USC-SIPI REPORT # 144

Structural Stability of Unsupervised Learning in Feedback Networks

By

Bart Kosko

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
Powell Hall of Engineering
University Park/MC-0272
Los Angeles, CA 90089 U.S.A.

April 1989

Submitted for Publication: April, 1989

STRUCTURAL STABILITY OF UNSUPERVISED LEARNING IN FEEDBACK NETWORKS

Bart Kosko
Department of Electrical Engineering–Systems
Signal and Image Processing Institute
University of Southern California
Los Angeles, California 90089-0272

Abstract

Structural stability is insensitivity to perturbations. Global stability, in contrast, is convergence to fixed points for all inputs and all parameters. Globally stable neural networks need not be structurally stable, need not be robust. Shaking can distort, destroy, or prevent equilibria. Then large-scale hardware implementation becomes dubious, and biological plausibility decreases. A large class of unsupervised nonlinear feedback neural networks, adaptive bidirectional associative memory (ABAM) models, is proven structurally stable. This is achieved by extending the ABAM models to the random-process domain as systems of stochastic differential equations, and appending scaled Brownian diffusions. This much larger family of models, random ABAM (RABAM) models, is then proved globally stable. Intuitively, RABAM equilibria are ABAM equilibria that randomly vibrate. Included in the ABAM family of structurally stable models are Hopfield circuits, Hodgkin-Huxley networks, competitive-learning networks, and ART-2 networks. All RABAM models permit Brownian annealing. The extent of RABAM system "vibration" is characterized by the RABAM Noise Suppression Theorem: The mean-squared activation and synaptic velocities, $E[\dot{x}_i^2]$, $E[\dot{y}_j^2]$, and $E[\dot{m}_{ij}^2]$, decrease exponentially quickly to their lower bounds, the respective temperature-scaled noise "variances," $T_i\sigma_i^2, T_j\sigma_j^2$, and $T_{ij}\sigma_{ij}^2$. This suggests that many feedback neural network models are more biologically "realistic" than they are often criticized as being. For, the many neuronal and synaptic parameters missing from such neural network models are now included, but as net random unmodeled effects. They simply do not affect the structure of realtime global computations.

I. Structural Stability with the Stochastic Calculus

Structural stability is insensitivity to small perturbations[9,24]. Shaking a dynamical system in equilibrium can distort or destroy the equilibrium. Shaking a dynamical system tending toward equilibrium can change the eventual equilibrium to a new equilibrium. Or it can prevent any equilibrium from being reached. Structurally stable dynamical systems are invariant under shaking, at least under mild shaking.

Are neural network dynamical systems structurally stable? This question may never arise when experimenting with small-scale software simulations of feedback networks. But it rises to the fore when considering large-scale hardware implementations of neural networks or when searching, or arguing, for biological correlates of modern neural network models. Thermal noise and component malfunction are inherent in VLSI implementations, especially analog VLSI implementations.

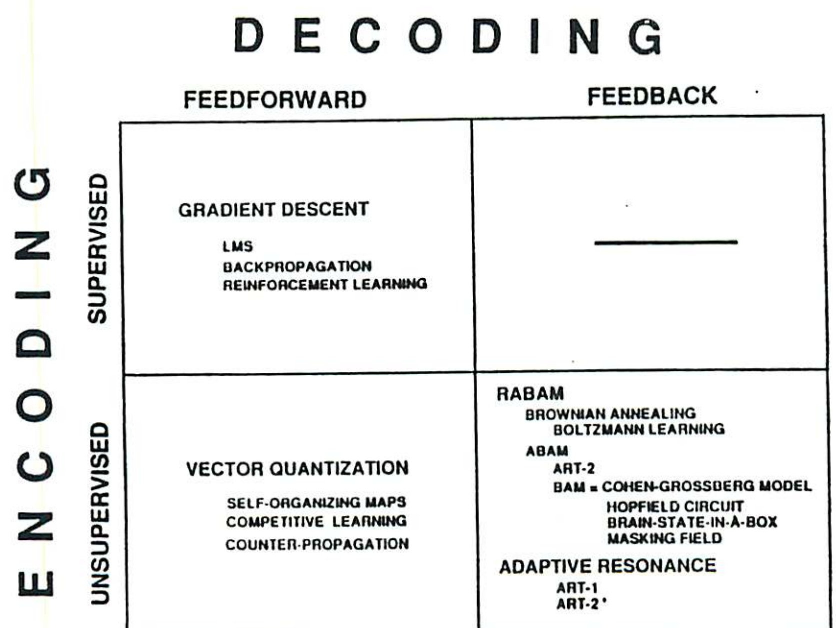
The biological need for the structural stability of unsupervised learning is even greater. Real synaptic junctions are embedded in numerous electro-chemical, molecular, hormonal, and glial processes. Yet neural network models summarize the gross "synaptic efficacy" of a synaptic junction with a single real number. It is no surprise that many neurobiologists find neural network models "unrealistic." If unsupervised learning is structurally stable, the plausibility, or "realism," of these "unrealistic" models increases. For instance, if feedback unsupervised learning systems are structurally stable in the sense of insensitivity to synaptic (and neuronal) noise, then the myrid missing synaptic processes that make current neural network models "unrealistic" are modeled but as net random unmodeled effects—effects that do not affect realtime global network computations.

