Risk-Sensitive and Robust Mean Field Games

Tamer Başar

Coordinated Science Laboratory
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL - 61801

IPAM Workshop on "Mean Field Games"
UCLA, Los Angeles, CA
August 28-September 1, 2017

August 31, 2017

Outline

- Introduction to nonzero-sum stochastic differential games (NZS SDGs) and Nash equilibrium: role of information structures
- Quick overview of risk-sensitive stochastic control (RS SC): equivalence to zero-sum stochastic differential games
- Mean field game approach to RS NZS SDGs with local state information–Problem 1 (P1)
- Mean field game approach to robust NZS SDGs with local state information–Problem 2 (P2)
- Connections between P1 and P2, and ϵ-Nash equilibrium
- Extensions and conclusions

General NZS SDGs

- N-player state dynamics described by a stochastic differential equation (SDE)

$$dx_t = f(t, x_t, u_t)dt + D(t)db_t\,, \quad x_{t|t=0} = x_0\,, \;\; u_t := (u_{1t}, \ldots, u_{Nt})$$

- Could also be in partitioned form: $x_t = (x_{1t}, \ldots, x_{Nt})$

$$dx_{it} = f_i(t, x_{it}, u_{it}; c_i(t, x_{-i,t}, u_{-i,t}))dt + D_i(t)db_{it}\,, \;\; i = 1, \ldots, N$$

- Information structures (control policy of player i: $\gamma_i \in \Gamma_i$):
  Closed-loop perfect state for all players: $u_{it} = \gamma_i(t; x_\tau, \tau \le t)\,,\; i = 1, \ldots, N$
  Partial (local) state: $u_{it} = \gamma_i(t; x_{i\tau}, \tau \le t)\,,\; i = 1, \ldots, N$
  Measurement feedback: $u_{it} = \gamma_i(t; y_{i\tau}, \tau \le t)\,,$
  $dy_{it} = h_i(t, x_{it}, x_{-i,t})dt + E_i(t)db_{it}\,,\; i = 1, \ldots, N$
- Loss function for player i (over $[t, T]$):

$$L_i(x_{[t,T]}, u_{[t,T]}) := q_i(x_T) + \int_t^T g_i(s, x_s, u_s)ds$$

  Take expectations (for horizon $[0, T]$) with $u = \gamma(\cdot)$: $J_i(\gamma_i, \gamma_{-i})$
- Nash equilibrium γ^*

$$J_i(\gamma_i^*, \gamma_{-i}^*) = \min_{\gamma_i \in \Gamma_i} J_i(\gamma_i, \gamma_{-i}^*)$$

Equilibrium solution

- State dynamics and loss functions

$$dx_t = f(t, x_t, u_t)dt + D(t)db_t\,, \quad x_{t|t=0} = x_0\,, \;\; u_t := (u_{1t}, \ldots, u_{Nt})$$

$$L_i(x_{[t,T]}, u_{[t,T]}) := q_i(x_T) + \int_t^T g_i(s, x_s, u_s)ds\,, \; i = 1, \ldots, N$$

- *Closed-loop perfect state for all players*: assume DD' is strongly positive

  Nash equilibrium exists and is unique, if the coupled PDEs below admit a unique smooth solution:

$$-V_{i,t}(t;x) = \min_v [V_{i,x} f(t, x, v, u^*_{-i}) + g_i(t, x, v, u^*_{-i})] + \frac{1}{2} Tr[V_{i,xx}(t;x)DD']$$

$$V_i(T;x) = q_i(x)\,, \quad u^*_{it} = \gamma^*_i(t, x(t))\,, \;\; i = 1, \ldots, N$$

- Other *dynamic* information structures (s.a. local state, decentralized, measurement feedback): Extremely challenging! Possibly infinite-dimensional (even if NE exists and is unique), even in linear-quadratic (LQ) NZS SDGs.

