ある種の極値問題の解法と，その適応・学習アルゴリズムへの応用

山本　祥弘

知能情報工学科

(1995年8月29日受理)

On a Method of Solutions of Some Extremum Problems and its Applications to Adaptive and Learning Algorithms

by

Yoshihiro YAMAMOTO

Department of Information and Knowledge Engineering

(Received August 29, 1995)

Extremum problems are usually solved by differentiation of the objective functions. In this paper, an algebraic method of solutions of some extremum problems are presented. First, simple mathematical problems are solved by the new method. Second, an adaptive and a learning algorithm are derived with the same approach. All these problems are demonstrated to illustrate the idea which is proposed in this paper. It is cleared that the new method is effective to treat the problems in an unified form.

Key words : Extremum Problem, Algebraic method, Adaptive algorithm, Learning algorithm

1。はじめに

工学に限らず、多くの数理的問題において、極値を求める問題が、古来から議論されている。その最も代表的な解法として、目的関数の導関数を零とすることは、よく知られている。例えば、最小２乗法も、この微分的方法を用いている。

本論では、ある種の極値問題に対する一つの新しい解法を示す。この提案する解法は、非常に簡単なものであるが、一言で表現するのは厄介であり、敢えて言えば、従来の微分的方法に対して、代数的方法といえる。そこで、その内容を、具体的問題およびその例題を通して説明する。最初に、第２章において、初等数学でよく知られている簡単な問題を取り上げる。第３章では、本研究の本来の目的である工学問題に応用する。一つは、ニューラルネットの学習アルゴリズムへの応用であり、他の一つは、適応アルゴリズムへの応用である。この二つは基本的には同一問題として扱うことができる。より正確には、学習アルゴリズムの基本形を、多層ニューラルネットへ拡張するのが学習アルゴリズムであり、時系列問題の特徴を生かして逐次形式に拡張するのが適応アルゴリズムである。これは筆者の最近の考え方であり、詳しくは、参考文献１）～４）を参照されたい。

提案する方法の特徴は、微分的方法において、導関数を零とする解が存在しない場合も含めて統一的に議論できることである。以下で示す問題は、筆者の興味ある問題に限定されているが、さらに多くの問題にも応用可能と考えられる。

2。数学問題への応用

幾つかの数学の簡単な問題を通して、提案する解法の考え方を示す。なお、右上付きＴは転値を表す。

［問１］ｎ次元ユークリッド空間R^nで与えられた超平面

$$c=\mathbf{a}^T\mathbf{x},\quad (\mathbf{x}\in R^n) \tag{2-1}$$

へ、点$\mathbf{x}=\mathbf{x}_0$から下ろした垂線の足$\mathbf{x}=\mathbf{x}^*$および$\mathbf{x}^*-\mathbf{x}_0$の大きさを求めよ。

（解）ベクトル$\mathbf{a}$は超平面に直交しているので、αをパラメータとして、

$$\mathbf{x}^*-\mathbf{x}_0=\alpha\mathbf{a} \tag{2-2}$$

とおく。この$\mathbf{x}^*$が超平面上の点であるために、$c=\mathbf{a}^T\mathbf{x}^*$に代入すると、

$$c=\mathbf{a}^T(\mathbf{x}_0+\alpha\mathbf{a}) \tag{2-3}$$

よりαが求まり、

$$\alpha=\frac{c-\mathbf{a}^T\mathbf{x}_0}{\mathbf{a}^T\mathbf{a}} \tag{2-4}$$

となる。従って、

$$\mathbf{x}^*=\mathbf{x}_0+\frac{c-\mathbf{a}^T\mathbf{x}_0}{\mathbf{a}^T\mathbf{a}}\mathbf{a} \tag{2-5}$$

と定まる。また、垂線の長さは、

$$\|\mathbf{x}^*-\mathbf{x}_0\|=\frac{|c-\mathbf{a}^T\mathbf{x}_0|}{\|\mathbf{a}\|} \tag{2-6}$$

