Least Mean Square Error Reconstruction Learning in Recurrent Network*

Bai-ling ZHANG, Qin ZHOU, Irwin KING *and* Lei XU
Department of Computer Science
The Chinese University of Hong Kong
Shatin, N.T. Hong Kong

Abstract

In this paper, we further studied the Least Mean Square Error Reconstruction (LMSER) principle for self-organization. When generalized to recurrent networks, some phenomena similar to those in Anti-Hebbian learning and the Competitive Hebbian Learning can simultaneously emerge. Particularly, when exposured to Gaussian environment, the converged connection pattern become feature detectors. As lateral interactions widely exist among real neurons, it is reasonable to believe that the best reconstruction is a natural principle for the self-organization in the brain. Comparing with the Competitive Hebbian Learning proposed by White, the LMSER algorithm in recurrent networks avoids the crucial problem of empirically determining weight-limit. Simulations confirmed that our algorithm can catch feature structures in the correlations of input vector component, which make it applicable in a number of practical problems. As an example, we demonstrated that the LMSER in recurrent networks can successfully develop a series of orientation selective masks which can be considered as spatial filters for input information.

1 Introduction

Since the renaissance of artifical neural networks, self-organization has been intensively studied, generally on the basis of Hebb rule. For example, linear neurons learning under an unsupervised Hebbian rule can learn to perform a linear statistical analysis of the input data, as was first shown by Oja[1], who proposed a learning which finds the first principal component of the covariance matrix of input data. Since then, Oja, Xu, Sanger and many others have devised numerious neural networks which find many components of this matrix[2],[3],[4],[5]. A unified principle—Least Mean Square Error Reconstruction (LMSER) principle was proposed in [3] for a variety of PCA related learnings. Based on the LMSER principle, many generalizations relevent to PCA type learning especially their nonlinear extensions have been obtained[2].

In this paper, we further studied and developed the LMSER principle in a more broad scenario. Though the LMSER-based learning can be established for arbitary system, detailed results have not yet been discovered except for single-layer and multi-layer feedforward networks. As the real neural inputs and outputs are temporal and the actual neurobiology of visual cortex involves extensive feedback circuitry, research about the principle and results of self-organization in dynamical feedback (recurrent) networks is of more significance than in pure feedforward network from biological perspective. From the engineering viewpoint, it is also significant to investigate the self-organization in recurrent networks governed under the LMSER principle, because most of practical problems involve temporality. In the literature, preliminary results had been obtained by applying dynamics to the development of oritentation selectivity in recurrent networks which represent the first stage of cortical visual processing in mammals. In this paper, we showed that self-organization in recurrent networks under the LMSER principle can automatically yield the oritentation selectivity, a result similar to those in some Anti-Hebbian learning[6] and the Competitive Hebbian Learning[8]. From an initially random set of weights, the LMSER algorithm in recurrent networks can converge to masks which have salicut structure determined by input distribution and the output nonlinearity.

*For correspondence, please contact Qin ZHOU, Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong.

2 LMSER Learning in Recurrent Network

To apply the LMSER principle to the learning in recurrent network, consider a specific network structure to avoid ambugity, as shown in Fig.1, which consists of a layer of feedforward connection and a fully-connected recurrent inter-output connection. Such a network structure has been studied by Foldiak[6], in which Hebb and anti-Hebbian modification rule was respectively applied to the feedforward weight and feedback inhibitory weight. Anti-Hebbian learning is directly set up on the the intuitive idea of decorrelating outputs.

Let $\mathbf{W}, \mathbf{V}$ represent the feedforward and recurrent weight matrix, respectively. The network recurrent dynamics can be described by

$$\dot{\mathbf{u}} = -\mathbf{u} + \mathbf{V}\mathrm{f}(\mathbf{u}) + \mathbf{W}^{\mathrm{T}}\mathbf{x} \tag{1}$$

where $\mathbf{u} \in \mathrm{R}^n$ denote the state of output node, f is sigmoidal nonlinear function. Before we derive LMSER learning in the recurrent network, two issues should be emphasized here. First, the network dynamics are assumed to be considerably faster than the updating dynamics of the learning algorithm in the following. Such an assumption is generally taken in studying dynamical recurrent networks. Second, unlike the symmetrical recurrent connectivity requirement in Foldiak's network, there is no such a priori restrictions on $\mathbf{V}$ except that we often take zero or negative constant self-connection, i.e., $v_{ii} = 0$, or $v_{ii} = -k$, mainly from the stability consideration.

