# Dynamical Behavior of Autoassociative Memory Performing Novelty Filtering

*Hanseok Ko

※This research was funded by the University Basic Research Program of the Ministry of Information and Communication.


## Abstract

This paper concerns the dynamical behavior, in probabilistic sense, of a feedforward neural network performing auto-association for novelty filtering. Networks of retinotopic topology having a one-to-one correspondence between input and output units can be readily trained using back-propagation algorithm, to perform autoassociative mappings. A novelty filter is obtained by subtracting the network output from the input vector. Then the presentation of a "familiar" pattern tends to evoke a null response; but any anomalous component is enhanced. Such a behavior exhibits a promising feature for enhancement of weak signals in additive noise. As an analysis of the novelty filtering, this paper shows that the probability density function of the weight converges to Gaussian when the input time series is statistically characterized by nonsymmetrical probability density functions. After output units are locally linearized, the recursive relation for updating the weight of the neural network is converted into a first-order random differential equation. Based on this equation it is shown that the probability density function of the weight satisfies the Fokker-Planck equation. By solving the Fokker-Planck equation, it is found that the weight is Gaussian distributed with time dependent mean and variance.


## I. Introduction

This paper concerns the probabilistic behavior of a novelty filter which employs an artificial neural network performing autoassociative task. Our novelty filter is based on a feedforward network featuring a one-to-one correspondence between input and output nodes and a single hidden layer of formally identical nonlinear units. A novelty filter is obtained by subtracting the network output from the input vector [1]. Then the presentation of a "familiar" pattern tends to evoke a null response; but any anomalous component is enhanced. This network topology, referred as "retinotopic" and "screen-like", has been applied with some success to problems of data compression where the comparatively narrow hidden layer encodes the input as a vector of activation values, which is readily decoded by forming its product with the back (output) end weight matrix. This technique utilizes a set of weights to adapt the received input data to some desired response. In this case, the desired responses are the blocks of input data, thus performing the auto-association task. Working in auto-association task, the novelty filter is rigorously investigated as a promising candidate to perform signal enhancement or time series prediction. This paper presents an analysis of the neural network's behavior, focusing on the weight changes. In particular, the probability density function (p.d.f) of the weight during training, as they change with time, was observed to be converging into Gaussian p.d.f. Our aim is to show analytically that the probability density function of the weight converges to Gaussian when the input time series is statistically characterized by nonsymmetrical probability density functions. In the structurally similar setting, Jacyna and Nguyen [2] have shown that the probability density function of the weight converges to Gaussian when the input time series is Gaussian distributed. The present paper extends this result and shows that the p.d.f of the weight converges to Gaussian even when the input time series is statistically characterized in non-symmetrical p.d.f. We first show the artificial neural network model as the building block for our novelty filter. Then, the recursive relation for updating the weight of the neural network is converted into a first-order random differential equation. Based on this equation, the probability density function of the weight satisfies the Fokker-Planck equation. By solving the Fokker-Planck equation, the weight is shown as Gaussian distributed with time-dependent mean and variance.

## II. Artificial Neural Network Model

Consider a stochastic phenomenon that is characterized by two random processes $x(t)$ and $y(t)$. The process $y(t)$, the output of the hidden units, is observed at time $t$ in response to the process applied as input. The relation-

* Department of Electronics Engineering Korea University
Manuscript Received: April 22, 1998.

ship between $x(t)$ and $y(t)$ is observed to be nonlinear with a transfer function $F(t)$ but can be approximated as closely as desired by the linear part of its power series expansion, i.e.,

$$y(t) \approx \eta_0 + \eta_1 w(t) x(t) \tag{1}$$

For $w(t)x(t)$ small with nonzero $\eta_1$. Here, $w(t)$ is a weight function determined by an adaptive control algorithm and a deterministic signal, $r(t)$, is called the desired response. For the asymmetric sigmoid, $F(t) = 1/(1+e^{w(t)x(t)})$, this gives $\eta_0 = 1/2$ and $\eta_1 = 1/4$; whereas for the symmetric sigmoid, $F(t) = (1-e^{w(t)x(t)})/(1+e^{w(t)x(t)})$, one has $\eta_0 = 0$ and $\eta_1 = 1/2$. For simplicity, we will use the asymmetric sigmoid throughout the analysis.

