Machine learning methods for mean field games and mean field control problems

Mathieu LAURIÈRE

NYU 上海 SHANGHAI 纽约大学

August 11, 2022
CIRM, Marseille CEMRACS 2022

Main questions for this talk

Q1: *How can we solve large games with complex structures?*

Part 1: Solving mean-field problems with deep learning

Q2: *How can large populations learn to coordinate?*

Part 2: Reinforcement learning with mean-field interactions

Outline

Outline

Mean Field Paradigm: number of agents $N \to \infty$

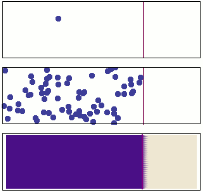
Main question: *How do global outcomes emerge from individual decisions?*

Mean Field Paradigm: number of agents $N \to \infty$

Main question: *How do global outcomes emerge from individual decisions?*
Large population $\Rightarrow$ individual interactions are **intractable**

Mean Field Paradigm: number of agents $N \to \infty$

Main question: *How do global outcomes emerge from individual decisions?*
Large population $\Rightarrow$ individual interactions are **intractable**
Assumption: perfect ***homogeneity*** & ***symmetry*** of the agents



Mean Field Paradigm: number of agents $N \to \infty$

Main question: *How do global outcomes emerge from individual decisions?*
Large population $\Rightarrow$ individual interactions are **intractable**
Assumption: perfect ***homogeneity*** & ***symmetry*** of the agents
Mean Field in statistical mechanics: particles (micro) $\to$ density function (macro)

Mean Field Paradigm: number of agents $N \to \infty$

Main question: *How do global outcomes emerge from individual decisions?*
Large population $\Rightarrow$ individual interactions are **intractable**
Assumption: perfect ***homogeneity*** & ***symmetry*** of the agents
Mean Field in statistical mechanics: particles (micro) $\to$ density function (macro)

Mean Field Paradigm: number of agents $N \to \infty$



Main question: *How do global outcomes emerge from individual decisions?*
Large population $\Rightarrow$ individual interactions are **intractable**
Assumption: perfect ***homogeneity*** & ***symmetry*** of the agents
Mean Field in statistical mechanics: particles (micro) $\to$ density function (macro)

Mix with optimization:

- mean field control: infinitely many cooperating agents
- mean field game: infinitely many competing players

Landscape of Research on MFG

Initiated by Lasry and Lions, and Huang *et al.* around 2006

Main research directions:

• **Modeling**: crowd motion, econ./finance, flocking, risk management, smart grid, energy production, distributed robotics, epidemic, . . .

Landscape of Research on MFG

Initiated by Lasry and Lions, and Huang *et al.* around 2006

Main research directions:

• **Modeling**: crowd motion, econ./finance, flocking, risk management, smart grid, energy production, distributed robotics, epidemic, . . .

• **Mean field** approach justification:

$\diamond$ N-agent problem $\rightarrow$ mean field: convergence

$\diamond$ N-agent problem $\leftarrow$ mean field: ϵ-optimality

Landscape of Research on MFG

Initiated by Lasry and Lions, and Huang *et al.* around 2006

Main research directions:

• **Modeling**: crowd motion, econ./finance, flocking, risk management, smart grid, energy production, distributed robotics, epidemic, . . .

• **Mean field** approach justification:

◇ N-agent problem $\rightarrow$ mean field: convergence

◇ N-agent problem $\leftarrow$ mean field: ϵ-optimality

• **Characterization** of the mean field problem solutions (**optimality conditions**):

◇ partial differential equations (PDE system)

◇ stochastic differential equations (SDE system)

◇ Master equation (PDE on Wasserstein space)

Landscape of Research on MFG

Initiated by Lasry and Lions, and Huang *et al.* around 2006

Main research directions:

- **Modeling**: crowd motion, econ./finance, flocking, risk management, smart grid, energy production, distributed robotics, epidemic, . . .
- **Mean field** approach justification:
  - ◇ N-agent problem $\rightarrow$ mean field: convergence
  - ◇ N-agent problem $\leftarrow$ mean field: ϵ-optimality
- **Characterization** of the mean field problem solutions (**optimality conditions**):
  - ◇ partial differential equations (PDE system)
  - ◇ stochastic differential equations (SDE system)
  - ◇ Master equation (PDE on Wasserstein space)
- **Computation** of solutions
  - ◇ "solving" numerically = *What is the optimal behavior?* (control rule & density flow)
  - ◇ crucial for applications
  - ◇ challenge: coupling between optimization & mean-field

Multi-Agent Control Problem

Assume there are N **identical** agents (***homogeneity***)

Agent i uses control $v^i(t, X_t^1, \ldots, X_t^N) \in \mathbb{R}^d$ and has state $X_t^i \in \mathbb{R}^d$ at time t, with

Multi-Agent Control Problem

Assume there are N **identical** agents (***homogeneity***)

Agent i uses control $v^i(t, X^1_t, \dots, X^N_t) \in \mathbb{R}^d$ and has state $X^i_t \in \mathbb{R}^d$ at time t, with

- initial position: $X^i_0 \sim m_0$
- and dynamics: $\underbrace{dX^i_t}_{\text{variation of position}} = \underbrace{v^i(t, \overbrace{X^1_t, \dots, X^N_t}^{\mathbf{x}_t})}_{\text{velocity}} \, dt + \underbrace{dW^i_t}_{\text{noise}} \quad \Big(+ \underbrace{dB_t}_{\text{(common noise)}} \Big)$

Multi-Agent Control Problem

Assume there are N **identical** agents (***homogeneity***)

Agent i uses control $v^i(t, X_t^1, \dots, X_t^N) \in \mathbb{R}^d$ and has state $X_t^i \in \mathbb{R}^d$ at time t, with

- initial position: $X_0^i \sim m_0$
- and dynamics:

$$\underbrace{dX_t^i}_{\substack{\text{variation}\\ \text{of position}}} = \underbrace{v^i(t, \overbrace{X_t^1, \dots, X_t^N}^{\mathbf{X}_t})}_{\text{velocity}}\, dt + \underbrace{dW_t^i}_{\text{noise}} \qquad \Big(+ \underbrace{dB_t}_{\text{(common noise)}}\Big)$$

Agent i pays running cost $f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right)$ typically increasing w.r.t. $\left(\mu_t^N, v_t^i\right)$ where the interaction is of **mean-field** type ***(symmetry)*** since it occurs only through

$$\mu_t^N = \frac{1}{N} \sum_{j=1}^N \delta_{X_t^j}$$

which is the **empirical distribution** of the agents' states ($\delta_x =$ Dirac mass at x)

Multi-Agent Control Problem

Assume there are N **identical** agents (***homogeneity***)

Agent i uses control $v^i(t, X_t^1, \dots, X_t^N) \in \mathbb{R}^d$ and has state $X_t^i \in \mathbb{R}^d$ at time t, with

- initial position: $\quad X_0^i \sim m_0$
- and dynamics:

$$\underbrace{dX_t^i}_{\substack{\text{variation}\\ \text{of position}}} = \underbrace{v^i(t, \overbrace{X_t^1, \dots, X_t^N}^{\mathbf{X}_t})}_{\text{velocity}}\, dt + \underbrace{dW_t^i}_{\text{noise}} \qquad \Big(+ \underbrace{dB_t}_{\text{(common noise)}} \Big)$$

Agent i pays running cost $f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right)$ typically increasing w.r.t. $\left(\mu_t^N, v_t^i\right)$ where the interaction is of **mean-field** type ***(symmetry)*** since it occurs only through

$$\mu_t^N = \frac{1}{N} \sum_{j=1}^N \delta_{X_t^j}$$

which is the **empirical distribution** of the agents' states ($\delta_x =$ Dirac mass at x)

The **social cost** is the average of all the individual costs:

$$J^N(v^1, \dots, v^N) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[\int_0^T \underbrace{f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right)}_{\text{running cost}} dt + \underbrace{g\left(X_T^i\right)}_{\text{terminal cost}} \right]$$

Goal: Find an optimal $\hat{v} = (\hat{v}^1, \dots, \hat{v}^N)$ **minimizing** J^N

Multi-Agent Control Problem

Assume there are N **identical** agents (***homogeneity***)

Agent i uses control $v^i(t, X_t^1, \ldots, X_t^N) \in \mathbb{R}^d$ and has state $X_t^i \in \mathbb{R}^d$ at time t, with

- initial position: $X_0^i \sim m_0$
- and dynamics: $$\underbrace{dX_t^i}_{\substack{\text{variation}\\ \text{of position}}} = \underbrace{v^i(t, \overbrace{X_t^1, \ldots, X_t^N}^{\mathbf{X}_t})}_{\text{velocity}}\, dt + \underbrace{dW_t^i}_{\text{noise}} \quad \underbrace{\left(+ \underbrace{dB_t}\right)}_{\text{(common noise)}}$$

Agent i pays running cost $f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right)$ typically increasing w.r.t. $\left(\mu_t^N, v_t^i\right)$ where the interaction is of **mean-field** type ***(symmetry)*** since it occurs only through

$$\mu_t^N = \frac{1}{N} \sum_{j=1}^{N} \delta_{X_t^j}$$

which is the **empirical distribution** of the agents' states ($\delta_x =$ Dirac mass at x)

The **social cost** is the average of all the individual costs:

$$J^N(v^1, \ldots, v^N) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\int_0^T \underbrace{f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right)}_{\text{running cost}} dt + \underbrace{g\left(X_T^i\right)}_{\text{terminal cost}} \right]$$

Goal: Find an optimal $\hat{v} = (\hat{v}^1, \ldots, \hat{v}^N)$ **minimizing** J^N

Rem.: Terminal cost and drift could involve μ_t^N too

Control with Mean Field Interactions: N-Agent & Asymptotic Versions

Optimal control of N agents: Find $(\hat{v}^1, \dots, \hat{v}^N)$ minimizing the social cost

$$J^N(v^1, \dots, v^N) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\int_0^T f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right) dt + g\left(X_T^i\right) \right],$$

where $\mu_t^N := \frac{1}{N} \sum_{j=1}^N \delta_{X_t^j}$ and

$$dX_t^j = v^j(t, \mathbf{X}_t)\, dt + dW_t^j, \quad X_0^j \text{ i.i.d } \sim m_0\,.$$

Control with Mean Field Interactions: N-Agent & Asymptotic Versions

Optimal control of N agents: Find $(\hat{v}^1, \dots, \hat{v}^N)$ minimizing the social cost

$$J^N(v^1, \dots, v^N) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\int_0^T f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right) dt + g\left(X_T^i\right)\right],$$

where $\mu_t^N := \frac{1}{N} \sum_{j=1}^N \delta_{X_t^j}$ and

$$dX_t^j = v^j(t, \mathbf{X}_t)\, dt + dW_t^j, \quad X_0^j \text{ i.i.d } \sim m_0 .$$

As $N \longrightarrow +\infty$, $\mu_t^N \longrightarrow \mu_t$ = deterministic distribution. **Asymptotic** problem:

Control with Mean Field Interactions: N-Agent & Asymptotic Versions

Optimal control of N agents: Find $(\hat{v}^1, \dots, \hat{v}^N)$ minimizing the social cost

$$J^N(v^1, \dots, v^N) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\int_0^T f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right) dt + g\left(X_T^i\right)\right],$$

where $\mu_t^N := \frac{1}{N} \sum_{j=1}^N \delta_{X_t^j}$ and

$$dX_t^j = v^j(t, \mathbf{X}_t)\, dt + dW_t^j, \quad X_0^j \text{ i.i.d } \sim m_0 \,.$$

As $N \longrightarrow +\infty$, $\mu_t^N \longrightarrow \mu_t$ = deterministic distribution. **Asymptotic** problem:

Mean field control (MFC): Find a control $\hat{v}$ minimizing

$$J(v) = \mathbb{E}\left[\int_0^T f(X_t^v, \mathcal{L}(X_t^v), v(t, X_t))\, dt + g(X_T^v)\right],$$

where $\mu_t = \mathcal{L}(X_t^v)$ is the **law** of X_t^v = state of a **representative player** with

$$dX_t^v = v(t, X_t^v)\, dt + dW_t, \quad X_0^v \sim m_0 \,.$$

Optimal control of N agents: Find $(\hat{v}^1, \dots, \hat{v}^N)$ minimizing the social cost

$$J^N(v^1, \dots, v^N) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\int_0^T f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right) dt + g\left(X_T^i\right)\right],$$

where $\mu_t^N := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^j}$ and

$$dX_t^j = v^j(t, \mathbf{X}_t)\, dt + dW_t^j, \quad X_0^j \text{ i.i.d } \sim m_0 \,.$$

As $N \longrightarrow +\infty$, $\mu_t^N \longrightarrow \mu_t$ = deterministic distribution. **Asymptotic** problem:

Mean field control (MFC): Find a control $\hat{v}$ minimizing

$$J(v) = \mathbb{E}\left[\int_0^T f(X_t^v, \mathcal{L}(X_t^v), v(t, X_t))\, dt + g(X_T^v)\right],$$

where $\mu_t = \mathcal{L}(X_t^v)$ is the **law** of X_t^v = state of a **representative player** with

$$dX_t^v = v(t, X_t^v)\, dt + dW_t, \quad X_0^v \sim m_0 \,.$$

Motivations:

- "$N \to \infty$": a large number of **cooperative** agents; **McKean-Vlasov** dynamics:

$$dX_t = b\Big(X_t, \mu_t^v, v(t, X_t)\Big)\, dt + dW_t$$

- Non-linear dependence on the law: e.g. risk measures:

$$\mathbb{E}[g(X_T, \mu_T)] = \mathrm{Var}(X_T) - \mathbb{E}[X_T]$$

Nash Equilibrium: When a player optimizes, the other players' controls are fixed

Nash Equilibrium: When a player optimizes, the other players' controls are fixed

Nash equilibrium between N players: Find $\hat{v} = (\hat{v}^1, \ldots, \hat{v}^N)$ such that

For each $i = 1, \ldots, N$, given $\hat{v}^{-i} = (\hat{v}^1, \ldots, \hat{v}^{i-1}, \hat{v}^{i+1}, \ldots, \hat{v}^N)$, $\hat{v}^i$ minimizes

$$v^i \mapsto J(v^i; \hat{v}^{-i}) = \mathbb{E}\left[\int_0^T f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right) dt + g\left(X_T^i\right)\right]$$

where $\mu_t^N = \frac{1}{N}\sum_{j\neq i} \delta_{X_t^j} + \frac{1}{N}\delta_{X_t^i}$ and

$$dX_t^i = v^i(t, \mathbf{X}_t)\, dt + dW_t^i, \qquad dX_t^j = \hat{v}^j(t, \mathbf{X}_t)\, dt + dW_t^j, \quad j \neq i$$

Games with Mean Field Interactions: N-Agent & Asymptotic Versions

Nash Equilibrium: When a player optimizes, the other players' controls are fixed

Nash equilibrium between N players: Find $\hat{v} = (\hat{v}^1, \ldots, \hat{v}^N)$ such that

For each $i = 1, \ldots, N$, given $\hat{v}^{-i} = (\hat{v}^1, \ldots, \hat{v}^{i-1}, \hat{v}^{i+1}, \ldots, \hat{v}^N)$, $\hat{v}^i$ minimizes

$$v^i \mapsto J(v^i; \hat{v}^{-i}) = \mathbb{E}\left[\int_0^T f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right)\, dt + g\left(X_T^i\right)\right]$$

where $\mu_t^N = \frac{1}{N}\sum_{j\neq i} \delta_{X_t^j} + \frac{1}{N}\delta_{X_t^i}$ and

$$dX_t^i = v^i(t, \mathbf{X}_t)\, dt + dW_t^i, \qquad dX_t^j = \hat{v}^j(t, \mathbf{X}_t)\, dt + dW_t^j, \quad j \neq i$$

As $N \longrightarrow +\infty$, $\mu_t^N \longrightarrow \mu_t$ which is *not influenced by* v^i. **Asymptotic** problem:

Games with Mean Field Interactions: N-Agent & Asymptotic Versions

Nash Equilibrium: When a player optimizes, the other players' controls are fixed

Nash equilibrium between N players: Find $\hat{v} = (\hat{v}^1, \dots, \hat{v}^N)$ such that

For each $i = 1, \dots, N$, given $\hat{v}^{-i} = (\hat{v}^1, \dots, \hat{v}^{i-1}, \hat{v}^{i+1}, \dots, \hat{v}^N)$, $\hat{v}^i$ minimizes

$$v^i \mapsto J(v^i; \hat{v}^{-i}) = \mathbb{E}\left[\int_0^T f\left(X_t^i, \mu_t^N, v^i(t, \mathbf{X}_t)\right) dt + g\left(X_T^i\right)\right]$$

where $\mu_t^N = \frac{1}{N}\sum_{j \neq i} \delta_{X_t^j} + \frac{1}{N}\delta_{X_t^i}$ and

$$dX_t^i = v^i(t, \mathbf{X}_t)\, dt + dW_t^i, \qquad dX_t^j = \hat{v}^j(t, \mathbf{X}_t)\, dt + dW_t^j, \quad j \neq i$$

As $N \longrightarrow +\infty$, $\mu_t^N \longrightarrow \mu_t$ which is *not influenced by* v^i. **Asymptotic** problem:

Mean field game (MFG): Find $(\hat{v}, \hat{\mu})$ = (control, flow of distributions) such that

(1) Given $\hat{\mu} = (\hat{\mu}_t)_{t \in [0,T]}$, the control $\hat{v}$ minimizes

$$v \mapsto J(v; \hat{\mu}) = \mathbb{E}\left[\int_0^T f(X_t^v, \hat{\mu}_t, v(t, X_t^v))\, dt + g(X_T^v)\right],$$

where $dX_t^v = v(t, X_t^v)\, dt + dW_t$, $X_0^v \sim m_0$
(2) $\hat{\mu}_t = \mathcal{L}(X_t^{\hat{v}})$ for all t.

(1) = standard **optimal control** problem for a representative player vs the population
(2) = **consistency** condition (fixed point): *"all the agents think in the same way"*

Outline

Introduction

Part 1: Solving Mean Field Problems with Deep Learning

- Direct approach for MFC
- MKV FBSDE system
- Mean Field PDE System

Part 2: Reinforcement Learning with Mean-Field Interactions

Conclusion

Numerical Methods for MFG

Methods based on a deterministic approach:

- Finite differences & Newton meth.: [Achdou, Capuzzo-Dolcetta'10; . . . ; Achdou, L.'15]
- Gradient descent: [L., Pironneau'14; Pfeiffer'16]
- Semi-Lagrangian scheme: [Carlini, Silva'14; Carlini, Silva'15]
- Augmented Lagrangian & ADMM: [Benamou, Carlier'14; Achdou, L.'16; Andreev'17]
- Primal-dual algo.: [Briceño-Arias, Kalise, Silva'18; BAKS + Kobeissi, L., Mateos González'18]
- Monotone operators: [Almulla et al.'17; Gomes, Saúde'18; Gomes, Yang'18]

Methods based on a probabilistic approach:

- Cubature: [Chaudru de Raynal, Garcia Trillos'15]
- Recursion: [Chassagneux et al.'17; Angiuli et al.'18]
- MC+Regression: [Balata, Huré, L., Pham, Pimentel'18]

Numerical Methods for MFG

Methods based on a deterministic approach:

- Finite differences & Newton meth.: [Achdou, Capuzzo-Dolcetta'10; . . . ; Achdou, L.'15]
- Gradient descent: [L., Pironneau'14; Pfeiffer'16]
- Semi-Lagrangian scheme: [Carlini, Silva'14; Carlini, Silva'15]
- Augmented Lagrangian & ADMM: [Benamou, Carlier'14; Achdou, L.'16; Andreev'17]
- Primal-dual algo.: [Briceño-Arias, Kalise, Silva'18; BAKS + Kobeissi, L., Mateos González'18]
- Monotone operators: [Almulla et al.'17; Gomes, Saúde'18; Gomes, Yang'18]

Methods based on a probabilistic approach:

- Cubature: [Chaudru de Raynal, Garcia Trillos'15]
- Recursion: [Chassagneux et al.'17; Angiuli et al.'18]
- MC+Regression: [Balata, Huré, L., Pham, Pimentel'18]

Limitations:

- dimensionality (state in dimension ≤ 3)
- structure of the problem (simple costs, dynamics and noises)

Numerical Methods for MFG

Methods based on a deterministic approach:

- Finite differences & Newton meth.: [Achdou, Capuzzo-Dolcetta'10; . . . ; Achdou, L.'15]
- Gradient descent: [L., Pironneau'14; Pfeiffer'16]
- Semi-Lagrangian scheme: [Carlini, Silva'14; Carlini, Silva'15]
- Augmented Lagrangian & ADMM: [Benamou, Carlier'14; Achdou, L.'16; Andreev'17]
- Primal-dual algo.: [Briceño-Arias, Kalise, Silva'18; BAKS + Kobeissi, L., Mateos González'18]
- Monotone operators: [Almulla et al.'17; Gomes, Saúde'18; Gomes, Yang'18]

Methods based on a probabilistic approach:

- Cubature: [Chaudru de Raynal, Garcia Trillos'15]
- Recursion: [Chassagneux et al.'17; Angiuli et al.'18]
- MC+Regression: [Balata, Huré, L., Pham, Pimentel'18]

Limitations:

- dimensionality (state in dimension ≤ 3)
- structure of the problem (simple costs, dynamics and noises)

Recent progress: extending the toolbox with tools from **machine learning**:

- approximation without a grid (mesh-free methods): **opt. control** & **distribution**

  $\rightarrow$ [Carmona, L.; Al-Aradi *et al.*; Fouque *et al.*; Germain *et al.*; Ruthotto *et al.*; Agram *et al.*; . . .]
- even when the **dynamics / cost are not known** (model-free methods)

  $\rightarrow$ [Guo *et al.*; Subramanian *et al.*; Elie *et al.*; Carmona *et al.*; Pham *et al.*; . . .]

Outline

Approximation Result for MFC

MFC:

Minimize over $v(\cdot,\cdot)$

$$J(v(\cdot,\cdot)) = \mathbb{E}\left[\int_0^T f(X_t, \mu_t, v(t, X_t))\,dt + g(X_T)\right],$$

where $\mu_t = \mathcal{L}(X_t)$ with

$$X_0 \sim m_0\,, \quad dX_t = v(t, X_t)\,dt + \,dW_t$$

Approximation Result for MFC

MFC: (1) Finite pop.,

Minimize over **decentralized** controls $v(\cdot, \cdot)$ with N agents

$$J^N(v(\cdot,\cdot)) = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\int_0^T f\left(X_t^i, \mu_t^N, v(t, X_t^i)\right)\, dt + g\left(X_T^i\right)\right],$$

with $\mu_t^N = \frac{1}{N}\sum_{j=1}^N \delta_{X_t^j}$,

$$X_0^j \sim m_0\,, \quad dX_t^j = v(t, X_t^j)\, dt + dW_t^j$$

Approximation Result for MFC

MFC: (1) Finite pop., (2) neural network φ_θ,

Minimize over **neural network** parameters θ with N agents

$$J^N(\theta) = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\int_0^T f\left(X_t^i, \mu_t^N, \varphi_\theta(t, X_t^i)\right)\, dt + g\left(X_T^i\right)\right],$$

with $\mu_t^N = \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^j}$,

$$X_0^j \sim m_0\,, \quad dX_t^j = \varphi_\theta(t, X_t^j)\, dt + dW_t^j$$

Approximation Result for MFC

MFC: (1) Finite pop., (2) neural network φ_θ, (3) time discretization

Minimize over **neural network** parameters θ with N agents and N_T time steps

$$J^{N,N_T}(\theta) = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{N_T-1} f\left(X_n^i, \mu_n^N, \varphi_\theta(t_n, X_n^i)\right)\Delta t + g\left(X_{N_T}^i\right)\right],$$

with $\mu_n^N = \frac{1}{N}\sum_{j=1}^{N}\delta_{X_n^j}$,

$$X_0^j \sim m_0\,, \quad X_{n+1}^j - X_n^j = \varphi_\theta(t_n, X_n^j)\Delta t + \Delta W_n^j$$