である。

（例１）$n=2$、すなわち、$c=a_1x_1+a_2x_2$、$\mathbf{x}_0=(x_{10},\ x_{20})^T$のときを考える。これは、高校時代より親しんできた問題であり、式(2-5)、(2-6)より

$$\mathbf{x}^*=\mathbf{x}_0+\frac{c-(a_1x_{10}+a_2x_{20})}{a_1^2+a_2^2}\mathbf{a} \tag{2-7}$$

$$\|\mathbf{x}^*-\mathbf{x}_0\|=\frac{|c-(a_1x_{10}+a_2x_{20})|}{(a_1^2+a_2^2)^{1/2}} \tag{2-8}$$

と簡単に求められる。式(2-5)、(2-6)が、任意のｎに対する公式となっているのが特徴である。

さらに、［問１］を一般化して、次の問題を考える。

［問２］R^nで与えられた余次元ｐ（$1\leqq p\leqq n$）の超平面

$$\mathbf{c}=A^T\mathbf{x},\quad (\mathbf{x}\in R^n) \tag{2-9}$$

へ、点$\mathbf{x}=\mathbf{x}_0$から下ろした垂線の足$\mathbf{x}=\mathbf{x}^*$および$\mathbf{x}^*-\mathbf{x}_0$の大きさを求めよ。ただし、$A=(\mathbf{a}_1,\dots,\mathbf{a}_p)$であり、$\mathrm{rank}A=p$とする。

（解）各ベクトル$\mathbf{a}_j$（$j=1,\dots,p$）は超平面に直交しているので、

$$\mathbf{x}^*-\mathbf{x}_0=\alpha_1\mathbf{a}_1+\cdots+\alpha_p\mathbf{a}_p=A\boldsymbol{\alpha} \tag{2-10}$$

とおく。この$\mathbf{x}^*$が超平面上の点であるために、$\mathbf{c}=A^T\mathbf{x}^*$に代入すると、

$$\mathbf{c}=A^T(\mathbf{x}_0+A\boldsymbol{\alpha}) \tag{2-11}$$

より$\boldsymbol{\alpha}$が求まり、

$$\boldsymbol{\alpha}=(A^TA)^{-1}(\mathbf{c}-A^T\mathbf{x}_0) \tag{2-12}$$

となる。ただし、A^TAは正則、従って、Ａの列ベクトル$\mathbf{a}_1,\dots,\mathbf{a}_p$は線形独立としている。これより、

$$\mathbf{x}^*=\mathbf{x}_0+A(A^TA)^{-1}(\mathbf{c}-A^T\mathbf{x}_0) \tag{2-13}$$

$$\|\mathbf{x}^*-\mathbf{x}_0\|=\|A(A^TA)^{-1}(\mathbf{c}-A^T\mathbf{x}_0)\| \tag{2-14}$$

となる。ここで$p=1$のとき、式(2-13)、(2-14)はそれぞれ式(2-5)、(2-6)となる。

（例２）$n=3$、$p=2$、すなわち、

$c_1 = \mathbf{a}^T\mathbf{x}$、$c_2 = \mathbf{b}^T\mathbf{x}$;　(2-15)

$\mathbf{c} = (c_1,\ c_2)^T$、$A = (\mathbf{a},\ \mathbf{b})$

とすると、

$$
\begin{aligned}
&A(A^TA)^{-1}(\mathbf{c} - A^T\mathbf{x}_0) \\
&= \{[(c_1 - \mathbf{a}^T\mathbf{x}_0)\mathbf{b}^T\mathbf{b} - (c_2 - \mathbf{b}^T\mathbf{x}_0)\mathbf{a}^T\mathbf{b}]\mathbf{a} \\
&\quad + [(c_2 - \mathbf{b}^T\mathbf{x}_0)\mathbf{a}^T\mathbf{a} \\
&\qquad - (c_1 - \mathbf{a}^T\mathbf{x}_0)\mathbf{a}^T\mathbf{b}]\mathbf{b}\}/d \qquad (2\text{-}16)
\end{aligned}
$$

$d = \mathbf{a}^T\mathbf{a}\,\mathbf{b}^T\mathbf{b} - (\mathbf{a}^T\mathbf{b})^2$

