NOTES ON MEAN FIELD GAMES //3

Guangyu Hou

Contents

1 Mean Field Games Theory

1.1 Routes From physics

- The main difference, indeed the challenge, is to take account not only of the ability of agents to take decisions, but especially the capacity for strategic interaction.
- It is no longer a statistic on the domain of particle states, but rather a statistic on the domain of agent states and hence on the domain of strategies and information.
- Mean field games theory proposes to use the tools of physics but to use them inside the classical economic axiomatic, to explain (and not only to describe) phenomenon.

1.2 Routes From Game Theory

- Things are simplified, though we are considering stuff from N-player games when N tends to infinity, at least for a wide range of games that are symmetrical as far as players are concerned, as the number of players increases. Each player is progressively lost in the crowd in the eyes of other players under the process.
- Each player has become infinitesimal amidst the mass of other players, and constructs his strategies from his own state and from the state of the infinite mass of his co-players, who in turn simultaneously construct their strategies in the same way.
- Players of the same kind can be interchanged without altering the game: a form of anonymity of contexts where nothing is dependent on the individual.

1.3 Routes From Economics

- Agents have little concerned about each others: everyone looks only to his own interest and to market prices.
- *Hypothesis of Rational expectations* (the only level at which the existence of others applies): A theory is viewed as credible only if each agent can check whether by putting himself in the place of others he would find the behavior predicted by the theory. Prices mediate all social interactions.
- Yet in many cases there are other economic effects which give rise to other interactions between agents: externality, public goods, etc. These interactions between agents are the main interests of economists. They want to understand how prices form through rational behaviors and the consequence of externality effects.
- (Also, economists are interested in the evolution of an economy and hence they have been spending a lot of time on anticipations and the way prices or, more generally, behaviors form in an intertemporal context.)
- With a forward/backward structure, we try not only to describe but also to explain a phenomenon using the economic toolbox of utility maximization and rational expectations.

1.4 Versus N-player modeling

The agents of mean field games theory are less sophisticated than the players of N-player game theory since they base their strategies only *on the statistical state of the mass of co-agents*.

Positive Effect: *Efficiency* (possibilities of deploying the power of differential calculus) and *Widening the field of application* (the modeling of the renewal of player generations, the use of the corrective term in $1/N$ allows to introduce a "social" dimension in regard to players when statistical data on other players emerge as fundamental constituents of individual strategies.).

1.5 Initial simple Model

1.5.1 Background

- A meeting is scheduled for a certain time t which only starts several minutes after the scheduled time.
- The actual time T when the meeting starts depends on the arrival of its participants.
- A rule sets the start of the meeting at the point when a certain quorum is reached. (which is a form of strategic interaction)

> Three times.
>
> - the scheduled time of the meeting t.
> - the time agent i planned to arrive τ^i, and the actual arrival time $\tilde{\tau}^i = \tau^i + \sigma^i \tilde{\epsilon}^i$ where $\tilde{\epsilon}^i$ is a normal noise with variance 1, specific to each agent.
> - note m_0 the distribution of σ^i in the population.
> - the actual time the meeting will start T

To decide one's intended arrival time τ^i, each will optimize a total cost $c(t, T, \tilde{\tau})$ which is assumed to be made of three components: reputation effect + personal inconvenience + waiting time cost. (simple version, convex)

- *reputation effect*: a cost of lateness in relation to scheduled time t, $c_1(t, T, \tilde{\tau}) = \alpha[\tilde{\tau} - t]_+$
- *personal inconvenience*: a cost of lateness in relation to the actual starting time of the meeting T, $c_2(t, T, \tilde{\tau}) = \beta[\tilde{\tau} - T]_+$
- *waiting time cost*: time lost waiting to reach time T, $c_3(t, T, \tilde{\tau}) = \gamma[T - \tilde{\tau}]_+$

1.5.2 Resolution

Each agent aims to minimize his expected total cost by assuming T to be known. T is a *priori* a random variable. But we are considering an infinite number of players, the "law of large numbers" will imply T (the mean field) is deterministic. For agent i, the problem is therefore:

$$\tau^i = \arg\min \mathbb{E}[c(t, T, \tilde{\tau}^i)], \;\; \tilde{\tau}^i = \tau^i + \sigma^i \tilde{\epsilon}^i$$