The desired response $r(t)$ provides a target for adjusting the weight function $w(t)$. By comparing $y(t)$, an estimate of the desired response $r(t)$, with the actual value of the desired response $r(t)$, an estimation error is produced. Denoting this error by $e(t)$, we have:

$$e(t) = r(t) - y(t) = r(t) - [\eta_0 + \eta_1 w(t) x(t)] \tag{2}$$

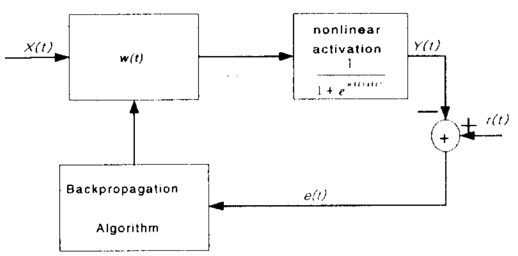


Figure 1. Neural Network (Nonlinear Adaptive Filter) Model.

where the weight function $w(t)$ is adjusted continuously to keep the mean squared error as small as possible. A signal-flow graph of this model is shown in Figure 1. In general, both adaptive filters [3] and Neural Networks [4] are built around this model.

Various algorithms have been developed for implementing the adaptive process. In this paper, a widely used algorithm known as the backpropagation(BP) algorithm [5] is used to adjust the weight function $w(t)$. According to this algorithm, the weight function for time instant $(n+1)T$ (where $T$ is the sampling time interval) is determined by:

$$w[(n+1)T] = w(nT) + \mu T y' x(nT) e(nT) \tag{3}$$

where

$$y' = \frac{\partial y(z)}{\partial z}\bigg|_{z = x(t)w(t)} = \eta_1$$

and where $\mu$ is a small fixed constant that governs the stability and rate of the algorithm. The first order derivative $y'(z) = \eta_1$ is obtained by the linear approximation made in Equation (1). It follows that:

$$w[(n+1)T] = w(nT) + \mu\eta_1 T x(nT) e(nT) \tag{4}$$

Letting $T \to 0$, the weight function $w(t)$ is a random process determined by the following equation:

$$\frac{dw(t)}{dt} = \mu\eta_1 r(t) x(t) - \mu\eta_0\eta_1 x(t) - \mu\eta_1^2 w(t) x^2(t) \tag{5}$$

Now assume that the desired response $r(t)$ is a constant, i.e., $r(t) = r$, since the corresponding target for various training patterns presented at the input remains the same at each epoch. Then, Equation (5) becomes:

$$\begin{aligned} \frac{dw(t)}{dt} &= \mu\eta_1(r-\eta_0)x(t) - \mu\eta_1^2 w(t) x^2(t) \\ &= \mu\beta_0 x(t) - \mu\beta_1 w(t) x^2(t) \end{aligned} \tag{6}$$

where $\beta_0$ and $\beta_1$ replace the constants, $\eta_1(r-\eta_0)$ and $\eta_1^2$, respectively. This implies that $w(t)$ is a continuous Markov process because $\frac{dw(t)}{dt}$, the instantaneous change in $w(t)$, depends on the present value of $w(t)$ and the present value of the input process $x(t)$. Therefore, at any given time $t$, the probability density function $p(w;t)$ of $w(t)$ satisfies the following partial differential equation [6][7]:

$$\frac{\partial p(w;t)}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial w^n} [Q_n(w)\, p(w;t)] \tag{7}$$

where $Q_n(w)$, the n-th order derivative moment of $w(t)$, is defined by;

$$Q_n(w) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E\{\int_t^{t+\Delta t} [\frac{dw(u)}{du} du]^n |_{w(t)=w} du\} \tag{8}$$

If $Q_n(w) = 0$ for $n \geq 3$, the partial differential equation in Equation (7) is generally called the Fokker-Planck equation. In the next section, $Q_1(w)$ and $Q_2(w)$ are computed for the case of input time series whose statistics is characterized with a nonsymmetric p.d.f. In particular, the p.d.f. is assumed to be in the form of a skewed-Gaussian described as follows :

$$n_{ns}(t) = n(t) + \varepsilon n^2(t)$$

where $n(t)$ is Gaussian. The two types of probability density function are shown in Figure 2. It was shown [2] that $Q_n$ is approximately zero for $n \geq 3$ when $\varepsilon$ is small. This means that changes in the process occur slowly enough (small $\mu$) that moments higher than the second vanish more rapidly than $\triangle t$ as the latter approaches zero. Hence, to a good approximation, the p.d.f. $p(w;t)$ of $u(t)$ with such a non-Gaussian input satisfies the Fokker-Planck equation. We discuss this Fokker-Planck equation in the next two sections. It will be shown that the density $p(w;t)$ is approximately Gaussian with time-dependent mean and variance.