Approximation Result for MFC

MFC: (1) Finite pop., (2) neural network φ_θ, (3) time discretization

Minimize over **neural network** parameters θ with N agents and N_T time steps

$$J^{N,N_T}(\theta) = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{N_T-1} f\left(X_n^i, \mu_n^N, \varphi_\theta(t_n, X_n^i)\right)\Delta t + g\left(X_{N_T}^i\right)\right],$$

with $\mu_n^N = \frac{1}{N}\sum_{j=1}^N \delta_{X_n^j}$,

$$X_0^j \sim m_0\,, \quad X_{n+1}^j - X_n^j = \varphi_\theta(t_n, X_n^j)\Delta t + \Delta W_n^j$$

Theorem: Convergence rate of the approximation [Carmona, L.'20]

Under suitable assumptions (in particular regularity of the value function),

$$\left|\inf_{v(\cdot,\cdot)} J(v(\cdot,\cdot)) - \inf_{\theta} J^{N,N_T}(\theta)\right| \leq \epsilon_1(N) + \epsilon_2(\dim(\theta)) + \epsilon_3(N_T)$$

Approximation Result for MFC

MFC: (1) Finite pop., (2) neural network φ_θ, (3) time discretization

Minimize over **neural network** parameters θ with N agents and N_T time steps

$$J^{N,N_T}(\theta) = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{N_T-1} f\left(X_n^i, \mu_n^N, \varphi_\theta(t_n, X_n^i)\right)\Delta t + g\left(X_{N_T}^i\right)\right],$$

with $\mu_n^N = \frac{1}{N}\sum_{j=1}^N \delta_{X_n^j}$,

$$X_0^j \sim m_0\,, \quad X_{n+1}^j - X_n^j = \varphi_\theta(t_n, X_n^j)\Delta t + \Delta W_n^j$$

Theorem: Convergence rate of the approximation [Carmona, L.'20]

Under suitable assumptions (in particular regularity of the value function),

$$\left|\inf_{v(\cdot,\cdot)} J(v(\cdot,\cdot)) - \inf_{\theta} J^{N,N_T}(\theta)\right| \leq \epsilon_1(N) + \epsilon_2(\dim(\theta)) + \epsilon_3(N_T)$$

Implementation: **Stochastic Gradient Descent**

Loss function = cost: $J^{N,N_T}(\theta) = \mathbb{E}[\mathbb{L}(\varphi_\theta, \xi)]$
One sample: $\xi = \left(X_0^j, (\Delta W_n^j)_{n=0,\dots,N_T-1}\right)_{j=1,\dots,N}$

Approximation Result for MFC

MFC: (1) Finite pop., (2) neural network φ_θ, (3) time discretization

Minimize over **neural network** parameters θ with N agents and N_T time steps

$$J^{N,N_T}(\theta) = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{N_T-1} f\left(X_n^i, \mu_n^N, \varphi_\theta(t_n, X_n^i)\right)\Delta t + g\left(X_{N_T}^i\right)\right],$$

with $\mu_n^N = \frac{1}{N}\sum_{j=1}^{N}\delta_{X_n^j}$,

$$X_0^j \sim m_0\,, \quad X_{n+1}^j - X_n^j = \varphi_\theta(t_n, X_n^j)\Delta t + \Delta W_n^j$$

Theorem: Convergence rate of the approximation [Carmona, L.'20]

Under suitable assumptions (in particular regularity of the value function),

$$\left|\inf_{v(\cdot,\cdot)} J(v(\cdot,\cdot)) - \inf_{\theta} J^{N,N_T}(\theta)\right| \leq \epsilon_1(N) + \epsilon_2(\dim(\theta)) + \epsilon_3(N_T)$$

Implementation: **Stochastic Gradient Descent**

Loss function = cost: $J^{N,N_T}(\theta) = \mathbb{E}[\mathbb{L}(\varphi_\theta, \xi)]$
One sample: $\xi = \left(X_0^j, (\Delta W_n^j)_{n=0,\dots,N_T-1}\right)_{j=1,\dots,N}$

- Generalizes standard stochastic control problems (no MF); [...; Gobet, Munos'05; Han, E'16]
- Related work with mean field: [Fouque, Zhang'19; Germain *et al.*'19; ...]

Approximation Result: Sketch of Proof

Proposition 1 (N agents & decentralized controls):

Under suitable assumptions, there exists a decentralized control $\hat{v}$ s.t. *(d = dimension of X_t)*

$$\left| \inf_{v(\cdot)} J(v(\cdot)) - J^N(\hat{v}(\cdot)) \right| \leq \epsilon_1(N) \in \widetilde{O}\left(N^{-1/d}\right).$$

Proof: propagation of chaos type argument [Carmona, Delarue'18]

Approximation Result: Sketch of Proof

Proposition 1 (N agents & decentralized controls):

Under suitable assumptions, there exists a decentralized control $\hat{v}$ s.t. *(d = dimension of X_t)*

$$\left| \inf_{v(\cdot)} J(v(\cdot)) - J^N(\hat{v}(\cdot)) \right| \leq \epsilon_1(N) \in \widetilde{O}\left(N^{-1/d}\right).$$

Proof: propagation of chaos type argument [Carmona, Delarue'18]

Proposition 2 (approximation by neural networks): Under suitable assumptions

There exists a set of parameters θ for a one-hidden layer $\hat{\varphi}_\theta$ s.t.

$$\left| J^N(\hat{v}(\cdot)) - J^N(\hat{\varphi}_\theta(\cdot)) \right| \leq \epsilon_2(\dim(\theta)) \in O\left(\dim(\theta)^{-\frac{1}{3(d+1)}}\right).$$

Proof: Key difficulty: approximate $\hat{v}(\cdot)$ by $\hat{\varphi}_\theta(\cdot)$ while controlling $\|\nabla \hat{\varphi}_\theta(\cdot)\|$ by $\|\nabla \hat{v}(\cdot)\|$

$\rightarrow$ universal approximation without rate of convergence is not enough

$\rightarrow$ approximation rate for the derivative too, e.g. from [Mhaskar, Micchelli'95]

Approximation Result: Sketch of Proof

Proposition 1 (N agents & decentralized controls):

Under suitable assumptions, there exists a decentralized control $\hat{v}$ s.t. *(d = dimension of X_t)*

$$\left| \inf_{v(\cdot)} J(v(\cdot)) - J^N(\hat{v}(\cdot)) \right| \leq \epsilon_1(N) \in \widetilde{O}\left(N^{-1/d}\right).$$

Proof: propagation of chaos type argument [Carmona, Delarue'18]

Proposition 2 (approximation by neural networks): Under suitable assumptions

There exists a set of parameters θ for a one-hidden layer $\hat{\varphi}_\theta$ s.t.

$$\left| J^N(\hat{v}(\cdot)) - J^N(\hat{\varphi}_\theta(\cdot)) \right| \leq \epsilon_2(\dim(\theta)) \in O\left(\dim(\theta)^{-\frac{1}{3(d+1)}}\right).$$

Proof: Key difficulty: approximate $\hat{v}(\cdot)$ by $\hat{\varphi}_\theta(\cdot)$ while controlling $\|\nabla \hat{\varphi}_\theta(\cdot)\|$ by $\|\nabla \hat{v}(\cdot)\|$

$\rightarrow$ universal approximation without rate of convergence is not enough

$\rightarrow$ approximation rate for the derivative too, e.g. from [Mhaskar, Micchelli'95]

Proposition 3 (Euler-Maruyama scheme):

For a specific neural network $\hat{\varphi}_\theta(\cdot)$,

$$\left| J^N(\hat{\varphi}_\theta(\cdot)) - J^{N,N_T}(\hat{\varphi}_\theta(\cdot)) \right| \leq \epsilon_3(N_T) \in O\left(N_T^{-1/2}\right).$$

Key point: $O(\cdot)$ independent of N and n_U

Proof: analysis of **strong error rate** for Euler scheme (reminiscent of [Bossy, Talay'97])

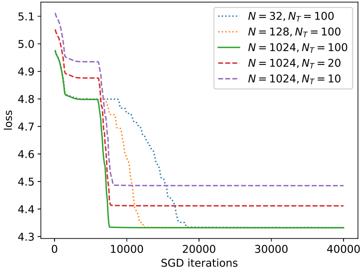
Numerical Illustration: LQ MFC

Benchmark to assess **empirical convergence of SGD:** LQ problem with explicit sol.

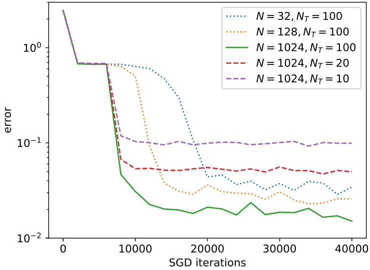
Example: Linear dynamics, quadratic costs of the type

$$f(x, \mu, v) = \underbrace{(\bar{\mu} - x)^2}_{\text{distance to mean position}} + \underbrace{v^2}_{\text{cost of moving}}, \qquad \bar{\mu} = \underbrace{\int \mu(\xi) d\xi}_{\text{mean position}}, \qquad g(x) = x^2$$

Numerical example with $d = 10$:



total cost (= loss function)



L^2-error on the control

(More details in [Carmona, L.'20])

Q: *What about mean-field Nash equilibria?*

Forward-backward mean-field systems

- **Forward-backward structure:**
  - ◇ Forward evolution of the state / density
  - ◇ Backward evolution of the control / value function

Q: *What about mean-field Nash equilibria?*

Forward-backward mean-field systems

- **Forward-backward structure:**
  - ◇ Forward evolution of the state / density
  - ◇ Backward evolution of the control / value function

- **SDE system:**
  - ◇ Deep BSDE method [E, Jentzen, Han'18] $\rightarrow$ [Carmona, L.'20]

Q: *What about mean-field Nash equilibria?*

Forward-backward mean-field systems

- **Forward-backward structure:**
  - ⋄ Forward evolution of the state / density
  - ⋄ Backward evolution of the control / value function

- **SDE system:**
  - ⋄ Deep BSDE method [E, Jentzen, Han'18] → [Carmona, L.'20]

- **PDE system:**
  - ⋄ Deep Galerkin Method [Sirignano, Spiliopoulos'18] → [Carmona, L.'20]

Reminder:

Nash Eq.: When a player optimizes, the other players' controls are fixed

Mean field game (MFG): Find $(\hat{v}, \hat{\mu}) =$ (control, flow of distribution) such that

(1) Given $\hat{\mu} = (\hat{\mu}_t)_{t \in [0,T]}$, the control $\hat{v}$ minimizes

$$v \mapsto J(v; \hat{\mu}) = \mathbb{E}\left[\int_0^T f(X_t^v, \hat{\mu}_t, v(t, X_t^v))\, dt + g(X_T^v)\right],$$

where $dX_t^v = v(t, X_t^v)\, dt + dW_t$,

(2) $\hat{\mu}_t = \mathcal{L}(X_t^{\hat{v}})$ for all t.

