Bounded Rationality Equilibrium Learning in Mean Field Games

Yannick Eich, Christian Fabian, Kai Cui, Heinz Koeppl

Dept. of Electrical Engineering and Information Technology, Technische Universität Darmstadt
{yannick.eich, heinz.koeppl}@tu-darmstadt.de

Abstract

Mean field games (MFGs) tractably model behavior in large agent populations. The literature on learning MFG equilibria typically focuses on finding Nash equilibria (NE), which assume perfectly rational agents and are hence implausible in many realistic situations. To overcome these limitations, we incorporate bounded rationality into MFGs by leveraging the well-known concept of quantal response equilibria (QRE). Two novel types of MFG QRE enable the modeling of large agent populations where individuals only noisily estimate the true objective. We also introduce a second source of bounded rationality to MFGs by restricting the agents' planning horizon. The resulting novel receding horizon (RH) MFGs are combined with QRE and existing approaches to model different aspects of bounded rationality in MFGs. We formally define MFG QRE and RH MFGs and compare them to existing equilibrium concepts such as entropy-regularized NE. Subsequently, we design generalized fixed-point iteration and fictitious play algorithms to learn QRE and RH equilibria. After a theoretical analysis, we give different examples to evaluate the capabilities of our learning algorithms and outline practical differences between the equilibrium concepts.

Introduction

Learning equilibria in multi-agent games is of great practical interest but hard to scale to many agents (Daskalakis, Goldberg, and Papadimitriou 2009; Deng et al. 2023). Mean field games (MFGs) allow scaling to arbitrarily many exchangeable agents at fixed complexity. MFGs are of recent interest as a tractable method to learn approximate equilibria of rational, selfish agents (Guo et al. 2019; Cui and Koeppl 2021; Xie et al. 2021; Laurière et al. 2022; Anahtarci, Kariksiz, and Saldi 2023). Thus, MFGs are applied in various settings ranging from finance to engineering (Djehiche, Tcheukam, and Tembine 2017; Achdou et al. 2020; Carmona 2020).

A common concept in multi-agent learning is the Nash equilibrium (NE), where each player's strategy is optimal given others', leading to no incentive for agents to change strategies. The optimality notion inherent in NE assumes full rationality of the individual agents.

However, in many real-world situations individuals may not behave perfectly rational due to limited information processing capabilities, psychological factors, social considerations or other factors. Deviations from perfect rationality are described by the fundamental concept of bounded rationality (Simon 1955, 1979; Kahneman and Tversky 1982; Selten 1990; Gigerenzer and Selten 2002; Kahneman 2013). Bounded rationality implies that for many real-world scenarios NE are insufficient due to their rigorous perfect rationality assumption. Instead of NE, we require a more realistic equilibrium concept accounting for partially irrational agents.

A popular game-theoretic approach to modeling bounded rationality of agents are quantal response equilibria (QRE) (McKelvey and Palfrey 1995, 1998) which are used, e.g., in economics (Breitmoser, Tan, and Zizzo 2010), robust RL (Reddi et al. 2024) and for efficient NE approximation (Gemp, Marris, and Piliouras 2024). Intuitively, in a QRE agents perceive rewards perturbed by noise and act optimally with respect to these perturbed rewards. In our work, we extend QRE to the domain of MFGs to model the behavior of a large number of agents who deviate from perfect rationality.

Meanwhile, on the control-theoretic side, a common approximately optimal control method is model predictive control (MPC) (Kouvaritakis and Cannon 2016), also known as receding horizon control. To further enhance modeling of bounded rationality in MFGs, we incorporate a receding horizon method, where agents make decisions based on a limited future time horizon, reflecting more realistic decision-making processes. In contrast to MPC-based variants of MFGs such as (Inoue et al. 2021), we analyze the resulting novel receding horizon equilibria and instead focus on *learning* such equilibria, in a *discrete-time* setting.

Beyond realism, introducing bounded rationality yields possible tractability advantages. NE computation for MFGs can be hard, motivating the search for alternative equilibrium notions. We show that under certain assumptions, QRE can be computed using a fixed-point iteration (FPI). Moreover, QRE solutions can be seen as NE approximations with arbitrarily accurate design (Eibelshäuser and Poensgen 2019). Recently, different equilibria have been introduced as NE approximations in MFGs (Cui and Koeppl 2021). We compare QRE with these equilibria theoretically and empirically and provide a new algorithm to compute QRE which extends to these equilibria. For receding horizon equilibria, we develop novel algorithms effective in theory and practice.

Our main contributions are:

- We formulate QRE for MFGs to incorporate bounded rationality for a more realistic MFG framework;
- We integrate a receding horizon method tailored to the limited lookahead capacity of realistic agents;
- We give theoretical and empirical results to put MFG QRE in context to existing equilibrium concepts;
- We generalize the known fictitious play (FP) and FPI algorithms for NE to learn QRE and other equilibria;
- We provide empirical examples to demonstrate the capabilities of our learning algorithms.

Equilibria in MFGs

In this section, we first describe finite games in discrete time and their corresponding MFGs. We then define common and new equilibrium notions as solution concepts and desired results of multi-agent equilibrium learning algorithms, which are introduced thereafter. Proofs and additional details can be found in the full preprint version (Eich et al. 2024).

Notation: Denote by $\mathcal{P}(\mathcal{X})$ the space of probability measures on finite set $\mathcal{X}$, equipped with the L_1 norm $\|\cdot\|$ unless noted otherwise. Equip products of metric spaces with the sup *metric. Further, let $[N] \coloneqq \{1, \dots, N\}$ for $N \in \mathbb{N}$.*

Finite Agent Games

For the finite N-agent game of practical interest, consider agents $i \in [N]$ endowed with random states x_t^i and actions u_t^i at all times $t \in \mathcal{T} \coloneqq \{0, 1, \dots, T-1\}$ up to time horizon $T \in \mathbb{N}$. Let $\mathcal{X}$ and $\mathcal{U}$ be the finite state and action spaces for agents, respectively. The empirical mean field (MF) $\mu_t^N \coloneqq \frac{1}{N}\sum_{i=1}^N \mathbf{1}_{x_t^i}$ can be understood as a histogram of agent states. Each agent i implements stochastic Markovian policies $\pi^i \in \Pi \equiv \mathcal{P}(\mathcal{U})^{\mathcal{X}\times\mathcal{T}}$ depending on the current time and local agent state. For some initial state distribution μ_0 with $x_0^i \sim \mu_0$, for all agents i define state-action dynamics

$$u_t^i \sim \pi_t^i(u_t^i \mid x_t^i), \quad x_{t+1}^i \sim p_t(x_{t+1}^i \mid x_t^i, u_t^i, \mu_t^N) \tag{1}$$

given some transition kernels $p_t \colon \mathcal{X} \times \mathcal{U} \times \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$.

Competitive agents aim to optimize their own objective while predicting other agents' behavior. The objective notion depends on the chosen equilibrium notion, but typically uses functions $r_t \colon \mathcal{X} \times \mathcal{U} \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ resulting in rewards $r_t(x_t^i, u_t^i, \mu_t^N)$ at all times $t \in \mathcal{T}$ to maximize.