Under the attainment of steady-state conditions, Eq.(1) can be replaced by following equilibrium equation:

$$\mathbf{u}^* = \mathbf{V}\mathrm{f}(\mathbf{u}^*) + \mathbf{W}^{\mathrm{T}}\mathbf{x} \tag{2}$$

In the following discussions, we will hold same argument provided in [7]. Specifically, we assume the following conditions have already satisfied. First, the nonlinear function is a strictly monotonic increasing, differentiable and bounded above and below, i.e., sigmoidal type functions. Second, the elements of equilibrium points lie in the saturation region of the sigmoidal nonlinearities. This requirement can be met without difficulty by appropriately design the gain (slope at the origin) of the sigmoid. Third, a given equilibrium point $\mathbf{u}^*$ being analyzed is an isolated equilibrium point.

From the LMSER principle, a cost function $J(\mathbf{W})$ for best reconstruction can be written as :

$$\begin{aligned} J(\mathbf{W}) &= \frac{1}{2}\sum_{k=1}^{L}(x_k - \hat{x_k})^2 \\ &= \frac{1}{2}\sum_{k=1}^{L} e_k^2 \end{aligned} \tag{3}$$

where $\hat{\mathbf{x}}$ is the reconstruction vector from the stable state of output layer, i.e., $\hat{\mathbf{x}} = \mathbf{W}\mathbf{u}^*$. By employing a gradient descent approach to minimize J, the weight changes should take the form as

$$\begin{aligned} \Delta\mathbf{w}(\mathrm{m}) &= -\frac{\partial J(\mathbf{W})}{\partial \mathbf{w}(\mathrm{m})} \\ &= \sum_{k=1}^{L}(x_k - \hat{x}_k)\frac{\partial \hat{x}_k}{\partial \mathbf{w}(\mathrm{m})} \end{aligned} \tag{4}$$

$$\begin{aligned} \Delta\mathbf{v}(\mathrm{m}) &= -\frac{\partial J(\mathbf{W})}{\partial \mathbf{v}(\mathrm{m})} \\ &= \sum_{k=1}^{L}(x_k - \hat{x}_k)\frac{\partial \hat{x}_k}{\partial \mathbf{v}(\mathrm{m})} \end{aligned} \tag{5}$$

From the equilibrium equation Eq.(2) we can get

$$\frac{\partial \mathbf{u}^*}{\partial \mathrm{w_j(m)}} = (I - \mathrm{VG})^{-1}[\underbrace{0,\cdots,0}_{\text{m-1}}, \mathrm{x_j}, 0, \cdots, 0]^{\mathrm{T}} \tag{6}$$

$$\frac{\partial \mathbf{u}^*}{\partial \mathrm{v_i(m)}} = (I - \mathrm{VG})^{-1}[\underbrace{0,\cdots,0}_{\text{m-1}}, \mathrm{f(u_i^*)}, 0, \cdots, 0]^{\mathrm{T}} \tag{7}$$

where $G = diag[f'(u_1^*), f'(u_2^*), \cdots, f'(u_M^*)], f'(u_i^*) = \frac{\partial f(u_i)}{\partial u_i}\mid_{u_i=u_i^*}$. It is assumed here that $I - \mathrm{VG}$ is a nonsingular matrix, which is generally satisfied as the derivative of sigmoidal function is very small. Then from $\hat{\mathrm{x}} = \mathbf{W}\mathbf{u}^*$, Eq.(4) and (5) are directly available.

It is clear that matrix inversion has appeared and this is usually cubersome. Fortunately, following the same argument as in [7], in recurrent networks, it is often desirable and feasible to design the equilibrium points in the saturation region of the sigmoidal nonlinear function. From this consideration, we can appropriately adjust the sigmoidal functions with large gains, thus making the elements of diagonal matrix G very small. Under this condition, we can simplify the learning algorithm (4),(5) as:

$$\begin{aligned} \Delta \mathbf{w}(\mathrm{m}) &= \mu_1(u_m^* \mathbf{e} + \mathbf{e}^{\mathrm{T}}\mathbf{w}(\mathrm{m})\mathbf{x}) && (8) \\ \Delta \mathbf{v}(\mathrm{m}) &= \mu_2 \mathbf{e}^{\mathrm{T}}\mathbf{w}(\mathrm{m})\mathrm{f}(\mathbf{u}^*) && (9) \end{aligned}$$

Or writing it in matrix form

$$\begin{aligned} \Delta \mathbf{W} &= \mu_1(\mathbf{e}\mathbf{u}^* + \mathbf{x}(\mathbf{e}^{\mathrm{T}}\mathbf{W})) && (10) \\ \Delta \mathbf{V} &= \mu_2 f(\mathbf{u}^*)(\mathbf{e}^{\mathrm{T}}\mathbf{W}) && (11) \end{aligned}$$