## III. Fokker-Planck Equation

The input process $x(t)$ is assumed to have the form :

$$x(t) = s + n_{ns}(t)$$

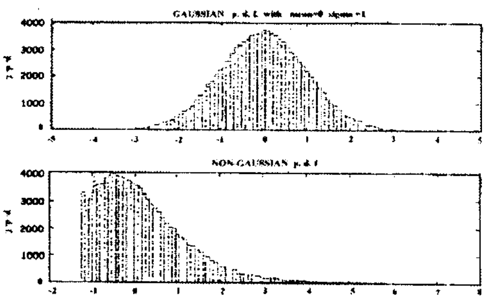

Figure 2. Gaussian VS. Non-Gaussian P.D.F.

where $s$ is a constant and $n_{ns}(t) = n(t) + \varepsilon n^2(t)$. Here $n(t)$ is a zero-mean stationary band-limited white Gaussian process with bandwidth $2B$ and power spectral density $N_0/2$. Therefore,

$$x(t) = s + n(t) + \varepsilon n^2(t) \tag{9}$$

We need the following facts to compute $Q_1(w)$ and $Q_2(w)$ :

$$E[n^2(t)] = N_0 B \tag{10}$$

$$E[n^2(u)\, n(v)] = E[n(u)\, n^2(v)] = 0 \tag{11}$$

$$E[n^3(u)\, n(v)] = 3(N_0 B) E[n(u)\, n(v)] \tag{12}$$

$$E[n^4(t)] = 3(N_0 B)^2 \tag{13}$$

$$\begin{aligned} E[n^2(u)\, n^2(v)] &= E[n(u)\, n(v)] E[n(v)\, n(v)] \\ &\quad + 2E[n(u)\, n(v)] E[n(u)\, n(v)] \\ &= (N_0 B)^2 + 2(E[n(u)\, n(v)]^2) \end{aligned} \tag{14}$$

$$E[n^2(u)\, n^4(v)] = 3(N_0 B)^3 \tag{15}$$

$$\begin{aligned} E[n^3(u)\, n^3(v)] &= 9(N_0 B)^2 [n(u)\, n(v)] \\ &\quad + 6(E[n(u)\, n(v)])^3 \end{aligned} \tag{16}$$

$$\begin{aligned} E[n^4(u)\, n^4(v)] &= 9(N_0 B)^4 + 72(N_0 B)^2 \\ &\quad (E[n(u)\, n(v)])^2 + 24(E[n(u)\, n(v)]) \end{aligned} \tag{17}$$

$$\begin{aligned} &\int_t^{t+\triangle t} \int_t^{t+\triangle t} \int_t^{t+\triangle t} E[n(u)\, n(v)]\, dudv \\ &= \int_t^{t+\triangle t} \int_t^{t+\triangle t} [\frac{N_0}{2} \frac{\sin 2\pi B(u-v)}{\pi(u-v)}]\, dudv \\ &= (1/2)(N_0)(\triangle t), \qquad \text{as} \quad B \to \infty \end{aligned} \tag{18}$$

$$\begin{aligned} &\int_t^{t+\triangle t} \int_t^{t+\triangle t} E[n(u)^2\, n(v)^2]\, dudv = (N_0 B)^2 \triangle t^2 \\ &+ 2\int_t^{t+\triangle t} \int_t^{t+\triangle t} \frac{N_0}{2} [\frac{\sin 2\pi B(u-v)}{\pi(u-v)}]^2\, dudv \\ &= (N_0 B)^2 \triangle t^2 + B(N_0)^2 \triangle t, \qquad \text{as} \quad B \to \infty \end{aligned} \tag{19}$$

$$\begin{aligned} &\int_t^{t+\triangle t} \int_t^{t+\triangle t} E[n(u)\, n(v)]]^3\, dudv \\ &= (3/8) N_0^3 B^2 \triangle t, \qquad \text{as} \quad B \to \infty \end{aligned} \tag{20}$$

$$\begin{aligned} &\int_t^{t+\triangle t} \int_t^{t+\triangle t} E[n(u)\, n(v)]]^4\, dudv \\ &= (1/4) N_0^4 B^3 \triangle t, \qquad \text{as} \quad B \to \infty \end{aligned} \tag{21}$$

The result in (10) follows directly from the definition of the bandlimited, white-noise process.

