ε-Nash Mean Field Games

From Mean Field Games to Weak KAM Dynamics

Peter E. Caines

McGill University

Mathematics Institute: University of Warwick

May, 2012

Part I

Co-Authors

Minyi Huang

Roland Malhamé

Collaborators & Students

Arman Kizilkale

Arthur Lazarte

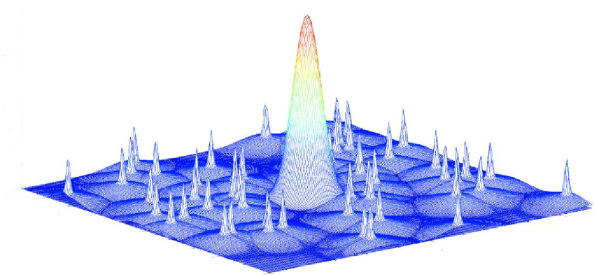
Zhongjing Ma

Mojtaba Nourian

Overview

- Overall Objective:
  - Develop a theory of decentralized decision-making in stochastic dynamical systems with many competing or cooperating agents
- Outline:
  - A motivating control problem from code division multiple access (CDMA) uplink power control
  - Motivational notions from statistical mechanics
  - The basic concepts of Mean Field (MF) control and game theory
  - The Nash Certainty Equivalence (NCE) methodology
  - Main NCE results for Linear-Quadratic-Gaussian (LQG) systems
  - Nonlinear MF Systems
  - Adaptive NCE System Theory
  - Adaptation based leader-follower stochastic dynamic games

Base Station & Individual Agents



Part 1 – CDMA Power Control

- Lognormal channel attenuation: $1 \le i \le N$

  i_{th} channel: $dx_i = -a(x_i + b)dt + \sigma dw_i, \qquad 1 \le i \le N$

  Transmitted power = channel attenuation × power
  $= e^{x_i(t)} p_i(t)$
  (Charalambous, Menemenlis; 1999)

  Signal to interference ratio (Agent i) at the base station
  $= e^{x_i} p_i / \left[(\beta/N) \sum_{j \neq i}^{N} e^{x_j} p_j + \eta \right]$

- How to optimize all the individual SIR's?
  - Self defeating for everyone to increase their power
  - Humans display the "Cocktail Party Effect": Tune hearing to frequency of friend's voice (E. Colin Cherry)

Part 1 – CDMA Power Control

- Can maximize $\sum_{i=1}^{N} SIR_i$ with centralized control.
(HCM, 2004)

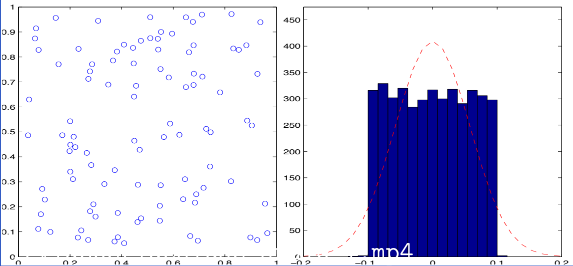
- Since centralized control is not feasible for complex systems, how can such systems be optimized using decentralized control?

- Idea: Use large population properties of the system together with basic notions of game theory.

The Statistical Mechanics Connection

Part 2 – Statistical Mechanics

- A foundation for thermodynamics was provided by the Statistical Mechanics of Boltzmann, Maxwell and Gibbs.
- Basic Ideal Gas Model describes the interaction of a huge number of essentially identical particles.
- SM describes the aggregate of the very complex individual particle trajectories in terms of the PDEs governing the continuum limit of the mass of particles.



Animation of Particles

Start from the equations for the perfectly elastic (i.e. hard) sphere mechanical collisions of each pair of particles:

Velocities before collision: $\mathbf{v}, \mathbf{V}$
Velocities after collision: $\mathbf{v}' = \mathbf{v}'(\mathbf{v}, \mathbf{V}, t), \quad \mathbf{V}' = \mathbf{V}'(\mathbf{v}, \mathbf{V}, t)$

These collisions satisfy the conservation laws, and hence:

$$\text{Conserv. of} \begin{cases} \text{Momentum} & m(\mathbf{v}' + \mathbf{V}') = m(\mathbf{v} + \mathbf{V}) \\ \\ \text{Energy} & \frac{1}{2} m \left(\|\mathbf{v}'\|^2 + \|\mathbf{V}'\|^2 \right) = \frac{1}{2} m \left(\|\mathbf{v}\|^2 + \|\mathbf{V}\|^2 \right) \end{cases}$$

Part 2 – Boltzmann's Equation

The assumption of Statistical Independence of particles (Propagation of Chaos Hypothesis) gives Boltzmann's PDE for the behaviour of an infinite population of particles

$$\frac{\partial p_t(\mathbf{v},\mathbf{x})}{\partial t} + \nabla_x p_t(\mathbf{v},\mathbf{x}) \cdot \mathbf{v} = \iiint Q(\theta,\psi) d\theta d\psi \|\mathbf{v} - \mathbf{V}\| \cdot$$

$$\Big(p_t(\mathbf{v}',\mathbf{x}) p_t(\mathbf{V}',\mathbf{x}) - p_t(\mathbf{v},\mathbf{x}) p_t(\mathbf{V},\mathbf{x}) \Big) d^3\mathbf{V}$$

$$\mathbf{v}' = \mathbf{v}'(\mathbf{v},\mathbf{V},t), \quad \mathbf{V}' = \mathbf{V}'(\mathbf{v},\mathbf{V},t)$$

Entropy H:

$$H(t) \stackrel{\text{def}}{=} -\iiint p_t(\mathbf{v},x)(\log p_t(\mathbf{v},x))\,d^3x d^3\mathbf{v}$$

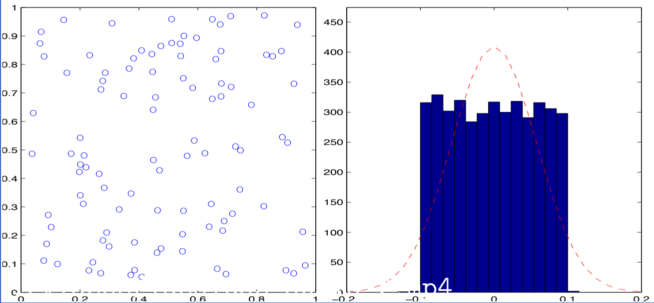
The H-Theorem: $$\frac{dH(t)}{dt} \geq 0$$

$H_\infty \stackrel{\text{def}}{=} \sup_{t\geq 0} H(t)$ occurs at $p_\infty = N(\mathbf{v}_\infty, \Sigma_\infty)$

(The Maxwell Distribution)

Part 2 – Statistical Mechanics: Key Intuition

Extremely complex large population particle systems
may have simple continuum limits
with distinctive bulk properties.