となる。例えば、$\mathbf{a} = (0,\ 1,\ 1)^T$、$\mathbf{b} = (0,\ 1,\ -1)^T$、$c_1 = c_2 = 0$、従って、超平面が x_1 軸の場合、

$A(A^TA)^{-1}(\mathbf{c} - A^T\mathbf{x}_0) = (0,\ -x_{20},\ -x_{30})^T$、

$\mathbf{x}^* = (x_{10},\ 0,\ 0)^T$、　(2-17)

$\|\mathbf{x}^* - \mathbf{x}_0\| = (x_{20}{}^2 + x_{30}{}^2)^{1/2}$

となり、幾何学的イメージと一致する。

［問３］［問２］と同じ問題で、$p > n$ の場合はどうなるか。ただし、$\mathrm{rank}\,A = n$ とする。

（解）前問の解と同様に、

$\mathbf{x}^* - \mathbf{x}_0 = A\boldsymbol{\alpha}$　(2-18)

とおき、$\mathbf{c} = A^T\mathbf{x}^*$ に代入すると

$\mathbf{c} = A^T(\mathbf{x}_0 + A\boldsymbol{\alpha})$　(2-19)

従って、

$A^TA\boldsymbol{\alpha} = \mathbf{c} - A^T\mathbf{x}_0$　(2-20)

となる。このとき $p > n$ より行列 A^TA は正則であり得ない。そこで、上式左から A をかけて

$AA^TA\boldsymbol{\alpha} = A(\mathbf{c} - A^T\mathbf{x}_0)$　(2-21)

とする。このとき AA^T は正則であるので、

$A\boldsymbol{\alpha} = (AA^T)^{-1}A(\mathbf{c} - A^T\mathbf{x}_0)$　(2-22)

となり、これより、

$\mathbf{x}^* - \mathbf{x}_0 = (AA^T)^{-1}A(\mathbf{c} - A^T\mathbf{x}_0)$　(2-23)

あるいは、

$\mathbf{x}^* = (AA^T)^{-1}A\mathbf{c}$　(2-24)

となる。

$\mathbf{c} = A^T\mathbf{x}^*$ を満たす $\mathbf{x}^*$ は存在しないが、上に求めた解は誤差最小の解であること、すなわち、

$J = \dfrac{1}{2}\|\mathbf{c} - A^T\mathbf{x}\|^2$　(2-25)

を最小とする解であることが、一般化逆行列の理論から言える。あるいは、実際に式(2-25)を $\mathbf{x}$ で微分することより、式(2-24)の解を得ることができる。従って、$\mathbf{x}^*$ が初期点 $\mathbf{x}_0$ に無関係となることも当然である。

（例３）$n = 2$、$p = 3$，すなわち、

$c_1 = \mathbf{a}_1{}^T\mathbf{x}$、$c_2 = \mathbf{a}_2{}^T\mathbf{x}$、$c_3 = \mathbf{a}_3{}^T\mathbf{x}$　(2-26)

$A = (\mathbf{a}_1,\ \mathbf{a}_2,\ \mathbf{a}_3)$、$\mathbf{c} = (c_1,\ c_2,\ c_3)^T$

のとき、

$$
\begin{aligned}
\mathbf{x}^* &= (AA^T)^{-1}A\mathbf{c} \\
&= (\mathbf{a}_1\mathbf{a}_1{}^T + \mathbf{a}_2\mathbf{a}_2{}^T + \mathbf{a}_3\mathbf{a}_3{}^T)^{-1} \\
&\qquad \cdot (c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3) \qquad (2\text{-}27)
\end{aligned}
$$

となる。例えば、$\mathbf{c} = (0,\ 0,\ 1)^T$、$\mathbf{a}_1 = (1,\ 0)^T$、$\mathbf{a}_2 = (0,\ 1)^T$、$\mathbf{a}_3 = (1,\ 1)^T$、すなわち、３直線が x_1 軸、x_2 軸および直線 $x_1 + x_2 = 1$ のとき、

$$
\mathbf{x}^* = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad (2\text{-}28)
$$