The results (11) through (17) are obtained by using the formulas for the third-order moments of a Gaussian process. In (18) through (21), we use the following result [8] :

$$\int_{-\infty}^{\infty} \frac{1}{\pi} [\frac{\sin \phi}{\phi}]^n\, d\phi = \begin{cases} 1, & n = 1, 2 \\ 3/4, & n = 3 \\ 2/3, & n = 4 \end{cases} \tag{22}$$

The next step in our development is to evaluate the quantities $Q_1(w)$ and $Q_2(w)$ based on the twelve facts shown above. From the definition of $Q_1(w)$ and $Q_2(w)$ in the previous section, we have :

$$\begin{aligned} Q_1(w) &= \lim_{\triangle t \to 0} \frac{1}{\triangle t} [\mu\beta_0 r \int_t^{t+\triangle t} x(u)\, du \\ &\qquad - \mu\beta_1 w \int_t^{t+\triangle t} x^2(u)\, du] \\ &= \mu(\beta_0 r[s + \varepsilon N_0 B] \\ &\qquad - \beta_1 N_0 B w[SNR + (2\varepsilon s + 1) + 3\varepsilon^2 N_0 B]\} \end{aligned} \tag{23}$$

where $SNR = s^2/N_0 B$ is the signal-to-noise ratio. Now

$$Q_2(w) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[\mu\beta_0 r \int_t^{t+\Delta t} x(u)\,du - \mu\beta_1 w \int_t^{t+\Delta t} x^2(u)\,du]^2 \quad (24)$$

After some tedious computations and simplifications, Equation (24) becomes

$$Q_2(w) = \mu^2\beta_0^2 r^2(\frac{N_0}{2} + \epsilon^2 BN_0^2) + 2N_0^2 B\mu^2\beta_1^2 w^2 [\frac{\epsilon^2 s}{N_0 B} + SNR + 3\epsilon s + 9\epsilon^2 N_0 \frac{B}{2} + 5\epsilon^4 N_0^2 B^2] - 2\mu^2\beta_0\beta_1 rw[sN_0 + \epsilon(5+4\epsilon s)BN_0^2] \quad (25)$$

Substituting $Q_1(w)$ and $Q_2(w)$ given by Equation (24) and (25), respectively into Equation (7) with $Q_n(w) = 0$ for $n \geq 3$, we have the Fokker-Planck equation for $p(w;t)$:

$$\frac{\partial p(w;t)}{\partial t} = -\mu\frac{\partial}{\partial w}\{(\beta_0 r[s+\epsilon N_0 B] - \beta_1 wN_0 B[SNR+(2\epsilon s+1)+3\epsilon^4 N_0 B]\} p(w;t)\} + \mu^2\frac{\partial^2}{\partial w^2}\{(\beta_0^2 r^2(\frac{N_0}{2} + \epsilon^2 BN_0^2) + 2N_0^2 B\beta_1^2 w^2[\frac{\epsilon^2 s}{N_0 B} + SNR + 3\epsilon s + 9\epsilon^2 N_0 \frac{B}{2} + 5\epsilon^4 N_0^2 B^2] - 2\beta_0\beta_1 rw[sN_0 + \epsilon(5+4\epsilon s)BN_0^2]\} p(w;t)\} \quad (26)$$

It should be emphasized that the derivation of Equation (26) assumes the noise process $n(t)$ has a large bandwidth $(B \to \infty)$. If we take the initial value of $w(t)$ to be $w_0$ then the initial condition for the above partial differential equation is:

$$p(w;0) = \delta(w - w_0) \quad (27)$$

where $\delta(w)$ is Dirac-delta function

## IV. Solution to the Fokker-Planck Equation

In this section, we examine the approximate solution to the Fokker-Planck equation assuming that the fixed constant $\mu$ in Equation (26) is sufficiently small. This assumption allows us to transform Equation (26) into a boundary layer equation. The choice of scale is predicated on the asymptotic behavior of the solution within the boundary layer [9].

We rewrite Equation (26) to bring out explicit dependence on the parameter $\mu$ and by replacing the tedious constant terms with temporary constants. We let the new constant terms with temporary constants. We let the new constants to be:

$$A_1 = \beta_1[SNR + (2\epsilon s+1) + 3\epsilon^2 N_0 B]$$

$$A_2 = 2N_0^2 Bbeta_1^2[\frac{\epsilon^3 s}{N_0 B} + SNR + 3\epsilon s + 1) + 3\epsilon s + 9\epsilon^2 N_0(\frac{B}{2}) + 5\epsilon^2 N_0(\frac{B}{2}) + 5\epsilon^4 N_0^2 B^2]$$

$$A_3 = 2\mu^2\beta_0\beta_1 r[sN_0 + 3N_0^3 \frac{B^2}{4} + \epsilon(2s\epsilon+1)BN_0^2]$$

Then we rewrite Equation (26) as:

$$\frac{\partial p(w;t)}{\partial t} = -\mu\frac{\partial}{\partial w}\{[\beta_0 r(s+\epsilon N_0 B) - wA_1 N_0 B] p(w;t)\} + \mu^2\frac{\partial^2}{\partial w^2}\{[\beta_0^2 r^2(\frac{N_0}{2} + \epsilon^2 BN_0^2) + w^2 A_2 - wA_3] p(w;t)\}$$