Mean Field Games

MFGs are the limit of finite N-agent games with $N \to \infty$ and approximate many-agent finite games well. By a law of large numbers, the empirical MF is essentially replaced by its deterministic limiting MF. The idea of MFGs is to find approximate (symmetric) equilibria, which are otherwise hard to find in finite games with many agents (Deng et al. 2023). MFGs assume all agents to symmetrically play the same policy $\pi^* \in \Pi$ – the equilibrium solution. Whenever agent i deviates from π^* and instead uses some policy π, this corresponds to the policy tuple $(\pi, \underline{\pi}^{-i})$, where $\underline{\pi}^{-i} = (\pi^*, \dots, \pi^*)$ denotes all but the i-th policy in the finite game. Hence, in the limit as $N \to \infty$ we have

$$u_t \sim \pi_t(u_t \mid x_t), \quad x_{t+1} \sim p_t(x_{t+1} \mid x_t, u_t, \mu_t) \tag{2}$$

for the representative deviating agent. Here, the empirical MFs μ_t^N are replaced by the deterministic limiting MF $\mu \coloneqq (\mu_t)_{t\in\mathcal{T}} \in \mathcal{M} \subseteq \mathcal{P}(\mathcal{X})^{\mathcal{T}}$, given by the probability law μ of any other agent playing the assumed equilibrium policy π^*. Further, $\mathcal{M}$ is the space of all obtainable MFs. We write $\mu = \Gamma_{\mathcal{M}}(\pi^*)$, defined by fixed initial μ_0 and the recursion

$$\mu_{t+1}(x') = \sum_{x\in\mathcal{X}} \mu_t(x) \sum_{u\in\mathcal{U}} \pi_t^*(u \mid x) p_t(x' \mid x, u, \mu_t).$$

Notions of Non-Cooperative Equilibria

As discussed, there are many equilibrium notions. Here, we focus on non-cooperative equilibria where agents optimize over independent policies to maximize their own objective.

Nash equilibria. First, we have the standard objective of any agent i given as

$$J^i(\hat{\pi}, \underline{\pi}^{-i}) = \mathbb{E}\left[\sum_{t\in\mathcal{T}} r(x_t^i, u_t^i, \mu_t^N)\right],$$

which, for $\underline{\pi}^{-i} = \times_{j\neq i}\pi$, in the limiting MFG yields

$$J(\hat{\pi}, \pi) = \mathbb{E}\left[\sum_{t\in\mathcal{T}} r(x_t, u_t, \mu_t)\right], \quad \mu = \Gamma_{\mathcal{M}}(\pi) \tag{3}$$

where only agent i deviates from policy π to $\hat{\pi}$. If agents are rational and anticipate other agents' decisions, all agents should use policies such that none can improve their objective by deviating from the equilibrium. This leads to the well-known Nash equilibrium.

Definition 1 (Exploitability)**.** *In the finite game, $\mathcal{E}^N(\underline{\pi}) \coloneqq \max_{i\in[N]} \sup_{\hat{\pi}\in\Pi}\{J^i(\hat{\pi}, \underline{\pi}^{-i}) - J^i(\underline{\pi})\}$ is the exploitability of a policy tuple $\underline{\pi} \in \Pi^N$. The limiting exploitability $\mathcal{E}(\pi)$ of a policy $\pi \in \Pi$ is $\mathcal{E}(\pi) \coloneqq \max_{\hat{\pi}\in\Pi} J(\hat{\pi}, \pi) - J(\pi, \pi)$.*

Definition 2 (Approximate NE)**.** *For any $\epsilon > 0$, an ϵ-approximate NE (ϵ-NE) is a policy tuple $\underline{\pi} \in \Pi^N$ with $\mathcal{E}^N(\underline{\pi}) \leq \epsilon$. An exact NE is an ϵ-NE with $\epsilon = 0$.*

The resulting limiting NE thus becomes a policy that performs optimally against itself, i.e. when all other agents also use the same policy (Saldi, Basar, and Raginsky 2018).

Definition 3 (Mean Field NE)**.** *A Mean Field NE (MFNE) is a policy $\pi^* \in \Pi$ such that $\pi^* \in \arg\max_{\pi\in\Pi} J(\pi, \pi^*)$.*

MFNE are well-known to constitute approximate NE in large finite N-agent games, in the sense of a negligible exploitability, rigorously motivating MFGs and MFNE under mild continuity assumptions of the game.

Assumption 1. *The transition kernel P and reward function r are continuous in their MF argument.*

Proposition 1 (Saldi, Basar, and Raginsky (2018, Thm. 3.3, 4.1))**.** *Under Assm. 1, a MFNE π^* exists, and yields a finite game ϵ-NE $\underline{\pi}^* = (\pi^*, \dots, \pi^*)$, with $\epsilon \to 0$ as $N \to \infty$.*

The maximization of Eq. (3) for a fixed MF μ involves the optimal state-action value function given by the Bellman recursion $Q^* \equiv Q^{\mu,*} = \Gamma_{Q^*}(\mu)$ (suppressing μ) defined as

$$Q_t^*(x,u) = r(x,u,\mu_t) + \sum_{x' \in \mathcal{X}} p_t(x' \mid x,u,\mu_t) \max_{u' \in \mathcal{U}} Q_{t+1}^*(x',u'), \quad (4)$$

with $Q_{T-1}^*(x,u) = r(x,u,\mu_{T-1})$. The optimal policy is then obtained by maximizing Q^* with respect to action u, for which we write $\pi^* = \Gamma_\Pi^*(Q^*)$. We can then rewrite Def. 3 as the fixed-point equation $\pi^* = \Gamma_\Pi^*(\Gamma_{Q^*}(\Gamma_\mathcal{M}(\pi^*)))$.

Regularized equilibria. A common alternative to MFNE is to use regularized control (Geist, Scherrer, and Pietquin 2019; Belousov and Peters 2019) (typically entropy regularization). The idea is to replace the objective in Eq. (3) by an entropy-regularized one, which maximizes the entropy $\mathcal{H}(\hat{\pi}_t(\cdot \mid x_t)) := -\sum_{u \in \mathcal{U}} \hat{\pi}_t(u \mid x_t) \log \hat{\pi}_t(u \mid x_t)$ of policies in encountered states x_t,

$$J_\alpha^{\mathrm{RE}}(\hat{\pi}, \pi) := \mathbb{E}\left[\sum_{t \in \mathcal{T}} r(x_t,u_t,\mu_t) + \alpha \mathcal{H}(\hat{\pi}_t(\cdot \mid x_t))\right]$$

with $\mu = \Gamma_\mathcal{M}(\pi)$, and temperature $\alpha > 0$. Accordingly, we define regularized equilibria (RE) similar to MFNE.