となる。この解は、

$$
\begin{aligned}
J &= (c_1 - \mathbf{a}_1{}^T\mathbf{x})^2 + (c_2 - \mathbf{a}_2{}^T\mathbf{x})^2 \\
&\qquad + (c_3 - \mathbf{a}_3{}^T\mathbf{x})^2 \\
&= x_1{}^2 + x_2{}^2 + (1 - x_1 - x_2)^2 \qquad (2\text{-}29)
\end{aligned}
$$

の、x_1、x_2 に関する第１偏導関数を零とした連立一次方程式からも得られる。

［問４］R^n で与えられた超曲面

$c = f(\mathbf{x})$　(2-30)

へ、点 $\mathbf{x} = \mathbf{x}_0$ から下ろした垂線の足 $\mathbf{x} = \mathbf{x}^*$ を求めよ。ただし、$f(\mathbf{x})$ は連続微分可能とする。ここに、曲面への垂線とは、その点での接平面と垂線とが直交することを意味する。この問題はまた、曲面への最短距離を求める問題でもある。

（解）超曲面の点 $\mathbf{x} = \mathbf{x}^*$ における法線ベクトル $\mathbf{n}$ は

$\mathbf{n} = f'(\mathbf{x}^*)$　(2-31)

である。これより、求める垂線は α を未知パラメータとして、

$\mathbf{x}^* = \mathbf{x}_0 + \alpha f'(\mathbf{x}^*)$　(2-32)

と表される。そこで、式(2-30)と(2-32)を連立させて解くことができれば、それが求める解である。

（例４）超曲面として、n次元超球

$\sum_{i=1}^{n} x_i{}^2 = \mathbf{x}^T\mathbf{x} = r^2$　(2-33)

を考える。法線ベクトルは、

$f'(\mathbf{x}) = 2\mathbf{x}$　(2-34)

であるので、式(2-32)は、

$\mathbf{x}^* = \mathbf{x}_0 + \alpha \cdot 2\mathbf{x}^*$　(2-35)

である。これより

$\mathbf{x}^* = \dfrac{1}{1 - 2\alpha}\mathbf{x}_0$　(2-36)

となり、これを式(2-33)に代入することより、

$$1-2\alpha=\pm\frac{\|\mathbf{x}_0\|}{r} \tag{2-37}$$

従って、

$$\mathbf{x}^*=\pm\frac{r}{\|\mathbf{x}_0\|}\mathbf{x}_0 \tag{2-38}$$

が得られる。また、

$$\|\mathbf{x}^*-\mathbf{x}_0\|=\left|1\pm\frac{r}{\|\mathbf{x}_0\|}\right|\|\mathbf{x}_0\| \tag{2-39}$$

なる当然の結果を得る。複号の一つは超球への最短距離を、一方は最長距離を示している。

この［問4］は非線形系への拡張であり、上に記した結果だけからは、解が求まるのは限られた問題である。実際、多くの問題で、高次代数方程式あるいは非線形方程式の解を仮定しなければならず、解析的には一般に困難である。今後さらに検討してみたい。

3。工学問題への応用

ニューラルネットワークは、近年さまざまな分野で応用され、その有効性も確認されているが、しかし、まだ幾つかの問題点を抱えている。局所解へのトラップと併せて、学習速度の遅いこともその一つである。多くの分野では、学習は、いわゆるオフラインで実行されるので、学習速度はたいして問題とならないが、制御系の設計、特に、適応制御の分野では、オンラインでの実行が要求されるので、学習速度の問題は致命的となる。しかし、非線形系に対するニューラルネットの有効性は非常に魅力的であるので、従来の学習アルゴリズムに代わる新しい方法を開発し、適応制御の設計にニューラルネットを応用してみたいというのが、筆者の最近の課題である。

