J. Math. Biology (1982) 15: 267–273

Journal of
Mathematical
Biology

A Simplified Neuron Model as a Principal Component Analyzer

Erkki Oja

University of Kuopio, Institute of Mathematics, 70100 Kuopio 10, Finland

Abstract. A simple linear neuron model with constrained Hebbian-type synaptic modification is analyzed and a new class of unconstrained learning rules is derived. It is shown that the model neuron tends to extract the principal component from a stationary input vector sequence.

Key words: Neuron models – Synaptic plasticity – Stochastic approximation

1. Introduction

In neuron models of the last decades since the work of McCulloch and Pitts (1943), many roles have been assigned to individual neurons from computational machines to analog signal processors. In many recent models, the feature detecting function has been emphasized (Cooper et al., 1979; von der Malsburg, 1973; Nass and Cooper, 1973; Perez et al., 1975; Takeuchi and Amari, 1979). Most of these works are strongly based on computer simulations. The present correspondence points out a mathematical finding related to the feature detecting role of model neurons: a new synaptic modification law, derived as a limit process from an earlier well-known formulation of the Hebbian-type modification, leads to a behaviour where the unit is able to extract from its input the statistically most significant factor. The behaviour is in close correspondence with a statistical technique known as principal component analysis or Karhunen-Loève feature extraction.

The neuron model considered here is as follows. The neuron receives a set of n scalar-valued inputs $\xi_1, \ldots, \xi_n$ (which may be assumed to represent firing frequencies in presynaptic fibers; in some models, the zero level is defined so that negative values for the effective inputs become possible) through n synaptic junctions with coupling strengths $\mu_1, \ldots, \mu_n$. The unit sends out an efferent signal η. According to many models of neural networks, the input-output relationship is linearized to read

$$\eta = \sum_{i=1}^{n} \mu_i \xi_i. \tag{1}$$

In most recent models, the junction strengths μ_i have been assumed variable in time according to some version of the Hebbian hypothesis: the efficacies grow stronger when both the pre- and postsynaptic signals are strong. (For a discussion of the

0303–6812/82/0015/0267/$01.40

linear law (1) as well as the synaptic modification, see Kohonen et al. (1981).) However, as the basic Hebbian scheme would lead to unrealistic growth of the efficacies, a saturation or normalization has usually been assumed. This leads typically to the following type of "learning equation":

$$\mu_i(t+1) = \frac{\mu_i(t) + \gamma\eta(t)\xi_i(t)}{\{\sum_{i=1}^{n} [\mu_i(t) + \gamma\eta(t)\xi_i(t)]^2\}^{1/2}}, \tag{2}$$

where γ is a positive scalar.

The form of the denominator is due to the use of Euclidean vector norm. Now the sum of the squares of $\mu_i(t)$ remains equal to one. This particular form of normalization is very convenient from a mathematical point of view. The actual physiological reason for a normalization would be the competition of the synapses of a given neuron over some limited resource factors that are essential in efficacy growth (e.g., number of receptor molecules, surface area of the postsynaptic membrane, or energy resources). It should be realized that the synaptic efficacies $\mu_i(t)$ need not be linearly related to such resource factors and thus in the denominator of Eq. (2) there is no reason to prefer e.g. the linear sum to some other form like the one suggested there. In Sec. 4, a mathematical analysis is made of a more general class of learning equations with normalization, of which (2) is an example.

In the following we show that both Eq. (2) and another learning scheme, derivable from it by a limit process but also plausible in physiological terms, produces a very specific type of behaviour for the model neuron: if the input vectors $[\xi_1(t), \ldots, \xi_n(t)]^T$ for $t = 1, 2, \ldots$ are regarded as a vector-valued stochastic process, then the neural unit tends to become a *principal component analyzer* for the input process.

2. A New Law of Synaptic Modification

Assume that the gain or plasticity coefficient γ is small. Then (2) can be expanded as a power series in γ (for details, see Sec. 4), yielding

$$\mu_i(t+1) = \mu_i(t) + \gamma\eta(t)[\xi_i(t) - \eta(t)\mu_i(t)] + O(\gamma^2). \tag{3}$$

Neglecting the $O(\gamma^2)$ term, proportional to the second and higher powers of γ, we have obtained a new interesting learning scheme. On the right hand side of (3), $\gamma\eta(t)\xi_i(t)$ represents the usual "Hebbian" increment. However, with the extra term, the increment $\mu_i(t+1) - \mu_i(t)$ becomes now $\gamma\eta(t)\xi_i'(t)$, with $\xi_i'(t) = \xi_i(t) - \eta(t)\mu_i(t)$ the *effective input* to the unit. Each junction strength $\mu_i(t)$ tends to grow according to its afferent $\xi_i(t)$, but the growth is controlled by an internal feedback in the neuron, $-\eta(t)\mu_i(t)$. This term is related to the "forgetting" or leakage terms often used in learning rules, with a difference that the leakage becomes stronger with stronger response $\eta(t)$. In view of some neurobiological models of synaptic growth, in which the efficacy of a synapse is explained in terms of postsynaptic factors, e.g. relative amounts of receptor molecules at postsynaptic sites (Stent, 1973), this kind of control factor may not be unrealistic.