LQ NZS SDGs with perfect state information

$$dx(t) = \left[Ax + \sum_{i=1}^{N} B_i u_i\right] dt + D db(t), \quad i \in \mathcal{N} = \{1, 2, \ldots, N\}$$

$$J_i(\gamma_i, \gamma_{-i}) = \mathbb{E}\left[|x(T)|^2_{W_i} + \int_0^T \left[|x(t)|^2_{Q_i(t)} + \sum_{j=1}^{N} |u_j(t)|^2_{R_{ij}}\right] dt\right]$$

- The feedback Nash equilibrium: $\gamma_i^*(t, x(t)) = -R_{ii}^{-1} B_i^\top Z_i(t) x(t)$, $i \in \mathcal{N}$
- The coupled RDEs with $Z_i \geq 0$ and $Z_i(T) = W_i$

$$-\dot{Z}_i = F^\top Z_i + Z_i F + Q_i + \sum_{j=1}^{N} Z_j B_j R_{jj}^{-1} R_{ij} R_{jj}^{-1} B_j^\top Z_j, \quad i \in \mathcal{N}$$

$$F := A - \textstyle\sum_{j=1}^{N} B_i R_{ii}^{-1} B_i^\top Z_i$$

- When information is local state, or imperfect measurement (even if shared by all players), existence and characterization an open problem
- Any hope for N sufficiently large? MFG approach provides the answer!

Risk-sensitive (RS) formulation of the NZS SDG

Replace J_i with

$$J_i(\gamma_i, \gamma_{-i}) = \frac{2}{\theta} \ln E\Big\{ \exp \frac{\theta}{2} L_i(x_{[0,T]}, u_{[0,T]}) \Big\}$$

where $\theta > 0$ is the risk sensitivity parameter, and as before

$$L_i(x_{[t,T]}, u_{[t,T]}) := q_i(x_T) + \int_t^T g_i(s, x_s, u_s) ds$$

and

$$u_{it} = \gamma_i(\cdot)\,, \quad i = 1, \ldots, N$$

Nash equilibrium is defined as before, and the same difficulties with regard to information structures arise as before.

Digression: Risk-sensitive (RS) stochastic control

State dynamics :

$$dx_t = f(t, x_t, u_t)\, dt + \sqrt{\epsilon}\, D\, db_t\,; \quad x_{t|t=0} = x_0$$

$b_t, t \geq 0$, standard Wiener process; $\epsilon > 0$;

$u_t \in U,\ t \geq 0$ (**state FB control law** $\mu \in \mathcal{M}$)

Objective : Choose μ to minimize : ($\theta > 0$)

$$J(\mu; t, x_t) = \frac{2\epsilon}{\theta} \ln E\{ \exp \frac{\theta}{2\epsilon} L(x_{[t,T]}, u_{[t,T]}) \}$$

$$L(x_{[t,T]}, u_{[t,T]}) := q(x_T) + \int_t^T g(s, x_s, u_s) ds$$

$\psi(t; x)$ – **value function associated with**

$$E\{ \exp \frac{\theta}{2\epsilon} \left[q(x_T) + \int_t^T g(s, x_s, u_s)\, ds \right] \}$$

$$\Rightarrow \qquad V(t; x) := \inf_{\mu \in \mathcal{M}} J(\mu; t, x) =: \frac{2\epsilon}{\theta} \ln \psi(t; x)\,,$$

DP and Itô differentiation rule $\Rightarrow$

$$-V_t(t;x) = \inf_{u \in U} \left\{ V_x(t;x)\, f(t,x,u) + g(t,x,u) \right\}$$

$$+ \frac{1}{4\gamma^2} |DV_x'(t;x)|^2 + \frac{\epsilon}{2} \mathrm{Tr}\left[V_{xx}\, DD' \right]$$

$$V(T;x) \equiv q(x) \qquad \left(\gamma^{-2} := \theta\right)$$

If $U = \mathbf{R}^{m_1}$, f linear in u, and g quadratic in u :

$$f(x,u) = f_0(t,x) + B(t,x)u\,;\ g(t,x,u) = g_0(t,x) + |u|^2$$

Optimal control law:

$$u^*(t) = \mu^*(t,x) = -\frac{1}{2}\, B'(t,x)\, V_x'(t;x)\,, \quad 0 \le t \le t_f$$

$\Rightarrow$ HJB equation :

$$-V_t = V_x\, f_0(t,x) + g_0(t,x) - \frac{1}{4}\left[|BV_x'|^2 - \gamma^{-2} |DV_x'|^2 \right]$$

$$+ \frac{\epsilon}{2} \mathrm{Tr}\left[V_{xx}(t;x) DD' \right]\,; \qquad V(T;x) \equiv q(x)$$

