力学系とニューラルネット

上坂　吉則

1.　はじめに

表題にある"力学系"と"ニューラルネット"はそれぞれ単独で膨大な内容をもっている分野であり，両者の共通部分をとったとしてもなお1冊の本になるくらいである．ここではその中で，関数の最小値を探索するという話題に絞って両者の絡み合いを考えてみることにする．

はじめに1変数の実数値関数：

(1)　$E(x)=\frac{1}{2}(x-2)^2$

の最小点を求めるという，ごく簡単な問題を考えてみよう．最小点が $x=2$ であることは，微分法を使うまでもなく，すぐわかる．

この問題をあえて少し難しく求めてみることにする．つまり，変数 x を時間 t の関数と考えて，与えられた関数 E から作られる次のような力学系：

(2)　$\frac{dx}{dt}=-\frac{dE}{dx}=-(x-2)=-x+2$

を考えてみる．この右辺から容易にわかるように $x<2$ ($x=2, x>2$) のとき dx/dt は正（零，負）となる．したがって上の力学系の状態（つまり，微分方程式の解の値）はどんな初期状態（初期値）から出発しても，つねに2に近づく．この2は明らかに関数 E の最小点である．このように，関数の最小点は，多くの場合，力学系の状態が漸近していく $t\to\infty$ における状態（これをその力学系の**漸近安定点**という）として求めることができる．

次に上の関数 E の閉区間 $[-1,+1]$ における最小点を求めることを考えてみよう．この区間では E は明らかに単調減少であるから，最小点は $x=1$ である．しかし，上の力学系(2)をそのまま使うと具合が悪い．つまり，初期状態を開区間 $(-1,+1)$ のどこにとっても状態は最小点に近づいてはゆくが，ついにはそこを通過していってしまう．そこで区間の両端で状態が停止するような**ストッパー**を上の力学系に入れてみる：

(3)　$\frac{dx}{dt}=-(1-x^2)\frac{dE}{dx}=(1-x^2)(2-x).$

そうすると区間 $[-1,+1]$ 上では依然として $dx/dt>0$ であるから状態は最小点 $x=1$ に近づいてゆき，充分近くにくると，式(3)の因子 $1-x^2$ が働いて，dx/dt は充分小さくなり，$t\to\infty$ で最小点の所に停止する．このように因子 $1-x^2$ は状態が最小点を通り過ぎてしまわないようにするためのブレーキの役目を果している．

このようなストッパーをもった力学系だと，仮に2点から成る集合 $\{-1,+1\}$ の上での E の最小点を求めよといわれても大丈夫である．しかし，関数 E が

(4)　$E(x)=\frac{1}{2}\left(x-\frac{1}{2}\right)^2$

となると困ってしまう．実際，このときの力学系は

(5)　$\frac{dx}{dt}=-(1-x^2)\frac{dE}{dx}=(1-x^2)\left(\frac{1}{2}-x\right)$

となるから，初期状態が何であっても，状態は $x=1/2$ に近づいてしまい，集合 $\{-1,+1\}$ 上での最小点は得られないというわけである．

しかし，少し落ちついて考えてみると，集合 $\{-1,+1\}$ の上で最小点を求めるのであるから，関数 E の定義域も集合 $\{-1,+1\}$ としてよい．このとき定義域の上ではつねに $x^2=1$ であるから，式(4)の E は

(6)　$E(x)=\frac{1}{2}\left(x^2-x+\frac{1}{4}\right)=\frac{1}{2}\left(\frac{5}{4}-x\right)$

