Mean field games and related tools: State of research and applications

Alexander Aurell

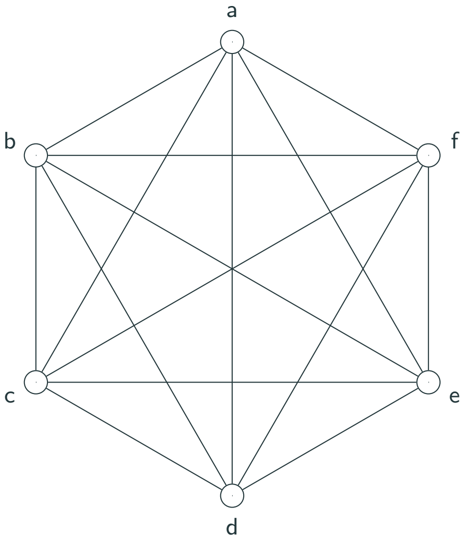
Silo AI Research Club

December 2, 2021

Introduction:
What is a mean field game?

What is a mean field game?

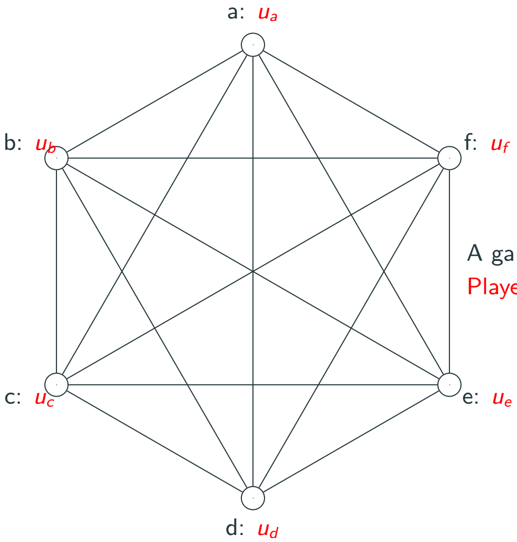
A finite game. Nodes represent players, edges interactions.



A game with 6 players, $a, \ldots, f$, over a complete graph (each is influenced by all).

What is a mean field game?

To each player we associate an **action**.

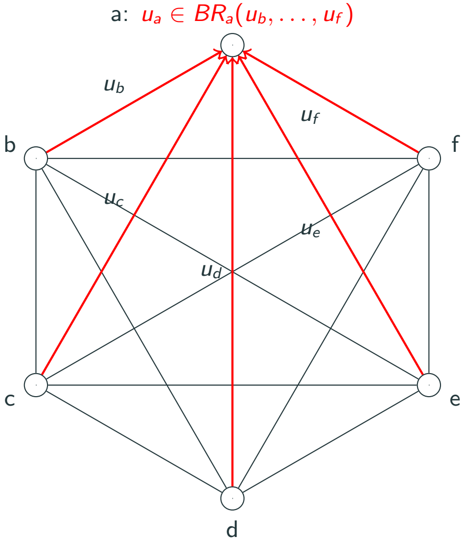


A game with 6 players: $a, \ldots, f$.

Players take actions: $u_a, \ldots, u_f$

What is a mean field game?

Game theory: the study of strategic interactions among rational agents.



A game with 6 players: $a, \ldots, f$.
Players take actions: $u_a, \ldots, u_f$
Aims to find the best response (defined from a preference structure).

What is a mean field game?

An **action profile** $(u_a, \ldots, u_f)$ is a **Nash equilibrium** when

no player can gain from a unilateral deviation from the profile.

$\rightarrow$ Nash 1950, 1951,

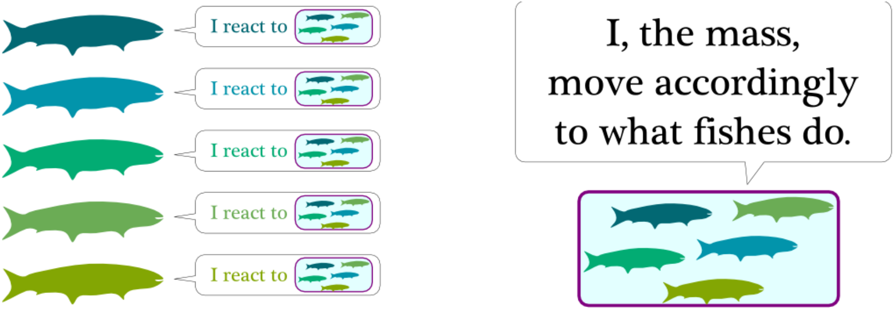
Most common setups lead to a system of coupled (in)equalities,
growing quickly with the number of players in the game.

