, μ^n) are assumed to be equilibria, we have $(m^n, \mu^n) = \Gamma(R_\lambda(\xi^n))$, so that $(m^n, \mu^n)_n \subset \Gamma(\mathcal{A})$. By compactness of $\mathcal{A} \times \Gamma(\mathcal{A})$ (see Lemma 3.3), we can extract a subsequence (not relabeled) of $(\xi^n, m^n, \mu^n)_n$ and a limit point (ξ^0, m^0, μ^0) such that $(\xi^n, m^n, \mu^n) \rightarrow (\xi^0, m^0, \mu^0)$. Thanks to Lemma 2, we have that $\xi^0 \in R_0(m^0, \mu^0)$. Moreover, by repeating the arguments in the proof of Theorem 3.6, we have that $\Gamma(\xi^0) = (m^0, \mu^0)$. This, in turn implies that $\xi^0 \in R_0(m^0, \mu^0) = R_0(\Gamma(\xi^0)) = \mathcal{R}_0(\xi^0)$, so that (ξ^0, m^0, μ^0) is an equilibrium of the 0-SC-MFG. This completes the proof of Claim (1).

When the additional Assumption 3.7 holds, by Theorem 3.9 we have the uniqueness of the equilibrium of the 0-SC-MFG. Hence, by the previous argument, any sequence $(\xi^{\lambda_n}, m^{\lambda_n}, \mu^{\lambda_n})$ converges to (ξ^0, m^0, μ^0) , thus proving Claim (2). \square

4. FICTITIOUS PLAY ALGORITHMS

In this section, we introduce two novel fictitious play algorithms for computing the unique mean-field equilibrium λ -SC-MFG in the case when $\lambda > 0$, and we establish their convergence under different data assumptions.

Note that the convergence of iterative numerical schemes – such as fictitious play algorithms under known model parameters – serves as a fundamental foundation for analyzing the convergence of RL algorithms in environments with unknown parameters. In many popular RL methods, such as Q-learning and actor-critic algorithms [60], the known quantities in these schemes (e.g., the value function) are simply replaced by their corresponding estimates, computed from available observations.

Algorithm 1: Fictitious play algorithm

Data: A number of steps n for the equilibrium approximation; a control $\bar{\xi}^0 \in \mathcal{A}$ and $(\bar{\mu}^0, \bar{m}^0) := \Gamma(\bar{\xi}^0) \in \mathcal{R}_\lambda$;

Result: Approximate MFG Nash equilibrium

```

1 for  $k = 0, 1, \dots, n - 1$  do
2   Set  $\xi^{k+1} := \arg \max_{\xi \in \mathcal{A}} J^\lambda(\xi, \bar{m}^k, \bar{\mu}^k)$ ;
3   Set  $\bar{\xi}^{k+1} := \frac{k}{k+1} \bar{\xi}^k + \frac{1}{k+1} \xi^{k+1} = \frac{1}{k+1} \sum_{\ell=1}^{k+1} \xi^\ell$ ;
4   Set  $(\bar{\mu}^{k+1}, \bar{m}^{k+1}) := \Gamma(\bar{\xi}^{k+1})$ .
5 end
```

Observe that, due to the linear dependence of the measure (m, μ) with respect to ξ , we have that $(\bar{\mu}^n, \bar{m}^n) = \frac{1}{n} \sum_{k=1}^n (\mu^{\xi^k}, m^{\xi^k})$.

In the next two subsections, we establish the convergence of the fictitious play algorithm $(\bar{m}^n, \bar{\mu}^n)_n$ to the equilibria of the λ -SC-MFG, under suitable structural conditions on the data, and for suitable choice of the initialization.

4.1. Fictitious play under the Lasry-Lions condition. We start with the fictitious play algorithm under the Lasry-Lions monotonicity condition.

This fictitious-play algorithm is novel in the context of SC-MFGs and also brings new contributions compared to the fictitious-play algorithm developed in [34] for OS-MFGs. In particular, it extends to cases where the profit-functional is non-linear with respect to *therandomized stopping*, a setting in which the linear programming approach developed in [34] fails.

To simplify the presentation, we set up some notation. First, notice that for a given $(m, \mu) \in \Gamma(\mathcal{A})$ and $\xi \in \mathcal{A}$, by using the identification between controls and associated measures given by (3.1), we can rewrite J^λ as follows:

$$J^\lambda(\xi, m, \mu) = \int_0^T \int_{\mathbb{R}^d} \left(f(t, x, m_t) m_t^\xi(dx) dt + g(t, x, \mu) \mu^\xi(dt, dx) \right) + \lambda \mathbb{E} \int_0^T \mathcal{E}(\xi_t) dt.$$

We now introduce the bilinear form notation

$$\langle f(m), m' \rangle := \int_0^T \int_{\mathbb{R}^d} f(t, x, m_t) m'_t(dx) dt, \quad \langle g(\mu), \mu' \rangle := \int_{[0, T] \times \mathbb{R}^d} g(t, x, \mu) \mu'(dt, dx),$$

where $(\mu, m), (\mu', m') \in \mathcal{P}_p([0, T] \times \mathbb{R}) \times M_p$. Therefore, J^λ writes as follows:

$$J^\lambda(\xi, m, \mu) = \langle f(m), m^\xi \rangle + \langle g(\mu), \mu^\xi \rangle + \lambda \mathbb{E} \int_0^T \mathcal{E}(\xi_t) dt.$$

We assume the following conditions hold:

Assumption 4.1. *There exist constants $c_f > 0$ and $c_g > 0$ such that for all $t \in [0, T]$, $x, x' \in \mathbb{R}^d$, $m, m' \in \mathcal{P}_p^{sub}(\mathbb{R}^d)$, $\mu, \mu' \in \mathcal{P}_p([0, T] \times \mathbb{R}^d)$,*

$$|f(t, x, m) - f(t, x, m')| \leq c_f(1 + |x|) \int_{\mathbb{R}^d} (1 + |z|^p) |m - m'| (dz),$$

$$|f(t, x, m) - f(t, x, m') - f(t, x', m) + f(t, x', m')| \leq c_f |x - x'| \int_{\mathbb{R}^d} (1 + |z|^p) |m - m'| (dz),$$

$$|g(t, x, \mu) - g(t, x, \mu')| \leq c_g(1 + |x|) \int_{[0, T] \times \mathbb{R}^d} (1 + |z|^p) |\mu - \mu'| (ds, dz),$$

$$|g(t, x, \mu) - g(t, x, \mu') - g(t', x', \mu) + g(t', x', \mu')| \leq c_g(|t - t'| + |x - x'|) \int_{[0, T] \times \mathbb{R}^d} (1 + |z|^p) |\mu - \mu'| (ds, dz),$$

where $|m|$ denotes the total variation measure of m .

Before proving the main result on the convergence of the fictitious play algorithm, we first recall from [34] some estimates (on the linear part of the payoff functional, as well as on the admissible measures), which will be needed in the proof of Theorem 4.5 below.

Lemma 4.2. *(Estimates on distances) For all $n \geq 1$, we have the following estimates:*

(i) *There exists a constant $C_1 > 0$ such that*

$$|d_1^M(\bar{m}^{n+1}, \bar{m}^n)| \leq \frac{C_1}{n}, \quad |d_1(\bar{\mu}^{n+1}, \bar{\mu}^n)| \leq \frac{C_1}{n}.$$

(ii) *There exists a constant $C_2 > 0$ (independent of n) such that for all $t \in [0, T]$:*

$$(4.1) \quad \int_{[0, T] \times \mathbb{R}} (1 + |x|^p) |\bar{\mu}^{n+1} - \bar{\mu}^n| (dt, dx) \leq \frac{C_2}{n}, \quad \int_{[0, T] \times \mathbb{R}} (1 + |x|^p) |\bar{m}_t^{n+1} - \bar{m}_t^n| (dt, dx) \leq \frac{C_2}{n}.$$

Lemma 4.3 *(Estimates on the rewards). There exist constants $C_f > 0$ and $C_g > 0$ such that for all $n \geq 1$*

$$\begin{aligned} \langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+2} - m^{n+1} \rangle &\leq \frac{C_f}{n} d_1^M(m^{n+1}, m^{n+2}), \\ \langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \mu^{n+2} - \mu^{n+1} \rangle &\leq \frac{C_g}{n} d_1(\mu^{n+1}, \mu^{n+2}). \end{aligned}$$

Lemma 4.4. *(Proximity between two successive best responses) The following convergence results hold:*

$$(4.2) \quad \lim_{n \rightarrow \infty} d_1(\mu^n, \mu^{n+1}) = 0, \quad \lim_{n \rightarrow \infty} d_1^M(m^n, m^{n+1}) = 0.$$

Proof. We follow the same arguments as in [34] by using the continuous map $\Gamma \circ R_\lambda$ from $(\mathcal{M}, d_1^M \otimes d_1)$ to $(\mathcal{M}, d_1^M \otimes d_1)$. \square

We now give the main convergence result of this section.