It is worthy to mention that the above discussion is based on a reconstruction scheme $\hat{\mathrm{x}} = \mathbf{W}\mathbf{u}^*$. Apparently, another scheme may be considered, that is $\hat{\mathrm{x}} = \mathbf{W}\mathrm{f}(\mathbf{u}^*)$. In this case, the above derivation can be directly extended and following results are obtained

$$\begin{aligned} \Delta \mathbf{w}(\mathrm{m}) &= \mu_1(f(u_m^*)\mathbf{e} + \mathbf{e}^{\mathrm{T}}\mathbf{w}(\mathrm{m})\mathbf{x}\mathrm{f}'(\mathrm{u_m^*})) && (12) \\ \Delta \mathbf{v}(\mathrm{m}) &= \mu_2 \mathbf{e}^{\mathrm{T}}\mathbf{w}(\mathrm{m})\mathrm{f}'(\mathbf{u}^*)\mathrm{f}(\mathrm{u_m^*}) && (13) \end{aligned}$$

It is easy to find that, as most elements of equilibrium points are forced to locate in the saturation region of sigmoidal nonlinearity, Eq.(13) and the second term in Eq.(12) will be near zero, therefore its performance will be quite different. As can be expected, more interesting results should be induced from (10) and (11) comparing with (12) and (13). From the space limitation consideration, we will mainly consider (10),(11) in this paper while leave more detailed discussion in other places.

In the simple learning rules (10),(11) , both feedforward connections $\mathbf{W}$ and recurrent connections $\mathbf{V}$ are adapted by the reconstruction error $\mathbf{e}$, which is the only available information in the network. Here no artificially designed anti-Hebb term is necessary for the weight development. As a result of the competition among the equilibrium points, the network outputs have a tendency to become distinct. In other words, decorrelation or independency in some sense will be obtained. In our experiment we have found that under the LMSER principle lateral connections in the recurrent network will be developed to be inhibitory, or most of them inhibitory. This means that the LMSER learning can force the output to be decorrelated. When presented with input consisting of 2-dimensional Gaussian signals, the feed-forward weight in recurrent network learned from LMSER all converged to masks which have salient structure determined by input distribution and the output nonlinearity, from an initially random set of weights.

About the self-organization capability in recurrent networks under the LMSER principle, many theoretical analysis remain to be taken. The most interesting aspect is the symmetry-breaking phenomena caused by output nonlinearity. As pointed out in [3], it is the sigmoidal type nonlinearity that makes the output units selective or sensitive to the features in input data. Instead of theoretical analysis which seems difficult at the moment, we'll turn our attention to experiment in the following section.

3 Simulations

As a demonstration of LMSER and DLMSER, we take the problem of learning to respond to randomly placed Gausian-shaped spots. The data generation scheme and traing was similar to that used by White in the Competitive Hebbian Learning. In the experiment, there were k^2 inputs units, treated as a $k \times k$ square array. Each input vector was a random located Gaussian spot, with center at arbitary position except that they must be two input units away from the nearest edge in the input array. The strength of each component of an input vector was determined by how far that input unit was from the corresponding spot center. The Gaussian spot for the following demonstrations was given by $e^{-(r-r_0)^2}$, with r_0 being the center location. The first motivation of this test is identical to that in Competitive Hebbian Learning, i.e., a self-organization principle should learn to respond to the structure present in the set of input vectors. With random Gaussian spots, the only structure is the two-dimensional structure of the input array, along with the weakness of the training vectors near the edges of the array. As been experimently confirmed by White, a single Hebbian node can learn just this, an essentially symmetric response which is strong in the center of the input array, and weak around the edges. For a multi-nodes network, a self-organization rule should learn to share the input space and develop distinct regions of strong response. Therefore, in the simulations, we are mainly interested in cases of more than one output nodes.

We tested the feature detection problem for the LMSER algorithm in recurrent network. The output nonlinearity is sigmoidal function $f(t) = tanh(\beta t), \beta = 3$. 100 input units with 10×10 square array was tested. The average brightness of 1000 Gaussian spot was calculated beforehand and then substracted from each random Gaussian spot during training. Initial weights were set to small random values. When presented with 10×10 array with three output nodes, the converged feedforward weight patterns were displayed in Fig.2(a)$\sim$(c). When the output nodes increase to 16, we use 16×16 input array which can be expanded into a vector of 256 components. The random Gaussian spot was generated in same manner as in previous experiment. With such two dimensional Gaussian input, the converged two dimensional masks were illustrated in Fig.3$\sim$4, corresponding to different sigmoidal nonlinearity $\beta = 0.5$ and $\beta = 1$, respectively. Comparing with the PCA masks studied by many workers, the connection pattern developed from LMSER in recurrent network is much less ordered. For example, some edge extraction masks has not appeared. These masks are not much reminiscent of the two dimensional Gabor filters described by Daugman. Although many features of biological vision may not yet be well predicated by our experiment results, their highly structured self-organized patterns is still very interesting, particularly from the viewpoint of feature extraction in pattern recognition.

