システム/制御/情報, Vol. 58, No. 11, pp. 462–467, 2014

解　説

報酬駆動型システムにおける報酬の設計と報酬による最適化

奥原　浩之*

1. はじめに

個体の最適な応答の結果として，集団にとって効率的な選択が実現されることは望ましい．しかしながら，環境によっては，個体の合理的な意思決定で集団にとって望ましい状況が実現されるとは限らない．そこで，環境を人為的に設計することで，設計者が望ましいと考える集団として効率的な選択を，個体の最適な応答により自律分散的に実現する仕組みが求められる．

このような背景から，特定の戦略の選択を促進・抑制するメカニズムを，ペイオフの観点から分析することが求められる [1]．遺伝子，神経回路，進化ゲームに関する力学 [1–4] は，意思決定の主体が戦略を選び行動した結果，報酬を原動力として駆動していると見なすことができ，ペイオフにもとづく効用関数の変化を報酬として入力している報酬駆動型システムと捉えることができる．

本解説では，その中でも，まず効用関数の最大化が適者生存則など [5,6] に関連していることを紹介する．つぎに，与えられたペイオフのもとで均衡が漸近安定とならない場合に，ペイオフを変更することで漸近安定化 [7,8] する報酬の設計法を概説する．そして，逐次得られる限られたペイオフのもとで，最もペイオフが高い戦略を探索する報酬による最適化 [9,10] について概観する．

2. メカニズムデザインとペイオフ

2.1 非協力ゲームとペイオフ双行列

本解説で想定しているメカニズムの設計について，二人非協力有限ゲームにおける標準型ゲームを例にして簡単に述べる．

意思決定の主体であるプレイヤーの集合が $P=\{A,B\}$，プレイヤー $i(\in P)$ の戦略の集合を $S_i=\{s_i^1,s_i^2\}$ とし，選択した行動を $s_i(\in S_i)$ とする．プレイヤー i が戦略 π を選択することを s_i^π で表す．プレイヤー A,B がゲームをプレーすることで得られる利得であるペイオフ双行列 $\boldsymbol{G}=(\boldsymbol{V}_{AB},\boldsymbol{V}_{BA})$ を第 1 表に示す．

ここで，

$$\boldsymbol{V}_{AB}=\begin{pmatrix} v_{AB}^{11} & v_{AB}^{12} \\ v_{AB}^{21} & v_{AB}^{22} \end{pmatrix},\quad \boldsymbol{V}_{BA}=\begin{pmatrix} v_{BA}^{11} & v_{BA}^{21} \\ v_{BA}^{12} & v_{BA}^{22} \end{pmatrix} \tag{1}$$

第 1 表　ペイオフ双行列の例

	s_B^1	s_B^2
s_A^1	$(v_{AB}^{11},v_{BA}^{11})$	$(v_{AB}^{12},v_{BA}^{21})$
s_A^2	$(v_{AB}^{21},v_{BA}^{12})$	$(v_{AB}^{22},v_{BA}^{22})$

である．

ペイオフの値によって，囚人のジレンマ，鷹鳩ゲームなどの異なる状況を表現することになる．

プレイヤーが N 人となる場合を考える．このとき，プレイヤーの選択した行動 $\boldsymbol{s}^*=[s_1^*s_2^*s_3^*\cdots s_N^*]^{\mathrm{T}}\in\Re^{N\times 1}$ が互いに最適な応答

$$s_i^*\in\arg\max_{s_i\in S_i} u_i\left(s_i,\boldsymbol{s}_{\bar{i}}^*\right)\qquad(\forall i)\tag{2}$$

となるとき，Nash 均衡であるという．ここで，$\boldsymbol{s}_{\bar{i}}^*$ は $\boldsymbol{s}^*$ から s_i^* を除いた $\left[s_1^*\cdots s_{i-1}^*s_{i+1}^*\cdots s_N^*\right]^{\mathrm{T}}\in\Re^{(N-1)\times 1}$ であり，効用関数 u_i は戦略空間 $S\left(=\prod_{i\in N}S_i\right)$ からペイオフへの写像を与える．

たとえば，第 1 表の場合，よく知られているように，$\boldsymbol{V}_{AB}=\boldsymbol{V}_{BA}^{\mathrm{T}}$ で，$v_{AB}^{21}>v_{AB}^{11}>v_{AB}^{22}>v_{AB}^{12}$ かつ $2v_{AB}^{11}>v_{AB}^{21}+v_{AB}^{12}$ であるなら，Nash 均衡は $\left(s_A^2,s_B^2\right)$ となる．このことは，個人の最適な応答が集団にとって効率的な選択 $\left(s_A^1,s_B^1\right)$ とならない囚人のジレンマの状況を表している．

2.2 混合戦略と複製方程式

プレイヤー i が戦略 π を頻度 x_i^π で選択する混合戦略の場合を考え，$\boldsymbol{x}_i=\left[x_i^1x_i^2x_i^3\cdots x_i^{n_i}\right]^{\mathrm{T}}\in\Re^{n_i\times 1}$ とする．ここで，n_i はプレイヤー i の戦略の数であり $\sum_{\pi=1}^{n_i}x_i^\pi=1$ である．このとき，純戦略 s_i^π の選択は $x_i^\pi=1$ であるので，$\boldsymbol{x}_i$ において $x_i^\pi=1$ かつ $x_i^{\pi'}=0$ $(\pi'\neq\pi)$ とした $\boldsymbol{e}_i^\pi=[0\cdots010\cdots0]^{\mathrm{T}}\in\Re^{N\times 1}$ と表せる．

プレイヤー i とそれ以外のプレイヤー $j(\neq i)$ が，ゲームをプレーするときの効用関数 u_i が期待利得で与えられるとすると，

$$u_i(\boldsymbol{x}_i,\,\boldsymbol{x}_{\bar{i}})=\sum_{j=1,j\neq i}^{N}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{V}_{ij}\boldsymbol{x}_j\tag{3}$$

* 大阪大学大学院　情報科学研究科

Key Words: reward-driven system, design, optimization.