An interesting thing about (3) is that due to this control factor, $\sum_{i=1}^{n} \mu_i(t)^2$ tends to be bounded and close to one even though no explicit normalization appears in the equation. Note that the right hand side depends on the other $\mu_j(t), j \neq i$, only through the common response $\eta(t)$, if the term $O(\gamma^2)$ is neglected.

3. Asymptotic Analysis

We introduce vector notation by writing $m(t) = [\mu_1(t), \ldots, \mu_n(t)]^T$ and $x(t) = [\xi_1(t), \ldots, \xi_n(t)]^T$, whereby $m(t)$ and $x(t)$ are time-dependent n-dimensional real column vectors. Both are assumed to be stochastic. Now (1) reads

$$\eta(t) = m(t)^T x(t) \tag{4}$$

while (2) reads

$$m(t+1) = [m(t) + \gamma\eta(t)x(t)]/\|m(t) + \gamma\eta(t)x(t)\| \tag{5}$$

and (3) becomes

$$\begin{aligned} m(t+1) &= m(t) + \gamma\eta(t)[x(t) - \eta(t)m(t)] + O(\gamma^2) \\ &= m(t) + \gamma[x(t)x(t)^T m(t) - m(t)^T x(t)x(t)^T m(t)m(t)] + O(\gamma^2). \end{aligned} \tag{6}$$

If $x(t)$ and $m(t)$ are statistically independent, (6) can be averaged to read

$$E\{m(t+1)|m(t)\} = m(t) + \gamma[Cm(t) - (m(t)^T Cm(t))m(t)] + O(\gamma^2) \tag{7}$$

with $C = E\{x(t)x(t)^T\}$.

Recent techniques of stochastic approximation theory are available for analyzing learning equations of the type given here (see Ljung, 1977 and Kushner and Clark, 1978). Without going rigorously into the details, it can be shown that if the distribution of $x(t)$ satisfies some not unrealistic assumptions and the gain γ is not constant but allowed to decrease to zero in a special way (e.g., proportionally to $1/t$), then both (5) and (6) can be approximated by a differential equation

$$\frac{d}{dt} z(t) = Cz(t) - (z(t)^T Cz(t))z(t) \tag{8}$$

which is the continuous-time counterpart of (7). The approximation is in the sense that the asymptotic paths of Eq. (8) and the stochastic equations (5) and (6) are close with a large probability and eventually the solution $m(t)$ of (5) and (6) tends (with probability one) to a uniformly asymptotically stable solution of (8) (see e.g. Theorem 2.3.1 of Kushner and Clark, 1978).

Of course, Eq. (8) might also be taken as the synaptic learning equation directly, with $z(t) = [\mu_1(t), \ldots, \mu_n(t)]^T$, if a continuous-time framework is used from the start. This contains the assumption that synaptic modification is slow compared to statistical variations in the input data in order to warrant the use of $E\{x(t)x(t)^T\}$ in (8).

A similar algorithm, arising in the context of digital signal processing and numerical methods of mathematical statistics, has been analyzed in detail elsewhere by Oja and Karhunen (1981). The following can be proven:

Theorem. *In* (8), *let* C *be positive semidefinite with the largest eigenvalue of multiplicity one, and let* c *be the corresponding normalized eigenvector (either of the two possible choices). Then if* $z(0)^T c > 0$ (< 0), $z(t)$ *tends to* c $(-c)$ *as* $t \to \infty$. *The points* c *and* $-c$ *are uniformly asymptotically (exponentially) stable.*

Proof. The uniform asymptotic stability is a direct consequence of standard results (see Theorem 2.4 of Hale, 1969). The domains of attraction of the stable points can be determined by expanding $z(t)$ in terms of eigenvectors of C,

$$z(t) = \sum_{i=1}^{n} \zeta_i(t) c_i, \qquad \text{with} \qquad c_1 = c,$$

by defining $\theta_i(t) = \zeta_i(t)/\zeta_1(t)$ and by deriving the linear differential equation

$$\dot{\theta}_i(t) = (\lambda_i - \lambda_1)\theta_i(t)$$

with λ_i and λ_1 eigenvalues of C. A detailed proof is provided in Oja and Karhunen (1981).