Theorem 4.5. *Under Assumptions 2.1, 3.7 and 4.1, for any initialization $(\bar{\xi}^0, \bar{m}^0, \bar{\mu}^0)$, the sequence $(\bar{\xi}^n, \bar{m}^n, \bar{\mu}^n)_n$ defined in Algorithm 1 converges to the unique equilibrium distribution $(\hat{\xi}^\lambda, \hat{m}^\lambda, \hat{\mu}^\lambda)$.*

Proof. The proof is organized in two steps.

Step 1. We first introduce the *exploitability errors* $(\varepsilon_n)_{n \geq 1}$ which are defined as follows:

$$\begin{aligned} \varepsilon_n &:= J^\lambda(\xi^{n+1}, \bar{m}^n, \bar{\mu}^n) - J^\lambda(\bar{\xi}^n, \bar{m}^n, \bar{\mu}^n) \\ (4.3) \quad &= \langle f(\bar{m}^n), m^{n+1} - \bar{m}^n \rangle + \langle g(\bar{\mu}^n), \mu^{n+1} - \bar{\mu}^n \rangle + \lambda \mathbb{E} \left[\int_0^T (\mathcal{E}(\xi_t^{n+1}) - \mathcal{E}(\bar{\xi}_t^n)) dt \right]. \end{aligned}$$

By the optimality of ξ^{n+1} , it is easy to observe that $\varepsilon_n \geq 0$, for all $n \geq 1$.

We write $\varepsilon_{n+1} - \varepsilon_n = \varepsilon_n^{(1)} + \varepsilon_n^{(2)}$, where

$$(4.4) \quad \begin{aligned} \varepsilon_n^{(1)} &:= \langle f(\bar{m}^n), \bar{m}^n \rangle + \langle g(\bar{\mu}^n), \bar{\mu}^n \rangle - \langle f(\bar{m}^{n+1}), \bar{m}^{n+1} \rangle - \langle g(\bar{\mu}^{n+1}), \bar{\mu}^{n+1} \rangle \\ &+ \lambda \mathbb{E} \left[\int_0^T (\mathcal{E}(\bar{\xi}_t^n) - \mathcal{E}(\bar{\xi}_t^{n+1})) dt \right], \end{aligned}$$

$$(4.5) \quad \begin{aligned} \varepsilon_n^{(2)} &:= \langle f(\bar{m}^{n+1}), m^{n+2} \rangle + \langle g(\bar{\mu}^{n+1}), \mu^{n+2} \rangle - \langle f(\bar{m}^n), m^{n+1} \rangle - \langle g(\bar{\mu}^n), \mu^{n+1} \rangle \\ &+ \lambda \mathbb{E} \left[\int_0^T (\mathcal{E}(\xi_t^{n+2}) - \mathcal{E}(\xi_t^{n+1})) dt \right]. \end{aligned}$$

In the following $C > 0$ denotes a constant (independent of n) that may vary from line to line. We first estimate $\varepsilon_n^{(1)}$. First, by the concavity of the entropy functional $\mathcal{E}(\cdot)$, we get

$$(4.6) \quad \begin{aligned} \mathbb{E} \left[\int_0^T (\mathcal{E}(\bar{\xi}_t^n) - \mathcal{E}(\bar{\xi}_t^{n+1})) dt \right] &= \mathbb{E} \left[\int_0^T \left(\mathcal{E}(\bar{\xi}_t^n) - \mathcal{E} \left(\frac{n}{n+1} \bar{\xi}_t^n + \frac{1}{n+1} \xi_t^{n+1} \right) \right) dt \right] \\ &\leq \frac{1}{n+1} \mathbb{E} \left[\int_0^T (\mathcal{E}(\bar{\xi}_t^n) - \mathcal{E}(\xi_t^{n+1})) dt \right]. \end{aligned}$$

Then, by using the estimates from Lemma 4.3, we have

$$\begin{aligned} -\frac{1}{n+1} \langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+1} - \bar{m}^n \rangle &\leq \frac{1}{n+1} \langle |f(\bar{m}^{n+1}) - f(\bar{m}^n)|, m^{n+1} + \bar{m}^n \rangle \\ &\leq \frac{C}{n^2}. \end{aligned}$$

We deduce that

$$\begin{aligned} &\langle f(\bar{m}^n), \bar{m}^n \rangle - \langle f(\bar{m}^{n+1}), \bar{m}^{n+1} \rangle = \langle f(\bar{m}^n), \bar{m}^n \rangle \\ &\quad - \langle f(\bar{m}^{n+1}), \bar{m}^n + \frac{1}{n+1} (m^{n+1} - \bar{m}^n) \rangle \\ &= \langle f(\bar{m}^n) - f(\bar{m}^{n+1}), \bar{m}^n \rangle - \frac{1}{n+1} \langle f(\bar{m}^{n+1}), m^{n+1} - \bar{m}^n \rangle \\ &\leq \langle f(\bar{m}^n) - f(\bar{m}^{n+1}), \bar{m}^n \rangle - \frac{1}{n+1} \langle f(\bar{m}^n), m^{n+1} - \bar{m}^n \rangle + \frac{C}{n^2}. \end{aligned}$$

Similar to the above inequality, we also derive

$$\begin{aligned} &\langle g(\bar{\mu}^n), \bar{\mu}^n \rangle - \langle g(\bar{\mu}^{n+1}), \bar{\mu}^{n+1} \rangle \\ &\leq \langle g(\bar{\mu}^n) - g(\bar{\mu}^{n+1}), \bar{\mu}^n \rangle - \frac{1}{n+1} \langle g(\bar{\mu}^n), \mu^{n+1} - \bar{\mu}^n \rangle + \frac{C}{n^2}. \end{aligned}$$

Therefore, by combining the last three inequalities, we derive

$$\begin{aligned} \varepsilon_n^{(1)} &= \langle f(\bar{m}^n), \bar{m}^n \rangle + \langle g(\bar{\mu}^n), \bar{\mu}^n \rangle - \langle f(\bar{m}^{n+1}), \bar{m}^{n+1} \rangle - \langle g(\bar{\mu}^{n+1}), \bar{\mu}^{n+1} \rangle \\ &\quad + \lambda \mathbb{E} \left[\int_0^T (\mathcal{E}(\bar{\xi}_t^n) - \mathcal{E}(\bar{\xi}_t^{n+1})) dt \right] \\ &\leq \langle f(\bar{m}^n) - f(\bar{m}^{n+1}), \bar{m}^n \rangle + \langle g(\bar{\mu}^n) - g(\bar{\mu}^{n+1}), \bar{\mu}^n \rangle - \frac{\varepsilon_n}{n+1} + \frac{C}{n^2}. \end{aligned}$$

We shall now proceed with the estimation of $\varepsilon_n^{(2)}$. First notice that, since

$$(4.7) \quad J^\lambda(\xi^{n+1}, \bar{m}^n, \bar{\mu}^n) \geq J^\lambda(\xi, \bar{m}^n, \bar{\mu}^n), \quad \forall \xi \in \mathcal{A},$$

it follows that

$$\begin{aligned}
(4.8) \quad & \langle f(\bar{m}^n), m^{n+1} \rangle + \langle g(\bar{\mu}^n), \mu^{n+1} \rangle + \lambda \mathbb{E} \int_0^T \mathcal{E}(\xi_t^{n+1}) dt \\
& \geq \langle f(\bar{m}^n), m^{n+2} \rangle + \langle g(\bar{\mu}^n), \mu^{n+2} \rangle + \lambda \mathbb{E} \int_0^T \mathcal{E}(\xi_t^{n+2}) dt.
\end{aligned}$$

We therefore have

$$\begin{aligned}
\varepsilon_n^{(2)} &= \langle f(\bar{m}^{n+1}), m^{n+2} \rangle + \langle g(\bar{\mu}^{n+1}), \mu^{n+2} \rangle + \lambda \mathbb{E} \left[\int_0^T (\mathcal{E}(\xi_t^{n+2})) \right] \\
&\quad - \langle f(\bar{m}^n), m^{n+1} \rangle - \langle g(\bar{\mu}^n), \mu^{n+1} \rangle - \lambda \mathbb{E} \left[\int_0^T \mathcal{E}(\xi_t^{n+1}) dt \right] \\
&\leq \langle f(\bar{m}^{n+1}), m^{n+2} \rangle + \langle g(\bar{\mu}^{n+1}), \mu^{n+2} \rangle - \langle f(\bar{m}^n), m^{n+2} \rangle - \langle g(\bar{\mu}^n), \mu^{n+2} \rangle \\
&= \langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+2} \rangle + \langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \mu^{n+2} \rangle \\
&= \langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+1} \rangle + \langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \mu^{n+1} \rangle \\
&\quad + \langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+2} - m^{n+1} \rangle + \langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \mu^{n+2} - \mu^{n+1} \rangle.
\end{aligned}$$