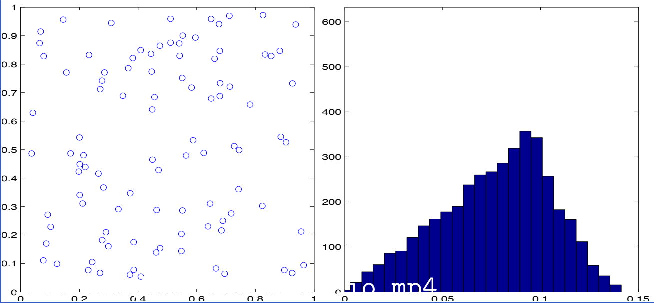
Animation of Particles

Part 2 – Statistical Mechanics

Control of Natural Entropy Increase

Feedback Control Law (Non-physical Interactions):

- At each collision, total energy of each pair of particles is shared equally while physical trajectories are retained.
- Energy is conserved.



Animation of Particles

Part 2 – Key Intuition

- A sufficiently large mass of individuals may be treated as a continuum.

- Local control of particle (or agent) behaviour can result in (partial) control of the continuum.

Part 3 – Game Theoretic Control Systems

- Massive game theoretic control systems: Large ensembles of partially regulated competing agents
- Fundamental issue: The relation between the actions of each individual agent and the resulting mass behavior

Part 3 – Basic LQG Game Problem

Individual Agent's Dynamics:

$$dx_i = (a_i x_i + bu_i)dt + \sigma_i dw_i, \quad 1 \leq i \leq N.$$

(scalar case only for simplicity of notation)

- x_i: state of the ith agent
- u_i: control
- w_i: disturbance (standard Wiener process)
- N: population size

Individual Agent's Cost:

$$J_i(u_i, \nu) \triangleq E \int_0^\infty e^{-\rho t}[(x_i - \nu)^2 + r u_i^2] dt$$

Basic case: $\nu \triangleq \gamma.(\frac{1}{N}\sum_{k \neq i}^N x_k + \eta)$

More General case: $\nu \triangleq \Phi(\frac{1}{N}\sum_{k \neq i}^N x_k + \eta)$ Φ Lipschitz

Main feature:

- Agents are coupled via their costs
- Tracked process ν:
  - (i) stochastic
  - (ii) depends on other agents' control laws
  - (iii) not feasible for x_i to track all x_k trajectories for large N

Part 3 – Large Popn. Models with Game Theory Features

- Economic models: Cournot-Nash equilibria (Lambson)
- Advertising competition: game models (Erickson)
- Wireless network res. alloc.: (Alpcan et al., Altman, HCM)
- Admission control in communication networks: (Ma, MC)
- Public health: voluntary vaccination games (Bauch & Earn)
- Biology: stochastic PDE swarming models (Bertozzi et al.)
- Sociology: urban economics (Brock and Durlauf et al.)
- Renewable Energy: Charging control of of PEVs (Ma et al.)

Part 3 – Background & Current Related Work

Background:

- 40+ years of work on stochastic dynamic games and team problems: Witsenhausen, Varaiya, Ho, Basar, *et al.*

Current Related Work:

- Industry dynamics with many firms: Markov models and Oblivious Equilibria (Weintraub, Benkard, Van Roy, Adlakha, Johari & Goldsmith, 2005, 2008 -)
- Mean Field Games: Stochastic control of many agent systems with applications to finance (Lasry et al., Cardaliaguet, Capuzzo-Dolcetta, Buckdahn, 2006-)
- Mean Field Control of Oscillators & Mean Field Particle Filters (Yin/Yang, Mehta, Meyn, Shanbhag, 2009-)
- Mean Field Games for Nonlinear Markov Processes (Kolokoltsov, Li, Wei, 2011-)
- Mean Field MDP Games on Networks: Exchangeability hypothesis; propagation of chaos in the popn. limit; evolutionary games. (Tembine et al., 2009-)
- Discrete Mean Field Games (Gomes, Mohr, Souza, 2009-)

Part 3 – Preliminary Optimal LQG Tracking

LQG Tracking: Take x^* (bounded continuous) for scalar model:

$$dx_i = a_i x_i dt + b u_i dt + \sigma_i dw_i$$

$$J_i(u_i, x^*) = E\int_0^\infty e^{-\rho t}[(x_i - x^*)^2 + r u_i^2]dt$$

Riccati Equation: $\rho \Pi_i = 2a_i \Pi_i - \frac{b^2}{r}\Pi_i^2 + 1, \quad \Pi_i > 0$

Set $\beta_1 = -a_i + \frac{b^2}{r}\Pi_i$, $\beta_2 = -a_i + \frac{b^2}{r}\Pi_i + \rho$, and assume $\beta_1 > 0$

Mass Offset Control: $\rho s_i = \frac{ds_i}{dt} + a_i s_i - \frac{b^2}{r}\Pi_i s_i - x^*.$

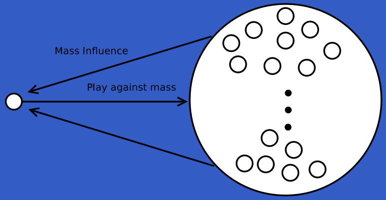
Optimal Tracking Control: $u_i = -\frac{b}{r}(\Pi_i x_i + s_i)$

- Boundedness condition on x^* implies existence of unique solution s_i.

Part 3 – Key Intuition

When the tracked signal is replaced by the deterministic mean state of the mass of agents:

Agent's feedback = feedback of agent's local stochastic state



+

feedback of deterministic mass offset

Think Globally, Act Locally (Geddes, Alinsky, Rudie-Wonham)

Part 3 – LQG-NCE Equation Scheme

The Fundamental NCE Equation System

Continuum of Systems: $a \in \mathcal{A}$; common b for simplicity

$$\rho s_a = \frac{ds_a}{dt} + a s_a - \frac{b^2}{r}\Pi_a s_a - x^*$$

$$\frac{d\overline{x}_a}{dt} = (a - \frac{b^2}{r}\Pi_a)\overline{x}_a - \frac{b^2}{r}s_a,$$

$$\overline{x}(t) = \int_{\mathcal{A}} \overline{x}_a(t) dF(a),$$

$$x^*(t) = \gamma(\overline{x}(t) + \eta) \qquad t \geq 0$$

$$\textbf{Riccati Equation:} \quad \rho\Pi_a = 2a\Pi_a - \frac{b^2}{r}\Pi_a^2 + 1, \quad \Pi_a > 0$$

- Individual control action $u_a = -\frac{b}{r}(\Pi_a x_a + s_a)$ is optimal w.r.t tracked x^*.
- Does there exist a solution $(\overline{x}_a, s_a, x^*; a \in \mathcal{A})$?
  Yes: Fixed Point Theorem