Even when γ does not tend to zero as time grows but remains small, Eq. (8) is a good approximation to the averaged equation (7). We can conclude that, neglecting statistical fluctuations, the synaptic vector $m(t)$ of either (5) or (6) will tend to the dominant eigenvector c of input correlation matrix C. In statistical literature, the normalized linear combination of the data components having maximum variance is called the principal component of the data (Anderson, 1958). For zero-mean data, the principal component is exactly $c^T x$, or the response of the model neuron after convergence. Even for nonzero means for the components of $x(t)$, this linear combination has the largest quadratic mean, since $E\{\eta(t)^2\} = E\{(m(t)^T x(t))^2\}$ $= m(t)^T C m(t)$ is maximized when $m(t) = c$. So the magnitude of the response $\eta(t)$ tends to be maximal on the average when the input vector belongs to the same statistical sample as the input vectors occurring during the training.

This becomes especially clear if we assume that for all t, $x(t) = x + v(t)$ where x is a fixed unit vector and $v(t)$ is symmetrically distributed zero-mean noise. Then $C = xx^T + E\{vv^T\} = xx^T + \sigma^2 I$ with σ^2 the variance of the components of $v(t)$. The largest eigenvalue of C is $1 + \sigma^2$ and the corresponding eigenvector is $c = x$. The model neuron then becomes a *matched filter* for the data, since $\lim_{t\to\infty} z(t) = x$ in (8).

4. A General Approach to Analyzing Learning Rules with Constraints

The analysis presented above is a special case of an approach that may have wider applicability. Assume that the junction strength $\mu_i(t)$ varies according to

$$\mu_i(t+1) = \frac{\tilde{\mu}_i(t+1)}{\omega[\tilde{\mu}_1(t+1), \ldots, \tilde{\mu}_n(t+1)]} \tag{9}$$

with

$$\tilde{\mu}_i(t+1) = \mu_i(t) + \gamma\varphi[\mu_i(t), \xi_i(t), \eta(t)]. \tag{10}$$

There $\varphi[\cdot]$ is the basic increment at step t; however, the new values $\tilde{\mu}_i(t+1)$ must be normalized according to some constraint function $\omega[\cdot]$ to obtain the $\mu_i(t+1)$.

Eq. (2) was a special case with

$$\varphi[\mu_i(t), \xi_i(t), \eta(t)] = \xi_i(t)\eta(t), \tag{11}$$

$$\omega[\tilde{\mu}_1(t+1), \ldots, \tilde{\mu}_n(t+1)] = \left[\sum_{i=1}^{n} \tilde{\mu}_i(t+1)^2\right]^{1/2}. \tag{12}$$

Now the ω function satisfies

$$\omega[\delta\tilde{\mu}_1(t+1), \ldots, \delta\tilde{\mu}_n(t+1)] = \delta\omega[\tilde{\mu}_1(t+1), \ldots, \tilde{\mu}_n(t+1)]$$
$$\text{for any scalar } \delta. \tag{13}$$

Let us assume that (13) holds in general for ω. Obviously it is true for any function of the form $[\sum \tilde{\mu}_i(t+1)^p]^{1/p}$.

Eqs. (9) and (10) may be difficult to analyze due to the normalization. For small gain factor γ, an equivalent form is again obtained, which would easily result in a limiting differential equation. Observe first that, from (9) and (13),

$$\omega(\mu_1, \ldots, \mu_n) = \omega\left[\frac{\tilde{\mu}_1}{\omega(\tilde{\mu}_1, \ldots, \tilde{\mu}_n)}, \ldots, \frac{\tilde{\mu}_n}{\omega(\tilde{\mu}_1, \ldots, \tilde{\mu}_n)}\right]$$
$$= \frac{1}{\omega(\tilde{\mu}_1, \ldots, \tilde{\mu}_n)}\omega(\tilde{\mu}_1, \ldots, \tilde{\mu}_n) = 1 \tag{14}$$

which holds for all t. This is the explicit constraint on the junction strengths $\mu_i(t)$. Then we obtain