By using the estimates of Lemma 4.3, we obtain

$$\begin{aligned}
\langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+2} - m^{n+1} \rangle &\leq \frac{C_f}{n} d_1^M(m^{n+1}, m^{n+2}), \\
\langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \mu^{n+2} - \mu^{n+1} \rangle &\leq \frac{C_g}{n} d_1(\mu^{n+1}, \mu^{n+2}).
\end{aligned}$$

Therefore

$$\begin{aligned}
\varepsilon_n^{(2)} &\leq \langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+1} \rangle + \langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \mu^{n+1} \rangle \\
&\quad + \frac{C}{n} (d_1^M(m^{n+1}, m^{n+2}) + d_1(\mu^{n+1}, \mu^{n+2})).
\end{aligned}$$

We set

$$\delta_n := C \left[d_1^M(m^{n+1}, m^{n+2}) + d_1(\mu^{n+1}, \mu^{n+2}) + \frac{1}{n} \right],$$

and we derive

$$\begin{aligned}
\varepsilon_{n+1} - \varepsilon_n &= \varepsilon_n^{(1)} + \varepsilon_n^{(2)} \leq \langle f(\bar{m}^n) - f(\bar{m}^{n+1}), \bar{m}^n \rangle + \langle g(\bar{\mu}^n) - g(\bar{\mu}^{n+1}), \bar{\mu}^n \rangle - \frac{\varepsilon_n}{n+1} \\
&\quad + \langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+1} \rangle + \langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \mu^{n+1} \rangle + \frac{\delta_n}{n} \\
&= \langle f(\bar{m}^{n+1}) - f(\bar{m}^n), m^{n+1} - \bar{m}^n \rangle + \langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \mu^{n+1} - \bar{\mu}^n \rangle - \frac{\varepsilon_n}{n+1} + \frac{\delta_n}{n} \\
&= (n+1) \left[\langle f(\bar{m}^{n+1}) - f(\bar{m}^n), \bar{m}^{n+1} - \bar{m}^n \rangle + \langle g(\bar{\mu}^{n+1}) - g(\bar{\mu}^n), \bar{\mu}^{n+1} - \bar{\mu}^n \rangle \right] \\
&\quad - \frac{\varepsilon_n}{n+1} + \frac{\delta_n}{n} \\
&\leq -\frac{\varepsilon_n}{n+1} + \frac{\delta_n}{n},
\end{aligned}$$

where the last inequality comes from the Lasry–Lions monotonicity condition. Observe that $\delta_n \rightarrow 0$ by Lemma 4.4. By Lemma 3.1 in [43], we conclude that $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. The result follows.

Step 2. We show now that the sequence $(\xi^n, \bar{m}^n, \bar{\mu}^n)$ converges to the unique λ -SC-MFG equilibrium $(\hat{\xi}^\lambda, \hat{m}^\lambda, \hat{\mu}^\lambda)$. By compactness of the set $\mathcal{A} \times \Gamma(\mathcal{A})$, from any subsequence $(\bar{\xi}^{k_n}, \bar{m}^{k_n}, \bar{\mu}^{k_n})$ we can subtract a further subsequence (not relabeled) such that $(\bar{\xi}^{k_n}, \bar{m}^{k_n}, \bar{\mu}^{k_n})$ converges in the

topology $\tau_2^w \otimes d_p \otimes W_p$ to some $(\bar{\xi}, \bar{m}, \bar{\mu})$, where τ_2^w represents the topology associated to the weak convergence in $\mathbb{L}^2(\Omega \times [0, T])$. First, by using similar arguments as for the proof of (3.9), we can establish that

$$(4.9) \quad (\bar{m}, \bar{\mu}) = \Gamma(\bar{\xi}).$$

Then, by optimality of $\xi^{k_{n+1}}$, we get

$$(4.10) \quad J^\lambda(\xi^{k_{n+1}}, \bar{m}^{k_n}, \bar{\mu}^{k_n}) \geq J^\lambda(\xi, \bar{m}^{k_n}, \bar{\mu}^{k_n}), \text{ for all } \xi \in \mathcal{A}.$$

By definition of ε_{k_n} , we derive

$$(4.11) \quad J^\lambda(\xi^{k_n}, \bar{m}^{k_n}, \bar{\mu}^{k_n}) \geq J^\lambda(\xi, \bar{m}^{k_n}, \bar{\mu}^{k_n}) - \varepsilon_{k_n}, \text{ for all } \xi \in \mathcal{A}.$$

By using similar arguments as the ones used to establish (3.6), (3.7) and (3.8) and by using the convergence of (ε_n) to 0 when $n \rightarrow \infty$, we get

$$(4.12) \quad J^\lambda(\bar{\xi}, \bar{m}, \bar{\mu}) \geq J^\lambda(\xi, \bar{m}, \bar{\mu}), \text{ for all } \xi \in \mathcal{A}.$$

From the above inequality and relation (4.9), we conclude that $(\bar{\xi}, \bar{m}, \bar{\mu})$ is a λ -SC-MFG equilibrium. By uniqueness of the equilibrium, it coincides with $(\hat{\xi}^\lambda, \hat{m}^\lambda, \hat{\mu}^\lambda)$. Since from any subsequence of $(\xi^n, \bar{m}^n, \bar{\mu}^n)$ we can subtract a further subsequence which converges to the unique λ -SC-equilibrium, we conclude that the whole sequence converges to the unique MFG equilibrium. \square

4.2. Fictitious play in the supermodular case. We now provide convergence of the fictitious play algorithm when the so-called supermodularity condition holds true (see, e.g., [27, 29, 64]). Such a property naturally appears in many financial and economic applications, as for example in models for bank runs (see [13] and Example 4.7 below). The interested reader is referred to the textbook [63] for further details on supermodular games.

We thus enforce the following structural condition:

Assumption 4.6 (Supermodularity). *The following hold true:*

- (1) For any $(m, \mu) \in \mathcal{M}$, $R_\lambda(m, \mu)$ is a singleton.
- (2) For each $(t, x) \in [0, T] \times \mathbb{R}^d$, the function $f(t, x, \cdot)$ is nondecreasing with respect to the measure argument in the following sense: for any two measures $m, \bar{m} \in \mathcal{P}_p^{\text{sub}}(\mathbb{R}^d)$, if $m(A) \leq \bar{m}(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$, then

$$f(t, x, m) \leq f(t, x, \bar{m}).$$

- (3) We have $g(x, \mu) = \tilde{g}(x, \langle \psi, \mu \rangle)$, for some functions $\tilde{g} \in \mathcal{C}^2(\mathbb{R}^d \times \mathbb{R})$ and $\psi \in \mathcal{C}^{1,2}([0, T] \times \mathbb{R}^d)$ such that $(\partial_t + \mathcal{L})\psi \geq 0$ and $y \mapsto \mathcal{L}\tilde{g}(x, y)$ is nondecreasing for any $x \in \mathbb{R}^d$.

For the sake of illustration, we discuss some key examples.

Example 4.7. An example in which the supermodularity condition is satisfied is when the function g does not depend on μ (in this case, Condition (3) above is not needed) and $f(t, x, m) = f_0(t, x, 1 - m(\mathbb{R}^d))$, with $f_0(t, x, \cdot)$ nonincreasing for any t, x . This setting represents supermodular MFGs with equilibrium condition

$$\hat{\tau} \in \arg \max_{\tau} \mathbb{E} \left[\int_0^\tau f_0(t, X_t, \hat{m}_t) dt + g(X_\tau) \right] \quad \text{and} \quad \hat{m}_t = \mathbb{P}[\hat{\tau} \leq t],$$

which are also referred to as MFGs of timing. The supermodularity condition is typical of models for bank runs [13]. Natural examples are of additive type $f_0(t, x, y) = f_1(x) - f_2(y)$ or of multiplicative type $f_0(t, x, y) = -f_1(x)f_2(y)$ (with $f_1 \geq 0$), for nondecreasing functions f_1, f_2 .

Example 4.8. *The underlying process evolves as a one dimensional geometric Brownian motion*

$$dZ_t = Z_t(b_0 dt + \sigma_0 dW_t), \quad Z_0 = z_0 > 0, \quad \mathbb{P}\text{-a.s.},$$

with $b_0 \in \mathbb{R}$, $\sigma_0 > 0$, and the payoffs are

$$f(t, z, m) = \int_{\mathbb{R}} (z + y)m(dy) \quad \text{and} \quad g(t, z, \mu) = \int_{\mathbb{R} \times [0, T]} (t + s)\mu(ds, dy).$$

With respect to the previous notation, the state process becomes $X_t := (t, Z_t)$ so that time-dependence in g is allowed.