Proposed MF Solution to the Large Population LQG Game Problem

The Finite System of N Agents with Dynamics:

$$dx_i = a_i x_i dt + b u_i dt + \sigma_i dw_i, \qquad 1 \leq i \leq N, \qquad t \geq 0$$

Let $u_{-i} \triangleq (u_1, \cdots, u_{i-1}, u_{i+1}, \cdots, u_N)$; then the individual cost

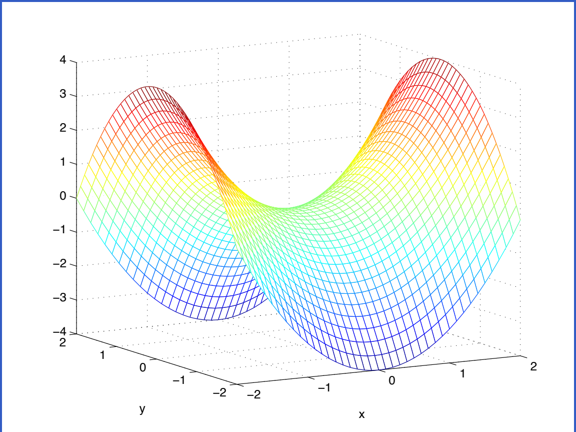
$$J_i(u_i, u_{-i}) \triangleq E \int_0^\infty e^{-\rho t} \{ [x_i - \gamma(\frac{1}{N} \sum_{k \neq i}^{N} x_k + \eta)]^2 + r u_i^2 \} dt$$

Algorithm: For ith agent with parameter (a_i, b) compute:

- x^* using NCE Equation System
- $\begin{cases} \rho \Pi_i = 2 a_i \Pi_i - \frac{b^2}{r} \Pi_i^2 + 1 \\ \rho s_i = \frac{ds_i}{dt} + a_i s_i - \frac{b^2}{r} \Pi_i s_i - x^* \\ u_i = -\frac{b}{r}(\Pi_i x_i + s_i) \end{cases}$

Part 3 – Saddle Point Nash Equilibrium

- Agent y is a maximizer
- Agent x is a minimizer



Part 3 – Nash Equilibrium

The Information Pattern:

$\mathcal{F}_i \triangleq \sigma(x_i(\tau); \tau \le t)$ $\mathcal{F}^N \triangleq \sigma(x_j(\tau); \tau \le t, 1 \le j \le N)$

$\mathcal{F}_i$ adapted control: $\mathcal{U}_{loc,i}$ $\mathcal{F}^N$ adapted control: $\mathcal{U}$

The Equilibria:

The set of controls $\mathcal{U}^0 = \{u_i^0;\, u_i^0$ adapted to $\mathcal{U}_{loc,i},\, 1 \le i \le N\}$ generates a *Nash Equilibrium* w.r.t. the costs $\{J_i; 1 \le i \le N\}$ if, for each i,

$$J_i(u_i^0, u_{-i}^0) = \inf_{u_i \in \mathcal{U}} J_i(u_i, u_{-i}^0)$$

ϵ-Nash Equilibria:

Given $\varepsilon > 0$, the set of controls $\mathcal{U}^0 = \{u_i^0; 1 \le i \le N\}$ generates an *ε-Nash Equilibrium* w.r.t. the costs $\{J_i; 1 \le i \le N\}$ if, for each i,

$$J_i(u_i^0, u_{-i}^0) - \varepsilon \le \inf_{u_i \in \mathcal{U}} J_i(u_i, u_{-i}^0) \le J_i(u_i^0, u_{-i}^0)$$

Part 3 – NCE Control: First Main Result

Theorem: (MH, PEC, RPM, 2003)

Subject to technical conditions, the NCE Equations have a unique solution for which the NCE Control Algorithm generates a set of controls $\mathcal{U}_{nce}^N = \{u_i^0; 1 \le i \le N\}, \ 1 \le N < \infty$, where

$$u_i^0 = -\frac{b}{r}(\Pi_i x_i + s_i)$$

which are s.t.

(i) All agent systems $S(A_i), \ 1 \le i \le N,$ are second order stable.

(ii) $\{\mathcal{U}_{nce}^N; 1 \le N < \infty\}$ yields an ε-Nash equilibrium for all ε, i.e. $\forall \varepsilon > 0 \ \exists N(\varepsilon)$ s.t. $\forall N \ge N(\varepsilon)$

$$J_i(u_i^0, u_{-i}^0) - \varepsilon \le \inf_{u_i \in \mathcal{U}} J_i(u_i, u_{-i}^0) \le J_i(u_i^0, u_{-i}^0),$$

where $u_i \in \mathcal{U}$ is adapted to $\mathcal{F}^N$.

□

Part 3 – NCE Control: Key Observations

- The information set for NCE Control is minimal and completely local since Agent A_i's control depends on:
  (i) Agent A_i's own state: $x_i(t)$
  (ii) Statistical information $F(\theta)$ on the dynamical parameters of the mass of agents.

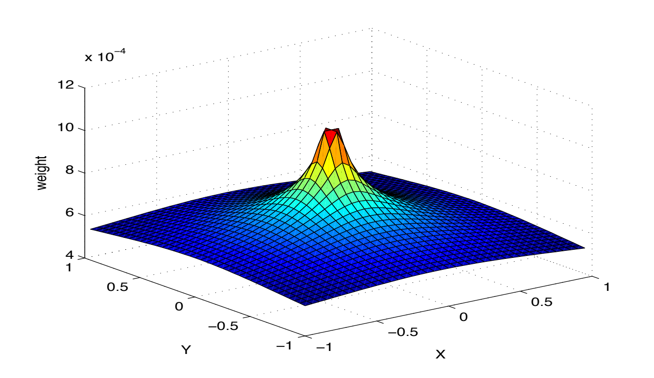
  Hence NCE Control is truly decentralized.

- All trajectories are statistically independent for all finite population sizes N.

