

ROBUST MEAN-FIELD CONTROL UNDER COMMON NOISE UNCERTAINTY

MATHIEU LAURIÈRE, ARIEL NEUFELD, AND KYUNGHYUN PARK

ABSTRACT. We propose and analyze a framework for discrete-time robust mean-field control problems under common noise uncertainty. In this framework, the mean-field interaction describes the collective behavior of infinitely many cooperative agents’ state and action, while the common noise—a random disturbance affecting all agents’ state dynamics—is uncertain. A social planner optimizes over open-loop controls on an infinite horizon to maximize the representative agent’s worst-case expected reward, where worst-case corresponds to the most adverse probability measure among all candidates inducing the unknown true law of the common noise process. We refer to this optimization as a robust mean-field control problem under common noise uncertainty. We first show that this problem arises as the asymptotic limit of a cooperative N -agent robust optimization problem, commonly known as propagation of chaos. We then prove the existence of an optimal open-loop control by linking the robust mean field control problem to a lifted robust Markov decision problem on the space of probability measures and by establishing the dynamic programming principle and Bellman–Isaac fixed point theorem for the lifted robust Markov decision problem. Finally, we complement our theoretical results with numerical experiments motivated by distribution planning and systemic risk in finance, highlighting the advantages of accounting for common noise uncertainty.

1. INTRODUCTION

Mean-field control problems [10, 17], also known as optimal control of McKean–Vlasov dynamics, have emerged as a fundamental framework for optimizing the behavior of large populations of *cooperative* agents. By considering a social planner or central controller managing an infinite (or very large) number of homogeneous agents, mean-field control problems capture a wide range of scenarios including in economics and finance (e.g., [16, 21, 33, 36]), and robotics (e.g., [25, 31, 48, 52]).

One significant extension of the mean-field control paradigm is the inclusion of *common noise*—a random disturbance affecting the dynamics of all agents (e.g., [22, 26, 27, 57, 58, 64]). This feature has become prominent because it captures systemic, correlated randomness (such as macroeconomic shocks or environmental disturbances) that affects the entire population simultaneously. In particular, accounting for common noise enhances the realism of mean-field control problems’ applications in financial engineering, including portfolio optimization, optimal liquidation, or systemic risk (e.g., [2, 18, 62]), as well as in economics, including contract theory or the production of exhaustible resources (e.g., [3, 32, 39]).

However, mean-field control problems with common noise inevitably face a key challenge: *model uncertainty*. When a social planner implements a mean-field control problem with common noise,

Date: November 7, 2025.

Key words: mean-field control, common noise uncertainty, robust optimization, propagation of chaos, Markov decision process, dynamic programming.

Funding: M. Laurière acknowledges the support of the grant “AI-driven Initiative to Promote Research Paradigm Reform and Empower Discipline Advancement.” Computing resources were provided by NYU Shanghai HPC. A. Neufeld acknowledges the support of the MOE AcRF Tier 2 Grant MOE-T2EP20222-0013. K. Park acknowledges the support of the National Research Foundation of Korea (grant DOI: RS-2025-02633175).

it is likely that there is a margin for potential inaccuracies in the model parameters or distributions governing the common noise process. Crucially, because the common noise process affects all agents simultaneously, even small modeling errors in the common noise process can have wide-spread impact across our prediction of the system’s evolution or our computation of the optimal control. This motivates the need for a *robust framework*—also known as the worst-case or Knightian approach (e.g., [23, 29, 37, 38])—in which the social planner seeks an optimal policy that performs robustly under uncertain dynamics of the common noise.

In this article, we aim to propose and analyze a discrete-time robust mean-field control problem under *common noise uncertainty*. The starting point for our problem is based on the two recent works by Carmona et al. [22] and Motte and Pham [57], where infinite time-horizon discounted mean-field control problems with common noise are considered. Both two works establish the correspondence between the conditional McKean–Vlasov dynamics for the representative agent’s state (that typically appear in mean-field control problems with common noise) and the *lifted* Markov decision process on the space of probability measures on the state space. This correspondence enables to articulate dynamic programming Bellman fixed point equations, leading to derive optimal open-loop (and closed-loop Markov) policies for mean-field control problems. Furthermore, [57] establishes the propagation of chaos result which connects the mean-field control problem to a social planner’s optimization problem with a large but finite number of cooperative agents. This ensures that the optimal open-loop policy for the mean-field control problem can be a useful approximation of the optimal policy for such large but finite cooperative agents problems.

Building on [22, 57], we introduce a probabilistic framework for robust mean-field control problems under common noise uncertainty. This framework is designed to encompass both the finite cooperative N -agent system and the conditional McKean–Vlasov dynamics when the common noise distribution is unknown (see Section 2.2). In contrast with the fixed probability measure setting in [22, 57] which induces a single law for the common noise, we construct a set of probability measures, allowing the common noise to have multiple laws within a prescribed uncertainty measures set (see Definition 2.2). This extension is inspired by the robust Markov decision framework of [50, 59, 61], which enables to specify a wide range of different uncertainty sets of probability measures and associated transition kernels.

Using this framework, we establish three main results. First, we prove a propagation of chaos result linking the finite N -agent robust control problem to its mean-field (infinite-agent) counterpart under common noise uncertainty. Under mild regularity conditions on the system and reward functions, we show that the N -agent robust control problem converges to the robust mean-field control problem as $N \rightarrow \infty$ (see Theorem 2.9). This implies that the optimal open-loop policy obtained from the robust mean-field control problem serves as an approximately optimal policy for the finitely many N -agent robust control problem. The proof is based on the Wasserstein convergence rates for empirical measures [14, 35]. In this regard, our propagation of chaos result can be viewed as a robust analog of the results in [57, 58].

Second, we establish a dynamic programming principle for the robust mean-field control problem by *lifting* it to the space of probability measures on the state space. To that end, we show that the conditional McKean–Vlasov state dynamics under common noise uncertainty corresponds to a lifted robust Markov decision process on the space of probability measures (see Proposition 2.12). This correspondence allows us to derive the Bellman–Isaacs fixed-point equations for the value function in the lifted space of distributions. The proof relies on Berge’s maximum theorem to construct local (i.e., one time-step) optimal control and worst-case common noise measure (see Proposition 2.15), and the Banach fixed-point theorem to establish the existence and uniqueness of a fixed point for the Bellman–Isaacs operator (see Proposition 2.16). We then construct an optimal open-loop policy for the robust mean-field control problem by aggregating the local

optimizers (see Theorem 2.21). A crucial toolkit in this construction is the use of an extrinsic randomization source with an atomless distribution (see Assumption 2.18), which also appears in [22]. This randomization not only facilitates the implementation of randomized policies in a decentralized manner but also ensures that each agent’s distribution of controls aligns with the law of optimal policy prescribed by the social planner. While the existence of a randomization source is not explicitly assumed in [57], a randomization hypothesis on the initial information is imposed therein, which in turn induces a structure from which a randomization source naturally exists; see Remark 3.1 therein.

Third, we introduce a closed-loop Markov policy formulation of the robust mean-field control problem. We establish the equivalence between open-loop and closed-loop formulations (Corollary 2.28) and obtain an optimal closed-loop Markov policy. This result can be considered as a robust analog of the main results in [22].

Finally, in order to illustrate all our theoretical results, we provide two numerical examples (see Section 3). In the first example, inspired by Example 1 of [22], the central planner’s goal is to steer the population distribution towards a target distribution. In the second example, inspired by the systemic risk model of [18], the central planner’s goal is to stabilize a financial system and avoid that too many institutions default. In both examples, we underscore the importance and benefits of incorporating common noise uncertainty into mean-field control frameworks.

Related literature. Classic mean-field control problems have been described predominantly in continuous time (see, e.g., [8, 11, 15, 24, 28, 33, 36, 51, 64–66, 68]). Several works [26, 27, 34, 49] have rigorously established the connection between mean-field control and large systems of controlled processes in continuous time settings.

Notably, robust mean-field control problems in continuous-time settings, involving uncertainty in the drift or volatility of the common noise, have been investigated in [45, 69, 70]. The conceptual structure of the arguments in [45] bears certain similarities to ours: in the paper, a centralized control problem under volatility uncertainty of the common noise (analogous to our lifted robust Markov decision problem) is tackled, and then decentralized strategies for the population of agents (analogous to our construction of optimal open-loop policies for the robust mean-field control problem) are obtained. Nevertheless, there are key differences. In particular, the continuous-time works rely on the theory of forward-backward stochastic differential equations, which are not suitable in the discrete-time setting we consider. Instead, our analysis requires a measure-theoretic construction of optimal controls and a derivation of the dynamic programming principle on the space of probability measures. Most notably, while the aforementioned works do not establish a propagation of chaos result, our article provides the first such result under common noise uncertainty.

Several works on mean-field game and control problems have introduced robustness via min–max formulations (e.g., [19, 20, 44, 54, 74]). However, these models do not consider common noise but idiosyncratic noise which is uncertain. In contrast, our framework explicitly accounts for common noise uncertainty, which introduces fundamentally different technical and conceptual challenges. While extending the model to include both idiosyncratic and common noise uncertainty is of clear interest, such an extension leads to significant technical obstacles that invalidate key arguments used in establishing the propagation of chaos result and the lifted dynamic programming principle. This is beyond the scope of the present paper, and we leave it for future work.

Moving away from the above continuous time settings to discrete time settings, some works [40–42, 51, 63] have explored dynamic programming principles for discrete time mean-field control problems, but without considering common noise. More relevant to our setting, recent works—including those we benchmark against [22, 57] and others such as [4, 9, 58]—have investigated

discrete-time mean-field control problems with common noise. Notably, a recent work [50] by two of the authors of the present article proposes a framework for discrete time mean-field Markov games under model uncertainty. In contrast, we focus on a cooperative control setting (as opposed to a game-theoretic equilibrium) and consider the model uncertainty in the law of the common noise process. This leads to a different optimization structure and our lifted dynamic programming formulation on the space of measures is specifically tailored to this social control setting. Furthermore, our propagation of chaos result has no analogue in [50], whose results concern approximate Nash equilibria rather than centralized performance guarantees.

Finally, for completeness, we note that a substantial body of work has focused on robust Markov decision processes under model uncertainty, which also underpin our lifted dynamic programming result on the space of probability measures (see, e.g., [5, 7, 30, 53, 55, 56, 59–61, 71–73] in the optimal control literature, and [43] in the economics literature).

2. MAIN RESULTS

2.1. Notation and preliminaries. Throughout this article, we work with Polish spaces. If X is such a space with corresponding metric d_X , we denote by $\mathcal{B}_X$ its Borel σ -algebra and by $\mathcal{P}(X)$ the set of all Borel probability measures on X . Let $C_b(X; \mathbb{R})$ be the set of all bounded and continuous functions $f : X \rightarrow \mathbb{R}$, endowed with the supremum norm $\|f\|_\infty := \sup_{x \in X} |f(x)|$ where $|\cdot|$ denotes the Euclidean norm. For any $L \geq 0$, we denote by $\text{Lip}_{b,L}(X; \mathbb{R}) \subset C_b(X; \mathbb{R})$ the set of all L -Lipschitz continuous functions.

We equip $\mathcal{P}(X)$ with the topology induced by weak convergence, i.e., for any $\mu \in \mathcal{P}(X)$ and any $(\mu^n)_{n \in \mathbb{N}} \subseteq \mathcal{P}(X)$, we have

$$(2.1) \quad \mu^n \rightharpoonup \mu \text{ as } n \rightarrow \infty \Leftrightarrow \lim_{n \rightarrow \infty} \int_X f(x) \mu^n(dx) = \int_X f(x) \mu(dx) \text{ for any } f \in C_b(X; \mathbb{R}).$$

If X is compact, then the weak topology given in (2.1) is equivalent to the topology induced by the 1-Wasserstein distance $\mathcal{W}_{\mathcal{P}(X)}(\cdot, \cdot)$ which we recall to be the following: For any $\mu, \hat{\mu} \in \mathcal{P}(X)$, denote by $\text{Cpl}_{X \times X}(\mu, \hat{\mu}) \subset \mathcal{P}(X \times X)$ the subset of all couplings with marginals $\mu, \hat{\mu}$. Then the 1-Wasserstein distance between μ and $\hat{\mu}$ is defined by

$$\mathcal{W}_{\mathcal{P}(X)}(\mu, \hat{\mu}) := \inf_{\Gamma \in \text{Cpl}_{X \times X}(\mu, \hat{\mu})} \int_{X \times X} d_X(x, y) \Gamma(dx, dy).$$

For each $t \in \mathbb{N}$, we use the abbreviation $X^t := X \times \cdots \times X$ for the t -times Cartesian product of the set X . Given a sequence $(x_0, \dots, x_t) \in X^{t+1}$ and $0 \leq s \leq t$, we use the following abbreviation $x_{s:t} := (x_s, \dots, x_t)$. Then we endow X^{t+1} with the corresponding product topology induced by the following metric: for every $x_{0:t}, \tilde{x}_{0:t} \in X^{t+1}$,

$$d_{X^{t+1}}(x_{0:t}, \tilde{x}_{0:t}) := \sum_{i=0}^t d_X(x_i, \tilde{x}_i).$$

The same convention applies to a finite Cartesian product of (possibly different) Polish spaces.

For two Polish spaces X and Y , the term ‘kernel’ refers to a Borel measurable map $\lambda : X \ni x \mapsto \lambda(dy|x) \in \mathcal{P}(Y)$. For every $\mu \in \mathcal{P}(X)$ and kernel λ , we write $\mu \hat{\otimes} \lambda \in \mathcal{P}(X \times Y)$ for the measure given by: for every $B \in \mathcal{B}_{X \times Y}$, $\mu \hat{\otimes} \lambda(B) := \int_{X \times Y} \mathbf{1}_{\{(x,y) \in B\}} \lambda(dy|x) \mu(dx)$. Moreover for every $\nu \in \mathcal{P}(Y)$, we write $\mu \otimes \nu \in \mathcal{P}(X \times Y)$ for the product measure.

Finally, given $\mu \in \mathcal{P}(X)$ we use the notation $\mathcal{L}_\mu(\mathcal{Z})$ for the law of a random variable $\mathcal{Z}$ under μ and use $\mathcal{L}_\mu(\mathcal{Z}|\mathcal{Y})$ for the conditional law of $\mathcal{Z}$ given a random variable $\mathcal{Y}$ under μ . The same convention applies to a σ -field. We denote by $\delta_x \in \mathcal{P}(X)$ the Dirac measure at the point $x \in X$.

2.2. Propagation of chaos under common noise uncertainty. In this section, we specify what we mean by the discrete-time N -agent model and mean-field control (MFC) model under common noise uncertainty. We then establish the convergence of the N -agent model to the MFC model as the number of agents N goes to infinity.

To that end, we begin by defining a canonical space for the mean-field models with infinitely many indistinguishable agents.

Denote by G the initial information space and by Θ the randomization source space. Moreover denote by E and E^0 idiosyncratic and common noise spaces, respectively. On the space defined by

$$\Omega := \left\{ \omega := ((g^i)_{i \in \mathbb{N}}, (\theta_t^i)_{t \geq 0, i \in \mathbb{N}}, (e_t^i)_{t \geq 1, i \in \mathbb{N}}, (e_t^0)_{t \geq 1}) : \begin{array}{l} (g^i, \theta_t^i) \in G \times \Theta, \text{ for } t \geq 0, i \in \mathbb{N}; \\ (e_t^i, e_t^0) \in E \times E^0, \text{ for } t \geq 1, i \in \mathbb{N} \end{array} \right\},$$

we denote, for every $\omega \in \Omega$,

$$(2.2) \quad \begin{aligned} (\gamma^i(\omega), \vartheta_0^i(\omega)) &:= (g^i, \theta_0^i) \in G \times \Theta & i \in \mathbb{N}, \\ (\vartheta_t^i(\omega), \varepsilon_t^i(\omega)) &:= (\theta_t^i, e_t^i) \in \Theta \times E & t \geq 1, i \in \mathbb{N}, \\ \varepsilon_t^0(\omega) &:= e_t^0 \in E^0 & t \geq 1, \end{aligned}$$

so that γ^i and $(\vartheta_t^i)_{t \geq 0}$ represent the initial state information of agent $i \in \mathbb{N}$ and her randomization source process, respectively. Moreover, $(\varepsilon_t^i)_{t \geq 1}$ represents her idiosyncratic noise process and $(\varepsilon_t^0)_{t \geq 1}$ represents the common noise process for all agents.

In what follows, we describe a set of probability measures on the space Ω , which captures model uncertainty in the common noise process.

Definition 2.1 (Filtrations). Consider the following filtrations: for each $i \in \mathbb{N}$

- $\mathbb{F}^0 := (\mathcal{F}_t^0)_{t \geq 0}$ is given by $\mathcal{F}_t^0 := \sigma(\varepsilon_{1:t}^0)$ for all $t \geq 1$ with $\mathcal{F}_0^0 = \{\emptyset, \Omega\}$.
- $\mathbb{F}^i := (\mathcal{F}_t^i)_{t \geq 0}$ is given by $\mathcal{F}_0^i := \sigma(\gamma^i)$ and $\mathcal{F}_t^i := \sigma(\gamma^i, \vartheta_{0:t-1}^i, \varepsilon_{1:t}^i, \varepsilon_{1:t}^0)$ for all $t \geq 1$.
- $\mathbb{G}^i := (\mathcal{G}_t^i)_{t \geq 0}$ is given by $\mathcal{G}_t^i := \mathcal{F}_t^i \vee \sigma(\vartheta_t^i)$ for all $t \geq 0$ so that $\mathbb{F}^i \subseteq \mathbb{G}^i$.

Here $\mathcal{F}_t^0$ represents the common noise information shared by all agents at time t . Both $\mathcal{F}_t^i$ and $\mathcal{G}_t^i$ represent the information of agent i at time t , where $\mathcal{G}_t^i$ includes the *current* randomization source ϑ_t^i , while $\mathcal{F}_t^i$ does not.

Definition 2.2 (Measures). Fix $\lambda_\gamma \in \mathcal{P}(G)$, $\lambda_\vartheta \in \mathcal{P}(\Theta)$, and $\lambda_\varepsilon \in \mathcal{P}(E)$.

- (i) Let $\mathfrak{P}^0 \subseteq \mathcal{P}(E^0)$ be a non-empty subset of Borel probability measures on E^0 . Then denote by $\mathcal{K}^0$ the set of all $(p_t)_{t \geq 1}$ consisting of a measure and sequence of kernels such that

$$p_1 \in \mathfrak{P}^0; \quad p_t : (E^0)^{t-1} \ni e_{1:t-1}^0 \mapsto p_t(de_t^0 | e_{1:t-1}^0) \in \mathfrak{P}^0 \quad \text{for all } t \geq 2,$$

inducing *model uncertainty* in the law of the common noise process $(\varepsilon_t^0)_{t \geq 1}$.

- (ii) Denote by $\mathcal{Q} \subseteq \mathcal{P}(\Omega)$ the subset of all Borel probability measures $\mathbb{P}$ on Ω induced by some $(p_t)_{t \geq 1} \in \mathcal{K}^0$ in the sense that for every $B_0 \in \bigvee_{i \in \mathbb{N}} \mathcal{G}_0^i$ and $B_1 \in \bigvee_{i \in \mathbb{N}} \mathcal{G}_1^i$

$$\mathbb{P}\{(\gamma^i, \vartheta_0^i)_{i \in \mathbb{N}} \in B_0\} = \hat{Q}_0(B_0), \quad \mathbb{P}\{((\gamma^i, \vartheta_{0:1}^i, \varepsilon_1^i)_{i \in \mathbb{N}}, \varepsilon_1^0) \in B_1\} = (\hat{Q}_0 \otimes \hat{Q}^{p_1})(B_1),$$

where

$$\hat{Q}_0((dg^i, d\theta_0^i)_{i \in \mathbb{N}}) := \bigotimes_{i \in \mathbb{N}} \{(\lambda_\gamma \otimes \lambda_\vartheta)(dg^i, d\theta_0^i)\} \in \mathcal{P}((G \times \Theta)^{\mathbb{N}})$$

$$\hat{Q}^{p_1}((d\theta_1^i, de_1^i)_{i \in \mathbb{N}}, de_1^0) := \bigotimes_{i \in \mathbb{N}} \{(\lambda_\vartheta \otimes \lambda_\varepsilon)(d\theta_1^i, de_1^i)\} p_1(de_1^0) \in \mathcal{P}((\Theta \times E)^{\mathbb{N}} \times E^0),$$

whereas for every $t \geq 2$ and $B_t \in \bigvee_{i \in \mathbb{N}} \mathcal{G}_t^i$

$$\mathbb{P}\{((\gamma^i, \vartheta_{0:t}^i, \varepsilon_{1:t}^i)_{i \in \mathbb{N}}, \varepsilon_{1:t}^0) \in B_t\} = (\hat{Q}_0 \otimes \hat{Q}^{p_1} \hat{\otimes} \hat{Q}^{p_2} \hat{\otimes} \dots \hat{\otimes} \hat{Q}^{p_t})(B_t),$$

where $\hat{Q}^{p_t} : (E^0)^{t-1} \ni e_{1:t-1}^0 \mapsto \hat{Q}^{p_t}((d\theta_t^i, de_t^i)_{i \in \mathbb{N}}, de_t^0 | e_{1:t-1}^0) \in \mathcal{P}((\Theta \times E)^{\mathbb{N}} \times E^0)$ is defined by

$$\hat{Q}^{p_t}((d\theta_t^i, de_t^i)_{i \in \mathbb{N}}, de_t^0 | e_{1:t-1}^0) := \bigotimes_{i \in \mathbb{N}} \{(\lambda_\vartheta \otimes \lambda_\varepsilon)(d\theta_t^i, de_t^i)\} p_t(de_t^0 | e_{1:t-1}^0).$$

Remark 2.3. By Ionescu–Tulcea’s theorem (see, e.g., [46, Theorem 6.17]), the set $\mathcal{Q}$ given in Definition 2.2 is well-defined and the following hold: for every $\mathbb{P} \in \mathcal{Q}$ w.r.t. some $(p_t)_{t \geq 1} \in \mathcal{K}^0$

- (i) $(\gamma^i)_{i \in \mathbb{N}}$, $(\vartheta_t^i)_{t \geq 0, i \in \mathbb{N}}$, $(\varepsilon_t^i)_{t \geq 1, i \in \mathbb{N}}$, and $(\varepsilon_t^0)_{t \geq 1}$ are mutually independent.
- (ii) $(\gamma^i)_{i \in \mathbb{N}}$ is independent and identically distributed (i.i.d.) with law λ_γ . Moreover, $(\vartheta_t^i)_{t \geq 0, i \in \mathbb{N}}$ is i.i.d. with law λ_ϑ , and $(\varepsilon_t^i)_{t \geq 1, i \in \mathbb{N}}$ is i.i.d. with law λ_ε .
- (iii) ε_1^0 is independent of $\bigvee_{i \in \mathbb{N}} \mathcal{G}_0^i$ with law p_1 , whereas for every $t \geq 2$ ε_t^0 is conditionally independent of $\bigvee_{i \in \mathbb{N}} \mathcal{G}_{t-1}^i$ given $\mathcal{F}_{t-1}^0$ (see [46, Lemma 6.9]), satisfying

$$\mathcal{L}_{\mathbb{P}}(\varepsilon_t^0 | \mathcal{F}_{t-1}^0) = p_t(\cdot | \varepsilon_{1:t-1}^0) \quad \mathbb{P}\text{-a.s.}$$

We note that when $\mathfrak{P}^0$ is a singleton (i.e., *without uncertainty*), the resulting probabilistic framework coincides with the setting in [22, Section 2.1.2] and is also similar to the one in [57, Section 2].

We introduce a dynamical system of mean-field models with indistinguishable N -agents under common noise uncertainty and define the corresponding robust optimization problem. To this end, let us introduce the following elementary components:

Definition 2.4. Let S and A be nonempty compact Polish spaces, representing the state and action spaces, respectively.

- (i) $F : S \times A \times \mathcal{P}(S \times A) \times E \times E^0 \rightarrow S$ is a Borel measurable transition function describing the dynamics of each of the N -agents as well as the mean-field model.
- (ii) $r : S \times A \times \mathcal{P}(S \times A) \rightarrow \mathbb{R}$ is a Borel measurable one-step reward function.
- (iii) $\beta \in [0, 1)$ is a discount factor.

Definition 2.5 (N -agent model). Recall that for each $i \in \mathbb{N}$, $\mathcal{F}_0^i = \sigma(\gamma^i)$ (see Definition 2.1). Denote for every $i \in \mathbb{N}$ by $L_{\mathcal{F}_0^i}^0(S)$ the set of all $\mathcal{F}_0^i$ measurable random variables with values in S .

- (i) Denote by Π the set of all open-loop policies $(\pi_t)_{t \geq 0}$ in the sense that $\pi_t : G \times \Theta^{t+1} \times E^t \times (E^0)^t \rightarrow A$ is a Borel measurable function for all $t \geq 0$. Given $(\pi_t) \in \Pi$, the action process of agent $i \in \mathbb{N}$ is given by the open-loop control

$$a_t^{i,\pi} := \pi_t(\gamma^i, \vartheta_{0:t}^i, \varepsilon_{1:t}^i, \varepsilon_{1:t}^0) \quad t \geq 1, \quad \text{with } a_0^{i,\pi} := \pi_0(\gamma^i, \vartheta_0^i).$$

In other words, $(a_t^{i,\pi})_{t \geq 0}$ is a $\mathbb{G}^i$ adapted process (see Definition 2.1).

- (ii) Fix the initial state $\xi^i \in L_{\mathcal{F}_0^i}^0(S)$ of agent i . Given $N \in \mathbb{N}$ and $(\pi_t)_{t \geq 0} \in \Pi$, the state and action processes of agent $i = 1, \dots, N$ in the N -agent model under $\mathbb{P} \in \mathcal{Q}$ are given by

$$(2.3) \quad \begin{cases} s_0^{i,N,\pi} := \xi^i, \\ s_{t+1}^{i,N,\pi} := F(s_t^{i,N,\pi}, a_t^{i,\pi}, \frac{1}{N} \sum_{j=1}^N \delta_{(s_t^{j,N,\pi}, a_t^{j,\pi})}, \varepsilon_{t+1}^i, \varepsilon_{t+1}^0) \quad t \geq 0. \end{cases}$$

Here we observe that both the law of the initial state and action $(s_0^{i,N,\pi}, a_0^{i,\pi})$ and the law of the idiosyncratic noise process $(\varepsilon_t^i)_{t \geq 0}$ do not depend the choice of $\mathbb{P} \in \mathcal{Q}$ (see Definition 2.2 (iii)). In contrast, the law of $(s_t^{i,N,\pi}, a_t^{i,\pi})$ for $t \geq 1$ depends on this choice, due to the model uncertainty in $(\varepsilon_t^0)_{t \geq 1}$.

- (iii) The contribution of agent i to the social planner’s gain over an infinite horizon under $\mathbb{P} \in \mathcal{Q}$ is defined by

$$R^{i,N,\pi} := \sum_{t=0}^{\infty} \beta^t r(s_t^{i,N,\pi}, a_t^{i,\pi}, \frac{1}{N} \sum_{j=1}^N \delta_{(s_t^{j,N,\pi}, a_t^{j,\pi})}) \quad i = 1, \dots, N.$$

Then the social planner's worst-case expected gain under the common noise uncertainty is

$$\mathcal{J}^{N,\pi} := \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}}[R^{N,\pi}] \quad \text{where} \quad R^{N,\pi} := \frac{1}{N} \sum_{i=1}^N R^{i,N,\pi},$$

and the resulting N -agent optimization problem is given by $V^N := \sup_{\pi \in \Pi} \mathcal{J}^{N,\pi}$. This problem is a robust analog of the classical N -agent optimization problem of [22, 57].

In light of the propagation of chaos argument, we expect and aim to show that the asymptotic version of the N -agent problem in Definition 2.5, as $N \rightarrow \infty$, is given by the following:

Definition 2.6 (MFC model). For each $i \in \mathbb{N}$, let $\xi^i \in L_{\mathcal{F}_0^i}^0(S)$ be the fixed initial state of agent i ; see Definition 2.5 (ii).

- (i) Given $(\pi_t)_{t \geq 0} \in \Pi$, the state process of agent $i \in \mathbb{N}$ in the infinite population model under $\mathbb{P} \in \mathcal{Q}$ is governed by the conditional McKean–Vlasov dynamics:

$$(2.4) \quad \begin{cases} s_0^{i,\pi,\mathbb{P}} := \xi^i, \\ s_{t+1}^{i,\pi,\mathbb{P}} := F(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi}, \mathbb{P}_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}^0, \varepsilon_{t+1}^i, \varepsilon_{t+1}^0) \quad t \geq 0, \end{cases}$$

where $(a_t^{i,\pi})_{t \geq 0}$ is the open-loop control of agent i as defined in Definition 2.5 (i), and $\mathbb{P}_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}^0$ is the conditional joint law of $(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})$ under $\mathbb{P}$ given the common noise trajectory $\varepsilon_{1:t}^0$, i.e.,

$$\mathbb{P}_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}^0 := \mathcal{L}_{\mathbb{P}}((s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi}) | \varepsilon_{1:t}^0) \quad t \geq 1$$

with the convention that $\mathbb{P}_{(s_0^{i,\pi,\mathbb{P}}, a_0^{i,\pi})}^0 := \mathcal{L}_{\mathbb{P}}((s_0^{i,\pi,\mathbb{P}}, a_0^{i,\pi}))$. Analogously, for every $t \geq 1$ let $\mathbb{P}_{s_t^{i,\pi,\mathbb{P}}}^0$ be the conditional law of $s_t^{i,\pi,\mathbb{P}}$ under $\mathbb{P}$ given the common noise trajectory $\varepsilon_{1:t}^0$ with the convention that $\mathbb{P}_{s_0^{i,\pi,\mathbb{P}}}^0 := \mathcal{L}_{\mathbb{P}}(s_0^{i,\pi,\mathbb{P}})$.

- (ii) The contribution of agent i to the social planner's gain under $\mathbb{P} \in \mathcal{Q}$ is defined by

$$R^{i,\pi,\mathbb{P}} := \sum_{t=0}^{\infty} \beta^t r(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi}, \mathbb{P}_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}^0) \quad i \in \mathbb{N}.$$

Then the social planner's worst-case expected gain under the common noise uncertainty is

$$(2.5) \quad \mathcal{J}^{\pi} := \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}}[R^{\pi,\mathbb{P}}], \quad \text{where} \quad R^{\pi,\mathbb{P}} := \mathbb{E}^{\mathbb{P}^0}[R^{i,\pi,\mathbb{P}}] = \mathbb{E}^{\mathbb{P}^0}[R^{1,\pi,\mathbb{P}}] \quad i \in \mathbb{N},$$

where $\mathbb{E}^{\mathbb{P}^0}[\cdot]$ denotes the conditional expectation under $\mathbb{P}$ given $(\varepsilon_t^0)_{t \geq 0}$ and the quantity $R^{\pi,\mathbb{P}}$ is independent of the choice of i due to the indistinguishability of agents. The resulting robust MFC problem is then defined as $V := \sup_{\pi \in \Pi} \mathcal{J}^{\pi}$.

The main goal of this section is to rigorously connect the N -agent model in Definition 2.5 with the MFC model in Definition 2.6.

We impose the following conditions on the basic components given in Definition 2.4.

Assumption 2.7. The following conditions hold:

- (i) There exists some $C_F > 0$ such that for every $s, \tilde{s} \in S$, $a \in A$, $\Lambda, \tilde{\Lambda} \in \mathcal{P}(S \times A)$, and $e^0 \in E^0$

$$\int_E d_S(F(s, a, \Lambda, e, e^0), F(\tilde{s}, a, \tilde{\Lambda}, e, e^0)) \lambda_{\varepsilon}(de) \leq C_F (d_S(s, \tilde{s}) + \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda})),$$

where λ_{ε} is given in Definition 2.5 (i).

(ii) There exists $C_r > 0$ such that for every $s, \tilde{s} \in S$, $a \in A$, and $\Lambda, \tilde{\Lambda} \in \mathcal{P}(S \times A)$

$$|r(s, a, \Lambda)| \leq C_r, \quad |r(s, a, \Lambda) - r(\tilde{s}, a, \tilde{\Lambda})| \leq C_r(d_S(s, \tilde{s}) + \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda})).$$

(iii) β is in $[0, 1 \wedge (2C_F)^{-1})$.

