One Dimensional Optimization using Learning Network

Taishin Chung and Zeungnam Bien**

* Department of Electrical Engineering, KAIST, 373-1 Kusong-dong, Yusong-gu, Taejon, 305-701, Korea

Abstract

One dimensional optimization problem is considered. we propose a method to find the global minimum of one-dimensional function with no gradient information but only the finite number of input-output samples. We construct a learning network which has a good learning capability and of which global maximum(or minimum) can be calculated with simple calculation. By teaching this network to approximate the given function with minimal samples, we can get the global minimum of the function. We verify this method using some typical examples.

I. Introduction

To find the global maximum(or minimum), we adopt a special learning network so-called OBLN (Orthogonal Basis based Learning Network) which is designed to have some useful characteristics. One of these characteristics is that we can get the global maximum(or minimum) of this network by using only a simple calculation with respect to each values of learning parameters. The approach is to teach this network to learn the target function with minimal samples, so that the global maximum(or minimum) of target function can be achieved through the learning network.

In this paper, we suggest a learning network which satisfies the spec. mentioned above, and prove a theorem which is concerned with the number of samples which is required to teach the network so that it can mimic the target function completely.

II. OBLN-Orthogonal Basis based Learning Network.

The structure of the OBLN is the linear combination of some orthogonal functions as follows.

$$N(x,w)=\sum_{i} w_i \cdot \phi_i(x)$$

where w is the learning parameter and the ϕ is the orthogonal bases defined on [0,T]. The orthogonal basis set is generated from a special function set called base function set through the Gram-schmidt procedure.

The base function set B is a set of functions which are real valued function defined on [0,T], and satisfies following equations.

$$B=\left\{\bar{\phi}_i \mid \quad \bar{\phi}_i:[0,T]\rightarrow R, \quad \bar{\phi}_i(t)=\bar{\phi}_1(i\cdot t), \quad i=0,1,2,\dots\right\}$$

$$\bar{\phi}_1(t)=\begin{cases} -t\cdot(t-\frac{T}{2}) & , \ 0<t<\frac{T}{2} \\ (t-\frac{T}{2})\cdot(t-T) & , \ \frac{T}{2}<t<T \end{cases}$$

Fig. 1 shows the plot of the base function set.

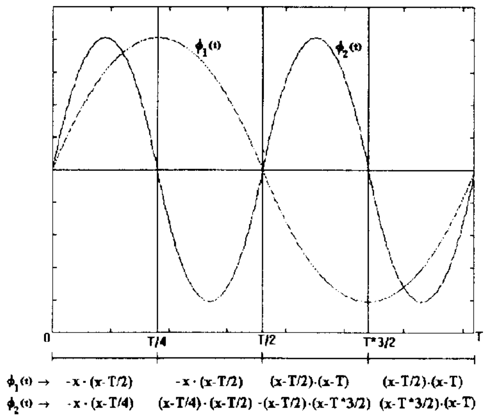

Fig. 1 Shape of base function

The network can be represented using the base function set instead of basis function set by basis change as follows

$$N(x,w)=\sum_i w_i\cdot\phi_i(x)=\sum_i \bar{w}_i\cdot\bar{\phi}_i(x)$$

where the $\bar{w}_i$ can be calculated from the w_i using the following relation.

$$w_i = Q\,\bar{w}_i$$

$$\bar{w}_i = Q^{-1}\,w_i$$

where ith column of the matrix Q is the representation of the $\bar{\phi}_i$ with respect to ϕ_i, and this can be expressed as follows.

$$Q_{i-thcolumn}=\left[\bar{\phi}_i|\phi_1 \quad \bar{\phi}_i|\phi_2 \quad \Lambda \quad \bar{\phi}_i|\phi_{i-1} \quad \|\bar{\phi}_i\|\right]$$

where $\bar{\phi}_i|\phi_j$ denotes the inner product of $\bar{\phi}_i$ and ϕ_j.