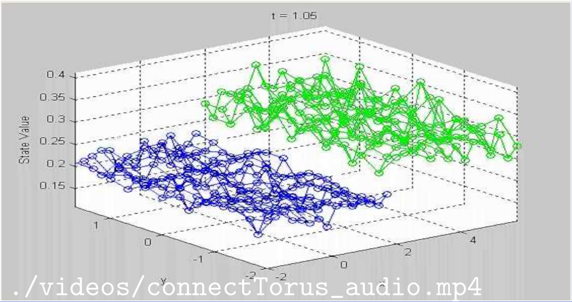
Part 4 – Localization of Influence

Consider the 2-D interaction:

- Partition $[-1,1] \times [-1,1]$ into a 2-D lattice
- Weight decays with distance by the rule $\omega^{(N)}_{p_i p_j} = c|p_i - p_j|^{-\lambda}$ where c is the normalizing factor and $\lambda \in (0,2)$



Part 4 – Separated and Linked Populations



2-D System

Part 5 – Nonlinear MF Systems

- In the infinite population limit, a representative agent satisfies a controlled McKean-Vlasov Equation:

$$dx_t = f[x_t, u_t, \mu_t]dt + \sigma dw_t, \quad 0 \le t \le T$$

with $f[x, u, \mu_t] = \int_{\mathbb{R}} f(x, u, y)\mu_t(dy), \quad x_0, \mu_0$ given and $\mu_t(\cdot) =$ distribution of population states at $t \in [0, T]$.

- In the infinite population limit, individual Agents' Costs:

$$J(u, \mu) \triangleq E \int_0^T L[x_t, u_t, \mu_t]dt,$$

where $L[x, u, \mu_t] = \int_{\mathbb{R}} L(x, u, y)\mu_t(dy)$.

Part 5 – Mean Field and McK-V-HJB Theory

- Mean Field Triple (HMC, 2006, LL, 2006-07):

[MF-HJB] $$-\frac{\partial V}{\partial t} = \inf_{u \in U} \left\{ f[x,u,\mu_t] \frac{\partial V}{\partial x} + L[x,u,\mu_t] \right\} + \frac{\sigma^2}{2} \frac{\partial^2 V}{\partial x^2}$$
$$V(T,x) = 0, \qquad (t,x) \in [0,T) \times \mathbb{R}$$

[MF-FPK] $$\frac{\partial p(t,x)}{\partial t} = -\frac{\partial \{ f[x,u,\mu] p(t,x) \}}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(t,x)}{\partial x^2}$$

[MF-BR] $$u_t = \varphi(t, x | \mu_t), \quad (t,x) \in [0,T] \times \mathbb{R}$$

- Closed-loop McK-V equation:

$$dx_t = f[x_t, \varphi(t, x | \mu_\cdot), \mu_t] dt + \sigma dw_t, \quad 0 \le t \le T$$

Part 5 – Mean Field and McK-V-HJB Theory

- Mean Field Triple (HMC, 2006, LL, 2006-07):

[MF-HJB] $$-\frac{\partial V}{\partial t} = \inf_{u\in U}\left\{ f[x,u,\mu_t]\frac{\partial V}{\partial x} + L[x,u,\mu_t]\right\} + \frac{\sigma^2}{2}\frac{\partial^2 V}{\partial x^2}$$

$$V(T,x)=0, \qquad (t,x)\in[0,T)\times\mathbb{R}$$

[MF-FPK] $$\frac{\partial p(t,x)}{\partial t} = -\frac{\partial\{f[x,u,\mu]p(t,x)\}}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p(t,x)}{\partial x^2}$$

[MF-BR] $$u_t = \varphi(t,x|\mu_t), \quad (t,x)\in[0,T]\times\mathbb{R}$$

- Closed-loop McK-V equation:

$$dx_t = f[x_t,\varphi(t,x|\mu.),\mu_t]dt + \sigma dw_t, \quad 0\le t\le T$$

Yielding Nash Certainty Equivalence Principle expressed in terms of McKean-Vlasov HJB Equation, hence achieving the highest Great Name Frequency possible for a Systems and Control Theory result.

Part 6 – Adaptive NCE Theory:

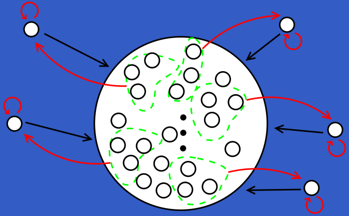
Certainty Equivalence Stochastic Adaptive Control (SAC) replaces unknown parameters by their recursively generated estimates

Key Problem:

To show this results in asymptotically optimal system behaviour in the $\epsilon-$Nash sense

Known:

- $\mathbf{Q}$
- $\mathbf{R}$



Observed:

- $x_i(t)$
- $u_i(t)$
- $\{x_j, u_j; j \in Obs_i(N)\}$

Estimated:

- $\hat{\mathbf{A}}_i$, $\hat{\mathbf{B}}_i$
- $F_{\hat{\zeta}}(\theta)$

- A_i observes a random subset $\mathrm{Obs}_i(N)$ of all agents s.t.
  $$|\mathrm{Obs}_i(N)| \to \infty, \ |\mathrm{Obs}_i(N)|/N \to 0 \text{ as } N \to \infty$$
- $\theta_i^{\mathrm{T}} = (\mathbf{A}_i, \mathbf{B}_i)$
- $F_\zeta = F_\zeta(\theta), \quad \theta \in \mathbf{\Theta} \subset\subset \mathbb{R}^{n(2n+m)}, \quad \zeta \in P \subset\subset \mathbb{R}^p$

Part 6 – NCE-SAC Cost Function

Each agent's Long Run Average (LRA) Cost Function:

$$J_i(\hat{u}_i, \hat{u}_{-i}) = \limsup_{T\to\infty} \frac{1}{T} \int_0^T \left\{ [x_i(t) - m_i(t)]^{\mathrm{T}} \mathbf{Q} [x_i(t) - m_i(t)] + \hat{u}_i^{\mathrm{T}}(t) \mathbf{R} \hat{u}_i(t) \right\} dt$$

$$1 \le i \le N, \quad a.s.$$

Part 6 – NCE-SAC Control Algorithm

For agent A_i, $t \geq 0$:

(I) Self Parameter Identification:
Solve the RWLS Equations for the dynamical parameters $[\hat{\mathbf{A}}_{i,t}, \hat{\mathbf{B}}_{i,t}]$:

(II) Popn. Parameter Identification:

(a) Solve the RWLS equations for the dynamical parameters $\{\hat{\mathbf{A}}_{j,t},\ \hat{\mathbf{B}}_{j,t},\ j \in Obs_i(N)\}$

(b) Solve the MLE equation at $\hat{\theta}_{i,t}^{[1:N_0]} = [\hat{\mathbf{A}}_{j,t}, \hat{\mathbf{B}}_{j,t}]$, $j \in Obs_i(N)$ to estimate ζ^0 via $\hat{\zeta}_{i,t}^N = \arg\min_{\zeta \in P} L(\hat{\theta}_{i,t}^{[1:N_0]}; \zeta)$, $\quad N_0 = |Obs_i(N)|$, and solve the set of NCE Equations for all $\theta \in \mathbf{\Theta}$ generating $x^*\left(\tau, \hat{\zeta}_{i,t}^N\right)$, $\tau \geq t$.