Definition 4 (RE)**.** *A RE is a policy $\pi^* \in \Pi$ with $\pi^* = \arg\max_{\pi \in \Pi} J_\alpha^{\mathrm{RE}}(\pi, \pi^*)$.*

Note that for fixed $\mu \in \mathcal{M}$, it is known (e.g., Cui and Koeppl (2021)) that the optimal policy $\hat{\pi}^{\mu,\alpha}$ is

$$\hat{\pi}_t^{\mu,\alpha}(u \mid x) = \frac{\exp\left(\frac{1}{\alpha}\tilde{Q}_t^{\mu,\alpha}(x,u)\right)}{\sum_{u' \in \mathcal{U}} \exp\left(\frac{1}{\alpha}\tilde{Q}_t^{\mu,\alpha}(x,u')\right)} \quad (5)$$

and write $\hat{\pi}^{\mu,\alpha} = \Gamma_\Pi^\alpha(\tilde{Q}^{\mu,\alpha})$ for such softmax policies, given the soft state-action value function $\tilde{Q}_t^{\mu,\alpha}(x,u)$. We also write $\tilde{Q}_t^{\mu,\alpha} = \Gamma_{\tilde{Q}}^\alpha(\mu)$ for the soft state-action value function given μ, defined through the smooth-maximum Bellman recursion

$$\tilde{Q}_t^{\mu,\alpha}(x,u) = r(x,u,\mu_t) + \sum_{x' \in \mathcal{X}} p_t(x' \mid x,u,\mu_t) \cdot \alpha \log\left(\sum_{u' \in \mathcal{U}} \exp\left(\frac{1}{\alpha}\tilde{Q}_{t+1}^{\mu,\alpha}(x',u')\right)\right) \quad (6)$$

with $\tilde{Q}_{T-1}^{\mu,\alpha}(x,u) = r(x,u,\mu_{T-1})$. For readability, we omit super- and subscript μ, α where it is clear from context. Now, we can rewrite Def. 4 as $\pi^* = \Gamma_\Pi^\alpha(\Gamma_{\tilde{Q}}^\alpha(\Gamma_\mathcal{M}(\pi^*)))$.

There are various reasons for regularization, such as improved tractability in learning MFG equilibria (Anahtarci, Kariksiz, and Saldi 2023; Cui and Koeppl 2021; Laurière et al. 2022; Li et al. 2024), exploratory properties in RL settings (Guo, Xu, and Zariphopoulou 2022), and robustness against model uncertainty (Eysenbach and Levine 2021).

Quantal response equilibria. To incorporate bounded rationality, we assume agents act suboptimally and do not exactly optimize an objective. Building on economics literature (Breitmoser, Tan, and Zizzo 2010; Eibelshäuser and Poensgen 2019), we introduce Markov QRE as an MFG equilibrium notion where agents only noisily estimate the state-action value function. Depending on the noise, agents act independently according to their own estimates, to the best of their knowledge. In the mentioned literature, the QRE definition is based on the state-action values $Q^\pi \equiv Q^{\mu,\pi} = \Gamma_{Q^\pi}(\mu,\pi)$ of a policy π under current MF μ, given by the recursion

$$Q_t^\pi(x,u) = r(x,u,\mu_t) + \sum_{x' \in \mathcal{X}} p_t(x' \mid x,u,\mu_t) \sum_{u' \in \mathcal{U}} \pi_t(u' \mid x') Q_{t+1}^\pi(x',u'),$$

with $Q_{T-1}^\pi(x,u) = r(x,u,\mu_{T-1})$. We extend this to the optimal state-action value function Q^* defined in Eq. (4) and denote the resulting equilibria as Q$^\pi$RE and Q*RE, respectively. For Q$^\pi$RE, the noisy state-action value function is

$$\hat{Q}_t^\pi(x,u) = Q_t^\pi(x,u) + \epsilon_t(x,u),$$

given policy π, where ϵ_t is sampled from a distribution $p(\epsilon_t)$. We then describe the set of realizations of ϵ_t, where agents in state x perceive action u as optimal, called response set, as

$$\mathcal{R}_{t,x,u} = \{\epsilon_t \colon \hat{Q}_t^\pi(x,u) > \hat{Q}_t^\pi(x,u') \quad \forall u' \neq u\}.$$

The probability that agents choose action u corresponds to the probability mass of the respective response set. Next, we define the corresponding equilibrium.

Definition 5 (Q$^\pi$RE)**.** *A (Markov) Q$^\pi$RE is a policy π^* s.t.*

$$\pi_t^*(u \mid x) = \int_{\mathcal{R}_{t,x,u}} p(\epsilon_t) \mathrm{d}\epsilon_t$$

for all $t \in \mathcal{T}, x \in \mathcal{X}, u \in \mathcal{U}$, and $\mu = \Gamma_\mathcal{M}(\pi^*)$.

For the Q*RE we define the noisy estimates of Q^* as

$$\hat{Q}_t^*(x,u) = Q_t^*(x,u) + \epsilon_t(x,u).$$

We equivalently define the resulting response set $\mathcal{R}_{t,x,u}^*$ and the corresponding equilibrium.

Definition 6 (Q*RE)**.** *A (Markov) Q*RE is a policy π^* s.t.*

$$\pi_t^*(u \mid x) = \int_{\mathcal{R}_{t,x,u}^*} p(\epsilon_t) \mathrm{d}\epsilon_t$$

for all $t \in \mathcal{T}, x \in \mathcal{X}, u \in \mathcal{U}$, and $\mu = \Gamma_\mathcal{M}(\pi^*)$.

If the noise follows a Gumbel distribution with parameter λ, Q$^\pi$RE and Q*RE policies can be computed analytically for fixed μ, as softmax policies $\pi^* = \Gamma_\Pi^{1/\lambda}(Q^\pi)$ and $\pi^* = \Gamma_\Pi^{1/\lambda}(Q^*)$, respectively, where $\alpha = 1/\lambda$. The special case of Q$^\pi$RE leads to the MFG analogue of the so-called logit equilibrium, considered in economics (Breitmoser, Tan, and Zizzo 2010; Eibelshäuser and Poensgen 2019), while the special case of Q*RE leads to the Boltzmann equilibrium (Guo et al. 2019; Cui and Koeppl 2021).

Definition 7 (Logit Q$^\pi$RE)**.** *A Logit Q$^\pi$RE (LQ$^\pi$RE) is a policy $\pi^* \in \Pi$ such that $\pi^* = \Gamma_\Pi^{1/\lambda}(\Gamma_{Q^\pi}(\Gamma_\mathcal{M}(\pi^*), \pi^*))$.*

Definition 8 (Boltzmann equilibrium)**.** *A BE is a policy $\pi^* \in \Pi$ such that $\pi^* = \Gamma_\Pi^{1/\lambda}(\Gamma_{Q^*}(\Gamma_\mathcal{M}(\pi^*)))$.*

In the following, Q$^\pi$RE and Q*RE usually refer to their special cases of LQ$^\pi$RE and BE. We show that such equilibria are guaranteed to exist. For BE see (Cui and Koeppl 2021).

Proposition 2. *For any $\lambda > 0$, a Q$^\pi$RE exists under Assm. 1.*

Other MFG equilibrium concepts. The literature contains many equilibrium concepts for MFGs. One example are (coarse) correlated equilibria (Campi and Fischer 2022; Muller et al. 2022a,b) where agents obtain advice from a mediator to align their actions. The resulting correlation mechanism enables efficient equilibria calculation under moderate assumptions compared to NE. Furthermore, there are Stackelberg equilibria for MFGs (Elie, Mastrolia, and Possamaï 2019; Carmona and Wang 2021; Carmona, Dayanıklı, and Laurière 2022; Vasal and Berry 2022) where one principal tries to optimally incentivize a mean field of infinitely many agents. Since a detailed discussion of these and many more MFG equilibrium concepts is beyond the scope of our paper, we focus on NE, Q$^\pi$RE, Q*RE, and RE instead and refer to the above references and Fudenberg and Tirole (1991) for a general, not MFG specific overview of equilibrium concepts.