By carrying out the differentiations on the right side of the equation, we obtain:

$$\frac{\partial p(w;t)}{\partial t} = \mu[2\mu A_2 + A_1 N_0 B] p(w;t) + \mu[-\beta_0 r(s+\epsilon N_0 B) + wA_1 N_0 B + 2\mu(2wA_2 - A_3)]\frac{\partial p(w;t)}{\partial w} + \mu^2[-\beta_0^2 r^2(\frac{N_0}{2} + \epsilon^2 BN_0^2) + w^2 A_2 - wA_3]\frac{\partial^2 p(w;t)}{\partial w^2} \quad (28)$$

By the method of separation of variables [10], we assume a solution of the form $p(w,t) = W(w)T(t) \neq 0$. Substituting this expression into Equation (28), it can be shown that:

$$\frac{T'(t)}{T(t)} = \mu^2[\beta_0^2 r^2(\frac{N_0}{2} + \epsilon^2 BN_0^2) + w^2 A_2 - wA_3]\frac{W''(w)}{W(w)} + \mu[-\beta_0 r(s+\epsilon N_0 B) + wA_1 N_0 B + 2\mu(2wA_2 - A_3)]\frac{W'(w)}{W(w)} + \mu(2\mu A_2 + A_1 N_0 B) \quad (29)$$

To have bounded solutions in the time domain, a separation constant for (29) has to be negative number [10]. That is

$$T'(t) + \alpha^2 T(t) = 0 \quad (30)$$

and

$$0 = \mu^2[\beta_0^2 r^2(\frac{N_0}{2} + \epsilon^2 BN_0^2) + w^2 A_2 - wA_3]\frac{\partial^2 W(w)}{\partial w^2} + \mu[-\beta_0 r(s+\epsilon N_0 B) + wA_1 N_0 B + 2\mu(2wA_2 - A_3)]\frac{\partial W(w)}{\partial w} + [\alpha^2 + \mu(2\mu A_2 + A_1 N_0 B)]W(w) \quad (31)$$

where $\alpha^2$ is a seperation constant.

To transform Equation(31) into a boundary layer equation, we define

$$w_\infty \equiv \frac{\beta_0 r(s+\varepsilon N_0 B)/(\beta_1 N_0 B)}{2\varepsilon s+3\varepsilon^2 N_0 B+1+SNR} \tag{32}$$

It will be shown at the end of this section that $w_\infty$ is the average value of the weight function $w(t)$ at steady state $(t \to \infty)$. This steady-state average weight can be determined independent of the Fokker-Planck equation by choosing the weight, which minimizes the mean-squared error $E[e^2(t)]$.

We now develop a uniformly valid matched asymptotic expansion for $W(\omega)$. Near $\omega = \omega_\infty$, varies rapidly from small values of $\mu$. This indicates the formation of a boundary layer in a neighborhood about the steady-state average weight $\omega_\infty$. We expand this layer by rescaling the weight such that:

$$\varsigma = \frac{\omega - \omega_\infty}{\mu^\lambda} \tag{33}$$

where $\lambda > 0$. Substituting the rescaled weight into Equation (31), we find that:

$$\begin{aligned} 0 = {} & \mu^{2-2\lambda}[\beta_0^2 r^2(\frac{N_0}{2}+\varepsilon^2 BN_0^2) \\ & +(\omega_\infty^2+\varsigma^2\mu^{2\lambda}+2\varsigma\mu^\lambda\omega_\infty)A_2-(\varsigma\mu^\lambda+\omega_\infty)A_3]\frac{d^2W(\varsigma)}{d\varsigma^2} \\ & +\mu^{1-\lambda}\{-\beta_0 r(s+\varepsilon N_0 B)+(\varsigma\mu^\lambda+\omega_\infty)A_1 N_0 B \\ & +2\mu[2(\varsigma\mu^\lambda+\omega_\infty)A_2-A_3]\}\frac{dW(\varsigma)}{d(\varsigma)} \\ & +[\alpha^2+\mu(2\mu A_2+A_1 N_0 B)]W(\varsigma) \end{aligned} \tag{34}$$