For every $N \in \mathbb{N}$, we define the following quantity

$$(2.6) \quad M_N := \sup_{t \geq 0} \sup_{\pi \in \Pi} \sup_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}} \left[\mathcal{W}_{\mathcal{P}(S \times A)} \left(\frac{1}{N} \sum_{i=1}^N \delta_{(s_t^{i, \pi, \mathbb{P}}, a_t^{i, \pi})}, \mathbb{P}_{(s_t^{1, \pi, \mathbb{P}}, a_t^{1, \pi})}^0 \right) \right],$$

where for each $j = 1, \dots, N$, $(s_t^{j, \pi, \mathbb{P}}, a_t^{j, \pi})_{t \geq 0}$ are the state and action processes of agent j under $\mathbb{P}$ in the MFC model, and for each $t \geq 0$ $\mathbb{P}_{(s_t^{1, \pi, \mathbb{P}}, a_t^{1, \pi})}^0$ is the conditional joint law of $(s_t^{1, \pi, \mathbb{P}}, a_t^{1, \pi})$ under $\mathbb{P}$ given the common noise $\varepsilon_{1:t}^0$ (see Definition 2.6). By the indistinguishability of the N agents, $\mathbb{P}_{(s_t^{1, \pi, \mathbb{P}}, a_t^{1, \pi})}^0$ can equivalently be replaced by $\mathbb{P}_{(s_t^{j, \pi, \mathbb{P}}, a_t^{j, \pi})}^0$ for any $j \in \mathbb{N}$.

The following estimates on the sequence $(M_N)_{N \in \mathbb{N}}$, as defined in (2.6), follow from standard applications of the non asymptotic bounds for the convergence rate of empirical measures in Wasserstein distance (see [35, Theorem 1], [14, Corollary 1.2]).

Lemma 2.8. *Denote by $\Delta_{S \times A} \in [0, \infty)$ the diameter of $S \times A$. Then the following hold:*

(i) *If $S \times A \subset \mathbb{R}^d$ for some $d \in \mathbb{N}$, then for any $q > 2$ there exists some constant $C > 0$ (that depends only on d and q) such that for every $N \in \mathbb{N}$,*

$$M_N \leq C \Delta_{S \times A} \cdot \alpha(N) < \infty,$$

where $\alpha : \mathbb{N} \ni N \mapsto \alpha(N) \in (0, \infty)$ is given as follows: $\alpha(N) := N^{-1/2}$ for $d = 1$; $\alpha(N) := N^{-1/2} \log(1 + N)$ for $d = 2$; $\alpha(N) := N^{-1/d} \log(1 + N)$ for $d \geq 3$.

(ii) *If for every $\delta > 0$ there exist some constants $k_{S \times A} > 0$ and $q > 2$ such that the minimal number of balls with radius δ covering $S \times A$, denoted by $\underline{n}(S \times A, \delta) \in \mathbb{N}$, satisfies $\underline{n}(S \times A, \delta) \leq k_{S \times A} (\Delta_{S \times A} \cdot \delta^{-1})^q$, then there exists some $C > 0$ (that depends only on $k_{S \times A}$ and q) such that for every $N \in \mathbb{N}$,*

$$M_N \leq C \Delta_{S \times A} \cdot N^{-\frac{1}{q}} < \infty.$$

By using Lemma 2.8, we can obtain a rate of convergence when approximating the N -agent model by the MFC model under model uncertainty in the common noise process.

Theorem 2.9. *Suppose that Assumption 2.7 holds. Moreover, we assume that $\Delta_{S \times A}$ satisfies one of the two settings imposed in Lemma 2.8. Then it holds that for every $N \in \mathbb{N}$, $i = 1, \dots, N$, and $t \geq 0$*

$$(2.7) \quad \sup_{\pi \in \Pi} \sup_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}} [d_S(s_t^{i, N, \pi}, s_t^{i, \pi, \mathbb{P}})] = O(M_N),$$

$$(2.8) \quad \sup_{\pi \in \Pi} \sup_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}} \left[\mathcal{W}_{\mathcal{P}(S \times A)} \left(\frac{1}{N} \sum_{j=1}^N \delta_{(s_t^{j, N, \pi}, a_t^{j, \pi})}, \mathbb{P}_{(s_t^{i, \pi, \mathbb{P}}, a_t^{i, \pi})}^0 \right) \right] = O(M_N),$$

where $O(\cdot)$ is the Landau symbol. Moreover, there exists some constant $C > 0$ (that depends only on C_F, C_r and β) such that for $N \in \mathbb{N}$ sufficiently large

$$(2.9) \quad \sup_{\pi \in \Pi} |\mathcal{J}^{N, \pi} - \mathcal{J}^{\pi}| \leq C M_N,$$

which ensures that $|V^N - V| = O(M_N)$. Consequently, any ε -optimal policy for the robust MFC problem V (see Definition 2.6) is $O(\varepsilon)$ -optimal for the N -agent robust optimization problem V^N (see Definition 2.5) if N is sufficiently large such that $M_N = O(\varepsilon)$. Conversely, any ε -optimal policy for V^N is $O(\varepsilon)$ -optimal for V if $N \in \mathbb{N}$ is sufficiently large such that $M_N = O(\varepsilon)$.

The proofs of Lemma 2.8 and Theorem 2.9 can be found in Section 4.

Remark 2.10. Theorem 2.9 can be viewed as a robust analog of [57, Theorem 2.1]. The overall proof roadmap follows the arguments in the reference, where the convergence rate of the empirical measure (see Lemma 2.8) plays a key role. Moreover, the Lipschitz conditions on the one-step reward and system functions in Assumption 2.7 (i), (ii) (denoted as $\mathbf{Hf}_{\text{lip}}$ and $\mathbf{HF}_{\text{lip}}$ therein), together with a certain condition on the discount factor (similar to Assumption 2.7 (iii)), are imposed. While our setting is more rigid due to the uncertainty measures set $\mathcal{Q}$, we are able to obtain the propagation of chaos result by establishing the convergence rate of the empirical measure uniformly over all probability measures $\mathbb{P} \in \mathcal{Q}$.

2.3. Lifted robust Markov decision processes on the space of probability measures. Theorem 2.9 shows that the robust MFC model in Definition 2.6 serves as a macroscopic approximation of the robust N -agent optimization model in Definition 2.5. By definition of the conditional McKean-Vlasov dynamics (2.4) and the social planner's worst-case expected gain (2.5), we can without loss of generality consider only one representative agent.

Accordingly, we suppress the index $i \in \mathbb{N}$ representing individual agents, and denote the representative agent's components as follows: the initial information is given by γ , the randomization source process by $(\vartheta_t)_{t \geq 0}$, the idiosyncratic noise by $(\varepsilon_t)_{t \geq 1}$, and the information processes by

$$(2.10) \quad \begin{aligned} \mathbb{F} &:= (\mathcal{F}_t)_{t \geq 0} \quad \text{with } \mathcal{F}_0 := \sigma(\gamma) \text{ and } \mathcal{F}_t := \sigma(\gamma, \vartheta_{0:t-1}, \varepsilon_{1:t}, \varepsilon_{1:t}^0) \text{ for all } t \geq 1; \\ \mathbb{G} &:= (\mathcal{G}_t)_{t \geq 0} \quad \text{with } \mathcal{G}_t := \mathcal{F}_t \vee \sigma(\vartheta_t) \text{ for all } t \geq 0 \text{ so that } \mathbb{F} \subseteq \mathbb{G}, \end{aligned}$$

see Definition 2.1. The initial state is then given by $\xi \in L_{\mathcal{F}_0}^0(S)$.

Moreover, we define by

$$(2.11) \quad \mathcal{A} := \left\{ a := (a_t)_{t \geq 0} : \begin{array}{l} a \text{ is } \mathbb{G} \text{ adapted and satisfies } a_t = \pi_t(\gamma, \vartheta_{0:t}, \varepsilon_{1:t}, \varepsilon_{1:t}^0) \text{ for } t \geq 1 \\ \text{and } a_0 = \pi_0(\gamma, \vartheta_0) \text{ w.r.t. some } \pi \in \Pi \end{array} \right\},$$

the set of open-loop controls of the representative agent (see Definition 2.5 (i) for the notation Π).

Given $a \in \mathcal{A}$, the state process of the representative agent in the infinite population model under $\mathbb{P} \in \mathcal{Q}$ evolves according to the conditional McKean-Vlasov dynamics:

$$(2.12) \quad s_{t+1}^{\xi, a, \mathbb{P}} := F(s_t^{\xi, a, \mathbb{P}}, a_t, \mathbb{P}_{(s_t^{\xi, a, \mathbb{P}}, a_t)}^0, \varepsilon_{t+1}, \varepsilon_{t+1}^0) \quad \text{for } t \geq 0, \quad \text{with } s_0^{\xi, a, \mathbb{P}} := \xi,$$

where $\mathbb{P}_{(s_t^{\xi, a, \mathbb{P}}, a_t)}^0$ is the conditional joint law of $(s_t^{\xi, a, \mathbb{P}}, a_t)$ under $\mathbb{P}$ given $\varepsilon_{1:t}^0$ for $t \geq 1$, with the convention that $\mathbb{P}_{(s_0^{\xi, a, \mathbb{P}}, a_0)}^0 := \mathcal{L}_{\mathbb{P}}((s_0^{\xi, a, \mathbb{P}}, a_0))$. Here we note that $(s_t^{\xi, a, \mathbb{P}})_{t \geq 0}$ is $\mathbb{F}$ adapted and $(\mathbb{P}_{(s_t^{\xi, a, \mathbb{P}}, a_t)}^0)_{t \geq 0}$ is $\mathbb{F}^0$ adapted (see Lemma 4.1 (ii)).

Then the social planner's worst-case expected gain under the common noise uncertainty is

$$(2.13) \quad \mathcal{J}^a(\xi) := \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}}[R^{a, \mathbb{P}}(\xi)], \quad \text{where } R^{a, \mathbb{P}}(\xi) := \mathbb{E}^{\mathbb{P}^0} \left[\sum_{t=0}^{\infty} \beta^t r(s_t^{\xi, a, \mathbb{P}}, a_t, \mathbb{P}_{(s_t^{\xi, a, \mathbb{P}}, a_t)}^0) \right].$$

Accordingly, the robust MFC problem of the social planner is defined by

$$(2.14) \quad V(\xi) := \sup_{a \in \mathcal{A}} \mathcal{J}^a(\xi), \quad \xi \in L_{\mathcal{F}_0}^0(S).$$

This formulation coincides with Definition 2.6 (by suppressing the agent index i).

We now show how the robust MFC problem given in (2.14) can be *lifted* to a robust Markov decision process (MDP) under model uncertainty in the space of probability measures. Given $\xi \in L_{\mathcal{F}_0}^0(S)$, $a \in \mathcal{A}$, and $\mathbb{P} \in \mathcal{Q}$, we define the following $\mathbb{F}^0$ adapted processes:

$$(2.15) \quad (\mu_t^{\xi, a, \mathbb{P}})_{t \geq 0} := (\mathbb{P}_{s_t^{\xi, a, \mathbb{P}}}^0)_{t \geq 0} \subseteq \mathcal{P}(S),$$

$$(2.16) \quad (\Lambda_t^{\xi, a, \mathbb{P}})_{t \geq 0} := (\mathbb{P}_{(s_t^{\xi, a, \mathbb{P}}, a_t)}^0)_{t \geq 0} \subseteq \mathcal{P}(S \times A).$$

We refer to (2.15) and (2.16) as the lifted state and lifted action processes, respectively. Note that the lifted processes satisfy the following marginal constraint: $\mathbb{P}$ -a.s.,

$$(2.17) \quad \text{pj}_S(\Lambda_t^{\xi, a, \mathbb{P}}) = \mu_t^{\xi, a, \mathbb{P}} \quad \text{for all } t \geq 0,$$

where $\text{pj}_S : \mathcal{P}(S \times A) \ni \Lambda \mapsto \text{pj}_S(\Lambda) := \Lambda(\cdot \times A) \in \mathcal{P}(S)$ denotes the projection function that maps Λ onto its marginal on S .

Based on this observation, we first characterize the dynamics of the lifted state processes. To that end, let us introduce some notation and functions defined on the spaces of probability measure, $\mathcal{P}(S)$ and $\mathcal{P}(S \times A)$ (we refer to them as the ‘lifted’ spaces), which is convenient to characterize the dynamics and then to obtain the lifted dynamic programming principle.

Definition 2.11. Let $\lambda_\varepsilon \in \mathcal{P}(E)$ be given in Definition 2.2. Moreover, let F and r be the transition function and one-step reward function, respectively, as defined in Definition 2.4 (i).

(i) Denote by

$$\mathfrak{U} : \mathcal{P}(S) \ni \mu \mapsto \mathfrak{U}(\mu) := \{\Lambda \in \mathcal{P}(S \times A) : \text{pj}_S(\Lambda) = \mu\} \subseteq \mathcal{P}(S \times A)$$

the correspondence (i.e., a set-valued map) inducing the marginal constraint on S . Moreover, denote by $\text{gr}(\mathfrak{U})$ the graph of $\mathfrak{U}$, i.e., $\text{gr}(\mathfrak{U}) := \{(\mu, \Lambda) \in \mathcal{P}(S) \times \mathcal{P}(S \times A) : \Lambda \in \mathfrak{U}(\mu)\}$.

(ii) Denote by $\bar{F} : \text{gr}(\mathfrak{U}) \times E^0 \ni (\mu, \Lambda, e^0) \mapsto \bar{F}(\mu, \Lambda, e^0) \in \mathcal{P}(S)$ the lifted transition function given by

$$\bar{F}(\mu, \Lambda, e^0)(ds') := ((\Lambda \otimes \lambda_\varepsilon) \circ F(\cdot, \cdot, \Lambda, \cdot, e^0)^{-1})(ds'),$$

i.e., the push-forward of $\Lambda \otimes \lambda_\varepsilon \in \mathcal{P}(S \times A \times E)$ by $F(\cdot, \cdot, \Lambda, \cdot, e^0) : S \times A \times E \rightarrow S$.

(iii) Let $\bar{p} : \text{gr}(\mathfrak{U}) \times \mathcal{P}(E^0) \ni (\mu, \Lambda, p) \mapsto \bar{p}(d\mu' | \mu, \Lambda, p) \in \mathcal{P}(\mathcal{P}(S))$ be a kernel defined by

$$\bar{p}(d\mu' | \mu, \Lambda, p) := (p \circ \bar{F}(\mu, \Lambda, \cdot)^{-1})(d\mu'),$$

i.e., the push-forward of $p \in \mathcal{P}(E^0)$ by $\bar{F}(\mu, \Lambda, \cdot) : E^0 \rightarrow \mathcal{P}(S)$.

(iv) Denote by $\bar{r} : \text{gr}(\mathfrak{U}) \ni (\mu, \Lambda) \mapsto \bar{r}(\mu, \Lambda) \in \mathbb{R}$ the lifted reward function defined by

$$\bar{r}(\mu, \Lambda) := \int_{S \times A} r(s, a, \Lambda) \Lambda(ds, da).$$

The following lemma shows that indeed $(\mu_t^{\xi, a, \mathbb{P}})_{t \geq 0}$ given in (2.15) can be seen as an MDP on the space of probability measures.

Proposition 2.12. Let $\bar{F}$ and $\bar{p}$ be given in Definition 2.11. Let $\xi \in L_{\mathcal{F}_0}^0(S)$, $a \in \mathcal{A}$, and $\mathbb{P} \in \mathcal{Q}$ be given where $\mathbb{P}$ is induced by some couple $(p_t)_{t \geq 1} \in \mathcal{K}^0$ (see Definition 2.2). Then the lifted state and action processes $(\mu_t^{\xi, a, \mathbb{P}})_{t \geq 0}$ and $(\Lambda_t^{\xi, a, \mathbb{P}})_{t \geq 0}$ (see (2.15), (2.16)) satisfy for every $t \geq 0$, $\mathbb{P}$ -a.s.

$$(2.18) \quad \mu_{t+1}^{\xi, a, \mathbb{P}} = \bar{F}(\text{pj}_S(\Lambda_t^{\xi, a, \mathbb{P}}), \Lambda_t^{\xi, a, \mathbb{P}}, \varepsilon_{t+1}^0),$$

which implies that $\mathbb{P}$ -a.s.

$$(2.19) \quad \begin{aligned} \mathcal{L}_{\mathbb{P}}(\mu_1^{\xi, a, \mathbb{P}}) &= \bar{p}(\cdot | \text{pj}_S(\Lambda_0^{\xi, a, \mathbb{P}}), \Lambda_0^{\xi, a, \mathbb{P}}, p_1(\cdot)), \\ \mathcal{L}_{\mathbb{P}}(\mu_{t+1}^{\xi, a, \mathbb{P}}) &= \bar{p}(\cdot | \text{pj}_S(\Lambda_t^{\xi, a, \mathbb{P}}), \Lambda_t^{\xi, a, \mathbb{P}}, p_{t+1}(\cdot | \varepsilon_{1:t}^0)) \quad \text{for all } t \geq 1. \end{aligned}$$

The proof of Proposition 2.12 can be found in Section 5.

Remark 2.13. Let $\xi \in L_{\mathcal{F}_0}^0(S)$, $a \in \mathcal{A}$, and $\mathbb{P} \in \mathcal{Q}$ be given. Note that for every $t \geq 0$,

$$(2.20) \quad \begin{aligned} \mathbb{E}^{\mathbb{P}}[r(s_t^{\xi, a, \mathbb{P}}, a_t, \Lambda_t^{\xi, a, \mathbb{P}})] &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}\left[\int_{S \times A} r(\tilde{s}, \tilde{a}, \Lambda_t^{\xi, a, \mathbb{P}}) \Lambda_t^{\xi, a, \mathbb{P}}(d\tilde{s}, d\tilde{a}) \mid \mathcal{F}_t^0\right]\right] \\ &= \mathbb{E}^{\mathbb{P}}[\bar{r}(\text{pj}_S(\Lambda_t^{\xi, a, \mathbb{P}}), \Lambda_t^{\xi, a, \mathbb{P}})] = \mathbb{E}^{\mathbb{P}}[\bar{r}(\mu_t^{\xi, a, \mathbb{P}}, \Lambda_t^{\xi, a, \mathbb{P}})], \end{aligned}$$

where the first equality holds by $\mathcal{F}_t^0$ -measurability of $\Lambda_t^{\xi, a, \mathbb{P}}$ (see Lemma 4.1 (ii)), the second equality follows from the definition of $\bar{r}$ (see Definition 2.11 (iv)), and the third equality follows from the marginal constraint (2.17).¹

Moreover, since r is bounded and $\beta < 1$ (see Assumption 2.7), by the dominated convergence theorem we can rewrite $\mathcal{J}^a(\xi)$ (given in (2.13)) by

$$(2.21) \quad \mathcal{J}^a(\xi) = \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}}\left[\sum_{t=0}^{\infty} \beta^t \bar{r}(\mu_t^{\xi, a, \mathbb{P}}, \Lambda_t^{\xi, a, \mathbb{P}})\right].$$

Using Proposition 2.12—particularly the MDP given in (2.19) and the representations (2.20) and (2.21) in Remark 2.13—we can view the robust MFC problem (2.14) as a robust MDP with state and action processes $(\mu_t^{\xi, a, \mathbb{P}}, \Lambda_t^{\xi, a, \mathbb{P}})_{t \geq 0}$ given in (2.15) and (2.16). This leads us to consider the following Bellman-Isaacs operator $\mathcal{T}$ defined on $C_b(\mathcal{P}(S); \mathbb{R})$: for every $\bar{V} \in C_b(\mathcal{P}(S); \mathbb{R})$

$$(2.22) \quad \mathcal{T}\bar{V}(\mu) := \sup_{\Lambda \in \mathfrak{U}(\mu)} \left\{ \bar{r}(\mu, \Lambda) + \beta \inf_{p \in \mathfrak{P}^0} \int_{\mathcal{P}(S)} \bar{V}(\mu') \bar{p}(d\mu' | \mu, \Lambda, p) \right\} \quad \mu \in \mathcal{P}(S),$$

where $\mathfrak{P}^0$ is given in Definition 2.2 (i), and $\mathfrak{U}$, $\bar{r}$ and $\bar{p}$ are given in Definition 2.11.

Following the framework of the ‘local to global paradigm’ for robust MDP problems (see, e.g., [50, 60, 61]), we first aim to characterize the local (i.e., one time-step) optimizers of the Bellman-Isaacs operator $\mathcal{T}$, and subsequently establish the fixed point theorem. This will then enable us to construct the global optimizers of the robust MFC problem (2.14).

To that end, we impose the following conditions on the basic components given in Definition 2.4. These conditions are (slightly) stronger than those in Assumption 2.7, as they contain certain regularity on the arguments in A and E^0 along with others on the arguments in S and $\mathcal{P}(S \times A)$. However, they allow us to have some useful properties on the lifted functions and mappings given in Definition 2.11, which are similar to and appear in a framework for robust MDP problems under model uncertainty (see, e.g., [50, 60, 61]).

Assumption 2.14. The following conditions hold:

- (i) The subset $\mathfrak{P}^0$ (see Definition 2.2 (i)) is compact.
- (ii) There is some $\bar{C}_F > 0$ such that² for every $(s, a, \Lambda, e^0), (\tilde{s}, \tilde{a}, \tilde{\Lambda}, \tilde{e}^0) \in S \times A \times \mathcal{P}(S \times A) \times E^0$

$$\int_E d_S(F(s, a, \Lambda, e, e^0), F(\tilde{s}, \tilde{a}, \tilde{\Lambda}, e, \tilde{e}^0)) \lambda_e(de) \leq \bar{C}_F d_{S \times A \times \mathcal{P}(S \times A) \times E^0}((s, a, \Lambda, e^0), (\tilde{s}, \tilde{a}, \tilde{\Lambda}, \tilde{e}^0)).$$
- (iii) The reward function r is Lipschitz continuous, in the sense that there is some $\bar{C}_r > 0$ such that for every $(s, a, \Lambda), (\tilde{s}, \tilde{a}, \tilde{\Lambda}) \in S \times A \times \mathcal{P}(S \times A)$

$$|r(s, a, \Lambda) - r(\tilde{s}, \tilde{a}, \tilde{\Lambda})| \leq \bar{C}_r d_{S \times A \times \mathcal{P}(S \times A)}((s, a, \Lambda), (\tilde{s}, \tilde{a}, \tilde{\Lambda})).$$
- (iv) β is in $[0, 1 \wedge (2\bar{C}_F)^{-1})$.

¹Since $(\mu_t^{\xi, a, \mathbb{P}}, \Lambda_t^{\xi, a, \mathbb{P}}) \in \text{gr}(\mathfrak{U})$, $\mathbb{P}$ -a.s., for all $t \geq 0$, the term $\bar{r}(\mu_t^{\xi, a, \mathbb{P}}, \Lambda_t^{\xi, a, \mathbb{P}})$ is well-defined in the $\mathbb{P}$ -a.s. sense.

²As noted in Section 2.1, the product space $S \times A \times \mathcal{P}(S \times A) \times E^0$ is endowed with the corresponding product topology induced by the following metric: for every $(s, a, \Lambda, e^0), (\tilde{s}, \tilde{a}, \tilde{\Lambda}, \tilde{e}^0) \in S \times A \times \mathcal{P}(S \times A) \times E^0$,

$$d_{S \times A \times \mathcal{P}(S \times A) \times E^0}((s, a, \Lambda, e^0), (\tilde{s}, \tilde{a}, \tilde{\Lambda}, \tilde{e}^0)) := d_S(s, \tilde{s}) + d_A(a, \tilde{a}) + \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}) + d_{E^0}(e^0, \tilde{e}^0).$$

The same convention applies to $S \times A \times \mathcal{P}(S \times A)$ appearing in (iii).

In the following proposition, we characterize the local optimizers of the Bellman-Isaacs operator $\mathcal{T}$ given in (2.22). To that end, we recall that given $L \geq 0$, $\text{Lip}_{b,L}(\mathcal{P}(S); \mathbb{R}) \subset C_b(\mathcal{P}(S); \mathbb{R})$ is the set of all L -Lipschitz continuous functions defined on $\mathcal{P}(S)$.

Proposition 2.15. *Suppose that Assumption 2.14 (i)–(iii) are satisfied. Then the following holds: For every $L \geq 0$ and every $\bar{V} \in \text{Lip}_{b,L}(\mathcal{P}(S); \mathbb{R})$,*

(i) *(Local minimizer) There exists a measurable selector $\bar{p}^* : \mathcal{P}(S \times A) \ni \Lambda \mapsto \bar{p}^*(\Lambda) \in \mathfrak{P}^0$ such that for every $\Lambda \in \mathcal{P}(S \times A)$*

$$(2.23) \quad \int_{\mathcal{P}(S)} \bar{V}(\mu') \bar{p}(d\mu' | \text{pJ}_S(\Lambda), \Lambda, \bar{p}^*(\Lambda)) = \inf_{p \in \mathfrak{P}^0} \int_{\mathcal{P}(S)} \bar{V}(\mu') \bar{p}(d\mu' | \text{pJ}_S(\Lambda), \Lambda, p).$$

(ii) *(Local maximizer) There exists a measurable selector $\bar{\pi}^* : \mathcal{P}(S) \ni \mu \mapsto \bar{\pi}^*(\mu) \in \mathfrak{U}(\mu)$ satisfying that for every $\mu \in \mathcal{P}(S)$*

$$(2.24) \quad \bar{r}(\mu, \bar{\pi}^*(\mu)) + \beta \inf_{p \in \mathfrak{P}^0} \int_{\mathcal{P}(S)} \bar{V}(\mu') \bar{p}(d\mu' | \mu, \bar{\pi}^*(\mu), p) = \mathcal{T}\bar{V}(\mu).$$

We now apply the Banach fixed-point theorem (see, e.g., [6, Theorem A 3.5]) for the Bellman-Isaacs operator $\mathcal{T}$ given in (2.22).

Proposition 2.16. *Suppose that Assumption 2.14 is satisfied, and let $\bar{L} \geq 2\bar{C}_r/(1-2\beta\bar{C}_F)$. Then it holds that $\mathcal{T}(\text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})) \subseteq \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$, and for every $\bar{V}^1, \bar{V}^2 \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$*

$$(2.25) \quad \|\mathcal{T}\bar{V}^1 - \mathcal{T}\bar{V}^2\|_\infty \leq \beta \|\bar{V}^1 - \bar{V}^2\|_\infty.$$

In particular, there exists a unique $\bar{V}^ \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$ satisfying that $\mathcal{T}\bar{V}^* = \bar{V}^*$. Moreover, it holds for every $\bar{V} \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$ that $\bar{V}^* = \lim_{n \rightarrow \infty} \mathcal{T}^n \bar{V}$.*

The proofs of Propositions 2.15 and 2.16 can be found in Section 5.

2.4. Verification theorem. This section aims to establish that the fixed point $\bar{V}^*$ of the Bellman-Isaacs operator $\mathcal{T}$ (see Proposition 2.16) coincides with the robust MFC problem V of the representative agent (see (2.14)) in the sense that³ $V(\xi) = \bar{V}(\mathcal{L}(\xi))$ for all $\xi \in L_{\mathcal{F}_0}^0(S)$.

To that end, we first construct a measure in $\mathcal{Q}$ for each open-loop control in $\mathcal{A}$ (see (2.11)), using the local minimizer from Proposition 2.15 (i). This will later be used in the verification theorem to derive a worst-case measure in $\mathcal{Q}$ by suitably choosing an optimal control in $\mathcal{A}$.

Lemma 2.17. *Suppose that Assumption 2.14 is satisfied. Let $\xi \in L_{\mathcal{F}_0}^0(S)$ be the initial state of the representative agent. Then for every $a \in \mathcal{A}$, there exists $\mathbb{P}^{\xi,a} \in \mathcal{Q}$ induced by some $(p_t^{\xi,a})_{t \geq 1} \in \mathcal{K}^0$ (see Definition 2.2) such that $\mathbb{P}^{\xi,a}$ -a.s.*

$$(2.26) \quad \begin{aligned} \mathcal{L}_{\mathbb{P}^{\xi,a}}(\varepsilon_1^0) &= p_1^{\xi,a} = \bar{p}^*(\underline{\Lambda}_0^{\xi,a}), \\ \mathcal{L}_{\mathbb{P}^{\xi,a}}(\varepsilon_{t+1}^0 | \mathcal{F}_t^0) &= p_{t+1}^{\xi,a}(\cdot | \varepsilon_{1:t}^0) = \bar{p}^*(\underline{\Lambda}_t^{\xi,a}) \quad \text{for all } t \geq 1, \end{aligned}$$

where $\bar{p}^*$ is the local minimizer given in Proposition 2.15 (i), $\underline{\Lambda}_0^{\xi,a}$ is the joint law of $(s_0^{\xi,a,\mathbb{P}^{\xi,a}}, a_0)$ under $\mathbb{P}^{\xi,a}$, and for $t \geq 1$ $\underline{\Lambda}_t^{\xi,a}$ is the conditional joint law of $(s_t^{\xi,a,\mathbb{P}^{\xi,a}}, a_t)$ under $\mathbb{P}^{\xi,a}$ given $\varepsilon_{1:t}^0$. Consequently, we have

$$(2.27) \quad \mathcal{L}_{\mathbb{P}^{\xi,a}}(\mu_{t+1}^{\xi,a}) = \bar{p}(\cdot | \text{pJ}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}, \bar{p}^*(\underline{\Lambda}_t^{\xi,a})), \quad \mathbb{P}^{\xi,a}\text{-a.s.}, \quad \text{for all } t \geq 0,$$

where $\bar{p}$ is given in Definition 2.11, and $\mu_{t+1}^{\xi,a}$ is the conditional law of $s_{t+1}^{\xi,a,\mathbb{P}^{\xi,a}}$ under $\mathbb{P}^{\xi,a}$ given $\varepsilon_{1:t+1}^0$.

³By construction of the set $\mathcal{Q}$ (see Definition 2.2 (ii)), the law of $\xi \in L_{\mathcal{F}_0}^0(S)$ is invariant w.r.t. the choice of supporting probability measure $\mathbb{P} \in \mathcal{Q}$. Therefore, we can and do write $\mathcal{L}(\xi) := \mathcal{L}_{\mathbb{P}}(\xi) \in \mathcal{P}(S)$ for any $\mathbb{P} \in \mathcal{Q}$.

We now construct an open-loop control in $\mathcal{A}$, using the local maximizer from Proposition 2.15 (ii). Then we will verify that this open-loop control is indeed a maximizer of the robust MFC problem given in (2.14).

We impose the following condition.

Assumption 2.18. $\lambda_\vartheta \in \mathcal{P}(\Theta)$ given in Definition 2.2 is atomless.

Remark 2.19. Assumption 2.18 also appears in [22] (see Section 2.1.2). Moreover, [57] incorporates this assumption by assuming the existence of a uniform random variable that is independent of the given initial state (see Section 3 therein). This assumption is crucial for constructing an optimal control/policy from the lifted dynamic programming results presented in both references—and consequently in this article as well. In particular, we often use the following properties.

Since $\mathcal{L}_\mathbb{P}(\vartheta) = \lambda_\vartheta$ for all $\mathbb{P} \in \mathcal{Q}$ (see Remark 2.3 (ii)), Assumption 2.18 implies the existence of a sequence $(h_t)_{t \geq 0}$ of Borel measurable functions $h_t : \Theta \rightarrow [0, 1]$ such that under any $\mathbb{P} \in \mathcal{Q}$,

$$(h_t(\vartheta_t))_{t \geq 0} \text{ is i.i.d. with law } \mathcal{U}_{[0,1]},$$

i.e., uniform distribution on $[0, 1]$; see [13, Theorem 9.2.2]. Since all the agents are indistinguishable, such a sequence exists for each agent $i \in \mathbb{N}$, and we denote it by $(h_t^i)_{t \geq 0}$.

Lemma 2.20. Suppose that Assumptions 2.14 and 2.18 are satisfied. Let $\xi \in L_{\mathcal{F}_0}^0(S)$ be the initial state of the representative agent. Then there exists $a^* \in \mathcal{A}$ such that for every $\mathbb{P} \in \mathcal{Q}$,

$$(2.28) \quad \Lambda_t^{\xi, a^*, \mathbb{P}} = \bar{\pi}^*(\mu_t^{\xi, a^*, \mathbb{P}}), \quad \mathbb{P}\text{-a.s.}, \text{ for all } t \geq 0,$$

where $\bar{\pi}^*$ is the local maximizer given in Proposition 2.15 (ii), and $(\mu_t^{\xi, a^*, \mathbb{P}})_{t \geq 0}$ and $(\Lambda_t^{\xi, a^*, \mathbb{P}})_{t \geq 0}$ are given in (2.15) and (2.16), respectively, under $(a^*, \mathbb{P})$.

We are now ready to state the verification theorem for the constructed open-loop control and probability measure in the preceding two lemmas. The proofs of the theorem and preceding lemmas are provided in Section 6.