Receding horizon equilibria. Next, we describe a second method to model bounded rationality that can be combined with the previously defined equilibrium concepts. We introduce receding horizon (RH) equilibria to model limited lookahead capacity of agents by considering a shorter horizon for the underlying Bellman recursion. They describe the behaviour of agents with a model predictive control (MPC) (Kouvaritakis and Cannon 2016), where decisions are based on a shortened future horizon, and therefore allows for more realistic or practical MFG models.

The previously defined objectives J, e.g. for NE or RE, are sums over the whole time horizon $\mathcal{T}$. In the RH scenario, however, agents plan ahead only the next H time steps beyond the current time $t \in \mathcal{T}$. Like in MPC, we assume that agents apply the first action of the resulting policy and then repeat the optimization for the next time step. A RH equilibrium is thus an ensemble of sequential MFG equilibria.

For RH MFG, define the respective RH NE as follows. For an agent at time t, the RH objective given the MF policy π is

$$J_t^H(\hat{\pi}, \pi) \coloneqq \mathbb{E}\left[\sum_{t'=t}^{\min(T,t+H)} r\left(x_{t'}, u_{t'}, \mu_{t,t'}^H\right)\right],$$

with $\mu_t^H = \Gamma_{\mathcal{M},t}^H(\pi)$, defined by initial $\mu_{t,t}^H$ and the recursion

$$\mu_{t,t'+1}^H(x') = \sum_{x\in\mathcal{X}} \mu_{t,t'}^H(x) \sum_{u\in\mathcal{U}} \pi_{t'}(u \mid x) p_{t'}(x' \mid x, u, \mu_{t,t'}^H).$$

Definition 9 (RH NE)**.** *For a horizon $H \in \mathbb{N}$, a RH NE is a policy ensemble $(\pi_t^{*,H})_{t\in\mathcal{T}} \in \Pi^T$ such that for all $t \in \mathcal{T}$*

$$\pi_t^{*,H} \in \arg\max_{\pi\in\Pi} J_t^H(\pi, \pi_t^{*,H}), \text{ with } \mu_t^H = \Gamma_{\mathcal{M},t}^H\left(\pi_t^{*,H}\right),$$

where the initial MF for each MFG is the MF of the previous MFG after one time step, i.e. $\mu_{t,t}^H = \mu_{t-1,t}^H$ for all $t > 0$ and $\mu_{0,0}^H = \mu_0$. $(J_t^H)_{t\in\mathcal{T}}$ is the corresponding RH objective ensemble. Since MFs may deviate in practice and the horizon moves forward by one after each time step t, agents only implement the first entry $\pi_{t,t}^{*,H}$ of each policy. Thus, the implemented equilibrium policy $\pi^{**,H} \in \Pi$ results from the policy ensemble $(\pi_t^{*,H})_{t\in\mathcal{T}} \in \Pi^T$ by taking

$$\pi_t^{**,H} = \pi_{t,t}^{*,H},$$

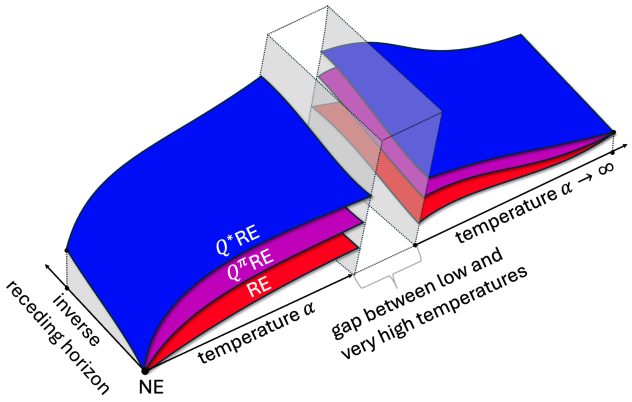

Figure 1: A visualization of the resulting equilibrium policies (one-dimensional for illustration) of QRE and RE over temperature and receding horizon. In the limit of low temperature and infinite horizon, all concepts become the MFNE. In the limit of infinite temperature, all solutions become the constant uniform policy.

for each $t \in \mathcal{T}$, i.e., the diagonal $\pi^{**,H} = \operatorname{diag}((\pi_t^{*,H})_{t\in\mathcal{T}})$.

The RH concept extends to regularized equilibria by changing the corresponding objective. Accordingly, we define an approximate ϵ-RH RE such that for any t, $\pi_t^{*,H}$ is ϵ-optimal instead of exactly maximizing $\pi \mapsto J_t^H(\pi, \pi_t^{*,H})$, and we define their exploitability as the sum of exploitabilities in each sub-MFG at time t. In a similar fashion RH QRE are defined in Eich et al. (2024, Appendix D).

Connections between Equilibrium Notions

In the following, we compare the notions of equilibria introduced in the prequel. In all mentioned equilibria, π^* can be written down using the equivalent fixed-point equations

$$\text{(NE)} \quad \pi^* = \Gamma_\Pi^*(\Gamma_{Q^*}(\Gamma_{\mathcal{M}}(\pi^*))), \tag{7a}$$

$$\text{(Q}^\pi\text{RE)} \quad \pi^* = \Gamma_\Pi^{1/\lambda}(\Gamma_{Q^\pi}(\Gamma_{\mathcal{M}}(\pi^*), \pi^*)), \tag{7b}$$

$$\text{(Q}^*\text{RE)} \quad \pi^* = \Gamma_\Pi^{1/\lambda}(\Gamma_{Q^*}(\Gamma_{\mathcal{M}}(\pi^*))), \tag{7c}$$

$$\text{(RE)} \quad \pi^* = \Gamma_\Pi^{\alpha}(\Gamma_{\tilde{Q}}^{\alpha}(\Gamma_{\mathcal{M}}(\pi^*))). \tag{7d}$$

The decomposed definition enables a comparison between different equilibria. If we choose $\alpha = 1/\lambda$, the Q$^\pi$RE, Q*RE and RE policies are all of α-softmax form. Therefore, although the equilibria have various differing derivations, their special cases considered here are connected.

Distinctiveness of equilibrium concepts. In general, the equilibrium notions are distinct. Here, we look at the special cases of entropy-regularized RE and Q$^\pi$RE / Q*RE with Gumbel noise (LQRE / BE) where the difference between Q$^\pi$RE, Q*RE and RE is the usage of *policy*, *optimal* and *soft* state-action value functions respectively. As $\alpha \to 0$, RE are known to essentially become NE in the sense of exploitability in the finite system (Cui and Koeppl 2021). Furthermore, as $1/\lambda \to 0$, the Gumbel noise vanishes in Q$^\pi$RE / Q*RE. Thus, in the low temperature limit, all equilibrium notions are equivalent which is visualized in Figure 1. On the other hand,

as $\alpha = 1/\lambda \to \infty$, by Assm. 1, each solution tends to the uniform policy. However, for intermediate temperatures α, the equilibria are distinct, see Figure 1. Analogously, for RH equilibria we have convergence to standard non-RH equilibria as the horizon becomes large, $H \to \infty$.