A further special case : LEQG Problem

$$f_0(t,x) = A(t)x\,, \quad g_0(t,x) = \frac{1}{2}\,x'Qx\,,\ Q \geq 0$$

$$q(x) = (1/2)\,x'Q_f x$$

$\Rightarrow$ **Explicit solution:**

$$V(t;x) = \frac{1}{2}\,x'Z(t)x + \ell^{\epsilon}(t)\,, \quad t \geq 0$$

$$\dot{Z} + A'Z + ZA + Q - Z\,(\,BB' - \gamma^{-2}DD'\,)\,Z = 0$$

$$\ell^{\epsilon}(t) = \frac{\epsilon}{2}\int_t^T \mathrm{Tr}\,[\,Z(s)\,D(s)D'(s)\,]\,ds$$

$$\Rightarrow \quad u^*(t) = \mu^*(t,x) = -B'(t)\,Z(t)\,x\,, \quad 0 \leq t \leq t_f$$

A class of stochastic differential games

Two Players : **Player 1:** u_t ; **Player 2:** w_t

$$dx_t = f(x_t, u_t)\,dt + Dw_t\,dt + \sqrt{\epsilon}\,D\,db_t\,; \quad x_0$$

$$J(\mu, \nu; t, x_t) := E\Big\{q(x_T) + \int_t^T g(s, x_s, u_s)\,ds - \gamma^2 \int_t^T |w_s|^2\,ds\Big\}$$

Upper-Value (UV) Function :

$$\bar{W}(t; x) = \inf_{\mu} \sup_{\nu} J(\mu, \nu; t, x)$$

HJI UV equation :

$$\inf_{u \in U} \sup_{w \in \mathbf{R}^{m_2}} \Big\{\bar{W}_t + \bar{W}_x\,(\,f + Dw\,) + g - \gamma^2 |w|^2 + \frac{\epsilon}{2}\mathrm{Tr}\big[\bar{W}_{xx} DD'\big]\Big\} = 0$$

HJI UV equation :

$$\inf_{u\in U}\sup_{w\in\mathbf{R}^{m_2}}\left\{\bar{W}_t+\bar{W}_x\,(\,f+Dw\,)+g-\gamma^2|w|^2\right.$$
$$\left.+\frac{\epsilon}{2}\mathrm{Tr}\left[\bar{W}_{xx}DD'\right]\right\}=0$$

Isaacs condition holds $\Rightarrow$ **Value Function :**

$$-W_t(t;x)=\inf_{u\in U}\left\{W_x(t;x)\,f(t,x,u)+g(t,x,u)\right\}$$
$$+\frac{1}{4\gamma^2}|DW_x'(t;x)|^2+\frac{\epsilon}{2}\,\mathrm{Tr}\left[W_{xx}(t;x)DD'\right];$$
$$W(T;x)\equiv q(x)$$

- IDENTICAL with V for all permissible ϵ,γ
- The same holds for the time-average case (LQ)

LQ RS MFGs, Problem 1 (**P1**)

- Stochastic differential equation (SDE) for agent i, $1 \leq i \leq N$

$$dx_i(t) = (A(\theta_i)x_i(t) + B(\theta_i)u_i(t))dt + \sqrt{\mu}D(\theta_i)db_i(t)$$

- The risk-sensitive cost function for agent i with $\delta > 0$

$$\textbf{P1}: \quad J_{1,i}^N(u_i, u_{-i}) = \limsup_{T\to\infty} \frac{\delta}{T} \log \mathbb{E}\Big\{ e^{\frac{1}{\delta}\phi_i^1(x, f_N, u)} \Big\}$$

$$\phi_i^1(x, f_N, u) := \int_0^T \|x_i(t) - \frac{1}{N}\sum_{i=1}^N x_i(t)\|_Q^2 + \|u_i(t)\|_R^2 dt$$

- Risk-sensitive control: Robust control w.r.t. the risk parameter δ

$$J_{1,i}^N(u_i, u_{-i}) = \limsup_{T\to\infty} \frac{1}{T}\Big[\mathbb{E}\{\phi_i^1\} + \frac{1}{2\delta} var\{\phi_i^1\} + o(\frac{1}{\delta}) \Big]$$

- $f_N(t) := \frac{1}{N}\sum_{i=1}^N x_i(t)$: Mean Field term (mass behavior)
- Agents are coupled with each other through the mean field term

Tembine-Zhu-TB, TAC (59, 4, 2014); Moon-TB, CDC (2014), TAC (62, 3, 2017)

LQ Robust MFGs, Problem 2 (**P2**)

- Stochastic differential equation (SDE) for agent i, $1 \leq i \leq N$

$$dx_i(t) = (A(\theta_i)x_i(t) + B(\theta_i)u_i(t) + D(\theta_i)v_i(t))dt + \sqrt{\mu}D(\theta_i)db_i(t)$$