とも書けることになる．この段階で改めて E の定義域を $[-1,+1]$ と考えることにすると，式(5)の力学系は

(7)　$\frac{dx}{dt}=-(1-x^2)\frac{dE}{dx}=\frac{1}{2}(1-x^2)>2$

となり，今度は状態 x は単調に増加して最小点 $x=1$ に漸近する．

この最後の例における最小点を求める過程は4つのステップに整理することができる：

1.　変数の値が ±1 の2値をとる実数値関数 E（式(4)）が与えられる．

2.　$x^2=1$ という関係を利用して，E の定義域を $[-1,+1]$ とみなしたときの極小点1/2を取り除いた形

うえさか　よしのり　東京理科大学　理工学部
〒278 野田市山崎2641

に変形する（式(6)）．

3．この式(6)の定義域を$[-1,+1]$と考えて，力学系(7)を作る．

4．この力学系の漸近安定点として最小点を求める．

つまり，この過程は，大げさにいえば，2値変数の関数の最小値問題の解法を与えているということになる．

1985年，カリフォルニア工科大学のHopfieldとTankがある種のニューラルネットによって2値変数の2次関数の最小値問題が解けることを示唆し，大きな関心を呼んだ[3]このアイデアの数学的なエッセンスは．上に述べた解法の4つのステップに，1点を除いて，すべて言いつくされているといってよい．

その1点とは最小にしたい関数が1変数ではなく多変数だということである．関数Eが1変数ならば$E(-1)$と$E(+1)$を計算してその大小を較べれば，最小値問題は難なく解けるわけで，力学系などを大げさに持ち出す必要はない．しかし，n個の変数の関数の最小値問題を考えるとなると，変数$x=(x_1,\cdots,x_n)$の取り得る場合の数が2^n通りと指数関数的に増大し，問題はいちじるしく困難なものになる．この難点を解消しようとするのがニューラルネットによる力学的な解法である．

それには，上の解法を多変数の場合にも通用するように，話をややペダンチックに飾り立てればよい．以下ではその"飾り"の部分を主として説明し，そののちこの種の力学系の特別の場合がニューラルネットによって実現されることを見ることにする．

2. 最小値探索のための力学系

集合$X=\{x\,|\,x=(x_1,\cdots,x_n)^t, x_i=\pm1\}$の上で定義された実数値関数$F$の最小点を探索する問題を考える．つまり，2値をとる多変数の関数の最小値問題である．ここにtはベクトルの転置を表わす．Fをこの問題の**目的関数**と呼ぶことにする．

本論に入る前に，目的関数の標準形について考えておくことにする．この種の関数はつねに多項式の形で書けることは明らかであろう．実際，パラメータ$r=(r_1,\cdots,r_n)$，$r_i=0$ or 1をもつX上の2^r個の関数：

$$\varphi(x;r)=\left(\frac{1+x_1}{2}\right)^{r_1}\cdots\left(\frac{1+x_n}{2}\right)^{r_n} \tag{8}$$

の全体はXから実数集合への関数全体が作る線形空間の基底になっていることから，Fは$\varphi(x;r)$の線形結合で書ける．したがってFは$x_1,\cdots,x_n$の多項式となるが，それだけではなく式(8)の形から容易にわかるように，$x_1,\cdots,x_n$の内の1つ，たとえばx_i，を除いて他の変数を固定するとFはx_iに関して1次式になるといういちじるしい性質をもっている．このことをFは**多重1次形式**を成しているという．別の言い方をすれば，各変数x_iに関してそのべき乗$x_i{}^p$がいっさい含まれていないということにほかならない．

組合せ的最適化問題の多くは2値多変数の実数値関数Fの最小値問題に帰着できる[4]．このとき目的関数Fが必ずしも多重1次形式にはなってないかもしれない．しかし，変数x_iは±1しか取らないから，$x_i{}^2=1$がつねに成り立っている．この関係を繰り返しFに適用すれば，やがてはx_iのべき乗をすべてFから排除することができる．したがって一般性を失なうことなく，今後は多重1次形式の最小値問題を考えることにする．与えられた目的関数を多重1次形式化する過程は上に述べたステップの2に相当する．