Theorem 2.21. Suppose that Assumptions 2.14 and 2.18 are satisfied. Let $\bar{L} \geq 2C_r/(1 - 2\beta C_F)$ be given, and let $\bar{V}^* \in \text{Lip}_{b, \bar{L}}(\mathcal{P}(S); \mathbb{R})$ be such that $\mathcal{T}\bar{V}^* = \bar{V}^*$ (see Proposition 2.16). Moreover, let $a^* \in \mathcal{A}$ be such that (2.28) holds for every $\mathbb{P} \in \mathcal{Q}$ (see Lemma 2.20). Moreover, let $\mathcal{J}^{a^*}$ and V be given in (2.13) and (2.14), respectively. Then, for every $\xi \in L_{\mathcal{F}_0}^0(S)$ the following hold:

- (i) $\bar{V}^*(\mathcal{L}(\xi)) = V(\xi)$, where $\mathcal{L}(\xi) \in \mathcal{P}(S)$ is the law of ξ (see Footnote 3).
- (ii) $a^* \in \mathcal{A}$ and $\mathbb{P}^{\xi, a^*} \in \mathcal{Q}$ induced by $(p_t^{\xi, a^*})_{t \geq 1} \in \mathcal{K}^0$ satisfying (2.26), (2.27) (see Lemma 2.17) are optimal in the sense that

$$(2.29) \quad V(\xi) = \mathcal{J}^{a^*}(\xi) = \mathbb{E}^{\mathbb{P}^{\xi, a^*}} [R^{a^*, \mathbb{P}^{\xi, a^*}}(\xi)].$$

Remark 2.22. As a consequence of Theorems 2.9 and 2.21, under Assumptions 2.14 and 2.18 the optimal open-loop policy $\pi^* \in \Pi$ of the robust MFC problem V (see Definition 2.6)—which can be obtained from the optimal open-loop control $a^* \in \mathcal{A}$ in Theorem 2.21 of the representative robust MFC problem $V(\xi)$ in (2.14)—serves as an approximate of the N -agent optimization problem V^N (see Definition 2.5) when $N \in \mathbb{N}$ is sufficiently large.

Lastly, we note that computing the local optimizers from the lifted dynamic programming principle (given in Proposition 2.15) is crucial for deriving the optimal open-loop control of the robust MFC problem. In particular, this step involves implementation of Q -learning (or policy iteration) algorithms for the lifted dynamic programming principle and analyzing their convergence, together with the discretization error arising from the lifted state and action spaces. While we defer these aspects to future research, in Section 3 we present some numerical examples based on a value iteration type scheme to implement the lifted dynamic programming principle.

2.5. Connection with a closed-loop Markov policy framework. In this section, we introduce the notion of a closed-loop Markov policy for the robust MFC problem. In particular, following [22, Definition 10], we consider a relaxed version of the robust MFC problem in Definition 2.6, in which individual agents are allowed to sample their actions randomly according to a policy specified by the social planner.

As in Sections 2.3 and 2.4, we suppress the index $i \in \mathbb{N}$ representing individual agents and consider the following representation agent's robust MFC problem with closed-loop Markov policies.

Definition 2.23. Let $\mathcal{Q}$ be the uncertainty measures set given in Definition 2.2. Moreover, let $\mathbb{F}, \mathbb{G}$ be the filtrations given in (2.10), and let $\mathbb{F}^0$ be the filtration generated by the common noise.

- (i) Denote by Π^c the set of all closed-loop Markov policies $\pi^c := (\pi_t^c)_{t \geq 0}$ such that for every $t \geq 0$ the kernel

$$\pi_t^c : S \times \mathcal{P}(S) \ni (s, \mu) \mapsto \pi_t^c(da|s, \mu) \in \mathcal{P}(A)$$

induces a randomized action given a couple of a state and a probability measure on S .

- (ii) Let $\xi \in L_{\mathcal{F}_0}^0(S)$ be the fixed initial state. Assume that for any $(\pi^c, \mathbb{P}) \in \Pi^c \times \mathcal{Q}$, the state and action processes $(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})_{t \geq 0}$ for the representative agent in the infinite population model satisfy that⁴ $(s_t^{\xi, \pi^c, \mathbb{P}})_{t \geq 0}$ is $\mathbb{F}$ -adapted, $(a_t^{\pi^c, \mathbb{P}})_{t \geq 0}$ is $\mathbb{G}$ -adapted, and they satisfy

$$(2.30) \quad \begin{aligned} s_{t+1}^{\xi, \pi^c, \mathbb{P}} &:= F(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}}, \mathbb{P}_{(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})}^0, \varepsilon_{t+1}, \varepsilon_{t+1}^0) \quad \text{for } t \geq 0, \quad \text{with } s_0^{\xi, \pi^c, \mathbb{P}} := \xi, \\ \mathcal{L}_{\mathbb{P}}(a_t^{\pi^c, \mathbb{P}} | \mathcal{F}_t) &= \pi_t^c(\cdot | s_t^{\xi, \pi^c, \mathbb{P}}, \mathbb{P}_{(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})}^0) \quad \mathbb{P}\text{-a.s.} \quad \text{for } t \geq 0, \end{aligned}$$

where $\mathbb{P}_{(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})}^0$ is the conditional joint law of $(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})$ under $\mathbb{P}$ given $\varepsilon_{1:t}^0$ for $t \geq 1$, with the convention that $\mathbb{P}_{(s_0^{\xi, \pi^c, \mathbb{P}}, a_0^{\pi^c, \mathbb{P}})}^0 := \mathcal{L}_{\mathbb{P}}((s_0^{\xi, \pi^c, \mathbb{P}}, a_0^{\pi^c, \mathbb{P}}))$. In analogy, $\mathbb{P}_{s_t^{\xi, \pi^c, \mathbb{P}}}^0$ is the conditional law of $s_t^{\xi, \pi^c, \mathbb{P}}$ under $\mathbb{P}$ given $\varepsilon_{1:t}^0$ for $t \geq 1$ with $\mathbb{P}_{s_0^{\xi, \pi^c, \mathbb{P}}}^0 := \mathcal{L}_{\mathbb{P}}(s_0^{\xi, \pi^c, \mathbb{P}})$.

- (iii) Accordingly, the robust MFC problem under closed-loop Markov policies is

$$(2.31) \quad V^c(\xi) := \sup_{\pi^c \in \Pi^c} \mathcal{J}^{\pi^c}(\xi), \quad \xi \in L_{\mathcal{F}_0}^0(S),$$

where $\mathcal{J}^{\pi^c}(\xi)$ is defined as $\mathcal{J}^{\pi^c}(\xi) := \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}}[R^{\pi^c, \mathbb{P}}(\xi)]$ with

$$R^{\pi^c, \mathbb{P}}(\xi) := \mathbb{E}^{\mathbb{P}^0} \left[\sum_{t=0}^{\infty} \beta^t r(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}}, \mathbb{P}_{(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})}^0) \right].$$

Remark 2.24. Under Assumption 2.18, the conditional McKean-Vlasov dynamics with closed-loop Markov policies, as given in Definition 2.23 (ii), are well-defined. Indeed, by using the random variable $h_t(\vartheta_t) \sim \mathcal{U}_{[0,1]}$ (see Remark 2.19) and the Blackwell–Dubins function $\rho_A : \mathcal{P}(A) \times [0, 1] \rightarrow A$ (see Lemma A.2), we can define, for any $\pi^c \in \Pi^c$ and $\mathbb{P} \in \mathcal{Q}$,

$$a_t^{\pi^c, \mathbb{P}} := \rho_A(\pi_t^c(\cdot | s_t^{\xi, \pi^c, \mathbb{P}}, \mathbb{P}_{(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})}^0), h_t(\vartheta_t)) \quad t \geq 0.$$

By the same arguments presented for the proof of Lemma 4.1 (ii), we note that $s_t^{\xi, \pi^c, \mathbb{P}}$ is $\mathcal{F}_t$ measurable and $\mathbb{P}_{s_t^{\xi, \pi^c, \mathbb{P}}}^0$ is $\mathcal{F}_t^0$ measurable. Consequently, $a_t^{\pi^c, \mathbb{P}}$ is $\mathcal{G}_t$ measurable by the construction above. Furthermore, since $\mathcal{F}_t$ is independent of ϑ_t , the property of ρ_A ensures that $a_t^{\pi^c, \mathbb{P}}$ satisfies the distributional constraint given in (2.30).

⁴We refer to Remark 2.24 for the well-posedness of $(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})_{t \geq 0}$ defined as in Definition 2.23 (ii).

We aim to show that the robust MFC problem V^c given in (2.31) coincides with the open-loop robust MFC problem V given in (2.14). This equivalence will be established by demonstrating that $V^c(\xi) = \bar{V}^*(\mathcal{L}(\xi))$ for all $\xi \in L_{\mathcal{F}_0}^0(S)$, where $\bar{V}^*$ is the fixed point of the Bellman–Isaacs operator $\mathcal{T}$ given in Proposition 2.16, and $\mathcal{L}(\xi) \in \mathcal{P}(S)$ is the law of ξ (see Footnote 3).

To this end, and following the approach in Section 2.3, we begin by examining the dynamics of the lifted state and action processes, defined as follows: for every $\pi^c \in \Pi^c$ and $\mathbb{P} \in \mathcal{Q}$,

$$(2.32) \quad \begin{aligned} (\mu_t^{\xi, \pi^c, \mathbb{P}})_{t \geq 0} &:= (\mathbb{P}_{s_t^{\xi, \pi^c, \mathbb{P}}}^0)_{t \geq 0} \subseteq \mathcal{P}(S), \\ (\Lambda_t^{\xi, \pi^c, \mathbb{P}})_{t \geq 0} &:= (\mathbb{P}_{(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})}^0)_{t \geq 0} \subseteq \mathcal{P}(S \times A). \end{aligned}$$

Here we note that both processes are $\mathbb{F}^0$ adapted (see Lemma 4.1).

Lemma 2.25. *Suppose that Assumptions 2.14 and 2.18 are satisfied. Let $\pi^c \in \Pi^c$ be given and let $\mathbb{P} \in \mathcal{Q}$ be induced by some $(p_t)_{t \geq 1} \in \mathcal{K}^0$ (see Definition 2.2). Then,*

$$(2.33) \quad \Lambda_t^{\xi, \pi^c, \mathbb{P}} = \mu_t^{\xi, \pi^c, \mathbb{P}} \hat{\otimes} \pi_t^c(\cdot | \cdot, \mu_t^{\xi, \pi^c, \mathbb{P}}) \quad \mathbb{P}\text{-a.s. for all } t \geq 0.$$

Consequently, it holds that $\mathbb{P}$ -a.s.

$$(2.34) \quad \begin{aligned} \mathcal{L}_{\mathbb{P}}(\mu_1^{\xi, \pi^c, \mathbb{P}}) &= \bar{p}(\cdot | \mu_0^{\xi, \pi^c, \mathbb{P}}, \mu_0^{\xi, \pi^c, \mathbb{P}} \hat{\otimes} \pi_0^c(\cdot | \cdot, \mu_0^{\xi, \pi^c, \mathbb{P}}), p_1(\cdot)), \\ \mathcal{L}_{\mathbb{P}}(\mu_{t+1}^{\xi, \pi^c, \mathbb{P}}) &= \bar{p}(\cdot | \mu_t^{\xi, \pi^c, \mathbb{P}}, \mu_t^{\xi, \pi^c, \mathbb{P}} \hat{\otimes} \pi_t^c(\cdot | \cdot, \mu_t^{\xi, \pi^c, \mathbb{P}}), p_{t+1}(\cdot | \varepsilon_{1:t}^0)) \quad \text{for all } t \geq 1. \end{aligned}$$

Then, as in Lemma 2.17, we construct a measure in $\mathcal{Q}$ for each closed-loop policy in Π^c (see Definition 2.23), using the local minimizer from Proposition 2.15 (i).

Lemma 2.26. *Suppose that Assumptions 2.14 and 2.18 are satisfied. For every $\pi^c \in \Pi^c$, there exists $\underline{\mathbb{P}}^{\xi, \pi^c} \in \mathcal{Q}$ induced by some $(\underline{p}_t^{\xi, \pi^c})_{t \geq 1} \in \mathcal{K}^0$ (see Definition 2.2) such that $\underline{\mathbb{P}}^{\xi, \pi^c}$ -a.s.*

$$(2.35) \quad \begin{aligned} \mathcal{L}_{\underline{\mathbb{P}}^{\xi, \pi^c}}(\varepsilon_1^0) &= \underline{p}_1^{\xi, \pi^c} = \bar{p}^*(\underline{\Lambda}_0^{\xi, \pi^c}), \\ \mathcal{L}_{\underline{\mathbb{P}}^{\xi, \pi^c}}(\varepsilon_{t+1}^0 | \mathcal{F}_t^0) &= \underline{p}_{t+1}^{\xi, \pi^c}(\cdot | \varepsilon_{1:t}^0) = \bar{p}^*(\underline{\Lambda}_t^{\xi, \pi^c}) \quad \text{for all } t \geq 1, \end{aligned}$$

where $\bar{p}^*$ is the local minimizer in Proposition 2.15 (i), $\underline{\Lambda}_0^{\xi, \pi^c}$ is the joint law of $(s_0^{\xi, \pi^c, \underline{\mathbb{P}}^{\xi, \pi^c}}, a_0^{\pi^c, \underline{\mathbb{P}}^{\xi, \pi^c}})$ under $\underline{\mathbb{P}}^{\xi, \pi^c}$, and for $t \geq 1$ $\underline{\Lambda}_t^{\xi, \pi^c}$ is the conditional joint law of $(s_t^{\xi, \pi^c, \underline{\mathbb{P}}^{\xi, \pi^c}}, a_t^{\pi^c, \underline{\mathbb{P}}^{\xi, \pi^c}})$ under $\underline{\mathbb{P}}^{\xi, \pi^c}$ given $\varepsilon_{1:t}^0$. Consequently, we have

$$(2.36) \quad \mathcal{L}_{\underline{\mathbb{P}}^{\xi, \pi^c}}(\mu_{t+1}^{\xi, \pi^c}) = \bar{p}(\cdot | \text{pj}_S(\underline{\Lambda}_t^{\xi, \pi^c}), \underline{\Lambda}_t^{\xi, \pi^c}, \bar{p}^*(\underline{\Lambda}_t^{\xi, \pi^c})), \quad \underline{\mathbb{P}}^{\xi, \pi^c}\text{-a.s., for all } t \geq 0,$$

where $\bar{p}$ is given in Definition 2.11, and $\underline{\mu}_{t+1}^{\xi, \pi^c}$ is the conditional law of $s_{t+1}^{\xi, \pi^c, \underline{\mathbb{P}}^{\xi, \pi^c}}$ given $\varepsilon_{1:t+1}^0$.

The proofs of Lemma 2.25 and 2.26 are presented in Section 7.

Remark 2.27. While the construction of $(\underline{p}_t^{\xi, \pi^c})_{t \geq 1} \in \mathcal{K}^0$ given in Lemma 2.26 proceeds inductively (as in the proof of Lemma 2.17), the arguments differ from those used therein. This is due to the fact that a closed-loop Markov policy $\pi^c \in \Pi^c$ does not determine a fixed action process, but a randomly sampled one. For this, we rely on the Blackwell–Dubins function given in Lemma A.2 together with Remark 2.19 and some measure-theoretic arguments.

Finally, we conclude that the robust MFC problem under the closed-loop Markov policy framework coincides with the fixed point $\bar{V}$, and hence with the robust MFC problem under the open-loop policy framework.

Corollary 2.28. *Suppose that Assumptions 2.14 and 2.18 are satisfied. Let $\bar{L} \geq 2C_r/(1-2\beta C_F)$ be given, and let $\bar{V}^* \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$ be such that $\mathcal{T}\bar{V}^* = \bar{V}^*$ (see Proposition 2.16). Define*

$$(2.37) \quad \pi_{\text{loc}}^{c,*} : S \times \mathcal{P}(S) \ni (s, \mu) \mapsto \pi_{\text{loc}}^{c,*}(\cdot | s, \mu) := \mathcal{K}_{S \times A}(\cdot | s, \bar{\pi}^*(\mu), \mu) \in \mathcal{P}(A),$$

i.e., the universal disintegration kernel of $\bar{\pi}^(\mu)$ w.r.t. $\text{pj}_S(\bar{\pi}^*(\mu)) = \mu$ (see Lemma A.3) so that*

$$(2.38) \quad \bar{\pi}^*(\mu) = \mu \hat{\otimes} \pi_{\text{loc}}^{c,*}(\cdot | \cdot, \mu).$$

Define $\pi^{c,} := (\pi_t^{c,*})_{t \geq 0} \in \Pi^c$ by $\pi_t^{c,*} := \pi_{\text{loc}}^{c,*}$ for every $t \geq 0$ (i.e., stationary closed-loop Markov policy). Moreover, let V^c and $\mathcal{J}^{\pi^{c,*}}$ be given in (2.31), and let V be given in (2.14). Then, for every $\xi \in L_{\mathcal{F}_0}^0(S)$ the following hold:*

- (i) $\bar{V}^*(\mathcal{L}(\xi)) = V^c(\xi) = V(\xi)$, where $\mathcal{L}(\xi) \in \mathcal{P}(S)$ is the law of ξ (see Footnote 3).
- (ii) $\pi^{c,*} \in \Pi^c$ and $\mathbb{P}^{\xi, \pi^{c,*}}$ induced by $(p_t^{\xi, \pi^{c,*}})_{t \geq 1} \in \mathcal{K}^0$ satisfying (2.35), (2.36) (see Lemma 2.26) are optimal in the sense that

$$(2.39) \quad V^c(\xi) = \mathcal{J}^{\pi^{c,*}}(\xi) = \mathbb{E}^{\mathbb{P}^{\xi, \pi^{c,*}}} [R^{\pi^{c,*}, \mathbb{P}^{\xi, \pi^{c,*}}}(\xi)].$$

3. NUMERICAL EXAMPLES

In this section, we apply our robust MFC framework under common noise uncertainty to illustrative examples in distribution matching and financial systemic risk, thereby emphasizing the critical role of incorporating common noise uncertainty into the analysis. In both examples, the algorithm implementing the lifted dynamic programming principle in Proposition 2.15 together with the verification theorem in Theorem 2.21 (or Corollary 2.28) builds upon the value iteration algorithm for the robust MDP framework of [61, Section 4.4.1].

3.1. Example 1: Distribution matching. We first consider an example inspired by Example 1 in [22], in which the goal for the central planner is to make the population distribution match a given target distribution. Common noise makes the task harder because it may randomly shift the distribution.

To be specific, consider the following basic elements (recall Definition 2.4):⁵

- $S = \{1, 2, \dots, |S|\}$ representing a one-dimensional grid world with $|S|$ states; in the experiments, we use $|S| = 7$ states.
- $A = \{-1, 0, 1\}$, where the actions are interpreted respectively as moving to the left, staying or moving to the right.
- $E = \{0\}$, which means that there is no idiosyncratic noise.
- $E^0 = \{-1, 0, 1\}$, where the common noise values are interpreted as the actions but they affect the whole population.
- $F : S \times A \times \mathcal{P}(S \times A) \times E \times E^0 \rightarrow S$ is given by

$$F(s, a, \Lambda, e, e^0) = \max(1, \min(|S|, s + a + e^0)),$$

which represents the fact that the agent's movement is determined by her action and the common noise, and the agent remains at 1 (resp. 7) if she tries to move to the left (resp. right) of this state.

- $r : S \times A \times \mathcal{P}(S \times A) \rightarrow \mathbb{R}$ is given by

$$r(s, a, \Lambda) = \|\text{pj}_S(\Lambda) - \mu^*\|_2^2 = \sum_{s \in S} |\text{pj}_S(\Lambda)(s) - \mu^*(s)|^2,$$

where $\mu^* \in \mathcal{P}(S)$ is a fixed target distribution which is part of the model's definition.

- $\beta = 0.4$ is the discount factor so that Assumptions 2.7 (iii) and 2.14 (iv) are satisfied.

⁵The code is provided for the sake of completeness at <https://github.com/mlauriere/RobustMFMDP>.

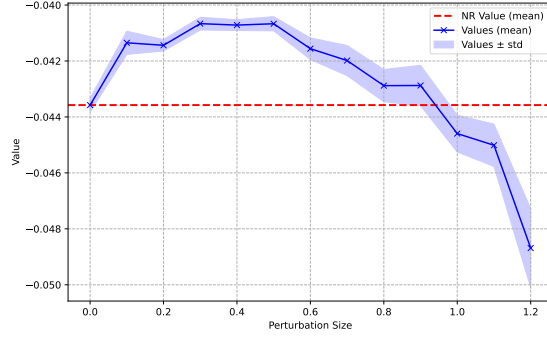


FIGURE 1. Values achieved under p_{true} when using the optimal policy for the MFC under p_{ref} (red dashed line) or the robust MFC under the uncertainty level $\delta_{\text{perturb}} \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2\}$ (blue curve) in Example 1. Shaded areas represents $\pm$ standard deviation over 8 independent runs.

For the common noise probability measure, we consider the following situation:

- The true common noise distribution $p_{\text{true}} \in \mathcal{P}(E^0)$ is given by

$$(3.1) \quad p_{\text{true}} := v_{\text{true},1}\delta_{\{\varepsilon^0=-1\}} + v_{\text{true},2}\delta_{\{\varepsilon^0=0\}} + v_{\text{true},3}\delta_{\{\varepsilon^0=1\}},$$

with some probability vector $v_{\text{true}} := (v_{\text{true},1}, v_{\text{true},2}, v_{\text{true},3}) \in [0, 1]^3$, i.e., a simplex.

- However, we consider that the central planner does not know this true distribution; she has estimated the common noise distribution to be approximately equal to a reference probability measure $p_{\text{ref}} \in \mathcal{P}(E^0)$ with the corresponding probability vector $v_{\text{ref}} \in [0, 1]^3$.

As a baseline, the central planner can learn a policy π_{ref} which is optimal for the MFC model with common noise distribution p_{ref} . Alternatively, she can solve the robust MFC problem and learn a policy π_{robust} which may be suboptimal for the model with p_{ref} but which performs better than π_{ref} in the true model with common noise distribution p_{true} .

We consider the uncertainty set $\mathfrak{P}^0$ which consists of all perturbed measures $p \in \mathcal{P}(E^0)$ of the reference measure p_{ref} , whose corresponding probability vector $v \in [0, 1]^3$ is

$$(3.2) \quad v := \text{renorm}(\max(0, v_{\text{ref}} + v_{\text{perturb}})),$$

where $v_{\text{perturb}} \in \mathbb{R}^3$ is a perturbation vector constructed as follows: each coordinate is sampled uniformly from $[-\delta_{\text{perturb}}, \delta_{\text{perturb}}]$, with a small $\delta_{\text{perturb}} > 0$ representing the uncertainty level. The average of the 3 coordinates is then subtracted to each coordinate to ensure that the average of v_{perturb} over coordinates is 0. Under this construction, Assumption 2.14 (i) is satisfied.

We implement the above model with: $v_{\text{true}} = (0.2, 0.7, 0.1)$, $v_{\text{ref}} = (0, 1, 0)$ and δ_{perturb} varying between 0.0 and 0.8. Figure 1 shows that for moderately small δ_{perturb} , the robust policy performs better than the non-robust policy. For large values of δ_{perturb} however, the robust policy yields a smaller value: being robust against a large set of possible common noise distributions prevents the policy from performing well on the true distribution. The results are averaged over 8 different runs and the plots shows the mean value and its standard deviation.

Figure 2 shows three realizations of trajectories, starting from random initial distributions. We display a few time steps between 0 and 20. We observe that the learnt policy uses the actions with varying proportions depending on the individual state and also depending on the current population distribution. Overall, it uses mostly action 1 (resp. -1) when the state is below (resp. above) the middle state because the target distribution is centered around the middle state.

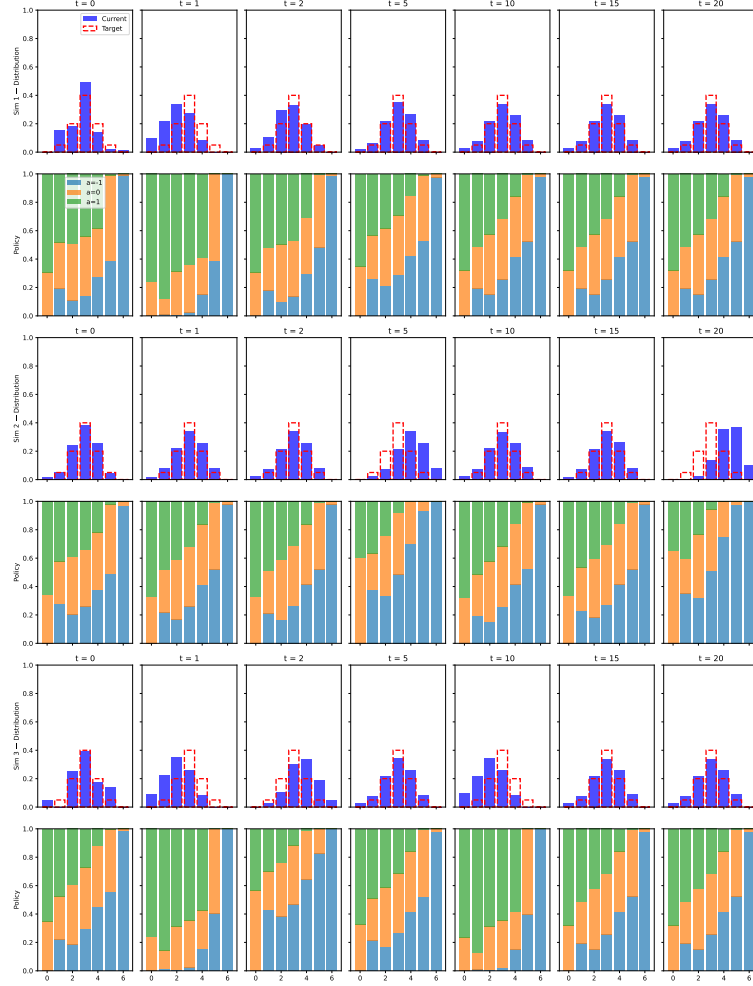


FIGURE 2. Three sample trajectories of the population distribution and corresponding action distribution for each state in Example 1. The target distribution to be matched is shown by dashed red lines.

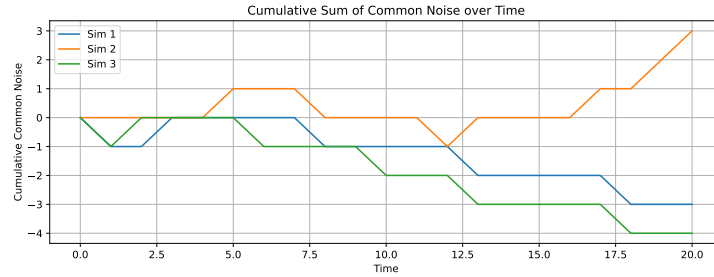


FIGURE 3. The three trajectories of common noise associated with Figure 2

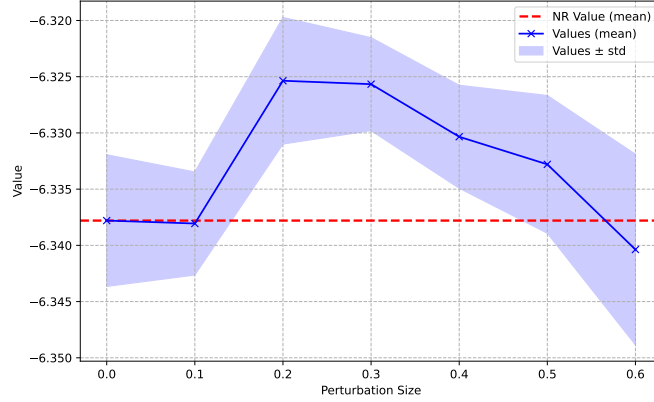


FIGURE 4. Value achieved under p_{true} when using the optimal policy for the MFC with p_{ref} (red dashed line) or the optimal policy for the robust MFC with $\delta_{\text{perturb}} \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ (blue curve) in Example 2. Shaded areas represents $\pm$ standard deviation over 8 independent runs.

The fact that the target distribution is not perfectly matched is due to the impact of the common noise, whose trajectories are displayed in Figure 3. Notice that for the second simulation, the common noise takes several positive values on time steps 17, 18 and 19, leaving little time for the population distribution to adapt and shift back to the target distribution (recall that the possible actions are $\{-1, 0, 1\}$, just as the possible common noise values).

3.2. Financial systemic risk. We now consider an example inspired by the systemic risk model proposed by [18]. In this model, the agents are financial institutions, represented by a state which is their log-reserve. They interact by borrowing and lending to each other, or to a central bank. Their evolution is impacted by a common noise which can be interpreted as macroscopic events affecting the whole economy. If a financial institution touches a given threshold, it defaults. There are two main differences between the model we present below and the original model one: first, the model of [18] was a mean field game (corresponding to non-cooperative players) while we consider a mean field control problem (corresponding to cooperative players); furthermore, the original model was written in continuous space and time whereas we consider a discrete space and time model for the sake of numerical experiments. However, the main ideas underpinning the model are similar. The central planner is to make the population distribution match a given target distribution.

To be specific, consider the following basic elements (recall Definition 2.4):

- $S = \{s_{\min}, s_{\min} + 1, \dots, s_{\max}\}$, which represents a one-dimensional grid world with $|S| = s_{\max} - s_{\min} + 1$ states; in the experiments, we use $s_{\min} = -1$, $s_{\max} = 4$, $|S| = 5$ states.
- $A = \{-1, 0, 1\}$, which corresponds to lending (if negative) or borrowing (if positive) units.
- $E = \{-1, 0, 1\}$, which corresponds to idiosyncratic noise. Moreover, the probability vector of its law $\lambda_e \in \mathcal{P}(E)$ is given by $(0.05, 0.9, 0.05)$.
- $E^0 = \{-2, -1, 0, 1, 2\}$, which corresponds to common noise affecting the whole population.
- $F : S \times A \times \mathcal{P}(S \times A) \times E \times E^0 \rightarrow S$ is given by

$$F(s, a, \Lambda, e, e^0) = \max(s_{\min}, \min(s_{\max}, s + a + e + e^0)) \quad \text{if } s > s_{\min},$$

and $F(s_{\min}, a, \Lambda, e, e^0) = s_{\min}$, which represents the fact that the agent's log-reserve evolution is determined by her action, the individual noise and the common noise, the agent

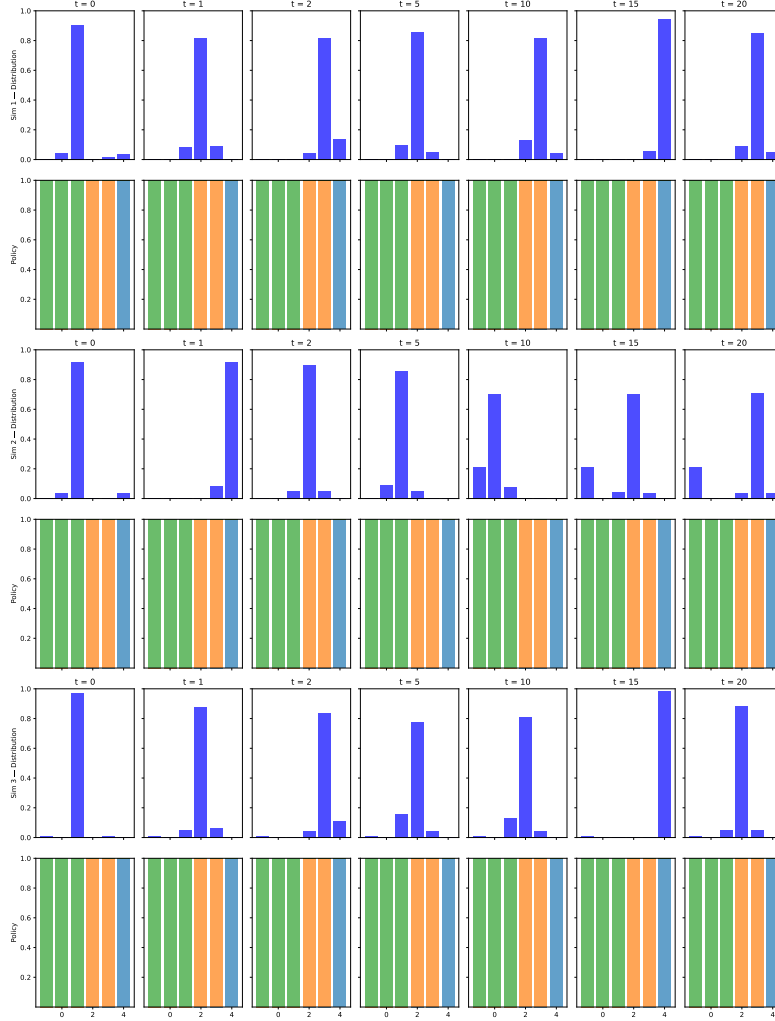


FIGURE 5. Three sample trajectories of the population distribution and corresponding action distribution for each state in Example 2.

remains at 1 (resp. 7) if she tries to move to the left (resp. right) of this state, and the agent remains stuck at $s = 1$ if she ever reaches this state.