となる．ここで，$\boldsymbol{x}_{\bar{i}}=\left[\boldsymbol{x}_1^{\mathrm{T}}\cdots\boldsymbol{x}_{i-1}^{\mathrm{T}}\boldsymbol{x}_{i+1}^{\mathrm{T}}\cdots\boldsymbol{x}_N^{\mathrm{T}}\right]^{\mathrm{T}}\in\Re^{(M-n_i)\times 1}$，$\boldsymbol{V}_{ij}\in\Re^{n_i\times n_j}$ であり，$M=\sum_{i=1}^{N}n_i$ である．

このとき，任意の有限標準形ゲームには混合戦略の範疇で

$$u_i\left(\boldsymbol{x}_i^*,\,\boldsymbol{x}_{\bar{i}}^*\right)\geq u_i\left(\boldsymbol{e}_i^{\pi},\,\boldsymbol{x}_{\bar{i}}^*\right)\qquad(\forall i,\,\forall\pi)\qquad(4)$$

を必要十分条件とする Nash 均衡 $\boldsymbol{x}^*=\left[\boldsymbol{x}_1^{*\mathrm{T}}\boldsymbol{x}_2^{*\mathrm{T}}\boldsymbol{x}_3^{*\mathrm{T}}\cdots\boldsymbol{x}_N^{*\mathrm{T}}\right]^{\mathrm{T}}\in\Re^{M\times 1}$ が存在するが，必ずしも唯一とは限らない．

いま，戦略 π の選択による利得が期待利得 $u_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})$ より高いようであれば，プレイヤー i は戦略の頻度 x_i^{π} を増加させ，低いようであれば減少させると考える．あるいは，期待利得に比べて利得が高い戦略を選択した集団の割合が複製により増加して，低い戦略を選択した集団の割合が淘汰により減少すると考える．

ゲームの動学的な状況を表す進化ゲームでは，複製方程式はリプリケータダイナミクスとして，

$$\dot{x}_i^{\pi}=x_i^{\pi}\left(u_i(\boldsymbol{e}_i^{\pi},\boldsymbol{x}_{\bar{i}})-u_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})\right)\qquad(\forall i,\,\forall\pi)\qquad(5)$$

とできる [1]．

たとえば，第1表の場合，プレイヤー A の混合戦略は $\left(x_A^1,x_A^2\right)=(0,1)$，プレイヤー B の混合戦略は $\left(x_B^1,x_B^2\right)=(0,1)$ となり Nash 均衡である純戦略の組 $\left(s_A^2,s_B^2\right)$ に収束することが示される．

3. 報酬駆動型システムの概要

3.1 遺伝子・神経回路網のモデル

ここで，ある種の遺伝子，結合振動子，神経回路網のモデルと勾配系やゲームのモデルにおけるアナロジーについて述べておく．

まず，遺伝子 A^{π} の相対頻度 x^{π} が従う連続時間の淘汰・突然変異のモデルの一つに

$$\dot{x}^{\pi}=x^{\pi}\left(u^{\pi}(\boldsymbol{e}^{\pi},\boldsymbol{x})-u(\boldsymbol{x},\boldsymbol{x})\right)+\sum_{\pi'=1}^{n}\left(\epsilon^{\pi\pi'}x^{\pi'}-\epsilon^{\pi'\pi}x^{\pi}\right)\qquad(\forall\pi)\qquad(6)$$

がある．ここで，$\boldsymbol{x}=\left[x^1x^2x^3\cdots x^n\right]^{\mathrm{T}}\in\Re^{n\times 1}$，$u(\boldsymbol{x},\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x}$ は集団平均適応度，行列 $\boldsymbol{M}\in\Re^{n\times n}$ の $\pi\pi'$ 要素はマルサス的適応度である．$\boldsymbol{e}^{\pi}=[0\cdots010\cdots0]^{\mathrm{T}}\in\Re^{n\times 1}$ は $\boldsymbol{x}$ において $x^{\pi}=1$ かつ $x^{\pi'}=0\,(\pi'\neq\pi)$ としたものである．$\epsilon^{\pi\pi'}$ は突然変異率であり，$\epsilon^{\pi\pi'}\geq 0$，$\sum_{\pi=1}^{n}\epsilon^{\pi\pi'}=1\ (\forall\pi')$ を満たしている．右辺第一項が淘汰，第二項が突然変異を表す [1]．

つぎに，非線形振動する素子を結合した振動子が従う連続時間の結合振動子のモデルの一つに

$$\dot{\boldsymbol{x}}_i=\boldsymbol{f}(\boldsymbol{x}_i)+D\sum_{j=1}^{N}(\boldsymbol{x}_j-\boldsymbol{x}_i)\qquad(\forall i)\qquad(7)$$

がある．右辺第一項は非線形項，第二項は拡散係数 D をもつ線形項である．微分方程式 $\dot{\boldsymbol{x}}_i=\boldsymbol{f}(\boldsymbol{x}_i)$ が写像関数 $\boldsymbol{x}_i^{(k+1)}=\boldsymbol{g}\left(\boldsymbol{x}_i^{(k)}\right)$ をもつと仮定する．このとき，サンプリング間隔を T $(t=kT)$，$\hat{D}=\dfrac{1-\exp(-NDT)}{N}$ として，離散時間のモデル

$$\boldsymbol{x}_i^{(k+1)}=\boldsymbol{g}\left(\boldsymbol{x}_i^{(k)}\right)+\hat{D}\sum_{j=1}^{N}\left(\boldsymbol{g}\left(\boldsymbol{x}_j^{(k)}\right)-\boldsymbol{g}\left(\boldsymbol{x}_i^{(k)}\right)\right)\qquad(8)$$

が得られる [2]．

離散時間の結合振動子のモデルの両辺に非線形変換を施すと

$$\boldsymbol{g}\left(\boldsymbol{x}_i^{(k+1)}\right)=\boldsymbol{g}\left((1-\epsilon)\boldsymbol{g}(\boldsymbol{x}_i^{(k)})+\frac{\epsilon}{N}\sum_{j=1}^{N}\boldsymbol{g}(\boldsymbol{x}_j^{(k)})\right)\qquad(9)$$

となる．ここで，$n_i=n\,(\forall i)$ であり，$\boldsymbol{V}_{ii}=\left(1-\epsilon+\dfrac{\epsilon}{N}\right)I_n$，$\boldsymbol{V}_{ij}=\dfrac{\epsilon}{N}I_n\ (j\neq i)$ として，変数変換 $\boldsymbol{y}_i^{(k)}=\boldsymbol{g}\left(\boldsymbol{x}_i^{(k)}\right)$ を施すと，神経回路網のモデル

$$\begin{aligned}\boldsymbol{y}_i^{(k+1)}&=\boldsymbol{g}\left(\sum_{j=1}^{N}\boldsymbol{V}_{ij}\boldsymbol{y}_j^{(k)}\right)\\&=\boldsymbol{g}\left(\frac{\partial\tilde{u}_i\left(\boldsymbol{y}_i^{(k)},\,\boldsymbol{y}_{\bar{i}}^{(k)}\right)}{\partial y_i^{\pi(k)}}\right)\end{aligned}\qquad(10)$$