さて，目的関数Fの最小点を求めるために，上のステップ3にならって，Fの定義域をXからn次元超立方体$C\equiv[-1,+1]^n$に拡大した関数Eを考え，式(7)に相当する力学系：

$$\frac{dx_i}{dt}=-(1-x_i{}^2)\frac{\partial E}{\partial x_i},\quad i=1,\cdots,n \tag{9}$$

を用意する．関数Eを目的関数Fの**エネルギー**と呼ぶことにする．力学系(9)の軌道の上では

$$\frac{dE}{dt}=\sum_{i=1}^{n}\frac{\partial E}{\partial x_i}\frac{dx_i}{dt}=-\sum_{i=1}^{n}(1-x_i{}^2)\left(\frac{\partial E}{\partial x_i}\right)^2\leq 0 \tag{10}$$

となるから，関数Eは軌道上で時間とともに減少していくことになる．しかし，Eは多重1次形式であるから，いわゆる極小点はもたない．したがってEは減りっぱなしということになるが，式(10)のストッパー$1-x_i{}^2$のおかげで，立方体Cのどこかの頂点に漸近する可能性はある．もし目的関数Fの最小点x^*に漸近するならば，この力学系によってFの最小値問題が解けるわけで，めでたしということになる．

しかし，実は残念ながら，状態が漸近する点は最小点x^*以外にもたくさん存在するのである．そのことを正確に述べるために，Fの極小点という概念を導入しよう．

目的関数Fの最小点x^*とはいうまでもなく

$$\forall x\in X: F(x^*)\leq F(x) \tag{11}$$

を満たすXの点のことである．それに対して極小点という概念は必ずしも一意には定まらないが，ここでは次のように定める．まず，点$x=(x_1,\cdots,x_n)^t$の近傍$B(x)$を

$$B(x)\equiv\{(x_1,\cdots,x_{i-1},-x_i,x_{i+1},\cdots,x_n)^t \mid i=1,\cdots,n\} \tag{12}$$

と定め，点 $x^*\in X$ が

(13)　$\forall x\in B(x^*) : F(x^*)<F(x)$

を満たすとき x^* を F の**極小点**（の 1 つ）であるということにする．幾何学的には次のようなイメージになる．いま，Xを n 次元ユークリッド空間の中の原点を中心にもつ 1 辺が 2 の超立方体の頂点の集合と同一視しよう．このとき，x^* はこの立方体のどこかの頂点に対応するが，この頂点につながっている辺のもう一方の頂点を集めたものが x^* の近傍 $B(x^*)$ であり，その各頂点における F の値より $F(x^*)$ の方が小さいというわけである．

このような極小点の定義を用いると，エネルギーEとFの極小点 x^* との間に

(14)　$x_i^*\dfrac{\partial E}{\partial x_i}(x^*)<0,\quad i=1,\cdots,n$

という関係が成り立つ．

実際，E は多重 1 次形式であるから，変数 x_i に着目すると，適当な $n-1$ 変数の関数 E_1, E_2 が存在して

(15)　$$E(x_1,\cdots,x_n)=E_1(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n)x_i+E_2(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n)$$

という形に書ける．この式の x_i に $-x_i$ を代入したものを用意し辺々引くと，$\partial E/\partial x_i=E_1$ に注意して

(16)　$$E(x_1,\cdots,x_{i-1},x_i,x_{i+1},\cdots,x_n)-E(x_1\cdots,x_{i-1},-x_i,x_{i+1},\cdots,x_n)=2x_i\frac{\partial E}{\partial x_i}$$

が得られる．この式の x に極小点 x^* を代入すると

(17)　$E(x^*)-E(y^*)=2x_i^*\dfrac{\partial E}{\partial x_i}(x^*)$

となる．ここに y^* は x^* の近傍の点である．したがって上の極小点の定め方から，式(17)の左辺が負となることがわかる．

式(14)に注意すると，力学系の状態は一般に目的関数Fの極小点に漸近することを確認できる．それには，微分方程式の定性的理論［2］に従って，極小点 x^* における力学系 (9) のヤコビ行列：