What hopes do we have to compute Nash equilibria
when the number of players in the game is **very large**?

What is a mean field game?

Mean field game (MFG) theory restates game theory as an
interaction of each player with the distribution of other players.

What is a mean field game?



The mean field game **hypothesis:**
The player population is homogeneous: each player is representative.

The mean field game **consistency condition:**
The best response of the representative player reacting to the mass behavior applied to all agents generates the same mass behavior.

The MFG equation

The MFG equation: Problem formulation – Vanilla Optimal Control

Action: $\alpha : t \mapsto \alpha_t \in A$

State equation (SDE):

$$dX_t = b(t, X_t, \alpha_t)dt + \sigma(t, X_t, \alpha_t)dW_t$$
$$X_0 \sim m_0$$

(Finite horizon) Cost of using α:

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

The MFG equation: Problem formulation

Action: $\alpha : t \mapsto \alpha_t \in A$ Evolution of the mass: $m : t \mapsto m_t$

State equation (SDE):

$$dX_t = b(t, X_t, m_t, \alpha_t)dt + \sigma(t, X_t, m_t, \alpha_t)dW_t$$
$$X_0 \sim m_0$$

(Finite horizon) Cost of using α facing the mass m:

$$J(\alpha; m) = \mathbb{E}\left[\int_0^T f(t, X_t, \alpha_t, m_t)dt + g(X_T, m_T)\right]$$

The MFG equation: Problem formulation

Action: $\alpha : t \mapsto \alpha_t \in A$ Evolution of the mass: $m : t \mapsto m_t$

State equation (SDE):

$$dX_t = b(t, X_t, m_t, \alpha_t)dt + \sigma(t, X_t, m_t, \alpha_t)dW_t$$
$$X_0 \sim m_0$$

(Finite horizon) Cost of using α facing the mass m:

$$J(\alpha; m) = \mathbb{E}\left[\int_0^T f(t, X_t, \alpha_t, m_t)dt + g(X_T, m_T)\right]$$

Consistency condition:

1. Fix a distribution flow $m : t \mapsto m_t$
2. Solve the stochastic control problem $\hat{a} = \arg\min_\alpha J(\alpha; m)$
3. Determine the distribution flow $\hat{m} : t \mapsto \hat{m}_t$ such that $\hat{m}_t$ is the distribution of $\hat{X}_t = X_t(\hat{\alpha}, \hat{m})$ at all times t:

$$\hat{m}_t = \mathcal{L}(\hat{X}_t)$$

The MFG equation: McKean-Vlasov SDE

Consistency condition:

1. Fix a distribution flow $m : t \mapsto m_t$
2. Solve the stochastic control problem $\hat{a} = \arg\min_\alpha J(\alpha; m)$
3. Determine the distribution flow $\hat{m} : t \mapsto \hat{m}_t$ such that $\hat{m}_t = \mathcal{L}(\hat{X}_t)$.

The MFG solution is a **fixed point**!

$$\text{Find } \hat{a} \text{ solving step 2 given } \hat{m} \rightleftarrows \text{Find } \hat{m} \text{ solving step 3 given } \hat{\alpha}$$

Solving step 2 (Maximum principle, HJB,...) then enforcing the fixed point leads to a **McKean-Vlasov forward-backward SDE**: the MFG equation.

There is an analytical counterpart in the form of a coupled forward-backward nonlinear PDE system: Kolmogorov (forward in time) and HJB (backward in time).

Research summary

History, development, and horizon

Research

→ Related work in economic theory (Aumann 196x, Jovanovic & Rosenthal 1989)

→ **Connecting MFGs with game theory** (Lasry & Lions 2006, Huang et al 2006)

→ First extensions: Discrete space/time, infinite time horizon, extended MFGs

Research

- $\rightarrow$ Related work in economic theory (Aumann 196x, Jovanovic & Rosenthal 1989)
- $\rightarrow$ **Connecting MFGs with game theory** (Lasry & Lions 2006, Huang et al 2006)
- $\rightarrow$ First extensions: Discrete space/time, infinite time horizon, extended MFGs

- $\rightarrow$ Other state equations: Boundary conditions, degeneracy, common noise
- $\rightarrow$ Other types of players: Multi-population, Major players
- $\rightarrow$ Mean-field optimal control, mean-field type games
- $\rightarrow$ Numerical methods: classical, machine learning
- $\rightarrow$ Optimal stopping, sparse control, impulse control