(1) = standard **optimal control** problem for a representative player vs the population

(2) = **consistency** condition (fixed point): *"all the agents think in the same way"*

Optimality Condition for MFG: forward-backward system

At **equilibrium**, X evolves according to: $X_0 \sim m_0$, $dX_t = \hat{v}(t, X_t)\, dt + dW_t$.
The evolution of its distribution $\hat{\mu}_t = \mathcal{L}(X_t)$ is given by a **Fokker-Planck** PDE:

$$\underbrace{\hat{\mu}(t=0,x) = m_0(x),}_{\text{initial condition}} \qquad \partial_t \hat{\mu}(t,x) \quad = \quad - \quad \underbrace{\partial_x \left(\hat{\mu}(t,x)\, \hat{v}(t,x)\right)}_{\text{advection}} \quad + \quad \frac{1}{2} \underbrace{\partial_{xx} \hat{\mu}(t,x)}_{\text{diffusion}}$$

Optimality Condition for MFG: forward-backward system

At **equilibrium**, X evolves according to: $X_0 \sim m_0$, $dX_t = \hat{v}(t, X_t)\, dt + dW_t$.
The evolution of its distribution $\hat{\mu}_t = \mathcal{L}(X_t)$ is given by a **Fokker-Planck** PDE:

$$\underbrace{\hat{\mu}(t=0, x) = m_0(x),}_{\text{initial condition}} \qquad \partial_t \hat{\mu}(t,x) \quad = \quad - \quad \underbrace{\partial_x \left(\hat{\mu}(t,x)\, \hat{v}(t,x)\right)}_{\text{advection}} \quad + \quad \frac{1}{2}\underbrace{\partial_{xx}\hat{\mu}(t,x)}_{\text{diffusion}}$$

How can we characterize the ***best response*** *(= opt. control) of a typical player?*

$$\hat{v}(\cdot,\cdot) = \operatorname*{argmin}_{v(\cdot,\cdot)} J(v(\cdot,\cdot); \hat{\mu}) = \operatorname*{argmin}_{v(\cdot,\cdot)} \mathbb{E}\left[\int_0^T f(X_t, \hat{\mu}_t, v(t, X_t))\, dt + g(X_T)\right]$$

Optimality Condition for MFG: forward-backward system

At **equilibrium**, X evolves according to: $X_0 \sim m_0$, $dX_t = \hat{v}(t, X_t)\, dt + dW_t$.
The evolution of its distribution $\hat{\mu}_t = \mathcal{L}(X_t)$ is given by a **Fokker-Planck** PDE:

$$\underbrace{\hat{\mu}(t=0,x) = m_0(x),}_{\text{initial condition}} \qquad \partial_t \hat{\mu}(t,x) \quad = \quad - \underbrace{\partial_x \left(\hat{\mu}(t,x)\, \hat{v}(t,x)\right)}_{\text{advection}} \quad + \quad \frac{1}{2} \underbrace{\partial_{xx} \hat{\mu}(t,x)}_{\text{diffusion}}$$

*How can we characterize the **best response** (= opt. control) of a typical player?*

$$\hat{v}(\cdot,\cdot) = \operatorname*{argmin}_{v(\cdot,\cdot)} J(v(\cdot,\cdot); \hat{\mu}) = \operatorname*{argmin}_{v(\cdot,\cdot)} \mathbb{E}\left[\int_0^T f(X_t, \hat{\mu}_t, v(t, X_t))\, dt + g(X_T)\right]$$

Picard iterations for MFG

Start with an initial guess $\mu^{(0)}$. Repeat for $k = 0, 1, \ldots$: Given $\mu^{(k)}$,

(1) Compute $v^{(k+1)}$ $=$ best response against $\mu^{(k)}$

(2) Compute $\mu^{(k+1)}$ $=$ mean-field flow associated to $v^{(k+1)}$

Converges if $\mu^{(k)} \mapsto \mu^{(k+1)}$ is a strict contraction (very restrictive . . .)

Optimality Condition for MFG: forward-backward system

At **equilibrium**, X evolves according to: $X_0 \sim m_0$, $dX_t = \hat{v}(t, X_t)\, dt + dW_t$.
The evolution of its distribution $\hat{\mu}_t = \mathcal{L}(X_t)$ is given by a **Fokker-Planck** PDE:

$$\underbrace{\hat{\mu}(t=0,x) = m_0(x),}_{\text{initial condition}} \qquad \partial_t \hat{\mu}(t,x) \quad = \quad - \underbrace{\partial_x\left(\hat{\mu}(t,x)\,\hat{v}(t,x)\right)}_{\text{advection}} \quad + \quad \frac{1}{2}\underbrace{\partial_{xx}\hat{\mu}(t,x)}_{\text{diffusion}}$$

How can we characterize the ***best response*** *(= opt. control) of a typical player?*

$$\hat{v}(\cdot,\cdot) = \operatorname*{argmin}_{v(\cdot,\cdot)} J(v(\cdot,\cdot); \hat{\mu}) = \operatorname*{argmin}_{v(\cdot,\cdot)} \mathbb{E}\left[\int_0^T f(X_t, \hat{\mu}_t, v(t, X_t))\, dt + g(X_T)\right]$$

(1) Dynamic programming: $\hat{v}(\cdot,\cdot)$ is given in terms of the **value function** $\hat{u}(\cdot,\cdot)$ which solves the **Hamilton-Jacobi-Bellman** PDE

$$\underbrace{-\partial_t \hat{u}(t,x)}_{\text{backward evolution}} \quad = \quad \underbrace{\hat{H}(x, \mu(t,\cdot), \partial_x \hat{u}(t,x))}_{\text{Hamiltonian}} + \frac{1}{2}\partial_{xx}\hat{u}(t,x), \qquad \underbrace{\hat{u}(t=T,x) = g(x)}_{\text{terminal condition}}$$

where $\hat{H}(x,m,q) := \min_{a \in \mathbb{R}^d}\big(f(x,m,a) + q \cdot a\big)$.

Optimality Condition for MFG: forward-backward system

At **equilibrium**, X evolves according to: $X_0 \sim m_0$, $dX_t = \hat{v}(t, X_t)\, dt + dW_t$.
The evolution of its distribution $\hat{\mu}_t = \mathcal{L}(X_t)$ is given by a **Fokker-Planck** PDE:

$$\underbrace{\hat{\mu}(t=0,x) = m_0(x),}_{\text{initial condition}} \qquad \partial_t \hat{\mu}(t,x) \quad = \quad - \underbrace{\partial_x \left(\hat{\mu}(t,x)\, \hat{v}(t,x)\right)}_{\text{advection}} \quad + \quad \frac{1}{2} \underbrace{\partial_{xx} \hat{\mu}(t,x)}_{\text{diffusion}}$$

How can we characterize the ***best response*** *(= opt. control) of a typical player?*

$$\hat{v}(\cdot,\cdot) = \operatorname*{argmin}_{v(\cdot,\cdot)} J(v(\cdot,\cdot); \hat{\mu}) = \operatorname*{argmin}_{v(\cdot,\cdot)} \mathbb{E}\left[\int_0^T f(X_t, \hat{\mu}_t, v(t, X_t))\, dt + g(X_T)\right]$$

(1) Dynamic programming: $\hat{v}(\cdot,\cdot)$ is given in terms of the **value function** $\hat{u}(\cdot,\cdot)$ which solves the **Hamilton-Jacobi-Bellman** PDE

$$\underbrace{-\partial_t \hat{u}(t,x)}_{\text{backward evolution}} \quad = \quad \underbrace{\hat{H}(x, \mu(t,\cdot), \partial_x \hat{u}(t,x))}_{\text{Hamiltonian}} + \frac{1}{2}\partial_{xx}\hat{u}(t,x), \qquad \underbrace{\hat{u}(t=T,x) = g(x)}_{\text{terminal condition}}$$

where $\hat{H}(x,m,q) := \min_{a\in\mathbb{R}^d}\big(f(x,m,a) + q\cdot a\big)$.

(2) Or: Stoch. Maximum Principle: $\hat{v}(t, X_t)$ is characterized in terms of $X_t, \mathcal{L}(X_t)$ and the **adjoint state** $Y_t \in \mathbb{R}^d$, which solves the **backward** SDE

$$dY_t \quad = \quad -\partial_x \hat{H}(X_t, \hat{\mu}_t, Y_t) dt + Z_t \cdot dW_t, \qquad Y_T = \partial_x g(X_T)$$

Optimality Condition for MFG: forward-backward system

At **equilibrium**, X evolves according to: $X_0 \sim m_0$, $dX_t = \hat{v}(t, X_t)\,dt + dW_t$.
The evolution of its distribution $\hat{\mu}_t = \mathcal{L}(X_t)$ is given by a **Fokker-Planck** PDE:

$$\underbrace{\hat{\mu}(t=0,x) = m_0(x),}_{\text{initial condition}} \qquad \partial_t \hat{\mu}(t,x) \quad = \quad - \quad \underbrace{\partial_x \left(\hat{\mu}(t,x)\,\hat{v}(t,x)\right)}_{\text{advection}} \quad + \quad \frac{1}{2}\underbrace{\partial_{xx}\hat{\mu}(t,x)}_{\text{diffusion}}$$

How can we characterize the ***best response*** *(= opt. control) of a typical player?*

$$\hat{v}(\cdot,\cdot) = \underset{v(\cdot,\cdot)}{\operatorname{argmin}}\, J(v(\cdot,\cdot);\hat{\mu}) = \underset{v(\cdot,\cdot)}{\operatorname{argmin}}\, \mathbb{E}\left[\int_0^T f(X_t, \hat{\mu}_t, v(t, X_t))\,dt + g(X_T)\right]$$

(1) Dynamic programming: $\hat{v}(\cdot,\cdot)$ is given in terms of the **value function** $\hat{u}(\cdot,\cdot)$ which solves the **Hamilton-Jacobi-Bellman** PDE

$$\underbrace{-\partial_t \hat{u}(t,x)}_{\text{backward evolution}} \quad = \quad \underbrace{\hat{H}(x, \mu(t,\cdot), \partial_x \hat{u}(t,x))}_{\text{Hamiltonian}} + \frac{1}{2}\partial_{xx}\hat{u}(t,x), \qquad \underbrace{\hat{u}(t=T,x) = g(x)}_{\text{terminal condition}}$$

where $\hat{H}(x,m,q) := \min_{a\in\mathbb{R}^d}\left(f(x,m,a) + q\cdot a\right)$.

(2) Or: Stoch. Maximum Principle: $\hat{v}(t, X_t)$ is characterized in terms of $X_t, \mathcal{L}(X_t)$ and the **adjoint state** $Y_t \in \mathbb{R}^d$, which solves the **backward** SDE

$$dY_t \quad = \quad -\partial_x \hat{H}(X_t, \hat{\mu}_t, Y_t)dt + Z_t \cdot dW_t, \qquad Y_T = \partial_x g(X_T)$$

$\Rightarrow$ **forward**-**backward** SDE or PDE system

ML for FBSDE of McKean-Vlasov type

Solutions of MFG (and MFC) can be characterized by **MKV FBSDEs** of the form

$$\begin{cases} dX_t = B(t, X_t, \mathcal{L}(X_t), Y_t)dt + dW_t, & X_0 \sim m_0 & \to \text{state} \\ dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t)dt + Z_t \cdot dW_t, & Y_T = G(X_T, \mathcal{L}(X_T)) & \to \text{control/cost} \end{cases}$$

ML for FBSDE of McKean-Vlasov type

Solutions of MFG (and MFC) can be characterized by **MKV FBSDEs** of the form

$$\begin{cases} dX_t = B(t, X_t, \mathcal{L}(X_t), Y_t)dt + dW_t, & X_0 \sim m_0 & \to \text{state} \\ dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t)dt + Z_t \cdot dW_t, & Y_T = G(X_T, \mathcal{L}(X_T)) & \to \text{control/cost} \end{cases}$$

Idea: rewrite as **optimal control of 2 forward SDEs** *([Ma,Yong], "Sannikov's trick", . . .)*

ML for FBSDE of McKean-Vlasov type

Solutions of MFG (and MFC) can be characterized by **MKV FBSDEs** of the form

$$\begin{cases} dX_t = B(t, X_t, \mathcal{L}(X_t), Y_t)dt + dW_t, & X_0 \sim m_0 & \to \text{state} \\ dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t)dt + Z_t \cdot dW_t, & Y_T = G(X_T, \mathcal{L}(X_T)) & \to \text{control/cost} \end{cases}$$

Idea: rewrite as **optimal control of 2 forward SDEs** *([Ma,Yong], "Sannikov's trick", ...)*

