Application of Potential Game for Power Control in Wireless Networks and Network Formation

Goto Tatsuhiko
FL10_02_01

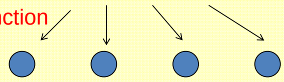
10th,May,2010

Review (Game theoretic approach)

Global objective function

Local objective function

agents



A_i :Action set $\quad A=\prod_{p_i\in p} A_i$:set of joint action

$$a_{-i} = (a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n)$$

joint action $a = (a_i, a_{-i})$

Local objective function $U_i : A \to R$

Control design
1. Designing the player objective function
2. Learning dynamics (repeated game)
   (ex)single stage memory dynamics

Review (Potential game)

Global planner $\phi : A \to \Re$ (potential function)

aligned

Make player's objective function U_i

$$U_i(a_i'', a_{-i}) - U_i(a_i', a_{-i}) = \phi(a_i'', a_{-i}) - \phi(a_i', a_{-i})$$

Changing in the player's objective function
=
Changing in the potential function

Every agent select an action to maximize their objective function

---

Outline

Near-Optimal Power Control in Wireless Networks: A Potential Game Approach

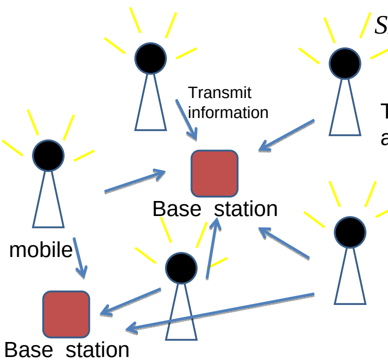
Utku Ozan Candogan, Ishai Menache, Asuman Ozdaglar and Pablo A. Parrilo Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA, 02139

- Background
- Model
- Modified utilities
- Near optimal dynamics
- Convergence analysis
- Simulation Result

---

background

CDMA system



Mobiles' rates

$$SINR_m(P)=\frac{h_{mm}P_m}{N_0+\sum_{k\neq m} h_{km}P_k}$$

The signals of other users are noise

rate $r_m(P)=\log(1+\gamma\, SINR_m(P))$

cost $c_m p_m$

net utility $u_m(p)=r_m(p)-\zeta_m c_m p_m$

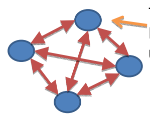
Maximize net utility

---

approach

Situation : Self interested

Interfere with each other



Thinking of Maximize own utility

Object : Design of power control

Tradeoff
rate ⟷ cost
$u_m(p)=r_m(p)-\zeta_m c_m p_m$

Centralized optimization problem with full information
→ difficult for complex

General distributed power control

Potential game approach

Original game g → Approximation game $\tilde{g}$

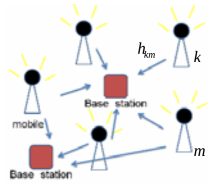
Best response dynamics
Converge within neighborhood of the optimal operating point

Set optimal price c^*
Converge to the optimal operating point

Outline

- Background
- Model
- Modified utilities
- Near optimal dynamics
- Convergence analysis
- Simulation Result

model



mobiles $M=\{1,\cdots,M\}$

Power allocation $P=\{p_1,\cdots,p_M\}$

SINR

$$SINR_m(P)=\frac{h_{mm}P_m}{N_0+\sum_{k\neq m}h_{km}P_k}$$

h_{km} :gain between user k and transmitter m's base station

rate

$$r_m(P)=\log(1+\gamma\, SINR_m(P))$$

cost $c_m p_m$

p_m : transmission power

$$0\leq p_{\min m}\leq p_m\leq \bar{p}_m$$

Objective function (net utility)

$$u_m(p)=r_m(p)-\zeta_m c_m p_m$$

User specific rate vs money Tradeoff coefficient

Power game (definition)

Power game

$$g=<M,\{u_m\}_{m\in M},\{p_m\}_{m\in M}>$$

m's objective function m's action set

Self interested

$$\max_{\tilde{p}_m\in P_m} u_m(\tilde{p}_m,p_{-m})$$

Nash equilibrium (NE)

$$u_m(p)\geq u_m(\tilde{p}_m,p_{-m})\qquad \forall\tilde{p}_m\in p_m,\forall m\in M$$

ε − Nash equilibrium

$$u_m(p)\geq u_m(q_m,p_{-m})-\varepsilon\qquad \forall q_m\in p_m,\forall m\in M$$

Central planner wishes to impose some performance objective

System Utility

$$\max_{p\in P} U_0(p)$$

(ex) $U_0(p)=\sum_m r_m(p)$

Sum rate objective

Optimal solution p^* (desired operating point)

Modified utilites

modified Utility

$$r_m(P)=\log(1+\gamma\, SINR_m(P))$$

$$SINR_m(P)=\frac{h_{mm}P_m}{N_0+\sum_{k\neq m}h_{km}P_k}$$

$$\tilde{u}_m(p)=\tilde{r}_m(p)-\zeta_m c_m p_m$$

$$\tilde{r}_m(P)=\log(\gamma\, SINR_m(P))$$

Approximation is good

Spreading gain $\gamma>>1$

or

$h_{mm}>>h_{mk}$

Can make potential function

$$\phi(p)=\sum_m \log(p_m)-\zeta_m c_m p_m$$

$$(\phi(p_m,p_{-m})-\phi(q_m,p_{-m})=\tilde{u}_m(p_m,p_{-m})-\tilde{u}_m(q_m,p_{-m}))$$

Strictly concave → unique NE

Potential game

$$\tilde{g}=<M,\{\tilde{u}_m\}_{m\in M},\{p_m\}_{m\in M}>$$

Assigning prices

Assigning prices c^* to coincide with NE of $\tilde{g}$ and p^*

$$\tilde{u}_m(p)=\tilde{r}_m(p)-\zeta_m c_m p_m$$

[Theorem]