Before discussing the main result of this subsection, we recall that the controls ξ are interpreted as probability measures; in particular, ξ_t is the probability of stopping before time t . In this sense, if $\xi_t \geq \bar{\xi}_t$ for any $t \in [0, T]$, \mathbb{P} -a.s., then $\tau^\xi \leq \tau^{\bar{\xi}}$ \mathbb{P} -a.s., and we say that ξ is earlier than $\bar{\xi}$ (or that $\bar{\xi}$ is later than ξ). Given two equilibria (ξ, m, μ) and $(\bar{\xi}, \bar{m}, \bar{\mu})$, we say that the first equilibrium is earlier (resp. later) than the second one if ξ is earlier (resp. later) than $\bar{\xi}$. Given $\lambda \geq 0$, the earliest equilibrium $(\underline{\xi}^\lambda, \underline{m}^\lambda, \underline{\mu}^\lambda)$ and the latest equilibrium $(\bar{\xi}^\lambda, \bar{m}^\lambda, \bar{\mu}^\lambda)$ are such that $\underline{\xi}_t^\lambda \geq \xi_t \geq \bar{\xi}_t^\lambda$ for any $t \in [0, T]$, \mathbb{P} -a.s., for any other equilibrium (ξ, m, μ) .

The following theorem discusses the existence and approximation of the earliest and latest equilibria.

Theorem 4.9. *Under Assumptions 2.1 and 4.6, the earliest equilibrium $(\underline{\xi}^\lambda, \underline{m}^\lambda, \underline{\mu}^\lambda)$ and the latest equilibrium $(\bar{\xi}^\lambda, \bar{m}^\lambda, \bar{\mu}^\lambda)$ exist. Moreover, the following statements hold true:*

- (1) *With initialization $\bar{\xi}_t^0 = 1$ for any $t \in [0, T]$, the sequence $\bar{\xi}^n$ is nonincreasing, $\bar{\xi}^{n+1} \leq \bar{\xi}^n$, and $(\bar{\xi}^n, \bar{m}^n, \bar{\mu}^n)$ converges to the earliest equilibrium $(\underline{\xi}^\lambda, \underline{m}^\lambda, \underline{\mu}^\lambda)$.*
- (2) *With initialization $\bar{\xi}_t^0 = 0$ for any $t \in [0, T]$, the sequence $\bar{\xi}^n$ is nondecreasing, $\bar{\xi}^{n+1} \geq \bar{\xi}^n$, and $(\bar{\xi}^n, \bar{m}^n, \bar{\mu}^n)$ converges to the latest equilibrium $(\bar{\xi}^\lambda, \bar{m}^\lambda, \bar{\mu}^\lambda)$.*

Proof. We limit ourself to show the claim on the existence and approximation of the earliest equilibrium, as the proof for the latest equilibrium follows by the same argument. The proof is divided into three steps.

Step 1. In this step we show the some monotonicity properties of the maps Γ and R_λ .

We first show anti-monotonicity property of the consistency map Γ . Observe that, if $\xi_t \leq \xi'_t$, $\forall t \in [0, T]$, \mathbb{P} -a.s., from the definition of $m^\xi, m^{\xi'}$ have

$$m_t^\xi(A) = \mathbb{E}[\mathbb{1}_A(X_t)(1 - \xi_t)] \geq \mathbb{E}[\mathbb{1}_A(X_t)(1 - \xi'_t)] = m_t^{\xi'}(A), \quad \forall A \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0, T].$$

Moreover, for ψ as in Assumption 4.6, by using integration by parts and that $\xi_T = 1$, we have

$$\langle \psi, \mu^\xi \rangle := \mathbb{E} \left[\int_{[0, T]} \psi(t, X_t) d\xi_t \right] = \mathbb{E} \left[\psi(T, X_T) - \int_0^T \xi_t (\partial_t + \mathcal{L}) \psi(t, X_t) dt \right],$$

and the analogous expression can be obtained for ξ' . Thus, since $(\partial_t + \mathcal{L})\psi \geq 0$, we find

$$\begin{aligned} \langle \psi, \mu^\xi \rangle &= \mathbb{E} \left[\psi(T, X_T) - \int_0^T \xi_t (\partial_t + \mathcal{L}) \psi(t, X_t) dt \right] \\ &\geq \mathbb{E} \left[\psi(T, X_T) - \int_0^T \xi'_t (\partial_t + \mathcal{L}) \psi(t, X_t) dt \right] = \langle \psi, \mu^{\xi'} \rangle. \end{aligned}$$

To summarize, we have found that

$$(4.13) \quad \text{if } \xi_t \leq \xi'_t, \forall t, \mathbb{P}\text{-a.s.}, \quad \text{then } m_t^\xi(A) \geq m_t^{\xi'}(A), \forall A, t, \text{ and } \langle \psi, \mu^\xi \rangle \geq \langle \psi, \mu^{\xi'} \rangle,$$

which is the desired anti-monotonicity of Γ .

We next show the anti-monotonicity of the best reply map R_λ ; that is, for any m, m' ,
(4.14) if $m_t(A) \geq m'_t(A)$, $\forall A, t$, and $\langle \psi, \mu \rangle \geq \langle \psi, \mu' \rangle$, then $R_\lambda(m, \mu)_t \leq R_\lambda(m', \mu')_t$, $\forall t$, \mathbb{P} -a.s.

To see this, set $\xi := R_\lambda(m, \mu)$, $\xi' := R_\lambda(m', \mu')$, and define the processes

$$\xi \wedge \xi' := (\min\{\xi_t, \xi'_t\})_t \quad \text{and} \quad \xi \vee \xi' := (\max\{\xi_t, \xi'_t\})_t.$$

For a generic process $\zeta \in \mathcal{A}$, using integration by parts for the integral in $d\zeta$, we find

$$J^\lambda(\zeta, m, \mu) = \mathbb{E} \left[- \int_0^T (f(t, X_t, m_t) + \mathcal{L}g(X_t, \mu)) \zeta_t dt + \int_0^T (f(t, X_t, m_t) + \lambda \mathcal{E}(\zeta_t)) dt + g(X_T, \mu) \right],$$

and, defining $\hat{f}(t, x, m, \mu) = f(t, x, m) + \mathcal{L}g(x, \mu)$, we rewrite the latter expression as

$$J^\lambda(\zeta, m, \mu) = \mathbb{E} \left[- \int_0^T \hat{f}(t, X_t, m_t, \mu) \zeta_t dt + \int_0^T (f(t, X_t, m_t) + \lambda \mathcal{E}(\zeta_t)) dt + g(X_T, \mu) \right].$$

Now, noticing that $\xi'_t - \xi_t \wedge \xi'_t = \xi_t \vee \xi'_t - \xi_t$ and that $\mathcal{E}(\xi_t \vee \xi'_t) - \mathcal{E}(\xi'_t) = \mathcal{E}(\xi_t) - \mathcal{E}(\xi_t \wedge \xi'_t)$, we first find

$$\begin{aligned} J^\lambda(\xi \vee \xi', m', \mu') - J^\lambda(\xi', m', \mu') \\ &= \mathbb{E} \left[\int_0^T (\hat{f}(t, X_t, m'_t, \mu') (\xi'_t - \xi_t \vee \xi'_t) + \lambda (\mathcal{E}(\xi_t \vee \xi'_t) - \mathcal{E}(\xi'_t))) dt \right] \\ &= \mathbb{E} \left[\int_0^T (\hat{f}(t, X_t, m'_t, \mu') (\xi_t \wedge \xi'_t - \xi_t) + \lambda (\mathcal{E}(\xi_t) - \mathcal{E}(\xi_t \wedge \xi'_t))) dt \right]. \end{aligned}$$

Thus, if $m_t(A) \geq m'_t(A)$, $\forall A, t$, and $\langle \psi, \mu \rangle \geq \langle \psi, \mu' \rangle$, from Conditions 2 and 3 in Assumption 4.6 we have

$$\begin{aligned} J^\lambda(\xi \vee \xi', m', \mu') - J^\lambda(\xi', m', \mu') \\ &\geq \mathbb{E} \left[\int_0^T (\hat{f}(t, X_t, m_t, \mu) (\xi'_t \wedge \xi_t - \xi_t) + \lambda ((\mathcal{E}(\xi_t) - \mathcal{E}(\xi_t \wedge \xi'_t)))) dt \right] \\ &= J^\lambda(\xi, m, \mu) - J^\lambda(\xi \wedge \xi', m, \mu). \end{aligned}$$

Hence, from the optimality of ξ' for (m', μ') and the fact that $\xi \vee \xi' \in \mathcal{A}$, we deduce that

$$0 \geq J^\lambda(\xi \vee \xi', m', \mu') - J^\lambda(\xi', m', \mu') \geq J^\lambda(\xi, m, \mu) - J^\lambda(\xi \wedge \xi', m, \mu).$$

This in turn implies that $\xi \wedge \xi' = R_\lambda(m, \mu)$, and, by uniqueness of the optimizer as in Condition 1 in Assumption 4.6, we conclude that $\xi \wedge \xi' = \xi$, thus proving (4.14).