本章では、適応および学習アルゴリズムの導出を通して、提案する考え方が工学問題にも応用可能であることを示す。

3.1 ニューラルネットワークの学習アルゴリズム

本節で扱うニューラルネットは階層型であり、その学習は教師付き学習である。ただし、ここでは2層回路に限定して議論する。2層回路の学習規則としては従来、デルタルール[5]が知られている。これを多層回路に一般化したのが誤差逆伝搬法であることは、多くの文献からも知ることができる。いま、s入力－r出力の2層回路モデルを次のように定める。

$$c_j=f(z_j) \tag{3-1}$$

$$z_j=\mathbf{w}_j{}^T\mathbf{a}=\sum_{i=1}^{s}w_{ij}a_i,\quad j=1,\ldots,r \tag{3-2}$$

ここに、c_jは出力、a_iは入力であり、w_{ij}が重み係数である。また、非線形関数fの逆関数の存在を仮定する。式(3-1)、(3-2)をベクトル行列表現すれば、

$$\mathbf{c}=f(\mathbf{z}),\quad \mathbf{z}=\mathbf{W}^T\mathbf{a} \tag{3-3}$$

となる。行列Wは$s\times r$型であり、$\mathbf{W}=(\mathbf{w}_1,\ldots,\mathbf{w}_r)$である。いま、教師信号を$d_j$あるいはベクトル表現で**d**とするとき、出力誤差の評価として、

$$J=\frac{1}{2}(\mathbf{d}-\mathbf{c})^T(\mathbf{d}-\mathbf{c})=\frac{1}{2}\sum_{j=1}^{r}(d_j-c_j)^2 \tag{3-4}$$

を考える。従来の学習規則の考え方は、この評価Jを減少させる方向に重みパラメータに修正を加える、いわゆる勾配法(最急降下法)であった。これは、評価Jを重みWで微分しても、導関数を零とする解が得られないからである。そこで、本論の考え方は、Jの導関数を扱うのではなく、直接、評価Jを零とする解を求めることである。明らかに、$\mathbf{d}=\mathbf{c}$のとき評価Jは零となるが、非線形関数fによって、$\mathbf{d}=\mathbf{c}$とするWは決定できない。しかし、fの逆関数の存在を仮定すれば、$\mathbf{d}=\mathbf{c}$とする代わりに

$$f^{-1}(\mathbf{d})=\mathbf{z} \tag{3-5}$$

とすることができれば、$\mathbf{d}=\mathbf{c}$、従って、評価Jを零とすることになる。非線形関数fが0－1タイプの2値関数のような不連続関数であれば、この議論は実行不可能となるが、実際には、例えば、シグモイド関数のような連続関数で十分近似できることが分かっているので、逆関数の存在を仮定しても一般性を失わないことになる。以上より、考えている問題は、

$$f^{-1}(\mathbf{d})=\mathbf{W}^T\mathbf{a} \tag{3-6}$$

を満たすWを決定する問題となる。これはまさに、前章の［問2］の形式となっており、同様な方法で解くことができる。すなわち、Wの修正量⊿Wを

$$\varDelta\mathbf{W}=\mathbf{a}\boldsymbol{\phi}^T \tag{3-7}$$

とおき、

$$f^{-1}(\mathbf{d})=(\mathbf{W}+\varDelta\mathbf{W})^T\mathbf{a}=\mathbf{W}^T\mathbf{a}+\boldsymbol{\phi}\mathbf{a}^T\mathbf{a} \tag{3-8}$$

から未知ベクトル$\boldsymbol{\phi}$が求まり、従って、

$$\varDelta\mathbf{W}=\frac{\mathbf{a}}{\mathbf{a}^T\mathbf{a}}(f^{-1}(\mathbf{d})-\mathbf{W}^T\mathbf{a})^T \tag{3-9}$$

と、求める修正量⊿Wが決定される。因みに、修正量を式(3-7)のようにおくのは、評価Jを減少させる方向であるからとも言える。実際、式(3-4)をWで微分すると、

$$\frac{dJ}{dW}=\mathbf{a}(-f_1{}'(z_1)(d_1-c_1),\ldots) \qquad (3\text{-}10)$$

となり、左辺の各列ベクトルは、ベクトル**a**に比例していることがわかる。

以上の結果は、一組の入出力データに対する逐次修正アルゴリズムであるが、複数のデータの組に対する一括処理の学習アルゴリズムも類似な方法で求めることができる。そのためにデータの組に添字を付して、

$$\mathbf{c}_p=f(\mathbf{z}_p)、\ \mathbf{z}_p=\mathbf{W}_p{}^T\mathbf{a}_p \qquad (3\text{-}11)$$