Reformulation as a **MFC problem**

Minimize over $y_0(\cdot)$ and $\mathbf{z}(\cdot) = (z_t(\cdot))_{t \geq 0}$

$$J(y_0(\cdot), \mathbf{z}(\cdot)) = \mathbb{E}\left[\| Y_T^{y_0, \mathbf{z}} - G(X_T^{y_0, \mathbf{z}}, \mathcal{L}(X_T^{y_0, \mathbf{z}})) \|_2 \right],$$

under the constraint that $(X^{y_0, \mathbf{z}}, Y^{y_0, \mathbf{z}})$ solve: $\forall t \in [0, T]$

$$\begin{cases} dX_t = B(t, X_t, \mathcal{L}(X_t), Y_t)dt + dW_t, & X_0 \sim \mu_0, \\ dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t)dt + z_t(X_t) \cdot dW_t, & Y_0 = y_0(X_0). \end{cases}$$

ML for FBSDE of McKean-Vlasov type

Solutions of MFG (and MFC) can be characterized by **MKV FBSDEs** of the form

$$\begin{cases} dX_t = B(t, X_t, \mathcal{L}(X_t), Y_t)dt + dW_t, & X_0 \sim m_0 & \to \text{state} \\ dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t)dt + Z_t \cdot dW_t, & Y_T = G(X_T, \mathcal{L}(X_T)) & \to \text{control/cost} \end{cases}$$

Idea: rewrite as **optimal control of 2 forward SDEs** *([Ma,Yong], "Sannikov's trick", ...)*

Reformulation as a **MFC problem**

Minimize over $y_0(\cdot)$ and $\mathbf{z}(\cdot) = (z_t(\cdot))_{t \geq 0}$

$$J(y_0(\cdot), \mathbf{z}(\cdot)) = \mathbb{E}\left[\|Y_T^{y_0,\mathbf{z}} - G(X_T^{y_0,\mathbf{z}}, \mathcal{L}(X_T^{y_0,\mathbf{z}}))\|_2 \right],$$

under the constraint that $(X^{y_0,\mathbf{z}}, Y^{y_0,\mathbf{z}})$ solve: $\forall t \in [0, T]$

$$\begin{cases} dX_t = B(t, X_t, \mathcal{L}(X_t), Y_t)dt + dW_t, & X_0 \sim \mu_0, \\ dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t)dt + z_t(X_t) \cdot dW_t, & Y_0 = y_0(X_0). \end{cases}$$

$\to$ MFC: can apply direct approach, replacing $y_0(\cdot), z(\cdot,\cdot)$ by NN

ML for FBSDE of McKean-Vlasov type

Solutions of MFG (and MFC) can be characterized by **MKV FBSDEs** of the form

$$\begin{cases} dX_t = B(t, X_t, \mathcal{L}(X_t), Y_t)dt + dW_t, & X_0 \sim m_0 & \to \text{state} \\ dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t)dt + Z_t \cdot dW_t, & Y_T = G(X_T, \mathcal{L}(X_T)) & \to \text{control/cost} \end{cases}$$

Idea: rewrite as **optimal control of 2 forward SDEs** *([Ma,Yong], "Sannikov's trick", ...)*

Reformulation as a **MFC problem**

Minimize over $y_0(\cdot)$ and $\mathbf{z}(\cdot) = (z_t(\cdot))_{t\geq 0}$

$$J(y_0(\cdot), \mathbf{z}(\cdot)) = \mathbb{E}\left[\|Y_T^{y_0,\mathbf{z}} - G(X_T^{y_0,\mathbf{z}}, \mathcal{L}(X_T^{y_0,\mathbf{z}}))\|_2\right],$$

under the constraint that $(X^{y_0,\mathbf{z}}, Y^{y_0,\mathbf{z}})$ solve: $\forall t \in [0, T]$

$$\begin{cases} dX_t = B(t, X_t, \mathcal{L}(X_t), Y_t)dt + dW_t, & X_0 \sim \mu_0, \\ dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t)dt + z_t(X_t) \cdot dW_t, & Y_0 = y_0(X_0). \end{cases}$$

$\to$ MFC: can apply direct approach, replacing $y_0(\cdot), z(\cdot, \cdot)$ by NN

Extends [Han, Jentzen, E'17] for FBSDE without mean-field interactions

Example: MFG for inter-bank borrowing/lending [Carmona, Fouque, Sun]
X = log-monetary reserve, α = rate of borrowing/lending to central bank, cost:

$$J(\alpha; \bar{m}) = \mathbb{E}\left[\int_0^T \left[\frac{1}{2}\alpha_t^2 - q\alpha_t(\bar{m}_t - X_t) + \frac{\epsilon}{2}(\bar{m}_t - X_t)^2\right] dt + \frac{c}{2}(\bar{m}_T - X_T)^2\right]$$

where $\bar{m} = (\bar{m}_t)_{t\geq 0}$ is the cond. mean given W^0 of the population states, and

$$dX_t = [a(\bar{m}_t - X_t) + \alpha_t]dt + \sigma\left(\sqrt{1-\rho^2}dW_t + \rho\, dW_t^0\right)$$

Numerical Illustration: LQ-MFG with common noise

Example: MFG for inter-bank borrowing/lending [Carmona, Fouque, Sun]
X = log-monetary reserve, α = rate of borrowing/lending to central bank, cost:

$$J(\alpha;\bar{m}) = \mathbb{E}\left[\int_0^T \left[\frac{1}{2}\alpha_t^2 - q\alpha_t(\bar{m}_t - X_t) + \frac{\epsilon}{2}(\bar{m}_t - X_t)^2\right] dt + \frac{c}{2}(\bar{m}_T - X_T)^2\right]$$

where $\bar{m} = (\bar{m}_t)_{t\geq 0}$ is the cond. mean given W^0 of the population states, and

$$dX_t = [a(\bar{m}_t - X_t) + \alpha_t]dt + \sigma\left(\sqrt{1-\rho^2}dW_t + \rho\, dW_t^0\right)$$

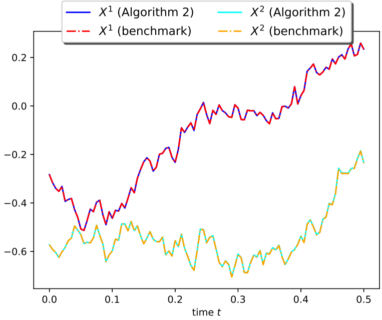
The Nash equilibrium can be characterized by the FBSDE system:

$$\begin{cases} dX_t = \underbrace{[(a+q)(\bar{m}_t - X_t) - Y_t]}_{\partial_y H} dt + \sigma\left(\sqrt{1-\rho^2}dW_t + \rho\, dW_t^0\right), \quad X_0 \sim m_0 \\ dY_t = \underbrace{(a+q)Y_t + (\epsilon - q^2)(\bar{m}_t - X_t)}_{-\partial_x H} dt + Z_t \cdot dW_t + Z_t^0 \cdot dW_t^0, \quad Y_T = c(X_T - \bar{m}_T) \end{cases}$$

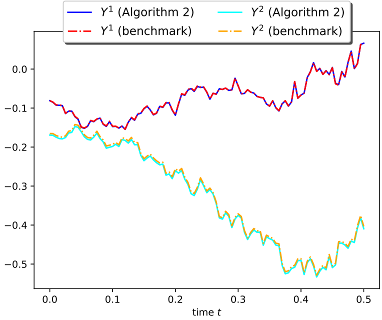
DL for FBSDE system VS (semi) analytical solution (LQ structure)

Numerical Illustration: LQ-MFG with common noise

DL for FBSDE system VS (semi) analytical solution (LQ structure)

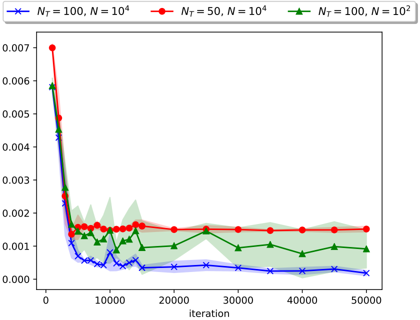


Samples of X

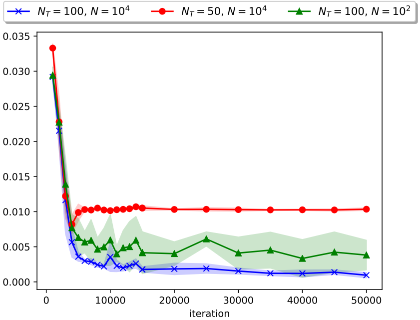


Samples of Y

DL for FBSDE system VS (semi) analytical solution (LQ structure)



L^2 error on X



L^2 error on Y

Outline

MFG: PDE system

MFG: If $(\hat{m}, \hat{v})$ solves the MFG, then $(\hat{m}(t,x), \hat{v}(t,x)) = \Big(m(t,x), \hat{\mathbf{v}}(x, m(t), \nabla u(t,x))\Big)$

with $$\hat{\mathbf{v}}(x, m(t), \nabla u(t,x)) = \operatorname{argmin}_{a \in \mathbb{R}^k} \Big(f(x, m(t), a) + \nabla u(t,x) \cdot b(x, m(t), a) \Big),$$

where (m, u) solve the PDE system

$$\begin{cases} 0 = \partial_t m(t,x) - \nu \Delta m(t,x) + \operatorname{div}\Big(m(t,x) \partial_q \hat{H}(x, m(t), \nabla u(t,x)) \Big) \\ 0 = \partial_t u(t,x) + \nu \Delta u(t,x) + \hat{H}(x, m(t), \nabla u(t,x)) \\ m(0,x) = m_0(x), \quad u(T,x) = g(x, m(T)) \end{cases}$$

with

$$\hat{H}(x, m, q) := \min_{a \in \mathbb{R}^k} \Big(f(x, m, a) + q \cdot b(x, m, a) \Big).$$

MFG: PDE system

MFG: If $(\hat{m}, \hat{v})$ solves the MFG, then $(\hat{m}(t,x), \hat{v}(t,x)) = \Big(m(t,x), \hat{\mathbf{v}}(x, m(t), \nabla u(t,x))\Big)$

with $\quad \hat{\mathbf{v}}(x, m(t), \nabla u(t,x)) = \operatorname{argmin}_{a \in \mathbb{R}^k} \Big(f(x, m(t), a) + \nabla u(t,x) \cdot b(x, m(t), a) \Big),$

where (m, u) solve the PDE system

$$\begin{cases} 0 = \partial_t m(t,x) - \nu \Delta m(t,x) + \operatorname{div}\Big(m(t,x) \partial_q \hat{H}(x, m(t), \nabla u(t,x)) \Big) \\ 0 = \partial_t u(t,x) + \nu \Delta u(t,x) + \hat{H}(x, m(t), \nabla u(t,x)) \\ m(0,x) = m_0(x), \quad u(T,x) = g(x, m(T)) \end{cases}$$

with

$$\hat{H}(x, m, q) := \min_{a \in \mathbb{R}^k} \Big(f(x, m, a) + q \cdot b(x, m, a) \Big).$$

Deep Galerkin Method [Sirignano, Spiliopoulos]:

→ application to **MFGs**: see [Al-Aradi *et al.*; Carmona, L.; Cao, Guo, L.]

- replace unknown functions by deep NN
- try to minimize the squared residual
- by sampling points in the domain

MFG: PDE system

MFG: If $(\hat{m}, \hat{v})$ solves the MFG, then $(\hat{m}(t,x), \hat{v}(t,x)) = \Big(m(t,x), \hat{\mathbf{v}}(x, m(t), \nabla u(t,x))\Big)$

with $$\hat{\mathbf{v}}(x, m(t), \nabla u(t,x)) = \mathrm{argmin}_{a \in \mathbb{R}^k} \Big(f(x, m(t), a) + \nabla u(t,x) \cdot b(x, m(t), a) \Big),$$

where (m, u) solve the PDE system

$$\begin{cases} 0 = \partial_t m(t,x) - \nu \Delta m(t,x) + \mathrm{div}\Big(m(t,x) \partial_q \hat{H}(x, m(t), \nabla u(t,x)) \Big) \\ 0 = \partial_t u(t,x) + \nu \Delta u(t,x) + \hat{H}(x, m(t), \nabla u(t,x)) \\ m(0,x) = m_0(x), \quad u(T,x) = g(x, m(T)) \end{cases}$$

with
$$\hat{H}(x, m, q) := \min_{a \in \mathbb{R}^k} \Big(f(x, m, a) + q \cdot b(x, m, a) \Big).$$

Deep Galerkin Method [Sirignano, Spiliopoulos]:

→ application to **MFGs**: see [Al-Aradi *et al.*; Carmona, L.; Cao, Guo, L.]