- $r : S \times A \times \mathcal{P}(S \times A) \rightarrow \mathbb{R}$ is given by

$$r(s, a, \Lambda) = -a^2 + qa(m(\Lambda) - s)^2 - 0.5\epsilon(m(\Lambda) - s)^2 + (m(\Lambda) - s_{\text{target}})^2,$$

where $m(\Lambda)$ is given by $m(\Lambda) := \int_S s' p_{j_S}(\Lambda)(ds')$ (i.e., the first moment of the state), the constants q, ϵ are non-positive and satisfy $q^2 \leq \epsilon$, and s_{target} is a target state taken equal to 2 in the experiments. The first term is a cost of borrowing/lending, the second and third terms have a mean-reverting effect, and the last term means that the regulator has a target level for the mean of the log-reserves. Here, q represents the incentive to borrowing or lending. We refer to [18] for more details.

- $\beta = 0.15$ is the discount factor so that Assumptions 2.7 (iii) and 2.14 (iv) are satisfied.

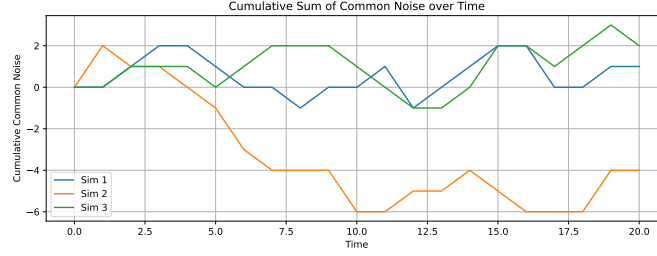


FIGURE 6. Three sample trajectories of common noise, associated to the three distribution trajectories presented in Figure 5.

For the common noise probability measure, we proceed as in the previous example of Section 3.1. The true common noise measure is denoted by $p_{\text{true}} \in \mathcal{P}(E^0)$ (as in (3.1), but now represented by a 5-dimensional probability vector $v_{\text{true}} \in [0, 1]^5$). The central planner does not know this true measure and instead relies on a reference probability measure $p_{\text{ref}} \in \mathcal{P}(E^0)$ with corresponding probability vector $v_{\text{ref}} \in [0, 1]^5$. We then compare, in the true model with p_{true} , the performance of π_{ref} (an optimal policy for the model with common noise distribution p_{ref}) and the performance of π_{robust} (a robust policy for p_{ref}). The uncertainty set $\mathfrak{P}^0$ is defined as in (3.2), but adapted to the 5-dimensional setting so that Assumption 2.14 (i) also holds.

We implement the above model with: $v_{\text{true}} = (0.1, 0.2, 0.4, 0.2, 0.1)$, $v_{\text{ref}} = (0, 0, 1, 0, 0)$ and δ_{perturb} varying between 0.0 and 0.6. Figure 4 shows that for moderately small δ_{perturb} , the robust policy performs better than the non-robust policy. For large values of δ_{perturb} however, the robust policy yields a smaller value: being robust against a large set of possible common noise distributions prevents the policy from performing well on the true distribution. The results are averaged over 15 different runs and the plots shows the mean value and its standard deviation. Figure 5 shows three realizations of trajectories, starting from random initial distributions. We display a few time steps between 0 and 20. We observe that the learnt policy is pure at the agent level, meaning that in each state, the agent uses one action with probability 1. In fact, the agent uses actions that tend to make the state move towards state 2 or 3. The distribution concentrates (but not completely due to the idiosyncratic noise which tends to make the agent spread). Moreover, the peak is not always at state 2 or 3 due to the impact of the common noise, whose trajectories are displayed in Figure 6.

4. PROOF OF RESULTS IN SECTION 2.2

We begin by verifying the measurability of the state dynamics appearing in both models. We recall the filtrations given in Definition 2.1.

Lemma 4.1. *For any $\pi \in \Pi$ and $\mathbb{P} \in \mathcal{Q}$, the following statements hold:*

- (i) *For every $N \in \mathbb{N}$, $i = 1, \dots, N$, and $t \geq 0$, $s_t^{i,N,\pi}$ given in (2.3) is $(\bigvee_{j=1}^N \mathcal{F}_t^j)$ measurable.*
- (ii) *For every $i \in \mathbb{N}$ and $t \geq 0$, $s_t^{i,\pi,\mathbb{P}}$ in (2.4) is $\mathcal{F}_t^i$ measurable, and both $\mathbb{P}_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}^0$ and $\mathbb{P}_{s_t^{i,\pi,\mathbb{P}}}^0$ are $\mathcal{F}_t^0$ measurable.*

Proof. We start proving (i). Let $N \in \mathbb{N}$ and $i = 1, \dots, N$ be given. The statement is shown via an induction over $t \geq 0$: Since $s_0^{i,N,\pi} = \xi^i \in L_{\mathcal{F}_0^i}^0(S)$ (see Definition 2.5), the claim for $t = 0$ holds.

Now assume that the induction claim holds for some $t \geq 0$. Note that $s_{t+1}^{i,N,\pi}$ satisfies

$$s_{t+1}^{i,N,\pi} = F(s_t^{i,N,\pi}, a_t^{i,\pi}, \frac{1}{N} \sum_{j=1}^N \delta_{(s_t^{j,N,\pi}, a_t^{j,\pi})}, \varepsilon_{t+1}^i, \varepsilon_{t+1}^0)$$

where the first three terms are $(\bigvee_{j=1}^N \mathcal{G}_t^j)$ measurable because of the induction assumption and the definition of the open-loop control $\alpha_t^{i,\pi}$ in Definition 2.5 (i), and the fact that $\bigvee_{j=1}^N \mathcal{F}_t^j \subset \bigvee_{j=1}^N \mathcal{G}_t^j$. Hence by the Borel measurability of F , $s_{t+1}^{i,N,\pi}$ is $(\bigvee_{j=1}^N \mathcal{F}_{t+1}^j)$ measurable (see Definition 2.1).

By the induction hypothesis, the statement in (i) holds for all $t \geq 0$.

The part (ii) is also shown via an induction over t given any $i \in \mathbb{N}$. Since $s_0^{i,\pi,\mathbb{P}} = \xi^i \in L_{\mathcal{F}_0^i}^0(S)$ (see Definition 2.6), $s_0^{i,\pi,\mathbb{P}}$ is $\mathcal{F}_0^i$ measurable. Moreover, since $\mathcal{F}_0^0$ is trivial, both $\mathbb{P}_{(s_0^{i,\pi,\mathbb{P}}, a_0^{i,\pi})}^0$ and $\mathbb{P}_{s_0^{i,\pi,\mathbb{P}}}^0$ are $\mathcal{F}_t^0$ measurable obviously.

We assume that the claim holds for some $t \geq 0$. Note that $s_{t+1}^{i,\pi,\mathbb{P}}$ satisfies

$$s_{t+1}^{i,\pi,\mathbb{P}} = F(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi}, \mathbb{P}_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}^0, \varepsilon_{t+1}^i, \varepsilon_{t+1}^0),$$

where the first three terms are $\mathcal{G}_t^i$ measurable because of the induction assumption and the fact that $\mathcal{F}_t^0 \subset \mathcal{G}_t^i$. Hence by the Borel measurability of F , $s_{t+1}^{i,\pi,\mathbb{P}}$ is $\mathcal{F}_{t+1}^i$ measurable (see Definition 2.1).

Moreover, since $a_{t+1}^{i,\pi}$ is $\mathcal{G}_{t+1}^i$ measurable and $(\gamma^i, \vartheta_{0:t+1}^i, \varepsilon_{1:t+1}^i)$ is independent of $\varepsilon_{1:t+1}^0$ (see Remark 2.3 (i)), we apply Lemma A.1 (ii) to have that both $\mathbb{P}_{(s_{t+1}^{i,\pi,\mathbb{P}}, a_{t+1}^{i,\pi})}^0$ and $\mathbb{P}_{s_{t+1}^{i,\pi,\mathbb{P}}}^0$ are $\mathcal{F}_{t+1}^0$ measurable. By the induction hypothesis, the statement in (ii) holds. $\square$

4.1. Proof of Lemma 2.8. We start proving (i). Let $q > 2$ be given. Note that by Lemma 4.1 (ii), the definition of open-loop controls (see Definition 2.5 (i)), and recalling that $\mathbb{F}^i \subset \mathbb{G}^i$ for any $i \in \mathbb{N}$ ($s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi}$) is $\mathcal{G}_t^i$ measurable.

Moreover, since the private components $(\gamma^i)_{i \in \mathbb{N}}$, $(\vartheta_t^i)_{t \geq 0, i \in \mathbb{N}}$, and $(\varepsilon_t^i)_{t \geq 1, i \in \mathbb{N}}$ are mutually independent (see Remark 2.3 (i)) and all agents are indistinguishable, it holds for every $t \geq 0$, $\pi \in \Pi$, and $\mathbb{P} \in \mathcal{Q}$ that $(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})_{i \in \mathbb{N}}$ is (conditionally) i.i.d. given the common noise information $\mathcal{F}_t^0$ with law $\mathbb{P}_{(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi})}^0$. Therefore, it follows from [35, Theorem 1] that

$$\mathbb{E}^{\mathbb{P}^0} \left[\mathcal{W}_{\mathcal{P}(S \times A)} \left(\frac{1}{N} \sum_{i=1}^N \delta_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}, \mathbb{P}_{(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi})}^0 \right) \right] \leq C (K_q(\mathbb{P}_{(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi})}^0))^{1/q} \alpha(N),$$

where $C > 0$ does not depends on $\mathbb{P}^0$ and N but on d and q , $\alpha(\cdot)$ is defined as in the statment, and $K_q(\mathbb{P}_{(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi})}^0)$ is given by

$$K_q(\mathbb{P}_{(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi})}^0) := \int_{S \times A} |(s, a)|^q \mathbb{P}_{(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi})}^0(ds, da).$$

Since $S \times A$ is a compact subset of $\mathbb{R}^d$, the above quantitiy is uniformly bounded by $(\Delta_{S \times A})^q$ for every $t \geq 0$, $\pi \in \Pi$, and $\mathbb{P} \in \mathcal{Q}$. Hence the estimate in part (i) holds.

Last, we prove (ii). Let $q > 2$ be given. In part (i), we have verified that for every $t \geq 0$, $\pi \in \Pi$, and $\mathbb{P} \in \mathcal{Q}$, $(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})_{i \in \mathbb{N}}$ is (conditionally) i.i.d. given $\mathcal{F}_t^0$ with law $\mathbb{P}_{(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi})}^0$.

Hence, we can apply [14, Corollary 1.2] to obtain that for every $t \geq 0$, $\pi \in \Pi$, and $\mathbb{P} \in \mathcal{Q}$

$$\mathbb{E}^{\mathbb{P}^0} \left[\mathcal{W}_{\mathcal{P}(S \times A)} \left(\frac{1}{N} \sum_{i=1}^N \delta_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}, \mathbb{P}_{(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi})}^0 \right) \right] \leq c \left(\frac{2}{q-2} \right)^{\frac{2}{q}} (k_{S \times A})^{\frac{1}{q}} \Delta_{S \times A} N^{-\frac{1}{q}},$$

with some $c \leq 64/3$. Therefore, we can obtain the estimate in part (ii), as claimed. $\square$

4.2. Proof of Theorem 2.9. For notational simplicity, throughout this proof, denote for every $N \in \mathbb{N}$, $i = 1, \dots, N$, $t \geq 0$, $\pi \in \Pi$, and $\mathbb{P} \in \mathcal{Q}$ by

$$\begin{aligned} \Lambda_t^{N,\pi} &:= \frac{1}{N} \sum_{j=1}^N \delta_{(s_t^{j,N,\pi}, a_t^{j,\pi})}, & \Lambda_t^{N,\infty,\pi,\mathbb{P}} &:= \frac{1}{N} \sum_{j=1}^N \delta_{(s_t^{j,\pi,\mathbb{P}}, a_t^{j,\pi})}, \\ \tilde{\Lambda}_t^{i,\pi,\mathbb{P}} &:= \mathbb{P}_{(s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi})}^0. \end{aligned}$$

Let $N \in \mathbb{N}$ and $i = 1, \dots, N$ be given. We first prove (2.7) and (2.8). The proof uses an induction over $t \geq 0$: Since $s_0^{i,N,\pi} = s_0^{i,\pi,\mathbb{P}}$ for every $\pi \in \Pi$, and $\mathbb{P} \in \mathcal{Q}$ (see Definitions 2.5 and 2.6), the claim for $t = 0$ holds.

Now assume that the induction claim holds true for some $t \geq 1$. Let $\pi \in \Pi$ and $\mathbb{P} \in \mathcal{Q}$ be given. Since $\bigvee_{j=1}^N \mathcal{F}_t^j \subset \bigvee_{j=1}^N \mathcal{G}_t^j$ (see Definition 2.1), both $s_t^{i,N,\pi}$ given in (2.3) and $s_t^{i,\pi,\mathbb{P}}$ given in (2.4) are $(\bigvee_{j=1}^N \mathcal{G}_t^j)$ measurable (see Lemma 4.1). Moreover, $a_t^{i,\pi}$ is $\mathcal{G}_t^i$ measurable (see Definition 2.5 (i)).

Since ε_{t+1}^i is independent of $\bigvee_{j=1}^N \mathcal{G}_t^j$ and ε_{t+1}^0 (see Remark 2.3 (i), (ii)), we can have the following conditioning

$$(4.1) \quad \mathbb{E}^\mathbb{P}[d_S(s_{t+1}^{i,N,\pi}, s_{t+1}^{i,\pi,\mathbb{P}})] = \mathbb{E}^\mathbb{P}[D^{i,\mathbb{P}}(s_t^{i,N,\pi}, s_t^{i,\pi,\mathbb{P}}, a_t^{i,\pi,\mathbb{P}}, \Lambda_t^{N,\pi}, \tilde{\Lambda}_t^{i,\pi,\mathbb{P}}, e^0)],$$

where for every $(s, \tilde{s}) \in S$, $a \in A$, $\Lambda, \tilde{\Lambda} \in \mathcal{P}(S \times A)$, and $e^0 \in E^0$

$$(4.2) \quad \begin{aligned} D^{i,\mathbb{P}}(s, \tilde{s}, a, \Lambda, \tilde{\Lambda}, e^0) &:= \int_E d_S(F(s, a, \Lambda, e, e^0), F(\tilde{s}, a, \tilde{\Lambda}, e, e^0)) \lambda_\varepsilon(de) \\ &\leq C_F(d_S(s, \tilde{s}) + \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda})), \end{aligned}$$

where the inequality follows from Assumption 2.7 (i).

On the other hand, it holds that

$$(4.3) \quad \begin{aligned} \mathbb{E}^\mathbb{P}[\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda_t^{N,\pi}, \tilde{\Lambda}_t^{i,\pi,\mathbb{P}})] &\leq \mathbb{E}^\mathbb{P}[\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda_t^{N,\pi}, \Lambda_t^{N,\infty,\pi,\mathbb{P}})] + \mathbb{E}^\mathbb{P}[\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda_t^{N,\infty,\pi,\mathbb{P}}, \tilde{\Lambda}_t^{i,\pi,\mathbb{P}})] \\ &\leq \mathbb{E}^\mathbb{P}[d_S(s_t^{i,N,\pi}, s_t^{i,\pi,\mathbb{P}})] + M_N, \end{aligned}$$

where the second inequality follows from the definition of M_N given in (2.6) and the fact that $\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda_t^{N,\pi}, \Lambda_t^{N,\infty,\pi,\mathbb{P}}) \leq \frac{1}{N} \sum_{j=1}^N d_S(s_t^{j,N,\pi}, s_t^{j,\pi,\mathbb{P}})$ together with the indistinguishability.

Combining (4.1) with (4.2) and (4.3), we have that

$$(4.4) \quad \mathbb{E}^\mathbb{P}[d_S(s_{t+1}^{i,N,\pi}, s_{t+1}^{i,\pi,\mathbb{P}})] \leq C_F(2\mathbb{E}^\mathbb{P}[d_S(s_t^{i,N,\pi}, s_t^{i,\pi,\mathbb{P}})] + M_N).$$

Since the estimate (4.4) holds for any $\pi \in \Pi$ and $\mathbb{P} \in \mathcal{Q}$, by the induction hypothesis we have that the estimate (2.7) holds for all $t \geq 0$, as claimed.

Moreover, since the estimate (4.3) holds for any $\pi \in \Pi$ and $\mathbb{P} \in \mathcal{Q}$, by using (2.7) we have that the other estimate (2.8) holds for all $t \geq 0$, as claimed. As $N \in \mathbb{N}$ and $i = 1, \dots, N$ are given arbitrary, we can conclude that (2.7) and (2.8) hold for all $N \in \mathbb{N}$, $i = 1, \dots, N$, and $t \geq 0$.

We now prove (2.9). Note that for every $N \in \mathbb{N}$ and $\pi \in \Pi$

$$(4.5) \quad \begin{aligned} |\mathcal{J}^{N,\pi} - \mathcal{J}^\pi| &= \left| \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^\mathbb{P} \left[\frac{1}{N} \sum_{i=1}^N R^{i,N,\pi} \right] - \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^\mathbb{P} \left[\frac{1}{N} \sum_{i=1}^N R^{i,\pi,\mathbb{P}} \right] \right| \\ &\leq \sup_{\mathbb{P} \in \mathcal{Q}} \frac{1}{N} \sum_{i=1}^N \mathbb{E}^\mathbb{P} [|R^{i,N,\pi} - R^{i,\pi,\mathbb{P}}|] = \sup_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^\mathbb{P} [|R^{1,N,\pi} - R^{1,\pi,\mathbb{P}}|] \\ &\leq \sum_{t=0}^{\infty} \beta^t \sup_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^\mathbb{P} [|r(s_t^{1,N,\pi}, a_t^{1,\pi}, \Lambda_t^{N,\pi}) - r(s_t^{1,\pi,\mathbb{P}}, a_t^{1,\pi}, \tilde{\Lambda}_t^{1,\pi,\mathbb{P}})|] =: \mathbf{I}^{N,\pi}, \end{aligned}$$

where the equalities follow from the indistinguishability and the last inequality holds because r is bounded (see Assumption 2.7 (ii)).

Moreover, by the Lipschitz continuity of $r(\cdot, a, \cdot) : S \times \mathcal{P}(S \times A) \rightarrow \mathbb{R}$ for any $a \in A$ (see Assumption 2.7 (ii)), for every $N \in \mathbb{N}$ and $\pi \in \Pi$

$$(4.6) \quad \begin{aligned} \mathbf{I}^{N,\pi} &\leq C_r \sum_{t=0}^{\infty} \beta^t \sup_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}} \left[d_S(s_t^{1,N,\pi}, s_t^{1,\pi,\mathbb{P}}) + \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda_t^{N,\pi}, \tilde{\Lambda}_t^{1,\pi,\mathbb{P}}) \right] \\ &\leq C_r \left(2 \sum_{t=0}^{\infty} \beta^t \delta_t^N + \frac{M_N}{1-\beta} \right), \end{aligned}$$

where $\delta_t^N := \sup_{\pi \in \Pi} \sup_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}}[d_S(s_t^{1,N,\pi}, s_t^{1,\pi,\mathbb{P}})]$ for $t \geq 0$.

Since the estimate given in (4.5) coincides with that of [57, Theorem 2.1]—specifically Eq. (2.17) therein—and Assumption 2.7 (iii) ensures that $2\beta C_r < 1$, we can follow the same calculations as in the proof of [57, Theorem 2.1] (replacing K_F with C_r). This yields that $\sum_{t=0}^{\infty} \beta^t \delta_t^N \leq CM_N$ for some constant $C > 0$ (that do not depend on N and π); see also [57, Remark 2.4].

Combining this with (4.5) and (4.6) establishes the estimate in (2.9). $\square$

5. PROOF OF RESULTS IN SECTION 2.3

5.1. Proof of Proposition 2.12. We first prove (2.18). For simplicity, denote for every $t \geq 0$ by

$$(5.1) \quad \mu_t := \mu_t^{\xi,a,\mathbb{P}}, \quad \Lambda_t := \Lambda_t^{\xi,a,\mathbb{P}}, \quad \nu_{t+1} := \mathcal{L}_{\mathbb{P}}(\varepsilon_{t+1}^0 | \mathcal{F}_t^0).$$

Since μ_{t+1} is $\mathcal{F}_{t+1}^0$ measurable, it is sufficient to show that for any bounded Borel measurable functions $\hat{g} : (E^0)^{t+1} \rightarrow \mathbb{R}$ and $\hat{f} : S \rightarrow \mathbb{R}$,

$$(5.2) \quad \mathbb{E}^{\mathbb{P}}[\hat{g}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi,a,\mathbb{P}})] = \mathbb{E}^{\mathbb{P}} \left[\hat{g}(\varepsilon_{1:t+1}^0) \int_S \hat{f}(s') \bar{\mathbb{F}}(\text{pj}_S(\Lambda_t), \Lambda_t, \varepsilon_{t+1}^0)(ds') \right],$$

where we note that $(\text{pj}_S(\Lambda_t), \Lambda_t) \in \text{gr}(\mathfrak{U})$ (see Definition 2.11 (i)).

Note that by Remark 2.3 (i) and (ii), ε_{t+1} is independent of $\varepsilon_{1:t+1}^0, s_t, a_t, \mathbb{P}_{(s_t, a_t)}^0$ (since they are all $\mathcal{G}_t \vee \sigma(\varepsilon_{t+1}^0)$ measurable) with $\mathcal{L}_{\mathbb{P}}(\varepsilon_{t+1}) = \lambda_{\varepsilon}$. Moreover, by (2.12) and Fubini's theorem (noting that $\hat{g}$ and $\hat{f}$ are both bounded)

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}[\hat{g}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi,a,\mathbb{P}})] &= \mathbb{E}^{\mathbb{P}} \left[\mathbb{E}^{\mathbb{P}} \left[\hat{g}(\varepsilon_{1:t+1}^0) \hat{f}(\mathbb{F}(s_t^{\xi,a,\mathbb{P}}, a_t, \Lambda_t, \varepsilon_{t+1}, \varepsilon_{t+1}^0)) \right] \middle| e = \varepsilon_{t+1} \right] \\ &= \int_E \mathbb{E}^{\mathbb{P}} \left[\hat{g}(\varepsilon_{1:t+1}^0) \hat{f}(\mathbb{F}(s_t^{\xi,a,\mathbb{P}}, a_t, \Lambda_t, e, \varepsilon_{t+1}^0)) \right] \lambda_{\varepsilon}(de) =: \text{I}. \end{aligned}$$

Note that $\varepsilon_{1:t}^0, s_t^{\xi,a,\mathbb{P}}, a_t$, and Λ_t are all $\mathcal{G}_t$ measurable. Since ε_{t+1}^0 is conditionally independent of $\mathcal{G}_t$ given $\mathcal{F}_t^0$ (see Remark 2.3 (iii)), by definition of ν_{t+1} (see (5.1))

$$\begin{aligned} \text{I} &= \int_E \mathbb{E}^{\mathbb{P}} \left[\mathbb{E}^{\mathbb{P}} \left[\int_{E^0} \hat{g}(\varepsilon_{1:t}^0, e^0) \hat{f}(\mathbb{F}(s_t^{\xi,a,\mathbb{P}}, a_t, \Lambda_t, e, e^0)) \nu_{t+1}(de^0) \right] \middle| \mathcal{F}_t^0 \right] \lambda_{\varepsilon}(de) \\ &= \int_E \mathbb{E}^{\mathbb{P}} \left[\int_{E^0} \hat{g}(\varepsilon_{1:t}^0, e^0) \mathbb{E}^{\mathbb{P}} \left[\hat{f}(\mathbb{F}(s_t^{\xi,a,\mathbb{P}}, a_t, \Lambda_t, e, e^0)) \right] \middle| \mathcal{F}_t^0 \right] \nu_{t+1}(de^0) \lambda_{\varepsilon}(de) =: \text{II}. \end{aligned}$$

Moreover by definition of Λ_t (see (5.1)) and Fubini's theorem

$$\begin{aligned} \text{II} &= \int_E \mathbb{E}^{\mathbb{P}} \left[\int_{E^0} \hat{g}(\varepsilon_{1:t}^0, e^0) \int_{S \times A} \hat{f}(\mathbb{F}(s, a, \Lambda_t, e, e^0)) \Lambda_t(ds, da) \nu_{t+1}(de) \right] \lambda_{\varepsilon}(de) \\ &= \mathbb{E}^{\mathbb{P}} \left[\hat{g}(\varepsilon_{1:t+1}^0) \int_{S \times A \times E} \hat{f}(\mathbb{F}(s, a, \Lambda_t, e, \varepsilon_{t+1}^0)) \Lambda_t(ds, da) \lambda_{\varepsilon}(de) \right]. \end{aligned}$$

By definition of $\bar{F}$ (see Definition 2.11 (ii)), the last term above is equal to the second term given in (5.2), as claimed.

We now prove (2.19). Note that by Remark 2.3 (iii) $(\nu_t)_{t \geq 0}$ given in (5.1) satisfies $\mathbb{P}$ -a.s.

$$\nu_1 = p_1, \quad \nu_t = p_t(\cdot | \varepsilon_{1:t-1}^0) \quad \text{for all } t \geq 2,$$

where $(p_t)_{t \geq 1} \in \mathcal{K}^0$ induces the measure $\mathbb{P} \in \mathcal{Q}$.

Let $t \geq 1$. It is sufficient to show that for any bounded Borel measurable function $\tilde{f} : \mathcal{P}(S) \rightarrow \mathbb{R}$

$$(5.3) \quad \mathbb{E}^\mathbb{P}[\tilde{f}(\mu_{t+1})] = \mathbb{E}^\mathbb{P} \left[\int_{\mathcal{P}(S)} \tilde{f}(\mu') \bar{p}(d\mu' | \text{pj}_S(\Lambda_t), \Lambda_t, \nu_{t+1}) \right].$$

By (2.18), we have $\mu_{t+1} = \bar{F}(\text{pj}_S(\Lambda_t), \Lambda_t, \varepsilon_{t+1}^0)$ $\mathbb{P}$ -a.s.. Moreover, since ε_{t+1}^0 is conditionally independent of $(\text{pj}_S(\Lambda_t), \Lambda_t)$ given $\mathcal{F}_t^0$ (as Λ_t is $\mathcal{G}_t$ measurable) with $\mathcal{L}_\mathbb{P}(\varepsilon_{t+1}^0 | \mathcal{F}_t^0) = \nu_{t+1}$, it follows that

$$\mathbb{E}^\mathbb{P}[\tilde{f}(\mu_{t+1})] = \mathbb{E}^\mathbb{P} \left[\mathbb{E}^\mathbb{P}[\tilde{f}(\bar{F}(\text{pj}_S(\Lambda_t), \Lambda_t, \varepsilon_{t+1}^0)) | \mathcal{F}_t^0] \right] = \mathbb{E}^\mathbb{P} \left[\int_{E^0} \tilde{f}(\bar{F}(\text{pj}_S(\Lambda_t), \Lambda_t, e^0)) \nu_{t+1}(de^0) \right].$$

By definition of $\bar{p}$ (see Definition 2.11 (iii)), the claim (5.3) holds.

For the case $t = 0$, note that $\mathcal{L}_\mathbb{P}(\varepsilon_1^0) = p_1$ and $\Lambda_0 \in \mathcal{P}(S \times A)$ is deterministic. Thus, it is straightforward to verify that (2.19) holds also for $t = 0$.

This completes the proof. $\square$

5.2. Proof of Proposition 2.15. In what follows, we often make use of the following coupling result along with the continuity of the projection map $\text{pj}_S : \mathcal{P}(S \times A) \rightarrow \mathcal{P}(S)$.

Lemma 5.1. *The following properties hold:*

- (i) *For every $(\mu, \zeta), (\tilde{\mu}, \tilde{\zeta}) \in \mathcal{P}(S) \times \mathcal{P}(A)$ and every $\Lambda \in \text{Cpl}_{S \times A}(\mu, \zeta)$, there exists a coupling $\tilde{\Lambda}^* \in \text{Cpl}_{S \times A}(\tilde{\mu}, \tilde{\zeta})$ such that*

$$\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}^*) \leq \mathcal{W}_{\mathcal{P}(S)}(\mu, \tilde{\mu}) + \mathcal{W}_{\mathcal{P}(A)}(\zeta, \tilde{\zeta}).$$

- (ii) *For every $\Lambda, \tilde{\Lambda} \in \mathcal{P}(S \times A)$, it holds that*

$$\mathcal{W}_{\mathcal{P}(S)}(\text{pj}_S(\Lambda), \text{pj}_S(\tilde{\Lambda})) \leq \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}).$$

Thus $\text{pj}_S : \mathcal{P}(S \times A) \rightarrow \mathcal{P}(S)$ is continuous.

Proof. We start by proving (i). Let $(\mu, \zeta), (\tilde{\mu}, \tilde{\zeta}) \in \mathcal{P}(S) \times \mathcal{P}(A)$ and $\Lambda \in \text{Cpl}_{S \times A}(\mu, \zeta)$. Denote by

$$(5.4) \quad \Gamma \in \text{Cpl}_{S \times S}(\mu, \tilde{\mu}), \quad \Upsilon \in \text{Cpl}_{A \times A}(\zeta, \tilde{\zeta})$$

the optimal couplings for $\mathcal{W}_{\mathcal{P}(S)}(\mu, \tilde{\mu})$ and $\mathcal{W}_{\mathcal{P}(A)}(\zeta, \tilde{\zeta})$, respectively (whose existence is ensured by [67, Theorem 4.1]). Then we define $\Xi \in \mathcal{P}((S \times A)^2)$ by

$$\Xi(ds, da, d\tilde{s}, d\tilde{a}) := \Upsilon_\zeta(d\tilde{a}|a) \Lambda_\mu(da|s) \Gamma(ds, d\tilde{s}),$$

where $\Lambda_\mu : S \ni s \mapsto \Lambda_\mu(da|s) \in \mathcal{P}(A)$ denotes a disintegrating kernel of Λ with respect to its marginal $\mu = \text{pj}_S(\Lambda)$, i.e.,

$$(5.5) \quad \Lambda(ds, da) = \Lambda_\mu(da|s) \mu(ds).$$

In a similar manner, $\Upsilon_\zeta : A \ni a \mapsto \Upsilon_\zeta(d\tilde{a}|a) \in \mathcal{P}(A)$ denotes a disintegrating kernel of Υ with respect to its marginal $\zeta = \text{pj}_A(\Upsilon)$.

Then, by (5.4) and (5.5), it holds that $\int_{(\tilde{s}, \tilde{a}) \in S \times A} \Xi(ds, da, d\tilde{s}, d\tilde{a}) = \Lambda(ds, da)$. Moreover by setting $\tilde{\Lambda}^\diamond(ds, d\tilde{a}) := \int_{(s, a) \in S \times A} \Xi(ds, da, d\tilde{s}, d\tilde{a})$, we have that

$$\tilde{\Lambda}^\diamond \in \text{Cpl}_{S \times A}(\tilde{\mu}, \tilde{\zeta}), \quad \Xi \in \text{Cpl}_{(S \times A)^2}(\Lambda, \tilde{\Lambda}^\diamond).$$

This implies that

$$\begin{aligned} \inf_{\tilde{\Lambda} \in \text{Cpl}_{S \times A}(\tilde{\mu}, \tilde{\zeta})} \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}) &\leq \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}^\diamond) \leq \int_{(S \times A)^2} d_{S \times A}((s, a), (\tilde{s}, \tilde{a})) \Xi(ds, da, d\tilde{s}, d\tilde{a}) \\ &= \int_{S \times S} d_S(s, \tilde{s}) \Gamma(ds, d\tilde{s}) + \int_{A \times A} d_A(a, \tilde{a}) \Upsilon(da, d\tilde{a}) \\ &= \mathcal{W}_{\mathcal{P}(S)}(\mu, \tilde{\mu}) + \mathcal{W}_{\mathcal{P}(A)}(\zeta, \tilde{\zeta}), \end{aligned}$$

where the last equality follows from the optimality of Γ and Υ (see (5.4)).