と式(3-3)を書き直す。そして、p番目のデータからM組のデータまでを一括表現すれば、

$$\mathbf{C}_p=f(\mathbf{Z}_p)、\ \mathbf{Z}_p=\mathbf{W}_p{}^T\mathbf{A}_p \qquad (3\text{-}12)$$

と表される。ここに、

$$\begin{aligned}\mathbf{C}_p&=(\mathbf{c}_p,\mathbf{c}_{p+1},\ldots,\mathbf{c}_{p+M-1})\\ \mathbf{Z}_p&=(\mathbf{z}_p,\mathbf{z}_{p+1},\ldots,\mathbf{z}_{p+M-1}) \qquad (3\text{-}13)\\ \mathbf{A}_p&=(\mathbf{a}_p,\mathbf{a}_{p+1},\ldots,\mathbf{a}_{p+M-1})\end{aligned}$$

である。行列$\mathbf{C}_p$、$\mathbf{Z}_p$、$\mathbf{A}_p$のサイズはそれぞれ$r\times M$, $r\times M$、$s\times M$である。そこで、

$$\varDelta\mathbf{W}_p=\mathbf{A}\Phi、\quad \Phi:未知行列 \qquad (3\text{-}14)$$

とおき、

$$\begin{aligned}f^{-1}(\mathbf{D}_p)&=(\mathbf{W}_p+\varDelta\mathbf{W}_p)^T\mathbf{A}_p\\ &=\mathbf{W}_p{}^T\mathbf{A}_p+\Phi^T\mathbf{A}_p{}^T\mathbf{A}_p \qquad (3\text{-}15)\end{aligned}$$

従って、

$$\mathbf{A}_p{}^T\mathbf{A}_p\Phi=(f^{-1}(\mathbf{D}_p)-\mathbf{W}_p{}^T\mathbf{A}_p)^T \qquad (3\text{-}16)$$

からΦを決定すればよい。ここに$r\times M$型行列$\mathbf{D}_p$は教師信号ベクトルのM個の組で

$$\mathbf{D}_p=(\mathbf{d}_p,\mathbf{d}_{p+1},\ldots,\mathbf{d}_{p+M-1}) \qquad (3\text{-}17)$$

である。いま、行列$\mathbf{A}_p$がフルランクを持っているとしても、行列$\mathbf{A}_p{}^T\mathbf{A}_p$が正則であるとは限らないので、以下の二つの場合に分けられる。

(1) $\mathbf{A}_p{}^T\mathbf{A}_p$が正則のとき：

これは、行列$\mathbf{A}_p$が列フルランクを持つ場合で、$M\leqq s$であることが必要である。このとき式(3-16)からΦが得られ、

$$\varDelta\mathbf{W}_p=\mathbf{A}_p(\mathbf{A}_p{}^T\mathbf{A}_p)^{-1}(f^{-1}(\mathbf{D}_p)-\mathbf{W}_p{}^T\mathbf{A}_p)^T \qquad (3\text{-}18)$$

と決定される。

(2) $\mathbf{A}_p\mathbf{A}_p{}^T$が正則のとき：

これは、行列$\mathbf{A}_p$が行フルランクを持つ場合で、$s\leqq M$であることが必要である。このとき、式(3-16)に左から$\mathbf{A}_p$をかけて$\mathbf{A}_p\Phi$が求まる。これより、

$$\varDelta\mathbf{W}_p=(\mathbf{A}_p\mathbf{A}_p{}^T)^{-1}\mathbf{A}_p(f^{-1}(\mathbf{D}_p)-\mathbf{W}_p{}^T\mathbf{A}_p)^T \qquad (3\text{-}19)$$