We propose the stochastic calculus[19,23] to establish the structural stability of unsupervised feedback neural models. The more formal approach of differential topology[9], using transversality techniques, tends to be extremely abstract and computationally unwieldy. For the differential topological approach considers all behavior of all related functions (open dense sets of functions). The approach of stochastic differential equations, in contrast, only considers the statistically relevant behavior of bunches of functions. Innumerable pathologies are eliminated in one stroke. This allows quantities to be manipulated as in the deterministic calculus, at least under appropriate conditions. The stochastic approach, though, does involve new conceptual and computational complexity. Solutions to deterministic differential equations are functions. Solutions to stochastic differential equations are random processes[23].

Below the stochastic calculus is used to prove the structural stability of a large class of nonlinear adaptive dynamical systems, including, but not limited to, many popular neural models. This family of autoassociative and heteroassociative models is called the family of *random adaptive bidirectional associative memory (RABAM)* models.

The "random" in RABAM refers to the stochastic calculus setting, including determinisitic cases degenerately. This includes annealing of noise. "Adaptive" refers to the unsupervised learning laws that, when combined with the general neuronal state nonlinear dynamics, yield structural stability. These learning laws include, but are not limited to, the signal Hebb law and the competitive law, discussed below. Nonadaptive models, such as the Hopfield circuit, are again obtained degenerately.

"Bidirectional" refers to the globally stable two-way flow of neuronal signal information between any two layers of a neural network[14,15] – provided, as in the ART-2 model[1], that the forward flow is through the adaptive matrix M and the backward flow is through the transpose matrix M^T. Bidirectional flow reduces to unidirectional flow within a single field of neurons when $M = M^T$, as in the Hopfield circuit.



NEURAL NETWORK TAXONOMY

Figure 1

"Associative memory" refers to the nonlinear dynamics of the neurons —how they transduce net inputs into bounded signals—and, broadly construed, the joint nonlinear dynamics of the synapses as well. A general taxonomy of neural networks is given in Figure 1. If feedback loops are pruned from RABAM models, the feedforward unsupervised neural algorithms, used ultimately for some form of vector quantization, are RABAM models as well.

II. The RABAM Diffusion and Noise Models

RABAM models are ABAM models[16] cast as random processes and perturbed by Brownian diffusions. In general these time-varying diffusions or "noise" processes can be scaled with deterministic "annealing" schedules[3]. In fullest generality, the *diffusion version* of the RABAM model is described by the following stochastic dynamical system:

$$dx_i = -a_i(x_i)[b_i(x_i) - \sum_j S_j(y_j)m_{ij}]dt + \sqrt{T_i}\, dB_i, \tag{1}$$

$$dy_j = -a_j(y_j)[b_j(y_j) - \sum_i S_i(x_i)m_{ij}]dt + \sqrt{T_j}\, dB_j, \tag{2}$$

$$dm_{ij} = [-m_{ij} + S_i(x_i)S_j(y_j)]dt + \sqrt{T_{ij}}\, dB_{ij} \tag{3}$$

or the signal Hebb law (3) can be replaced with the competitive learning law (4),

$$dm_{ij} = S_j[S_i - m_{ij}]dt + \sqrt{T_{ij}}\, dB_{ij}, \tag{4}$$

if the signal function S_j is reasonably steep, or with other learning laws, such as the differential Hebb or differential competitive learning laws under appropriate conditions. The logistic signal function discussed below is "reasonably steep."