Step 2. In this step we show the monotonicity of $\bar{\xi}^n$ by an induction argument. To simplify the notation, we will write $m \geq m'$ instead of $m_t(A) \geq m'_t(A)$, $\forall A, t$. Moreover, $(m^k, \mu^k) := \Gamma(\xi^k)$, so that (due to the linear dependence of the measure (m, μ) with respect to ξ) we have the relation

$$(\bar{\mu}^n, \bar{m}^n) = \frac{1}{n} \sum_{k=1}^n (\mu^{\xi^k}, m^{\xi^k}) = \frac{1}{n} \sum_{k=1}^n (\mu^k, m^k),$$

that will be used several times in the sequel.

We proceed with an induction argument. Since $\bar{\xi}^0 \equiv 1$ and $(\bar{m}^0, \bar{\mu}^0) = (m^{\xi^0}, \mu^{\xi^0})$, we have

$$\begin{aligned} \xi^1 &= R_\lambda(\bar{m}^0, \bar{\mu}^0) \leq \bar{\xi}^0, \\ (m^1, \mu^1) &= (m^{\xi^1}, \mu^{\xi^1}), \\ \bar{m}^1 &= m^1 = m^{\xi^1} \geq m^{\bar{\xi}^0} = \bar{m}^0, \\ \langle \psi, \bar{\mu}^1 \rangle &= \langle \psi, \mu^1 \rangle = \langle \psi, \mu^{\xi^1} \rangle \geq \langle \psi, \mu^{\bar{\xi}^0} \rangle = \langle \psi, \bar{\mu}^0 \rangle. \end{aligned}$$

Hence, by the monotonicity of R_λ in (4.14) and of Γ in (4.13), we have

$$\begin{aligned}\xi^2 &= R_\lambda(\bar{m}^1, \bar{\mu}^1) \leq R_\lambda(\bar{m}^0, \bar{\mu}^0) = \xi^1, \\ m^2 &= m^{\xi^2} \geq m^{\xi^1} = m^1, \\ \langle \psi, \mu^2 \rangle &= \langle \psi, \mu^{\xi^2} \rangle \geq \langle \psi, \mu^{\xi^1} \rangle = \langle \psi, \mu^1 \rangle, \\ \bar{m}^2 &= \frac{1}{2}(m^2 + m^1) \geq m^1 = \bar{m}^1, \\ \langle \psi, \bar{\mu}^2 \rangle &= \langle \psi, \frac{1}{2}(\mu^2 + \mu^1) \rangle \geq \langle \psi, \mu^1 \rangle \geq \langle \psi, \bar{\mu}^0 \rangle.\end{aligned}$$

Assume that for a generic n we have

$$\begin{aligned}\xi^n &\leq \xi^{n-1} \leq \dots \leq \xi^1, \\ m^n &= m^{\xi^n} \geq m^{n-1} \geq \dots \geq m^1, \\ \langle \psi, \mu^n \rangle &= \langle \psi, \mu^{\xi^n} \rangle \geq \langle \psi, \mu^{n-1} \rangle \geq \dots \geq \langle \psi, \mu^1 \rangle, \\ \bar{m}^n &= \frac{1}{n} \sum_{k=1}^n m^k \geq \bar{m}^{n-1} \geq \dots \geq \bar{m}^1, \\ \langle \psi, \bar{\mu}^n \rangle &= \langle \psi, \frac{1}{n} \sum_{k=1}^n \mu^k \rangle \geq \langle \psi, \bar{\mu}^{n-1} \rangle \geq \dots \geq \langle \psi, \bar{\mu}^1 \rangle.\end{aligned}$$

Then, the monotonicity of R_λ in (4.14) and of Γ in (4.13), again gives

$$\begin{aligned}\xi^{n+1} &= R_\lambda(\bar{m}^n, \bar{\mu}^n) \leq R_\lambda(\bar{m}^{n-1}, \bar{\mu}^{n-1}) = \xi^n, \\ m^{n+1} &= m^{\xi^{n+1}} \geq m^{\xi^n} = m^n \geq \dots \geq m^1, \\ \langle \psi, \mu^{n+1} \rangle &= \langle \psi, \mu^{\xi^{n+1}} \rangle \geq \langle \psi, \mu^{\xi^n} \rangle = \langle \psi, \mu^n \rangle \geq \dots \geq \langle \psi, \mu^1 \rangle.\end{aligned}$$

Moreover, from the last two inequalities we deduce that

$$m^{n+1} \geq \bar{m}^n = \frac{1}{n} \sum_{k=1}^n m^k \quad \text{and} \quad \langle \psi, \mu^{n+1} \rangle \geq \langle \psi, \bar{\mu}^n \rangle = \langle \psi, \frac{1}{n} \sum_{k=1}^n \mu^k \rangle,$$

which allows to conclude that

$$\begin{aligned}\bar{m}^{n+1} &= \frac{1}{n+1}(m^{n+1} + n\bar{m}^n) \geq \bar{m}^n, \\ \langle \psi, \bar{\mu}^{n+1} \rangle &= \langle \psi, \frac{1}{n+1}(\mu^{n+1} + n\bar{\mu}^n) \rangle \geq \langle \psi, \bar{\mu}^n \rangle,\end{aligned}$$

thus completing the induction argument.

Finally, the monotonicity of the sequence $\bar{\xi}^n$ follows from the monotonicity of ξ^n .

Step 3. We next define (thanks to the monotonicity of ξ^n and $\bar{\xi}^n$) the limit points

$$\underline{\xi} := \inf_n \xi^n = \inf_n \bar{\xi}^n, \quad \underline{m} := m^{\underline{\xi}}, \quad \text{and} \quad \underline{\mu} := \mu^{\underline{\xi}},$$

and show that these coincide with the minimal equilibria $(\underline{\xi}^\lambda, \underline{m}^\lambda, \underline{\mu}^\lambda)$.

Notice that the sequence $(\bar{\xi}^n)_n$ converges strongly. Moreover, by monotonicity, it is easy to see that the consistency map Γ is continuous along this sequence. Therefore, we have

$$(\underline{m}, \underline{\mu}) = (m^{\underline{\xi}}, \mu^{\underline{\xi}}) = \Gamma(\underline{\xi}) = \lim_n \Gamma(\bar{\xi}^n) = \lim_n (\bar{m}^n, \bar{\mu}^n),$$

so that the sequence $(\bar{m}^n, \bar{\mu}^n)_n$ converges to $(\underline{m}, \underline{\mu})$ as well. Hence, the continuity of R_λ in turn implies that

$$\underline{\xi} = \lim_n R_\lambda(\bar{m}^n, \bar{\mu}^n) = R_\lambda(\underline{m}, \underline{\mu}),$$

which proves the optimality. This show that $(\underline{\xi}, \underline{m}, \underline{\mu})$ is a mean-field equilibrium.

Finally, we show the minimality of $(\underline{\xi}, \underline{m}, \underline{\mu})$. Let $(\hat{\xi}, \hat{m}, \hat{\mu})$ be another mean-field equilibrium. By definition of $\bar{\xi}^0$, we have $\hat{\xi} \leq \bar{\xi}^0$. Hence, by monotonicity of Γ in (4.13), we have $m^{\hat{\xi}} = \hat{m} \geq m^1 = m^{\xi^1} = \bar{m}^1$ and $\langle \psi, \mu^{\hat{\xi}} \rangle = \langle \psi, \hat{\mu} \rangle \geq \langle \psi, \mu^1 \rangle = \langle \psi, \mu^{\xi^1} \rangle = \langle \psi, \bar{\mu}^1 \rangle$. Hence, using the monotonicity of R_λ so that $\hat{\xi} = R_\lambda(\hat{m}, \hat{\mu}) \leq R_\lambda(\bar{m}^1, \bar{\mu}^1) = \xi^2$ by monotonicity of R_λ . Proceeding by induction, we conclude that

$$\hat{\xi} \leq \bar{\xi}^n, \quad \forall n,$$

so that $\hat{\xi} \leq \underline{\xi} = \inf_n \bar{\xi}^n$, thus giving the desired minimality. \square

APPENDIX A. CONNECTION BETWEEN THE λ -SINGULAR CONTROL MFG PROBLEM AND A λ -OPTIMAL STOPPING MFG PROBLEM

In this part, we present the connection between the singular mean-field problem λ -SC-MFG and a *new* type of OS-MFG problems, that we introduce below and we call λ -OS-MFG (λ -optimal stopping MFG). Let $\bar{\mathcal{T}}$ denote the set of $\mathbb{F}^{x_0, W, U}$ -stopping times with values in $[0, T]$, where the filtration $\mathbb{F}^{x_0, W, U}$ has been introduced in Section 2.3. For ease of exposition, we restrict ourselves to the relative entropy $\mathcal{E}(z) = -z \log z$; the analysis for a general entropy function satisfying Assumption 2.3 is analogous.