Because the adaptation parameter $\mu$ of the backpropagation algorithm has to be a small number $(\mu \to 0)$, the choice of $\lambda$ in Equation (34) is not arbitrary. For example, if $\lambda > 1$ then Equation (34) becomes $d^2W/d\varsigma^2 = 0$ as $\mu \to 0$. This implies that $W(\varsigma)$ must be constant for all values of $\varsigma$ so that it is a bounded solution as dictated by the property of a probability density function. Now suppose $\lambda = 1$. This choice of $\lambda$ admits sinusoidal solutions as $\mu \to 0$ and the corresponding probability density functions are not admissible. Similarly, if $\lambda < 1$ and $\lambda \neq 1/2$ then solutions to Equation (34) are either trivial or nonadmissible. When $\lambda = 1/2$ we arrive at a candidate differential equation which describes the dynamics within the boundary layer as $\mu \to 0$. In particular:

$$\begin{aligned} & [\beta_0^2 r^2(\frac{N_0}{2}+\varepsilon^2 BN_0)+\omega_\infty^2 A_2-\omega_\infty A_3]\frac{d^2W(\varsigma)}{d\varsigma^2} \\ & +\varsigma A_1 N_0 B\frac{dW(\varsigma)}{d\varsigma}+[\rho^2+A_1 N_0 B]W(\varsigma)=0 \end{aligned} \tag{35}$$

where $\rho^2 = \alpha^2/\mu$. This choice of $\rho^2$ is valid since $\alpha^2$ is an arbitrary separation constant. Equation (35) is the dominant $O(1)$ boundary layer equation which results by ignoring all terms of order $\sqrt{\mu}$ or higher in Equation (34) after dividing both sides of Equation (34) by $\mu$.[1] Equation (35) can be simplified by using the steady-state average weight given by Equation (32). The resulting differential equation is:

$$\begin{aligned} & \frac{d^2W(\varsigma)}{d\varsigma^2}+2\gamma^2\varsigma+ \\ & [2\gamma^2+\frac{\rho^2}{\beta_0^2 r^2(\frac{N_0}{2}+\varepsilon^2 BN_0^2)+\omega_\infty^2 A_2-\omega_\infty A_3}]W(\varsigma)=0 \end{aligned} \tag{36}$$

where:

$$\gamma = \frac{N_0 B A_1}{2[\beta_0^2 r^2(\frac{N_0}{2}+\varepsilon^2 BN_0^2)+\omega_\infty^2 A_2-\omega_\infty A_3]}$$

We can now solve Equation (36) by introducing the following substitions:

$$W(\zeta) = V(\zeta)\exp(-\gamma^2\zeta^2) \tag{37}$$

and

$$\xi = \gamma\zeta \tag{38}$$

The canonical differential equation for $V(\xi)$ can be shown to be:

$$\frac{d^2V(\xi)}{d\xi^2}-2\xi\frac{dV(\xi)}{d\xi}+\left[\frac{\rho^2}{A_1 N_0 B}\right]V(\xi)=0 \tag{39}$$

which is independent of the desired response $r$. Equation (39) is the singular Sturm-Liouville equation with eigenvalues equal to $2n$ for nonnegative integer $n$ [10]. In other words, by setting $\rho^2 = \rho_n^2 = 2nA_1 N_0 B$, Equation (39) admits nontrival solutions. With this substitution, Equation (39) becomes a well known differential equation which has the Hermite polynomials $H_n(\xi)$ of degree $n$ as solutions [10]. Therefore, using Equation (37), the complete dominant $O(1)$ boundary layer solutions to Equation (31) can be expressed in terms of $H_n(\xi)$ as:

$$W_n(\xi) = B_n \exp(-\zeta^2\gamma^2)H_n(\gamma\zeta) \tag{40}$$

1) The landau symbol $O()$ has the following meaning: $f(\varepsilon) = O[g(\varepsilon)]$ as $\varepsilon \to 0$ if there exits a positive number A independent of $\varepsilon$ such that $|f(\varepsilon)| \le A|g(\varepsilon)|$ for all $|\varepsilon| < \varepsilon_0$, where $\varepsilon_0 > 0$.

where $B_n$ is an arbitrary normalization constant.

Equation (40) describes the inner boundary layer solution to Equation (31). The "thickness" of the layer is of order $\sqrt{\mu}$. In order to obtain a uniformly valid asymptotic expansion for the weight probability density function on the interval $(-\infty, \infty)$, the solution outside the boundary layer must be examined. It is shown [2] that the solution to Equation (31) is zero to $O(1)$ outside of the boundary layer.