Combining this with the compactness of $\text{Cpl}_{S \times A}(\tilde{\mu}, \tilde{\zeta})$ (see [67, Theorem 4.1 & Lemma 4.4]), one can choose $\tilde{\Lambda}^* \in \text{Cpl}_{S \times A}(\tilde{\mu}, \tilde{\zeta})$ so that

$$\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}^*) = \inf_{\tilde{\Lambda} \in \text{Cpl}_{S \times A}(\tilde{\mu}, \tilde{\zeta})} \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}) \leq \mathcal{W}_{\mathcal{P}(S)}(\mu, \tilde{\mu}) + \mathcal{W}_{\mathcal{P}(A)}(\zeta, \tilde{\zeta}),$$

as claimed.

Next we prove the part (ii). Let $\Lambda, \tilde{\Lambda} \in \mathcal{P}(S \times A)$. Denote by $\Xi^* \in \text{Cpl}_{(S \times A)^2}(\Lambda, \tilde{\Lambda})$ the optimal coupling for $\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda})$. By setting $h(s, a) := s$ for every $(s, a) \in S \times A$ (i.e., a projection map onto S), denote by

$$\Xi^\diamond := (\Xi^* \circ (h \times h)^{-1}) \in \mathcal{P}(S \times S)$$

the push-forward of Ξ^* by the map $(h \times h) : (S \times A)^2 \rightarrow S^2$.

Clearly $\Xi^\diamond$ is in $\text{Cpl}_{S \times S}(\text{pj}_S(\Lambda), \text{pj}_S(\tilde{\Lambda}))$. Thus,

$$\mathcal{W}_{\mathcal{P}(S)}(\text{pj}_S(\Lambda), \text{pj}_S(\tilde{\Lambda})) \leq \int_{S \times S} d_S(s, \tilde{s}) \Xi^\diamond(ds, d\tilde{s}) = \int_{(S \times A)^2} d_S(h(s, a), h(\tilde{s}, \tilde{a})) \Xi^*(ds, da, d\tilde{s}, d\tilde{a}).$$

Moreover, since $d_S(h(s, a), h(\tilde{s}, \tilde{a})) = d_S(s, \tilde{s}) \leq d_{S \times A}((s, a), (\tilde{s}, \tilde{a}))$ for every $(s, a), (\tilde{s}, \tilde{a}) \in S \times A$, by the optimality of $\Xi^* \in \text{Cpl}_{(S \times A)^2}(\Lambda, \tilde{\Lambda})$, the assertion for the part (ii) holds, as claimed. $\square$

The following lemma provides useful properties of the lifted functions defined in Definition 2.11.

Lemma 5.2. *Suppose that Assumption 2.14 (ii), (iii) are satisfied. Let $\mathfrak{U}$, $\bar{\mathbb{F}}$, $\bar{r}$ be given in Definition 2.11. Then the following hold:*

- (i) $\mathfrak{U}$ is non-empty, compact-valued and continuous.⁶
- (ii) $\bar{\mathbb{F}}$ satisfies that for every $(\mu, \Lambda, e^0), (\tilde{\mu}, \tilde{\Lambda}, \tilde{e}^0) \in \text{gr}(\mathfrak{U}) \times E^0$,

$$\mathcal{W}_{\mathcal{P}(S)}(\bar{\mathbb{F}}(\mu, \Lambda, e^0), \bar{\mathbb{F}}(\tilde{\mu}, \tilde{\Lambda}, \tilde{e}^0)) \leq \bar{C}_F(2\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}) + d_{E^0}(e^0, \tilde{e}^0)).$$

- (iii) $\bar{r}$ is bounded. Furthermore, for every $(\mu, \Lambda), (\tilde{\mu}, \tilde{\Lambda}) \in \text{gr}(\mathfrak{U})$

$$|\bar{r}(\mu, \Lambda) - \bar{r}(\tilde{\mu}, \tilde{\Lambda})| \leq 2\bar{C}_r \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}).$$

Proof. We start by proving (i). Both the non-emptiness and the compact-valuedness of $\mathfrak{U}$ are clear. Indeed, for every $\mu \in \mathcal{P}(S)$ one can consider the Dirac measure $\delta_{\tilde{a}}(da) \in \mathcal{P}(A)$ at some $\tilde{a} \in A$ to obtain that $\delta_{\tilde{a}}(da)\mu(ds) \in \mathfrak{U}(\mu)$. Therefore $\mathfrak{U}(\mu)$ is non-empty.

Moreover, since $\text{pj}_S : \mathcal{P}(S \times A) \rightarrow \mathcal{P}(S)$ is continuous (see Lemma 5.1 (ii)) and $\mathcal{P}(S \times A)$ is compact (as $S \times A$ is compact), $\mathfrak{U}(\mu) \subseteq \mathcal{P}(S \times A)$ is compact for every $\mu \in \mathcal{P}(S)$, as claimed.

We now claim that $\mathfrak{U}$ is both upper and lower hemicontinuous. Let $\mu \in \mathcal{P}(S)$ be given.

Recalling that $\text{gr}(\mathfrak{U}) = \{(\mu, \Lambda) \in \mathcal{P}(S) \times \mathcal{P}(S \times A) \mid \Lambda \in \mathfrak{U}(\mu)\}$, let us consider a sequence $(\mu^{(n)}, \Lambda^{(n)})_{n \in \mathbb{N}} \in \text{gr}(\mathfrak{U})$ such that $\mu^{(n)} \rightarrow \mu$ as $n \rightarrow \infty$. Since the subset $\text{gr}(\mathfrak{U}) \subseteq \mathcal{P}(S) \times \mathcal{P}(S \times A)$

⁶A correspondence between topological spaces is continuous if it is both lower- and upper-hemicontinuous (see, e.g., [1, Definition 17.2, p. 558]).

is compact (by the continuity of $\text{pj}_S : \mathcal{P}(S \times A) \rightarrow \mathcal{P}(S)$ and the compactness of $\mathcal{P}(S) \times \mathcal{P}(S \times A)$), there exists a subsequence

$$(\mu^{(n_k)}, \Lambda^{(n_k)})_{k \in \mathbb{N}} \subseteq (\mu^{(n)}, \Lambda^{(n)})_{n \in \mathbb{N}} \quad \text{s.t.} \quad (\mu^{(n_k)}, \Lambda^{(n_k)}) \rightharpoonup (\mu^*, \Lambda^*) \text{ as } k \rightarrow \infty$$

with some $(\mu^*, \Lambda^*) \in \text{Gr}(\mathfrak{U})$. Combined with the limit $\mu^{(n)} \rightharpoonup \mu = \mu^*$, this ensures that $(\Lambda^{(n)})_{n \in \mathbb{N}}$ has a limit point $\Lambda^* \in \mathfrak{U}(\mu) = \mathfrak{U}(\mu^*)$. Thus, by [1, Theorem 17.20], $\mathfrak{U}$ is upper hemicontinuous.

It remains to show the lower hemicontinuity of $\mathfrak{U}$. First note that for every $\mu \in \mathcal{P}(S)$ the set $\mathfrak{U}(\mu) \subseteq \mathcal{P}(S \times A)$ can be represented by

$$(5.6) \quad \mathfrak{U}(\mu) = \bigcup_{\zeta \in \mathcal{P}(A)} \text{Cpl}_{S \times A}(\mu, \zeta).$$

Then we claim that $\text{Cpl}_{S \times A} : \mathcal{P}(S) \times \mathcal{P}(A) \ni (\mu, \zeta) \mapsto \text{Cpl}_{S \times A}(\mu, \zeta) \subseteq \mathcal{P}(S \times A)$ is lower-hemicontinuous. To that end, let $(\mu, \zeta) \in \mathcal{P}(S) \times \mathcal{P}(A)$ and $\Lambda \in \text{Cpl}_{S \times A}(\mu, \zeta)$ be given, and consider a sequence $(\mu^{(n)}, \zeta^{(n)})_{n \in \mathbb{N}} \subseteq \mathcal{P}(S) \times \mathcal{P}(A)$ such that $(\mu^{(n)}, \zeta^{(n)}) \rightharpoonup (\mu, \zeta)$ as $n \rightarrow \infty$.

By Lemma 5.1, for every $n \in \mathbb{N}$ there exists $\Lambda^{(n),*} \in \text{Cpl}_{S \times A}(\mu^{(n)}, \zeta^{(n)})$ such that

$$\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \Lambda^{(n),*}) \leq \mathcal{W}_{\mathcal{P}(S)}(\mu, \mu^{(n)}) + \mathcal{W}_{\mathcal{P}(A)}(\zeta, \zeta^{(n)}).$$

Combined with the limit $(\mu^{(n)}, \zeta^{(n)}) \rightharpoonup (\mu, \zeta)$, this ensures that $\Lambda^{(n),*} \rightharpoonup \Lambda$ as $n \rightarrow \infty$. Thus, by [1, Theorem 17.21], $\text{Cpl}_{S \times A}$ is lower hemicontinuous.

Moreover, by the lower hemicontinuity of $\text{Cpl}_{S \times A}$ and the representation given in (5.6), [1, Theorem 17.27] asserts that $\mathfrak{U}$ is lower hemicontinuous. Therefore, $\mathfrak{U}$ is continuous, as claimed.

Now we prove the part (ii). Let $(\mu, \Lambda, e^0, (\tilde{\mu}, \tilde{\Lambda}, \tilde{e}^0)) \in \text{gr}(\mathfrak{U}) \times \mathcal{P}(E) \times E^0$. For simplicity, let

$$(5.7) \quad \mu' := \bar{F}(\mu, \Lambda, e^0), \quad \tilde{\mu}' := \bar{F}(\tilde{\mu}, \tilde{\Lambda}, \tilde{e}^0).$$

Then, set $\text{id}_E : E \ni e \mapsto \text{id}_E(e) := (e, e) \in E^2$. Then we denote the diagonal coupling of λ_ε by

$$(5.8) \quad \Xi_1 := \lambda_\varepsilon \circ (\text{id}_E(\cdot))^{-1} \in \text{Cpl}_{E \times E}(\lambda_\varepsilon, \lambda_\varepsilon)$$

so that $\mathcal{W}_{\mathcal{P}(E)}(\lambda_\varepsilon, \lambda_\varepsilon) = \int_{E \times E} d_E(e, \tilde{e}) \Xi_1(de, d\tilde{e}) = 0$.

Furthermore, we denote the optimal coupling for $\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda})$ (see [67, Theorem 4.1]) by

$$(5.9) \quad \Xi_2 \in \text{Cpl}_{(S \times A)^2}(\Lambda, \tilde{\Lambda}).$$

Using the couplings Ξ_1 and Ξ_2 , we define a coupling $\Xi_3 \in \text{Cpl}_{(S \times A \times E)^2}(\Lambda \otimes \lambda_\varepsilon, \tilde{\Lambda} \otimes \lambda_\varepsilon)$ by

$$(5.10) \quad \Xi_3(ds, da, de, d\tilde{s}, d\tilde{a}, d\tilde{e}) := \Xi_1(de, d\tilde{e}) \Xi_2(ds, da, d\tilde{s}, d\tilde{a}).$$

By the definition of $\bar{F}$ (see Definition 2.11 (ii)) and the setting (5.7), it holds that

$$\Xi_3 \circ (F(\cdot, \cdot, \Lambda, \cdot, e^0) \times F(\cdot, \cdot, \tilde{\Lambda}, \cdot, \tilde{e}^0))^{-1} \in \text{Cpl}_{S \times S}(\mu', \tilde{\mu}'),$$

i.e., the push-forward of Ξ_3 by $F(\cdot, \cdot, \Lambda, \cdot, e^0) \times F(\cdot, \cdot, \tilde{\Lambda}, \cdot, \tilde{e}^0) : (S, A, E)^2 \rightarrow S^2$.

Then it holds that

$$(5.11) \quad \begin{aligned} \mathcal{W}_{\mathcal{P}(S)}(\mu', \tilde{\mu}') &\leq \int_{S \times S} d_S(s, s') (\Xi_3 \circ (F(\cdot, \cdot, \Lambda, \cdot, e^0) \times F(\cdot, \cdot, \tilde{\Lambda}, \cdot, \tilde{e}^0))^{-1})(ds, ds') \\ &= \int_{(S \times A \times E)^2} d_S(F(s, a, \Lambda, e, e^0), F(\tilde{s}, \tilde{a}, \tilde{\Lambda}, \tilde{e}, \tilde{e}^0)) \Xi_3(ds, da, de, d\tilde{s}, d\tilde{a}, d\tilde{e}) \\ &= \int_{(S \times A)^2} \int_E d_S(F(s, a, \Lambda, e, e^0), F(\tilde{s}, \tilde{a}, \tilde{\Lambda}, e, \tilde{e}^0)) \lambda_\varepsilon(de) \Xi_2(ds, da, d\tilde{s}, d\tilde{a}) =: \text{I}, \end{aligned}$$

where the last line follows from the definition of Ξ_1 and Ξ_3 (see (5.8), (5.10)) and by applying Fubini's theorem (noting that F maps into the compact space S).

By Assumption 2.14 (ii) and the triangle inequality,

$$\begin{aligned} \mathbf{I} &\leq \bar{C}_F \left(\int_{(S \times A)^2} d_{S \times A}((s, a), (\tilde{s}, \tilde{a})) \Xi_2(ds, da, d\tilde{s}, d\tilde{a}) + \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}) + d_{E^0}(e^0, \tilde{e}^0) \right) \\ &= \bar{C}_F (2\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}) + d_{E^0}(e^0, \tilde{e}^0)), \end{aligned}$$

where the last equality follows from the optimality of Ξ_2 (see (5.9)).

Combined with (5.11), this ensures the estimates for $\bar{F}$ to hold.

We next prove the part (iii). Since S , A , and $\mathcal{P}(S \times A)$ are all compact and r is continuous (by Assumption 2.7 (i) and Assumption 2.14 (iii)), $\bar{r}$ is bounded. We prove its $2\bar{C}_r$ -Lipschitz continuity. Let $(\mu, \Lambda), (\tilde{\mu}, \tilde{\Lambda}) \in \text{gr}(\mathfrak{U})$ be given. Then it follows from Assumption 2.14 (iii) and the triangle inequality that for every $\Xi \in \text{Cpl}_{S \times A}(\Lambda, \tilde{\Lambda})$

$$\begin{aligned} |\bar{r}(\mu, \Lambda) - \bar{r}(\tilde{\mu}, \tilde{\Lambda})| &= \left| \int_{(S \times A)^2} (r(s, a, \Lambda) - r(\tilde{s}, \tilde{a}, \tilde{\Lambda})) \Xi(ds, da, d\tilde{s}, d\tilde{a}) \right| \\ &\leq \bar{C}_r \left(\int_{S \times A} d_{S \times A}((s, a), (\tilde{s}, \tilde{a})) \Xi(ds, da, d\tilde{s}, d\tilde{a}) + \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda, \tilde{\Lambda}) \right). \end{aligned}$$

By taking infimum over all $\Xi \in \text{Cpl}_{S \times A}(\Lambda, \tilde{\Lambda})$ into the above, we can obtain the estimate for $\bar{r}$.

This completes the proof. $\square$

Using the two preceding lemmas, we now proceed to prove Proposition 2.15.

Proof of Proposition 2.15. We start by proving (i). Let $L \geq 0$ and $\bar{V} \in \text{Lip}_{b,L}(\mathcal{P}(S); \mathbb{R})$ be given. Set $\mathcal{S} := \mathcal{P}(S \times A) \times \mathfrak{P}^0$. Recalling the definition of $\bar{p}$ (see Definition 2.11 (iii)), define $G : \mathcal{S} \ni (\Lambda, p) \mapsto G(\Lambda, p) \in \mathbb{R}$ by

$$(5.12) \quad G(\Lambda, p) := \int_{\mathcal{P}(S)} \bar{V}(\mu') \bar{p}(d\mu' | \text{pj}_S(\Lambda), \Lambda, p) = \int_{E^0} \bar{V}(\bar{F}(\text{pj}_S(\Lambda), \Lambda, e^0)) p(de^0).$$

We claim that G is continuous. Consider a sequence $(\Lambda^{(n)}, p^{(n)})_{n \in \mathbb{N}} \subseteq \mathcal{S}$ such that $(\Lambda^{(n)}, p^{(n)}) \rightharpoonup (\Lambda^*, p^*)$ as $n \rightarrow \infty$, with some $(\Lambda^*, p^*) \in \mathcal{S}$.

By the triangle inequality, for every $n \in \mathbb{N}$,

$$\begin{aligned} |G(\Lambda^{(n)}, p^{(n)}) - G(\Lambda^*, p^*)| &\leq |G(\Lambda^*, p^{(n)}) - G(\Lambda^*, p^*)| + |G(\Lambda^{(n)}, p^{(n)}) - G(\Lambda^*, p^{(n)})| \\ &=: \mathbf{I}^{(n)} + \mathbf{II}^{(n)}. \end{aligned}$$

We will show that $\mathbf{I}^{(n)}$ and $\mathbf{II}^{(n)}$ vanish as $n \rightarrow \infty$.

Since $\bar{V} \in \text{Lip}_{b,L}(\mathcal{P}(S); \mathbb{R})$ and $\bar{F}$ is continuous (see Lemma 5.2 (ii)), it holds that $g^*(\cdot) := \bar{V}(\bar{F}(\text{pj}_S(\Lambda^*), \Lambda^*, \cdot)) \in C_b(E_0; \mathbb{R})$. Combined with the limit $p^{(n)} \rightharpoonup p^*$, this ensures that

$$\lim_{n \rightarrow \infty} \mathbf{I}^{(n)} = \lim_{n \rightarrow \infty} \left| \int_{E^0} g^*(e^0) p^{(n)}(de^0) - \int_{E^0} g^*(\tilde{e}^0) p^*(d\tilde{e}^0) \right| = 0.$$

It remains to show the limit of $\mathbf{II}^{(n)}$. We use the L -Lipschitz continuity of $\bar{V}$, the estimate of $\bar{F}$ given in Lemma 5.2 (ii), and the limits $\Lambda^{(n)} \rightharpoonup \Lambda^*$ and $p^{(n)} \rightharpoonup p^*$ to obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{II}^{(n)} &\leq \lim_{n \rightarrow \infty} \int_{E^0} \left| \bar{V}(\bar{F}(\text{pj}_S(\Lambda^{(n)}), \Lambda^{(n)}, e^0)) - \bar{V}(\bar{F}(\text{pj}_S(\Lambda^*), \Lambda^*, e^0)) \right| p^{(n)}(de^0) \\ &\leq 2L\bar{C}_F \lim_{n \rightarrow \infty} \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda^{(n)}, \Lambda^*) = 0. \end{aligned}$$

Therefore G given in (5.12) is continuous, as claimed.

Since $\mathfrak{P}^0$ is compact (see Assumption 2.7 (i)) and G is continuous, an application of Berge's maximum theorem (see, e.g., [1, Theorem 17.31]) ensures the continuity of the map $\bar{J} : \mathcal{P}(S \times A) \ni \Lambda \mapsto \bar{J}(\Lambda) \in \mathbb{R}$ given by

$$(5.13) \quad \bar{J}(\Lambda) := \inf_{p \in \mathfrak{P}^0} \int_{\mathcal{P}(S)} \bar{V}(\mu') \bar{p}(d\mu' | \text{pj}_S(\Lambda), \Lambda, p),$$

and the existence of the measurable selector $\bar{p}^* : \mathcal{P}(S \times A) \ni \Lambda \mapsto \bar{p}^*(\Lambda) \in \mathfrak{P}^0$ satisfying (2.23).

We now prove the part (ii). In analogy to the part (i), the key idea is to apply Berge's maximum theorem. To that end, we first show that a map $H : \text{gr}(\mathfrak{U}) \in (\mu, \Lambda) \mapsto H(\mu, \Lambda) \in \mathbb{R}$ defined by

$$(5.14) \quad H(\mu, \Lambda) := \bar{r}(\mu, \Lambda) + \beta \cdot \bar{J}(\Lambda),$$

with $\bar{J} : \mathcal{P}(S \times A) \rightarrow \mathbb{R}$ defined in (5.13) is continuous. That will be achieved in two steps.

Consider a sequence $(\mu^{(n)}, \Lambda^{(n)})_{n \in \mathbb{N}} \subseteq \text{gr}(\mathfrak{U})$ such that $(\mu^{(n)}, \Lambda^{(n)}) \rightharpoonup (\mu^*, \Lambda^*)$ as $n \rightarrow \infty$, with some $(\mu^*, \Lambda^*) \in \text{gr}(\mathfrak{U})$. By the triangle inequality, it holds that for every $n \in \mathbb{N}$,

$$\begin{aligned} |H(\mu^{(n)}, \Lambda^{(n)}) - H(\mu^*, \Lambda^*)| &\leq |\bar{r}(\mu^{(n)}, \Lambda^{(n)}) - \bar{r}(\mu^*, \Lambda^*)| + \beta \cdot |\bar{J}(\Lambda^{(n)}) - \bar{J}(\Lambda^*)| \\ &=: \text{III}^{(n)} + \beta \cdot |\text{IV}^{(n)}|. \end{aligned}$$

The limit of $\text{III}^{(n)}$ is straightforward. Indeed, by Lemma 5.2 (iii) and the limit $\Lambda^{(n)} \rightharpoonup \Lambda^*$,

$$\lim_{n \rightarrow \infty} \text{III}^{(n)} \leq 2\bar{C}_r \lim_{n \rightarrow \infty} \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda^{(n)}, \Lambda^*) = 0.$$

It remains to show the limit of $|\text{IV}^{(n)}|$. Recalling the measurable selector $\bar{p}^*$ defined as in the part (i), denote by $p^* := \bar{p}^*(\Lambda^*) \in \mathfrak{P}^0$. Then it holds that

$$(5.15) \quad \bar{J}(\Lambda^*) = \int_{\mathcal{P}(S)} \bar{V}(\mu') \bar{p}(d\mu' | \text{pj}_S(\Lambda^*), \Lambda^*, p^*) = \int_{E^0} \bar{V}(\bar{F}(\mu^*, \Lambda^*, e^0)) p^*(de^0),$$

noting that $\text{pj}_S(\Lambda^*) = \mu^*$ as $(\mu^*, \Lambda^*) \in \text{gr}(\mathfrak{U})$.

On the other hand, as $p^* \in \mathfrak{P}^0$ does not necessarily optimize $\bar{J}(\Lambda^{(n)})$, it holds that

$$(5.16) \quad \bar{J}(\Lambda^{(n)}) \leq \int_{\mathcal{P}(S)} \bar{V}(\mu') \bar{p}(d\mu' | \text{pj}_S(\Lambda^{(n)}), \Lambda^{(n)}, p^*) = \int_{E^0} \bar{V}(\bar{F}(\mu^{(n)}, \Lambda^{(n)}, e^0)) p^*(de^0),$$

with $\text{pj}_S(\Lambda^{(n)}) = \mu^{(n)}$.

By (5.15) and (5.16), it holds that for every $n \in \mathbb{N}$ and every $\Gamma \in \text{Cpl}_{E^0 \times E^0}(p^*, p^*)$,

$$\begin{aligned} \text{IV}^{(n)} &\leq \int_{E^0} \bar{V}(\bar{F}(\mu^{(n)}, \Lambda^{(n)}, e^0)) p^*(de^0) - \int_{E^0} \bar{V}(\bar{F}(\mu^*, \Lambda^*, e^0)) p^*(de^0) \\ (5.17) \quad &= \int_{E^0 \times E^0} \left(\bar{V}(\bar{F}(\mu^{(n)}, \Lambda^{(n)}, e^0)) - \bar{V}(\bar{F}(\mu^*, \Lambda^*, \tilde{e}^0)) \right) \Gamma(de^0, d\tilde{e}^0) \\ &\leq 2L\bar{C}_F \cdot \left(\mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda^{(n)}, \Lambda^*) + \int_{E^0 \times E^0} d_{E^0}(e^0, \tilde{e}^0) \Gamma(de^0, d\tilde{e}^0) \right), \end{aligned}$$

where the last inequality follows from the L -Lipschitz continuity of $\bar{V}$ and the estimate of $\bar{F}$ given in Lemma 5.2 (ii).

By taking infimum over $\Gamma \in \text{Cpl}_{E^0 \times E^0}(p^*, p^*)$ in the last equation of (5.17), we have

$$(5.18) \quad \text{IV}^{(n)} \leq 2L\bar{C}_F \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda^{(n)}, \Lambda^*).$$

Using the same arguments as presented for (5.18), one can have the lower bound with the same constant, i.e., $\text{IV}^{(n)} \geq -2L\bar{C}_F \mathcal{W}_{\mathcal{P}(S \times A)}(\Lambda^{(n)}, \Lambda^*)$.

Combined with the limit $\Lambda^{(n)} \rightharpoonup \Lambda^*$, this ensures that $|\text{IV}^{(n)}|$ vanishes as $n \rightarrow \infty$. Therefore H given in (5.14) is continuous as claimed.

Since $\mathfrak{U}$ is non-empty, compact-valued, and continuous (see Lemma 5.2 (ii)) and H is continuous, an application of Berge's maximum theorem ensures the continuity of $\mathcal{T}\bar{V}$ (see (2.22)) and the existence of the measurable selector $\bar{\pi}^* : \mathcal{P}(S) \ni \mu \mapsto \bar{\pi}^*(\mu) \in \mathfrak{U}(\mu)$ satisfying (2.24). This completes the proof. $\square$

5.3. Proof of Proposition 2.16. Let $\bar{V} \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$. We claim that $\mathcal{T}\bar{V} \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$. From Lemma 5.2 (iii) and the fact that $\bar{V} \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$, the boundedness of $\mathcal{T}\bar{V}$ is straightforward. To verify the $\bar{L}$ -Lipschitz continuity of $\mathcal{T}\bar{V}$, let $\mu, \tilde{\mu} \in \mathcal{P}(S)$ and denote by

$$(5.19) \quad D(\mu, \tilde{\mu}) := \mathcal{T}\bar{V}(\mu) - \mathcal{T}\bar{V}(\tilde{\mu}).$$

Then let $\bar{\pi}^*(\mu) \in \mathfrak{U}(\mu)$ be the local maximizer of $\mathcal{T}\bar{V}(\mu)$ (see Proposition 2.15 (ii)). Then, denote by $\zeta^\diamond := \text{pj}_A(\bar{\pi}^*(\mu)) \in \mathcal{P}(A)$ the marginal of $\bar{\pi}^*(\mu) \in \mathfrak{U}(\mu) \subset \mathcal{P}(S \times A)$ on A . Since $\bar{\pi}^*(\mu) \in \text{Cpl}_{S \times A}(\mu, \zeta^\diamond)$, by Lemma 5.1 (i) there exists a coupling $\tilde{\Lambda}^\diamond \in \text{Cpl}_{S \times A}(\tilde{\mu}, \zeta^\diamond)$ such that

$$(5.20) \quad \mathcal{W}_{\mathcal{P}(S \times A)}(\bar{\pi}^*(\mu), \tilde{\Lambda}^\diamond) \leq \mathcal{W}_{\mathcal{P}(S)}(\mu, \tilde{\mu}).$$

Then since $\tilde{\Lambda}^\diamond \in \mathfrak{U}(\tilde{\mu})$ (which does not necessarily maximize $\mathcal{T}\bar{V}(\tilde{\mu})$), it holds that

$$D(\mu, \tilde{\mu}) \leq \bar{r}(\mu, \bar{\pi}^*(\mu)) - \bar{r}(\tilde{\mu}, \tilde{\Lambda}^\diamond) + \beta \cdot \bar{J}(\bar{\pi}^*(\mu)) - \beta \cdot \bar{J}(\tilde{\Lambda}^\diamond) =: D^1(\mu, \tilde{\mu}),$$

recalling $\bar{J} : \mathcal{P}(S \times A) \rightarrow \mathbb{R}$ defined in (5.13) (with noting that $\text{pj}_S(\bar{\pi}^*(\mu)) = \mu$ and $\text{pj}_S(\tilde{\Lambda}^\diamond) = \tilde{\mu}$).

Let $\bar{p}^*(\tilde{\Lambda}^\diamond) \in \mathfrak{P}^0$ be the local minimizers of $\bar{J}(\tilde{\Lambda}^\diamond)$ (see Proposition 2.15 (i)). Since they do not necessarily minimize $\bar{J}(\bar{\pi}^*(\mu))$, it holds that

$$(5.21) \quad \begin{aligned} D^1(\mu, \tilde{\mu}) &\leq \bar{r}(\mu, \bar{\pi}^*(\mu)) - \bar{r}(\tilde{\mu}, \tilde{\Lambda}^\diamond) + \beta \int_{E^0} \bar{V}(\bar{F}(\mu, \bar{\pi}^*(\mu), e^0)) \bar{p}^*(\tilde{\Lambda}^\diamond)(de^0) \\ &\quad - \beta \int_{\mathcal{P}(S)} \bar{V}(\bar{F}(\tilde{\mu}, \tilde{\Lambda}^\diamond, \tilde{e}^0)) \bar{p}^*(\tilde{\Lambda}^\diamond)(d\tilde{e}^0) =: D^2(\mu, \tilde{\mu}), \end{aligned}$$

recalling the definition of $\bar{p}$ given in Definition 2.11 (iii).

Let $\Gamma \in \text{Cpl}_{E^0 \times E^0}(\bar{p}^*(\tilde{\Lambda}^\diamond), \bar{p}^*(\tilde{\Lambda}^\diamond))$ be some arbitrary. Then, by the estimates for $\bar{r}$ and $\bar{F}$ (given in Lemma 5.2 (ii), (iii)) and $\bar{V} \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$, it holds that

$$(5.22) \quad \begin{aligned} D^2(\mu, \tilde{\mu}) &\leq |\bar{r}(\mu, \bar{\pi}^*(\mu)) - \bar{r}(\tilde{\mu}, \tilde{\Lambda}^\diamond)| + \beta \int_{E^0 \times E^0} |\bar{V}(\bar{F}(\mu, \bar{\pi}^*(\mu), e^0)) - \bar{V}(\bar{F}(\tilde{\mu}, \tilde{\Lambda}^\diamond, \tilde{e}^0))| \Gamma(de^0, d\tilde{e}^0) \\ &\leq 2\bar{C}_r \mathcal{W}_{\mathcal{P}(S \times A)}(\bar{\pi}^*(\mu), \tilde{\Lambda}^\diamond) \\ &\quad + \bar{C}_F \bar{L} \beta \left(2\mathcal{W}_{\mathcal{P}(S \times A)}(\bar{\pi}^*(\mu), \tilde{\Lambda}^\diamond) + \int_{E^0 \times E^0} d_{E^0}(e^0, \tilde{e}^0) \Gamma(de^0, d\tilde{e}^0) \right). \end{aligned}$$

For the last line of (5.22), we take infimum over all $\Gamma \in \text{Cpl}_{E^0 \times E^0}(\bar{p}^*(\tilde{\Lambda}^\diamond), \bar{p}^*(\tilde{\Lambda}^\diamond))$ and then use the estimate given in (5.20) to obtain

$$(5.23) \quad D^2(\mu, \tilde{\mu}) \leq (2\bar{C}_r + 2\bar{C}_F \bar{L} \beta) \mathcal{W}_{\mathcal{P}(S)}(\mu, \tilde{\mu}) \leq \bar{L} \mathcal{W}_{\mathcal{P}(S)}(\mu, \tilde{\mu}),$$

where the last inequality holds by the inequality $\bar{L} \geq 2\bar{C}_r / (1 - 2\bar{C}_F \beta)$ with $2\bar{C}_F \beta < 1$.

By (5.19), (5.21) and (5.23), we have that

$$\mathcal{T}\bar{V}(\mu) - \mathcal{T}\bar{V}(\tilde{\mu}) = D(\mu, \tilde{\mu}) \leq D^1(\mu, \tilde{\mu}) \leq D^2(\mu, \tilde{\mu}) \leq \bar{L} \mathcal{W}_{\mathcal{P}(S)}(\mu, \tilde{\mu}).$$

Since $\mu, \tilde{\mu} \in \mathcal{P}(S)$ are chosen arbitrary, one can have that $\mathcal{T}\bar{V}(\cdot)$ is $\bar{L}$ -Lipschitz continuous. Hence, we conclude that $\mathcal{T}\bar{V} \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$.

To verify (2.25), let $\bar{V}, \bar{W} \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$. By Proposition 2.15 (ii), for every $\mu \in \mathcal{P}(S)$

$$|\mathcal{T}\bar{V}(\mu) - \mathcal{T}\bar{W}(\mu)| \leq \beta \sup_{p \in \mathfrak{P}^0} \int_{\mathcal{P}(S)} |\bar{V}(\mu') - \bar{W}(\mu')| \bar{p}(d\mu' | \mu, \bar{\pi}^*(\mu), p) \leq \beta \|\bar{V} - \bar{W}\|_\infty,$$

which ensures (2.25) to hold.