We want to show that individual choices fully generate the realization of this time T supposing T is known, i.e. we're going to show the existence of a fixed point and in this case, the point is T.
The problem starts with examine agents' individual choices (to minimize the expected total cost). By the first-order condition and a strictly monotonic cumulative distribution function, we prove the existence and uniqueness of τ^i, a function of (t, T, σ^i),

$$\alpha \mathcal{N}\left(\frac{\tau^i - t}{\sigma^i}\right) + (\beta + \gamma)\mathcal{N}\left(\frac{\tau^i - T}{\sigma^i}\right) = \gamma$$

Because of the continuum and the law of large numbers, the distribution is transported by the application $\sigma^i \mapsto \tilde{\tau}^i$. If we note F the as a deterministic cumulative distribution function of the agents' real arrival times, we can establish a rule on the real starting time T, depending on the function $F(\cdot)$.
We then have to prove the existence and uniqueness of a fixed point. But we have reached the following statement.

- Starting from a value T, we obtain agents' optimal strategies $(\tau^i(\cdot; T))_i \;\; T \mapsto (\tau^i(\cdot; T))_i$
- Actual arrival time is based on the optimal strategies but affected by noise $(\tau^i(\cdot; T))_i \mapsto (\tilde{\tau}^i(\cdot; T))_i$
- From the law of large numbers and the hypothesis of the independence of agents' uncertainties, these arrival times are distributed according to a deterministic $F \;\; (\tilde{\tau}^i(\cdot; T))_i \mapsto F$
- T is deduced from F by the meeting starting time rule $\;\; F \mapsto T^*(F)$

We reach the scheme $T^{**} : T \mapsto (\tau^i(\cdot; T))_i \mapsto (\tilde{\tau}^i(\cdot; T))_i \mapsto F = F(\cdot, T) \mapsto T^*(F)$

If $\alpha, \beta, \gamma > 0$ and if $0 \notin \overline{\text{supp}(m_0)}$, then T^{**} is a contraction mapping of $[t; +\infty]$, and there's unique solution T to our problem.

Differentiate with respect to T the FOC defines τ^i,

$$\frac{d\tau^i}{dT}\left[\alpha \mathcal{N}'\left(\frac{\tau^i - t}{\sigma^i}\right) + (\beta + \gamma)\mathcal{N}'\left(\frac{\tau^i - T}{\sigma^i}\right)\right] = (\beta + \gamma)\mathcal{N}'\left(\frac{\tau^i - T}{\sigma^i}\right)$$

Since 0 is not in the support of m_0, we know $\frac{d}{dT}\tau(t, \sigma; T) \leq k < 1$. Hence, $\forall T, s, h > 0$,

$$F(s; T + h) = \mathbb{P}(\tau^i(\sigma^i); T + h) + \sigma^i \epsilon^i \leq s) \geq \mathbb{P}(\tau^i(\sigma^i); T) + kh + \sigma^i \epsilon^i \leq s) = F(s - kh; T)$$

Therefore,

$$T^*(F(\cdot;T+h)) \leq T^*(F(\cdot - kh;T)) \leq T^*(F(\cdot;T)) + kh$$

$\Rightarrow T^{**}(T+h) - T^{**}(T) \leq khandthisprovestheresultthroughthecontractionmappingtheorem. It'snotdifficultbutnecessarytoprovethefollowingpropertiesofthesettingrule$ T*

- $\forall F(\cdot), T^*(F(\cdot)) \geq t$; the meeting never starts before t
- **(Monotony)** For two cumulative distribution functions $F(\cdot)$ and $G(\cdot)$, if $F(\cdot) \leq G(\cdot)$, then $T^*(F(\cdot)) \geq T^*(G(\cdot))$
- **(Sub-additivity)** $\forall s > 0$, $T^*(F(\cdot - s)) - T^*(F(\cdot)) \leq s$

1.6 Variants

We will talk about a "geographical" model, i.e. the agents are initially distributed in different places and must come to where the meeting is being held.
The interest of this variant is that it will show how *coupled forward/backward PDEs*, which are the core of mean field game theory (in continuous time, with a continuous state space), emerge.