(III) Solve the NCE Control Law at $\hat{\theta}_{i,t}$ and $\hat{\zeta}_{i,t}^N$:

(a) $\hat{\mathbf{\Pi}}_t$: Solve the Riccati Equation at $\hat{\theta}_{i,t}$

(b) $\hat{s}(t)$: Solve the mass control offset at $\hat{\theta}_{i,t}$ and $x^*\left(\tau, \hat{\zeta}_{i,t}^N\right)$

(c) The control law from Certainty Equivalence Adaptive Control:

$$\hat{u}^0(t) = -\mathbf{R}^{-1}\hat{\mathbf{B}}_t^{\mathrm{T}}\left(\hat{\mathbf{\Pi}}_t x(t) + \hat{s}(t)\right) + \xi_k\left[\epsilon(t) - \epsilon(k)\right]$$

Dither weighting: $\xi_k^2 = \frac{\log k}{\sqrt{k}}$, $\quad k \geq 1 \qquad \epsilon(t) =$ Wiener Process

Part 6 – NCE-SAC - Self & Popn. Ident.

Theorem: (AK & PEC, 2010)

Hypotheses: Subject to the conditions above, assume each agent A_i:

(i) Observes a random subset $\mathrm{Obs}_i(N)$ of the total population N s.t. $|\mathrm{Obs}_i(N)| \to \infty,\ |\mathrm{Obs}_i(N)|/N \to 0, \quad N \to \infty,$

(ii) Estimates own parameter $\hat{\theta}_{i,t}$ via RWLS

(iii) Estimates the population distribution parameter $\hat{\zeta}_{i,t}^N$ via RMLE

(iv) Computes $\hat{u}_i^0(t)$ via the extended NCE equations plus dither, where $\bar{x}(\tau, \hat{\zeta}_{i,t}^N) = \int_{\boldsymbol{\Theta}} \bar{x}(\tau, \theta)\ dF_{\hat{\zeta}_{i,t}^N}(\theta)$.

□

Theorem: (AK & PEC, 2010)

Implications: Then, as $t \to \infty$ and $N \to \infty$:

(i) $\hat{\theta}_{i,t} \to \theta_i^0$ a.s. $1 \le i \le N$

(ii) $\hat{\zeta}_{i,t}^N \to \zeta^0 \in P$ w.p.1 and hence, $F_{\hat{\zeta}_{i,t}^N} \to F_{\zeta^0}$ a.s. (weak convergence on P), $1 \le i \le N$ and the set of controls $\{\hat{\mathcal{U}}_{nce}^N;\ 1 \le N < \infty\}$ is s.t.

(iii) Each $S(A_i), 1 \le i \le N$, is an $LRA - L^2$ stable system.

(iv) $\{\hat{\mathcal{U}}_{nce}^N; 1 \le N < \infty\}$ yields a (strong) ϵ-Nash equilibrium for all ϵ

(v) Moreover $J_i^\infty(\hat{u}_i, \hat{u}_{-i}) = J_i^\infty(u_i^0, u_{-i}^0)$ w.p.1, $1 \le i \le N$

□

Part 6 – NCE-SAC Simulation

- 400 Agents
- System matrices $\{A_k\}, \{B_k\}, \quad 1 \leq k \leq 400$

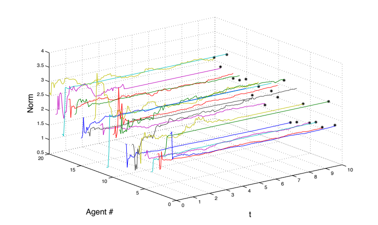
$$A \triangleq \begin{bmatrix} -0.2 + a_{11} & -2 + a_{12} \\ 1 + a_{21} & 0 + a_{22} \end{bmatrix} \quad B \triangleq \begin{bmatrix} 1 + b_1 \\ 0 + b_2 \end{bmatrix}$$

- Population dynamical parameter distribution a_{ij}'s and b_i's are independent.

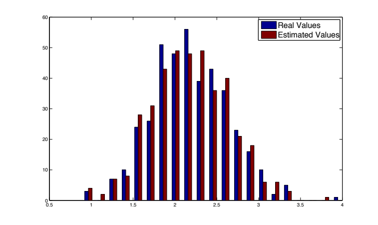
$$a_{ij} \sim N(0, 0.5) \qquad b_i \sim N(0, 0.5)$$

- Population distribution parameters:
  $\bar{a}_{11} = -0.2, \quad \sigma^2_{a_{11}} = 0.5, \quad \bar{b}_{11} = 1, \quad \sigma^2_{b_{11}} = 0.5$ etc.
- All agents performing individual parameter and population distribution parameter estimation
- Each of 400 agents observing its own 20 randomly chosen agents' outputs and control inputs

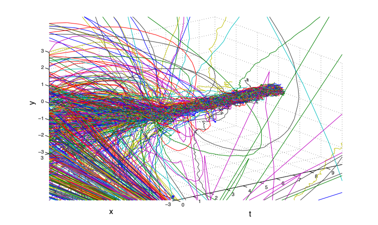
Part 6 – NCE-SAC Simulation



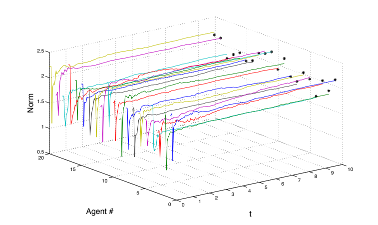
self parameter est.



self parameter est. histogram

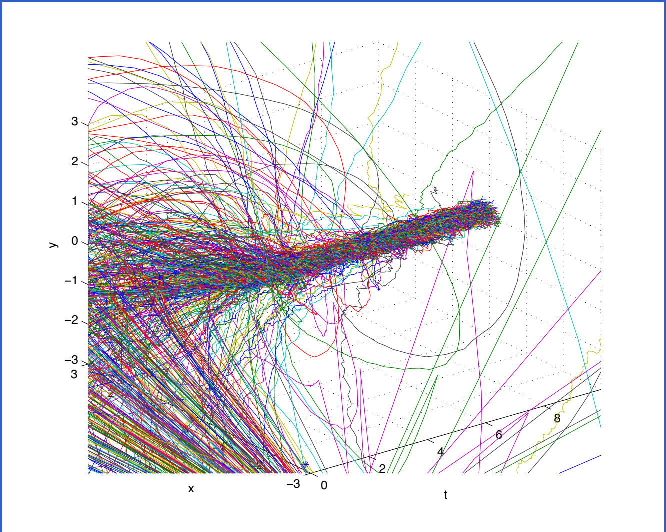


state trajectories



popn. parameter est.