(18)　$$J(x^*)\equiv\begin{bmatrix}\dfrac{\partial}{\partial x_1}\left\{-(1-x_1^2)\dfrac{\partial E}{\partial x_1}\right\} & \cdots & \dfrac{\partial}{\partial x_n}\left\{-(1-x_1^2)\dfrac{\partial E}{\partial x_1}\right\}\\ \vdots & \ddots & \vdots\\ \dfrac{\partial}{\partial x_1}\left\{-(1-x_n^2)\dfrac{\partial E}{\partial x_n}\right\} & \cdots & \dfrac{\partial}{\partial x_n}\left\{-(1-x_n^2)\dfrac{\partial E}{\partial x_n}\right\}\end{bmatrix}_{x=x^*}$$

の固有値の実部がすべて負であることを確認できればよい．この固有値は容易に計算できて，実は式(14)の左辺で与えられ，したがってすべて負である．つまり，Fの極小点はすべて漸近安定であって，力学系 (9) の状態はすべて目的関数の極小点に漸近するというわけである．

目的関数の極小点の中にはもちろん最小点も含まれている．したがって，初期状態を上手に選ぶことができれば，上の力学系によって 2 値をとる多変数の関数の最小値問題が解けるということになる．

しかしそのような初期状態を具体的にどのように設定すればよいかは，きわめて難しい問題であって，現在のところ立方体Cの中心付近，すなわち状態空間の原点近傍にとれば，多くの場合最小点に漸近する可能性が高いことが実験的に確認されているに留まっている［5］．

3. 擬似勾配系

ところで式 (9) においてストッパー，つまり，右辺の因子 $1-x_i^2$ がなければ，これは E をポテンシャルエネルギーとする**勾配系**と呼ばれる力学系になる．つまり，$x=(x_1,\cdots,x_n)^t$ に対して

(19)　$(\mathrm{grad}\,E)(x)\equiv\left(\dfrac{\partial E}{\partial x_1}(x),\cdots,\dfrac{\partial E}{\partial x_n}(x)\right)^t$

と書くことにすると，勾配系は

(20)　$\dfrac{dx}{dt}=-(\mathrm{grad}\,E)(x)$

なる連立の微分方程式に従う力学系である．

この種の力学系の性質は比較的簡単で，容易にわかるように，解軌道の上で

(21)　$$\frac{dE}{dt}=-(\mathrm{grad}\,E)(x)^t(\mathrm{grad}\,E)(x)=-\|(\mathrm{grad}\,E)(x)\|^2\leq 0$$

が成り立つ．したがって，もしEが孤立した普通の意味での極小点 x^* をもてば，E は x^* の近傍で狭義のリアプノフ関数となるから，x^* が漸近安定であることがいえる［2］．それ故，勾配系シミュレートすることによって，関数Eの最小点（一般に極小点）を求めることができるわけで，これがいわゆる**最急降下法**に他ならない．

一方，われわれの力学系を式(20)の形に書くために

(22)　$V(x)\equiv\begin{bmatrix}1-x_1^2 & & 0\\ & \ddots & \\ 0 & & 1-x_n^2\end{bmatrix}$

とおくと，式 (9) の力学系は

(23)　$\dfrac{dx}{dt}=-V(x)(\mathrm{grad}\,E)(x)$

と書くことができ，式(21)に相当する式は

(24)　$\dfrac{dE}{dt}=-(\mathrm{grad}\,E)(x)^tV(x)(\mathrm{grad}\,E)(x)$

となる．ところが，式 (9) から容易にわかるように，初期状態 $x(0)$ が超立方体C内から出発する軌道はCの中に留まり，Cにおいては $V(x)$ は半正定値であるから，この軌道の上で式(24)は非正である．したがって勾配系の場合と同じ議論によって，式(23)の平衡状態 x^* が E