Q$^\pi$RE is a first order approximation of RE. Although Q$^\pi$RE and RE are distinct, we find a connection between both through a principled approximation. Indeed, let $\alpha = 1/\lambda$, then the Q$^\pi$RE is obtained by solving a recursive first order approximation of the soft state-action value function.

Theorem 1. *Q$^\pi$RE are obtained from RE by first-order approximation of the smooth-maximum Bellman equation* (6)*.*

Proof of Thm. 1. Assume w.l.o.g. that $|\mathcal{U}| = n$ for some finite $n \in \mathbb{N}$ and define the function $g\colon \mathbb{R}^n \to \mathbb{R}$, $g(x_1, \dots, x_n) = \alpha \log\left(\sum_{i=1}^n \exp\left(x_i/\alpha\right)\right)$ such that

$$\frac{\partial}{\partial x_j} g(x_1, \dots, x_n) = \frac{\exp\left(x_j/\alpha\right)}{\sum_{i \le n} \exp\left(x_i/\alpha\right)}\,.$$

Then, the smooth-maximum Bellman recursion in the context of regularized equilibria can be rewritten as

$$\tilde{Q}_t^{\mu,\alpha}(x,u) = r(x,u,\mu_t) + \sum_{x' \in \mathcal{X}} p_t(x' \mid x,u,\mu_t) \cdot g\left(\tilde{Q}_{t+1,\alpha}^{\mu}(x',\cdot)\right)\,.$$

Similarly, the Bellman equation in the Q$^\pi$RE setup with Gumbel noise and $\lambda = 1/\alpha$ can be expressed as

$$Q_t^{\mu,*}(x,u) = r(x,u,\mu_t) + \sum_{x' \in \mathcal{X}} p_t(x' \mid x,u,\mu_t) \left\langle \nabla g(Q_{t+1}^{\mu,*}(x',\cdot)), Q_{t+1}^{\mu,*}(x',\cdot)\right\rangle$$

by using $\pi_t^*(\cdot|x) = \nabla g(Q_t^{\mu,*}(x,\cdot))$, and $g(\mathbf{0}) = \alpha \log n$. Thus, we interpret the Q$^\pi$RE Bellman equation as a first order approximation of the smooth-maximum Bellman recursion, with added error term $\alpha \log n \to 0$ as $\alpha \to 0$. Here, we recursively estimate $\tilde{Q}$ through Q^* backwards in $t \in \mathcal{T}$. □

Thm. 1 establishes a rigorous connection between Q$^\pi$RE and RE, in the common case of Gumbel noise and entropy regularization. While both concepts share similarities, our empirical evaluations demonstrate that they can yield different results in general, see the experiments section.

Learning Non-Cooperative MF Equilibria

There is a variety of methods for computing or learning NE in MFGs, each with its own limitations. Standard fixed-point iteration (FPI) is not guaranteed to converge, as it cannot be Lipschitz even in simple standard finite MFGs (Cui and Koeppl 2021, Thm. 2). Meanwhile, other algorithms such as fictitious play (FP) and Online Mirror Descent require monotonicity and their theory is currently limited to dynamics independent of the mean field (Perrin et al. 2020; Pérolat et al. 2022). Other recent ideas include an optimization-based approach (Guo, Hu, and Zhang 2024) and a regularization approach (Guo et al. 2019; Cui and Koeppl 2021), for which convergence via FPI is guaranteed for strong enough regularization. In the following, we generalize FPI and FP to our equilibrium concepts of interest.

Algorithm 1: Generalized Fixed-Point Iteration (GFPI).

1: Input: Temperature $\alpha = 1/\lambda > 0$, initial policy π^0, equilibrium type ET $\in \{\text{NE}, \text{Q}^\pi\text{RE}, \text{Q}^*\text{RE}, \text{RE}\}$.
2: Define Γ_Π and Γ_Q according to ET (Eq. 7).
3: **for** $k = 0, 1, \dots,$K-1 **do**
4: Evaluate $\pi^{k+1} \leftarrow \Gamma_\Pi(\Gamma_Q(\Gamma_{\mathcal{M}}(\pi^k), \pi^k))$.
5: **end for**
6: **return** π^K

Fixed point iteration. In the FPI approach, one repeatedly computes the result of the right-hand side of the fixed point equations (7). Start with some initial policy, e.g., $\pi_t^0(u \mid x) = 1/|\mathcal{U}|$ for all t, x, u. Then, for each iteration $k = 0, 1, \dots$, compute the resulting MF by solving the fixed point equation, the resulting value functions under the new MF, and finally the new policy, e.g. for RE as

$$\pi^{k+1} = \Gamma_\Pi^\alpha(\Gamma_{\tilde{Q}}^\alpha(\Gamma_{\mathcal{M}}(\pi^k)))\,.$$

For NE, Q$^\pi$RE and Q*RE the operators are changed according to the desired setting as in Eqs. (7). The overall generalized FPI (GFPI) algorithm is given in Alg. 1 and converges for sufficiently high temperatures under Lipschitz conditions.

Assumption 2. *The transition kernel P and reward function r are Lipschitz continuous in their MF argument.*

Proposition 3. *In MFGs, under Assm. 2, FPI converges to a Q$^\pi$RE / Q*RE / RE for sufficiently high $\alpha > 0$.*

Such a result is known (Cui and Koeppl 2021) for RE and BE (Q*RE) and extended towards Q$^\pi$RE here. As a result, the convergence of FPI follows for high temperatures.

Corollary 1. *By Banach's fixed-point theorem, regularized FPI converges to an equilibrium for sufficiently large $\alpha > 0$.*

Requiring sufficiently large α limits applicability of the GFPI algorithm. Therefore, alternate or more general learning algorithms are desired.

Fictitious play. One such algorithm for standard NE is the FP algorithm. Parallel to GFPI, we formulate the generalized FP (GFP) algorithm in Alg. 2 for learning MFG equilibria.

Algorithm 2: Generalized Fictitious Play (GFP).

1: Input: Temperature $\alpha > 0$, policy π^0, $\beta \in (0,1)$, equilibrium type ET $\in \{\text{NE}, \text{Q}^\pi\text{RE}, \text{Q}^*\text{RE}, \text{RE}\}$.
2: Define Γ_Π and Γ_Q according to ET (Eq. 7).
3: Initialize $\mu^0 \leftarrow \Gamma_{\mathcal{M}}(\pi^0)$ as the MF induced by π^0.
4: Initialize weighted sum of policies $\bar{\pi}^0 = \pi^0 \mu^0$
5: **for** $k = 0, 1, \dots, K-1$ **do**
6: Evaluate $\pi^{k+1} \leftarrow \Gamma_\Pi(\Gamma_Q(\mu^k, \pi^k))$.
7: Compute MF $\mu^{k+1} \leftarrow \Gamma_{\mathcal{M}}(\pi^{k+1})$ induced by π^{k+1}.
8: Compute $\bar{\pi}^{k+1} = \beta\bar{\pi}^{k+1} + (1-\beta)\mu^{k+1}\pi^{k+1}$.
9: Average MF $\mu^{k+1} \leftarrow (1-\beta)\mu^{k+1} + \beta\mu^k$.
10: Normalize $\pi^{k+1} \propto \bar{\pi}^{k+1}$.
11: **end for**
12: **return** π^K.