の一例とも見なせる [3]．ただし，

$$\tilde{u}_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})=u_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})+\frac{1}{2}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{V}_{ii}\boldsymbol{x}_i\qquad(11)$$

であり，$\boldsymbol{I}_n\in\Re^{n\times n}$ は単位行列である．

3.2 勾配系・N 人進化ゲームのモデル

ところで，プレイヤー i の状態が n_i 個の変数をもつベクトル $\boldsymbol{x}_i$ で表される N 人の勾配系モデルの一つに

$$\dot{x}_i^{\pi}=g(x_i^{\pi})\phi_i^{\pi}\qquad(\forall i,\,\forall\pi)\qquad(12)$$

がある．ここで，

$$\begin{aligned}\phi_i^{\pi}&=\sum_{j=1}^{N}\sum_{\pi'=1}^{n_j}v_{ij}^{\pi\pi'}x_j^{\pi'}+\theta_i^{\pi}\\&=\frac{\partial\hat{u}_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})}{\partial x_i^{\pi}}\end{aligned}\qquad(13)$$

であり，

$$\hat{u}_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})=\tilde{u}_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})+\boldsymbol{\theta}_i^{\mathrm{T}}\boldsymbol{x}_i\qquad(14)$$

である．ただし，$\boldsymbol{\theta}_i=\left[\theta_i^1\theta_i^2\theta_i^3\cdots\theta_i^{n_i}\right]^{\mathrm{T}}\in\Re^{n_i\times 1}$ であり，$g(x_i^{\pi})>0$ なら値 x_i^{π} の更新は効用関数 $\hat{u}_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})$ の値を増加させる．もちろん，均衡が漸近安定であるかどうかは，定常点近傍でのヤコビ行列の固有値により安定判別がなされる [4]．

【定理 1】 効用最大化と適者生存との関係

全プレイヤーの効用関数 $\hat{u}_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})$ の和の最大化は，任意のベクトル $\boldsymbol{y}_k \in \Re^{m_k\times 1}$，$(k=1,2,3,\cdots,S)$ に対する供給関数 $\eta(\boldsymbol{y}_k)$ と，パラメータ φ_i^π をもつ利用関数 $\xi(\boldsymbol{y}_k,\varphi_i^\pi)$ が与えられるとき，競争係数を

$$\lambda_{ij}^{\pi\pi'}=\lambda_{ji}^{\pi'\pi}=\frac{v_{ij}^{\pi\pi'}+v_{ji}^{\pi'\pi}}{2}=-\sum_{k=1}^{S}\xi(\boldsymbol{y}_k,\varphi_i^\pi)\xi(\boldsymbol{y}_k,\varphi_j^{\pi'}) \tag{15}$$

とし，適応係数を

$$\theta_i^\pi=\sum_{k=1}^{S}\eta(\boldsymbol{y}_k)\xi(\boldsymbol{y}_k,\varphi_i^\pi) \tag{16}$$

として，供給関数の利用関数による関数近似

$$\eta(\boldsymbol{y}_k)=\sum_{i=1}^{N}\sum_{\pi=1}^{n_i}x_i^\pi\xi(\boldsymbol{y}_k,\varphi_i^\pi) \tag{17}$$

と等価である．

（証明）

$$\begin{aligned}&-\frac{1}{2}\sum_{k=1}^{S}\left(\eta(\boldsymbol{y}_k)-\sum_{i=1}^{N}\sum_{\pi=1}^{n_i}x_i^\pi\xi(\boldsymbol{y}_k,\varphi_i^\pi)\right)^2\\&=-\frac{1}{2}\sum_{k=1}^{S}\eta^2(\boldsymbol{y}_k)+\sum_{i=1}^{N}\boldsymbol{\theta}_i^{\mathrm{T}}\boldsymbol{x}_i-\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\Lambda}_{ij}\boldsymbol{x}_j\\&=\sum_{i=1}^{N}\hat{u}_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}})-\frac{1}{2}\sum_{k=1}^{S}\eta^2(\boldsymbol{y}_k)\end{aligned} \tag{18}$$

ここで，$\boldsymbol{\Lambda}_{ij}\in\Re^{n_i\times n_j}$ は要素に $\lambda_{ij}^{\pi\pi'}$ をもつ．

関数近似の観点から，ベクトル x_i^π やパラメータ φ_i^π に対する適者生存型の更新則 [5] や，プレイヤー数や要素数が変化する動的な環境でも適用可能な自由エネルギー型の更新則 [6] が導かれている． □

さらに，勾配系モデルを N 人進化ゲームとするために，混合戦略の条件 $\sum_{\pi=1}^{n_i}x_i^\pi=1$ $(\forall i)$ を満たすダイナミクスを導出すると

$$\begin{aligned}\dot{x}_i^\pi&=x_i^\pi(\hat{u}_i(\boldsymbol{e}_i^\pi,\boldsymbol{x}_{\bar{i}})-\hat{u}_i(\boldsymbol{x}_i,\boldsymbol{x}_{\bar{i}}))\\&=x_i^\pi\left(\phi_i^\pi-\sum_{\pi=1}^{n_i}x_i^\pi\phi_i^\pi\right)\\&=\sum_{\pi'=1}^{n_i}g^{\pi'}(x_i^\pi)\phi_i^{\pi'}\qquad(\forall i,\ \forall\pi)\end{aligned} \tag{19}$$

となる．ここで，$\boldsymbol{\phi}_i=\left[\phi_i^1\phi_i^2\phi_i^3\cdots\phi_i^{n_i}\right]^{\mathrm{T}}\in\Re^{n_i\times 1}$ であり，

$$g^{\pi'}(x_i^\pi)=\begin{cases}x_i^\pi(1-x_i^\pi) & (\pi'=\pi)\\ -x_i^\pi x_i^{\pi'} & (\pi'\neq\pi)\end{cases} \tag{20}$$

である．このとき，初期値が混合戦略の条件を満たしていれば，$\sum_{\pi=1}^{n_i}\dot{x}_i^\pi=0$ $(\forall i)$ となる．