- The worst-case risk-neutral cost function for agent i

P2: $$J_{2,i}^N(u_i, u_{-i}) = \sup_{v_i \in \mathcal{V}_i} \limsup_{T \to \infty} \frac{1}{T}\mathbb{E}\{\phi_i^2(x, f_N, u, v)\}$$

$$\phi_i^2(x, f_N, u, v) := \int_0^T \|x_i(t) - \frac{1}{N}\sum_{i=1}^N x_i(t)\|_Q^2 + \|u_i(t)\|_R^2 - \gamma^2\|v_i(t)\|^2 dt$$

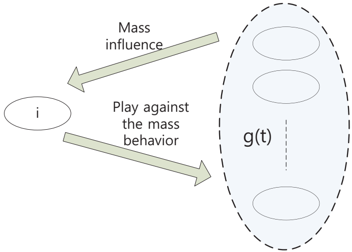
- v_i can be viewed as a fictitious player (or adversary) of agent i, which strives for a worst-case cost function for agent i
- Agents are coupled with each other through the mean field term

Mean Field Analysis for **P1** and **P2**

- Solve the individual local robust control problem with g instead of f_N

$$\textbf{P1: } \bar{J}_1(u,g) = \limsup_{T\to\infty} \frac{\delta}{T} \log \mathbb{E}\{\exp[\frac{1}{\delta}\int_0^T \|x(t)-g(t)\|_Q^2 + \|u(t)\|_R^2 dt]\}$$

$$\textbf{P2: } \bar{J}_2(u,v,g) = \limsup_{T\to\infty} \frac{1}{T} \mathbb{E}\{\int_0^T \|x(t)-g(t)\|_Q^2 + \|u(t)\|_R^2 - \gamma^2\|v(t)\|^2 dt\}$$



- Characterize g^* that is a *best estimate* of the mean field f_N
  - need to construct a *mean field system* $\mathcal{T}(g)(t)$
  - obtain a fixed point of $\mathcal{T}(g)(t)$, i.e., $g^* = \mathcal{T}(g^*)$

Robust Tracking Control for **P1** and **P2**

Proposition: Individual robust control problems for **P1** and **P2**

Suppose that (A, B) is stabilizable and $(A, Q^{1/2})$ is detectable. Suppose that for a fixed $\gamma = \sqrt{\delta/2\mu} > 0$, there is a matrix $P \geq 0$ that solves the following GARE

$$A^T P + PA + Q - P(BR^{-1}B^T - \frac{1}{\gamma^2} DD^T)P = 0$$

Then

- $H := A - BR^{-1}B^T P + \frac{1}{\gamma^2} DD^T P$ and $G := A - BR^{-1}B^T P$ are Hurwitz
- The robust decentralized controller: $\bar{u}(t) = -R^{-1}B^T P x(t) - R^{-1}B^T s(t)$ where $\frac{ds(t)}{dt} = -H^T s(t) + Qg(t)$
- The worst-case disturbance (**P2**): $\bar{v}(t) = \gamma^{-2} D^T P x(t) + \gamma^{-2} D^T s(t)$
- $s(t)$ has a unique solution in $\mathcal{C}_n^b$: $s(t) = -\int_t^\infty e^{-H^T(t-s)} Qg(s) ds$

Remark

- The two robust tracking problems are identical
- Related to the robust (H^∞) control problem w.r.t. γ

Mean Field Analysis for **P1** and **P2**

- $\bar{x}_\theta(t) = \mathbb{E}\{x_\theta(t)\}$ and we use $h \in \mathcal{C}_n^b$ for **P2**
- Mean field system for **P1** (with the robust decentralized controller)

$$\mathcal{T}(g)(t) := \int_{\theta\in\Theta, x\in X} \bar{x}_\theta(t) dF(\theta, x)$$

$$\bar{x}_\theta(t) = e^{G(\theta)t}x + \int_0^t e^{G(\theta)(t-\tau)} B(\theta)R^{-1}B^T(\theta) \times \left(\int_\tau^\infty e^{-H(\theta)^T(\tau-s)} Qg(s)ds \right) d\tau$$