Let p^* be the desired operating point .Then the prices c^* are given by

$$c_m^*=(\zeta_m p_m^*)^{-1}\qquad m\in M$$

(proof) $\phi(p)$ Strictly concave → unique NE

→ Maxima of $\phi(p)$ is NE

$$\phi(p)=\sum_m \log(p_m)-\zeta_m c_m^* p_m$$

$$c_m^*=(\zeta_m p_m^*)^{-1}$$

$$\frac{\partial\phi}{\partial p_m}=\frac{1}{p_m}-\frac{1}{p_m^*}\ \Rightarrow\ p=p^*\rightarrow\frac{\partial\phi}{\partial p_m}=0$$

⇒ p^* Global maximum of the potential

Near optimal dynamics

p^* is not NE of the game g with c^*

→ Converge neighbor of p^* ?

Best Response dynamics

$$p_m\leftarrow p_m+\alpha(\beta_m(p_{-m})-p_m)$$

Best Response $\beta_m(p_{-m})=\arg\max_{p_m\in P_m} u_m(p_m,p_{-m})$

most good action for user m

$$\dot{p}_m=\beta_m(p_{-m})-p_m$$

α is small

$\tilde{g}$ with $c=c^*$ $\xrightarrow{\text{BR}}$ Converge to p^*

(Lyapunov analysis)

How about g ?

Outline

- Background
- Model
- Modified utilities
- Near optimal dynamics
- Convergence analysis
- Simulation Result

Convergence analysis

Best Response of $\tilde{g}$ $\quad \tilde{\beta}_m(p_{-m}) = \arg\max_{p_m \in P_m} \tilde{u}_m(p_m, p_{-m})$

$= \arg\max_{p_m \in P_m} \phi(p_m, p_{-m})$ (From PG)

ε -equilibria of $\tilde{g}$ $\quad \tilde{I}_\varepsilon = \{ p \mid \tilde{u}_m(p_m, p_{-m}) \geq \tilde{u}_m(q_m, p_{-m}) - \varepsilon \}$

[Lemma]

The BR in g converge to $\tilde{I}_\varepsilon$ $\quad \varepsilon \leq \frac{1}{\gamma} \sum_{m \in M} \frac{1}{SINR_{\min m}}$

$SINR_{\min m}(P) = \frac{h_{mm} P_{\min m}}{N_0 + \sum_{k \neq m} h_{km} P_{\max k}}$

(proof) $\quad \bar{\phi}$: maximum value of ϕ

$V = \bar{\phi} - \phi \geq 0$: Lyapunov function

proof

$$-\dot{V} = \sum_{m \in M} \frac{\partial \phi}{\partial p_m} \left(\tilde{\beta}_m(p_{-m}) - p_m \right) + \sum_{m \in M} \frac{\partial \phi}{\partial p_m} \left(\beta_m(p_{-m}) - \tilde{\beta}_m(p_{-m}) \right)$$

$$\sum_{m \in M} \frac{\partial \phi}{\partial p_m} \left(\tilde{\beta}_m(p_{-m}) - p_m \right) \geq \tilde{u}_m\left(\tilde{\beta}_m(p_{-m}), p_{-m} \right) - \tilde{u}_m(p_m, p_{-m})$$

$$\left| \frac{\partial \phi}{\partial p_m} \left(\beta_m(p_{-m}) - \tilde{\beta}_m(p_{-m}) \right) \right| \leq \left(\frac{1}{p_{\min m}} - \frac{1}{\bar{p}_m} \right) \frac{\xi_m}{\gamma}$$

$$\left(\xi_m = \frac{N_0}{h_{mm}} + \sum_{k \neq m} \frac{h_{km}}{h_{mm}} \bar{p}_k \right)$$

$$-\dot{V} \geq \sum_{m \in M} \left(\tilde{u}_m\left(\tilde{\beta}_m(p_{-m}), p_{-m} \right) - \tilde{u}_m(p_m, p_{-m}) \right) - \sum_{m \in M} \left(\frac{1}{p_{\min m}} - \frac{1}{\bar{p}_m} \right) \frac{\xi_m}{\gamma}$$

$$\sum_{m \in M} \left(\tilde{u}_m\left(\tilde{\beta}_m(p_{-m}), p_{-m} \right) - \tilde{u}_m(p_m, p_{-m}) \right) \geq \sum_{m \in M} \frac{\xi_m}{\gamma p_{\min m}}$$

$$\dot{V} \leq -\sum_{m \in M} \frac{\xi_m}{\gamma \bar{p}_m}$$

proof

$$\sum_{m \in M} \left(\tilde{u}_m\left(\tilde{\beta}_m(p_{-m}), p_{-m} \right) - \tilde{u}_m(p_m, p_{-m}) \right) \geq \sum_{m \in M} \frac{\xi_m}{\gamma p_{\min m}}$$

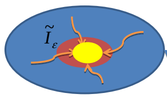
$$\dot{V} \leq -\sum_{m \in M} \frac{\xi_m}{\gamma \bar{p}_m}$$

Converge to this set from Lyapunov method

$$\tilde{u}_m\left(\tilde{\beta}_m(p_{-m}), p_{-m} \right) - \tilde{u}_m(p_m, p_{-m}) \leq \sum_{m \in M} \frac{\xi_m}{\gamma p_{\min m}}$$

$$\varepsilon \leq \frac{1}{\gamma} \sum_{m \in M} \frac{1}{SINR_{\min m}}$$

$$\left(\tilde{I}_\varepsilon = \{ p \mid \tilde{u}_m(p_m, p_{-m}) \geq \tilde{u}_m(q_m, p_{-m}) - \varepsilon \} \right)$$



$$\dot{V} \leq -\sum_{m \in M} \frac{\xi_m}{\gamma \bar{p}_m}$$

How far ?