For NE, the algorithm matches with the proven FP algorithm in Perrin et al. (2020), while for BE, RE the algorithm matches with initial experiments in Cui and Koeppl (2021). Thus, GFP is known to converge for NE under certain assumptions (Perrin et al. 2020). We extend the FP convergence proof in standard MFGs towards RE, and also towards RH equilibria by proposing suitable modifications, see the red plane in Fig. 1. As a side result, we make the existing proof for FP convergence more precise by explicitly verifying the usage of the envelope theorem, which is strictly speaking only applicable for any non-zero regularization, $\alpha > 0$. The proofs for the remaining cases are left to future work. The proof requires a standard monotonicity assumption on the considered MFG, as well as a common simplifying assumption on the system dynamics and is based on a continuous-time ODE version of the algorithm with time-continuous iterations as in Perrin et al. (2020). Then, the actual algorithm can be interpreted as a discretization of the continuous-time version.

Assumption 3 (Lasry and Lions (2007))**.** *The game is monotone, i.e.* $R_t(x,u,\mu) = r_t(x,\mu) + \bar{r}_t(x,u)$ *with differentiable* r_t *and* $\sum_{x\in\mathcal{X}} (\mu(x) - \mu'(x))\,(r_t(x,\mu) - r_t(x,\mu')) \leq 0$ *for all* $t \in \mathcal{T}$ *and all* $\mu, \mu' \in \mathcal{P}(\mathcal{X})$.

Assumption 4 (Perrin et al. (2020); Pérolat et al. (2022))**.** *The transition kernel does not depend on the MF.*

Theorem 2. *Under Assm. 3 and 4, the continuous-time version of GFP for RE converges to zero exploitability,* $\mathcal{E}^{\mathrm{RE}}(\bar{\pi}^\tau) := \max_{\pi'} J_\alpha^{\mathrm{RE}}(\pi', \bar{\pi}^\tau) - J_\alpha^{\mathrm{RE}}(\bar{\pi}^\tau, \bar{\pi}^\tau) \to 0$ *at rate* $\mathcal{O}(\frac{1}{\tau})$*, with algorithm run time* τ.

Computation of RH equilibria. RH equilibria can be computed in a sequential manner by applying the GFPI or GFP algorithm for the arising MFGs at each time step. The result of each algorithm yields the initial condition for the next MFG, which starts with the previous MF after one time step. We summarize the sequential RH-GFP variant in Alg. 3.

Theorem 3. *Under Assm. 3 and 4, the continuous-time versions of sequential RH-GFP for RE converges to zero exploitability as* $\tau \to \infty$.

Algorithm 3: Sequential RH-GFP.

1: Input: Temperature $\alpha > 0$, policy π^0, $\beta \in (0,1)$, equilibrium type ET $\in \{\mathrm{NE}, \mathrm{Q}^\pi\mathrm{RE}, \mathrm{Q}^*\mathrm{RE}, \mathrm{RE}\}$, receding horizon H.
2: Define Γ_Π and Γ_Q according to ET (Eq. 7) with horizon H.
3: **for** $t = 0, 1, \ldots, T-1$ **do**
4: **if** $t = 0$ **then** $\mu_{t,0}^0 = \mu_0$; **else** $\mu_{t,0}^0 = \mu_{t-1,1}^K$;
5: Initialize $\mu_t^0 \leftarrow \Gamma_{\mathcal{M}}(\pi^0)$ as the MF induced by π^0 with initial MF $\mu_{t,0}^0$.
6: Initialize weighted sum of policies $\bar{\pi}_t^0 = \pi^0 \mu_t^0$.
7: **for** $k = 0, 1, \ldots, K-1$ **do**
8: Evaluate lines 6-10 of Alg. 2.
9: **end for**
10: **end for**
11: **return** π_t^K for $t = 0, 1, \ldots, T-1$.

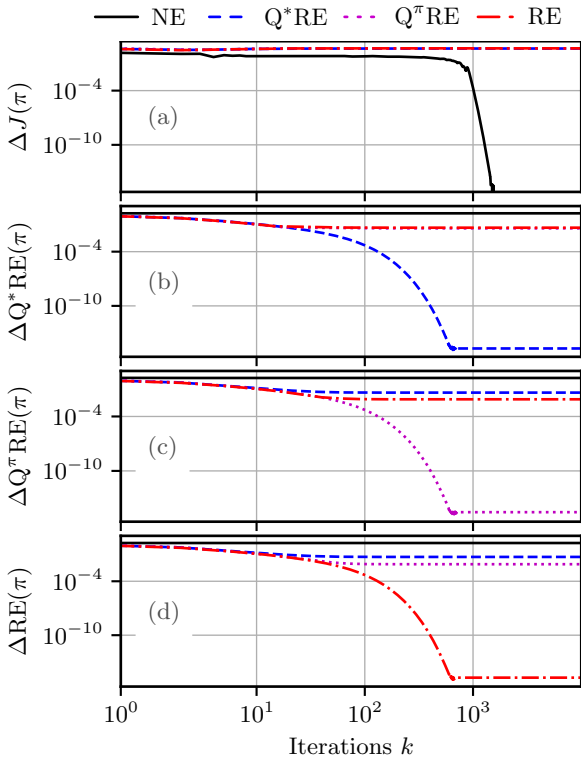

Figure 2: Convergence of GFP for the Susceptible-Infectious-Susceptible MFG with $\alpha = 1.0$ and $\beta = 0.95$. The GFP algorithms for Q^πRE, Q^*RE and RE show similar behaviour in the first iterations before converging to their respective equilibria.

This sequential approach, however, can be inefficient, especially for long horizons, since every MFG has to wait until the previous MFG converged. To circumvent this, we propose a parallel algorithm, where in each iteration for each MFG we change the initial condition to the previous MF after one time step. This parallelization is efficient, since the later MFGs start learning equilibria before their initial condition has converged. The parallel RH-GFP algorithm is summarized in Eich et al. (2024, Appendix F), where also experiments are provided, which highlight the efficiency compared to the sequential approach.

Experiments

In this section we evaluate our algorithms and analyze the different equilibria for several MFGs. We analyze the efficacy of our methods for a Susceptible-Infectious-Susceptible (SIS) problem and a sequential version of Rock-Paper-Scissor (RPS) game. Additionally, we evaluate random MFGs, similar to Pérolat et al. (2022), by creating random transition and reward matrices and adding a mean-field dependent function to the reward that promotes swarm avoiding behaviour. Detailed game descriptions are found in Eich et al. (2024, Appendix G). For code, see https://github.com/yannickeich/QRE-MFG.