Part 6 – NCE-SAC Animation



Animation of Trajectories

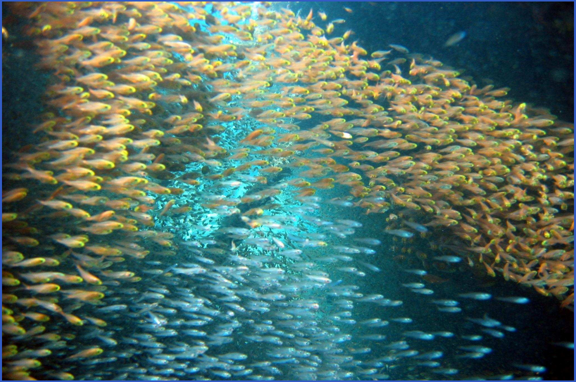
Part 7 – Adaptation based L-F dynamic games

Leader-Follower behaviour:

- is observed in humans [Dyer $et.al.$ 2009] and other species in nature [Couzin $et.al.$ 2005]
- is studied in many disciplines:
  - game theory [Simaan and Cruz 1973]
  - biology [Couzin $et.al.$ 2005]
  - networking [Wang and Slotine 2006]
  - flocking [Gu and Wang 2009]
  - among others.

Part 7 – An Example: Leadership in Animal Groups

Some individuals in the group have more information than others, for instance the location of food or migratory routes [Couzin $et.al.$ 2005]



Part 7 – Problem Formulation: Leaders

Leaders' Dynamics:

$$dz_l^L = [A_l z_l^L + B_l u_l^L]dt + C dw_l, \quad l \in \mathcal{L},\ t \geq 0$$

- $\mathcal{L}$: the set of leaders (L-agents), $N_L = |\mathcal{L}|$
- $z_l^L \in \mathbb{R}^n$: state of the $l-$ th Leader
- $u_l^L \in \mathbb{R}^m$: control input
- $w_l \in \mathbb{R}^p$: disturbance (standard Wiener process)
- $\theta_l \triangleq [A_l, B_l] \in \Theta_l$: dynamic parameter
- $\{z_l^L(0) : l \in \mathcal{L}\}$: initial states, mutually independent

Part 7 – Problem Formulation: Leaders

The LRA Cost Function for Leaders:

$$J_l^L \triangleq \limsup_{T\to\infty} \frac{1}{T}\int_0^T \left\{ \|z_l^L - \Phi^L\|^2 + \|u_l^L\|_R^2 \right\} dt$$

- $\|x\|_R \triangleq (x^T R x)^{1/2}$, $R > 0$ is a symmetric matrix
- $\Psi^L(\cdot) = \frac{1}{N_L}\sum_{i\in\mathcal{L}} z_i^L(\cdot)$
- $\Phi^L(\cdot) \triangleq \lambda h(\cdot) + (1-\lambda)\Psi^L(\cdot)$
- $\lambda \in [0,1]$, h: common reference trajectory of leaders

This cost function is based on a trade-off between moving towards a common reference trajectory, $h(\cdot)$, and staying near the leaders' centroid

Part 7 – Problem Formulation: Followers

Followers' Dynamics:

$$dz_f^F = [A_f z_f^F + B_f u_f^F]dt + C dw_f, \quad f \in \mathcal{F},\ t \geq 0$$

- $\mathcal{F}$: the set of Followers (F-agents)
- $z_f^F \in \mathbb{R}^n$: state of the $f-$th Follower
- $u_f^F \in \mathbb{R}^m$: control input
- $w_f \in \mathbb{R}^p$: disturbance (standard Wiener process)

Part 7 – Problem Formulation: Followers

The LRA Cost Function for Followers:

$$J_f^F \triangleq \limsup_{T\to\infty} \frac{1}{T} \int_0^T \left\{ \|z_f^F - \Psi^L(\cdot)\|^2 + \|u_f^F\|_R^2 \right\} dt$$

- $\Psi^L(\cdot) = \frac{1}{N_L} \sum_{i\in\mathcal{L}} z_i^L(\cdot)$

The followers react by tracking the centroid of the leaders

Part 7 – Leaders' MF (NCE) Equation Systems

Equilibria in infinite population of leaders:

$$
\begin{aligned}
\frac{d\bar{z}_l^L}{dt} &= \left(A_l - B_l R^{-1} B_l^T \Pi_{\theta_l}\right) \bar{z}_l^L - B_l R^{-1} B_l^T s_l^L, \\
\frac{ds_l^L}{dt} &= -\left(A_l - B_l R^{-1} B_l^T \Pi_{\theta_l}\right)^T s_l^L + \Phi^L, \\
r^{L,\infty}(t) &= \int_{\Theta_L} \bar{z}_{\theta_l}^L(t) dF^L(\theta_l), \\
\Phi^L(t) &= \lambda h(t) + (1-\lambda) r^{L,\infty}(t),
\end{aligned}
$$

Leaders' control action $u_l^L(t) \triangleq -R^{-1} B_l^T (\Pi_{\theta_l} z_l^L(t) + s_l^L(t))$ is optimal with respect to $\Phi^L(\cdot)$

Part 7 – MF Equation Systems: Followers

For each follower with $\theta_f = [A_f, B_f]$ when $N_L \to \infty$:

$$\frac{d\bar{z}_f^F}{dt} = \left(A_f - B_f R^{-1} B_f^T \Pi_{\theta_f}\right) \bar{z}_f^F - B_f R^{-1} B_f^T s_f^F,$$

$$\frac{ds_f^F}{dt} = -\left(A_f - B_f R^{-1} B_f^T \Pi_{\theta_f}\right)^T s_f^F + r^{L,\infty},$$

$$r^{L,\infty}(t) = \int_{\Theta_L} \bar{z}_{\theta_l}^L(t) dF^L(\theta_l),$$

Followers' control action $u_f^F(t) \triangleq -R^{-1} B_f^T (\Pi_{\theta_f} z_f^F(t) + s_f^F(t))$ is optimal with respect to $r^{L,\infty}(\cdot)$

Part 7 – Estimation Procedure for the Followers

Each adaptive follower is observing a random subset $\mathcal{M}$ of size M of the leaders' trajectories through the process $y(\cdot)$

$$dy^M = (\frac{1}{M}\sum_{i\in\mathcal{L}}^{M} z_i^L)dt + \frac{1}{M}\sum_{i=1}^{M} Ddv_i$$

- $\{v_i, 1 \leq i \leq M\}$: disturbance (standard Wiener processes)
- $\mathcal{M}$ is chosen by uniformly distributed random selection on $\mathcal{L}$
- $h(\cdot)$, is parameterized with δ from a finite set Δ
- WLG assume $\delta_1 \in \Delta$ is the true unobservable parameter

Part 7 – The Likelihood Ratio

For each Adaptive Follower, define:

- Likelihood Function [T. Duncan 1968]:

$$L_t^M(\delta) \triangleq \exp\{\int_0^t z_{\delta,s}^{L,M} dy_s^M - \frac{1}{2}\int_0^t \|z_{\delta,s}^{L,M}\|^2 ds\}, \quad t > 0$$

$z_{\delta,t}^{L,M} \triangleq \frac{1}{M}\sum_{i\in\mathcal{L}}^M z_{i,\delta}^L(t)$: the centroid of the leaders' states when the defining parameter of $h(\cdot)$ is $\delta \in \Delta$

- Likelihood Ratio:

$$x_i^j(t) \triangleq \frac{L_t^M(\delta_i)}{L_t^M(\delta_j)}, \qquad \delta_i, \delta_j \in \Delta, \quad t > 0$$

which depend explicitly upon the hypotheses δ_i and δ_j

Part 7 – Identifiability Condition

(**A1**) For all $K > 0$ there exists $0 < T_K < \infty$ such that

$$\int_0^t \|r^{L,\infty}_{\delta_i,s} - r^{L,\infty}_{\delta_j,s}\|^2 ds > K,\ \forall \delta_i, \delta_j \in \Delta,\ \delta_i \neq \delta_j,\ t > T_K,$$

$r^{L,\infty}_{\delta}(\cdot)$ is computed by the followers from the leader's MF (NCE) equation system with parameter $\delta \in \Delta$.

Part 7 – Maximum Likelihood Ratio Estimator

For an adaptive follower $f \in \mathcal{F}$ with observation size m, the Maximum Likelihood Ratio (MLR) estimator:

$$\hat{\delta}_f^m(t) \triangleq \{\delta \in \Delta | \frac{L_{t_k}^m(\delta_i)}{L_{t_k}^m(\delta)} < 1 \ \forall \delta_i \in \Delta, \delta_i \neq \delta\}$$

- $t \in [t_k, t_k + \tau_f)$
- τ_f is a pre-specific positive number
- $t_0, t_1, \cdots$ is an infinite sequence, $t_{k+1} - t_k = \tau_f$.

Part 7 – Main Estimation Theorem

Theorem: [Based on Caines 1975]

Under suitable assumptions, for each follower $f \in \mathcal{F}$ there exist non-random T_f, and, with probability one, $M_f(\omega)$, $0 < T_f, M_f(\omega) < \infty$, such that $\hat{\delta}_f^m(t) = \delta_1$ for all $t > T_f$ and $m > M_f(\omega)$. ■

Part 7 – Algorithm for Adaptive Followers

1. Estimation Phase:
   - By observing a sample population of the leaders each follower computes the LRs for alternative values in Δ
   - control laws:

$$\hat{u}_f^{F,\infty}(t) \triangleq -R^{-1} B_f^T (\Pi_{\theta_f} z_f^F(t) + s^F_{\hat{\delta}_f^m(t)}(t))$$

2. Lock-on Phase:
   - the MF (NCE) control laws will necessarily be computed with the true parameter of the reference trajectory

Part 7 – Optimality Property: Followers

Theorem: Under suitable assumptions each follower's adaptive MF (NCE) control strategy is almost surely ϵ_{N_L}-optimal with respect to the leaders' control strategies, i.e.

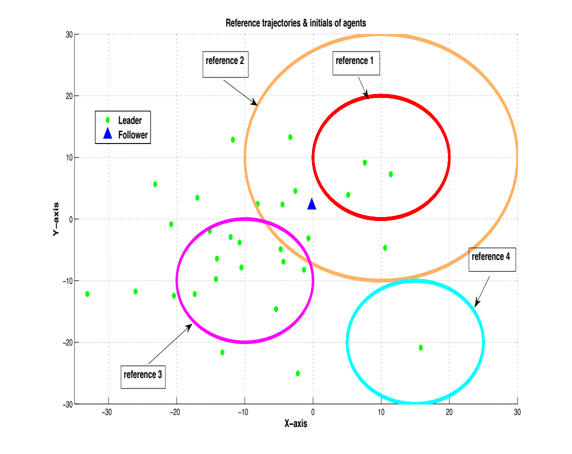
$$J_f^F(\hat{u}_f^{F,\infty}, u_{\delta_1}^{L,\infty}) - \epsilon_{N_L} \le \inf_{u\in\mathcal{U}} J_f^F(u, u_{\delta_1}^{L,\infty}) \le J_f^F(\hat{u}_f^{F,\infty}, u_{\delta_1}^{L,\infty})$$

almost surely and such that $\lim_{N_L\to\infty} \epsilon_{N_L} = 0,\ a.s.$ ■

Part 7 – Simulation

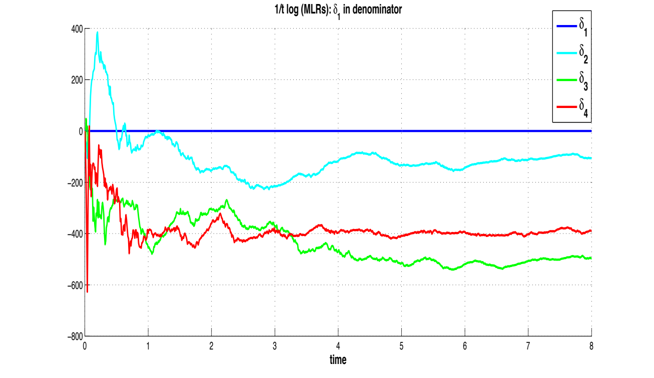
- 30 leaders and one adaptive follower
- $n = 2$, $\lambda = 0.5$, $C = D = 5I$, $R = 0.001I$, $\tau_f = 1$
- $A_f = \begin{bmatrix} -0.2 & 0.5 \\ -0.8 & 0.4 \end{bmatrix}$
- observation size of adaptive follower is 10
- A_l is chosen randomly from a normal probability distribution with zero mean and identity covariance
- The reference trajectories: $[a_1 + b_1 cos(wt)\ a_2 + b_2 sin(wt)]$, $t \in [0, \infty)$, where $\delta = (a_1, b_1, a_2, b_2, w) \in \Delta$
- Δ has four parameters

Part 7 – Simulation



Part 7 – Simulation (cnt)

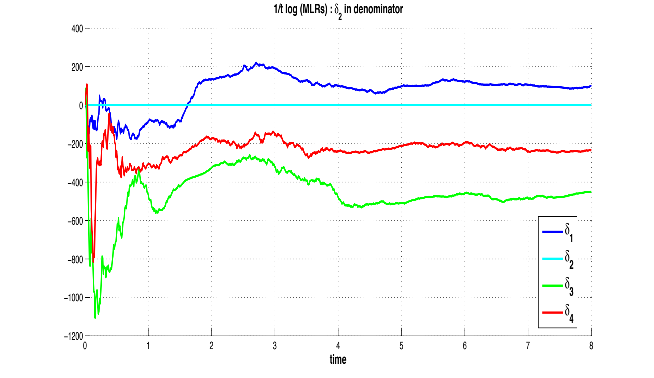
$$\hat{\delta}_f^m(t) \triangleq \{\delta \in \Delta | \frac{L_{t_k}^m(\delta_i)}{L_{t_k}^m(\delta)} < 1 \ \forall \delta_i \in \Delta, \delta_i \neq \delta\}$$