以上に述べた遺伝子，神経回路網や進化ゲームのモデルには，複製方程式から派生したモデルも含まれており，これらは共通して，戦略を選び行動した結果，ペイオフにもとづく効用関数の変化を入力として動作していると考えることもできる．本解説では，それらを報酬駆動型システムとして考えている．

4. 報酬の設計

4.1 非線形大域結合システム

ここで，遺伝子，神経回路網や進化ゲームの連続時間のモデルである式 (6)，(7)，(19) をまとめて取り扱うことができる非線形大域結合システムによるモデル化と漸近安定化に向けた制御，報酬の設計について議論する．

システムは N 個の局所システムから構成され，i 番目の局所システムの状態は n_i 個の変数をもつベクトル $\boldsymbol{x}_i$ で表される．そのとき，局所システムの状態方程式を

$$\dot{\boldsymbol{x}}_i=\sum_{j=1}^{N}\boldsymbol{G}_{ij}(\boldsymbol{x})\left\{\boldsymbol{f}(\boldsymbol{x}_j)+\sum_{k=1}^{N}c_{jk}a_{jk}(\boldsymbol{\Gamma}_{jk}\boldsymbol{x}_k-\boldsymbol{\Gamma}_{jj}\boldsymbol{x}_j)\right\}\quad(\forall i) \tag{21}$$

とする．ここで，$\boldsymbol{G}_{ij}(\boldsymbol{x})\in\Re^{n_i\times n_j}$，$\boldsymbol{f}(\boldsymbol{x}_j)$ は C^1 級で，$\Re^{n_j}\to\Re^{n_j}$ の非線形写像であり，$\boldsymbol{f}(\boldsymbol{0}_{n_j})=\boldsymbol{0}_{n_j}$ である．$\boldsymbol{x}=[\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_3^{\mathrm{T}}\cdots\boldsymbol{x}_N^{\mathrm{T}}]^{\mathrm{T}}\in\Re^{M\times 1}$ であり，$\boldsymbol{0}_{n_j}\in\Re^{n_j\times 1}$ は零ベクトルである．また，$c_{jk}\in\Re^{N\times N}$ は局所システム間の結合強度，$a_{jk}\in\Re^{N\times N}$ は局所システム間の結合情報であり，$a_{jk}=1$ は結合あり，$a_{jk}=0$ は結合なしを意味する．$\boldsymbol{\Gamma}_{jk}\in\Re^{n_j\times n_k}$ は局所システム変数間の相互作用である．

いま，$\tilde{a}_{jk}=a_{jk}$ $(k\neq j)$，$\tilde{a}_{jj}=-\sum_{k=1}^{N}a_{jk}$，$\tilde{c}_{jk}=c_{jk}$ $(k\neq j)$，$\tilde{c}_{jj}=\frac{1}{\tilde{a}_{jj}}\sum_{k=1}^{N}c_{jk}a_{jk}$ とすると

$$\dot{\boldsymbol{x}}_i=\sum_{j=1}^{N}\boldsymbol{G}_{ij}(\boldsymbol{x})\tilde{\boldsymbol{f}}(\boldsymbol{x}_j)\qquad(\forall i) \tag{22}$$

となる．ここで，

$$\tilde{\boldsymbol{f}}(\boldsymbol{x}_j)=\boldsymbol{f}(\boldsymbol{x}_j)+\sum_{k=1}^{N}\tilde{c}_{jk}\tilde{a}_{jk}\boldsymbol{\Gamma}_{jk}\boldsymbol{x}_k \tag{23}$$

である．

システム全体の状態方程式は

$$\dot{\boldsymbol{x}}=\boldsymbol{G}(\boldsymbol{x})\{\boldsymbol{f}(\boldsymbol{x})+\boldsymbol{\Gamma}_\alpha\boldsymbol{x}\} \tag{24}$$

となる．ここで，$\boldsymbol{G}(\boldsymbol{x})\in\Re^{M\times M}$ は ij ブロックに $\boldsymbol{G}_{ij}(\boldsymbol{x})\in\Re^{n_i\times n_j}$，$\boldsymbol{\Gamma}_\alpha\in\Re^{M\times M}$ は ij ブロックに $\alpha_{ij}\boldsymbol{\Gamma}_{ij}\in\Re^{n_i\times n_j}$

をもち，$\alpha_{ij}=\tilde{c}_{ij}\tilde{a}_{ij}$ である．また，$\boldsymbol{f}(\boldsymbol{x})=\left[\boldsymbol{f}(\boldsymbol{x}_1)^{\mathrm{T}}\ \boldsymbol{f}(\boldsymbol{x}_2)^{\mathrm{T}}\ \boldsymbol{f}(\boldsymbol{x}_3)^{\mathrm{T}}\cdots\boldsymbol{f}(\boldsymbol{x}_N)^{\mathrm{T}}\right]^{\mathrm{T}}\in\Re^{M\times 1}$ である．

4.2 漸近安定化

非線形大域結合システムの漸近安定化は，非線形関数の線形緩和とフィードバック制御により行われる．まず，局所システムの変数 $\boldsymbol{x}_i$ の均衡 $\boldsymbol{x}_i^*$ に対して領域

$$\boldsymbol{D}=\bigcup_{i=1}^{N}\boldsymbol{D}_i\quad \boldsymbol{D}_i=\{\boldsymbol{x}_i:\|\boldsymbol{x}_i-\boldsymbol{x}_i^*\|<\alpha,\ \alpha>0\}\tag{25}$$

を考える．

各局所システムに正定対称行列 $\boldsymbol{P}_i\in\Re^{n_i\times n_i}$ をもつ Lyapunov 関数 $V_i(\boldsymbol{x}_i)=\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{x}_i$ の存在を仮定し，Lyapunov 関数 $V(\boldsymbol{x})=\sum_{i=1}^{N}V_i(\boldsymbol{x}_i):\ \boldsymbol{D}\subseteq\Re^M\to\Re_+^M$ は C^1 級で $V(\boldsymbol{x}^*)=0$, $\boldsymbol{x}^*\in\boldsymbol{D}$ とする．

いま，非線形関数の線形緩和のために

$$\left\{\mathcal{L}_{\boldsymbol{h}_\mu}V\right\}(\boldsymbol{x})<0\tag{26}$$

を満たす Passivity degree $\boldsymbol{\mu}=[\mu_1\mu_2\mu_3\cdots\mu_N]^{\mathrm{T}}\in\Re^{N\times 1}$ を考える [7]．ただし，Lie 微分作用素は

$$\begin{aligned}\left\{\mathcal{L}_{\boldsymbol{h}_\mu}V\right\}(\boldsymbol{x})&=\sum_{i=1}^{N}\frac{\partial V(\boldsymbol{x})}{\partial\boldsymbol{x}_i}\boldsymbol{h}_{\mu_i}(\boldsymbol{x}_i)\\&=\nabla_{\boldsymbol{x}}V^{\mathrm{T}}\boldsymbol{h}_\mu(\boldsymbol{x})\end{aligned}\tag{27}$$