Now, for $\alpha^2 = \rho_n^2 = \mu\beta_n^2$, the general solution to Equation (30) may be written in the form

$$T_n(t) = K_n \exp[-\alpha_n^2 t] = K_n \exp[-2\mu\sigma_N^2 A_1 nt] \tag{41}$$

where $K_n$ is an arbitrary constant and

$$\sigma_N^2 = N_0 B \tag{42}$$

is the noise variance. Thus, using Equation (40) and (41), the functions

$$p_n(w;t) = W_n(w)\, T_n(t) \tag{43}$$

satisfy Equation (28). By the superposition principle, the general solution to Equation (28) is

$$p(w;t) = \sum_{n=0}^{\infty} p_n(w;t) \tag{44}$$

Therefore, we can now write the uniformly valid asymptotic solution to Equation (28) in the following form:

$$p(w;t) \approx \exp[\frac{-(w-w_\infty)^2}{\sqrt{2\sigma_\infty^2}}] \sum_{n=0}^{\infty} C_n \exp(-2\mu\sigma_N^2 A_1 nt) H_n \frac{-(w-w_\infty)^2}{\sqrt{2\sigma_\infty^2}} \tag{45}$$

where

$$\sigma_\infty^2 = \frac{\mu[\beta_0^2 r^2(\frac{N_0}{2} + \varepsilon^2 BN_0^2) + w_\infty^2 A_2 - w_\infty A_3]}{N_0 BA_1 r^2} \tag{46}$$

and $C_n = K_n B_n$ are normalizing constants. It will be shown at the end of this section that $\sigma_\infty^2$ is the steady-state variance of $w(t)$.

The constants are chosen to satisfy the initial condition $p(w;0) = \delta(w-w_0)$ in Equation(27). Invoking the orthogonality property of Hermite polynomials [10], the normalizing constants $C_n$ are:

$$C_n = \frac{1}{2^n n! \sqrt{2\pi\sigma_\infty^2}} H_n[\frac{w_0 - w_\infty}{\sqrt{2\sigma_\infty^2}}] \tag{47}$$

Upon substituting $C_n$ into Equation (45), the weight probability density function can be written as:

$$p(w;t) = p_\infty(w) \sum_{n=0}^{\infty} \frac{1}{2^n n!} \exp[-2\mu\sigma_N^2 A_1 nt] H_n(\frac{w-w_\infty}{\sqrt{2\sigma_\infty^2}}) H_n(\frac{w_0 - w_\infty}{\sqrt{2\sigma_\infty^2}})[1 + O(\sqrt{\mu})] \tag{48}$$

where

$$p_\infty(w) = \frac{1}{\sqrt{2\pi\sigma_\infty^2}} \exp(\frac{-(w-w_\infty)^2}{\sqrt{2\pi\sigma_\infty^2}}) \tag{49}$$

Because $H_0(x) = 1$ for any $x$, $p_\infty(w)$ is the weight probability density function at steady-state, i.e., $\lim_{t\to\infty} p(w;t) = p_\infty(w)$.

We now find a closed-form expression for Equation (48) by considering the following Hermite summation identity [11]:

$$\sum_{n=0}^{\infty} \frac{z^n}{2^n n!} H_n(x) H_n(y) = \frac{1}{\sqrt{1-z^2}} \exp[\frac{2zxy - z^2(x^2+y^2)}{1-z^2}] \tag{50}$$

where $|z| < 1$. This identity follows by using the Hermite generating function and the corresponding integral representation for the Hermite function. In order to directly associate Equations (50) with (48), we define:

$$z = e^{-\frac{t}{\tau}} \tag{51}$$

$$x = \frac{w - w_\infty}{\sqrt{2\sigma_\infty^2}} \tag{52}$$

and

$$y = \frac{w_0 - w_\infty}{\sqrt{2\sigma_\infty^2}} \tag{53}$$

where

$$\tau = \frac{1}{2\mu\sigma_N^2 A_1} \tag{54}$$

is the corresponding time constant for the stochastic process. Substituting Equation (50) into (48) and performing some algebraic manipulations, it can be shown that:

$$p(w;t) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left[\frac{-(w-w(t))^2}{2\sigma^2(t)}\right][1+O(\sqrt{\mu})] \quad (55)$$

where the time-dependent average weight value $w(t)$ is given by :

$$w(t) = w_\infty + \exp^{-\frac{t}{\tau}}(w_0 - w_\infty) \quad (56)$$

and the time-dependent weight variance

$$\sigma^2(t) = (1-\exp^{-2\frac{t}{\tau}})\,\sigma_\infty^2 \quad (57)$$

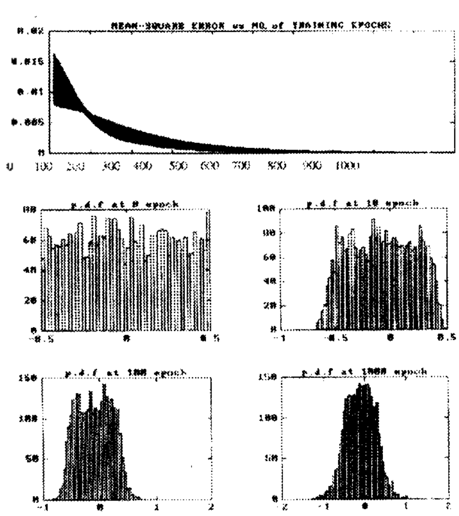

Figure 3. Evolution of Weight P.D.F's as $t \to \infty$.