の孤立した極小点ならば，それは漸近安定であることがいえる．そしてEが目的関数Fから作られたエネルギーである場合には，Fの極小点は式(23)の平衡状態でありしかもEのC内での孤立した極小点でもあることから，上と同じ結論が得られる．つまり，ヤコビ行列の固有値を計算しなくても，勾配系の議論を真似ればよかったわけである．

このように2値多変数の関数の最小値を探索する力学系は勾配系と非常によく似た力学系である．また，これまでの議論から容易に推察できるように，式(22)の$V(x)$は対角行列である必要はなく，要は軌道の存在領域で半正定値であれば同じ議論が成り立つことになる．そしてこのことから上の力学系を少し拡張することができる．

そのためにまず勾配系を少しばかり拡張しておくことにする．いま，$U\subseteq\mathscr{R}^n$ を開集合とし，EをU上で定義された連続2回微分可能な関数とする．また，VをU上で定義された関数 v_{ij} を成分とする n 次の正方行列：

(25) $$V(x)\equiv\begin{bmatrix} v_{11}(x) & \cdots & v_{1n}(x) \\ \vdots & \ddots & \vdots \\ v_{n1}(x) & \cdots & v_{nn}(x)\end{bmatrix}$$

でU上で半正定値であるとする．このとき，力学系：

(26) $$\frac{dx}{dt}=-V(x)(\mathrm{grad}\,E)(x)$$

をU上の**擬似勾配系**と呼ぶことにしよう．特に $V(x)$が単位行列のときが通常の勾配系に他ならない．上で述べたように，擬似勾配系の平衡状態 x^* がEの孤立した極小点ならば，x^* は漸近安定である．

4. ニューロダイナミックス

ここで力学系(9)に話を戻して，目的関数Fが2次関数の場合を考える：

(27) $$F(x)=-\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}x_ix_j-\sum_{i=1}^{n}b_ix_i.$$

ここに係数 a_{ii} は，一般性を失なうことなく，

(28) $$a_{ii}=0,\ a_{ij}=a_{ji}$$

と仮定してよいことに注意しておく．実際，式(27)において $a_{ii}\neq0$ ならば，関係 $x_i^2=1$ を用いてFを多重1次形式に直せばよいし，$a_{ij}\neq a_{ji}$ ならば$(a_{ij}+a_{ji})/2$を改めて a_{ij} とすればよいからである．

さて目的関数Fに対応するエネルギーEは

(29) $$E(x)=-\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}x_ix_j-\sum_{i=1}^{n}b_ix_i$$

であり，やはり式(28)が成り立つとしてよい．この式(28)に注意すると

(30) $$\frac{\partial E}{\partial x_i}=-\sum_{j=1}^{n}a_{ij}x_j-b_i,\ i=1,\cdots,n$$

と計算されるから，力学系(9)は

(31) $$\frac{dx_i}{dt}=(1-x_i^2)\left\{\sum_{j=1}^{n}a_{ij}x_j+b_i\right\},\ i=1,\cdots,n$$

となる．ここで新たに n 個の関数 u_i を

(32) $$x_i(t)\equiv\tanh u_i(t)$$

で導入する．そうするとこの式から導かれる関係：

(33) $$\frac{dx_i}{dt}=(1-x_i^2)\frac{du_i}{dt}$$

を使うことによって，上の力学系(31)は

(34) $$\frac{du_i}{dt}=\sum_{j=1}^{n}a_{ij}x_j+b_i,\ x_i=\tanh u_i,\ i=1,\cdots,n$$

という形に書き改めることができる．

ところがこの式は実はニューロン素子の時間的な挙動を支配するダイナミックスとみることができる（甘利，1978)．つまり，シナプスにおける重み係数が $a_{i1},\cdots,a_{in}$ でしきい値が $-b_i$ であるニューロンに入力 $x_1,\cdots,x_n$ が印加されたときの内部電位 u_i と出力 x_i は式(34)の微分方程式に従うというわけである．