Since $\beta < 1$ and $\mathcal{T}(\text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})) \subseteq \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$, $\mathcal{T}$ is a contraction on $\text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$. Hence, an application of the Banach's fixed point theorem ensures the existence and uniqueness of $\bar{V}^* \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$ such that for every $\bar{V} \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$, $\bar{V}^* = \mathcal{T}\bar{V}^* = \lim_{n \rightarrow \infty} \mathcal{T}^n \bar{V}$. This completes the proof. $\square$

6. PROOF OF RESULTS IN SECTION 2.4

We begin by presenting an observation that plays a key role in the proof of Lemmas 2.17 and 2.20. Recall the set $\mathcal{Q}$ given in Definition 2.2 and the filtration $\mathbb{G} = (\mathcal{G}_t)_{t \geq 0}$ given in (2.10).

Lemma 6.1. *Denote for every $t \geq 0$ by $L_{\mathcal{G}_t}^0(Z)$ the set of all $\mathcal{G}_t$ measurable random variables ζ_t with values in a compact Polish space Z . Then for every $\zeta_0 \in L_{\mathcal{G}_0}^0(Z)$ and $\mathbb{P}, \tilde{\mathbb{P}} \in \mathcal{Q}$, it holds that $\mathcal{L}_{\mathbb{P}}(\zeta_0) = \mathcal{L}_{\tilde{\mathbb{P}}}(\zeta_0)$. Furthermore, for every $t \geq 1$, $\zeta_t \in L_{\mathcal{G}_t}^0(Z)$, and $\mathbb{P}, \tilde{\mathbb{P}} \in \mathcal{Q}$, it holds that $\mathcal{L}_{\mathbb{P}}(\zeta_t | \varepsilon_{1:t}^0) = \mathcal{L}_{\tilde{\mathbb{P}}}(\zeta_t | \varepsilon_{1:t}^0)$, $\mathbb{P}$ -a.s..*

Proof. Without loss of generality, we consider the case $t \geq 1$, as the case $t = 0$ can be subsumed into it. Then, let $\zeta_t \in L_{\mathcal{G}_t}^0(Z)$ and $\mathbb{P}, \tilde{\mathbb{P}} \in \mathcal{Q}$ be given.

By the same arguments presented for the proof of Lemma 4.1 (ii), $\mathcal{L}_{\mathbb{P}}(\zeta_t | \varepsilon_{1:t}^0)$ and $\mathcal{L}_{\tilde{\mathbb{P}}}(\zeta_t | \varepsilon_{1:t}^0)$ are $\mathcal{F}_t^0$ measurable. Hence it suffices to show that for any bounded Borel measurable functions $\hat{g}_t : (E^0)^t \rightarrow \mathbb{R}$ and $\hat{f} : Z \rightarrow \mathbb{R}$,

$$\mathbb{E}^{\mathbb{P}}[\hat{g}_t(\varepsilon_{1:t}^0) \hat{f}(\zeta_t)] = \mathbb{E}^{\mathbb{P}}\left[\hat{g}_t(\varepsilon_{1:t}^0) \int_Z \hat{f}(\tilde{z}) \mathcal{L}_{\tilde{\mathbb{P}}}(\zeta_t | \varepsilon_{1:t}^0)(d\tilde{z})\right].$$

Note that since ζ_t is $\mathcal{G}_t$ measurable, there exists a Borel measurable function $\hat{l} : G \times \Theta^{t+1} \times E^t \times (E^0)^t \rightarrow Z$ such that $\zeta = \hat{l}(\gamma, \vartheta_{0:t}, \varepsilon_{1:t}, \varepsilon_{1:t}^0)$.

Moreover, since $\varepsilon_{1:t}^0$ is independent of $\gamma, \vartheta_{0:t}, \varepsilon_{1:t}$ (see Remark 2.3 (i)),

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}[\hat{g}_t(\varepsilon_{1:t}^0) \hat{f}(\zeta_t)] &= \mathbb{E}^{\mathbb{P}}[\hat{g}_t(\varepsilon_{1:t}^0) \hat{f}(\hat{l}(\gamma, \vartheta_{0:t}, \varepsilon_{1:t}, \varepsilon_{1:t}^0))] \\ &= \int_{(E^0)^t} \hat{g}_t(e_{1:t}^0) \mathbb{E}^{\mathbb{P}}[\hat{f}(\hat{l}(\gamma, \vartheta_{0:t}, \varepsilon_{1:t}, e_{1:t}^0))] \mathcal{L}_{\mathbb{P}}(\varepsilon_{1:t}^0)(de_{1:t}^0) \\ &= \int_{(E^0)^t} \hat{g}_t(e_{1:t}^0) \mathbb{E}^{\tilde{\mathbb{P}}}[\hat{f}(\hat{l}(\gamma, \vartheta_{0:t}, \varepsilon_{1:t}, e_{1:t}^0))] \mathcal{L}_{\tilde{\mathbb{P}}}(\varepsilon_{1:t}^0)(de_{1:t}^0) =: \text{I}_t, \end{aligned}$$

where the second equality holds by Fubini's theorem and the last equality follows from the fact that $\mathcal{L}_{\mathbb{P}}(\gamma, \vartheta_{0:t}, \varepsilon_{1:t}) = \mathcal{L}_{\tilde{\mathbb{P}}}(\gamma, \vartheta_{0:t}, \varepsilon_{1:t})$ (see Remark 2.3 (ii)).

Therefore, by definition of $\mathcal{L}_{\tilde{\mathbb{P}}}(\zeta_t | \varepsilon_{1:t}^0)$ and $\mathcal{L}_{\mathbb{P}}(\varepsilon_{1:t}^0)$,

$$\begin{aligned} \text{I}_t &= \int_{(E^0)^t} \hat{g}_t(e_{1:t}^0) \mathbb{E}^{\tilde{\mathbb{P}}}[\mathbb{E}^{\tilde{\mathbb{P}}}[\hat{f}(\zeta) | \varepsilon_{1:t}^0 = e_{1:t}^0]] \mathcal{L}_{\mathbb{P}}(\varepsilon_{1:t}^0)(de_{1:t}^0) \\ &= \int_{(E^0)^t} \hat{g}_t(e_{1:t}^0) \left(\int_Z \hat{f}(z) \mathcal{L}_{\tilde{\mathbb{P}}}(\zeta | \varepsilon_{1:t}^0 = e_{1:t}^0)(dz) \right) \mathcal{L}_{\mathbb{P}}(\varepsilon_{1:t}^0)(de_{1:t}^0) \\ &= \mathbb{E}^{\mathbb{P}}\left[\hat{g}_t(\varepsilon_{1:t}^0) \int_Z \hat{f}(\tilde{z}) \mathcal{L}_{\tilde{\mathbb{P}}}(\zeta_t | \varepsilon_{1:t}^0)(d\tilde{z})\right], \end{aligned}$$

as claimed. $\square$

6.1. Proof of Lemma 2.17. We first prove (2.26). Let $a \in \mathcal{A}$ be given. We will construct $\underline{p}_1^{\xi,a} \in \mathfrak{P}^0$ and the sequence of kernels $\underline{p}_t^{\xi,a} : (E^0)^{t-1} \ni e_{1:t-1}^0 \mapsto \underline{p}_t^{\xi,a}(e_t^0 | e_{1:t-1}^0) \in \mathfrak{P}^0$ for $t \geq 2$ to define $\underline{\mathbb{P}}^{\xi,a} \in \mathcal{Q}$ induced by $(\underline{p}_t^{\xi,a})_{t \geq 1} \in \mathcal{K}^0$.

Step 1: Let $\tilde{\mathbb{P}} \in \mathcal{Q}$ be some arbitrary. Then set

$$(6.1) \quad \tilde{s}_0 := \xi, \quad \tilde{\Lambda}_0 := \mathcal{L}_{\tilde{\mathbb{P}}}((\tilde{s}_0, a_0)),$$

and define by

$$(6.2) \quad \underline{p}_1^{\xi, a} := \bar{p}^*(\tilde{\Lambda}_0) \in \mathfrak{P}^0,$$

where $\bar{p}^*$ is given in Proposition 2.15 (i).

Next set

$$(6.3) \quad \tilde{s}_1 := F(\tilde{s}_0, a_0, \tilde{\Lambda}_0, \varepsilon_1, \varepsilon_1^0), \quad \tilde{\Lambda}_1 := \mathcal{L}_{\tilde{\mathbb{P}}}((\tilde{s}_1, a_1) | \varepsilon_1^0),$$

where $(\tilde{s}_0, \tilde{\Lambda}_0)$ are given in (6.1). We see that $(\tilde{s}_1, a_1)$ are $\mathcal{G}_1$ measurable (because $\tilde{s}_0 \in L_{\mathcal{F}_0}^0(S)$, $a_0 = \pi_0(\gamma, \vartheta_0)$, $a_1 = \pi_1(\gamma, \vartheta_{0:1}, \varepsilon_1, \varepsilon_1^0)$) and ε_1^0 is independent of $(\gamma, \vartheta_{0:1}, \varepsilon_1)$ (see Remark 2.3 (iii)).

Moreover, an application of Lemma A.1 (ii) implies that $\tilde{\Lambda}_1$ is $\mathcal{F}_1^0$ measurable, which ensures the existence of a Borel measurable function $l_1 : E^0 \rightarrow \mathcal{P}(S \times A)$ such that

$$(6.4) \quad l_1(\varepsilon_1^0) = \tilde{\Lambda}_1.$$

From this, define $\underline{p}_2^{\xi, a} : E^0 \ni e_1^0 \mapsto \underline{p}_2^{\xi, a}(\cdot | e_1^0) \in \mathcal{P}(E^0)$ by

$$(6.5) \quad \underline{p}_2^{\xi, a}(\cdot | e_1^0) := \bar{p}^*(l_1(e_1^0)) \in \mathfrak{P}^0.$$

Using the same arguments presented for (6.3)–(6.5), for every $t \geq 1$ we inductively set

$$(6.6) \quad \tilde{s}_t := F(\tilde{s}_{t-1}, a_{t-1}, \tilde{\Lambda}_{t-1}, \varepsilon_t, \varepsilon_t^0), \quad \tilde{\Lambda}_t := \mathcal{L}_{\tilde{\mathbb{P}}}((\tilde{s}_t, a_t) | \varepsilon_{1:t}^0),$$

where $(\tilde{s}_t, a_t)$ are $\mathcal{G}_t$ measurable, and $\tilde{\Lambda}_t$ is $\mathcal{F}_t^0$ measurable.

Hence, there exists a Borel measurable function $l_t : (E^0)^t \rightarrow \mathcal{P}(S \times A)$ such that

$$(6.7) \quad l_t(\varepsilon_{1:t}^0) = \tilde{\Lambda}_t.$$

From this, define $\underline{p}_{t+1}^{\xi, a} : (E^0)^t \ni e_{1:t}^0 \mapsto \underline{p}_{t+1}^{\xi, a}(\cdot | e_{1:t}^0) \in \mathcal{P}(E^0)$ by

$$(6.8) \quad \underline{p}_{t+1}^{\xi, a}(\cdot | e_{1:t}^0) := \bar{p}^*(l_t(e_{1:t}^0)) \in \mathfrak{P}^0.$$

Using $(\underline{p}_t^{\xi, a})_{t \geq 1} \in \mathcal{K}^0$, constructed via (6.2), (6.5), and (6.8), we define the measure $\underline{\mathbb{P}}^{\xi, a} \in \mathcal{Q}$ induced by this sequence. We underline that the existence of such a measure is ensured by Ionescu–Tulcea’s theorem (see Remark 2.3), and that the above inductive construction is invariant and can be carried out under any $\tilde{\mathbb{P}} \in \mathcal{Q}$.

Step 2: Recall for every $t \geq 0$, $\tilde{\Lambda}_t$ is the conditional joint law of $(\tilde{s}_t, a_t)$ given $\varepsilon_{1:t}^0$ under $\tilde{\mathbb{P}}$, as given in (6.1), (6.3), and (6.6). We claim that for every $t \geq 0$, $\underline{\mathbb{P}}^{\xi, a}$ -a.s.

$$(6.9) \quad s_t^{\xi, a, \underline{\mathbb{P}}^{\xi, a}} = \tilde{s}_t, \quad \underline{\Lambda}_t^{\xi, a} = \tilde{\Lambda}_t,$$

where $\underline{\Lambda}_t^{\xi, a}$ is the conditional joint law of $(s_t^{\xi, a, \underline{\mathbb{P}}^{\xi, a}}, a_t)$ given $\varepsilon_{1:t}^0$ under $\underline{\mathbb{P}}^{\xi, a}$.

The proof uses an induction over $t \geq 0$: For $t = 0$, clearly $s_0^{\xi, a, \underline{\mathbb{P}}^{\xi, a}} = \tilde{s}_0 = \xi \in L_{\mathcal{F}_0}^0(S)$. Moreover, since a_0 is $\mathcal{G}_0$ measurable (noting that $\mathcal{G}_0 = \sigma(\gamma, \vartheta_0)$) and $\mathcal{L}_{\underline{\mathbb{P}}^{\xi, a}}(\gamma, \vartheta_0) = \mathcal{L}_{\tilde{\mathbb{P}}}(\gamma, \vartheta_0)$ (see Remark 2.3 (ii)), it holds that $\underline{\Lambda}_0^{\xi, a} = \tilde{\Lambda}_0$.

Assume that the induction claim holds true for some $t \geq 0$. For the case $t+1$, by the conditional McKean-Vlasov dynamics given in (2.12) and the induction hypothesis for t , it holds that $\underline{\mathbb{P}}^{\xi, a}$ -a.s.,

$$(6.10) \quad \begin{aligned} s_{t+1}^{\xi, a, \underline{\mathbb{P}}^{\xi, a}} &= F(s_t^{\xi, a, \underline{\mathbb{P}}^{\xi, a}}, a_t, \underline{\Lambda}_t^{\xi, a}, \varepsilon_{t+1}, \varepsilon_{t+1}^0) \\ &= F(\tilde{s}_t, a_t, \tilde{\Lambda}_t, \varepsilon_{t+1}, \varepsilon_{t+1}^0) = \tilde{s}_{t+1}, \end{aligned}$$

where the second equality holds by the Borel measurability of F (see Definition 2.4 (i)), and the last equality holds by definition (6.6), as claimed.

We now show that $\underline{\Lambda}_{t+1}^{\xi,a} = \tilde{\Lambda}_{t+1}$, $\mathbb{P}^{\xi,a}$ -a.s.. By $\mathcal{F}_{t+1}^0$ -measurability of $(\Lambda_{t+1}^{\xi,a}, \tilde{\Lambda}_{t+1})$, it suffices to show that for any bounded Borel measurable functions $\hat{g}_{t+1} : (E^0)^{t+1} \rightarrow \mathbb{R}$ and $\hat{f} : S \times A \rightarrow \mathbb{R}$,

$$(6.11) \quad \mathbb{E}^{\mathbb{P}^{\xi,a}}[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi,a, \mathbb{P}^{\xi,a}}, a_{t+1})] = \mathbb{E}^{\mathbb{P}^{\xi,a}}\left[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \int_{S \times A} f(\tilde{s}, \tilde{a}) \tilde{\Lambda}_{t+1}(d\tilde{s}, d\tilde{a})\right].$$

Indeed, by (6.10),

$$\mathbb{E}^{\mathbb{P}^{\xi,a}}[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi,a, \mathbb{P}^{\xi,a}}, a_{t+1})] = \mathbb{E}^{\mathbb{P}^{\xi,a}}[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(\tilde{s}_{t+1}, a_{t+1})] =: \mathbf{I}^{t+1}.$$

Moreover, since $(\tilde{s}_{t+1}, a_{t+1})$ are $\mathcal{G}_{t+1}$ measurable (with $\mathcal{G}_{t+1} = \sigma(\gamma, \vartheta_{0:t+1}, \varepsilon_{1:t+1}, \varepsilon_{1:t+1}^0)$), an application of Lemma 6.1 ensures that $\mathbb{P}^{\xi,a}$ -a.s.,

$$\mathcal{L}_{\mathbb{P}^{\xi,a}}((\tilde{s}_{t+1}, a_{t+1}) | \varepsilon_{1:t+1}^0) = \mathcal{L}_{\tilde{\mathbb{P}}}((\tilde{s}_{t+1}, a_{t+1}) | \varepsilon_{1:t+1}^0) = \tilde{\Lambda}_{t+1},$$

which implies that $\mathbf{I}^{t+1}$ equals the second term given in (6.11), as claimed.

By induction hypothesis, the claim (6.9) holds for all $t \geq 0$.

Step 3: Recall that $\mathbb{P}^{\xi,a} \in \mathcal{Q}$ is the measure induced by $(p_t^{\xi,a})_{t \geq 1} \in \mathcal{K}^0$ given in (6.2), (6.5), and (6.8) (see Step 1). Then from Remark 2.3 (iii), it holds that $\mathbb{P}^{\xi,a}$ -a.s.

$$(6.12) \quad \begin{aligned} \mathcal{L}_{\mathbb{P}^{\xi,a}}(\varepsilon_1^0) &= p_1^{\xi,a} \in \mathfrak{P}^0, \\ \mathcal{L}_{\mathbb{P}^{\xi,a}}(\varepsilon_t^0 | \mathcal{F}_{t-1}^0) &= p_t^{\xi,a}(\cdot | \varepsilon_{1:t-1}^0) \in \mathfrak{P}^0 \text{ for all } t \geq 2. \end{aligned}$$

Moreover, since $\underline{\Lambda}_t^{\xi,a} = \tilde{\Lambda}_t$ $\mathbb{P}^{\xi,a}$ -a.s. for all $t \geq 0$ (see (6.9) in Step 2), it holds that $\mathbb{P}^{\xi,a}$ -a.s.

$$(6.13) \quad \underline{p}_1^{\xi,a} = \bar{p}^*(\underline{\Lambda}_0^{\xi,a}), \quad \underline{p}_t^{\xi,a}(\cdot | \varepsilon_{1:t-1}^0) = \bar{p}^*(\underline{\Lambda}_{t-1}^{\xi,a}) \text{ for all } t \geq 2,$$

which ensures (2.26) to hold, as claimed.

The proof for (2.27) is straightforward. Indeed, by (2.19) in Proposition 2.12 it holds that $\mathbb{P}^{\xi,a}$ -a.s.

$$\begin{aligned} \mathcal{L}_{\mathbb{P}^{\xi,a}}(\mu_1^{\xi,a}) &= \bar{p}(\cdot | \text{pj}_S(\underline{\Lambda}_0^{\xi,a}), \underline{\Lambda}_0^{\xi,a}, p_1^{\xi,a}(\cdot)) \\ \mathcal{L}_{\mathbb{P}^{\xi,a}}(\mu_{t+1}^{\xi,a}) &= \bar{p}(\cdot | \text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}, p_t^{\xi,a}(\cdot | \varepsilon_{1:t-1}^0)) \text{ for all } t \geq 1. \end{aligned}$$

Combined with (6.13), this ensures (2.27) to hold, as claimed. This completes the proof. $\square$

6.2. Proof of Lemma 2.20. We first introduce some kernels used for constructing $a^* \in \mathcal{A}$. We denote by

$$(6.14) \quad \mathcal{K}_{S \times A} : S \times \mathcal{P}(S \times A) \times \mathcal{P}(S) \ni (s, \Lambda, \mu) \mapsto \mathcal{K}_{S \times A}(\cdot | s, \Lambda, \mu) \in \mathcal{P}(A)$$

the universal disintegration kernel (see Lemma A.3). Then, we define a kernel

$$(6.15) \quad \psi^* : S \times \mathcal{P}(S) \ni (s, \mu) \mapsto \psi^*(\cdot | s, \mu) := \mathcal{K}_{S \times A}(\cdot | s, \bar{\pi}^*(\mu), \mu) \in \mathcal{P}(A),$$

where $\bar{\pi}^*$ is the local maximizer given in Proposition 2.15 (ii).

Moreover, denote by

$$(6.16) \quad \rho_A : \mathcal{P}(A) \times [0, 1] \ni (\eta, u) \mapsto \rho_A(\eta, u) \in A$$

the Blackwell–Dubins function of the action space A (see Lemma A.2).

Step 1. Let $\tilde{\mathbb{P}} \in \mathcal{Q}$ be some arbitrary. We will inductively construct $a^* \in \mathcal{A}$ over time $t \geq 0$. Let

$$(6.17) \quad \begin{aligned} \tilde{s}_0 &:= \xi, & \tilde{\mu}_0 &:= \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{s}_0), \\ a_0^* &:= \rho_A(\psi^*(\cdot | \tilde{s}_0, \tilde{\mu}_0), h_0(\vartheta_0)), & \tilde{\Lambda}_0 &:= \mathcal{L}_{\tilde{\mathbb{P}}}((\tilde{s}_0, a_0^*)), \end{aligned}$$

where $h_0 : \Theta \rightarrow [0, 1]$ is given in Remark 2.19 (so that $h_0(\vartheta_0) \sim \mathcal{U}_{[0,1]}$). In particular, since $\xi \in L_{\mathcal{F}_0}^0(S)$, $\tilde{s}_0$ is $\mathcal{F}_0$ measurable, and a_0^* is $\mathcal{G}_0$ measurable.

For every $t \geq 1$ we inductively define

$$(6.18) \quad \begin{aligned} \tilde{s}_t &:= F(\tilde{s}_{t-1}, a_{t-1}^*, \tilde{\Lambda}_{t-1}, \varepsilon_t, \varepsilon_t^0), & \tilde{\mu}_t &:= \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{s}_t | \varepsilon_{1:t}^0), \\ a_t^* &:= \rho_A(\psi^*(\cdot | \tilde{s}_t, \tilde{\mu}_t), h_t(\vartheta_t)), & \tilde{\Lambda}_t &:= \mathcal{L}_{\tilde{\mathbb{P}}}((\tilde{s}_t, a_t^*) | \varepsilon_{1:t}^0), \end{aligned}$$

where $h_t : \Theta \rightarrow [0, 1]$ is given in Remark 2.19 (ii) (so that $(h_u(\vartheta_u))_{0 \leq u \leq t}$ is i.i.d. with law $\mathcal{U}_{[0,1]}$). Moreover, by the same arguments presented for the proof of Lemma 4.1, $\tilde{s}_t$ is $\mathcal{F}_t$ measurable, while a_t^* is $\mathcal{G}_t$ measurable. Moreover, $(\tilde{\mu}_t, \tilde{\Lambda}_t)$ are $\mathcal{F}_t^0$ measurable.

Since $a^* = (a_t^*)_{t \geq 0}$ constructed via (6.17) and (6.18) is $\mathbb{G}$ adapted, it is in $\mathcal{A}$. We underline that the above inductive construction is invariant and can be carried out under any $\tilde{\mathbb{P}} \in \mathcal{Q}$.

Step 2. We claim that for every $\mathbb{P} \in \mathcal{Q}$,

$$(6.19) \quad s_t^{\xi, a^*, \mathbb{P}} = \tilde{s}_t, \quad \mu_t^{\xi, a^*, \mathbb{P}} = \tilde{\mu}_t, \quad \Lambda_t^{\xi, a^*, \mathbb{P}} = \tilde{\Lambda}_t, \quad \mathbb{P}\text{-a.s.}, \quad \text{for all } t \geq 0,$$

where $s_t^{\xi, a^*, \mathbb{P}}$, $\mu_t^{\xi, a^*, \mathbb{P}}$, and $\Lambda_t^{\xi, a^*, \mathbb{P}}$ are given in (2.12), (2.15) and (2.16), respectively, under $(a^*, \mathbb{P})$.

Let $\mathbb{P} \in \mathcal{Q}$ be given. The proof uses an induction over $t \geq 0$: For $t = 0$, clearly $s_0^{\xi, a^*, \mathbb{P}} = \tilde{s}_0 = \xi \in L_{\mathcal{F}_0}^0(S)$. Moreover, since a_0^* is $\mathcal{G}_0$ measurable (noting that $\mathcal{G}_0 = \sigma(\gamma, \vartheta_0)$) and $\mathcal{L}_{\mathbb{P}}(\gamma, \vartheta_0) = \mathcal{L}_{\tilde{\mathbb{P}}}(\gamma, \vartheta_0)$ (see Remark 2.3 (ii)), it holds that $\mu_0^{\xi, a^*, \mathbb{P}} = \tilde{\mu}_0$ and $\Lambda_0^{\xi, a^*, \mathbb{P}} = \tilde{\Lambda}_0$.

Assume that the induction claim holds true for some $t \geq 0$. For the case $t+1$, by the conditional McKean-Vlasov dynamics given in (2.12) and the induction hypothesis for t , it holds that $\mathbb{P}$ -a.s.,

$$(6.20) \quad \begin{aligned} s_{t+1}^{\xi, a^*, \mathbb{P}} &= F(s_t^{\xi, a^*, \mathbb{P}}, a_t^*, \Lambda_t^{\xi, a^*, \mathbb{P}}, \varepsilon_{t+1}, \varepsilon_{t+1}^0) \\ &= F(\tilde{s}_t, a_t^*, \tilde{\Lambda}_t, \varepsilon_{t+1}, \varepsilon_{t+1}^0) = \tilde{s}_{t+1}, \end{aligned}$$

where the second equality holds by the Borel measurability of F (see Definition 2.4 (i)), and the last equality holds by definition (6.18).

We now show that $\Lambda_{t+1}^{\xi, a^*, \mathbb{P}} = \tilde{\Lambda}_{t+1}$, $\mathbb{P}$ -a.s.. By $\mathcal{F}_{t+1}^0$ -measurability of $(\Lambda_{t+1}^{\xi, a^*, \mathbb{P}}, \tilde{\Lambda}_{t+1})$, it suffices to show that for any bounded Borel measurable functions $\hat{g}_{t+1} : (E^0)^{t+1} \rightarrow \mathbb{R}$ and $\hat{f} : S \times A \rightarrow \mathbb{R}$,

$$(6.21) \quad \mathbb{E}^{\mathbb{P}}[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi, a^*, \mathbb{P}}, a_{t+1}^*)] = \mathbb{E}^{\mathbb{P}}\left[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \int_{S \times A} f(\tilde{s}, \tilde{a}) \tilde{\Lambda}_{t+1}(d\tilde{s}, d\tilde{a})\right].$$

Indeed, by (6.20),

$$\mathbb{E}^{\mathbb{P}}[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi, a^*, \mathbb{P}}, a_{t+1}^*)] = \mathbb{E}^{\mathbb{P}}[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(\tilde{s}_{t+1}, a_{t+1}^*)] =: \mathbb{I}^{t+1}.$$

Moreover, as $(\tilde{s}_{t+1}, a_{t+1}^*)$ is $\mathcal{G}_{t+1}$ measurable, an application of Lemma 6.1 ensures that $\mathbb{P}$ -a.s.

$$\mathcal{L}_{\mathbb{P}}((\tilde{s}_{t+1}, a_{t+1}^*) | \varepsilon_{1:t+1}^0) = \mathcal{L}_{\tilde{\mathbb{P}}}((\tilde{s}_{t+1}, a_{t+1}^*) | \varepsilon_{1:t+1}^0) = \tilde{\Lambda}_{t+1},$$

which implies that $\mathbb{I}^{t+1}$ equals the second term in (6.21), as claimed.

Using the same arguments presented for (6.21), we have that $\mu_{t+1}^{\xi, a^*, \mathbb{P}} = \tilde{\mu}_{t+1}$ $\mathbb{P}$ -a.s.. Hence, by induction hypothesis, the claim (6.19) holds.

Step 3. Let $\mathbb{P} \in \mathcal{Q}$ be some arbitrary. Then we claim that (2.28) holds. Without loss of generality, we consider the case $t \geq 1$, as the case $t = 0$ can be subsumed into it.

By the $\mathcal{F}_t^0$ -measurability of $(\Lambda_t^{\xi, a^*, \mathbb{P}}, \mu_t^{\xi, a^*, \mathbb{P}})$, it suffices to show that for any bounded Borel measurable functions $\hat{g}_t : (E^0)^t \rightarrow \mathbb{R}$ and $\hat{f} : S \times A \rightarrow \mathbb{R}$,

$$(6.22) \quad \mathbb{E}^{\mathbb{P}}[\hat{g}_t(\varepsilon_{1:t}^0) \hat{f}(s_t^{\xi, a^*, \mathbb{P}}, a_t^*)] = \mathbb{E}^{\mathbb{P}}\left[\hat{g}_t(\varepsilon_{1:t}^0) \int_{S \times A} f(\tilde{s}, \tilde{a}) \bar{\pi}^*(\mu_t^{\xi, a^*, \mathbb{P}})(d\tilde{s}, d\tilde{a})\right],$$

where $\bar{\pi}^*$ is the local maximizer given in Proposition 2.15 (ii).

Since $\hat{g}(\varepsilon_{1:t}^0)$ is $\mathcal{F}_t^0$ measurable and it holds that $s_t^{\xi, a^*, \mathbb{P}} = \tilde{s}_t$, $\mathbb{P}$ -a.s. (see (6.19) in Step 2),

$$\begin{aligned} \mathbb{E}^\mathbb{P}[\hat{g}_t(\varepsilon_{1:t}^0) \hat{f}(s_t^{\xi, a^*, \mathbb{P}}, a_t^*)] &= \mathbb{E}^\mathbb{P}[\hat{g}_t(\varepsilon_{1:t}^0) \hat{f}(\tilde{s}_t, a_t^*)] \\ &= \mathbb{E}^\mathbb{P}[\hat{g}_t(\varepsilon_{1:t}^0) \mathbb{E}^\mathbb{P}[\hat{f}(\tilde{s}_t, a_t^*) | \mathcal{F}_t^0]] =: \mathbf{I}_t, \end{aligned}$$

where the last equality follows from the tower property with fact that $\mathcal{F}_t^0 \subset \mathcal{F}_t$.

Since $\tilde{s}_t$ is $\mathcal{F}_t$ measurable and $h_t(\vartheta_t) \sim \mathcal{U}_{[0,1]}$ is independent of $\mathcal{F}_t$ (noting that $\mathcal{F}_t$ does not contain the current randomization source ϑ_t),

$$\begin{aligned} \mathbf{I}_t &= \mathbb{E}^\mathbb{P} \left[\hat{g}_t(\varepsilon_{1:t}^0) \mathbb{E}^\mathbb{P} \left[\mathbb{E}^\mathbb{P} \left[\int_A \hat{f}(\tilde{s}_t, \tilde{a}) \psi^*(d\tilde{a} | \tilde{s}_t, \tilde{\mu}_t) \middle| \mathcal{F}_t \right] \middle| \mathcal{F}_t^0 \right] \right] \\ (6.23) \quad &= \mathbb{E}^\mathbb{P} \left[\hat{g}_t(\varepsilon_{1:t}^0) \mathbb{E}^\mathbb{P} \left[\int_{S \times A} \hat{f}(\tilde{s}, \tilde{a}) \mathcal{K}_{S \times A}(d\tilde{a} | \tilde{s}, \bar{\pi}^*(\tilde{\mu}_t), \tilde{\mu}_t) \tilde{\mu}_t(d\tilde{s}) \middle| \mathcal{F}_t^0 \right] \right] \\ &= \mathbb{E}^\mathbb{P} \left[\hat{g}_t(\varepsilon_{1:t}^0) \int_{S \times A} \hat{f}(\tilde{s}, \tilde{a}) \bar{\pi}^*(\tilde{\mu}_t)(d\tilde{s}, d\tilde{a}) \right], \end{aligned}$$

where the first equality follows from definition of a_t^* given in (6.18), the second equality follows from definition of $\psi^*(\cdot | \tilde{s}_t, \tilde{\mu}_t)$ (see (6.15)) and $\mathcal{F}_t^0$ -measurability of $\tilde{\mu}_t$, and the last equality follows from definition of the universal differentiation kernel $\mathcal{K}_{S \times A}$ (see (6.14)).

Moreover, since $\tilde{\mu}_t = \mu_t^{\xi, a^*, \mathbb{P}}$, $\mathbb{P}$ -a.s. (see (6.19) in Step 2), the last term in (6.23) equals the second term in (6.22), as claimed. This completes the proof. $\square$

6.3. Proof of Theorem 2.21. For notational simplicity, set $\mu := \mathcal{L}(\xi)$.

Step 1: We claim that for every $n \in \mathbb{N}$

$$(6.24) \quad \mathcal{I}_n^{\xi, a^*} := \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^\mathbb{P} \left[\sum_{t=0}^{n-1} \beta^t r(s_t^{\xi, a^*, \mathbb{P}}, a_t^*, \Lambda_t^{\xi, a^*, \mathbb{P}}) + \beta^n \bar{V}^*(\mu_n^{\xi, a^*, \mathbb{P}}) \right] \geq \bar{V}^*(\mu),$$

where for every $\mathbb{P} \in \mathcal{Q}$, let $(\mu_t^{\xi, a^*, \mathbb{P}})_{t \geq 0}$ and $(\Lambda_t^{\xi, a^*, \mathbb{P}})_{t \geq 0}$ be given by (2.15) and (2.16), respectively.