Definition A.1. A triple $(\hat{m}^\lambda, \hat{\mu}^\lambda, \hat{\theta}^\lambda)$ is said to be an equilibrium to the λ -OS-MFG problem if:

$$\begin{cases} \hat{\theta}^\lambda \in \arg \max_{\theta \in \bar{\mathcal{T}}} \mathbb{E} \left[\int_0^\theta (f(t, X_t, \hat{m}_t^\lambda) - \lambda(1 + \log U)) dt + g(X_\theta, \hat{\mu}^\lambda) \right] \\ \hat{m}_t^\lambda(A) = \mathbb{P} \left[X_t \in A, t \leq \hat{\theta}^\lambda \right], \quad A \in \mathcal{B}(\mathbb{R}^d) \\ \hat{\mu}^\lambda(B \times [0, t]) = \mathbb{P} \left[X_\theta \in B, \hat{\theta}^\lambda \leq t \right] \quad B \in \mathcal{B}(\mathbb{R}^d). \end{cases}$$

Observe that, for any stopping time $\theta \in \bar{\mathcal{T}}$, since θ is a $\mathbb{F}^{x_0, W, U}$ -stopping time, then there exists a measurable function, still denoted by θ , such that $\theta = \theta(x_0, W, U)$.

We now establish the first relation between the two game problems.

Theorem A.2. Let $(\hat{m}^\lambda, \hat{\mu}^\lambda, \hat{\xi}^\lambda)$ be an equilibrium to the λ -SC-MFG. Define

$$(A.1) \quad \hat{\theta}^\lambda := \theta^\lambda(\cdot, 1 - U),$$

where

$$(A.2) \quad \theta^\lambda(\cdot, u) := \inf\{t \geq 0 \mid \hat{\xi}_t^\lambda \geq u\}.$$

Then $(\hat{m}^\lambda, \hat{\mu}^\lambda, \hat{\theta}^\lambda)$ is an equilibrium to the λ -OS-MFG.

Proof. Let $(\hat{m}^\lambda, \hat{\mu}^\lambda, \hat{\xi}^\lambda)$ be an λ -SC-MFG equilibrium and let $\hat{\theta}^\lambda$ be given by (A.1). Observe that

$$(A.3) \quad \{t < \hat{\theta}^\lambda\} = \{\hat{\xi}_t^\lambda < 1 - U\}.$$

In view of this observation, we obtain:

$$\begin{aligned} \mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} f(t, X_t, \hat{m}_t^\lambda) dt \right] &= \mathbb{E} \left[\int_0^T f(t, X_t, \hat{m}_t^\lambda) \mathbb{1}_{\{t < \hat{\theta}^\lambda\}} dt \right] = \mathbb{E} \left[\int_0^T f(t, X_t, \hat{m}_t^\lambda) \mathbb{P}(U < 1 - \hat{\xi}_t^\lambda | \mathcal{F}_t^{x_0, W}) dt \right] \\ &= \mathbb{E} \left[\int_0^T f(t, X_t, \hat{m}_t^\lambda) (1 - \hat{\xi}_t^\lambda) dt \right] \end{aligned}$$

Compute now

$$\begin{aligned} \mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} (1 + \log(U)) dt \right] &= \mathbb{E} \left[\int_0^T \mathbb{E} \left[(1 + \log(U)) \mathbf{1}_{\{U < 1 - \hat{\xi}_t^\lambda\}} | \mathcal{F}_t^{x_0, W} \right] dt \right] \\ &= \mathbb{E} \left[\int_0^T \left(\int_0^{1 - \hat{\xi}_t^\lambda} (1 + \log(u)) du \right) dt \right] = \mathbb{E} \left[\int_0^T (1 - \hat{\xi}_t^\lambda) \log(1 - \hat{\xi}_t^\lambda) dt \right]. \end{aligned}$$

We get

$$\begin{aligned} &\mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} f(t, X_t, \hat{m}_t^\lambda) dt \right] - \lambda \mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} (1 + \log(U)) dt \right] = \\ (A.4) \quad &\mathbb{E} \left[\int_0^T f(t, X_t, \hat{m}_t^\lambda) (1 - \hat{\xi}_t^\lambda) dt \right] - \lambda \mathbb{E} \left[\int_0^T (1 - \hat{\xi}_t^\lambda) \log(1 - \hat{\xi}_t^\lambda) dt \right]. \end{aligned}$$

By applying the "change of variable" formula from Chapter 0, Proposition 4.9 in [57], we also have

$$(A.5) \quad \mathbb{E}[g(X_{\hat{\theta}^\lambda}, \hat{\mu}^\lambda)] = \mathbb{E} \left[\mathbb{E}[g(X_{\hat{\theta}^\lambda}, \hat{\mu}^\lambda) | \mathcal{F}_T^{x_0, W}] \right] = \mathbb{E} \left[\int_0^1 g(X_{\hat{\theta}^\lambda(\cdot, u)}, \hat{\mu}^\lambda) du \right] = \mathbb{E} \left[\int_0^T g(X_t, \hat{\mu}^\lambda) d\hat{\xi}^\lambda \right].$$

Thus, in view of (A.4) and (A.5), we derive:

$$\begin{aligned} &\mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} f(t, X_t, \hat{m}_t^\lambda) dt \right] - \lambda \mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} (1 + \log(U)) dt \right] + \mathbb{E}[g(X_{\hat{\theta}^\lambda}, \hat{\mu}^\lambda)] = \\ (A.6) \quad &\mathbb{E} \left[\int_0^T f(t, X_t, \hat{m}_t^\lambda) (1 - \hat{\xi}_t^\lambda) dt \right] - \lambda \mathbb{E} \left[\int_0^T (1 - \hat{\xi}_t^\lambda) \log(1 - \hat{\xi}_t^\lambda) dt \right] + \mathbb{E} \left[\int_0^T g(X_t, \hat{\mu}^\lambda) d\hat{\xi}^\lambda \right]. \end{aligned}$$

It remains to show that

$$\begin{aligned} &\mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} f(t, X_t, \hat{m}_t^\lambda) dt \right] - \lambda \mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} (1 + \log(U)) dt \right] + \mathbb{E}[g(X_{\hat{\theta}^\lambda}, \hat{\mu}^\lambda)] \geq \\ (A.7) \quad &\mathbb{E} \left[\int_0^\theta f(t, X_t, \hat{m}_t^\lambda) dt \right] - \lambda \mathbb{E} \left[\int_0^\theta (1 + \log(U)) dt \right] + \mathbb{E}[g(X_\theta, \hat{\mu}^\lambda)], \text{ for all } \theta \in \bar{\mathcal{T}}. \end{aligned}$$

We assume that $\hat{\theta}^\lambda \in \bar{\mathcal{T}}$ is not a maximiser and proceed by contradiction. Let $\bar{\theta} \in \bar{\mathcal{T}}$ be the smallest optimal stopping time. By assumption, we have

$$\begin{aligned} &\mathbb{E} \left[\int_0^{\bar{\theta}} f(t, X_t, m_t) dt \right] - \lambda \mathbb{E} \left[\int_0^{\bar{\theta}} (1 + \log(U)) dt \right] + \mathbb{E}[g(X_{\bar{\theta}}, \mu^\lambda)] \\ (A.8) \quad &> \mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} f(t, X_t, m_t) dt \right] - \lambda \mathbb{E} \left[\int_0^{\hat{\theta}^\lambda} (1 + \log(U)) dt \right] + \mathbb{E}[g(X_{\hat{\theta}^\lambda}, \hat{\mu}^\lambda)]. \end{aligned}$$

Since $\bar{\theta} = \bar{\theta}(x_0, W, U)$ is the smallest optimal stopping time, it follows by similar arguments as in Lemma 1 in [4] that the map $u \rightarrow \bar{\theta}(x_0, W, u)$ is a.s. decreasing, therefore we can define as in (A.1) the right-continuous generalized inverse $\bar{\xi}_t$ as follows:

$$\bar{\xi}_t := \sup\{u \in [0, 1] : \bar{\theta}(x_0, W, 1 - u) \leq t\}.$$

Observe that we have

$$\mathbb{E} \left[\int_0^{\bar{\theta}} \left(f(t, X_t, \hat{m}_t^\lambda) - \lambda(1 + \log U) \right) dt + g(X_{\bar{\theta}}, \hat{\mu}^\lambda) \right]$$

$$= \mathbb{E} \left[\int_0^T \left(f(t, X_t, \hat{m}_t^\lambda) - \lambda(1 + \log U) \right) \mathbb{1}_{\{t < \bar{\theta}\}} dt + g(X_{\bar{\theta}}, \hat{\mu}^\lambda) \right]$$