A characteristics of this network is that at each interval shown in (Fig. 1), the network can be expressed with only second order polinomials so that the calculation of the maximum(or minimum) value, which can be calculated from setting the derivative of the polinomials zero, drops into a problem of solving just a first order equation, that is, in the case of first interval,

$$\begin{aligned} N(w,t) &= \sum_i \bar{w}_i\cdot\bar{\phi}_i(t) \\ &= \sum_i \bar{w}_i\cdot(-1)\cdot t\cdot(t-\frac{T}{2\cdot i}) \end{aligned}$$

Then the optimal value in this interval can be calculated by solving the following equation.

$$\begin{aligned} \frac{\partial}{\partial t}N(w,t)\Big|_{t^*} &= \sum_i \bar{w}_i\cdot\frac{\partial}{\partial t}\bar{\phi}_i(t)\Big|_{t^*} \\ &= \sum_i \bar{w}_i\cdot\frac{\partial}{\partial t}\left[(-t)\cdot(t-\frac{T}{2\cdot i})\right]\Big|_{t^*} = \sum_i \bar{w}_i\cdot(-2\cdot t+\frac{T}{2\cdot i})\Big|_{t^*} \\ &= -2\cdot\left(\sum_i \bar{w}_i\right)\cdot t+\frac{T}{2}\cdot\left(\sum_i\frac{\bar{w}_i}{i}\right)\Big|_{t^*} = 0 \end{aligned}$$

Thus,

$$t^* = \frac{T\cdot\sum_i\frac{\bar{w}_i}{i}}{4\cdot\sum_i \bar{w}_i}$$

After the learing of the target function is finished, we can get the maxima(or minima) by solving the first order equations in each interval, and the largest(or smallest) value in these maxima(or minima) is the global maximum(or minimum).

III. Number of sampling

What remain are how we can learn this network efficiently and how many samples we need to teach the network completely. The following theorem gives the minimal numbers of sampling required.

Theorem 1.

Let $y \in H$ be a real-valued function on [0,T] and assuming that y can be represented by OBLN within W_N-th order, that is,

$$y(t) = \sum_{i=0}^{W_N} Y_i \cdot \phi_i(t)$$

where Y_i is learing parameter and ϕ_i is i-th order orthogonal basis which is generated from so-called base function set

$$B = \left\{ \bar{\phi}_i \mid \bar{\phi}_i : [0,T] \to R, \quad \bar{\phi}_i(t) = \bar{\phi}_1(i \cdot t), \quad i = 0,1,2,\dots \right\}$$

through Gram-schmidt process. If $\overline{\phi}_i$ can be represented within W_F-th order Fourier series, then the sufficient condition for complete sampling is

$$N > 2 \cdot W_F \cdot W_N$$

where N = T/Ts (Ts : sampling period).

The proof of above theorem is very similar with the sampling theory. Before proof, we are going to show two lemma which shows the frequency characteristics of orthogonal basis set.

Lemma 1

Let Y_n be the representation of y(t) with respect to a orthogonal basis $\phi(t)$, that is,

$$y(t) = \sum_{n=0}^{\infty} Y_n \cdot \phi_n(t)$$

Then the followings are hold.

$$(1) \quad F(y)^k = \sum_{n=0}^{\infty} Y_n \cdot \Phi_n^k$$

$$(2) \quad Y_n = \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot F(y)^k$$

$$(3) \quad = \sum_{n'=0}^{\infty} Y_{n'} \cdot \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot \Phi_{n'}^k$$

where

$F(y)^k$ is Fourier coefficient of $y(t)$

and Φ_n^k is Fourier coefficient of $\phi_n(t)$

Lemma 1 can be proved as following.

$$y(t) = \sum_{n=0}^{\infty} Y_n \cdot \phi_n(t) = \sum_{n=0}^{\infty} Y_n \cdot \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot e^{jk\omega t}$$

$$= \sum_{k=-\infty}^{\infty} \left(\sum_{n=0}^{\infty} Y_n \Phi_n^k\right) \cdot e^{jk\omega t}$$

$$= \sum_{k=-\infty}^{\infty} F(y)^k \cdot e^{jk\omega t} \quad \text{------------(1)}$$

$$Y_n = \int_{<T>} y(t)\overline{\phi}_n(t)dt = \int_{<T>} y(t) \cdot \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot e^{-jk\omega t} dt$$

$$= \sum_{k=-\infty}^{\infty} \Phi_n^k \int_{<T>} y(t) e^{-jk\omega t} dt$$

$$= \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot F(y)^k \quad \text{--------------(2)}$$

$$= \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot \sum_{n'=0}^{\infty} Y_{n'} \Phi_{n'}^k$$

$$= \sum_{n'=0}^{\infty} Y_{n'} \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot \Phi_{n'}^k \quad \text{-------------(3)}$$