We prove (6.24) via an induction over n . Before proceeding, note that for every $\mathbb{P} \in \mathcal{Q}$ and $t \geq 0$,

$$\begin{aligned} \mathbb{E}^\mathbb{P} [r(s_t^{\xi, a^*, \mathbb{P}}, a_t^*, \Lambda_t^{\xi, a^*, \mathbb{P}})] &= \mathbb{E}^\mathbb{P} [\bar{r}(\text{pj}_S(\Lambda_t^{\xi, a^*, \mathbb{P}}), \Lambda_t^{\xi, a^*, \mathbb{P}})] \\ (6.25) \quad &= \mathbb{E}^\mathbb{P} [\bar{r}(\mu_t^{\xi, a^*, \mathbb{P}}, \bar{\pi}^*(\mu_t^{\xi, a^*, \mathbb{P}}))], \end{aligned}$$

where the first equality holds by (2.20) in Remark 2.13 and the second equality follows from (2.28) in Lemma 2.20 and the fact that $\bar{\pi}^*(\mu) \in \mathfrak{U}(\mu)$ (see Proposition 2.15 (ii)).

Hence by the property (6.25), $\mathcal{I}_n^{\xi, a^*}$ given in (6.24) can be represented by

$$(6.26) \quad \mathcal{I}_n^{\xi, a^*} = \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^\mathbb{P} \left[\sum_{t=0}^{n-1} \beta^t \bar{r}(\mu_t^{\xi, a^*, \mathbb{P}}, \bar{\pi}^*(\mu_t^{\xi, a^*, \mathbb{P}})) + \beta^n \bar{V}^*(\mu_n^{\xi, a^*, \mathbb{P}}) \right].$$

Step 1a: For $n = 1$, let $\mathbb{P} \in \mathcal{Q}$ be induced by some $(p_t)_{t \geq 1} \in \mathcal{K}^0$ (see Definition 2.2).

We first note that $\mu_0^{\xi, a^*, \mathbb{P}} = \mu$ with trivial $\mathcal{F}_0^0$ and $\mathcal{L}_\mathbb{P}(\varepsilon_1^0) = p_1 \in \mathfrak{P}^0$ (see Remark 2.3 (iii)). Combined with (2.19) (see Proposition 2.12), this implies that

$$\begin{aligned} \mathbb{E}^\mathbb{P} [\bar{r}(\mu_0^{\xi, a^*, \mathbb{P}}, \bar{\pi}^*(\mu_0^{\xi, a^*, \mathbb{P}})) + \beta \bar{V}^*(\mu_1^{\xi, a^*, \mathbb{P}})] &= \bar{r}(\mu, \bar{\pi}^*(\mu)) + \beta \int_{\mathcal{P}(S)} \bar{V}^*(\mu') \bar{p}(d\mu' | \mu, \bar{\pi}^*(\mu), p_1) \\ (6.27) \quad &\geq \bar{r}(\mu, \bar{\pi}^*(\mu)) + \beta \inf_{p \in \mathfrak{P}^0} \int_{\mathcal{P}(S)} \bar{V}^*(\mu') \bar{p}(d\mu' | \mu, \bar{\pi}^*(\mu), p) \\ &= \mathcal{T} \bar{V}^*(\mu) = \bar{V}^*(\mu), \end{aligned}$$

where the last line follows from the optimality of $\bar{\pi}^*(\mu) \in \mathfrak{U}(\mu)$ for $\mathcal{T}\bar{V}^*(\mu)$ (see Proposition 2.15 (ii) for $\bar{V}^* \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$) and the fixed point result given in Proposition 2.16.

Since (6.27) holds for any $\mathbb{P} \in \mathcal{Q}$, by (6.26) we have that $\mathcal{I}_1^{\xi,a^*} \geq \bar{V}^*(\mu)$.

Step 1b: Assume that (6.24) holds for some $n \geq 1$. Let $\mathbb{P} \in \mathcal{Q}$ be induced by some $(p_t)_{t \geq 1} \in \mathcal{K}^0$.

Note that $\mu_n^{\xi,a^*,\mathbb{P}}$ and $\mathcal{L}_{\mathbb{P}}(\varepsilon_{n+1}^0 | \mathcal{F}_n^0)$ are $\mathcal{F}_n^0$ measurable and $\mathcal{L}_{\mathbb{P}}(\varepsilon_{n+1}^0 | \mathcal{F}_n^0) = p_{n+1}(\cdot | \varepsilon_{1:n}^0) \in \mathfrak{P}^0$ $\mathbb{P}$ -a.s. (see Remark 2.3 (ii)).

From this, we can use the same arguments presented for (6.27) to have that $\mathbb{P}$ -a.s.

$$\begin{aligned} & \mathbb{E}^{\mathbb{P}} [\bar{r}(\mu_t^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_n^{\xi,a^*,\mathbb{P}})) + \beta \bar{V}^*(\mu_{n+1}^{\xi,a^*,\mathbb{P}}) | \mathcal{F}_n^0] \\ &= \bar{r}(\mu_n^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_n^{\xi,a^*,\mathbb{P}})) + \beta \int_{\mathcal{P}(S)} \bar{V}^*(\mu') \bar{p}(d\mu' | \mu_n^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_n^{\xi,a^*,\mathbb{P}}), p_{n+1}(\cdot | \varepsilon_{1:n}^0)) \\ &\geq \bar{r}(\mu_n^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_n^{\xi,a^*,\mathbb{P}})) + \beta \inf_{p \in \mathfrak{P}^0} \int_{\mathcal{P}(S)} \bar{V}^*(\mu') \bar{p}(d\mu' | \mu_n^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_n^{\xi,a^*,\mathbb{P}}), p) \\ &= \mathcal{T}\bar{V}^*(\mu_n^{\xi,a^*,\mathbb{P}}) = \bar{V}^*(\mu_n^{\xi,a^*,\mathbb{P}}), \end{aligned}$$

which ensures that

$$\begin{aligned} (6.28) \quad & \mathbb{E}^{\mathbb{P}} \left[\sum_{t=0}^n \beta^t \bar{r}(\mu_t^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_t^{\xi,a^*,\mathbb{P}})) + \beta^{n+1} \bar{V}^*(\mu_{n+1}^{\xi,a^*,\mathbb{P}}) \right] \\ & \geq \mathbb{E}^{\mathbb{P}} \left[\sum_{t=0}^{n-1} \beta^t \bar{r}(\mu_t^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_t^{\xi,a^*,\mathbb{P}})) + \beta^n \bar{V}^*(\mu_n^{\xi,a^*,\mathbb{P}}) \right] \geq \mathcal{I}_n^{\xi,a^*} \geq \bar{V}^*(\mu), \end{aligned}$$

where the second inequality follows from definition of $\mathcal{I}_n^{\xi,a^*}$ given in (6.26) and the last inequality follows from assumption of the induction for n (see (6.24)).

As (6.28) holds for any $\mathbb{P} \in \mathcal{Q}$, we have $\mathcal{I}_{n+1}^{\xi,a^*} \geq \bar{V}^*(\mu)$. Therefore, by the induction hypothesis, (6.24) holds for every $n \in \mathbb{N}$. We conclude that the claim for Step 1 holds.

Step 2: We claim that $\bar{V}^*(\mu) \leq V(\xi)$. Since $\bar{r}$ and $\bar{V}^*$ is bounded and $\beta < 1$ (see Lemma 5.2 (iii) and $\bar{V}^* \in \text{Lip}_{b,\bar{L}}(\mathcal{P}(S); \mathbb{R})$), the dominated convergence theorem asserts that for every $\mu \in \mathcal{P}(S)$

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathcal{I}_n^{\xi,a^*} &\leq \inf_{\mathbb{P} \in \mathcal{Q}} \left\{ \limsup_{n \rightarrow \infty} \mathbb{E}^{\mathbb{P}} \left[\sum_{t=0}^{n-1} \beta^t \bar{r}(\mu_t^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_t^{\xi,a^*,\mathbb{P}})) \right] + \limsup_{n \rightarrow \infty} \mathbb{E}^{\mathbb{P}} [\beta^n |\bar{V}^*(\mu_n^{\xi,a^*,\mathbb{P}})|] \right\} \\ &= \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}} \left[\sum_{t=0}^{\infty} \beta^t \bar{r}(\mu_t^{\xi,a^*,\mathbb{P}}, \bar{\pi}^*(\mu_t^{\xi,a^*,\mathbb{P}})) \right] = \mathcal{J}^{a^*}(\xi) \leq V(\xi), \end{aligned}$$

where the second equality follows from (6.25) and the definition of $\mathcal{J}^{a^*}(\xi)$ (see (2.13)).

Combining this with (6.24) (as shown in Step 1), we conclude that

$$(6.29) \quad \bar{V}^*(\mu) \leq \limsup_{n \rightarrow \infty} \mathcal{I}_n^{\xi,a^*} \leq \mathcal{J}^{a^*}(\xi) \leq V(\xi),$$

as claimed.

Step 3: We claim that $V(\xi) \leq \bar{V}^*(\mu)$, which ensures the statement (i) to hold. For every $a \in \mathcal{A}$, let $\mathbb{P}^{\xi,a} \in \mathcal{Q}$ be induced by $(p_t^{\xi,a})_{t \geq 1} \in \mathcal{K}^0$ such that (2.26) and (2.27) given in Lemma 2.17 hold.

Then, define $\mathcal{V}^a(\xi)$ by

$$\mathcal{V}^a(\xi) := \mathbb{E}^{\mathbb{P}^{\xi,a}} \left[\sum_{t=0}^{\infty} \beta^t r(s_t^{\xi,a,\mathbb{P}^{\xi,a}}, a_t, \underline{\Lambda}_t^{\xi,a}) \right] = \mathbb{E}^{\mathbb{P}^{\xi,a}} \left[\sum_{t=0}^{\infty} \beta^t \bar{r}(\text{pjs}(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}) \right],$$

where $\underline{\Lambda}_0^{\xi,a}$ is the joint law of $(s_0^{\xi,a,\mathbb{P}^{\xi,a}}, a_0)$ under $\mathbb{P}^{\xi,a}$, for $t \geq 1$ $\underline{\Lambda}_t^{\xi,a}$ is the conditional joint law of $(s_t^{\xi,a,\mathbb{P}^{\xi,a}}, a_t)$ under $\mathbb{P}^{\xi,a}$ given $\varepsilon_{1:t}^0$, and the last equality follows from the same arguments presented for (6.25).

Then by definition of $\mathcal{J}^a(\xi)$ given in (2.13)

$$(6.30) \quad V(\xi) = \sup_{a \in \mathcal{A}} \mathcal{J}^a(\xi) \leq \sup_{a \in \mathcal{A}} \mathcal{V}^a(\xi).$$

Moreover, since $\bar{r}$ and $\bar{V}^*$ is bounded and $\beta < 1$, by the dominated convergence theorem to the sums $\sum_{t=0}^n \beta^t \bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a})$ $n \in \mathbb{N}$, we can have that for every $a \in \mathcal{A}$

$$(6.31) \quad \mathcal{V}^a(\xi) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}^{\mathbb{P}^{\xi,a}} [\bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}) + \beta \bar{V}^*(\underline{\mu}_{t+1}^{\xi,a}) - \beta \bar{V}^*(\underline{\mu}_t^{\xi,a})].$$

Then it follows from (2.27) in Lemma 2.17 that for every $t \geq 0$

$$\begin{aligned} \mathbb{E}^{\mathbb{P}^{\xi,a}} [\bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}) + \beta \bar{V}^*(\underline{\mu}_{t+1}^{\xi,a})] &= \mathbb{E}^{\mathbb{P}^{\xi,a}} \left[\mathbb{E}^{\mathbb{P}^{\xi,a}} [\bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}) + \beta \bar{V}^*(\underline{\mu}_{t+1}^{\xi,a}) \mid \mathcal{F}_t^0] \right], \\ &=: \mathbb{E}^{\mathbb{P}^{\xi,a}} [\bar{J}(\underline{\Lambda}_t^{\xi,a})] \end{aligned}$$

where $\bar{J}(\underline{\Lambda}_t^{\xi,a})$ is $\mathcal{F}_t^0$ measurable and satisfies

$$\begin{aligned} \bar{J}(\underline{\Lambda}_t^{\xi,a}) &= \bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}) + \beta \int_{\mathcal{P}(S)} \bar{V}^*(\tilde{\mu}) \bar{p}(d\tilde{\mu} \mid \text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}, \bar{p}^*(\underline{\Lambda}_t^{\xi,a})) \\ (6.32) \quad &= \bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}) + \beta \inf_{p \in \mathfrak{P}^0} \int_{\mathcal{P}(S)} \bar{V}^*(\tilde{\mu}) \bar{p}(d\tilde{\mu} \mid \text{pj}_S(\underline{\Lambda}_t^{\xi,a}), \underline{\Lambda}_t^{\xi,a}, p) \\ &\leq \mathcal{T}\bar{V}^*(\text{pj}_S(\underline{\Lambda}_t^{\xi,a})), \end{aligned}$$

where the equality holds by the local optimality $\bar{p}^*(\underline{\Lambda}_t^{\xi,a}) \in \mathfrak{P}^0$ (see Proposition 2.15 (i)) and the inequality holds by definition of $\mathcal{T}\bar{V}^*(\text{pj}_S(\underline{\Lambda}_t^{\xi,a}))$ (see (2.22))

Combining (6.30)–(6.32) with the marginal constraint (i.e., $\text{pj}_S(\underline{\Lambda}_t^{\xi,a}) = \underline{\mu}_t^{\xi,a}$ $\mathbb{P}^{\xi,a}$ -a.s.; see (2.17)), and the fixed point result (i.e., $\mathcal{T}\bar{V}^* = \bar{V}^*$; see Proposition 2.16), we conclude that

$$V(\xi) \leq \sup_{a \in \mathcal{A}} \sum_{t=0}^{\infty} (\beta^t \mathbb{E}^{\mathbb{P}^{\xi,a}} [\bar{V}^*(\underline{\mu}_t^{\xi,a})] - \beta^{t+1} \mathbb{E}^{\mathbb{P}^{\xi,a}} [\bar{V}^*(\underline{\mu}_{t+1}^{\xi,a})]) = \bar{V}^*(\mu),$$

where the last equality holds by the dominated convergence theorem and the fact that $\underline{\mu}_0^{\xi,a} = \mu$, as claimed.

Step 4: It remains to show that (2.29) holds. Recall that $a^* \in \mathcal{A}$ is such that (2.28) holds for every $\mathbb{P} \in \mathcal{Q}$ (see Lemma 2.20). Moreover, let $\mathbb{P}^{\xi,a^*} \in \mathcal{Q}$ is induced by $(p_t^{\xi,a^*})_{t \geq 1} \in \mathcal{K}^0$ satisfying (2.26) and (2.27) (see Lemma 2.17).

By applying the dominated convergence theorem to $\sum_{t=0}^n (\beta^t \bar{V}^*(\underline{\mu}_t^{\xi,a^*}) - \beta^{t+1} \bar{V}^*(\underline{\mu}_{t+1}^{\xi,a^*}))$ $n \in \mathbb{N}$,

$$(6.33) \quad \bar{V}^*(\mu) = \sum_{t=0}^{\infty} (\beta^t \mathbb{E}^{\mathbb{P}^{\xi,a^*}} [\bar{V}^*(\underline{\mu}_t^{\xi,a^*})] - \beta^{t+1} \mathbb{E}^{\mathbb{P}^{\xi,a^*}} [\bar{V}^*(\underline{\mu}_{t+1}^{\xi,a^*})]),$$

where $\underline{\mu}_t^{\xi,a^*}$ is the conditional law of $s_t^{\xi,a^*,\mathbb{P}^{\xi,a^*}}$ given $\varepsilon_{1:t}^0$.

Note that for every $\mu' \in \mathcal{P}(S)$

$$(6.34) \quad \bar{V}^*(\mu') = \mathcal{T}\bar{V}^*(\mu') = \bar{r}(\mu', \bar{\pi}^*(\mu')) + \beta \int_{\mathcal{P}(S)} \bar{V}^*(\tilde{\mu}') \bar{p}(d\tilde{\mu}' \mid \mu', \bar{\pi}^*(\mu'), \bar{p}^*(\bar{\pi}^*(\mu'))).$$

where the first equality follows from Proposition 2.16 and the second equality follows from the optimality of the local optimizers $\bar{\pi}^*$ and $\bar{p}^*$ given in Proposition 2.15.

From (6.34), it holds that for every $t \geq 0$

$$\begin{aligned} \mathbb{E}^{\mathbb{P}^{\xi, a^*}} [\bar{V}^*(\underline{\mu}_t^{\xi, a^*})] &= \mathbb{E}^{\mathbb{P}^{\xi, a^*}} [\mathcal{T}\bar{V}^*(\underline{\mu}_t^{\xi, a^*})] \\ &= \mathbb{E}^{\mathbb{P}^{\xi, a^*}} \left[\bar{r}(\underline{\mu}_t^{\xi, a^*}, \bar{\pi}^*(\underline{\mu}_t^{\xi, a^*})) + \beta \int_{\mathcal{P}(S)} \bar{V}^*(\tilde{\mu}') \bar{p}(d\tilde{\mu}' | \underline{\mu}_t^{\xi, a^*}, \bar{\pi}^*(\underline{\mu}_t^{\xi, a^*}), \bar{p}^*(\bar{\pi}^*(\underline{\mu}_t^{\xi, a^*}))) \right] \\ &= \mathbb{E}^{\mathbb{P}^{\xi, a^*}} \left[\bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi, a^*}), \underline{\Lambda}_t^{\xi, a^*}) + \beta \int_{\mathcal{P}(S)} \bar{V}^*(\tilde{\mu}') \bar{p}(d\tilde{\mu}' | \text{pj}_S(\underline{\Lambda}_t^{\xi, a^*}), \underline{\Lambda}_t^{\xi, a^*}, \bar{p}^*(\underline{\Lambda}_t^{\xi, a^*})) \right] =: \mathbf{I}_t, \end{aligned}$$

where $\underline{\Lambda}_0^{\xi, a^*}$ is the joint law of $(s_0^{\xi, a^*, \mathbb{P}^{\xi, a^*}}, a_0^*)$ under $\mathbb{P}^{\xi, a^*}$, for $t \geq 1$ $\underline{\Lambda}_t^{\xi, a^*}$ is the conditional joint law of $(s_t^{\xi, a^*, \mathbb{P}^{\xi, a^*}}, a_t^*)$ under $\mathbb{P}^{\xi, a^*}$ given $\varepsilon_{1:t}^0$, and the last equality follows from the fact that $\underline{\Lambda}_t^{\xi, a^*} = \bar{\pi}^*(\underline{\mu}_t^{\xi, a^*})$ $\mathbb{P}^{\xi, a^*}$ -a.s.; see Lemma 2.20, and the marginal constraint that $\text{pj}_S(\underline{\Lambda}_t^{\xi, a^*}) = \underline{\mu}_t^{\xi, a^*}$ $\mathbb{P}^{\xi, a^*}$ -a.s.; see (2.17).

Furthermore, by (2.27) in Lemma 2.17 for $(a^*, \mathbb{P}^{\xi, a^*})$, it holds that for every $t \geq 0$

$$\mathbf{I}_t = \mathbb{E}^{\mathbb{P}^{\xi, a^*}} [\bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi, a^*}), \underline{\Lambda}_t^{\xi, a^*}) + \beta \bar{V}^*(\underline{\mu}_{t+1}^{\xi, a^*})].$$

Combined with (6.33), this ensures that

$$\bar{V}^*(\mu) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}^{\mathbb{P}^{\xi, a^*}} [\bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi, a^*}), \underline{\Lambda}_t^{\xi, a^*})] = \mathbb{E}^{\mathbb{P}^{\xi, a^*}} \left[\sum_{t=0}^{\infty} \beta^t \bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi, a^*}), \underline{\Lambda}_t^{\xi, a^*}) \right].$$

Therefore, by the equality $\bar{V}^*(\mu) = V(\xi)$ (from Step 2 and Step 3), we conclude that

$$\begin{aligned} \bar{V}^*(\mu) = V(\xi) &= \sup_{a \in \mathcal{A}} \mathcal{J}^a(\xi) = \mathbb{E}^{\mathbb{P}^{\xi, a^*}} \left[\sum_{t=0}^{\infty} \beta^t \bar{r}(\text{pj}_S(\underline{\Lambda}_t^{\xi, a^*}), \underline{\Lambda}_t^{\xi, a^*}) \right] \\ &= \mathbb{E}^{\mathbb{P}^{\xi, a^*}} \left[\sum_{t=0}^{\infty} \beta^t r(s_t^{\xi, a^*, \mathbb{P}^{\xi, a^*}}, a_t^*, \underline{\Lambda}_t^{\xi, a^*}) \right] = \mathcal{J}^{a^*}(\xi), \end{aligned}$$

where the last line follows from the same arguments presented for (6.25), and the inequality (6.29) given in Step 2. This completes the proof. $\square$

7. PROOF OF RESULTS IN SECTION 2.5

7.1. Proof of Lemma 2.25. We first prove (2.33). For simplicity, denote for every $t \geq 0$ by

$$\mu_t := \mu_t^{\xi, \pi^c, \mathbb{P}}, \quad \Lambda_t := \Lambda_t^{\xi, \pi^c, \mathbb{P}}, \quad \nu_{t+1} := \mathcal{L}_{\mathbb{P}}(\varepsilon_{t+1}^0 | \mathcal{F}_t^0).$$

As the case for $t = 0$ can be subsumed into the others for $t \geq 1$, we consider the case $t \geq 1$. Since Λ_t and μ_t are $\mathcal{F}_t^0$ measurable, it is sufficient to show that for any bounded Borel measurable functions $g : (E^0)^t \rightarrow \mathbb{R}$ and $f : S \times A \rightarrow \mathbb{R}$,

$$\mathbb{E}^{\mathbb{P}} [g(\varepsilon_{1:t}^0) f(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})] = \mathbb{E}^{\mathbb{P}} \left[g(\varepsilon_{1:t}^0) \int_{S \times A} f(s', a') \pi_t^c(d\tilde{a} | \tilde{s}, \mu_t) \mu_t(d\tilde{s}) \right].$$

Note that $g(\varepsilon_{1:t}^0)$ is $\mathcal{F}_t^0$ measurable and $s_t^{\xi, \pi^c, \mathbb{P}}$ is $\mathcal{F}_t$ measurable. Hence, by the distributional constraint that $\mathcal{L}_{\mathbb{P}}(a_t^{\pi^c, \mathbb{P}} | \mathcal{F}_t) = \pi_t^c(\cdot | s_t^{\xi, \pi^c, \mathbb{P}}, \mu_t)$ $\mathbb{P}$ -a.s. (see (2.30)) and the tower property,

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} [g(\varepsilon_{1:t}^0) f(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}})] &= \mathbb{E}^{\mathbb{P}} \left[g(\varepsilon_{1:t}^0) \mathbb{E}^{\mathbb{P}} [f(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}}) | \mathcal{F}_t^c] | \mathcal{F}_t^0 \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[g(\varepsilon_{1:t}^0) \int_A f(s_t^{\xi, \pi^c, \mathbb{P}}, \tilde{a}) \pi_t^c(da' | s_t^{\xi, \pi^c, \mathbb{P}}, \mu_t) \right] =: \mathbf{I}_t. \end{aligned}$$

Moreover by the definition of μ_t and its $\mathcal{F}_t^0$ -measurability,

$$\begin{aligned} \mathbf{I}_t &= \mathbb{E}^{\mathbb{P}} \left[g(\varepsilon_{1:t}^0) \mathbb{E}^{\mathbb{P}} \left[\int_A f(s_t^{\xi, \pi^c, \mathbb{P}}, \tilde{a}) \pi_t^c(da' | s_t^{\xi, \pi^c, \mathbb{P}}, \mu_t) \mid \mathcal{F}_t^0 \right] \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[g(\varepsilon_{1:t}^0) \int_{S \times A} f(s', a') \pi_t^c(d\tilde{a} | \tilde{s}, \mu_t) \mu_t(d\tilde{s}) \right], \end{aligned}$$

as claimed.

Moreover, since $\text{pj}_S(\mu_t^{\xi, \pi^c, \mathbb{P}} \hat{\otimes} \pi_t^c(\cdot | \cdot, \mu_t^{\xi, \pi^c, \mathbb{P}})) = \mu_t^{\xi, \pi^c, \mathbb{P}}$, we can use the same arguments as in the proof of Proposition 2.12 (we refer to Section 5) to get that (2.34) holds $\mathbb{P}$ -a.s. $\square$

7.2. Proof of Lemma 2.26. We first prove (2.35). *Step 1:* Let $\pi^c \in \Pi^c$ be given, and let $\tilde{\mathbb{P}} \in \mathcal{Q}$ be some arbitrary. Then set

$$(7.1) \quad \begin{aligned} \tilde{s}_0 &:= \xi, & \tilde{\mu}_0 &:= \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{s}_0), \\ \tilde{a}_0 &:= \rho_A(\pi_0^c(\cdot | \tilde{s}_0, \tilde{\mu}_0), h_0(\vartheta_0)), \end{aligned}$$

where ρ_A is the Blackwell-Dubins function on A (see Lemma A.2) and h_0 is given in Remark 2.19. Here we note that $\tilde{s}_0$ is $\mathcal{F}_0$ measurable (as $\xi \in L_{\mathcal{F}_0}^0(S)$) and $\tilde{a}_0$ is $\mathcal{G}_0$ measurable.

Then we define by

$$(7.2) \quad \underline{p}_1^{\xi, \pi^c} := \bar{p}^*(\tilde{\mu}_0 \hat{\otimes} \pi_0^c(\cdot | \cdot, \tilde{\mu}_0)) \in \mathfrak{P}^0,$$

where $\bar{p}^*$ is given in Proposition 2.15 (i).

Next, for every $t \geq 1$ we inductively set

$$(7.3) \quad \begin{aligned} \tilde{s}_t &:= F(\tilde{s}_{t-1}, \tilde{a}_{t-1}, \tilde{\mu}_{t-1} \hat{\otimes} \pi_{t-1}^c(\cdot | \cdot, \tilde{\mu}_{t-1}), \varepsilon_t, \varepsilon_t^0), & \tilde{\mu}_t &:= \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{s}_t | \varepsilon_{1:t}^0), \\ \tilde{a}_t &:= \rho_A(\pi_t^c(\cdot | \tilde{s}_t, \tilde{\mu}_t), h_t(\vartheta_t)), \end{aligned}$$

Here, by using the same arguments presented for the proof of Lemma 4.1 (ii), we can deduce that $\tilde{s}_t$ is $\mathcal{F}_t$ measurable and $\tilde{a}_t$ is $\mathcal{G}_t$ measurable. Moreover, $(\tilde{\mu}_t, \tilde{\Lambda}_t)$ are $\mathcal{F}_t^0$ measurable.

From this, we can consider a Borel measurable function $l_t : (E^0)^t \rightarrow \mathcal{P}(S \times A)$ such that

$$(7.4) \quad l_t(\varepsilon_{1:t}^0) = \tilde{\mu}_t \hat{\otimes} \pi_t^c(\cdot | \cdot, \tilde{\mu}_t).$$

Then, define $\underline{p}_{t+1}^{\xi, \pi^c} : (E^0)^t \ni e_{1:t}^0 \mapsto \underline{p}_{t+1}^{\xi, \pi^c}(\cdot | e_{1:t}^0) \in \mathcal{P}(E^0)$ by

$$(7.5) \quad \underline{p}_{t+1}^{\xi, \pi^c}(\cdot | e_{1:t}^0) := \bar{p}^*(l_t(e_{1:t}^0)) \in \mathfrak{P}^0.$$

Therefore we can define by $\mathbb{P}^{\xi, \pi^c} \in \mathcal{Q}$ the measure induced by $(\underline{p}_t^{\xi, \pi^c})_{t \geq 1} \in \mathcal{K}^0$ given in (7.2) and (7.5).

Step 2: Recall $(\tilde{\mu}_t)_{t \geq 0}$ given in (7.1) and (7.3). We claim that $\mathbb{P}^{\xi, \pi^c}$ -a.s.

$$(7.6) \quad \underline{\mu}_t^{\xi, \pi^c} = \tilde{\mu}_t, \quad \text{for all } t \geq 0,$$

where $\underline{\mu}_0^{\xi, a}$ is the law of $s_0^{\xi, a, \mathbb{P}^{\xi, a}}$ under $\mathbb{P}^{\xi, a}$, and for $t \geq 1$ $\underline{\mu}_t^{\xi, a}$ is the conditional law of $s_t^{\xi, a, \mathbb{P}^{\xi, a}}$ under $\mathbb{P}^{\xi, a}$ given $\varepsilon_{1:t}^0$.

The proof uses an induction over $t \geq 0$: For $t = 0$, clearly $s_0^{\xi, \pi^c, \mathbb{P}^{\xi, a}} = \tilde{s}_0 = \xi \in L_{\mathcal{F}_0}^0(S)$. Moreover, since $\mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\gamma) = \mathcal{L}_{\tilde{\mathbb{P}}}(\gamma)$ (see Remark 2.3 (ii)), it holds that $\underline{\mu}_0^{\xi, \pi^c} = \tilde{\mu}_0$.

Assume that the induction claim holds for some $t \geq 0$. By $\mathcal{F}_{t+1}^0$ -measurability of $(\mu_{t+1}^{\xi, \pi^c}, \tilde{\mu}_{t+1})$, it suffices to show that for any bounded Borel measurable functions $\hat{g}_{t+1} : (E^0)^{t+1} \rightarrow \mathbb{R}$ and $\hat{f} : S \rightarrow \mathbb{R}$,

$$(7.7) \quad \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} [\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi, \pi^c, \mathbb{P}^{\xi, \pi^c}})] = \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \int_{S \times A} f(\tilde{s}) \tilde{\mu}_{t+1}(d\tilde{s}) \right].$$

Indeed, by the conditional McKean-Vlasov dynamics given in (2.30) and Fubini's theorem

$$(7.8) \quad \begin{aligned} \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} [\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi, \pi^c, \mathbb{P}^{\xi, \pi^c}})] &= \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(F(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}}, \underline{\Lambda}_t^{\xi, \pi^c}, \varepsilon_{t+1}, \varepsilon_{t+1}^0)) \right] \\ &= \int_E \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(F(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}}, \underline{\Lambda}_t^{\xi, \pi^c}, e, \varepsilon_{t+1}^0)) \right] \lambda_\varepsilon(de) =: \mathbf{I}_t, \end{aligned}$$

where the second equality holds since ε_{t+1} is independent of $\mathcal{G}_t \vee \sigma(\varepsilon_{t+1}^0)$ with $\mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\varepsilon_{t+1}) = \lambda_\varepsilon$ (see Remark 2.3 (i), (ii)).

Moreover, since ε_{t+1}^0 is conditionally independent of $\mathcal{G}_t$ given $\mathcal{F}_t^0$ (see Remark 2.3 (iii)) with $\mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\varepsilon_{t+1}^0 | \mathcal{F}_t^0) = p_{t+1}^{\xi, \pi^c}(de^0 | \varepsilon_{1:t}^0)$ (by definition of $\mathbb{P}^{\xi, \pi^c}$), and $s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}}$, and $\underline{\Lambda}_t^{\xi, \pi^c}$ are all $\mathcal{G}_t$ measurable, we have

$$(7.9) \quad \mathbf{I}_t = \int_E \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\int_{E^0} \left(\hat{g}_{t+1}(\varepsilon_{1:t}^0, e^0) D_{\mathcal{F}_t^0}(e, e^0) \right) p_{t+1}^{\xi, \pi^c}(de^0 | \varepsilon_{1:t}^0) \right] \lambda_\varepsilon(de)$$

where for every $(e, e^0) \in E \times E^0$

$$\begin{aligned} D_{\mathcal{F}_t^0}(e, e^0) &:= \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\hat{f}(F(s_t^{\xi, \pi^c, \mathbb{P}}, a_t^{\pi^c, \mathbb{P}}, \underline{\Lambda}_t^{\xi, \pi^c}, e, e^0)) \mid \mathcal{F}_t^0 \right] \\ &= \int_{S \times A} \hat{f}(F(s, a, \underline{\Lambda}_t^{\xi, \pi^c}, e, e^0)) \underline{\Lambda}_t^{\xi, \pi^c}(ds, da). \end{aligned}$$

Moreover, from (2.33) in Lemma 2.25 it holds for every $(e, e^0) \in E \times E^0$ that $\mathbb{P}^{\xi, \pi^c}$ -a.s.,

$$\begin{aligned} D_{\mathcal{F}_t^0}(e, e^0) &= \int_{S \times A} \hat{f}(F(s, a, (\underline{\mu}_t^{\xi, \pi^c} \hat{\otimes} \pi_t^c(\cdot | \cdot, \underline{\mu}_t^{\xi, \pi^c})), e, e^0)) (\underline{\mu}_t^{\xi, \pi^c} \hat{\otimes} \pi_t^c(\cdot | \cdot, \underline{\mu}_t^{\xi, \pi^c}))(ds, da) \\ &= \int_{S \times A} \hat{f}(F(s, a, (\tilde{\mu}_t \hat{\otimes} \pi_t^c(\cdot | \cdot, \tilde{\mu}_t)), e, e^0)) (\tilde{\mu}_t \hat{\otimes} \pi_t^c(\cdot | \cdot, \tilde{\mu}_t))(ds, da) \end{aligned}$$

where the second inequality follows from the induction assumption at t .