By definition of $(\bar{\xi}_t)$, it holds that

$$(A.9) \quad \{\bar{\theta}(\cdot, u) \leq t\} = \{\bar{\xi}_t \geq 1 - u\}.$$

In view of the above, and then by conditioning with respect to $\mathbb{F}^{x_0, W}$, using the measurability of $f(t, X_t, \hat{m}_t^\lambda)$ with respect to $\mathbb{F}^{x_0, W}$ and again the "change of variable" formula, we get

$$\begin{aligned} & \mathbb{E} \left[\int_0^T \left(f(t, X_t, \hat{m}_t^\lambda) - \lambda(1 + \log U) \right) \mathbb{1}_{\{t < \bar{\theta}\}} dt + g(X_{\bar{\theta}}, \hat{\mu}^\lambda) \right] \\ &= \mathbb{E} \left[\int_0^T \left(f(t, X_t, \hat{m}_t^\lambda)(1 - \bar{\xi}_t) dt \right) \right] - \mathbb{E} \left[\int_0^T \left(\int_0^1 \lambda(1 + \log u) \right) \mathbb{1}_{\{t < \bar{\theta}(x_0, W, u)\}} du dt + g(X_{\bar{\theta}}, \hat{\mu}^\lambda) \right] \\ &= \mathbb{E} \left[\int_0^T \left(f(t, X_t, \hat{m}_t^\lambda)(1 - \bar{\xi}_t) dt \right) \right] - \mathbb{E} \left[\int_0^T \left(\int_0^1 \lambda(1 + \log u) \right) \mathbb{1}_{\{1-u > \bar{\xi}_t\}} du dt + \int_0^1 g(X_{\bar{\theta}(W, u)}, \hat{\mu}^\lambda) du \right] \\ (A.10) \quad & \leq \mathbb{E} \left[\int_0^T \left(f(t, X_t, \hat{m}_t^\lambda)(1 - \bar{\xi}_t) dt \right) \right] - \mathbb{E} \left[\int_0^T \lambda(1 - \bar{\xi}_t) \log(1 - \bar{\xi}_t) dt \right] + \mathbb{E} \left[\int_0^T g(X_t, \hat{\mu}^\lambda) d\bar{\xi}_t \right]. \end{aligned}$$

From (A.4)-(A.8)-(A.10), it follows that

$$\begin{aligned} & \mathbb{E} \left[\int_0^T \left(f(t, X_t, \hat{m}_t^\lambda)(1 - \bar{\xi}_t) dt \right) \right] - \mathbb{E} \left[\int_0^T \lambda(1 - \bar{\xi}_t) \log(1 - \bar{\xi}_t) dt \right] + \mathbb{E} \left[\int_0^T g(X_t, \hat{\mu}^\lambda) d\bar{\xi}_t \right] \\ (A.11) \quad & > \mathbb{E} \left[\int_0^T f(t, X_t, m_t)(1 - \hat{\xi}_t^\lambda) dt \right] - \lambda \mathbb{E} \left[\int_0^T (1 - \hat{\xi}_t^\lambda) \log(1 - \hat{\xi}_t^\lambda) dt \right] + \mathbb{E} \left[\int_0^T g(X_t, \hat{\mu}^\lambda) d\hat{\xi}_t^\lambda \right], \end{aligned}$$

which contradicts the optimality of $\hat{\xi}^\lambda$. Therefore (A.7) holds true.

By using the same arguments as in the derivation of the consistency conditions (2.6) and (2.7), we also derive that

$$\hat{m}_t^\lambda(A) = \mathbb{P} \left[X_t \in A, t \leq \hat{\theta}^\lambda \right], \quad A \in \mathcal{B}(\mathbb{R}^d)$$

and

$$\hat{\mu}^\lambda(B \times [0, t]) = \mathbb{P} \left[X_{\hat{\theta}^\lambda} \in B, \hat{\theta}^\lambda \leq t \right], \quad B \in \mathcal{B}(\mathbb{R}^d).$$

□

We now present the converse result.

Theorem A.3. *Let $(\hat{m}^\lambda, \hat{\mu}^\lambda, \hat{\theta}^\lambda)$ be an equilibrium to the λ -OS-MFG. Assume that $u \rightarrow \hat{\theta}^\lambda(\cdot, u)$ is a.s. decreasing. Define*

$$(A.12) \quad \hat{\xi}_t^\lambda := \sup\{u \in [0, 1] : \hat{\theta}^\lambda(x_0, W, 1 - u) \leq t\}.$$

Then $(\hat{m}^\lambda, \hat{\mu}^\lambda, \hat{\xi}^\lambda)$ is an equilibrium to λ -SC-MFG.

Proof. The proof follows the same arguments as for the previous result, therefore we omit it. □

Notice that the assumed requirement that $u \rightarrow \hat{\theta}^\lambda(\cdot, u)$ is a.s. decreasing can be easily shown to be valid when $\hat{\theta}^\lambda(\cdot, u)$ is the smallest optimal stopping time for the optimal stopping problem of Definition A.1 (as it was already used in the proof of Theorem A.2).

Acknowledgements. Supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 317210226 - SFB 1283 and the NSF CAREER award DMS-2339240.

REFERENCES

- [1] R. AÏD, R. DUMITRESCU, AND P. TANKOV, *The entry and exit game in the electricity market: a mean-field approach*, Journal of Economic Dynamics & Control, 8 (2021), pp. 331–358.
- [2] B. ANAHTARCI, C. D. KARIKSIZ, AND N. SALDI, *Q-learning in regularized mean-field games*, Dynamic Games and Applications, 13 (2023), pp. 89–117.
- [3] B. ATA, J. M. HARRISON, AND N. SI, *Singular control of (reflected) Brownian motion: A computational method suitable for queueing applications*, arXiv preprint arXiv:2312.11823, (2023).
- [4] F. M. BALDURSSON AND I. KARATZAS, *Irreversible investment and industry equilibrium*, Finance and stochastics, 1 (1996), pp. 69–89.
- [5] A. BASSIÈRE, R. DUMITRESCU, AND P. TANKOV, *A mean-field game model of electricity market dynamics*, in Quantitative Energy Finance: Recent Trends and Developments, Springer, 2024, pp. 181–219.
- [6] C. BAYER, D. BELOMESTNY, P. HAGER, P. PIGATO, AND J. SCHOENMAKERS, *Randomized optimal stopping algorithms and their convergence analysis*, SIAM Journal on Financial Mathematics, 12 (2021), pp. 1201–1225.
- [7] C. BERTUCCI, *Optimal stopping in mean field games, an obstacle problem approach*, J. Math. Pures Appl., 120 (2018), pp. 165–194.
- [8] G. BOUVERET, R. DUMITRESCU, AND P. TANKOV, *Mean-field games of optimal stopping: a relaxed solution approach*, SIAM J. Control Optim., 58 (2020), pp. 1795–1821.
- [9] S. CAMPBELL AND Y. ZHANG, *A mean field game of sequential testing*, arXiv preprint arXiv:2403.18297, (2024).
- [10] ———, *A bayesian sequential soft classification problem for a brownian motion’s drift*, arXiv preprint arXiv:2501.11314, (2025).
- [11] L. CAMPI, T. DE ANGELIS, M. GHIO, AND G. LIVIERI, *Mean-field games of finite-fuel capacity expansion with singular controls*, Ann. Appl. Probab., 32 (2022), pp. 3674–3717.
- [12] R. CARMONA AND F. DELARUE, *Probabilistic Theory of Mean Field Games with Applications I-II*, Springer, 2018.
- [13] R. CARMONA, F. DELARUE, AND D. LACKER, *Mean field games of timing and models for bank runs*, Appl. Math. Optim., 76 (2017), pp. 217–260.
- [14] R. CARMONA AND M. LAURIÈRE, *Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games i: The ergodic case*, SIAM Journal on Numerical Analysis, 59 (2021), pp. 1455–1485.
- [15] R. CARMONA, M. LAURIÈRE, AND Z. TAN, *Model-free mean-field reinforcement learning: mean-field mdp and mean-field q-learning*, The Annals of Applied Probability, 33 (2023), pp. 5334–5381.
- [16] S. CHRISTENSEN, N. DEXHEIMER, AND C. STRAUCH, *Data-driven optimal stopping: A pure exploration analysis*, arXiv preprint arXiv:2312.05880, (2023).
- [17] S. CHRISTENSEN AND C. STRAUCH, *Nonparametric learning for impulse control problems—exploration vs. exploitation*, The Annals of Applied Probability, 33 (2023), pp. 1569–1587.
- [18] S. CHRISTENSEN, C. STRAUCH, AND L. TROTTNER, *Learning to reflect: A unifying approach for data-driven stochastic control strategies*, Bernoulli, 30 (2024), pp. 2074–2101.
- [19] S. CHRISTENSEN, A. H. THOMSEN, AND L. TROTTNER, *Data-driven rules for multidimensional reflection problems*, arXiv preprint arXiv:2311.06639, (2023).
- [20] J. CLAISSE, Z. REN, AND X. TAN, *Mean field games with branching*, The Annals of Applied Probability, 33 (2023), pp. 1034–1075.
- [21] A. COSSO AND L. PERELLI, *Mean field optimal stopping with uncontrolled state*, arXiv preprint arXiv:2503.04269, (2025).
- [22] M. DAI, Y. SUN, Z. Q. XU, AND X. Y. ZHOU, *Learning to optimally stop a diffusion process*, arXiv preprint arXiv:2408.09242, (2024).
- [23] T. DE ANGELIS AND G. STABILE, *On lipschitz continuous optimal stopping boundaries*, SIAM Journal on Control and Optimization, 57 (2019), pp. 402–436.
- [24] H. DENKERT, ROBERT PHAM AND X. WARIN, *Control randomisation approach for policy gradient and application to reinforcement learning in optimal switching*, arXiv preprint arXiv:2404.17939, (2024).
- [25] R. DENKERT AND U. HORST, *Extended mean-field control problems with multi-dimensional singular controls*, arXiv preprint arXiv:2308.04378, (2023).
- [26] R. DEVIDZE, P. KAMALARUBAN, AND A. SINGLA, *Exploration-guided reward shaping for reinforcement learning under sparse rewards*, Advances in Neural Information Processing Systems, 35 (2022), pp. 5829–5842.
- [27] J. DIANETTI, *Strong solutions to submodular mean field games with common noise and related mckean–vlasov fbsdes*, Ann. Appl. Probab., 35 (2025), pp. 1622–1667.
- [28] J. DIANETTI, R. DUMITRESCU, G. FERRARI, AND R. XU, *Reinforcement learning for mean-field games of optimal stopping*, work in progress, (2025).