To measure algorithm efficiency, we quantify the distance

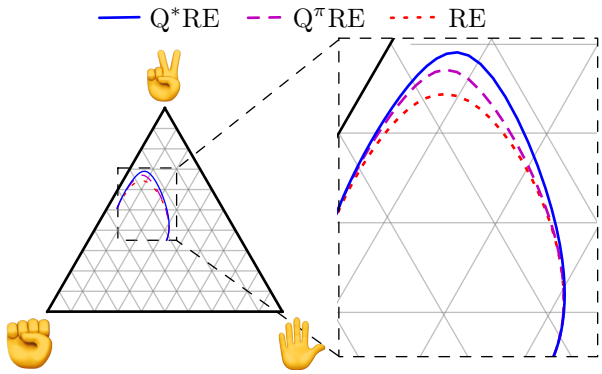

Figure 3: Action probabilities in the Rock-Paper-Scissor problem at $t = 0$ for the resulting Q*RE / Q$^\pi$RE / RE using GFP (Alg. 2) over various temperatures $\alpha = 1/\lambda$. As $\alpha \to \infty$, we always obtain the uniform policy in the center, while as $\alpha \to 0$, solutions converge to the Nash solution (to the left). In-between, solutions differ from each other, regardless of the temperature.

to a Q$^\pi$RE / Q*RE / RE, similar to exploitability in the NE case.

Definition 10 (Distance to equilibria)**.** *The distance of a policy $\pi \in \Pi$ to a Q$^\pi$RE / Q*RE / RE is defined as*

$$\Delta Q^\pi RE(\pi) \coloneqq \max_{t \in \mathcal{T}} \left\| \pi_t - \Gamma_\Pi^{1/\lambda}(\Gamma_{Q^\pi}(\Gamma_\mathcal{M}(\pi), \pi))_t \right\|,$$

$$\Delta Q^* RE(\pi) \coloneqq \max_{t \in \mathcal{T}} \left\| \pi_t - \Gamma_\Pi^{1/\lambda}(\Gamma_{Q^*}(\Gamma_\mathcal{M}(\pi)))_t \right\|,$$

$$\Delta \mathrm{RE}(\pi) \coloneqq \max_{t \in \mathcal{T}} \left\| \pi_t - \Gamma_\Pi^{\alpha}(\Gamma_{\tilde{Q}}^{\alpha}(\Gamma_\mathcal{M}(\pi)))_t \right\|.$$

Note that whenever ΔQ$^\pi$RE(π), ΔQ*RE(π) or ΔRE(π) is zero for a policy π, π is a Q$^\pi$RE / Q*RE / RE of the MFG.

First, we employ the GFP algorithm to compute Q$^\pi$RE, Q*RE and RE for the SIS MFG. Figure 2 illustrates the progress of the algorithms by displaying $\Delta J(\pi)$, ΔQ$^\pi$RE(π), ΔQ*RE(π) and ΔRE(π) over the iterations k. The results display the efficacy of the GFP algorithm, showing fast convergence to the equilibria. The exploitability plot highlights the bounded rationality of the other equilibria compared to the NE. In exchange for converging to their respective equilibria, GFP for Q$^\pi$RE, Q*RE and RE does not lead to zero $\Delta J(\pi)$. Additionally, Figure 2 highlights both the similarities and differences of Q$^\pi$RE, Q*RE and RE. Their respective algorithms behave similarly in the first iterations before leading to the distinct equilibria.

To further visualize the distinctiveness of the equilibrium concepts, we compute Q$^\pi$RE, Q*RE and RE with different temperatures $\alpha = 1/\lambda$ for the sequential RPS MFG using the GFP algorithm. We indicate the policies of the first time step of the resulting equilibria in the probability simplex in Figure 3. The comparison experimentally verifies our previous discussion of the connections between the equilibrium notions. The results highlight both the equivalence of the equilibrium concepts for $\alpha \to 0$ (representing the NE) and $\alpha \to \infty$ (representing the uniform policy) and their distictiveness for intermediate temperatures.

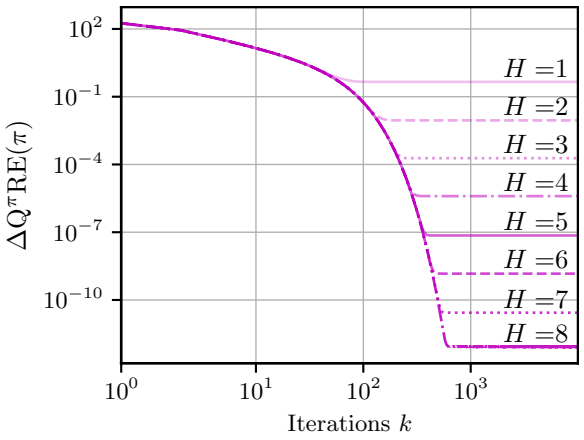

Figure 4: Comparison of the distance of various RH Q$^\pi$RE with different horizons H to the Q$^\pi$RE with total horizon $\mathcal{T}$ for a random MFG with $\alpha = 1.0$ and $\beta = 0.95$ over iterations k. The equilibria induced by assuming shorter lookahead capacities deviate more from the QRE with total horizon, demonstrating the impact of limited lookahead on equilibrium behavior.

Finally, we examine how combining Q$^\pi$RE with RH equilibria can effectively model agents with limited lookahead capacities. To do this, we employ the parallel RH-GFP algorithm to compute RH QRE of a random MFG for different receding horizons H. Figure 4 illustrates the distance of the various RH QRE to the Q$^\pi$RE with time horizon $\mathcal{T}$ over the iterations k. The results demonstrate that RH QRE with higher lookahead are closer to the Q$^\pi$RE. Conversely, assuming shorter lookahead capacities results in equilibria that deviate more from the Q$^\pi$RE, highlighting the efficacy of modeling bounded rationality with receding horizon.

For additional experiments that demonstrate the convergence of our proposed algorithms and provide further insights into the connections between the different equilibrium concepts, see Eich et al. (2024, Appendix H).

Conclusion

In this work, we introduced both QRE MFGs and RH MFGs, to incorporate bounded rationality in MFGs for a more realistic modeling of large agent populations. We compared these new equilbrium concepts to existing ones theoretically and empirically. Our analysis highlights the similarities and differences of the discussed equilibrium concepts. Furthermore, we designed general learning algorithms to compute equilibria efficiently and evaluated these algorithms on different problem settings. We hope that the novel RH and QRE MFGs combined with our learning algorithms help bring the theory closer to real-world scenarios with not perfectly rational agents. For future work, one could apply our theory and learning approach to research problems where bounded rationality is crucial, e.g. in economics or the social sciences.

Acknowledgements

This work has been co-funded by the LOEWE emergenCITY research promotion program of the federal state of Hessen, Germany, by the German Research Foundation (DFG) within the Collaborative Research Center (CRC) 1053 MAKI and project number 517777863, by the Federal Ministry of Education and Research as part of the Software Campus project RL4MFRP (funding code 01IS23067) and by the Hessian Ministry of Science and the Arts (HMWK) within the projects "The Third Wave of Artificial Intelligence - 3AI" and hessian.AI.

References

Achdou, Y.; Cardaliaguet, P.; Delarue, F.; Porretta, A.; Santambrogio, F.; Achdou, Y.; and Laurière, M. 2020. Mean field games and applications: Numerical aspects. *Mean Field Games: Cetraro, Italy 2019*, 249–307.

Anahtarci, B.; Kariksiz, C. D.; and Saldi, N. 2023. Q-learning in regularized mean-field games. *Dynamic Games and Applications*, 13(1): 89–117.

Belousov, B.; and Peters, J. 2019. Entropic regularization of Markov decision processes. *Entropy*, 21(7): 674.

Breitmoser, Y.; Tan, J. H.; and Zizzo, D. J. 2010. Understanding perpetual R&D races. *Economic Theory*, 44(3): 445–467.