How far the set of ε- equilibria of $\tilde{g}$ from p^*?

[Theorem] $\quad \left| \tilde{p}_m - p_m^* \right| \leq \bar{P}_m \sqrt{2\varepsilon} \qquad \tilde{p} \in \tilde{I}_\varepsilon$

ε -equilibria of $\tilde{g}$

Global maximum of the potential

$\tilde{I}_\varepsilon = \{ p \mid \tilde{u}_m(p_m, p_{-m}) \geq \tilde{u}_m(q_m, p_{-m}) - \varepsilon \}$

(proof)

$$\phi\left(p_m^*, \tilde{p}_{-m}\right) - \phi\left(\tilde{p}_m, \tilde{p}_{-m}\right) \leq \varepsilon$$

$$\left(\log(p_m^*) - \lambda_m p_m^* \right) - \left(\log(\tilde{p}_m) - \lambda_m \tilde{p}_m \right) \leq \varepsilon \qquad \left(\phi(p) = \sum_m \log(p_m) - \zeta_m c_m^* p_m \right)$$

$$f_m(p_m^*) - f_m(\tilde{p}_m) \leq \varepsilon \qquad \left(f_m = \log(p_m) - \lambda_m p_m \right)$$

proof

$$f_m = \log(p_m) - \lambda_m p_m$$

$$f_m(\tilde{p}_m) = f_m(p_m^*) + (\tilde{p}_m - p_m^*) \frac{\partial f_m(p_m^*)}{\partial p_m} + \frac{1}{2} (p_m^* - \tilde{p}_m)^2 \frac{\partial^2 f_m(p_m^* + \alpha(\tilde{p}_m - p_m^*))}{\partial p_m^2}$$

p^* Is desired operating point

$$\frac{\partial f_m(p_m^*)}{\partial p_m} = \frac{\partial \phi(p_m^*)}{\partial p_m} = 0$$

$$f_m(p_m^*) - f_m(\tilde{p}_m) = \frac{1}{2} (p_m^* - \tilde{p}_m)^2 \frac{1}{(p_m^* + \alpha(\tilde{p}_m - p_m^*))^2}$$

$$2\left(p_m^* + \alpha(\tilde{p}_m - p_m^*)\right)^2 \left(f_m(p_m^*) - f_m(\tilde{p}_m) \right) = (p_m^* - \tilde{p}_m)^2$$

$$2\varepsilon \bar{P}_m^2 \geq \left(p_m^* - \tilde{p}_m \right)^2 \qquad \left(f_m(p_m^*) - f_m(\tilde{p}_m) \leq \varepsilon, 0 < p_m^*, \tilde{p}_m \leq \bar{p}_m \right)$$

$$\left| \tilde{p}_m - p_m^* \right| \leq \bar{P}_m \sqrt{2\varepsilon}$$

Near optimal performance

Near optimal performance in terms of system utility

Performance loss decrease with small ε
increase with large L, L_m

(ex) $U_0(p) = \sum_m r_m(p)$
Sum rate objective

[Theorem] Let $\varepsilon \le \frac{1}{\gamma} \sum_{m \in M} \frac{1}{SINR_{\min m}}$ then

(1) U_0 is Lipschitz continuous function ,with L . Then

$$\left|U_0(p^*) - U_0(\tilde{p})\right| \le \sqrt{2\varepsilon} L \sqrt{\sum_{m \in M} \bar{P}_m^{\,2}}$$

(2) Assume that U_0 is a continuous differentiable function such that $\left|\frac{\partial U_0}{\partial p_m}\right| \le L_m$ Then $\left|U_0(p^*) - U_0(\tilde{p})\right| \le \sqrt{2\varepsilon} \sum_{m \in M} \bar{P}_m L_m$

Difference between p^* and $\tilde{p}$ ⇒ Difference between $U_0(p^*)$ and $U_0(\tilde{p})$

Outline

- Background
- Model
- Modified utilities
- Near optimal dynamics
- Convergence analysis
- Simulation Result

Simulation result

・Three users

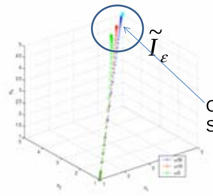
・desired operating point $p^* = [5,5,5]$

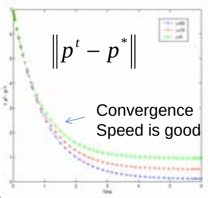
$r_m(P) = \log(1 + \gamma SINR_m(P))$ $\gamma \in \{5,10,15\}$

$SINR_m(P) = \frac{h_{mm} P_m}{N_0 + \sum_{k \ne m} h_{km} P_k}$

$N_0 = 1$ $h_{km} \in [0,2]$ $h_{mm} \in [2,4]$

Do the BR dynamics with c^*





⇒ Big γ is good

$\varepsilon \le \frac{1}{\gamma} \sum_{m \in M} \frac{1}{SINR_{\min m}}$

⇒ Monotonously decreasing

Simulation result (system utility)

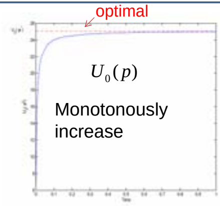
Sum rate objective

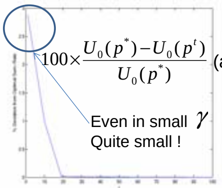
$U_0(p) = \sum_m r_m(p)$ ⇒ $\max_{p \in P} U_0(p)$

[Theorem]
U_0 is a continuous differentiable function
$\left|\frac{\partial U_0}{\partial p_m}\right| \le \frac{M-1}{p_{\min m}}$

⇒ $\left|U_0(p^*) - U_0(\tilde{p})\right| \le \sqrt{2\varepsilon}(M-1) \sum_{m \in M} \frac{\bar{P}_m}{P_{\min m}}$