であり，$\boldsymbol{x}\in\boldsymbol{D}$, $\boldsymbol{x}\neq\boldsymbol{x}^*$ に対して

$$\boldsymbol{h}_\mu(\boldsymbol{x})=\boldsymbol{G}(\boldsymbol{x})\{\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{\Gamma}_\mu(\boldsymbol{x}^*-\boldsymbol{x})\}\tag{28}$$

である．ここで，$\boldsymbol{\Gamma}_\mu\in\Re^{M\times M}$ は ii ブロックに $\mu_i\boldsymbol{\Gamma}_{ii}\in\Re^{n_i\times n_i}$ をもつブロック対角行列である．

Passivity degree $\boldsymbol{\mu}$ による線形緩和の後，限られた局所システム $(i=1,2,3,\cdots,m)$ に対して，フィードバックゲイン d_i (>0) を用いる Pinning 制御を行う [8]．

$$\tilde{\boldsymbol{f}}(\boldsymbol{x}_j)=\boldsymbol{f}(\boldsymbol{x}_j)+\sum_{k=1}^{N}\alpha_{jk}\boldsymbol{\Gamma}_{jk}\boldsymbol{x}_k+\boldsymbol{u}_i\quad(\forall i)\tag{29}$$

ただし，

$$\boldsymbol{u}_i=-(k_i-1)d_i\boldsymbol{\Gamma}_{jk}\boldsymbol{x}_i\qquad(i=1,2,3,\cdots,m)\tag{30}$$

$$\boldsymbol{u}_i=0\qquad(i=m+1,m+2,m+3,\cdots,N)\tag{31}$$

である．ここでは，局所システムの番号 i は状況に応じて振り直されていることに注意が必要である．

【補題 1】 非線形大域結合システムの漸近安定化

正定対称行列 $\boldsymbol{P}\in\Re^{M\times M}$ をもつ Lyapunov 関数 $V(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}$ と Passivity degree $\boldsymbol{\mu}$ の存在を仮定する．このとき，Pinning 制御を行い

$$\boldsymbol{M}(\boldsymbol{x})\prec 0\tag{32}$$

が成立するとき，非線形大域結合システムは漸近安定となる．ここで，

$$\begin{aligned}\boldsymbol{M}(\boldsymbol{x})=&\boldsymbol{P}\boldsymbol{G}(\boldsymbol{x})(\boldsymbol{\Gamma}_\alpha-\boldsymbol{\Gamma}_\mu-\boldsymbol{\Gamma}_K)\\&+(\boldsymbol{\Gamma}_\alpha-\boldsymbol{\Gamma}_\mu-\boldsymbol{\Gamma}_K)^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{x})\boldsymbol{P}\prec 0\end{aligned}\tag{33}$$

であり，$\boldsymbol{P}$ は ii ブロックに正定対称行列 $\boldsymbol{P}_i$ をもつブロック対角行列である．$\boldsymbol{\Gamma}_K\in\Re^{M\times M}$ は ii ブロックに $K_{ii}\boldsymbol{\Gamma}_{ii}\in\Re^{n_i\times n_i}$ をもつブロック対角行列であり，K_{ii} は $\boldsymbol{K}=\mathrm{diag}\{(k_1-1)d_1,(k_2-1)d_2,(k_3-1)d_3,\cdots,(k_m-1)d_m,0,0,0,\cdots,0\}\in\Re^{N\times N}$ の ii 要素である．

（証明）一般性を失うことなく $\boldsymbol{x}^*=\boldsymbol{0}$ を仮定する．

$$\begin{aligned}\dot{V}(\boldsymbol{x})&=\nabla_{\boldsymbol{x}}V^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{x})\{\boldsymbol{f}(\boldsymbol{x})+\boldsymbol{\Gamma}_\alpha\boldsymbol{x}-\boldsymbol{\Gamma}_K\boldsymbol{x}\}\\&<\nabla_{\boldsymbol{x}}V^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{x})(\boldsymbol{\Gamma}_\alpha-\boldsymbol{\Gamma}_\mu-\boldsymbol{\Gamma}_K)\boldsymbol{x}\\&=2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{G}(\boldsymbol{x})(\boldsymbol{\Gamma}_\alpha-\boldsymbol{\Gamma}_\mu-\boldsymbol{\Gamma}_K)\boldsymbol{x}\\&=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x})\boldsymbol{x}\end{aligned}\tag{34}$$

よって，$\boldsymbol{M}(\boldsymbol{x})\prec 0$ で漸近安定となる． □

4.3 N 人進化ゲームの漸近安定化

N 人進化ゲームのダイナミクスにおいて，$\boldsymbol{A}_{ij}=\boldsymbol{V}_{ij}-\frac{1}{N}\boldsymbol{\Theta}_{ij}$ とする．ただし，$\boldsymbol{\Theta}_{ij}=[\boldsymbol{\theta}_i\boldsymbol{\theta}_i\boldsymbol{\theta}_i\cdots\boldsymbol{\theta}_i]\in\Re^{n_i\times n_j}$ である．このとき，

$$\boldsymbol{\phi}_i=\sum_{j=1}^{N}\boldsymbol{A}_{ij}\boldsymbol{x}_j\qquad(\forall i)\tag{35}$$

となる．ここで，

$$\boldsymbol{G}_{ii}(\boldsymbol{x}_i)=\boldsymbol{I}(\boldsymbol{x}_i)\left\{\boldsymbol{I}_{n_i}-\left(\boldsymbol{1}_{n_i}\otimes\boldsymbol{x}_i^{\mathrm{T}}\right)\right\}\qquad(\forall i)\tag{36}$$

を考える．ただし，$\boldsymbol{I}(\boldsymbol{x}_i)=\mathrm{diag}\left\{x_i^1,x_i^2,x_i^3,\cdots,x_i^{n_i}\right\}\in\Re^{n_i\times n_i}$，$\boldsymbol{I}_{n_i}\in\Re^{n_i\times n_i}$ は単位行列，$\boldsymbol{1}_{n_i}=[1\ 1\ 1\ \cdots\ 1]^{\mathrm{T}}\in\Re^{n_i\times 1}$ であり，$\otimes$ は Kronecker 積である．