The probability density function in Equation (55) is Gaussian to $O(1)$ with a time-dependent mean and variance given by Equations (56) and (57) respectively. As expected, when $t \to \infty$, $w(t) \to w_\infty$ and $\sigma^2(t) \to \sigma_\infty^2$. Additionally, the rate at which the time-dependent density in Equation (55) approaches steady state is directly proportional to the fixed constant $\mu$, the noise variance $\sigma_N^2$, and the corresponding Signal-to-Noise Ratio (SNR) as shown explicitly in Equation (54).

Figure 3 shows the result of a training process together with the curve of $p(w;t)$ at t = 0, 500 and 1000 epochs using Equation (55). Here, we normalized the weight such that it is uniformly distributed with zero mean and its variance is equal to 1/12. Five non-Gaussian input pattern vectors of 100 samples each were generated using the random number generator in accordance to the formula in Equation (9). As shown in Figure 3, Equation (55) is a good approximation for the probability density function of the converged weight when the artificial neural network's backpropagation algorithm is operated in the auto-associative mode.

## V. Conclusions

This paper presented the dynamical behavior, in probabilistic sense, of a feedforward neural network performing auto-association for novelty filtering. Networks of retinotopic topology having a one-to-one correspondence between input and output units can be readily trained using backpropagation algorithm, to perform autoassociative mappings. As an analysis of the novelty filtering, the probability density function of the weight is shown to converge to Gaussian when the input time series is statistically characterized by nonsymmetrical probability density functions. After output units are locally linearized, the recursive relation for updating the weight of the neural network is converted into a first-order random differential equation. Based on this equation it is shown that the probability density function of the weight satisfies the Fokker-Planck equation. By solving the Fokker-Planck equation, it is found that the weight is Gaussian distributed with time dependent mean and variance, which closely approximates the actual converged weight of network in novelty filtering setting. The rate at which the time-dependent density approaches steady state is shown directly proportional to the fixed constant $\mu$, the noise variance $\sigma_N^2$, and the corresponding Signal-to-Noise Ratio (SNR).

## References


1. H. Ko and R. H. Baran, "Signal detectability enhancement with auto-associative backpropagation networks," *Neurocomputing*, pp. 219-236, 1994.
2. T. Nguyen and G. M. Jacyna, "Probabilistic behavior of a least-mean-square adaptive filter," Tech Rep. MTR-90W00142, The MITRE Corporation, Mclean, VA, 1990.
3. S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ : Prentive-Hall, 1986.
4. R. P. Lippman, "An introduction to computing with neural networks," *IEEE ASSP Magazine*, 1987.
5. D. E. Rummelhart, *Parallel Distributed Processing*, MIT Press, 1986.
6. A. J. Viterbi, *Principles of Coherent Communication*, New York, NY : McGraw-Hill, 1969.
7. T. T. Soong, *Random Differential Equations in Science and Engineering*, New York, NY : Academic Press, 1973.
8. I. S. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, New York, NY : Academic Press, Inc., 4th ed., 1965.
9. A. Nayfeh, *Perturbation Methods*, New York, NY : John Wiley and Sons, Inc., 1973.

10. T. Myint-U and L.Debbnath, *Partial Differential Equation for Scientists and Engineers*, New York, NY : North-Holland, 3rd ed., 1987.
11. P. Morse and H. Feshbach, *Methods of Theoretical Physics*, New York, NY : McGraw-Hill, 1953.

▲Hanseok Ko

Hanseok Ko received the B.S. degree from Carnegie-Mellon University in 1982, M.S. degree from the Johns Hopkins University in 1988, and Ph.D. degree from the Catholic University of America in 1992, all in electrical engineering. He also received the M.S. degree in Systems Engineering in 1986 from the University of Maryland, College Park. From 1982 to 1987, he was with the Search and Track Division of the Naval Surface Warfare Center, White Oak, Maryland, where his work involved radar signal and infrared image processing for the detection of low observable targets. From 1988 to 1994, he directed research and development projects for the 21st Century ship automation system with the Sensors and Electronics Division, NSWC. In addition, Dr. Ko was a part-time faculty member in the Dept of Electrical Engineering at UMBC from 1992 to 1995. In March of 1995, Dr. Ko joined the faculty of the Department of Electronics Engineering, Korea University and is currently Associate Professor. His professional interests include signal processing, data fusion, navigation, and tracking.