Research

→ Related work in economic theory (Aumann 196x, Jovanovic & Rosenthal 1989)

→ **Connecting MFGs with game theory** (Lasry & Lions 2006, Huang et al 2006)

→ First extensions: Discrete space/time, infinite time horizon, extended MFGs

→ Other state equations: Boundary conditions, degeneracy, common noise

→ Other types of players: Multi-population, Major players

→ Mean-field optimal control, mean-field type games

→ Numerical methods: classical, machine learning

→ Optimal stopping, sparse control, impulse control

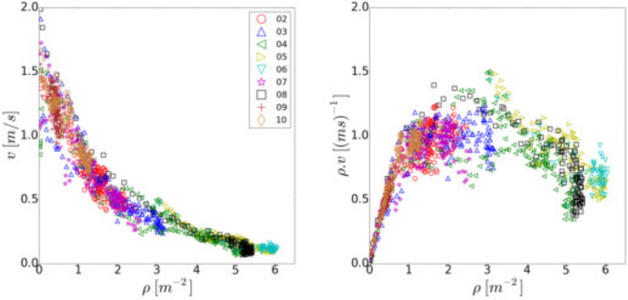
→ Other types of equilibria, partial information

→ Heterogeneous populations, games on random graphs, Graphon games

Applications of MFGs

Applications

→ **Crowd motion (evacuation, public space safety)**



(a) Density-velocity (b) Density-flow

Figure 1: *Fundamental diagrams for pedestrian traffic flow in controlled experiments* Wang et al 2019

→ Epidemiology (vaccination, testing, social distancing policy optimization)
→ Trading (optimal execution, HFT, crypto)
→ Markets (price formation, resource extraction)
→ Networks (communication, coordination of electric loads)

Applications

→ Crowd motion (evacuation, public space safety)

→ **Epidemiology (vaccination, testing, social distancing policy optimization)**

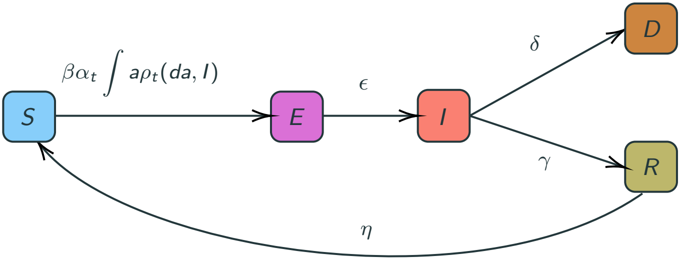


Figure 2: *Optimal incentives to mitigate epidemics: A Stackelberg mean field game approach*, Aurell et al 2021

→ Trading (optimal execution, HFT, crypto)

→ Markets (price formation, resource extraction)

→ Networks (communication, coordination of electric loads)

Applications

→ Crowd motion (evacuation, public space safety)

→ Epidemiology (vaccination, testing, social distancing policy optimization)

→ **Trading (optimal execution, HFT, crypto)**

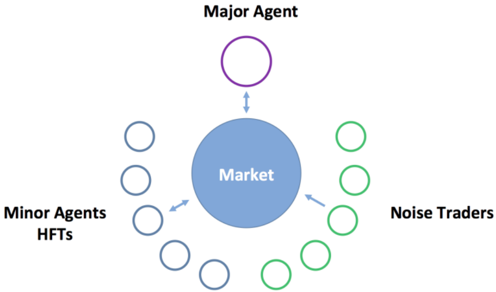


Figure 3: *Mean-Field Game Strategies for Optimal Execution*, Huang et al 2019

→ Resource management (smart energy grids, oil extraction)

→ and more (communication networks, ...)

Applications

- → Crowd motion (evacuation, public space safety)
- → Epidemiology (vaccination, testing, social distancing policy optimization)
- → Trading (optimal execution, HFT, crypto)
- → **Resource management (smart energy grids, oil extraction)**

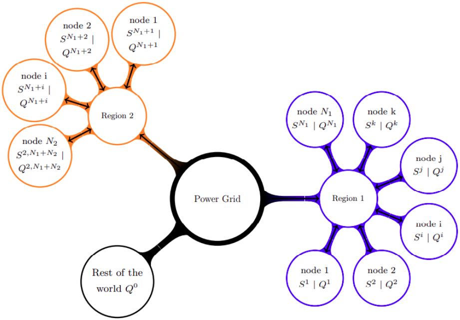


Figure 4: *An Extended Mean Field Game for Storage in Smart Grids*, Alasseur et al 2019

- → and more (communication networks, cyber security,)

Applications

- → Crowd motion (evacuation, public space safety)
- → Epidemiology (vaccination, testing, social distancing policy optimization)
- → Trading (optimal execution, HFT, crypto)
- → Resource management (smart energy grids, oil extraction)
- → **And more (wireless communication, systemic risk, . . .)**

What about machine learning?