Lemma 2

Let a orthogonal function sequence $\phi_1, \phi_2, \dots, \phi_n$ be generated from a sequence of base function $\overline{\phi}_1, \overline{\phi}_2, \ldots, \overline{\phi}_n$, where

$\overline{\phi}_i \in B = \left\{ \overline{\phi}_i \,\middle|\, \overline{\phi}_i : [0,T] \rightarrow R, \overline{\phi}_i(t) = \overline{\phi}_1(i \cdot t), i = 0,1,2,\ldots \right\}$

If ϕ_1 can be represented within W_F-order fourier series, then ϕ_n can be represented within $n \times W_F$-order fourier series.

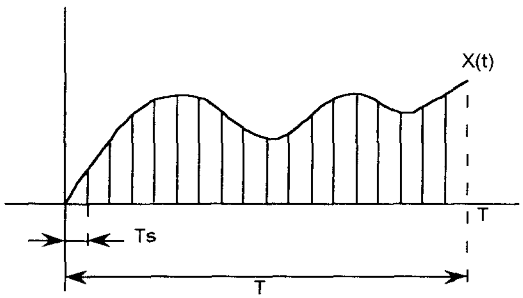

Fig. 2 Sampling the target function.

The proof of theorem 1 is shown in the following.

Proof>

Let x(t) be the target function, and y(t) be the sampled function with a priod Ts (Fig. 2), that is ,

$$y(t) = p(t) \cdot x(t)$$

where $p(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_s)$.

Then

$$Y_n = \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot F(y)^k$$

$$= \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot \sum_{k'=-\infty}^{\infty} F(p)^k \cdot F(x)^{k-k'}$$

where

$$F(p)^k = \begin{bmatrix} N/T & if\ k = N \cdot m \\ 0 & otherwise \end{bmatrix}, \quad m = \ldots,-2,-1,0,1,2,\ldots$$

Thus

$$Y_n = \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot \sum_{m=-\infty}^{\infty} \frac{N}{T} \cdot F(x)^{k-N\cdot m}$$

$$= \frac{N}{T} \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot F(x)^{k-N\cdot m}$$

$$= \frac{N}{T} \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \Phi_n^k \sum_{n'=0}^{\infty} x_{n'} \Phi_{n'}^{k-N\cdot m}$$

$$= \frac{N}{T} \sum_{m=-\infty}^{\infty} \sum_{n'=0}^{\infty} x_{n'} \sum_{k=-\infty}^{\infty} \Phi_n^k \cdot \Phi_{n'}^{k-N\cdot m}$$

We are going to choose N such that

$$\sum_{k=-\infty}^{\infty} \Phi_n^k \cdot \Phi_{n'}^{k-N\cdot m} = \sum_{k=-W_F}^{W_F} \Phi_n^k \cdot \Phi_{n'}^{k-N\cdot m} = 0 \ for\ m \neq 0,$$

and , this makes

$$Y_n = \frac{N}{T} \sum_{n'=0}^{\infty} x_{n'} \sum_{k=-\infty}^{\infty} \Phi_k^n \Phi_k^{n'} = \frac{N}{T} x_n$$

For this, the following condition is sufficient (Fig. 3, Lemma 2)

$$m \cdot N - n' \cdot W_F \;>\; n \cdot W_F$$

$$N \;>\; \frac{(n+n') \cdot W_F}{|m|} \qquad (N > 0, m \neq 0)$$

The minimum value of $|m|$ is 1, and the maximum value of n' is W_N and generally this is equal to the maximaum value of n because the maximum value of n is the order of OBLN. Thus

$$N \;>\; 2 \cdot W_F \cdot W_N$$

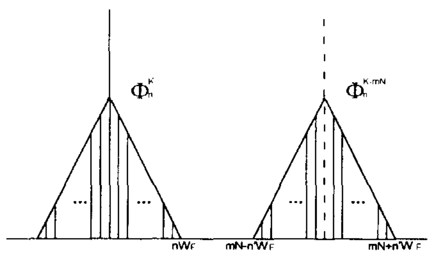

Fig. 3 Frequncy characteristic of sampled data

IV. Learning with OBLN

The learning problem is stated as follows. Given N data points $\{(x_i, d_i) \mid x_i \in R,\ d_i \in R\}$, find the value of learning parameters, so that OBLN satisfies following equation.