あるいは、

$$\begin{aligned}\mathbf{W}_{p+M}&=\mathbf{W}_p+\varDelta\mathbf{W}_p\\ &=(\mathbf{A}_p\mathbf{A}_p{}^T)^{-1}\mathbf{A}_p f^{-1}(\mathbf{D}_p)^T \qquad (3\text{-}20)\end{aligned}$$

と決定される。

ここで注意すべきことは、式(3-18)に対しては、

$$f^{-1}(\mathbf{D}_p)-(\mathbf{W}_p+\varDelta\mathbf{W}_p)^T\mathbf{A}_p=0 \qquad (3\text{-}21)$$

となっているが、式(3-19)または(3-20)に対しては、

$$(f^{-1}(\mathbf{D}_p)-(\mathbf{W}_p+\varDelta\mathbf{W}_p)^T\mathbf{A}_p)\mathbf{A}_p{}^T=0 \qquad (3\text{-}22)$$

としか一般にならないことである。

なお、多層回路に対する学習アルゴリズムは、この結果を基にして、さらに工夫が必要であるが、詳しくは参考文献2）を、そして一括処理に関しては4）を参照されたい。

3.2 適応アルゴリズム

従来、適応アルゴリズムと学習アルゴリズムとは何ら関係なく別個に議論されてきたが、適応も、実は学習を行っているのであり、学習した結果を適応的に利用しているものと理解できる。このことは、適応制御の分野では、とくに顕著である。従って、学習アルゴリズムを前節のように考えるとき、適応アルゴリズムも全く同様に得ることができる。言い換えれば、適応アルゴリズムは学習アルゴリズムと同じことであり、ただ、その時系列システムへの応用であると見なすことができる。以下、これを示そう。

いま、次のような回帰モデルで表されるシステムを考える。

$$y_k=\boldsymbol{\theta}^T\mathbf{v}_k \qquad (3\text{-}23)$$

ここに、y_kはスカラー出力、$\boldsymbol{\theta}$はN次元未知パラメータベクトル、$\mathbf{v}_k$はN次元既知信号ベクトルである。未知ベクトル$\boldsymbol{\theta}$を決定するための評価を式(3-23)の出力誤差として、

$$J_k=\frac{1}{2}\sum_{i=0}^{M-1}(y_{k-i}-\boldsymbol{\theta}^T\mathbf{v}_{k-i})^2 \qquad (3\text{-}24)$$

とする。ここにMは任意の正整数であり、現時点から過去Mステップにわたるデータを考慮している。学習アルゴリズムでは、一括処理に対応している。なお、式(3-24)による結果を、修正最小2乗適応アルゴリズム[6] (Truncated least squares adaptive algorithm)と呼んでおり、M＝kとすると、通常の最小2乗法となっている。式(3-24)をベクトル行列表現するために、以下の記号を導入する。

$$\mathbf{y}_k=(y_k,\ldots,y_{k-M-1}) \qquad (3\text{-}25)$$

$V_k=(v_k,\ldots,v_{k-M-1})$

y_kはM次元行ベクトル、V_kはN×M型行列である。このとき式(3-24)は

$$J_k=\frac{1}{2}(y_k-\theta^T V_k)(y_k-\theta^T V_k)^T \quad (3\text{-}26)$$

と表現される。従来の適応アルゴリズムは、式(3-24)あるいは(3-26)をθで微分し、正規方程式を導くことから求めている。しかし、本論では、提案する考え方に従い、

$$\varDelta\theta_{k-1}=\sum_{i=1}^{M-1}\phi_i v_{k-i}=V_k\phi \quad (3\text{-}27)$$

とおく。ただし、

$$\begin{aligned}&\varDelta\theta_{k-1}=\theta_k-\theta_{k-1}\\&\phi=(\phi_0,\ldots,\phi_{M-1})^T\end{aligned} \quad (3\text{-}28)$$

である。そして、評価J_kを零とするように、

$$y_k=\theta_k{}^T V_k=(\theta_{k-1}+\varDelta\theta_{k-1})^T V_k \quad (3\text{-}29)$$