Furthermore, since $\tilde{s}_t$ is $\mathcal{G}_t$ measurable (noting that $\mathcal{F}_t \subset \mathcal{G}_t$), an application of Lemma 6.1 ensures that $\tilde{\mu}_t = \mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\tilde{s}_t | \mathcal{F}_t^0)$ $\mathbb{P}^{\xi, \pi^c}$ -a.s.. This implies that $\mathbb{P}^{\xi, \pi^c}$ -a.s.

$$(7.10) \quad \begin{aligned} D_{\mathcal{F}_t^0}(e, e^0) &= \int_{S \times A} \hat{f}(F(s, a, (\tilde{\mu}_t \hat{\otimes} \pi_t^c(\cdot | \cdot, \tilde{\mu}_t)), e, e^0)) (\mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\tilde{s}_t | \mathcal{F}_t^0) \hat{\otimes} \pi_t^c(\cdot | \cdot, \tilde{\mu}_t))(ds, da) \\ &= \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\int_A \hat{f}(F(\tilde{s}_t, a, (\tilde{\mu}_t \hat{\otimes} \pi_t^c(\cdot | \cdot, \tilde{\mu}_t)), e, e^0)) \pi_t^c(da | \tilde{s}_t, \tilde{\mu}_t) \mid \mathcal{F}_t^0 \right] \\ &= \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\hat{f}(F(\tilde{s}_t, \tilde{a}_t, (\tilde{\mu}_t \hat{\otimes} \pi_t^c(\cdot | \cdot, \tilde{\mu}_t)), e, e^0)) \mid \mathcal{F}_t^0 \right], \end{aligned}$$

where the last equality holds by definition of $\tilde{a}_t$ given in (7.3) (which follows from the property of the Blackwell-Dubins function and the fact that $\mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(h_t(\vartheta_t)) = \mathcal{U}_{[0,1]}$; see Remark 2.19).

Combining (7.9) with (7.9) and (7.8), we hence have

$$\begin{aligned} \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} [\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(s_{t+1}^{\xi, \pi^c, \mathbb{P}^{\xi, \pi^c}})] &= \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(F(\tilde{s}_t, \tilde{a}_t, (\tilde{\mu}_t \hat{\otimes} \pi_t^c(\cdot | \cdot, \tilde{\mu}_t)), \varepsilon_{t+1}, \varepsilon_{t+1}^0)) \right] \\ &= \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} [\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \hat{f}(\tilde{s}_{t+1})] \\ &= \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\hat{g}_{t+1}(\varepsilon_{1:t+1}^0) \int_S \hat{f}(s) \mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\tilde{s}_{t+1} | \varepsilon_{1:t+1}^0)(ds) \right], \end{aligned}$$

where the last line holds by definition of $\tilde{s}_{t+1}$ given in (7.3).

Moreover, since $\tilde{s}_{t+1}$ is $\mathcal{G}_{t+1}$ measurable, another application of Lemma 6.1 ensures that

$$\mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\tilde{s}_{t+1} | \varepsilon_{1:t+1}^0) = \tilde{\mu}_{t+1}, \quad \mathbb{P}^{\xi, \pi^c}\text{-a.s.},$$

which ensures (7.7) to hold, as claimed.

By the induction hypothesis, (7.6) holds for all $t \geq 0$.

Step 3: Recall that $\mathbb{P}^{\xi, \pi^c} \in \mathcal{Q}$ is the measure induced by $(p_t^{\xi, \pi^c})_{t \geq 1} \in \mathcal{K}^0$ given in (7.2) and (7.5) (see Step 1). Then from Remark 2.3 (iii), it holds that $\mathbb{P}^{\xi, a}$ -a.s.

$$(7.11) \quad \begin{aligned} \mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\varepsilon_1^0) &= p_1^{\xi, \pi^c} \in \mathfrak{P}^0, \\ \mathcal{L}_{\mathbb{P}^{\xi, \pi^c}}(\varepsilon_t^0 | \mathcal{F}_{t-1}^0) &= p_t^{\xi, \pi^c}(\cdot | \varepsilon_{1:t-1}^0) \in \mathfrak{P}^0 \text{ for all } t \geq 2. \end{aligned}$$

Moreover, by (7.6) in Step 2 and (2.33) in Lemma 2.25, it holds that $\mathbb{P}^{\xi, \pi^c}$ -a.s.

$$(7.12) \quad p_1^{\xi, \pi^c} = \bar{p}^*(\underline{\Lambda}_0^{\xi, \pi^c}), \quad p_t^{\xi, \pi^c}(\cdot | \varepsilon_{1:t-1}^0) = \bar{p}^*(\underline{\Lambda}_{t-1}^{\xi, \pi^c}) \quad \text{for all } t \geq 2,$$

which ensures (2.35) to hold, as claimed.

A direct consequence of (2.34) ensures (2.36) to hold, as claimed. This completes the proof. $\square$

7.3. Proof of Corollary 2.28. As the essential arguments of the proof closely follow those of Theorem 2.21, we provide the outline of the proof and omit some details here.

Step 1. For notational simplicity, set $\mu := \mathcal{L}(\xi)$. We first consider for every $n \in \mathbb{N}$

$$\mathcal{I}_n^{\xi, \pi^{c,*}} := \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}} \left[\sum_{t=0}^{n-1} \beta^t r(s_t^{\xi, \pi^{c,*}, \mathbb{P}}, a_t^{\pi^{c,*}, \mathbb{P}}, \Lambda_t^{\xi, \pi^{c,*}, \mathbb{P}}) + \beta^n \bar{V}^*(\mu_n^{\xi, \pi^{c,*}, \mathbb{P}}) \right],$$

where for each $\mathbb{P} \in \mathcal{Q}$, $(\mu_t^{\xi, \pi^{c,*}, \mathbb{P}})_{t \geq 0}$ and $(\Lambda_t^{\xi, \pi^{c,*}, \mathbb{P}})_{t \geq 0}$ are given in (2.32).

Note that by (2.33) in Lemma 2.25 and definition of $\pi_t^{c,*} = \pi_{\text{loc}}^{c,*}$ given in (2.37) together with the property (2.38), it holds for every $\mathbb{P} \in \mathcal{Q}$ that $\mathbb{P}$ -a.s.,

$$\bar{\pi}^*(\mu_t^{\xi, \pi^{c,*}, \mathbb{P}}) = \Lambda_t^{\xi, \pi^{c,*}, \mathbb{P}} \quad \text{for all } t \geq 0.$$

From this, using the same arguments presented for (6.25), we have that for every $n \in \mathbb{N}$

$$\mathcal{I}_n^{\xi, \pi^{c,*}} = \inf_{\mathbb{P} \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}} \left[\sum_{t=0}^{n-1} \beta^t \bar{r}(\mu_t^{\xi, \pi^{c,*}, \mathbb{P}}, \Lambda_t^{\xi, \pi^{c,*}, \mathbb{P}}) + \beta^n \bar{V}^*(\mu_n^{\xi, \pi^{c,*}, \mathbb{P}}) \right].$$

Hence, from the representation of the Markov decision process of the lifted state process in (2.34) (see Lemma 2.25), we can use the same arguments presented for Steps 1 and 2 in the proof of Theorem 2.21 (that relies on the local optimality of $\bar{\pi}^*(\mu_t^{\xi, \pi^{c,*}, \mathbb{P}})$ to $\mathcal{T}\bar{V}^*(\mu_t^{\xi, \pi^{c,*}, \mathbb{P}})$ in Proposition 2.15 (ii) and the fixed point theorem in Proposition 2.16; see Section 6) to have

$$\bar{V}^*(\mu) \leq \limsup_{n \rightarrow \infty} \mathcal{I}_n^{\xi, \pi^{c,*}} \leq \mathcal{J}^{\pi^{c,*}}(\xi) \leq V^c(\xi).$$

Step 2. For every $\pi^c \in \Pi^c$, let $\mathbb{P}^{\xi, \pi^c} \in \mathcal{Q}$ be induced by some $(p_t^{\xi, \pi^c})_{t \geq 1} \in \mathcal{K}^0$ satisfying (2.26) and (2.35) (see Lemma 2.26). Then define $\mathcal{V}^{\pi^c}(\xi)$ by

$$\mathcal{V}^{\pi^c}(\xi) := \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\sum_{t=0}^{\infty} \beta^t r(s_t^{\xi, \pi^c, \mathbb{P}^{\xi, \pi^c}}, a_t^{\pi^c, \mathbb{P}^{\xi, \pi^c}}, \Lambda_t^{\xi, \pi^c}) \right] = \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} \left[\sum_{t=0}^{\infty} \beta^t \bar{r}(\text{pjs}(\underline{\Lambda}_t^{\xi, \pi^c}), \underline{\Lambda}_t^{\xi, \pi^c}) \right],$$

where $\underline{\Lambda}_t^{\xi, \pi^c}$ is the conditional joint law of $(s_t^{\xi, \pi^c, \mathbb{P}^{\xi, \pi^c}}, a_t^{\pi^c, \mathbb{P}^{\xi, \pi^c}})$ under $\mathbb{P}^{\xi, \pi^c}$ given $\varepsilon_{1:t}^0$.

By the local optimality of $\bar{p}^*(\underline{\Lambda}_t^{\xi, \pi^c})$ to $\mathcal{T}\bar{V}^*(\text{pjs}(\underline{\Lambda}_t^{\xi, \pi^c}))$ (see Proposition 2.15 (i)), we can use the same arguments presented for Step 3 in the proof of Theorem 2.21 to have

$$V^c(\xi) \leq \sup_{\pi^c \in \Pi^c} \mathcal{V}^{\pi^c}(\xi) \leq \sup_{\pi^c \in \Pi^c} \sum_{t=0}^{\infty} \left(\beta^t \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} [\bar{V}^*(\underline{\mu}_t^{\xi, \pi^c})] - \beta^{t+1} \mathbb{E}^{\mathbb{P}^{\xi, \pi^c}} [\bar{V}^*(\underline{\mu}_{t+1}^{\xi, \pi^c})] \right) = \bar{V}^*(\mu),$$

where $\underline{\mu}_t^{\xi, \pi^c}$ is the conditional law of $s_t^{\xi, \pi^c, \mathbb{P}^{\xi, \pi^c}}$ under $\mathbb{P}^{\xi, \pi^c}$ given $\varepsilon_{1:t}^0$.

Therefore, we have obtained that $\bar{V}^*(\mu) = V^c(\xi)$, as claimed. In fact, $\bar{V}^*(\mu) = V(\xi)$ follows from Theorem 2.21. Hence the statement (i) holds.

Step 3. Lastly, we consider $\underline{\mathbb{P}}^{\xi, \pi^{c,*}} \in \mathcal{Q}$ which is induced by $(\underline{p}_t^{\xi, \pi^{c,*}})_{t \geq 1} \in \mathcal{K}^0$ satisfying (2.35) and (2.36) (see Lemma 2.26). Then by definition of $\pi^{c,*}$ and of $\underline{\mathbb{P}}^{\xi, \pi^{c,*}}$ (noting that both satisfy the local optimality given in Proposition 2.15), it holds that for every $t \geq 0$

$$\begin{aligned} \mathbb{E}^{\underline{\mathbb{P}}^{\xi, \pi^{c,*}}}[\bar{V}^*(\underline{\mu}_t^{\xi, \pi^{c,*}})] &= \mathbb{E}^{\underline{\mathbb{P}}^{\xi, \pi^{c,*}}}[\mathcal{T}\bar{V}^*(\underline{\mu}_t^{\xi, \pi^{c,*}})] \\ &= \mathbb{E}^{\underline{\mathbb{P}}^{\xi, \pi^{c,*}}} \left[\bar{r}(\text{pj}_S(\underline{\Delta}_t^{\xi, \pi^{c,*}}), \underline{\Delta}_t^{\xi, \pi^{c,*}}) + \beta \int_{\mathcal{P}(S)} \bar{V}^*(\tilde{\mu}') \bar{p}(d\tilde{\mu}' | \text{pj}_S(\underline{\Delta}_t^{\xi, \pi^{c,*}}), \underline{\Delta}_t^{\xi, \pi^{c,*}}, \bar{p}^*(\underline{\Delta}_t^{\xi, \pi^{c,*}})) \right]. \end{aligned}$$

Hence by using the same arguments presented for Step 4 of the proof of Theorem 2.21, we deduce that (2.39) holds. This completes the proof. $\square$

APPENDIX A. SUPPLEMENTARY STATEMENTS

Let us provide some elementary observations on conditional laws.

Lemma A.1. *Fix a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$. Let X be Borel space and Y be measurable space. For every random elements $\mathcal{X}$ and $\mathcal{Y}$ with values in X and Y , respectively, the following hold:*

- (i) *There exists a kernel $k^{\mathcal{X}|\mathcal{Y}} : Y \ni y \mapsto k^{\mathcal{X}|\mathcal{Y}}(dx|y) \in \mathcal{P}(X)$ such that for every $B \in \mathcal{B}(X)$, $\tilde{\mathbb{P}}(\mathcal{X} \in B|\mathcal{Y}) = k^{\mathcal{X}|\mathcal{Y}}(B|\mathcal{Y})$ $\tilde{\mathbb{P}}$ -a.s., and $k^{\mathcal{X}|\mathcal{Y}}$ is unique $\mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{Y})$ -a.e.. As a consequence, $k^{\mathcal{X}|\mathcal{Y}}(\cdot|\mathcal{Y})$ is $\sigma(\mathcal{Y})$ measurable and we denote for every $\tilde{\omega} \in \tilde{\Omega}$*

$$\mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{X}|\mathcal{Y})(\tilde{\omega}) := k^{\mathcal{X}|\mathcal{Y}}(\cdot|\mathcal{Y})(\tilde{\omega}),$$

i.e., a conditional law of $\mathcal{X}$ given $\mathcal{Y}$; see, e.g., [46, Section 6, p.106–107].

- (ii) *If $\mathcal{X}$ is given by $\mathcal{X} = \varphi(\mathcal{Y}, \mathcal{Z})$, where $\varphi : Y \times Z \rightarrow X$ is a measurable function and $\mathcal{Z}$ is a random element in Z and independent of $\mathcal{Y}$, then $\mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{X}|\mathcal{Y}) = \mathcal{L}_{\tilde{\mathbb{P}}}(\varphi(y, \mathcal{Z}))|_{y=\mathcal{Y}}$ and $\mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{X}|\mathcal{Y})$ is $\sigma(\mathcal{Y})$ measurable.*

Proof. Part (i) is shown in [46, Theorem 6.3]. We proceed to prove (ii), which is a consequence of (i) with an application of Fubini's theorem. Clearly, it is sufficient to show that for any bounded measurable function $g : Y \rightarrow \mathbb{R}$ and bounded Borel measurable function $f : X \rightarrow \mathbb{R}$,

$$\mathbb{E}^{\tilde{\mathbb{P}}} \left[g(\mathcal{Y}) \int_X f(x') \mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{X}|\mathcal{Y})(dx') \right] = \mathbb{E}^{\tilde{\mathbb{P}}} \left[g(\mathcal{Y}) \int_X f(x') \mathcal{L}_{\tilde{\mathbb{P}}}(\varphi(y, \mathcal{Z}))|_{y=\mathcal{Y}}(dx') \right].$$

Indeed, by definition of the conditional law $\mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{X}|\mathcal{Y})$ (given in (i)) it holds that

$$\mathbb{E}^{\tilde{\mathbb{P}}} \left[g(\mathcal{Y}) \int_X f(x') \mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{X}|\mathcal{Y})(dx') \right] = \mathbb{E}^{\tilde{\mathbb{P}}} [g(\mathcal{Y}) \mathbb{E}^{\tilde{\mathbb{P}}} [f(\mathcal{X})|\mathcal{Y}]] = \mathbb{E}^{\tilde{\mathbb{P}}} [g(\mathcal{Y}) f(\mathcal{X})] =: \text{I},$$

where the second equality follows from the $\sigma(\mathcal{Y})$ -measurability of $g(\mathcal{Y})$ and the tower property.

Moreover since $\mathcal{X} = \varphi(\mathcal{Y}, \mathcal{Z})$, and $\mathcal{Y}$ and $\mathcal{Z}$ are independent,

$$\begin{aligned} \text{I} &= \mathbb{E}^{\tilde{\mathbb{P}}} [g(\mathcal{Y}) \mathbb{E}^{\tilde{\mathbb{P}}} [f(\varphi(\mathcal{Y}, \mathcal{Z}))|\mathcal{Y}]] = \int_Y g(y) \mathbb{E}^{\tilde{\mathbb{P}}} [f(\varphi(y, \mathcal{Z}))] \mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{Y})(dy) \\ &= \int_Y g(y) \mathbb{E}^{\tilde{\mathbb{P}}} \left[\int_X f(x') \mathcal{L}_{\tilde{\mathbb{P}}}(\varphi(y, \mathcal{Z}))(dx') \right] \mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{Y})(dy) \\ &= \mathbb{E}^{\tilde{\mathbb{P}}} \left[g(\mathcal{Y}) \int_X f(x') \mathcal{L}_{\tilde{\mathbb{P}}}(\varphi(y, \mathcal{Z}))|_{y=\mathcal{Y}}(dx') \right], \end{aligned}$$

where the second equality follows from definition of $\mathcal{L}_{\tilde{\mathbb{P}}}(\varphi(y, \mathcal{Z}))$ and the last one follows from Fubini's theorem (since both f and g are bounded). The $\sigma(\mathcal{Y})$ -measurability of $\mathcal{L}_{\tilde{\mathbb{P}}}(\mathcal{X}|\mathcal{Y})$ follows from (i). This concludes the proof. $\square$

Lemma A.2 (Blackwell and Dubins [12]). *For any Polish space X , there exists a Borel measurable function $\rho_X : \mathcal{P}(X) \times [0, 1] \rightarrow X$ satisfying the following conditions:*

- (i) *for every $\lambda \in \mathcal{P}(X)$ and every uniform random variable $U \sim \mathcal{U}_{[0,1]}$, $\rho_X(\lambda, U)$ is distributed according to λ ;*
- (ii) *for almost every u , the map $\lambda \mapsto \rho_X(\lambda, u)$ is continuous w.r.t. the weak topology of $\mathcal{P}(X)$.*

We call ρ_X the Blackwell–Dubins function of the space X .

Lemma A.3 (Universal disintegration; see, e.g., [47, Corollary 1.26]). *For any Borel spaces X and Y , there exists a kernel $\mathcal{K}_{X \times Y} : X \times \mathcal{P}(X \times Y) \times \mathcal{P}(X) \ni (x, \lambda, \eta) \mapsto \mathcal{K}_{X \times Y}(\cdot | x, \lambda, \eta) \in \mathcal{P}(Y)$ such that for every $\lambda \in \mathcal{P}(X \times Y)$ and $\eta \in \mathcal{P}(X)$ satisfying $\text{pj}_X(\lambda) \ll \eta$, it holds that*

$$\lambda = \eta \hat{\otimes} \mathcal{K}_{X \times Y}(\cdot | \cdot, \lambda, \eta),$$

Moreover, $\mathcal{K}_{X \times Y}(\cdot | \cdot, \lambda, \eta)$ is unique η -a.e. for fixed λ and η .

REFERENCES

- [1] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis: A Hitchhiker's Guide*. Springer, 2006.
- [2] A. Balata, C. Huré, M. Laurière, H. Pham, and I. Pimentel. A class of finite-dimensional numerically solvable McKean–Vlasov control problems. *ESAIM: Proc. Surv.*, 65:114–144, 2019.
- [3] M. Basei and H. Pham. A weak martingale approach to linear-quadratic McKean–Vlasov stochastic control problems. *J. Optim. Theory Appl.*, 181:347–382, 2019.
- [4] N. Bäuerle. Mean field Markov decision processes. *Appl. Math. Optim.*, 88(1):12, 2023.
- [5] N. Bäuerle and A. Glauner. Distributionally robust Markov decision processes and their connection to risk measures. *Math. Oper. Res.*, 47(3):1757–1780, 2022.
- [6] N. Bäuerle and U. Rieder. *Markov decision processes with applications to finance*. Springer Science & Business Media, 2011.
- [7] E. Bayraktar and T. Chen. Nonparametric adaptive robust control under model uncertainty. *SIAM Journal on Control and Optimization*, 61(5):2737–2760, 2023.
- [8] E. Bayraktar, A. Cosso, and H. Pham. Randomized dynamic programming principle and Feynman–Kac representation for optimal control of McKean–Vlasov dynamics. *Trans. Amer. Math. Soc.*, 370(3):2115–2160, 2018.
- [9] E. Bayraktar and A. D. Kara. Infinite horizon average cost optimality criteria for mean-field control. *SIAM J. Control Optim.*, 62(5):2776–2806, 2024.
- [10] A. Bensoussan, J. Frehse, and P. Yam. *Mean field games and mean field type control theory*, volume 101. New York: Springer-Verlag, 2013.
- [11] A. Bensoussan, P. J. Graber, and S. C. P. Yam. Control on Hilbert spaces and application to some mean field type control problems. *Ann. Appl. Probab.*, 34(4):4085–4136, 2024.
- [12] D. Blackwell and L. E. Dubins. An extension of Skorohod’s almost sure representation theorem. *Proc. Amer. Math. Soc.*, 89(4):691–692, 1983.
- [13] V. I. Bogachev. *Measure Theory: Volume II*. Springer, 2007.
- [14] E. Boissard and T. Le Gouic. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. In *Ann. inst. Henri Poincaré (B) Probab. Stat.*, volume 50, pages 539–563, 2014.
- [15] M. Burzoni, V. Ignazio, A. M. Reppen, and H. M. Soner. Viscosity solutions for controlled McKean–Vlasov jump-diffusions. *SIAM J. Control Optim.*, 58(3):1676–1699, 2020.
- [16] R. Carmona. Applications of mean field games in financial engineering and economic theory. *preprint, arXiv:2012.05237*, 2020.
- [17] R. Carmona and F. Delarue. *Probabilistic theory of mean field games with applications I-II*. Springer, 2018.
- [18] R. Carmona, J.-P. Fouque, and L.-H. Sun. Mean field games and systemic risk. *Commun. Math. Sci.*, 13(4):911–933, 2015.
- [19] R. Carmona, K. Hamidouche, M. Laurière, and Z. Tan. Policy optimization for linear-quadratic zero-sum mean-field type games. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 1038–1043. IEEE, 2020.
- [20] R. Carmona, K. Hamidouche, M. Laurière, and Z. Tan. Linear-quadratic zero-sum mean-field type games: Optimality conditions and policy optimization. *J. Dyn. Games*, 8(4), 2021.
- [21] R. Carmona and M. Laurière. Deep learning for mean field games and mean field control with applications to finance. *preprint, arXiv:2107.04568*, 7, 2021.

- [22] R. Carmona, M. Laurière, and Z. Tan. Model-free mean-field reinforcement learning: mean-field MDP and mean-field Q -learning. *Ann. Appl. Probab.*, 33(6B):5334–5381, 2023.
- [23] Z. Chen and L. Epstein. Ambiguity, risk, and asset returns in continuous time. *Econometrica*, 70(4):1403–1443, 2002.
- [24] A. Cosso and H. Pham. Zero-sum stochastic differential games of generalized McKean–Vlasov type. *J. Math. Pures Appl.*, 129:180–212, 2019.
- [25] K. Cui, M. Li, C. Fabian, and H. Koeppl. Scalable task-driven robotic swarm control via collision avoidance and learning mean-field control. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1192–1199. IEEE, 2023.
- [26] M. F. Djete. Extended mean field control problem: A propagation of chaos result. *Electron. J. Probab.*, 27:1–53, 2022.
- [27] M. F. Djete, D. Possamaï, and X. Tan. McKean–Vlasov optimal control: limit theory and equivalence between different formulations. *Math. Oper. Res.*, 47(4):2891–2930, 2022.
- [28] M. F. Djete, D. Possamaï, and X. Tan. McKean–Vlasov optimal control: The dynamic programming principle. *Ann. Probab.*, 50(2):791–833, 2022.
- [29] J. Dow and S. R. da Costa Werlang. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica*, pages 197–204, 1992.
- [30] L. El Ghaoui and A. Nilim. Robust solutions to Markov decision problems with uncertain transition matrices. *Oper. Res.*, 53(5):780–798, 2005.
- [31] K. Elamvazhuthi and S. Berman. Mean-field models in swarm robotics: A survey. *Bioinsp. Biomim.*, 15(1):015001, 2019.
- [32] R. Elie, E. Hubert, T. Mastrolia, and D. Possamaï. Mean-field moral hazard for optimal energy demand response management. *Math. Finance*, 31(1):399–473, 2021.
- [33] M. Fischer and G. Livieri. Continuous time mean-variance portfolio optimization through the mean field approach. *ESAIM: Probab. Stat.*, 20:30–44, 2016.
- [34] M. Fornasier, S. Lisini, C. Orrieri, and G. Savaré. Mean-field optimal control as gamma-limit of finite agent controls. *Eur. J. Appl. Math.*, 30(6):1153–1186, 2019.
- [35] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields*, 162(3):707–738, 2015.
- [36] G. Fu, U. Horst, and X. Xia. A mean-field control problem of optimal portfolio liquidation with semimartingale strategies. *Math. Oper. Res.*, 49(4):2356–2384, 2024.
- [37] L. Garlappi, R. Uppal, and T. Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Rev. Financial Stud.*, 20(1):41–81, 2007.
- [38] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *J. Math. Econ.*, 18(2):141–153, 1989.
- [39] P. J. Graber. Linear quadratic mean field type control and mean field games with common noise, with application to production of an exhaustible resource. *Appl. Math. Optim.*, 74:459–486, 2016.
- [40] H. Gu, X. Guo, X. Wei, and R. Xu. Mean-field controls with Q-learning for cooperative MARL: convergence and complexity analysis. *SIAM J. Math. Data Sci.*, 3(4):1168–1196, 2021.
- [41] H. Gu, X. Guo, X. Wei, and R. Xu. Dynamic programming principles for mean-field controls with learning. *Oper. Res.*, 71(4):1040–1054, 2023.
- [42] H. Gu, X. Guo, X. Wei, and R. Xu. Mean-field multiagent reinforcement learning: A decentralized network approach. *Math. Oper. Res.*, 50(1):506–536, 2025.
- [43] L. P. Hansen and T. J. Sargent. *Robustness*. Princeton University Press, 2008.
- [44] J. Huang and M. Huang. Robust mean field linear-quadratic-gaussian games with unknown L^2 -disturbance. *SIAM J. Control Optim.*, 55(5):2811–2840, 2017.
- [45] J. Huang, B.-C. Wang, and J. Yong. Social optima in mean field linear-quadratic-Gaussian control with volatility uncertainty. *SIAM J. Control Optim.*, 59(2):825–856, 2021.
- [46] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications. Springer, New York, 2nd ed., 2002.
- [47] O. Kallenberg. *Random measures, theory and applications*. Springer, 2017.
- [48] A. Khamis, A. Hussein, and A. Elmogy. Multi-robot task allocation: A review of the state-of-the-art. *Cooperative robots and sensor networks 2015*, pages 31–51, 2015.
- [49] D. Lacker. Limit theory for controlled McKean–Vlasov dynamics. *SIAM J. Control Optim.*, 55(3):1641–1672, 2017.
- [50] J. Langner, A. Neufeld, and K. Park. Markov-Nash equilibria in mean-field games under model uncertainty. preprint, *arXiv:2410.11652*, 2024.

- [51] M. Laurière and O. Pironneau. Dynamic programming for mean-field type control. *J. Optim. Theory Appl.*, 169(3):902–924, 2016.
- [52] K. Lerman, A. Martinoli, and A. Galstyan. A review of probabilistic macroscopic models for swarm robotic systems. In *International workshop on swarm robotics*, pages 143–152. Springer, 2004.
- [53] M. Li, D. Kuhn, and T. Sutter. Policy gradient algorithms for robust MDPs with non-rectangular uncertainty sets. *preprint, arXiv:2305.19004*, 2023.
- [54] Y. Liang, B.-C. Wang, and H. Zhang. Robust mean field linear quadratic social control: Open-loop and closed-loop strategies. *SIAM J. Control Optim.*, 60(4):2184–2213, 2022.
- [55] Z. Liu, Q. Bai, J. Blanchet, P. Dong, W. Xu, Z. Zhou, and Z. Zhou. Distributionally robust Q -learning. In *International Conference on Machine Learning*, pages 13623–13643. PMLR, 2022.
- [56] C. I. Lu, J. Sester, and A. Zhang. Distributionally robust deep Q -learning. *preprint, arXiv:2505.19058*, 2025.
- [57] M. Motte and H. Pham. Mean-field Markov decision processes with common noise and open-loop controls. *Ann. Appl. Probab.*, 32(2):1421–1458, 2022.
- [58] M. Motte and H. Pham. Quantitative propagation of chaos for mean field Markov decision process with common noise. *Electron. J. Probab.*, 28:1–24, 2023.
- [59] A. Neufeld and J. Sester. Robust Q -learning algorithm for Markov decision processes under Wasserstein uncertainty. *Automatica*, 168:111825, 2024.
- [60] A. Neufeld and J. Sester. Non-concave distributionally robust stochastic control in a discrete time finite horizon setting. *To appear in Math. Finance, arXiv:2404.05230*, 2025+.
- [61] A. Neufeld, J. Sester, and M. Šikić. Markov decision processes under model uncertainty. *Math. Finance*, 33(3):618–665, 2023.
- [62] H. Pham. Linear quadratic optimal control of conditional McKean–Vlasov equation with random coefficients and applications. *Probab. Uncertain. Quant. Risk*, 1:1–26, 2016.
- [63] H. Pham and X. Wei. Discrete time McKean–Vlasov control problem: A dynamic programming approach. *Appl. Math. Optim.*, 74(3):487–506, 2016.
- [64] H. Pham and X. Wei. Dynamic programming for optimal control of stochastic McKean–Vlasov dynamics. *SIAM J. Control Optim.*, 55(2):1069–1101, 2017.
- [65] S. Sanjari and S. Yüksel. Optimal solutions to infinite-player stochastic teams and mean-field teams. *IEEE Trans. Autom. Control*, 66(3):1071–1086, 2020.
- [66] H. M. Soner and Q. Yan. Viscosity solutions for McKean–Vlasov control on a torus. *SIAM J. Control Optim.*, 62(2):903–923, 2024.
- [67] C. Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.
- [68] B.-C. Wang and J. Huang. Social optima in robust mean field LQG control. In *2017 11th Asian Control Conference (ASCC)*, pages 2089–2094. IEEE, 2017.
- [69] B.-C. Wang, J. Huang, and J.-F. Zhang. Social optima in robust mean field LQG control: From finite to infinite horizon. *IEEE Trans. Autom. Control*, 66(4):1529–1544, 2020.
- [70] B.-C. Wang and J.-F. Zhang. Social optima in mean field linear-quadratic-gaussian models with markov jump parameters. *SIAM J. Control Optim.*, 55(1):429–456, 2017.
- [71] W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Math. Oper. Res.*, 38(1):153–183, 2013.
- [72] H. Xu and S. Mannor. Distributionally robust Markov decision processes. *Math. Oper. Res.*, 37(2):288–301, 2012.
- [73] I. Yang. A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance. *IEEE Control Syst. Lett.*, 1(1):164–169, 2017.
- [74] M. A. U. Zaman, M. Lauriere, A. Koppel, and T. Başar. Robust cooperative multi-agent reinforcement learning: A mean-field type game perspective. In *6th Annual Learning for Dynamics & Control Conference*, pages 770–783. PMLR, 2024.

SHANGHAI FRONTIERS SCIENCE CENTER OF ARTIFICIAL INTELLIGENCE AND DEEP LEARNING, AND NYU-ECNU INSTITUTE OF MATHEMATICAL SCIENCES, NYU SHANGHAI
 Email address: mathieu.lauriere@nyu.edu

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY
 Email address: ariel.neufeld@ntu.edu.sg

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY
 Email address: kyunghyun.park@ntu.edu.sg