$\left(\left|U_0(p^*) - U_0(\tilde{p})\right| \le \sqrt{2\varepsilon} \sum_{m \in M} \bar{P}_m L_m \right)$





General type

$u_m \rightarrow \phi \rightarrow \hat{u}_m$

Find the most close potential function of the game by u_m

Find the most close objective function of ϕ to u_m

$$d^2(g) = \min_{\phi \in C_0} \left\| \delta_0 \phi - \sum_{m \in M} D_m u^m \right\|_2^2$$

$$\hat{u}^m = \underset{\bar{u}^m}{\operatorname{argmin}} \left\| u^m - \bar{u}^m \right\|_2^2$$

$$D_m \bar{u}_m = D_m \phi$$

D_m : Difference operator

$(D_m \phi)(p,q) = W^n(p,q)(\phi(q) - \phi(p))$ $W^n(p,q) = 1$ $p \ne q$ $W^n(p,q) = 0$ $p = q$

Potential game
$D_m u_m = D_m \phi$

⇒ $\sum_{m \in M} D_m u_m = \sum_{m \in M} D_m \phi = \delta_0 \phi$

combinatorial gradient operator $\delta_0 = \sum_{m \in M} D_m$

Theorem of Projection

[Theorem] Optimal projection

$$\phi = \left(\sum_{m \in \mathcal{M}} \Pi_m \right)^{\dagger} \sum_{m \in \mathcal{M}} \Pi_m u^m$$

$$\hat{u}^m = (I - \Pi_m) u^m + \Pi_m \left(\sum_{k \in \mathcal{M}} \Pi_k \right)^{\dagger} \sum_{k \in \mathcal{M}} \Pi_k u_k . \quad \Pi_m = D_m^* D_m$$

[Theorem]

Any equilibrium of $\tilde{g}$ is an ε–equilibrium of g .

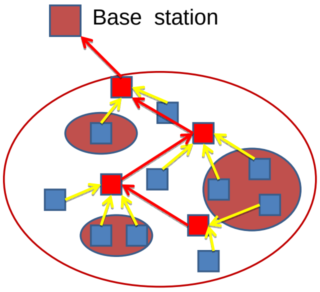
$\varepsilon \le \sqrt{2} d(g)$ $d^2(g) = \min_{\phi \in C_0} \left\| \delta_0 \phi - \sum_{m \in M} D_m u^m \right\|_2^2$

g : game

$\tilde{g}$: projection of the game g

$u_m(p) \ge u_m(q_m, p_{-m}) - \varepsilon$

Coverage considering wireless network



Searching agent

Information transmission agent

Information from (searching agent) to (information transmission agent)

Information from (information transmission agent) to (information transmission agent)

Object : optimal coverage and
make small transmission cost of information
→ Several Information transmission agent needed

Agent can change from (searching agent) to (information transmission agent) and from (information transmission agent) to (searching agent)

Outline

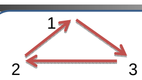
Distributed Dynamic Reinforcement of Efficient Outcomes in Multiagent Coordination and Network Formation*

Georgios C. Chasparis[†] and Jeff S. Shamma[†]

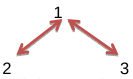
November 7, 2009

- Background
- Reinforcement learning
- Asymptotic stability analysis
- Dynamic Reinforcement
- Asymptotic stability of RADR
- Simulation Result

Distributed network formation



→ Efficient network



Non-efficient nash network

Want to make it!

Reward

$$R_i(\alpha)=\underbrace{\sum_{s\in I, s\neq i}\chi_\alpha(i\to s)}_{\text{connection reward}}-\underbrace{\kappa_0|\alpha_i|}_{\text{cost}}$$

$\chi_\alpha(i\leftarrow s)=1$: g_α contain the path $(i\leftarrow s)$

$\chi_\alpha(i\leftarrow s)=0$: otherwise

Actions of agent

(ex) $A_1=\{\{1\},\{2\},\{3\},\{2,3\}\}$

Nash network

α^* Nash network ⟺ $R_i(\alpha_i^*,\alpha_{-i}^*)\geq R_i(\alpha_i',\alpha_{-i}^*)\quad \alpha_i'\in A_i\setminus\alpha_i^*$

wheel network

Connected network is uniquely defined by a path $(i\leftarrow i)$

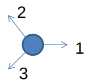
→ Every agent realizes its maximum possible utility

Reinforcement learning

Learning algorithm

$$x_i(k+1)=\underbrace{x_i(k)}_{\text{Propensities of action}}+\varepsilon(k)\underbrace{R_i(\alpha(k))}_{\text{reward}}\underbrace{[\alpha_i(k)-x_i(k)]}_{\text{Action at time k}}$$

(ex)



$x_i(k)=\begin{bmatrix}0.2\\0.3\\0.5\end{bmatrix}$ Do action 1 and get reward 1.5 → $x_i(k+1)=\begin{bmatrix}0.2\\0.3\\0.5\end{bmatrix}+\frac{1.5}{k+1}\begin{bmatrix}0.8\\-0.3\\-0.5\end{bmatrix}$

Probability of action selection

$$\sigma_i(k)=(1-\lambda)x_i(k)+\underbrace{\lambda\frac{1}{|A_i|}\mathbf{1}}_{\text{Random selection}}$$

$\varepsilon(k)=\frac{1}{k+1}$
Propensities are proportional to the cumulative reward

→ $x_i(k)$ evolves over the probability simplex $\Delta(|A_i|)\quad(\Delta(m)=\{v\in\Re^m\,|\,v\geq 0, 1^T v=1\})$