Deep learning, Reinforcement learning, GANs

Deep learning

Recall: finding an MFG equilibrium can be reduced to solving a McKean-Vlasov forward-backward system of SDEs (MKV FBSDE) which reads

$$dX_t = B(t, X_t, \mathcal{L}(X), Y_t)dt + \sigma dW_t$$
$$dY_t = -F(t, X_t, \mathcal{L}(X_t), Y_t, \sigma^T Z_t)dt + Z_t dW_t$$

with boundary conditions $X_0 \sim m_0$ and $Y_T = G(X_T, \mathcal{L}(X_T))$.

The backward equation for Y is forced on us by the optimization ("step 2") and makes solving the problem VERY HARD!

Path to numerically tractable problem:

1. Time-change ("Sannikov's Trick")
2. Discretization (Monte Carlo approximation)
3. Parameterization (Deep Learning)

1.

Controller chooses initial condition y_0 and volatility z of Y to reach the target.

Replace MVK FBSDE with the optimization problem:

$$J_{FBSDE}(y_0, z) = E\left[\|Y_T^{y_0,z} - G(X_T^{y_0,z}, \mathcal{L}(X_T^{y_0,z}))\|^2\right]$$

subject to

$$dX_t^{y_0,z} = B(t, X_t^{y_0,z}, \mathcal{L}(X_t^{y_0,z}), Y_t^{y_0,z})dt + \sigma dW_t, \quad X_0^{y_0,z} \sim m_0,$$
$$dY_t^{y_0,z} = -F(t, X_t^{y_0,z}, \mathcal{L}(X_t^{y_0,z}), Y_t^{y_0,z}, \sigma^T z_t)dt + z_t dW_t, \quad Y_0^{y_0,z} = y_0$$

2.

Discretize the state distribution:

$$J^N_{FBSDE}(y_0, z) = \frac{1}{N}\sum_{i=1}^{N} \| Y_T^{i,y_0,z} - G(X_T^{i,y_0,z}, m_T^{N,y_0,z}) \|^2$$

subject to (for $i = 1, \ldots, N$)

$$\begin{aligned} dX_t^{i,y_0,z} &= B(t, X_t^{i,y_0,z}, m_t^{N,y_0,z}, Y_t^{i,y_0,z})dt + \sigma dW_t^i, \quad X_0^{i,y_0,z} \sim m_0, \\ dY_t^{i,y_0,z} &= -F(t, X_t^{i,y_0,z}, m_t^{N,y_0,z}, Y_t^{i,y_0,z}, \sigma^T z_t)dt + z_t dW_t^i, \quad Y_0^{i,y_0,z} = y_0 \end{aligned}$$

where

$$m_t^{N,y_0,z} = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^{i,y_0,z}}$$

Deep learning

3.
Replace y_0, z with neural networks y_0^θ, z^θ parameterized by θ

$$J^N_{FBSDE}(\theta) = \frac{1}{N}\sum_{i=1}^N \| Y_T^{i,\theta} - G(X_T^{i,\theta}, m_T^{N,\theta})\|^2$$

subject to (for $i = 1, \dots, N$)

$$\begin{aligned} dX_t^{i,\theta} &= B(t, X_t^{i,\theta}, m_t^{N,\theta}, Y_t^{i,\theta})dt + \sigma dW_t^i, \quad X_0^{i,\theta} \sim m_0, \\ dY_t^{i,\theta} &= -F(t, X_t^{i,\theta}, m_t^{N,\theta}, Y_t^{i,\theta}, \sigma^T z_t^\theta)dt + z_t^\theta dW_t^i, \quad Y_0^{i,\theta} = y_0^\theta \end{aligned}$$

4. Discretize time, 5. Setup for SGD, 6. ...

Deep learning

→ SGD for MKV FBSDEs, Deep Galerkin methods for MF PDEs (Carmona & Lauriere, 2021)

→ SGD for pure-jump MKV FBSDEs (Aurell et al, 2021)