いま，N 人進化ゲームは，

$$\dot{\boldsymbol{x}}_i=\boldsymbol{G}(\boldsymbol{x})\boldsymbol{A}\boldsymbol{x}\tag{37}$$

で表せることがわかる．$\boldsymbol{A}\in\Re^{M\times M}$ は ij ブロックに $\boldsymbol{A}_{ij}\in\Re^{n_i\times n_j}$ をもつ．$\boldsymbol{G}(\boldsymbol{x})\in\Re^{M\times M}$ は ii ブロックに $\boldsymbol{G}_{ii}(\boldsymbol{x})\in\Re^{n_i\times n_i}$ をもつブロック対角行列である．

ここで，与えられたペイオフのもとで均衡が漸近安定とならない場合に，ペイオフを変更することで漸近安定化する報酬の設計を考える．

【定理 2】 報酬の設計

ペイオフ $\boldsymbol{A}$ による均衡 $\boldsymbol{x}^*$ が不安定であるとき，$\boldsymbol{x}^*$ を漸近安定化する報酬の設計法は $\boldsymbol{A}+\bar{\boldsymbol{K}}\bar{\boldsymbol{P}}^{-1}$ により与えられる．

（証明）$\boldsymbol{\Gamma}_\alpha=\boldsymbol{A}$, $\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{0}$ として，

$$\begin{aligned}\boldsymbol{M}(\boldsymbol{x}^*)=&\boldsymbol{P}\boldsymbol{G}(\boldsymbol{x}^*)(\boldsymbol{A}-\boldsymbol{\Gamma}_K)\\&+(\boldsymbol{A}-\boldsymbol{\Gamma}_K)^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{x}^*)\boldsymbol{P}\prec 0\end{aligned}\tag{38}$$

より，正定対称行列 $\boldsymbol{Q}$ を用いると，

$$\boldsymbol{G}(\boldsymbol{x}^*)\boldsymbol{A}\bar{\boldsymbol{P}}+\boldsymbol{G}(\boldsymbol{x}^*)\bar{\boldsymbol{K}} +\left(\boldsymbol{G}(\boldsymbol{x}^*)\boldsymbol{A}\bar{\boldsymbol{P}}+\boldsymbol{G}(\boldsymbol{x}^*)\bar{\boldsymbol{K}}\right)^{\mathrm{T}}+\boldsymbol{Q}\preceq 0 \quad (39)$$

となる．$\bar{\boldsymbol{P}}=\boldsymbol{P}^{-1}$，$\bar{\boldsymbol{K}}=\boldsymbol{\Gamma}_K\boldsymbol{P}^{-1}$ と $\boldsymbol{Q}$ に関する線形行列不等式より，$\bar{\boldsymbol{P}}$ と $\bar{\boldsymbol{K}}$ が得られれば，$\boldsymbol{A}-\boldsymbol{\Gamma}_K=\boldsymbol{A}+\bar{\boldsymbol{K}}\bar{\boldsymbol{P}}^{-1}$ で設計できる． □

上記の遺伝子，神経回路網や進化ゲームは，プレイヤー i が戦略 π を選び行動した結果，ペイオフにもとづく効用関数の変化を入力として動作していると考えるとき，プレイヤー j と戦略 π' に対するペイオフ $v_{ij}^{\pi\pi'}$ $(\forall j,\forall\pi')$ すべては既知とする報酬駆動型システムであるといえる．

5. 報酬による最適化

5.1 遺伝的アルゴリズム

複製方程式で表現できる遺伝子のモデルでは，淘汰と突然変異が考慮されていた．ここでは，さらに交叉と確率を考慮した遺伝子のモデルである以下のような簡易な遺伝的アルゴリズム（Simple Genetic Algorithm）について考える [9]．

[1] l ビットの0または1で遺伝子を表現する．N 個の個体 $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\cdots,\boldsymbol{x}_N\}$ を作成．$N\ll 2^l-1=n$ とする．

[2] 適合度関数 $f(\cdot)$，突然変異率 ϵ，α，交叉率 $\chi_{\boldsymbol{k}}$ の値と逆温度 β の初期値を与える．

[3] 突然変異する部分 $\boldsymbol{h}$ と交叉する部分 $\boldsymbol{k}$ を与える．

[4] 選択，突然変異，交叉を繰り返し世代交代する．

淘汰は，現世代において存在する遺伝子群 $\boldsymbol{p}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\cdots,\boldsymbol{x}_N\}$ のうち，遺伝子が選択される確率

$$S_\beta(\boldsymbol{x}'|\boldsymbol{p})=\frac{f(\boldsymbol{x}')^\beta}{\sum_{\boldsymbol{x}\in\boldsymbol{p}}f(\boldsymbol{x})^\beta} \qquad (\boldsymbol{x}'\in\boldsymbol{p}) \quad (40)$$

で表現される．ここで，$f(\boldsymbol{x})$ は遺伝子 $\boldsymbol{x}$ に対する適応度関数であり，$\beta>0$ は Boltzmann の逆温度である．

突然変異は，確率 $\mu_\beta=\epsilon\exp(-\alpha\beta)$ によりビットが反転する確率

$$m_\beta(\boldsymbol{x},\boldsymbol{h})=\mu_\beta^{\mathbf{1}^{\mathrm{T}}\boldsymbol{h}}(1-\mu_\beta)^{l-\mathbf{1}^{\mathrm{T}}\boldsymbol{h}} \quad (41)$$

で表現される．ここで，$\mathbf{1}\in\Re^{l\times 1}$ は要素すべてが1のベクトルであり，$\boldsymbol{h}\in\Re^{l\times 1}$ は突然変異で反転する部分のビットは1，それ以外は0で与えられる．$\boldsymbol{x}'=\boldsymbol{x}\oplus\boldsymbol{h}$ が突然変異後の状態を表す．ただし，$\oplus$ は排他的論理和である．

交叉は，遺伝子 $\boldsymbol{x}_i$ と $\boldsymbol{x}_j$ が交叉率 $\chi_{\boldsymbol{k}}$ で交叉することで，遺伝子 $\boldsymbol{x}'$ となる確率

$$C_{\boldsymbol{x}'}(\boldsymbol{x}_i,\boldsymbol{x}_j)=\sum_{\boldsymbol{k}}\frac{\chi_{\boldsymbol{k}}+\chi_{\bar{\boldsymbol{k}}}}{2} \times\delta\left[\boldsymbol{x}_i\odot\boldsymbol{k}\oplus\bar{\boldsymbol{k}}\odot\boldsymbol{x}_j=\boldsymbol{x}'\right] \quad (42)$$