- [29] J. DIANETTI, G. FERRARI, M. FISCHER, AND M. NENDEL, *A unifying framework for submodular mean field games*, Math. Oper. Res., 48 (2023), pp. 1679–1710.
- [30] J. DIANETTI, G. FERRARI, AND R. XU, *Exploratory optimal stopping: A singular control formulation*, arXiv preprint arXiv:2408.09335, (2024).
- [31] Y. DONG, *Randomized optimal stopping problem in continuous time and reinforcement learning algorithm*, SIAM J. Control Optim., 62 (2024), pp. 1590–1614.
- [32] R. DUMITRESCU, J. GUTIERREZ-PINEDA, AND P. TANKOV, *Fast-slow mean-field games*, 2025.
- [33] R. DUMITRESCU, M. LEUTSCHER, AND P. TANKOV, *Control and optimal stopping mean field games: a linear programming approach*, Electronic Journal of Probability, 26 (2021), pp. 1–49.
- [34] ———, *Linear programming fictitious play algorithm for mean field games with optimal stopping and absorption*, ESAIM Math. Model. Numer. Anal., 57 (2023), pp. 953–990.
- [35] ———, *Energy transition under scenario uncertainty: a mean-field game of stopping with common noise*, Mathematics and Financial Economics, 18 (2024), pp. 233–274.
- [36] D. FIROOZI AND S. JAIMUNGAL, *Exploratory lqg mean field games with entropy regularization*, Automatica, 139 (2022), p. 110177.
- [37] W. H. FLEMING AND H. M. SONER, *Controlled Markov Processes and Viscosity Solutions*, Springer Science & Business Media, 2006.
- [38] G. FU, *Extended mean field games with singular controls*, SIAM J. Control Optim., 61 (2023), pp. 283–312.
- [39] G. FU AND U. HORST, *Mean field games with singular controls*, SIAM J. Control Optim., 55 (2017), pp. 3833–3868.
- [40] X. GUO, A. HU, R. XU, AND J. ZHANG, *Learning mean-field games*, Advances in neural information processing systems, 32 (2019).
- [41] ———, *A general framework for learning mean-field games*, Mathematics of Operations Research, 48 (2023), pp. 656–686.
- [42] X. GUO, R. XU, AND T. ZARIPHOPOULOU, *Entropy regularization for mean field games with learning*, Mathematics of Operations research, 47 (2022), pp. 3239–3260.
- [43] S. HADIKHANLOO, *Learning in anonymous nonatomic games with applications to first-order mean field games*, arXiv preprint arXiv:1704.00378, (2017).
- [44] X. HAN, R. WANG, AND X. Y. ZHOU, *Choquet regularization for continuous-time reinforcement learning*, SIAM Journal on Control and Optimization, 61 (2023), pp. 2777–2801.
- [45] J. HARE, *Dealing with sparse rewards in reinforcement learning*, arXiv preprint arXiv:1910.09281, (2019).
- [46] X. HE, X. TAN, AND J. ZOU, *A mean-field version of bank-el karoui’s representation of stochastic processes*, arXiv preprint arXiv:2302.03300, (2023).
- [47] Y. M. KABANOV, *Hedging and liquidation under transaction costs in currency markets*, Finance Stoch., 3 (1999), pp. 237–248.
- [48] I. KARATZAS AND S. E. SHREVE, *Connections between optimal stopping and singular stochastic control I. Monotone follower problems*, SIAM J. Control Optim., 22 (1984), pp. 856–877.
- [49] I. KARATZAS, S. E. SHREVE, I. KARATZAS, AND S. E. SHREVE, *Methods of mathematical finance*, vol. 39, Springer, 1998.
- [50] J. LI AND G. ŽITKOVIĆ, *Existence, characterization, and approximation in the generalized monotone-follower problem*, SIAM J. Control Optim., 55 (2017), pp. 94–118.
- [51] Z. LIANG, X. LUO, AND X. YU, *A reinforcement learning framework for some singular stochastic control problems*, arXiv preprint arXiv:2506.22203, (2025).
- [52] M. NUTZ, *A mean field game of optimal stopping*, SIAM Journal on Control and Optimization, 56 (2018), pp. 1206–1221.
- [53] D. POSSAMAÏ AND M. TALBI, *Mean-field games of optimal stopping: master equation and weak equilibria*, arXiv preprint arXiv:2307.09278, (2023).
- [54] M. L. PUTERMAN, *Markov decision processes: discrete stochastic dynamic programming*, John Wiley & Sons, 2014.
- [55] M. RAO, Y. CHEN, B. C. VEMURI, AND F. WANG, *Cumulative residual entropy: A new measure of information*, IEEE Trans. Inform. Theory, 50 (2004), pp. 1220–1228.
- [56] A. M. REPPEN, H. M. SONER, AND V. TISSOT-DAGUETTE, *Neural optimal stopping boundary*, arXiv preprint arXiv:2205.04595, (2022).
- [57] D. REVUZ AND M. YOR, *Continuous Martingales and Brownian Motion*, vol. 293, Springer Science & Business Media, 2013.
- [58] H. M. SONER AND V. TISSOT-DAGUETTE, *Stopping times of boundaries: Relaxation and continuity*, arXiv preprint arXiv:2305.09766, (2023).
- [59] P. E. SOUGANIDIS, R. XU, AND Z. THALEIA, *Mean-field information acquisition*, (2025).

- [60] R. S. SUTTON, A. G. BARTO, ET AL., *Reinforcement learning: An introduction*, vol. 1, MIT press Cambridge, 1998.
- [61] M. TALBI, N. TOUZI, AND J. ZHANG, *Dynamic programming equation for the mean field optimal stopping problem*, SIAM Journal on Control and Optimization, 61 (2023), pp. 2140–2164.
- [62] N. TOUZI AND N. VIEILLE, *Continuous-time Dynkin games with mixed strategies*, SIAM J. Control Optim., 41 (2002), pp. 1073–1088.
- [63] X. VIVES, *Oligopoly Pricing: Old Ideas and New Tools*, MIT press, 1999.
- [64] X. VIVES AND O. VRAVOSINOS, *Strategic complementarity in games*, J. Math. Econom., 113 (2024), p. 103005.
- [65] X. YU, J. ZHANG, K. ZHANG, AND Z. ZHOU, *Major-minor mean field game of stopping: An entropy regularization approach*, arXiv preprint arXiv:2501.08770, (2025).