Asymptotic stability analysis

$$x_i(k+1)=x_i(k)+\varepsilon(k)R_i(\alpha(k))[\alpha_i(k)-x_i(k)]$$

$$\bar{g}_i(x(k))=\bar{r}_i(x(k))-\bar{R}_i(x(k))x_i(k)$$

$\bar{r}_i(x(k))=E[R_i(\alpha(k))\alpha_i(k)\,|\,x(k)]\quad \bar{R}_i(x(k))=E[R_i(\alpha(k))\,|\,x(k)]$

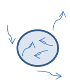
$$\xi_i(k)=R_i(\alpha(k))[\alpha_i(k)-x_i(k)]-\bar{g}_i(x(k))$$

$$x_i(k+1)=x_i(k)+\varepsilon(k)[\underbrace{\bar{g}_i(x(k))}_{\text{deterministic}}+\underbrace{\xi_i(k)}_{\text{noise}}]$$

[proposition3.1] $(\lambda>0)$

$$x(k+1)=x(k)+\varepsilon(k)[\bar{g}(x(k))+\xi(k)]\quad x(k)=\{x_1(k),\cdots,x_n(k)\}$$

$x(k)$ convergence to an invariant set of the $\dot{x}=\bar{g}(x)$
$A\subset\Delta$ be a locally Asymptotically stable set for $\dot{x}=\bar{g}(x)$ → $\Pr[\lim_{k\to\infty}x(k)\in A]>0$



Asymptotic stability analysis

$\dot{x}=\bar{g}(x)$ → Stationary point
$S=\{x\in X:\bar{g}(x)=\bar{r}(x)-\bar{R}(x)x=0\}$

[proposition3.2] $(\lambda>0)$

x^* linearly unstable stationary point of $\dot{x}=\bar{g}(x)$
→ $\Pr[\lim_{k\to\infty}x(k)=x^*]=0$

Expected reward $\bar{v}_i(j,x^*)=E[R_i(\alpha)\,|\,\alpha_i=j,x_{-i}=x_{-i}^*]\quad x^*\in S$

[proposition3.3] (Stationary point)

x^* stationary point of $\dot{x}=\bar{g}(x)$ ⟷ $\bar{v}_i(j,x^*)=c_i\quad j\in A_i$
Expected reward not change by one agent

proposition

[proposition3.4] (pure strategies) $(\lambda = 0)$

pure strategies profile → stationary point of $\dot{x} = \bar{g}(x)$

[proposition3.5] (sensitivity of pure strategies) $(\lambda > 0)$

x^* pure strategies profile and strict NE → Exist a differentiable function $\upsilon^*(\lambda)$,such $\tilde{x} = x^* + \upsilon^*(\lambda)$ stationary point of $\dot{x} = \bar{g}(x)$ $(\lim_{\lambda \to 0} \upsilon^*(\lambda) = \upsilon^*(0) = 0)$

[proposition3.6] (LAS) $(\lambda > 0)$

$\tilde{x}$ Locally asymptotically Stable point of $\dot{x} = \bar{g}(x)$ ↔ $\bar{v}_i(j^*, \tilde{x}) > \bar{v}_i(s, \tilde{x}) \quad \forall s \in A_i \setminus j^*$, $x_i^* = e_{j^*}$

[proposition3.7] $(\lambda > 0)$

$\bar{v}_i(j^*, \tilde{x}) > \bar{v}_i(s, \tilde{x}) \quad \forall s \in A_i \setminus j^*$ $\tilde{x}$:strict NE ⇒ $\Pr[\lim_{k \to \infty} x(k) = \tilde{x}] > 0$ (from p3.6)

$\bar{v}_i(j^*, x^*) < \bar{v}_i(s, x^*) \quad \exists s \in A_i \setminus j^*$ x^* :not NE ⇒ $\Pr[\lim_{k \to \infty} x(k) \in \underbrace{B_\delta(x^*)}_{\text{Neighbor of } x^*}] = 0$

Outline

- Background
- Reinforcement learning
- Asymptotic stability analysis
- Dynamic Reinforcement
- Asymptotic stability of RADR
- Simulation Result

Dynamic Reinforcement

Before depend only on the probability distribution ⇒ Also affected by the history of x_i

⇒ Goal is to investigate the effects on convergence to an Efficient pure equilibrium

Learning algorithm

$$x_i(k+1) = \underbrace{x_i(k)}_{\text{Propensities of action}} + \varepsilon(k)\underbrace{R_i(\alpha(k))}_{\text{reward}}[\underbrace{\alpha_i(k)}_{\text{Action at time k}} - x_i(k)]$$

Probability of action selection

$$\sigma_i(k) = \Pi_\Delta[(1-\lambda)(x_i(k) + \underbrace{u_i(k)}_{\text{Correspond to history of } x_i}) + \lambda \frac{1}{|A_i|}\mathbf{1}] \qquad \Pi_\Delta[x] = \arg\min_{s \in \Delta(m)} |x - s| \qquad \varepsilon(k) = \frac{1}{k+1}$$

$\boxed{u_i(k) = \gamma_i(\rho_i(k))(x_i(k) - y_i(k))}$ $\gamma_i(\rho_i(k))$: RADR parameter

$y_i(k+1) = y_i(k) + \varepsilon(k)(x_i(k) - y_i(k))$ Running average of x_i

$\rho_i(k+1) = \rho_i(k) + \varepsilon(k)(R_i(\alpha(k)) - \rho_i(k))$ Running average of $R_i(\alpha(k))$

Asymptotic stability of RADR

$z(k) = \begin{pmatrix} x(k) \\ y(k) \\ \rho(k) \end{pmatrix}$ ⇒ Relevant ODE $\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\rho} \end{pmatrix} = \begin{pmatrix} \bar{g}(z) \\ x - y \\ \bar{R}_i(z) - \rho \end{pmatrix}$ α is depend on y, ρ ⇒ linearization $\frac{d}{dt}\begin{pmatrix} \delta x(t) \\ \delta y(t) \\ \delta\rho(t) \end{pmatrix} = \tilde{A}^{\lambda,\gamma} \begin{pmatrix} \delta x(t) \\ \delta y(t) \\ \delta\rho(t) \end{pmatrix}$