(a) $\log(\frac{L_t^m(\delta_i)}{L_t^m(\delta_1)})$ for $\delta_i \in \Delta$

Part 7 – Simulation (cnt)

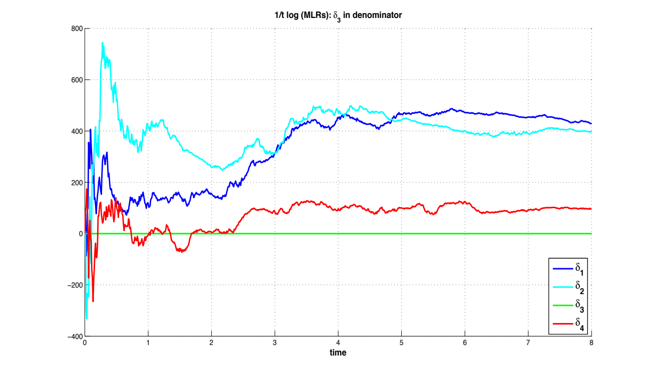
$$\hat{\delta}_f^m(t) \triangleq \{\delta \in \Delta | \frac{L_{t_k}^m(\delta_i)}{L_{t_k}^m(\delta)} < 1 \ \forall \delta_i \in \Delta, \delta_i \neq \delta\}$$



(b) $\log(\frac{L_t^m(\delta_i)}{L_t^m(\delta_2)})$ for $\delta_i \in \Delta$

Part 7 – Simulation (cnt)

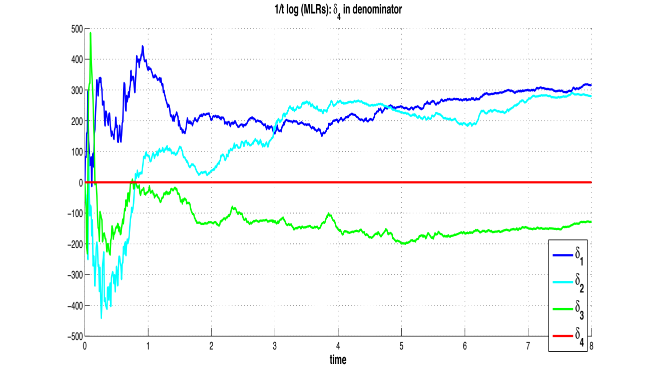
$$\hat{\delta}_f^m(t) \triangleq \{\delta \in \Delta | \frac{L_{t_k}^m(\delta_i)}{L_{t_k}^m(\delta)} < 1 \; \forall \delta_i \in \Delta, \delta_i \neq \delta\}$$



(c) $\log(\frac{L_t^m(\delta_i)}{L_t^m(\delta_3)})$ for $\delta_i \in \Delta$

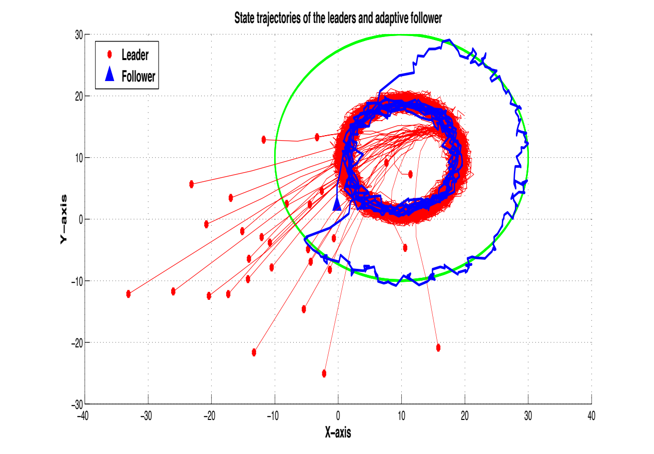
Part 7 – Simulation (cnt)

$$\hat{\delta}_f^m(t) \triangleq \{\delta \in \Delta | \frac{L_{t_k}^m(\delta_i)}{L_{t_k}^m(\delta)} < 1 \; \forall \delta_i \in \Delta, \delta_i \neq \delta\}$$

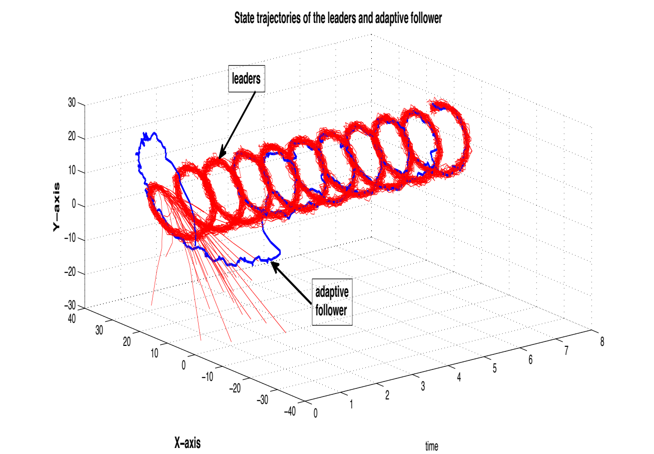


(d) $\log(\frac{L_t^m(\delta_i)}{L_t^m(\delta_4)})$ for $\delta_i \in \Delta$

Part 7 – Simulation (cnt)



Part 7 – Simulation (cnt)



Summary

- NCE Theory solves a class of decentralized decision-making problems with many competing agents.

- Asymptotic Nash Equilibria are generated by the NCE Equations.

- Key intuition:

  Single agent's control = feedback of stochastic local (rough) state + feedback of deterministic global (smooth) system behaviour

- NCE Theory extends to (i) localized problems, (ii) stochastic adaptive control, (iii) egoist-altruist, major agent-minor agent systems, (iv) flocking behaviour, (v) point processes in networks.

Future Directions

- Further development of Minyi Huang's large and small players extension of NCE Theory
- Further development of egoists and altruists version of NCE Theory
- Mean Field stochastic control of non-linear (McKean-Vlasov, YMMS) systems
- Extension of NCE (MF) SAC Theory in richer game theory contexts
- Development of MF Theory towards economic, renewable energy, biological applications
- Development of large scale cybernetics: Systems and control theory for competitive and cooperative systems

Thank You !

谢 谢