Campi, L.; and Fischer, M. 2022. Correlated equilibria and mean field games: a simple model. *Mathematics of Operations Research*, 47(3): 2240–2259.

Carmona, R. 2020. Applications of mean field games in financial engineering and economic theory. *arXiv preprint arXiv:2012.05237*.

Carmona, R.; Dayanıklı, G.; and Laurière, M. 2022. Mean field models to regulate carbon emissions in electricity production. *Dynamic Games and Applications*, 12(3): 897–928.

Carmona, R.; and Wang, P. 2021. Finite-state contract theory with a principal and a field of agents. *Management Science*, 67(8): 4725–4741.

Cui, K.; and Koeppl, H. 2021. Approximately solving mean field games via entropy-regularized deep reinforcement learning. In *International Conference on Artificial Intelligence and Statistics*, 1909–1917. PMLR.

Daskalakis, C.; Goldberg, P. W.; and Papadimitriou, C. H. 2009. The complexity of computing a Nash equilibrium. *Communications of the ACM*, 52(2): 89–97.

Deng, X.; Li, N.; Mguni, D.; Wang, J.; and Yang, Y. 2023. On the complexity of computing Markov perfect equilibrium in general-sum stochastic games. *National Science Review*, 10(1): nwac256.

Djehiche, B.; Tcheukam, A.; and Tembine, H. 2017. Mean-Field-Type Games in Engineering. *AIMS Electronics and Electrical Engineering*, 1(1): 18–73.

Eibelshäuser, S.; and Poensgen, D. 2019. Markov Quantal Response Equilibrium and a Homotopy Method for Computing and Selecting Markov Perfect Equilibria of Dynamic Stochastic Games. VfS Annual Conference 2019 (Leipzig): 30 Years after the Fall of the Berlin Wall - Democracy and Market Economy 203603, Verein für Socialpolitik / German Economic Association.

Eich, Y.; Fabian, C.; Cui, K.; and Koeppl, H. 2024. Bounded Rationality Equilibrium Learning in Mean Field Games. arXiv:2411.07099.

Elie, R.; Mastrolia, T.; and Possamaï, D. 2019. A tale of a principal and many, many agents. *Mathematics of Operations Research*, 44(2): 440–467.

Eysenbach, B.; and Levine, S. 2021. Maximum Entropy RL (Provably) Solves Some Robust RL Problems. In *International Conference on Learning Representations*.

Fudenberg, D.; and Tirole, J. 1991. *Game theory*. MIT press.

Geist, M.; Scherrer, B.; and Pietquin, O. 2019. A theory of regularized Markov decision processes. In *International Conference on Machine Learning*, 2160–2169. PMLR.

Gemp, I.; Marris, L.; and Piliouras, G. 2024. Approximating Nash Equilibria in Normal-Form Games via Stochastic Optimization. In *The Twelfth International Conference on Learning Representations*.

Gigerenzer, G.; and Selten, R. 2002. *Bounded rationality: The adaptive toolbox*. MIT press.

Guo, X.; Hu, A.; Xu, R.; and Zhang, J. 2019. Learning mean-field games. In *Advances in Neural Information Processing Systems*, 4966–4976.

Guo, X.; Hu, A.; and Zhang, J. 2024. MF-OMO: An optimization formulation of mean-field games. *SIAM Journal on Control and Optimization*, 62(1): 243–270.

Guo, X.; Xu, R.; and Zariphopoulou, T. 2022. Entropy regularization for mean field games with learning. *Mathematics of Operations Research*, 47(4): 3239–3260.

Inoue, D.; Ito, Y.; Kashiwabara, T.; Saito, N.; and Yoshida, H. 2021. Model Predictive Mean Field Games for Controlling Multi-Agent Systems. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 982–987. IEEE.

Kahneman, D. 2013. A perspective on judgment and choice: Mapping bounded rationality. *Progress in Psychological Science around the World. Volume 1 Neural, Cognitive and Developmental Issues.*, 1–47.

Kahneman, D.; and Tversky, A. 1982. The psychology of preferences. *Scientific American*, 246(1): 160–173.

Kouvaritakis, B.; and Cannon, M. 2016. Model Predictive Control. *Switzerland: Springer International Publishing*, 38: 13–56.

Lasry, J.-M.; and Lions, P.-L. 2007. Mean field games. *Japanese journal of mathematics*, 2(1): 229–260.

Laurière, M.; Perrin, S.; Geist, M.; and Pietquin, O. 2022. Learning mean field games: A survey. *arXiv preprint arXiv:2205.12944*.

Li, P.; Yu, R.; Wang, X.; and An, B. 2024. Transition-Informed Reinforcement Learning for Large-Scale Stackelberg Mean-Field Games. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

McKelvey, R. D.; and Palfrey, T. R. 1995. Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior*, 10(1): 6–38.

McKelvey, R. D.; and Palfrey, T. R. 1998. Quantal response equilibria for extensive form games. *Experimental Economics*, 1: 9–41.

Muller, P.; Elie, R.; Rowland, M.; Lauriere, M.; Perolat, J.; Perrin, S.; Geist, M.; Piliouras, G.; Pietquin, O.; and Tuyls, K. 2022a. Learning Correlated Equilibria in Mean-Field Games. *arXiv preprint arXiv:2208.10138*.

Muller, P.; Rowland, M.; Elie, R.; Piliouras, G.; Perolat, J.; Lauriere, M.; Marinier, R.; Pietquin, O.; and Tuyls, K. 2022b. Learning Equilibria in Mean-Field Games: Introducing Mean-Field PSRO. In *Proc. AAMAS*, 926–934.

Pérolat, J.; Perrin, S.; Elie, R.; Laurière, M.; Piliouras, G.; Geist, M.; Tuyls, K.; and Pietquin, O. 2022. Scaling Mean Field Games by Online Mirror Descent. In *Proc. AAMAS*, 1028–1037.

Perrin, S.; Pérolat, J.; Laurière, M.; Geist, M.; Elie, R.; and Pietquin, O. 2020. Fictitious play for mean field games: Continuous time analysis and applications. In *Advances in Neural Information Processing Systems*, volume 33, 13199–13213.

Reddi, A.; Tölle, M.; Peters, J.; Chalvatzaki, G.; and D'Eramo, C. 2024. Robust Adversarial Reinforcement Learning via Bounded Rationality Curricula. In *The Twelfth International Conference on Learning Representations*.

Saldi, N.; Basar, T.; and Raginsky, M. 2018. Markov–Nash Equilibria in Mean-Field Games with Discounted Cost. *SIAM Journal on Control and Optimization*, 56(6): 4256–4287.

Selten, R. 1990. Bounded rationality. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 146(4): 649–658.

Simon, H. A. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 99–118.

Simon, H. A. 1979. Rational decision making in business organizations. *The American Economic Review*, 69(4): 493–513.

Vasal, D.; and Berry, R. 2022. Master Equation for Discrete-Time Stackelberg Mean Field Games with a Single Leader. In *Proc. CDC*, 5529–5535. IEEE.

Xie, Q.; Yang, Z.; Wang, Z.; and Minca, A. 2021. Learning while playing in mean-field games: Convergence and optimality. In *International Conference on Machine Learning*, 11436–11447. PMLR.