$x_i(t) = \tilde{x}_i + N\delta x_i(t)$
$y_i(t) = \tilde{y}_i + N\delta y_i(t)$
$\delta\rho(t) = \rho(t) - \tilde{\rho}$

[Theorem 4.1] (LAS of RADR)

equilibrium $\tilde{z} = (\tilde{x}, \tilde{y}, \tilde{\rho})$ is LAS point of linearization ↔ RADR parameter $0 \le \gamma_i(\tilde{\rho}_i(k)) \le \frac{\bar{v}_i(j^*, \tilde{x}) + 1}{\bar{v}_i(s, \tilde{x})} - 1 \quad \forall s \ne j^*$

[Theorem 4.2] (RADR convergence)

$0 \le \gamma_i(\tilde{\rho}_i(k)) \le \frac{\bar{v}_i(j^*, \tilde{x}) + 1}{\bar{v}_i(s, \tilde{x})} - 1 \quad \forall s \ne j^*, \forall i$ ⇒ $\Pr[\lim_{k \to \infty} x(k) = \tilde{x}] > 0$

$\gamma_i(\tilde{\rho}_i(k)) \ge \frac{\bar{v}_i(j^*, \tilde{x}) + 1}{\bar{v}_i(s, \tilde{x})} - 1 \quad \exists s \ne j^*, \exists i$ ⇒ $\Pr[\lim_{k \to \infty} x(k) = \tilde{x}] = 0$

example

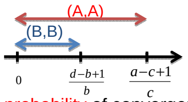
	2.A	2.B
1.A	a,a	b,c
1.B	c,b	d,d

Symmetric coordination game

(case1) $a > c, d > b, (a-c) > (d-b), a < d$

⇒ (A,A) risk dominant
(B,B) payoff dominant

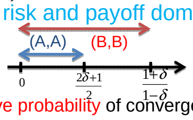
$0 \le \gamma_i(\tilde{\rho}_i(k)) \le \frac{\bar{v}_i(j^*, \tilde{x}) + 1}{\bar{v}_i(s, \tilde{x})} - 1 \quad \forall s \ne j^*$



⇒ $\frac{d-b+1}{b} < \gamma < \frac{a-c+1}{c}$ Positive probability of convergence to (A,A)
Zero probability of convergence to (B,B)

(case2) $a = 2 + 2\delta, b = 1 - \delta, c = 2, d = 1$

⇒ (A,A) risk and payoff dominant



$\frac{2\delta+1}{2} < \gamma < \frac{\delta+1}{1-\delta}$ ⇒ Positive probability of convergence to (B,B)
Zero probability of convergence to (A,A)

Example 2

	2.A	2.B
1.A	a_1, a_2	b,c
1.B	c,b	d_1, d_2

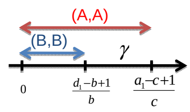
Asymmetric coordination game

$a_1 > a_2 > 0, d_1 = a_2, d_2 = a_1, b = c > 0, a_1 > c, d_1 > b$

⇒ (A,A) Desirable for agent 1 $(a_1 < a_2)$
(B,B) Desirable for agent 2 $(a_1 > a_2)$

Only agent 1

$\frac{d_1 - b + 1}{b} < \gamma < \frac{a_1 - c + 1}{c}$ ⇒



Positive probability of convergence to (A,A)
Zero probability of convergence to (B,B)

⇒ The agent that applies RADR destabilizes the less Desirable equilibrium in favor of the desirable one

Payoff dependent γ

	2.A	2.B
1.A	a,a	b,c
1.B	c,b	d,d

Symmetric coordination game

(A,A)
(B,B)

$0 \quad \frac{a-c+1}{c} \quad \frac{d-b+1}{b}$

Risk dominant but not Pay off dominant can't destabilized with constant γ

$$a>c>0, d>b>0, (a-c)<(d-b), a>d$$

(A,A) payoff dominant
(B,B) risk dominant
can't be destabilized

$$\gamma_i(\rho_i)=\frac{\gamma_0}{\rho_i^k} \quad \text{RADR parameter}$$

$$\kappa > \frac{\log\left(\frac{a-c+1}{d-b+1}\frac{b}{c}\right)}{\log\left(\frac{d}{a}\right)} \qquad d^{\kappa}\frac{d-b+1}{b} < \gamma_0 < a^{\kappa}\frac{a-c+1}{c}$$

Positive probability of convergence to (A,A)

Zero probability of convergence to (B,B)

The risk dominant equilibrium is no longer stable in favor of the payoff dominant equilibrium

Outline

- Background
- Reinforcement learning
- Asymptotic stability analysis
- Dynamic Reinforcement
- Asymptotic stability of RADR
- Simulation Result

simulation

3 agent

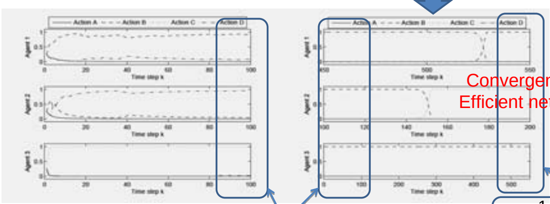
Actions of agent

$$A_1=\{A,B,C,D\}$$
$$A_1=\{\{1\},\{2\},\{3\},\{2,3\}\}$$
$$A_2=\{\{1\},\{2\},\{3\},\{1,3\}\}$$
$$A_3=\{\{1\},\{2\},\{3\},\{2,1\}\}$$

$$\gamma_i(\rho_i)=\frac{\gamma_0}{\rho_i^k} \quad \text{RADR parameter}$$

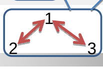
$$\kappa_0=\frac{1}{2} \qquad \gamma_i \in (2/3, 3/2)$$

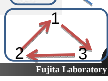
Non-efficient Nash network is unstable



Convergence to Efficient network!

$$\gamma_i(\rho_i)=0$$





Outline

Designing Games to Handle Coupled Constraints

Na Li and Jason R.Marden

- Background
- Noncooperative Game and State Based Game
- State Based Game
- Gradient play

Background

Coupled constraint game → challenging

In many systems , the emergent global behavior must satisfy a coupled constraint on the agents' actins

(ex) Consensus $\sum_{i\in N}\omega_i v_i(t)=\sum_{i\in N}\omega_i v_i(0)$ Formation control
Power control

problem
Whether can utility design be effective for dealing with these coupled constraint?