- replace unknown functions by deep NN → $m_{\theta_1}, u_{\theta_2}$
- try to minimize the squared residual → loss = $\int \int |\partial_t m_{\theta_1}(t,x) + \dots|^2 dt dx + \dots$
- by sampling points in the domain → sample (t_i, x_i)

Example: Crowd trading

Model of crowd trading [Cardaliaguet, Lehalle]:

$$\begin{cases} dS_t^{\bar{\mu}} = \gamma \bar{\mu}_t dt + \sigma dW_t & \text{(asset price)} \\ dQ_t^{v} = v_t dt & \text{(player's inventory)} \\ dX_t^{v,\bar{\mu}} = -v_t (S_t^{\bar{\mu}} + \kappa v_t) dt & \text{(player's wealth)} \end{cases}$$

Objective: given $\bar{\mu} = (\bar{\mu}_t)_t$, maximize

$$J(v; \bar{\mu}) = \mathbb{E}\left[X_T^{v,\bar{\mu}} + Q_T^{v} S_T^{\bar{\mu}} - A|Q_T^{v}|^2 - \phi \int_0^T |Q_t^{v}|^2 dt \right]$$

where: $\phi, A > 0 \Rightarrow$ penalty for holding inventory

Example: Crowd trading

Model of crowd trading [Cardaliaguet, Lehalle]:

$$\begin{cases} dS_t^{\bar\mu} = \gamma \bar\mu_t dt + \sigma dW_t & \text{(asset price)} \\ dQ_t^{v} = v_t dt & \text{(player's inventory)} \\ dX_t^{v,\bar\mu} = -v_t (S_t^{\bar\mu} + \kappa v_t) dt & \text{(player's wealth)} \end{cases}$$

Objective: given $\bar\mu = (\bar\mu_t)_t$, maximize

$$J(v; \bar\mu) = \mathbb{E}\left[X_T^{v,\bar\mu} + Q_T^{v} S_T^{\bar\mu} - A|Q_T^{v}|^2 - \phi \int_0^T |Q_t^{v}|^2 dt \right]$$

where: $\phi, A > 0 \Rightarrow$ penalty for holding inventory

Ansatz [Cartea, Jaimungal]: $V(t,x,s,q) = x + qsu(t,q)$, $\qquad v_t^*(q) = \frac{\partial_q u(t,q)}{2\kappa}$

where $u(\cdot)$ solves

$$-\gamma \bar\mu q = \partial_t u - \phi q^2 + \sup_{v} \{ v \partial_q u - \kappa v^2 \}, \qquad u(T,q) = -Aq^2$$

Example: Crowd trading

Model of crowd trading [Cardaliaguet, Lehalle]:

$$\begin{cases} dS_t^{\bar{\mu}} = \gamma \bar{\mu}_t dt + \sigma dW_t & \text{(asset price)} \\ dQ_t^{v} = v_t dt & \text{(player's inventory)} \\ dX_t^{v,\bar{\mu}} = -v_t (S_t^{\bar{\mu}} + \kappa v_t) dt & \text{(player's wealth)} \end{cases}$$

Objective: given $\bar{\mu} = (\bar{\mu}_t)_t$, maximize

$$J(v; \bar{\mu}) = \mathbb{E}\left[X_T^{v,\bar{\mu}} + Q_T^{v} S_T^{\bar{\mu}} - A|Q_T^{v}|^2 - \phi \int_0^T |Q_t^{v}|^2 dt \right]$$

where: $\phi, A > 0 \Rightarrow$ penalty for holding inventory

Ansatz [Cartea, Jaimungal]: $V(t,x,s,q) = x + qsu(t,q)$, $\qquad v_t^*(q) = \frac{\partial_q u(t,q)}{2\kappa}$

where $u(\cdot)$ solves

$$-\gamma \bar{\mu} q = \partial_t u - \phi q^2 + \sup_v \{ v \partial_q u - \kappa v^2 \}, \qquad u(T,q) = -Aq^2$$

Mean field term: at equilibrium

$$\bar{\mu}_t = \int v_t^*(q) m^*(t,dq) = \int \frac{\partial_q u^*(t,q)}{2\kappa} m^*(t,dq),$$

where m^* solves the KFP equation:

$$m(0,\cdot) = m_0, \qquad \partial_t m + \partial_q \left(m \frac{\partial_q u^*(t,q)}{2\kappa} \right) = 0$$

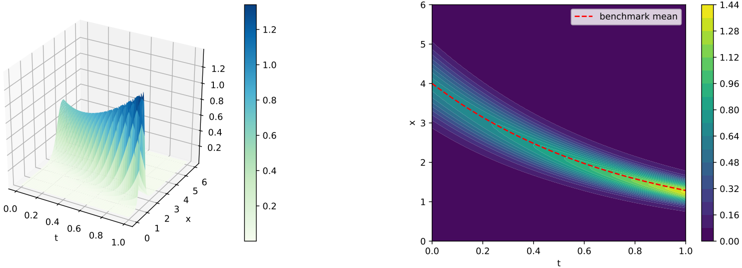
Example: Crowd trading

Forward-backward PDE system:

$$\begin{cases} -\gamma\bar{\mu}_t q = \partial_t u(t,q) - \phi q^2 + \dfrac{|\partial_q u(t,q)|^2}{4\kappa} \\ \partial_t m(t,q) + \partial_q \left(m(t,q) \dfrac{\partial_q u(t,q)}{2\kappa} \right) = 0 \\ \bar{\mu}_t = \displaystyle\int \dfrac{\partial_q u(t,q)}{2\kappa} m(t,q) dq \\ m(0,\cdot) = m_0, u(T,q) = -Aq^2. \end{cases}$$

Numerical Illustration: Crowd trading

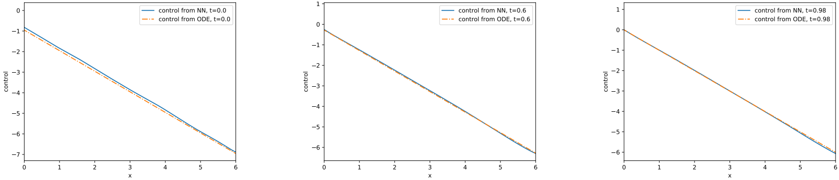
Trade crowding MFG example solved by DGM.



Evolution of the distribution m: surface (left) and contour (right).

Numerical Illustration: Crowd trading

Trade crowding MFG example solved by DGM.



Evolution of the optimal control v^* (3 different time steps).

Outline

Mean Field Framework

Generic Mean Field model: for a typical infinitesimal agent

- **Dynamics:** discrete time

$$X^{\alpha,\mu}_{n+1} = \varphi(X^{\alpha,\mu}_n, \alpha_n, \mu_n, \epsilon_{n+1}, \epsilon^0_{n+1}), \quad n \geq 0, \qquad X^{\alpha,\mu}_0 \sim \mu_0$$

  - ◇ $X^{\alpha,\mu}_n \in \mathcal{X} \subseteq \mathbb{R}^d$: state, $\alpha_n \in \mathcal{U} \subseteq \mathbb{R}^k$: action
  - ◇ $\epsilon_n \sim \nu$: idiosyncratic noise, $\epsilon^0_n \sim \nu^0$: common noise (random environment)
  - ◇ $\mu_n \in \mathcal{P}(\mathcal{X})$ is a state distribution

- **Cost:** $\mathbb{J}(\alpha; \mu) = \mathbb{E}_{\epsilon,\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n f\left(X^{\alpha,\mu}_n, \alpha_n, \mu_n\right)\right]$

Mean Field Framework

Generic Mean Field model: for a typical infinitesimal agent

- **Dynamics:** discrete time

$$X^{\alpha,\mu}_{n+1} = \varphi(X^{\alpha,\mu}_n, \alpha_n, \mu_n, \epsilon_{n+1}, \epsilon^0_{n+1}), \quad n \geq 0, \qquad X^{\alpha,\mu}_0 \sim \mu_0$$

◇ $X^{\alpha,\mu}_n \in \mathcal{X} \subseteq \mathbb{R}^d$: state, $\alpha_n \in \mathcal{U} \subseteq \mathbb{R}^k$: action

◇ $\epsilon_n \sim \nu$: idiosyncratic noise, $\epsilon^0_n \sim \nu^0$: common noise (random environment)

◇ $\mu_n \in \mathcal{P}(\mathcal{X})$ is a state distribution

- **Cost:** $\mathbb{J}(\alpha; \mu) = \mathbb{E}_{\epsilon,\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n f\left(X^{\alpha,\mu}_n, \alpha_n, \mu_n\right)\right]$

Two scenarios:

- **Cooperative (MFControl):** Find α^* minimizing $\alpha \mapsto J^{MFC}(\alpha) = \mathbb{J}(\alpha; \mu^\alpha)$

  where $\mu^\alpha_n = \mathbb{P}^0_{X^{\alpha,\mu^\alpha}_n}$

- **Non-Cooperative (MFGame):** Find $\hat{\alpha}$ minimizing $\alpha \mapsto J^{MFG}(\alpha; \hat{\mu}) = \mathbb{J}(\alpha; \hat{\mu})$

  where $\hat{\mu}_n = \mathbb{P}^0_{X^{\hat{\alpha},\hat{\mu}}_n}$

Q: *How to learn an optimal behavior when the model* (φ, f) *is not known?*

Two scenarios

1. Learning with cooperation

2. Learning with competition

1. Learning with cooperation
2. Learning with competition

From Optimal Control to Mean Field RL

$$\begin{tikzcd}[column sep=huge, row sep=huge]
\text{Mean-Field Control} \arrow[r, "\textit{unknown model}"] & \begin{array}{c}\text{Mean-Field}\\ \text{Reinforcement Learning}\end{array} \\
\text{Optimal Control} \arrow[u, "\substack{N\to\infty \\ MKV}"'] \arrow[r, "\textit{unknown model}"'] & \text{Reinforcement Learning} \arrow[u, "\substack{N\to\infty \\ MKV}"]
\end{tikzcd}$$

Approximate Policy Gradient

Idea 1: *Make the "direct approach" model-free*

Policy Gradient (PG) to minimize $J(\theta)$

- Control $\approx$ parameterized function
- Look for the optimal parameter θ^*
- Perform gradient descent on the space of parameters

Approximate Policy Gradient

Idea 1: *Make the "direct approach" model-free*

Policy Gradient (PG) to minimize $J(\theta)$

- Control $\approx$ parameterized function
- Look for the optimal parameter θ^*
- Perform gradient descent on the space of parameters

Hierarchy of three situations, more and more complex:

(1) access to the exact **(mean field) model**: $\theta_{k+1} = \theta_k - \eta \nabla J(\theta_k)$

Approximate Policy Gradient

Idea 1: *Make the "direct approach" model-free*

Policy Gradient (PG) to minimize $J(\theta)$

- Control $\approx$ parameterized function
- Look for the optimal parameter θ^*
- Perform gradient descent on the space of parameters

Hierarchy of three situations, more and more complex:

(1) access to the exact **(mean field) model**: $\theta_{k+1} = \theta_k - \eta \nabla J(\theta_k)$

(2) access to a **mean field simulator**:
$\rightarrow$ idem + gradient estimation (0^{th}-order opt.): $\theta_{k+1} = \theta_k - \eta \widetilde{\nabla} J(\theta_k)$

Approximate Policy Gradient

Idea 1: *Make the "direct approach" model-free*

Policy Gradient (PG) to minimize $J(\theta)$

- Control $\approx$ parameterized function
- Look for the optimal parameter θ^*
- Perform gradient descent on the space of parameters

Hierarchy of three situations, more and more complex:

(1) access to the exact **(mean field) model**: $\theta_{k+1} = \theta_k - \eta \nabla J(\theta_k)$