Vanilla:

$$\text{Expected reward: } \mathbb{E}\left[\sum_{n=0}^{N_T-1} f(X_{t_n}^{\alpha}, \alpha_{t_n}) + g(X_{t_n}^{\alpha})\right]$$

$$\text{Transition probability: } \mathbb{P}(X_{t_{n=1}}^{\alpha} = x' | X_{t_n}^{\alpha} = x, \alpha_{t_n} = a) = p(x'|x, a)$$

The optimal Q-function:

$$Q_{N_T}^*(x, a) = g(x)$$

$$Q_n^*(x, a) = f(x, a) + \sum_{x' \in \mathcal{X}} p(x'|x, a) \min_{a'} Q_{n+1}^*(x', a')$$

MFG:

Expected reward: $\mathbb{E}\left[\sum_{n=0}^{N_T-1} f(X_{t_n}^{\alpha,\mu}, \alpha_{t_n}, \mu_{t_n}) + g(X_T^{\alpha,\mu}, \mu_T)\right]$

Transition probability: $\mathbb{P}(X_{t_{n+1}}^{\alpha} = x' | X_{t_n}^{\alpha} = x, \alpha_{t_n} = a, \mu_{t_n} = m) = p(x'|x, a, m)$

The optimal Q-function (with $\mu = (\mu_{t_n})_{n=0\,T}^{N}$ frozen):

$$Q^*_{N_T,\mu}(x, a) = g(x, \mu_T)$$

$$Q^*_{n,\mu}(x, a) = f(x, a, \mu_{t_n}) + \sum_{x' \in \mathcal{X}} p(x'|x, a, \mu_{t_n}) \min_{a'} Q^*_{n+1,\mu}(x', a')$$

As expected: the optimal control $\hat{\alpha}_{t_n}(x) = \arg\max_a Q^*_{n,\mu}(x, a)$ depends on μ.

The system is not closed! Impose the consistency condition:

$$\hat{\mu}_{t_n} = \mathcal{L}(X_{t_n}^{\hat{\alpha}})$$

Simple approach:

1. Start with an initial guess $\mu^{(0)}$
2. Solve the backward equation $Q^{(k+1)} = Q^*_{\mu^{(k)}}$:

$$
\begin{aligned}
Q^{(k+1)}_{N_T}(x,a) &= g(x,\mu^{(k)}_T) \\
Q^{(k+1)}_{n}(x,a) &= f(x,a,\mu^{(k)}_{t_n}) + \sum_{x' \in \mathcal{X}} p(x'|x,a,\mu^{(k)}_{t_n}) \min_{a'} Q^*_{n+1}(x',a')
\end{aligned}
$$

3. Compute the optimal control given $\mu^{(k)}$:

$$
\hat{\alpha}^{(k+1)}_{t_n}(x) = Q^{(k+1)}_n(x,a)
$$

4. Solve the forward equation for the state distribution flow, assuming all agents use $\hat{\alpha}^{(k+1)}$:

$$
\begin{aligned}
\mu^{(k+1)}_{t_0}(x) &= \mu_{t_0}(x) \\
\mu^{(k+1)}_{t_{n+1}}(x) &= \sum_{x \in \mathcal{X}} p(x', x, \alpha^{(k+1)}_{t_n}(x'), \mu^{(k+1)}_{t_n})
\end{aligned}
$$

Reinforcement Learning

→ RL for stationary MFGs (Subramanian & Mahajan, 2019)

→ Mean field MDP and Q-learning (Carmona et al, 2021)

→ "Unified Two Timescales Mean Field Q-learning" (Angiuli et al, 2021)

→ "Mean Field PSRO" (Muller et al, 2021)

→ Entropy regularization (Cui & Koeppl, 2021)

Further connections with ML

Concave Utility Reinforcement Learning (CURL) extends RL from linear to concave utilities in the occupancy measure induced by the agent's policy.

Decision making in the face of a non-linear distribution dependence ... MFG?

"Our numerical illustrations suggest that it may be worth considering MFG algorithms for addressing CURL problems."
- Concave Utility Reinforcement Learning: the Mean-field Game viewpoint (Geist et al, 2021)

Generative adversarial networks (GAN): a minimax zero-sum two-player game with objective depending non-linearly on distributions.

A multipopulation MFG ... with two teams?

"MFGs have the structure of GANs, and GANs are MFGs under the Pareto Optimality criterion."
- Connecting GANs, mean-field games, and optimal transport (Cao et al, 2021)