Do all pure Nash equilibrium satisfy performance criteria that include coupled constraint?
→ Noncooperative game is not suitable

Adding an additional state (State based game)
→ Can satisfy performance criteria

Noncooperative Game and State Based Game

Usual consensus algorithm Sensitive for environment uncertain

$$v_i(t)=v_i(t-1)+\frac{\varepsilon}{\omega_i}\sum_{j\in N_i}(v_j(t-1)-v_i(t-1)) \Rightarrow \sum_{i\in N}\omega_i v_i(t)=\sum_{i\in N}\omega_i v_i(0)$$

Noncooperative game

Cost function

$$J_i(v_i,v_{-i})=\sum_{j\in N_i}\left\|v_i-v_j\right\|_2^2$$

a^* Pure Nash equilibrium

$$J_i(a_i^*,a_{-i}^*)=\min_{a_i\in A_i} J_i(a_i,a_{-i}^*)$$

problem
- Any $v_i=v'$ for all i is Nash equilibrium
- under the constraint like $\sum_{i\in N}\omega_i v_i(t)=\sum_{i\in N}\omega_i v_i(0)$ any feasible action sets is Nash equilibrium

Convergence to the weighted average is not possible

State Based Game

Noncooperative Game and State Based Game

State Based Game $x(t) \in X$:State

Cost function

$J_i : X \times A \to \Re$ → Select an action a_i by J_i

(ex) $a_i(t) \in \arg\min_{a_i \in A_i} J_i(x(t), a_i, a_{-i}(t-1))$

State transition function

$f : X \times A \to X$ $\quad x(t+1) = f(x(t), a(t))$

State Based Nash equilibrium

$[x^*, a^*]$ is State Based Nash equilibrium ↔ for every state $x \in \bar{X}(x^*, a^* : f)$

$J_i(x, a_i^*, a_{-i}^*) = \min_{a_i \in A_i} J_i(x, a_i, a_{-i}^*)$

Reachable State

$\bar{X}(x^0, a^0 : f) = \{x^0, x^1, x^2 \cdots\}$, $x^{k+1} = f(x^k, a^0)$

→ $[x^*, a^*]$ is State Based Nash equilibrium → $[\bar{x}, a^*]$ is State Based Nash equilibrium

State Based Game

Coupled constraint on player's value

⇓

Coupled constraint on player's value and bias

State $x = (v, b)$ Action $a_i = (\hat{v}_i, \hat{b}_i)$

Change bias using neighbor bias information

$$f_i^v(x, a) = v_i + \hat{v}_i$$

$$f_i^b(x, a) = b_i - \hat{v}_i + \frac{1}{\omega_i} \sum_{j \in N_i} (\hat{b}_j^i - \hat{b}_i^j)$$

→ $\sum_{i \in N} \omega_i v_i^k + \sum_{i \in N} \omega_i b_i^k = \sum_{i \in N} \omega_i v_i^0$

Can think constraint !

State Based Game

Goal is to make cost function which induces Nash equilibrium

$[x^*, a^*]$ → $x^* = (v^*, b^*)$

$v_i^* = \sum_{i \in N} \omega_i v_i(0) \qquad b_i^* = 0$

[theorem]

Cost function is $J_i(x, a) = J_i^v(x, a) + J_i^b(x, a)$

$$J_i^v(x, a) = \sum_{j \in N_i} (\omega_i + \omega_j) \left\| v_i + \hat{v}_i - (v_j + \hat{v}_j) \right\|_2^2$$

$$J_i^b(x, a) = \sum_{j \in N_i} \omega_j \left\| b_j - \hat{v}_j + \frac{1}{\omega_j} \sum_{k \in N_j / \{j\}} (\hat{b}_k^j - \hat{b}_j^k) \right\|_2^2$$

$[x, a]$ is State Based Nash equilibrium → $v_i = \sum_{i \in N} \omega_i v_i(0)$, $b_i = 0$

$\hat{v}_i = 0$, $\sum_{j \in N_i} (\hat{b}_i^j - \hat{b}_j^i) = 0$

Gradient Play

State Based Nash equilibrium → Desirable global behavior

Make learning algorithm which ensure that the agents to reach the state based Nash Equilibrium

Gradient Play

$$\hat{v}_i(t) = 2\varepsilon_i^v \left[\omega_i b_i(t) - \sum_{j \in N_i} (\omega_i + \omega_j)(v_i(t) - v_j(t)) \right]$$

$$\hat{b}_i^j(t) = 2\varepsilon_{ij}^b \left[b_i(t) - b_j(t) \right]$$

From Cost function, Potential Function is $\phi(x, a) = \phi^v(x, a) + \phi^b(x, a)$

$$\phi^v(x, a) = \frac{1}{2} \sum_{i \in N} \sum_{j \in N_i} (\omega_i + \omega_j) \left\| v_i + \hat{v}_i - (v_j + \hat{v}_j) \right\|_2^2$$

$$\phi^b(x, a) = \sum_{i \in N} \omega_i \left\| b_i - \hat{v}_i + \frac{1}{\omega_i} \sum_{k \in N_i / \{i\}} (\hat{b}_k^i - \hat{b}_i^k) \right\|_2^2$$