2 Stochastic Optimal Control

2.1 Setup

On finite time horizon, $T > 0$ and agents choose actions from a closed set $A \subset \mathbb{R}^k$. We also have two **standing assumptions**: the drift and volatility functions are measurable and satisfy a uniform Liptschitz condition in x, the objective functions $f : \mathbb{R}^d \times A \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ are continuous and bounded.
The *state process* X is a d-dimensional stochastic process controlled by the A valued process α and whose dynamics are given by

$$\begin{aligned} dX_t &= b\left(X_t, \alpha_t\right) dt + \sigma\left(X_t, \alpha_t\right) dW_t \\ X_0 &= x \end{aligned}$$

The controller steers the process X by choosing the control α, which is a process taking values in A. The goal of the controller is to choose α to maximize

$$\mathbb{E}\left[\int_0^T f(X_t, \alpha_t) dr + g(X_T)\right],$$

where f is called the *running reward/objective function*, and g is the *terminal reward/objective function*.
Note. These are far from necessary and mostly given for convenience, and it's adimitedly that the theorems we will develop are typical and cover narrow models in practice.
We will consider the following two natural families of controls: 1. **Open loop**: We denote

by $\mathbb{A}$ the set of $\mathbb{F}$-progressively measurable processes $\alpha = (\alpha_t)_{t\geq 0}$ such that

$$\mathbb{E}\left[\int_0^T \left[|b\left(0,\alpha_t\right)|^2 + |\sigma\left(0,\alpha_t\right)|^2\right] dt\right] < \infty$$

Under the above assumptions, classical theory ensures that the state equation has a unique strong solution. **2. Markovian controls:** $\mathbb{A}_M \subset \mathbb{A}$ consist of the set of Markovian controls. That is, $\alpha \in \mathbb{A}_M$ if $\alpha \in \mathbb{A}$ and there exists a measurable function $\hat{\alpha} : [0,T] \times \mathbb{R}^d \to A$ such that $\alpha_t = \hat{\alpha}\left(t, X_t\right)$.

One uncommon but natural alternative would be to work with *path-dependent controls*, of the form $\alpha_t = \hat{\alpha}(t, (X_s)_{s\in[0,t]})$. The terms *closed-loop control* and *feedback control* are variously used by different authors as synonymous with *Markovian control* or even sometimes the latter notion of *path-dependent control*.

2.2 Dynamic programming

Now we will focus on approach to solving a dynamic optimization problem. Suppose that we know how to solve the control problem when it "starts" at a time $t \in (0,T)$ and from any initial state $x \in \mathbb{R}^d$, and we denote the resulting optimal value by $v(t,x)$. Then, to solve the control problem starting from an earlier time $s \in (0,t)$ and starting from another state $y \in \mathbb{R}^d$, we may equivalently solve a control problem on the time horizon (s,t), using the function $x \mapsto v(t,x)$ as our terminal reward function.
To make this precise, we need to explain what it means to solve the control problem starting from $(t,x) \in [0,T] \times \mathbb{R}^d$. To this end, define the state process $X^{t,x} = (X_s^{t,x})_{s\in[t,T]}$. (To be more clear we should write $X^{t,x,\alpha}$), for any (t,x) by the SDE,

$$\begin{aligned} dX_s^{t,x} &= b\left(X_s^{t,x}, \alpha_s\right) ds + \sigma\left(X_s^{t,x}, \alpha_s\right) dW_s, \quad s \in [t,T] \\ X_t^{t,x} &= x \end{aligned}$$

For $\alpha \in \mathbb{A}$ and $(t,x) \in [0,T] \times \mathbb{R}^d$, define the reward functional

$$J(t,x,\alpha) := \mathbb{E}\left[\int_t^T f\left(X_r^{t,x}, \alpha_r\right) dr + g(X_T^{t,x})\right]$$

Under the standing assumptions, the SDE admits a unique strong solution. Moreover, since f and g are bounded from above, the integral and expectation in (5.3) are well-defined, and $J(t,x,\alpha) < \infty$, though it is possible that $J(t,x,\alpha) = -\infty$.
Finally, define the value function which is the optimal expected reward achievable starting at time t from the state x.

$$V(t,x) = \sup_{\alpha\in\mathbb{A}} J(t,x,\alpha)$$

Dynamic Programming Principle.
Fix $0 \le t < s \le T$ and $x \in \mathbb{R}^d$. Then

$$V(t,x) = \sup_{\alpha \in \mathbb{A}} \mathbb{E}\left[\int_t^T f\left(X_r^{t,x}, \alpha_r\right) dr + V(s, X_s^{t,x})\right]$$