$$d_i = N(x_i, w) = \sum_j w_j \cdot \phi_j(x_i)$$

From the theorem of previous chapter, if the order of highest component of target function is given, we can see the number of the minimal sampling, and with this, the target function can be interpolated completely. There are two kinds of learning method. First method is using the method to find the solution of a linear problem. To find the coefficients w, one simply evaluates above equation at the given data points. The following linear system is obtained:

$$d = Aw$$

where $A_{ij} = \phi_j(x_i)$. Then the solution is

$$w^* = A^{-1} \cdot d.$$

The other learning method is related with the gradient descent methods. The iterative techniques are based on refining estimates of the optimal weight vector and usually no matrix inversions are performed. The updating rule is shown in the following. If we let the cost function be

$$E = \sum_i (d_i - N(x_i, w))^2 = \sum_i \left(d_i - \sum_j w_j \phi_j(x_i) \right)^2,$$

then the updating rule is

$$w_n^{k+1} = w_n^k - \eta \cdot \left. \frac{\partial E}{\partial w_n} \right|_{w_n = w_n^k} = w_n^k + 2 \cdot \eta \cdot \sum_i \varepsilon_i \cdot \phi_n(x_i)$$

where $\varepsilon_i = d_i - N(x_i, w) = d_i - \sum_j w_j \phi_j(x_i)$

V. Experiments

Fig[4]-Fig[10] shows the result of learning and finding the global minimum of a target function shown below with various sampling period.

$$y^e(t) = \sum_{i=1}^{9} 0.1 \cdot \overline{\phi}(t)$$

The highest order of target function can be said 9, and the OBLN can be expressed as follows.

$$y(t) = \sum_{i=0}^{9} Y_i \cdot \phi_i(t)$$

If the highest order of fourier series of the first order base function is considered as 1 approximately, then the number of minimum sampling is 19(>2*9*1).

Fig.[4]-Fig.[7] shows the result with increasing the network order. In these figure, one can see that the better OBLNlearns the target function, the more precise the global minimum given by OBLN is. Fig[8] shows the learning result with marginal sampling. Fig[9] is the result with N samples and this is under-sampled case.

Fig[10] is another under-sampled case. This shows the result of learning with 14th order OBLN rather than 9 order. This means over-structured case and, more sampling (more than 23 =(9+14)*1) is needed.

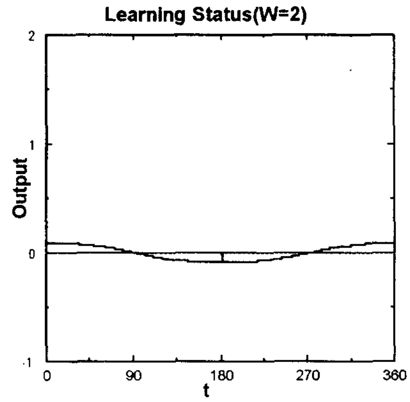

Fig. 4 Learning with small network 1and its gloval minima.

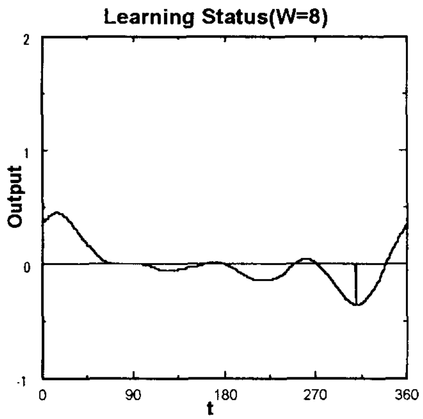

Fig. 5 Learning with small network 2 and its gloval minima.