(2) access to a **mean field simulator**:
$\rightarrow$ idem + gradient estimation (0^{th}-order opt.): $\theta_{k+1} = \theta_k - \eta \widetilde{\nabla} J(\theta_k)$

(3) access to a N-agent **population simulator**:
$\rightarrow$ idem + error on mean $\approx$ empirical mean (LLN): $\theta_{k+1} = \theta_k - \eta \widetilde{\nabla}^N J(\theta_k)$

Approximate Policy Gradient

Idea 1: *Make the "direct approach" model-free*

Policy Gradient (PG) to minimize $J(\theta)$

- Control $\approx$ parameterized function
- Look for the optimal parameter θ^*
- Perform gradient descent on the space of parameters

Hierarchy of three situations, more and more complex:

(1) access to the exact **(mean field) model**: $\theta_{k+1} = \theta_k - \eta \nabla J(\theta_k)$

(2) access to a **mean field simulator**:
$\rightarrow$ idem + gradient estimation (0^{th}-order opt.): $\theta_{k+1} = \theta_k - \eta \widetilde{\nabla} J(\theta_k)$

(3) access to a N-agent **population simulator**:
$\rightarrow$ idem + error on mean $\approx$ empirical mean (LLN): $\theta_{k+1} = \theta_k - \eta \widetilde{\nabla}^N J(\theta_k)$

Theorem: For **Linear-Quadratic** MFC [Carmona, L., Tan'19]

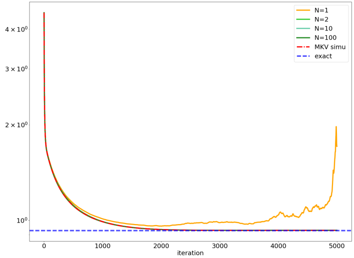
In each case, convergence holds at a linear rate:

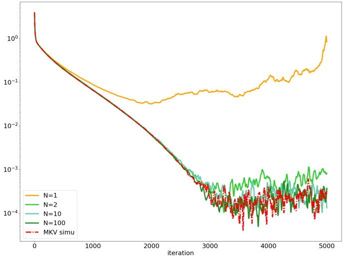
Taking $k \approx \mathcal{O}\big(\log(1/\epsilon)\big)$ is sufficient to ensure $J(\theta_k) - J(\theta^*) < \epsilon$.

Proof: builds on [Fazel et al.'18], analysis of perturbation of Riccati equations

Example: Linear dynamics, quadratic costs of the type:

$$f(x, \mu, v) = \underbrace{(\bar{\mu} - x)^2}_{\text{distance to mean position}} + \underbrace{v^2}_{\text{cost of moving}}, \qquad \bar{\mu} = \underbrace{\int \mu(\xi) d\xi}_{\text{mean position}},$$





Value of the MF cost

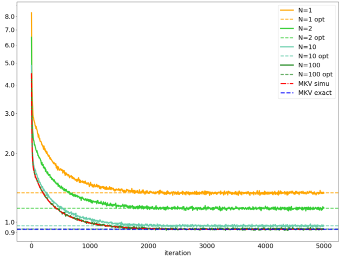
Rel. err. on MF cost

MF cost = cost in the mean field problem

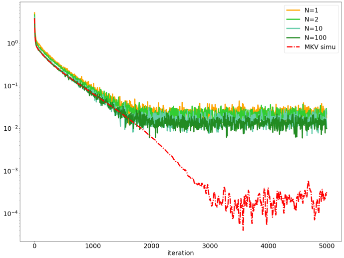
Numerical Illustration: LQMFC [Carmona, L., Tan'19]

Example: Linear dynamics, quadratic costs of the type:

$$f(x,\mu,v) = \underbrace{(\bar{\mu} - x)^2}_{\text{distance to mean position}} + \underbrace{v^2}_{\text{cost of moving}}, \qquad \bar{\mu} = \underbrace{\int \mu(\xi)d\xi}_{\text{mean position}},$$



Value of the social cost



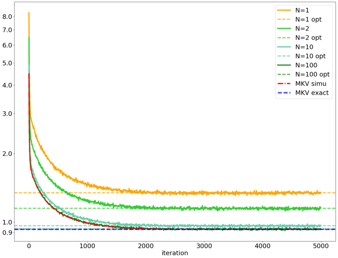
Rel. err. on social cost

Social cost = average over the N-agents

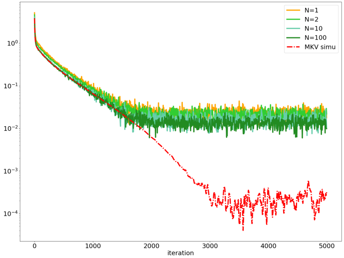
Numerical Illustration: LQMFC [Carmona, L., Tan'19]

Example: Linear dynamics, quadratic costs of the type:

$$f(x,\mu,v) = \underbrace{(\bar\mu - x)^2}_{\text{distance to mean position}} + \underbrace{v^2}_{\text{cost of moving}}, \qquad \bar\mu = \underbrace{\int \mu(\xi)d\xi}_{\text{mean position}},$$



Value of the social cost



Rel. err. on social cost

Social cost = average over the N-agents

Main take-away:

Trying to learn the mean-field regime solution can be efficient even for N small

Q: *Beyond the LQ setting?*
Idea 2: *Generalize Q-learning to the mean-field setting*

Q: *Beyond the LQ setting?*

Idea 2: *Generalize Q-learning to the mean-field setting*

$$\alpha^* \in \underset{\alpha}{\operatorname{argmin}}\, J^{MFC}(\alpha) = \mathbb{E}_{\epsilon,\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n f\left(X_n^{\alpha}, \alpha_n, \mu_n^{\alpha}\right)\right], \qquad \mu_n^{\alpha} = \mathbb{P}^0_{X_n^{\alpha}}$$

Q: *Beyond the LQ setting?*
Idea 2: *Generalize Q-learning to the mean-field setting*

$$\alpha^* \in \underset{\alpha}{\operatorname{argmin}}\, J^{MFC}(\alpha) = \mathbb{E}_{\epsilon,\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n f\left(X_n^\alpha, \alpha_n, \mu_n^\alpha\right)\right], \qquad \mu_n^\alpha = \mathbb{P}^0_{X_n^\alpha}$$

$$= \mathbb{E}_{\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n \underbrace{\int_{\mathcal{X}\times\mathcal{U}} f\left(x, a, \mu_n^\alpha\right) \mu_n^\alpha(dx, da)}_{\text{function of } \mu_n^\alpha}\right]$$

Q: *Beyond the LQ setting?*
Idea 2: *Generalize Q-learning to the mean-field setting*

$$\alpha^* \in \underset{\alpha}{\operatorname{argmin}}\, J^{MFC}(\alpha) = \mathbb{E}_{\epsilon,\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n f\left(X_n^\alpha, \alpha_n, \mu_n^\alpha\right)\right], \qquad \mu_n^\alpha = \mathbb{P}^0_{X_n^\alpha}$$

$$= \mathbb{E}_{\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n \underbrace{\int_{\mathcal{X}\times\mathcal{U}} f\left(x, a, \mu_n^\alpha\right) \mu_n^\alpha(dx, da)}_{\text{function of } \mu_n^\alpha}\right]$$

Dynamic Programming Principle (DPP):

- via the "lifted" problem for the population distribution μ^α (social planner's optim.)
- value function = function of the distribution μ

Q: *Beyond the LQ setting?*

Idea 2: *Generalize Q-learning to the mean-field setting*

$$\alpha^* \in \underset{\alpha}{\operatorname{argmin}}\, J^{MFC}(\alpha) = \mathbb{E}_{\epsilon,\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n f\left(X_n^\alpha, \alpha_n, \mu_n^\alpha\right)\right], \qquad \mu_n^\alpha = \mathbb{P}^0_{X_n^\alpha}$$

$$= \mathbb{E}_{\epsilon^0}\left[\sum_{n=0}^{\infty} \gamma^n \underbrace{\int_{\mathcal{X}\times\mathcal{U}} f\left(x, a, \mu_n^\alpha\right) \mu_n^\alpha(dx, da)}_{\text{function of } \mu_n^\alpha}\right]$$

Dynamic Programming Principle (DPP):

- via the "lifted" problem for the population distribution μ^α (social planner's optim.)
- value function = function of the distribution μ

MFMDP and Dynamic Programming

Mean Field Markov Decision Process: $(\bar{\mathcal{S}}, \bar{\mathcal{A}}, \bar{p}, \bar{r}, \gamma)$, where:

- State space: $\bar{\mathcal{S}} = \mathcal{P}(\mathcal{X})$
- Action space: $\bar{\mathcal{A}} = \mathcal{P}(\mathcal{X} \times \mathcal{U})$
- Transition: $\mu' = \bar{\Phi}(\mu, \bar{a}, \epsilon^0) \sim \bar{p}(\mu, \bar{a})$
- Reward: $\bar{r}(\mu, \bar{a}) = -\int_{\mathcal{X}\times\mathcal{U}} f(x, a, \mu)\bar{a}(dx, da)$

Goal: max. $\bar{V}^{\bar{\pi}}(\mu) = \mathbb{E}\left[\sum_{n=0}^{\infty} \gamma^n \bar{r}\left(\mu_n^{\bar{\pi}}, \bar{a}_n\right)\right]$, $\bar{a}_n \sim \bar{\pi}(\cdot|\mu_n^{\bar{\pi}})$, $\mu_{n+1}^{\bar{\pi}} \sim \bar{p}(\cdot|\mu_n^{\bar{\pi}}, \bar{a}_n)$, $\mu_0^{\bar{\pi}} = \mu$

MFMDP and Dynamic Programming

Mean Field Markov Decision Process: $(\bar{\mathcal{S}}, \bar{\mathcal{A}}, \bar{p}, \bar{r}, \gamma)$, where:

- State space: $\bar{\mathcal{S}} = \mathcal{P}(\mathcal{X})$
- Action space: $\bar{\mathcal{A}} = \mathcal{P}(\mathcal{X} \times \mathcal{U})$
- Transition: $\mu' = \bar{\Phi}(\mu, \bar{a}, \epsilon^0) \sim \bar{p}(\mu, \bar{a})$
- Reward: $\bar{r}(\mu, \bar{a}) = -\int_{\mathcal{X}\times\mathcal{U}} f(x, a, \mu)\bar{a}(dx, da)$

Goal: max. $\bar{V}^{\bar{\pi}}(\mu) = \mathbb{E}\left[\sum_{n=0}^{\infty} \gamma^n \bar{r}\left(\mu_n^{\bar{\pi}}, \bar{a}_n\right)\right]$, $\bar{a}_n \sim \bar{\pi}(\cdot|\mu_n^{\bar{\pi}})$, $\mu_{n+1}^{\bar{\pi}} \sim \bar{p}(\cdot|\mu_n^{\bar{\pi}}, \bar{a}_n)$, $\mu_0^{\bar{\pi}} = \mu$

Theorem: DPP for MFMDP [Carmona, L., Tan'20]

$$\bar{V}^*(\mu) := \sup_{\bar{\pi}} \bar{V}^{\bar{\pi}}(\mu) = \sup_{\bar{\pi}} \left\{ \int_{\bar{\mathcal{A}}} \left[\bar{r}(\mu, \bar{a}) + \gamma\, \mathbb{E}\left[\bar{V}^*\left(\bar{\Phi}(\mu, \bar{a}, \epsilon^0) \right) \right] \right] \bar{\pi}(d\bar{a}|\mu) \right\},$$

under suitable conditions, where the sup is over a subset of $\{\bar{\pi} : \bar{\mathcal{S}} \to \mathcal{P}(\bar{\mathcal{A}})\}$
Likewise for **mean field state-action value function** $\bar{Q}^*$

Proof based on double lifting [Bertsekas, Shreve'78]

MFMDP and Dynamic Programming

Mean Field Markov Decision Process: $(\bar{\mathcal{S}}, \bar{\mathcal{A}}, \bar{p}, \bar{r}, \gamma)$, where:

- State space: $\bar{\mathcal{S}} = \mathcal{P}(\mathcal{X})$
- Action space: $\bar{\mathcal{A}} = \mathcal{P}(\mathcal{X} \times \mathcal{U})$
- Transition: $\mu' = \bar{\Phi}(\mu, \bar{a}, \epsilon^0) \sim \bar{p}(\mu, \bar{a})$
- Reward: $\bar{r}(\mu, \bar{a}) = -\int_{\mathcal{X}\times\mathcal{U}} f(x, a, \mu)\bar{a}(dx, da)$

Goal: max. $\bar{V}^{\bar{\pi}}(\mu) = \mathbb{E}\left[\sum_{n=0}^{\infty} \gamma^n \bar{r}\left(\mu_n^{\bar{\pi}}, \bar{a}_n\right)\right]$, $\bar{a}_n \sim \bar{\pi}(\cdot|\mu_n^{\bar{\pi}})$, $\mu_{n+1}^{\bar{\pi}} \sim \bar{p}(\cdot|\mu_n^{\bar{\pi}}, \bar{a}_n)$, $\mu_0^{\bar{\pi}} = \mu$

Theorem: DPP for MFMDP [Carmona, L., Tan'20]

$$\bar{V}^*(\mu) := \sup_{\bar{\pi}} \bar{V}^{\bar{\pi}}(\mu) = \sup_{\bar{\pi}} \left\{ \int_{\bar{\mathcal{A}}} \left[\bar{r}(\mu, \bar{a}) + \gamma\, \mathbb{E}\left[\bar{V}^*\left(\bar{\Phi}(\mu, \bar{a}, \epsilon^0) \right) \right] \right] \bar{\pi}(d\bar{a}|\mu) \right\},$$

under suitable conditions, where the sup is over a subset of $\{\bar{\pi} : \bar{\mathcal{S}} \to \mathcal{P}(\bar{\mathcal{A}})\}$
Likewise for **mean field state-action value function** $\bar{Q}^*$

Proof based on double lifting [Bertsekas, Shreve'78]

DPPs for MFC: [L., Pironneau; Pham *et al.*; Gast *et al.*; Guo *et al.*; Possamai *et al.*;. . .]

Here: discrete time, infinite horizon, common noise, feedback controls.
$\rightarrow$ well-suited for **RL** $\rightarrow$ Mean-field Q-learning algorithm

Two scenarios

1. Learning with cooperation

2. Learning with competition

Picard fixed-point iterations:

$\mu^{(k)} \mapsto \alpha^{(k+1)} \mapsto \mu^{(k+1)}$

- $\alpha^{(k+1)}$ best response against $\mu^{(k)}$
- $\mu^{(k+1)}$ induced by $\alpha^{(k+1)}$

$\rightarrow$ *Convergence typically relies on strict contraction* [Caines *et al.*; Guo *et al.*; . . .]

Fictitious Play for MFG

Picard fixed-point iterations:

$\mu^{(k)} \mapsto \alpha^{(k+1)} \mapsto \mu^{(k+1)}$

- $\alpha^{(k+1)}$ best response against $\mu^{(k)}$
- $\mu^{(k+1)}$ induced by $\alpha^{(k+1)}$

$\rightarrow$ *Convergence typically relies on strict contraction* [Caines *et al.*; Guo *et al.*; …]

Fictitious Play [Brown'51; Robinson'51; …; Cardaliaguet, Hadikhanloo'15]

$\bar{\mu}^{(k)} \mapsto \alpha^{(k+1)} \mapsto \mu^{(k+1)} \mapsto \bar{\mu}^{(k+1)}$

- $\alpha^{(k+1)}$ best response against $\bar{\mu}^{(k)}$
- $\mu^{(k+1)}$ induced by $\alpha^{(k+1)}$
- $\bar{\mu}^{(k+1)} = \frac{k}{k+1}\bar{\mu}^{(k)} + \frac{1}{k+1}\mu^{(k+1)} = \frac{1}{k+1}\sum_{\ell=1}^{k+1}\mu^{(\ell)}$

$\rightarrow$ *Convergence typically under monotonicity condition*

Fictitious Play for MFG

Picard fixed-point iterations:

$\mu^{(k)} \mapsto \alpha^{(k+1)} \mapsto \mu^{(k+1)}$

- $\alpha^{(k+1)}$ best response against $\mu^{(k)}$
- $\mu^{(k+1)}$ induced by $\alpha^{(k+1)}$

$\rightarrow$ *Convergence typically relies on strict contraction* [Caines *et al.*; Guo *et al.*; ...]

Approximate Fictitious Play

- $\widetilde{\alpha}^{(k+1)}$ approximate best response against $\bar{\mu}^{(k)}$
- $\mu^{(k+1)}$ induced by $\widetilde{\alpha}^{(k+1)}$
- $\bar{\mu}^{(k+1)} = \frac{k}{k+1}\bar{\mu}^{(k)} + \frac{1}{k+1}\mu^{(k+1)} = \frac{1}{k+1}\sum_{\ell=1}^{k+1}\mu^{(\ell)}$

$\rightarrow$ *Convergence typically under monotonicity condition*

Theorem: Error propagation [Elie, Pérolat, L., Geist, Pietquin, AAAI'20]

Under Lasry-Lions monotonicity condition,

$$(\widetilde{\alpha}^{(k)}, \bar{\mu}^{(k)}) \xrightarrow[k\to+\infty]{} (\epsilon, \delta)\text{-Nash equilibrium}$$

Fictitious Play for MFG

Picard fixed-point iterations:

$\mu^{(k)} \mapsto \alpha^{(k+1)} \mapsto \mu^{(k+1)}$

- $\alpha^{(k+1)}$ best response against $\mu^{(k)}$
- $\mu^{(k+1)}$ induced by $\alpha^{(k+1)}$

$\rightarrow$ *Convergence typically relies on strict contraction* [Caines *et al.*; Guo *et al.*; ...]

Approximate Fictitious Play

- $\widetilde{\alpha}^{(k+1)}$ approximate best response against $\bar{\mu}^{(k)}$
- $\mu^{(k+1)}$ induced by $\widetilde{\alpha}^{(k+1)}$
- $\bar{\mu}^{(k+1)} = \frac{k}{k+1}\bar{\mu}^{(k)} + \frac{1}{k+1}\mu^{(k+1)} = \frac{1}{k+1}\sum_{\ell=1}^{k+1}\mu^{(\ell)}$

$\rightarrow$ *Convergence typically under monotonicity condition*

Theorem: Error propagation [Elie, Pérolat, L., Geist, Pietquin, AAAI'20]

Under Lasry-Lions monotonicity condition,

$$(\widetilde{\alpha}^{(k)}, \bar{\mu}^{(k)}) \xrightarrow[k\to+\infty]{} (\epsilon, \delta)\text{-Nash equilibrium}$$

RL for $\widetilde{\alpha}^{(k+1)}$: standard MDP parameterized by $\bar{\mu}^{(k)}$

Continuous Time Fictitious Play

Fictitious Play [Cardaliaguet, Hadikhanloo'15]: $\bar{\mu}^{(k)} \mapsto \alpha^{(k+1)} \mapsto \mu^{(k+1)} \mapsto \bar{\mu}^{(k+1)}$, with

$$\frac{\bar{\mu}^{(k+1)} - \bar{\mu}^{(k)}}{k+1} = \frac{1}{k+1}\left(\mu^{(k+1)} - \bar{\mu}^{(k)}\right)$$

Continuous Time Fictitious Play

- averaged distribution dynamics: $t \geq 1$,

$$\frac{d}{dt}\bar{\mu}^{(t)} = \frac{1}{t}\left(\mu^{(t)} - \bar{\mu}^{(t)}\right)$$

  where $\mu^{(t)} =$ induced by BR against $\bar{\mu}^{(t)}$
- averaged (mixed) policy dynamics: $\bar{\pi}^{(t)}$ generating $\bar{\mu}^{(t)}$

$\rightarrow$ *Rate of convergence*

Continuous Time Fictitious Play

Fictitious Play [Cardaliaguet, Hadikhanloo'15]: $\bar{\mu}^{(k)} \mapsto \alpha^{(k+1)} \mapsto \mu^{(k+1)} \mapsto \bar{\mu}^{(k+1)}$, with

$$\frac{\bar{\mu}^{(k+1)} - \bar{\mu}^{(k)}}{k+1} = \frac{1}{k+1}\left(\mu^{(k+1)} - \bar{\mu}^{(k)}\right)$$

Continuous Time Fictitious Play

- averaged distribution dynamics: $t \geq 1$,

$$\frac{d}{dt}\bar{\mu}^{(t)} = \frac{1}{t}\left(\mu^{(t)} - \bar{\mu}^{(t)}\right)$$

  where $\mu^{(t)}$ = induced by BR against $\bar{\mu}^{(t)}$
- averaged (mixed) policy dynamics: $\bar{\pi}^{(t)}$ generating $\bar{\mu}^{(t)}$

$\rightarrow$ *Rate of convergence*

Theorem: Convergence Rate [Perrin, Pérolat, L., Geist, Elie, Pietquin, NeurIPS'20]

Under Lasry-Lions monotonicity condition,

$$\mathcal{E}(\bar{\pi}^{(t)}) = O\left(1/t\right)$$

Exploitability: $\mathcal{E}(\pi) = \max_{\pi'} J(\pi'; \mu^{\pi}) - J(\pi; \mu^{\pi})$

Example: Systemic Risk

Systemic risk model of [Carmona, Fouque, Sun] with LQ structure & **common noise**:

$$J(a;(m_n)_n) = -\mathbb{E}\Big[\sum_{n=0}^{N_T}\Big(a_n^2 \underbrace{-q a_n(m_n - X_n)}_{\substack{\text{borrow if } X_n < m_n \\ \text{lend if } X_n > m_n}} + \kappa(m_n - X_n)^2\Big) + c(m_{N_T} - X_{N_T})^2\Big]$$

Subj. to: $X_{n+1} = X_n + [K(m_n - X_n) + a_n] + \epsilon_{n+1} + \epsilon^0_{n+1}$

At equilibrium: $m_n = \mathbb{E}[X_n|\epsilon^0]$, $n \geq 0$

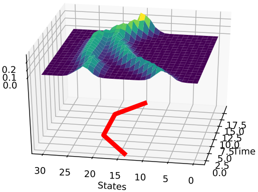
Example: Systemic Risk

Systemic risk model of [Carmona, Fouque, Sun] with LQ structure & **common noise**:

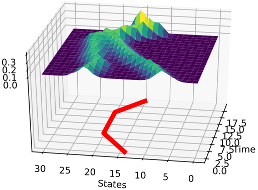
$$J(a;(m_n)_n) = -\mathbb{E}\Big[\sum_{n=0}^{N_T}\Big(a_n^2 \underbrace{-q a_n (m_n - X_n)}_{\substack{\text{borrow if } X_n < m_n \\ \text{lend if } X_n > m_n}} + \kappa (m_n - X_n)^2\Big) + c(m_{N_T} - X_{N_T})^2\Big]$$

Subj. to: $X_{n+1} = X_n + [K(m_n - X_n) + a_n] + \epsilon_{n+1} + \epsilon^0_{n+1}$

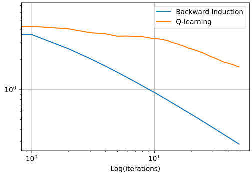
At equilibrium: $m_n = \mathbb{E}[X_n | \epsilon^0],\ n \geq 0$



Exact solution



Fictitious Play & RL



Exploitability

Outline

Summary

Q1: *How can we solve large games with complex structures?*

Part 1: Solving mean-field problems with deep learning

- Direct approach
- FB systems of SDEs
- FB systems of PDEs

Q2: *How can large populations learn to coordinate?*

Part 2: Reinforcement learning with mean-field interactions

- Learning with cooperation: PG / mean-field Q-learning
- Learning with competition: Fictitious Play & RL

Main directions for future research:

1. Bidirectional links with machine learning

- Machine learning for large population games
- Mean field view on machine learning

2. Breaking the barrier of homogeneity & symmetry

- Variety of agents
- Networked interactions
- PDEs on the Wasserstein space

One last example of MFG: Walk for the climate, Paris

Thank you