- Mean field system for **P2** (with the robust decentralized controller and the worst-case disturbance)

$$\mathcal{L}(h)(t) := \int_{\theta\in\Theta, x\in X} \bar{x}_\theta(t) dF(\theta, x)$$

$$\bar{x}_\theta(t) = e^{H(\theta)t}x + \int_0^t e^{H(\theta)(t-\tau)} \Big(B(\theta)R^{-1}B^T(\theta) - \gamma^{-2}D(\theta)D(\theta)^T \Big) \times \left(\int_\tau^\infty e^{-H^T(\theta)(\tau-s)} Qh(s)ds \right) d\tau$$

Mean Field Analysis for **P1** and **P2**

- $\mathcal{T}(g)(t)$ and $\mathcal{L}(h)(t)$ capture the mass behavior when N is large
- Simplest case

$$\lim_{N\to\infty} f_N(t) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i(t) = \mathbb{E}\{x_i(t)\} = \mathcal{T}(g)(t), \quad \text{SLLN}$$

- We need to seek g^* and h^* such that $g^* = \mathcal{T}(g^*)$ and $h^* = \mathcal{L}(h^*)$
- Sufficient condition (due to the contraction mapping theorem)

$$\textbf{P1}: \ \|R^{-1}\|\|Q\| \int_{\theta\in\Theta} \|B(\theta)\|^2 \Big(\int_0^\infty \|e^{G(\theta)\tau}\| d\tau\Big)\Big(\int_0^\infty \|e^{H(\theta)\tau}\| d\tau\Big) dF(\theta) < 1$$

$$\textbf{P2}: \ \int_{\theta\in\Theta} \Big(\int_0^\infty \|e^{H(\theta)t}\|^2 dt\Big)^2 \Big(\|B(\theta)\|^2\|R^{-1}\| + \gamma^{-2}\|D(\theta)\|^2\Big) dF(\theta) < 1$$

- $\lim_{k\to\infty} \mathcal{T}^k(g_0) = g^*$ for any $g_0 \in \mathcal{C}_n^b$
- $g^*(t)$ and $h^*(t)$ are best estimates of $f_N(t)$ when N is large
- Generally $g^* \not\equiv h^*$. But when $\gamma \to \infty$, $g^* \equiv h^*$

Main Results for **P1** and **P2**

Existence and Characterization of an ϵ-Nash equilibrium

- There exists an ϵ-Nash equilibrium with g^* (**P1**), i.e., there exist $\{u_i^*, 1 \leq i \leq N\}$ and $\epsilon_N \geq 0$ such that

$$J_{1,i}^N(u_i^*, u_{-i}^*) \leq \inf_{u_i \in \mathcal{U}_i^c} J_{1,i}^N(u_i, u_{-i}^*) + \epsilon_N,$$

  where $\epsilon_N \to 0$ as $N \to \infty$. For the uniform agent case, $\epsilon_N = O(1/\sqrt{N})$
- The ϵ-Nash strategy u_i^* is decentralized, i.e., u_i^* is a function of x_i and g^*
- Law of Large Numbers: g^* satisfies

$$\lim_{N\to\infty} \int_0^T \left\| \frac{1}{N} \sum_{i=1}^N x_i^*(t) - g^*(t) \right\|^2 dt = 0, \ \forall T \geq 0, \ \text{a.s.}$$

$$\lim_{N\to\infty} \limsup_{T\to\infty} \frac{1}{T} \int_0^T \left\| \frac{1}{N} \sum_{i=1}^N x_i^*(t) - g^*(t) \right\|^2 dt = 0, \ \text{a.s.}$$

  g^*: deterministic function and can be computed offline
- The same results also hold for **P2** with the worst-case disturbance

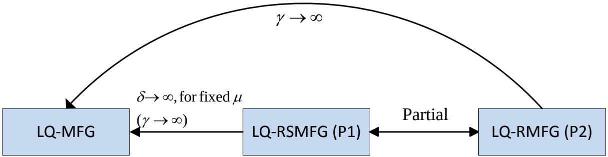
Main Results for **P1** and **P2**

Proof (sketch): Law of large numbers (first part)

$$\int_0^T \|f_N^*(t) - g^*(t)\|^2 dt \leq 2 \int_0^T \left\| \frac{1}{N} \sum_{i=1}^{N} (x_i^*(t) - \mathbb{E}\{x_i^*(t)\}) \right\|^2 dt$$
$$+ 2T \sup_{t \geq 0} \left\| \mathbb{E}\{x_i^*(t)\} - g^*(t) \right\|^2$$