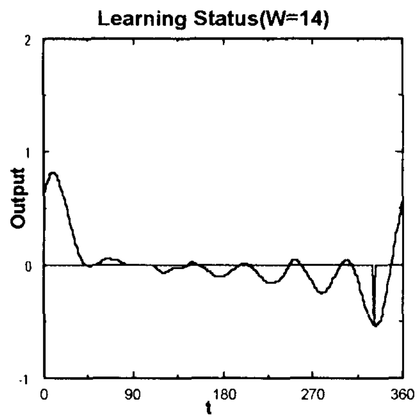

Fig. 6 Learning with small network 3 and its gloval minima.

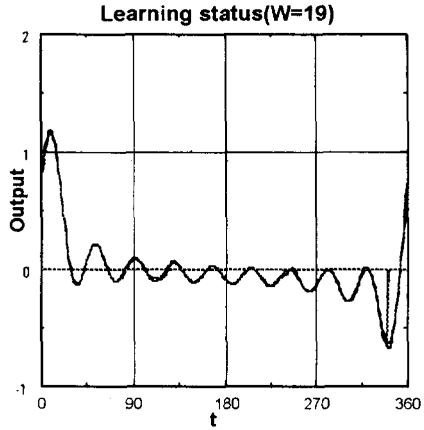

Fig. 7 Learning with full network.

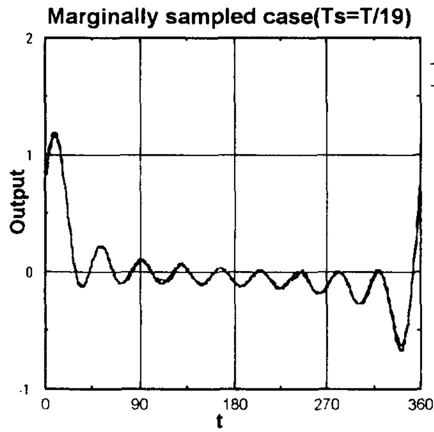

Fig. 8 Learning with marginal sampling

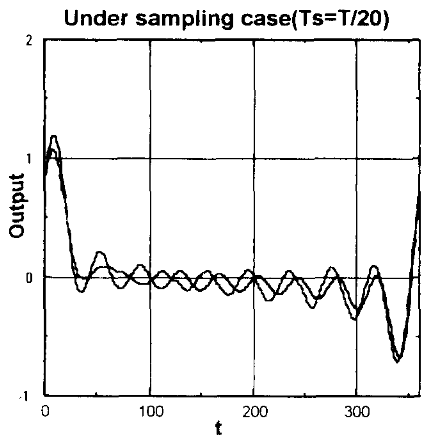

Fig. 9 Learning with under-sampling

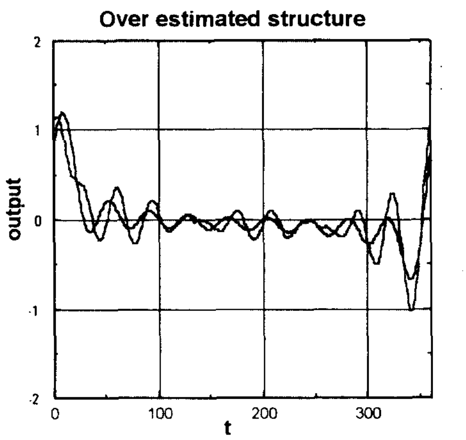

Fig. 10 Learning with over-structured network but same number of sampling

VI. Summary.

We proposed a method to find the global maximum(or minimum) of one dimensional real valued function. In this method, a learing network called OBLN, which can give its global maximum(or minimum) with simple calculation, is tought to mimic the target function and after the learning is complete, we can consider the global maximum(or minimum) of OBLN as that of target function. We also give a theorem which deal with the minimum number for complete learning. With this theorem, one can teach the OBLN and get the global optimal value without searching the whole input domain. Finally,the experimental results justify the proposed algorithm.

REFERENCE

[1] M.J.Grimble,Neurofuzzy Adaptive Modelling and Control,Prentice Hall,1994.

[2] D.G.Luenberger,Optimization by Vector Space Methods,John Wiley & Sons Inc.,1968.

[3] J.R.Hanna,Fourier Series and Intergrals of Boundary Value Problems,John Wiley & Sons Inc.,1982.

[4] M.Avriel,Nonlinear Programming-Analysis and Methods,Prentice Hall,1976.

[5] A.V.Oppenheim,Signals and Systems,Prentice Hall,1983.