いま，このようなニューロン素子を n 個用意し，その各出力をすべての素子の入力にフィードバックすることによって得られる回路，すなわち，相互結合型の回路を考える．そうするとこのニューラルネットの動作は他ならぬ式(34)の，したがって式(9)においてエネルギーが2次式である場合の力学系で表わされることになる．

関数Eはこのニューラルネットの**エネルギー**と呼ばれ，上で一般の場合にみたように，時間とともに減少する．この性質を利用して組合せ的最適化問題を解こうというのが Hopfield らのアイデアである．

5. おわりに

ニューラルネットによる力学系(31)は擬似勾配系の特殊な場合であることを上で見たが，そのストッパーを少し一般化すると生物系などによく現われる**レプリカ力学系**[6]と呼ばれるものの特殊な場合になっていることがわかる．そのことを簡単に紹介してこの稿を閉じることとしよう．

力学系(31)で $b_i=0$ とおいて，式(23)の形に書くと

(35) $$\frac{dx}{dt}=V(x)Ax$$

となる．ここにAは a_{ij} を i 行 j 列の成分とする行列であり，$V(x)$ は式(22)のストッパーである．このストッパーを少し変更して

(36) $V(x)\equiv\begin{bmatrix} v_{11}(x)\cdots v_{1n}(x) \\ \vdots \quad \ddots \quad \vdots \\ v_{n1}(x)\cdots v_{nn}(x)\end{bmatrix}$, $v_{ij}(x)\equiv x_i(\delta_{ij}-x_j)$

とおいてみる．ここに δ_{ij} はクロネッカーのデルタである．そうすると，この $V(x)$ は $n-1$ 単体：

(37) $\Delta\equiv\{(x_1,\cdots,x_n)^t|x_1+\cdots+x_n=1, x_i\geq 0\}$

の上で半正定値になることを示すことができるので，式(35)の $V(x)$ にこのストッパーを代入して得られる力学系はこの単体上での擬似勾配系になる．いま，この力学系を成分で書くと，式(36)に注意して

(38) $\dfrac{dx_i}{dt}=x_i[(Ax)_i-x^tAx]$, $i=1,\cdots,n$

が得られる．ここに $(Ax)_i$ はベクトル Ax の第 i 成分である．

これまでは行列 A として式(28)を満たすもの，すなわち，対角成分が零の対称行列を考えてきたが，A として一般の行列も許容すると考えた力学系がレプリケータと呼ばれるものである．したがってレプリケータは，一般には，もはや擬似勾配系というわけにはいかず，その振舞いも多様になる．そのため進化や生物行動や社会科学などの多くのダイナミックスに登場してくるが，ニューラルネットはこうした内容豊かな力学系のファミリに実は所属しているのである．

参 考 文 献

[1] 甘利 (1978)：神経回路網の数理，産業図書，323 pp.

[2] Hirsch, M.W. and Smale, S. (田村ほか訳) (1976)：力学系入門，岩波．

[3] Hopfield, J. J. and Tank, D. W. (1985)："Neural" Computation of Decisions in Optimization Problems, *Biological Cybernetics*, 52, 141-152.

[4] 上坂，尾関(1991)：パターン認識と学習のアルゴリズム，文一総合出版，第9章．

[5] Uesaka, Y. (1991)：Mathematical Aspects of Neuro-Dynamics for Combinatorial Optimization, *Trans. of the Institute of Electronics, Information and Communication of Japan*, E74, 6, 1368-1372.

[6] Zeeman, E.C. (1979)：Population Dynamics from Game Theory, in "Nitecki, Z. et al. (eds.) (1979)：Global Theory of Dynamical Systems, Springer-Verlarg, 499pp.", pp.471-497.