で表現される．ここで，$\boldsymbol{k}\in\Re^{l\times 1}$ は交叉する部分のビットは1，それ以外は0で与えられる．$\bar{\boldsymbol{k}}$ は $\boldsymbol{k}$ の否定であり，$\delta[\cdot]$ は真なら1，それ以外は0となる条件演算子である．ただし，$\odot$ は論理積である．

遺伝子 $\boldsymbol{x}$ が淘汰，突然変異，交叉を経て遺伝子 $\boldsymbol{x}'$ となる遷移確率 $\mathcal{M}(\boldsymbol{x}'|\boldsymbol{x})$ は

$$\begin{aligned}&\mathcal{M}(\boldsymbol{x}'|\boldsymbol{x})\\&=\sum_{\boldsymbol{y},\boldsymbol{h}_1,\boldsymbol{h}_2}C_{\boldsymbol{x}'}(m_\beta(S_\beta(\boldsymbol{x}|\boldsymbol{p}),\boldsymbol{h}_1),m_\beta(S_\beta(\boldsymbol{y}|\boldsymbol{p}),\boldsymbol{h}_2))\\&=S_\beta(\boldsymbol{x}|\boldsymbol{p})\sum_{\boldsymbol{y}}S_\beta(\boldsymbol{y}|\boldsymbol{p})\\&\quad\times\sum_{\boldsymbol{h}_1,\boldsymbol{h}_2}m_\beta(\boldsymbol{x},\boldsymbol{h}_1)m_\beta(\boldsymbol{y},\boldsymbol{h}_2)\sum_{\boldsymbol{k}}\frac{\chi_{\boldsymbol{k}}+\chi_{\bar{\boldsymbol{k}}}}{2}\\&\quad\times\delta\left[(\boldsymbol{x}\oplus\boldsymbol{h}_1)\odot\boldsymbol{k}\oplus\bar{\boldsymbol{k}}\odot(\boldsymbol{y}\oplus\boldsymbol{h}_2)=\boldsymbol{x}'\right]\end{aligned} \quad (43)$$

となる．

5.2 報酬駆動型システムとしての遺伝的アルゴリズム

遷移確率 $\mathcal{M}(\boldsymbol{x}'|\boldsymbol{x})$ が遺伝子 $\boldsymbol{x}$ から遺伝子 $\boldsymbol{x}'$ へ切り替えるために必要な制御入力 $\phi_{\boldsymbol{x}';\boldsymbol{x}}$

$$\mathcal{M}(\boldsymbol{x}'|\boldsymbol{x})=p(\boldsymbol{x}'|\boldsymbol{x})\exp(\phi_{\boldsymbol{x}';\boldsymbol{x}}) \quad (44)$$

によりモデル化され，遺伝的アルゴリズムが累積無限期間平均コスト関数

$$\mathcal{R}^\pi(\boldsymbol{x})=\lim_{t_f\to\infty}\mathrm{E}\left[\sum_{\tau=0}^{t_f-1}l(\boldsymbol{x}_{(\tau)},\pi(\boldsymbol{x}_{(\tau)}))\right] \quad (45)$$

の最小化 $\mathcal{R}(\boldsymbol{x})=\min_\pi\mathcal{R}^\pi(\boldsymbol{x})$ を評価していると考える．$p(\boldsymbol{x}'|\boldsymbol{x})$ は受動的遷移であり，$p(\boldsymbol{x}'|\boldsymbol{x})=0$ なら $\mathcal{M}(\boldsymbol{x}'|\boldsymbol{x})=0$ である．また，$\sum_{\boldsymbol{x}'}\mathcal{M}(\boldsymbol{x}'|\boldsymbol{x})=\sum_{\boldsymbol{x}'}p(\boldsymbol{x}'|\boldsymbol{x})=1$，$\pi(\boldsymbol{x}_{(\tau)})=\phi_{\boldsymbol{x}'_{(\tau)};\boldsymbol{x}_{(\tau)}}$ である．

コスト関数 $l(\boldsymbol{x},\boldsymbol{u})$ は報酬 $r(\boldsymbol{x}';\boldsymbol{x})$ と制御入力 $\phi_{\boldsymbol{x}';\boldsymbol{x}}$ を用いて

$$l(\boldsymbol{x},\boldsymbol{\phi})=\mathrm{E}_{\boldsymbol{x}'\sim\mathcal{M}(\cdot|\boldsymbol{x})}[r(\boldsymbol{x}';\boldsymbol{x})+\phi_{\boldsymbol{x}';\boldsymbol{x}}] \quad (46)$$

で定義されるものとする．

このとき，Bellman 方程式は

$$\mathcal{R}(\boldsymbol{x})=\min_{\phi\in\phi_{\boldsymbol{x}';\boldsymbol{x}}}\{l(\boldsymbol{x},\boldsymbol{u}) +\mathrm{E}_{\boldsymbol{x}'\sim\mathcal{M}(\cdot|\boldsymbol{x})}[\mathcal{R}(\boldsymbol{x}')]\}-c \quad (47)$$

となる．ここで，c は平均コストである．

最適制御入力は，$\psi(\boldsymbol{x}';\boldsymbol{x})=-r(\boldsymbol{x}';\boldsymbol{x})-\mathcal{R}(\boldsymbol{x}')$ として

$$\phi^*_{\boldsymbol{x}';\boldsymbol{x}}=\psi(\boldsymbol{x}';\boldsymbol{x})-\log\mathrm{E}_{\boldsymbol{x}'\sim p(\cdot|\boldsymbol{x})}[\exp(\psi(\boldsymbol{x}';\boldsymbol{x}))] \quad (48)$$

となる [10]．

ここで，逐次得られる，限られたペイオフである適応度のもとで，最も適応度が高い遺伝子の表現を探索する報酬による最適化を考える．

【定理 3】 報酬による最適化

簡易な遺伝的アルゴリズムでは，報酬 $r(\boldsymbol{x}';\boldsymbol{x})$ を遺伝

子 $\boldsymbol{x}$ から遺伝子 $\boldsymbol{x}'$ へ戦略を切り替えたときの適応度の変化

$$r(\boldsymbol{x}';\boldsymbol{x}) = \log p(\boldsymbol{x}'|\boldsymbol{x}) \left(\frac{f(\boldsymbol{x})}{f(\boldsymbol{x}')}\right)^{\beta} \tag{49}$$