に式(3-27)を代入する。その結果、

$$V_k{}^T V_k\phi=(y_k-\theta_{k-1}{}^T V_k)^T \quad (3\text{-}30)$$

となるが、式(3-16)に対すると同様に、二つの場合に分類される。

(1)M≦Nのとき：これは、V_kが列フルランクをもち、従って、行列$V_k{}^T V_k$が正則であるとき、

$$\varDelta\theta_{k-1}=V_k(V_k{}^T V_k)^{-1}(y_k-\theta_{k-1}{}^T V_k)^T \quad (3\text{-}31)$$

(2)N≦Mのとき：これは、V_kが行フルランクをもち、従って、行列$V_k V_k{}^T$が正則であるときである。式(3-30)に左からV_kをかけ、整理すれば、

$$\varDelta\theta_{k-1}=(V_k V_k{}^T)^{-1}V_k(y_k-\theta_{k-1}{}^T V_k)^T \quad (3\text{-}32)$$

となる。

次に、この結果を、出力がr次元ベクトルy_kの場合に拡張する。すなわち、システムは

$$y_k=\Theta^T v_k \quad (3\text{-}33)$$

であり、これを過去Mステップまでまとめて行列表現すれば、

$$Y_k=\Theta^T V_k \quad (3\text{-}34)$$

となる。ここに、ΘはN×r型の未知パラメータ行列であり、Y_kはr×M型行列で、

$$Y_k=(y_k,\ldots,y_{k-M-1}) \quad (3\text{-}35)$$

である。前と同様にして、

$$\varDelta\Theta_{k-1}=V_k\Phi \quad (3\text{-}36)$$

とおくと、

$$V_k{}^T V_k\Phi=(Y_k-\Theta_{k-1}{}^T V_k)^T \quad (3\text{-}37)$$

となり、前と同様に、

(1)M≦Nのとき：

$$\varDelta\Theta_{k-1}=V_k(V_k{}^T V_k)^{-1}(Y_k-\Theta_{k-1}{}^T V_k)^T \quad (3\text{-}38)$$

(2)N≦Mのとき：

$$\varDelta\Theta_{k-1}=(V_k V_k{}^T)^{-1}V_k(Y_k-\Theta_{k-1}{}^T V_k)^T \quad (3\text{-}39)$$

なる結果が得られる。

以上、適応アルゴリズムの導出を述べてきたが、ここに記した範囲に限定すれば、適応アルゴリズムと学習アルゴリズムは全く同一であることがわかる。しかし、適応アルゴリズムとしては、式(3-31),(3-32)あるいは(3-38),(3-39)だけでは不十分であり、さらに時系列の特性を生かして、より簡潔な逐次形式に変形しなければならない。特に、逆行列の計算を避けることが重要である。詳細は、参考文献3）を参照されたい。

4。おわりに

本論では、数学の初等的問題ならびに学習・適応アルゴリズムの導出問題を採り上げ、一貫して、一つの考え方から、それぞれの解を得ることができることを示してきた。採り上げた問題はすべて、極値問題に属するものであり、さらに多くの問題も、同じ考え方から解を求めることができるものと期待される。代数的方法とでも言える提案してきた考え方は、いわば、一般化逆行列の応用に分類されるものである。その特徴は、従来の微分的方法よりも、より広い範囲を統一的に扱うことができることである。

工学的応用としての適応・学習アルゴリズムは、そのアルゴリズム自身を示すのが本論の目的ではないので、詳細は参考文献に委ねることにする。

参考文献

1）山本：ニューロ回路の学習規則と適応アルゴリズム、システム制御情報学会論文誌、7-12, 533/535, 1994.
2）山本：教師付き学習の新しい学習規則、第6回自律分散システムシンポジウム資料、1/6, 1995.
3）山本：適応アルゴリズム～学習アルゴリズムとの統一～、第15回適応制御シンポジウム資料、87/90, 1995.
4）山本：一般化誤差逆伝搬法、情報処理学会第50回全国大会講演論文集、4Q－2, 1995.
5）D.E.Rumelhart & J.L.McCleland : Parallel Distributed Processing, Vol.1, MIT Press, 1989.
6）山本：修正最小2乗法による適応アルゴリズム、計測自動制御学会論文集、26-12, 22/27, 1990.