- The second part is zero (due to the fixed-point theorem)
- $e_i^*(t) = x_i^*(t) - \mathbb{E}\{x_i^*(t)\}$ is a mutually orthogonal random vector with $\mathbb{E}\{e_i^*(t)\} = 0$ and $\mathbb{E}\{\|e_i^*(t)\|^2\} < \infty$ for all i and $t \geq 0$
- Strong law of large numbers $\Rightarrow \lim_{N\to\infty} \|(1/N) \sum_{i=1}^N e_i^*(t)\| = 0$ for all $t \in [0, T]$
- $\|(1/N) \sum_{i=1}^N e_i^*(t)\|^2$, $N \geq 1$, is uniformly integrable on $[0, T]$ for all $T \geq 0$, we have the desired result

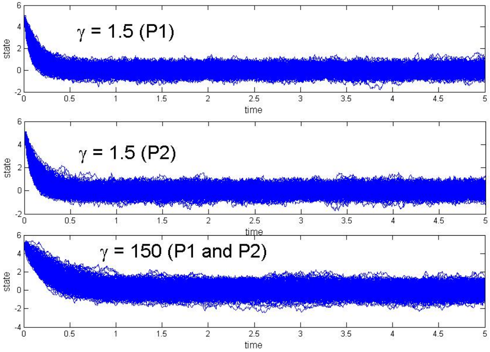
Partial Equivalence and Limiting Behaviors of **P1** and **P2**



- **P1** and **P2** share the same robust decentralized controller
- Partial equivalence: the mean field systems (and their fixed points) are different
- Limiting behaviors
  - Large deviation (small noise) limit ($\mu, \delta \to 0$ with $\gamma = \sqrt{\delta/2\mu} > 0$): The same results hold under this limit (SDE $\Rightarrow$ ODE)
  - Risk-neutral limit ($\gamma \to \infty$): The results are identical to that of the (risk-neutral) LQ mean field game ($g^* \equiv h^*$)

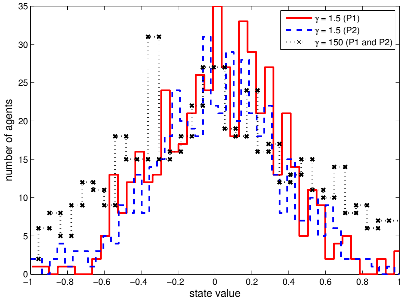
Simulations ($N = 500$)

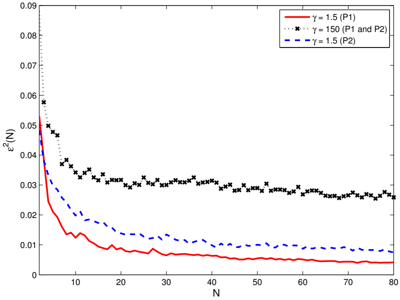
- $A_i = \theta_i$ is an i.i.d. uniform random variable with the interval $[2, 5]$, $B = D = Q = R = 1$, $\mu = 2 \Rightarrow \gamma_\theta^* = \gamma^* = 1$, $g^*(t) = 5.086e^{-8.49t}$, $h^*(t) = 5.1e^{-3.37t}$
- $\epsilon^2(N) := \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\int_0^T \|f_N^*(t) - g^*(t)\|^2 dt$
- γ determines robustness of the equilibrium (due to the individual robust control problems)



Simulations ($N = 500$)

- $A_i = \theta_i$ is an i.i.d. uniform random variable with the interval $[2, 5]$, $B = D = Q = R = 1$, $\mu = 2 \Rightarrow \gamma^*_\theta = \gamma^* = 1$, $g^*(t) = 5.086e^{-8.49t}$, $h^*(t) = 5.1e^{-3.37t}$
- $\epsilon^2(N) := \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\int_0^T \|f^*_N(t) - g^*(t)\|^2 dt$
- γ determines robustness of the equilibrium (due to the individual robust control problems)





Conclusions

- Decentralized (local state-feedback) ϵ-Nash equilibria for LQ risk-sensitive and LQ robust mean field games
- The equilibrium features robustness due to the local robust optimal control problem parametrized by γ
- LQ risk-sensitive and LQ robust mean field games
  - are partially equivalent ($g^* \not\equiv h^*$)
  - hold the same limiting behaviors as the one-agent case
- Extensions to heterogenous case and nonlinear dynamics are possible, but results are not as explicit; see, Tembine, Zhu, Başar, IEEE-TAC (59, 4, 2014) for RSMFG
- Imperfect state measurements
- RSMFGs on networks with agents interacting only with their neighbors
- Leader-Follower MFGs

THANKS !