で与えて駆動することで，累積無限期間平均コスト関数

$$\mathcal{R}^{\pi}(\boldsymbol{x}) = \lim_{t_f \to \infty} \mathrm{E}\left[\sum_{\tau=0}^{t_f-1} \mathrm{E}_{\boldsymbol{x}' \sim \mathcal{M}(\cdot|\boldsymbol{x})}\left[\log\left(\frac{f(\boldsymbol{x}_{(\tau)})}{f(\boldsymbol{x}'_{(\tau)})}\right)\right]\right] \tag{50}$$

の最小化を行っている．

（証明） コスト関数 $l(\boldsymbol{x},\boldsymbol{\phi})$ は

$$\begin{aligned} l(\boldsymbol{x},\boldsymbol{\phi}) &= \mathrm{E}_{\boldsymbol{x}' \sim \mathcal{M}(\cdot|\boldsymbol{x})}[v(\boldsymbol{x}';\boldsymbol{x}) + \phi_{\boldsymbol{x}';\boldsymbol{x}}] \\ &= \beta \mathrm{E}_{\boldsymbol{x}' \sim \mathcal{M}(\cdot|\boldsymbol{x})}\left[\log\left(\frac{f(\boldsymbol{x})}{f(\boldsymbol{x}')}\right)\right] - S_{\mathcal{M}} \end{aligned} \tag{51}$$

とできる．ここで，$S_{\mathcal{M}} = \mathrm{E}_{\boldsymbol{x}' \sim \mathcal{M}(\cdot|\boldsymbol{x})}[-\log \mathcal{M}(\boldsymbol{x}'|\boldsymbol{x})]$ はエントロピーである． □

ところで，最適制御入力 $\phi^*_{\boldsymbol{x}';\boldsymbol{x}}$ は，以下の Kullback-Leibler divergence

$$\begin{aligned} &\mathrm{KL}(\mathcal{M}(\boldsymbol{x}'|\boldsymbol{x}) \,\|\, p(\boldsymbol{x}'|\boldsymbol{x})\exp(\phi_{\boldsymbol{x}';\boldsymbol{x}})) \\ &= \mathrm{E}_{\boldsymbol{x}' \sim \mathcal{M}(\cdot|\boldsymbol{x})}\left[\frac{\mathcal{M}(\boldsymbol{x}'|\boldsymbol{x})}{p(\boldsymbol{x}'|\boldsymbol{x})\exp(\phi_{\boldsymbol{x}';\boldsymbol{x}})}\right] \end{aligned} \tag{52}$$

を最小化している．遷移確率のモデル化において，Kolmogorov-Gabor 多項式で表現されるポテンシャルの存在を仮定する場合には，ボルツマンマシンと類似の手法が適用可能である [11].

遺伝的アルゴリズムは，戦略を選び行動した結果，逐次，ペイオフである適応度が判明し，報酬にもとづき入力が得られる報酬駆動型システムであるといえる．

6. おわりに

本稿では，意思決定の主体が戦略を選び行動した結果，ペイオフにもとづく効用関数の変化を報酬と考え，それを入力として動作する報酬駆動型システムを考えた．まず，効用関数の最大化が適者生存則に関連していることを紹介した．つぎに，与えられたペイオフのもとで均衡が漸近安定とならない場合に，ペイオフを変更することで漸近安定化する報酬の設計法を概説した．そして，逐次得られる限られたペイオフのもとで，最もペイオフが高い戦略を探索する報酬による最適化について概観した．

(2014年6月2日受付)

参考文献

[1] J. Hofbauer and K. Sigmund: *The Theory of Evolution and Dynamical Systems*, Cambridge University Press (1988)

[2] T. Yamada and H. Fujisaka: Stability theory of synchronized motion in coupled-oscillator systems. II –The mapping approach–; *Progress of Theoretical Physics*, Vol. 70, No. 5, pp. 1240–1248 (1983)

[3] 金子, 津田: 複雑系のカオス的シナリオ, 朝倉書店, p. 127 (1996)

[4] 堀江, 相吉, 宮野: ゲーム理論的均衡解探索のための疑似勾配系モデルのニューラルネットワーク実現とその挙動; システム/制御/情報, Vol. 12, No. 11, pp. 680–690 (1999)

[5] 奥原, 尾崎: 適者生存型学習則を適用した競合動径基底関数ネットワーク; 電子情報通信学会論文誌, Vol. J80-DII, No. 12, pp. 3191–3199 (1997)

[6] K. Okuhara, K. Sasaki and S. Osaki: Reproductive and competitive radial basis function networks adaptable to dynamical environments; *Systems and Computers in Japan*, Vol. 31, No. 13, pp. 65–73 (2000)

[7] J. Xiang and G. Chen: On the V-stability of complex dynamical networks; *Automatica*, Vol. 43, pp. 1049–1057 (2007)

[8] X. Li, X. Wang and G. Chen: Pinning a complex dynamical network to its equilibrium; *IEEE Transactions on Circuits and Systems*, Vol. 51, No. 10, pp. 2074–2087 (2004)

[9] M. D. Vose: *The Simple Genetic Algorithm: Foundations and Theory*, MIT Press, (1999)

[10] E. Todorov: Efficient computation of optimal actions; *PNAS*, Vol. 106, No. 28, pp. 11478–11483 (2009)

[11] K. Okuhara, S. Osaki and M. Kijima: Learning to design synergetic computers with an expanded symmetric diffusion network; *Neural Computation*, Vol. 11, pp. 1475–1491 (1999)

著者略歴

奥原（おくはら） 浩之（こうじ） (正会員)

1996年3月広島大学大学院工学研究科システム工学専攻博士後期課程修了．同年4月同大学工学部助手，1998年10月広島県立大学経営学部講師，2000年同助教授，2006年4月大阪大学大学院情報科学研究科准教授となり現在に至る．2006年4月大阪大学金融保険教育センター兼任，2007年4月台湾国立成功大学客員准教授，2010年4月大学共同利用機関法人情報・システム研究機構統計数理研究所客員准教授．ソフトコンピューティング技術の開発，数理モデリングと最適化・制御の研究に従事．2006年スケジューリング学会技術賞受賞．IEEE，電子情報通信学会，日本オペレーションズ・リサーチ学会などの会員．