$$\begin{aligned}
&\omega[\tilde{\mu}_1(t+1), \ldots, \tilde{\mu}_n(t+1)] \\
&\quad = \omega[\mu_1(t) + \gamma\varphi(\mu_1(t), \xi_1(t), \eta(t)), \ldots, \mu_n(t) + \gamma\varphi(\mu_n(t), \xi_n(t), \eta(t))] \\
&\quad = \omega[\mu_1(t), \ldots, \mu_n(t)] + \gamma\frac{\partial\omega}{\partial\gamma}\bigg|_{\gamma=0} + O(\gamma^2) \\
&\quad = 1 + \gamma\frac{\partial\omega}{\partial\gamma}\bigg|_{\gamma=0} + O(\gamma^2),
\end{aligned}$$

and (9) yields for γ small

$$\mu_i(t+1) = \tilde{\mu}_i(t+1) - \gamma\tilde{\mu}_i(t+1)\frac{\partial\omega}{\partial\gamma}\bigg|_{\gamma=0} + O(\gamma^2)$$
$$= \mu_i(t) + \gamma\varphi[\mu_i(t), \xi_i(t), \eta(t)] - \gamma\mu_i(t)\frac{\partial\omega}{\partial\gamma}\bigg|_{\gamma=0} + O(\gamma^2). \tag{15}$$

This is the desired approximative form. The first two terms on the right are exactly the same as the right hand side of (10), and the third term reflects the effect of normalization.

In the example (11) and (12),

$$\left.\frac{\partial \omega}{\partial \gamma}\right|_{\gamma=0} = \frac{\partial}{\partial \gamma}\left\{\sum_{i=1}^{n}[\mu_i(t) + \gamma\xi_i(t)\eta(t)]^2\right\}^{1/2}\Bigg|_{\gamma=0}$$

$$= \sum_{i=1}^{n} \mu_i(t)\xi_i(t)\eta(t) = \eta(t)^2,$$

resulting in (3).

5. Discussion

The simple neuron model was considered above without any reference to its role in a nervous system. When a neural network is composed of such units, there will by necessity be lateral connections between the units. These were not included in the above analysis. In many model studies using related elements, inhibitory lateral connections have been assumed (Cooper et al., 1979; von der Malsburg, 1973; Nass and Cooper, 1975; Perez et al., 1975; Takeuchi and Amari, 1979) which have an effect of enhancing selectivity to the incoming patterns. Recently, Kohonen (1982) has introduced a model with an array of interconnected elements having some of the properties of the present model. He has shown how self-organization is possible in such a network, with the result that topological properties of the input space are inherited by the response distribution of the neural elements.

Acknowledgement. The author wishes to thank Professor T. Kohonen for many stimulating discussions on the subject of the paper.

References

Anderson, T. W.: An introduction to multivariate statistical analysis. New York: Wiley 1958

Cooper, L. N., Liberman, F., Oja, E.: A theory for the acquisition and loss of neuron specificity in visual cortex. Biol. Cyb. **33**, 9–28 (1979)

Hale, J. K.: Ordinary differential equations. New York: Wiley 1969

Kohonen, T.: Self-organized formation of topologically correct feature maps. Biol. Cyb. **43**, 59–69 (1982)

Kohonen, T., Lehtiö, P., Oja, E.: Storage and processing of information in distributed associative memory systems. In: Hinton, G., Anderson, J. A. (eds.). Parallel models of associative memory, pp. 105–143. Hillsdale: Erlbaum 1981

Kushner, H., Clark, D.: Stochastic approximation methods for constrained and unconstrained systems. New York: Springer 1978

Ljung, L.: Analysis of recursive stochastic algorithms. IEEE Trans. Automatic Control **AC-22**, 551–575 (1977)

von der Malsburg, C.: Self-organization of orientation sensitive cells in the striate cortex. Kybernetik **14**, 85–100 (1973)

McCulloch, W. S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophysics **5**, 115–133 (1943)

Nass, M., Cooper, L. N.: A theory for the development of feature detecting cells in visual cortex. Biol. Cyb. **19**, 1–18 (1975)

Oja, E., Karhunen, J.: On stochastic approximation of eigenvectors and eigenvalues of the expectation of a random matrix. Helsinki University of Technology, Report TKK-F-A458 (1981)

Perez, R., Glass, L., Shlaer, R. J.: Development of specificity in the cat visual cortex. J. Math. Biol. **1**, 275–288 (1975)

Stent, G. S.: A psychological mechanism for Hebb's postulate of learning. Proc. Natl. Acad. Sciences **70**, 997–1001 (1973)

Takeuchi, A., Amari, S.: Formation of topographic maps and columnar microstructures in nerve fields. Biol. Cyb. **35**, 63–72 (1979)

Received August 21, 1981/Revised June 21, 1982