4 Conclusions

In this paper, we studied the LMSER principle govered weight development in specified recurrent network. As feedback circuitry is considered to be ubiquitous in the brain, it is of particularly importance to study self-organization in dynamical recurrent networks from the biological viewpoint. It was discovered that each output unit in recurrent networks can adapt under LMSER rule to a changing environment and give useful response to input information, particularly, they can become detectors of mutually independent features contained in the presented patterns. As receptive fields are probably the most prominent and important computational mechanism employed by biological information processing system, these developed features detectors may be considered as an approximation to receptive fields at some stage in primary visual cortex.

References

[1] E.Oja, "Neural Networks, Principal Components, and Subspace," *International Journal of Neural Systems,*, Vol.1, pp. 61-68, 1989

[2] Lei Xu, " Theories for Unsupervised Learning: PCA and Its Nonlinear Extensions," *III World Congress on Computational Intelligence, Orlando, Florida*, Vol.II, pp.1252-1257, 1994

[3] Lei Xu, " Least Mean Square Error Reconstruction Principle for Self-Organizing Neural-Nets," *Neural Networks*, Vol.6, pp. 627-648, 1993.

[4] Lei Xu, " Self-Organizing Rules for Robust Principle Component Analysis," *Advances in Neural Information Processing Systems 5*, pp. 467-474, 1993.

[5] Terence D.Sanger, "Optimal Unsupervised Learning in a Single-Layer Liner Feedforward Neural Network," *Neural Networks*, Vol.2, pp. 459-473, 1989

[6] P.Foldiak, " Forming Sparse Representationsby Local Anti-Hebbian Learning," *Biological Cybernetics,* , Vol.64, pp.165-170, June 1990

[7] Subramania I.Sudharsanan, " Equilibrium Characterization of dynamical neural Networks and a Systematic Synthesis Procedure for Associative Memories," *IEEE Trans. Neural Networks*, Vol.2, pp. 509-521, 1991.

[8] Ray H. White, " Competitive Hebbian Learning: Algorithm and Demonstrations," *IEEE Trans. Neural Networks*, Vol.5, pp. 261-275, 1992.

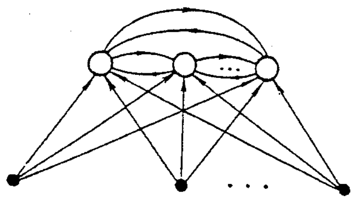

Fig.1

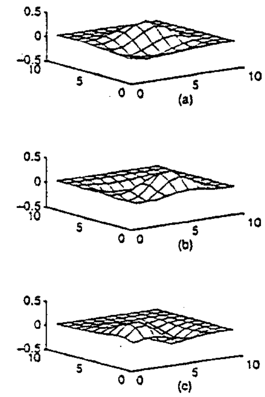

Fig.2

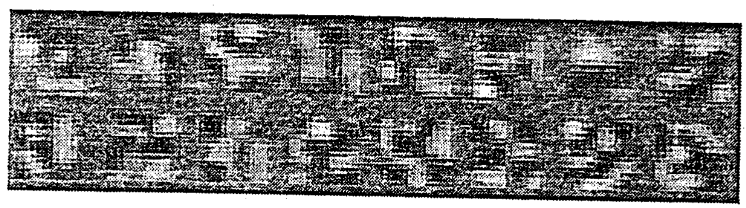

Fig.3

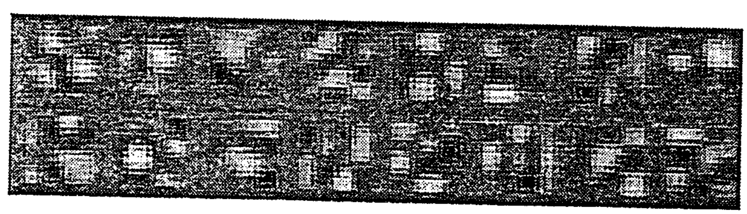

Fig.4

Fig.1 A recurrent network

Fig.2 Learning result from the LMSER (a)~(c)

Fig.3 The converged two dimensional masks with $\beta = 0.5$

Fig.4 The converged two dimensional masks with $\beta = 1$