Importantly, we will exploit the Markov or flow property, $X_u^{t,x} = X_u^{s, X_s^{t,x}}$, $t \le s \le u$. Also, to prove the $\ge$ direction, we have to use the measurable selection theorem to handle the measurablility issue of the new control

$$\widehat{\alpha}_r^{\epsilon}(\omega) = \begin{cases} \alpha_r(\omega) & r \le s \\ \alpha_r^{\epsilon,\omega}(\omega) & r > s. \end{cases}$$

Then, use the control $\widehat{\alpha}^e$ to define the state process $X^{t,x}$, and use the definition of the value function.

2.3 The verification theorem

Our point of view will be to identify a PDE which, if solvable in the classical sense, must coincide with the value function. We begin by reviewing the uncontrolled analogue, in which the "verification theorem" is nothing but the celebrated *Feynman-Kac formula*. (Assuming local martingale terms are true martingales.)

Feyman-Kac.
Let b and σ be "nice" coefficients. Let $X^{t,x}$ solve the SDE

$$\begin{aligned} dX_r^{t,x} &= b\left(X_r^{t,x}\right) dr + \sigma\left(X_r^{t,x}\right) \cdot dW_r \quad r \in (t,T] \\ X_t^{t,x} &= x \end{aligned}$$

Suppose v is a smooth, i.e. $C^{1,2}([0,T], \mathbb{R})$, solution of the PDE

$$\begin{aligned} &\partial_t v(t,x) + b(x) \cdot \nabla v(t,x) + \frac{1}{2} \operatorname{Tr}\left[\sigma\sigma^\top(x) \nabla^2 v(t,x)\right] + f(t,x) = 0 \\ &v(T,x) = g(x) \end{aligned}$$

Then v admits the representation

$$v(t,x) = \mathbb{E}\left[\int_t^T f(r, X_r^{t,x}) dr + g(X_T^{t,x})\right].$$

We can prove it by applying Ito's formula. Essentially, the PDE becomes an inequality, with equality holding only along the optimal control. Or, in other words, the PDE involves a pointwise optimization over the control variable. This leads to what is known as the Hamilton-Jacobi-Bellman (HJB) equation.
To simplify the notation, we introduce the *infinitesimal generator* of the controlled process.

Infinitesimal generator.
For a smooth function $\psi : \mathbb{R}^d \to \mathbb{R}$, and for $(x,a) \in \mathbb{R}^d \times A$, define

$$L^a\psi(x) = b(x,a) \cdot \nabla_x \psi(x) + \frac{1}{2} \operatorname{Tr}\left[\sigma\sigma^t(x,a)\nabla^2\psi(x)\right]$$

Verification theorem.
Suppose $v = v(t,x)$ is $C^{1,2}([0,T]\times \mathbb{R}^d)$ and satisfies $v(T,x) = g(x)$ along with the Hamilton-Jacobi-Bellman equation

$$\partial_t v(t,x) + \sup_{a\in A}\{L^a v(t,x) + f(x,a)\} = 0$$

where the operator L^a acts on the x variable of v. Assume also that there exists a measurable function $\widehat{\alpha} : [0,T] \times \mathbb{R}^d \to A$ such that

$$\widehat{\alpha}(t,x) \in \arg\max_{a\in A}[L^a v(t,x) + f(x,a)]$$

and the SDE

$$dX_t = b\left(X_t, \widehat{\alpha}\left(t, X_t\right)\right) dt + \sigma\left(X_t, \widehat{\alpha}\left(t, X_t\right)\right) dW_t$$

has a solution from any starting time and state. Then $v(t,x) = V(t,x)$ for all (t,x), and $\widehat{\alpha}\left(t, X_t\right)$ is an optimal control.

The theorem is absolutely crucial for the developments to come, as it provides a powerful recipe for finding an optimal control.

References

Professor Daniel Lacker, IEOR at Columbia University.Rough lecture notes from the Spring 2018 PhD course (IEOR E8100) on mean field games and interacting diffusion models. Olivier Gu´eant, Jean-Michel Lasry, Pierre-Louis Lions. Mean field games and applications