Gradient Play

[Theorem] Under Gradient play

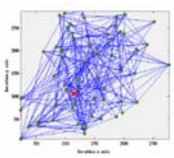
Stepsize ε_i^v , ε_{ij}^b is smaller than $C/2$ → $[x(t), a(t)]$ convergence exponentially to $[(v^*, 0), 0]$

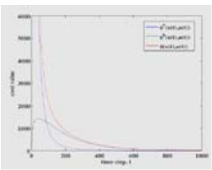
$C = \left\| \frac{\partial^2 \phi(x, a)}{\partial^2 a} \right\|_2$ $\qquad v_i^* = \sum_{i \in N} \omega_i v_i(0)$

Proof is using

$$J_i(x, a_i', a_{-i}) - J_i(x, a_i, a_{-i}) = \phi(x, a_i', a_{-i}) - \phi(x, a_i, a_{-i})$$

Simulation result





Main role

first ϕ^v

gradually ϕ^b

Outline

Game Theoretic learning algorithm for a spatial Coverage Problem

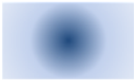
Ketan Savla , Emilio Frazzoli

- Environment and Purpose
- Potential Game

Environment and Purpose

Environment

workspace Q



φ : density function $\int_Q \varphi(q)dq = 1$

$P(t)$: service requests generated in $[0, t)$

Spatio-temporal Poisson point process

Process generating service requests is

$$\Pr[\ card\ ((P(s+t) - P(s)) \cap S) = k] = \frac{\exp(-\lambda t \varphi(s))(\lambda t \varphi(s))^k}{k!}$$

$\varphi(S) = \int_S \varphi(q)dq$

λ : temporal intensity

Expected number of targets generated in $\Delta t, S$

$$E[card\ ((P(t+\Delta t) - P(t)) \cap S)] = \lambda \Delta t \varphi(S)$$

Service request is fulfilled when one of agent move to target position

Agent model $\dot{p}_i(t) = u_i(t)$, $\|u_i(t)\| \le 1$

Environment and Purpose

Assumption

$\lambda \to 0^+$

- Agents stay most of time idle
- Go target position and come back reference point before next target emerge

Expected system time

$$E[T_j] = \int_Q \min \|p_i(t_j) - q\| \varphi(q)dq$$

System time under the policy π

$$\bar{T}_\pi = \lim_{j \to \infty} E[T_j] \qquad \pi = \{\pi_1, \ldots, \pi_m\}$$

Expected time a service request must wait

purpose

Want to make policy which achieve optimal performance

$\bar{T}_{opt} = \min_\pi \bar{T}_\pi$ with game theory

Usually use Voronoi centroidal low

$$p_i = \arg\min_{s \in \Re^n} \int_{V_i(p)} \|s - q\| \varphi(q)dq \qquad \bar{T}_{opt} = \min_{p \in Q^m} \sum_{i=1}^m \int_{V_i(p)} \|p_i - q\| \varphi(q)dq$$

Potential Game

Assumption

- Location of a resource is broadcast to all agent
- No explicit communication between agent

$r_i(q, \pi)$: agent i's reward

$$r_i(q,\pi) = \max\{0, \min_{j \ne i} \|\pi_j - q\| - \|\pi_i - q\|\}$$

positive only i is most close to q

utility

$$u_i(\pi_i, \pi_{-i}) = E_q[r_i(q,\pi)]$$

$$= \int_Q \max\{0, \min_{j \ne i} \|\pi_j - q\| - \|\pi_i - q\|\} \varphi(q)dq$$

$$= \int_{V_i(\pi)} (\min_{j \ne i} \|\pi_j - q\| - \|\pi_i - q\|) \varphi(q)dq$$ (from Voronoi cell)

Potential Game

utility $u_i(\pi_i, \pi_{-i}) = E_q[r_i(q,\pi)]$

$$= \sum_{j=1, j \ne i}^m \int_{V_j(\pi_{-i})} \|\pi_j - q\| \varphi(q)dq - \sum_{j=1}^m \int_{V_j(\pi)} \|\pi_j - q\| \varphi(q)dq$$

Potential function

$$\psi(\pi) = -\sum_{i=1}^m \int_{V_i(\pi)} \|\pi_i - q\| \varphi(q)dq$$

Equal to global objective function

$$u_g(\pi) = -\bar{T}_\pi = -\sum_{i=1}^m \int_{V_i(\pi)} \|\pi_i - q\| \varphi(q)dq$$

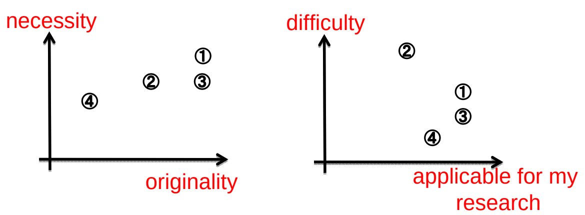
$\bar{T}_{opt} = \min_\pi \bar{T}_\pi$

From Voronoi Centoroidal low

Equilibrium strategy $\hat{p}^*$ $\left(\hat{p}_i^* = \arg\min_{p_i \in Q} \int_{V_i(\hat{p}^*)} \|p_i - q\| \varphi(q)dq \right)$ is an efficient pure Nash equilibrium for this game

Use learning algorithm to go to an efficient pure Nash equilibrium

summary



① Near-Optimal Power Control in Wireless Networks

② Distributed Dynamic Reinforcement of Efficient Outcome in Multiagent Coordination and Network Formation

③ Designing Games to Handle Coupled Constraints

④ Game Theoretic learning algorithm for a spatial Coverage Problem