The basic structure of the nonlinear dynamics in (1) and (2) is the general form proposed by Cohen and Grossberg[2]. This family of dynamical systems is what is meant by "Cohen-Grossberg dynamics" in the literature. The functions (random processes) a_i and a_j are general nonnegative "amplification" functions, often constant or linear in practice, which we shall sometimes take to be strictly positive. The functions b_i and b_j are essentially arbitrary nonlinear functions. Technically they must be constrained so as to keep the ABAM Lyapunov function below bounded. They appear in integrals in the Lyapunov function.

The signal functions S_i and S_j, of the ith and jth neurons in the respective n-dimensional neural field F_X and p-dimensional neural field F_Y, are bounded monotone nondecreasing functions. Again it is often convenient to only deal with strictly increasing signal functions, so that the derivative of the signal S_i with respect to the ith neuron's real-valued "activation," or membrane potential, x_i is strictly positive: $\frac{dS_i}{dx_i} = S_i' > 0$, where we have denoted the activation derivative with a prime. This is true of the popular logistic sigmoid signal function $S(x) = (1 + e^{-cx})^{-1}$ for positive constant $c > 0$, since $S' = c\, S\, (1 - S) > 0$.

The processes B_i, B_j, and B_{ij} are Brownian diffusions[23]. The stochastic differentials dB_i, dB_j, and dB_{ij} are assumed obtained from the standard limit techniques used to construct stochastic integrals, such as the popular Wiener, Ito, or Stratonovich stochastic integrals[19].

The "temperature" functions T_i, T_j, and T_{ij} are *deterministic* nonnegative "annealing schedules." They are control parameters that scale the system randomness. If they decrease with time, which we need not assume, they represent "cooling." The intuition is that infinite temperatures yield pure random search, while, in gradient systems, zero temperatures yield pure gradient descent[3]. Between these temperature extremes is a spectrum of scaled random hill climbing. Temperature functions are included here for generality. For most neural models, the temperature functions are constant functions.

The *noise version* of the RABAM model is given by the less rigorous, more intuitive, notation of scaled additive noise processes:

$$\begin{aligned}
\dot{x}_i &= -a_i(x_i)[b_i(x_i) - \sum_j S_j(y_j)m_{ij}] \;+\; \sqrt{T_i}\; n_i, && (5)\\
\dot{y}_j &= -a_j(y_j)[b_j(y_j) - \sum_i S_i(x_i)m_{ij}] \;+\; \sqrt{T_j}\; n_j, && (6)\\
\dot{m}_{ij} &= -m_{ij} \;+\; S_iS_j \;+\; \sqrt{T_{ij}}\; n_{ij}, && (7)
\end{aligned}$$

or with (7) replaced with a noisy competitive learning law,

$$\dot{m}_{ij} \;=\; S_j[S_i - m_{ij}] \;+\; \sqrt{T_{ij}}\; n_{ij}, \qquad (8)$$

if S_j is reasonably steep, or, again, with other unsupervised learning laws under appropriate conditions.

The noise version of the RABAM model is convenient but fictitious. We shall use it hereafter to describe the RABAM model, remembering all the while that it is shorthand for the diffusion version. For, with probability one, Brownian motion is continuous but nondifferentiable[23]. Brownian motion has no standard time derivative, in particular a time derivative that is white noise. Still the noise interpretation can be justified on formal, if approximate, grounds and it has served well a half century of estimation and control theory modeling[19].

The noise RABAM implicitly assumes that the scaled noise terms are statistically independent of the nonlinear "signal" terms to which they are added. This independence stems from the independence of "increments" of the Brownian motion used to construct the temperature-scaled stochastic differentials. To prove the RABAM Theorem, we need only assume that the noise processes are *uncorrelated* with the corresponding "signal" processes. We also make the customary assumptions that the noise processes are zero mean and have finite, though possibly time- varying, variances:

$$E[n_i] = E[n_j] = E[n_{ij}] = 0 \text{ for all } i \text{ and } j, \tag{9}$$

$$V(n_i) = \sigma_i^2 < \infty,\ \sigma_j^2 < \infty,\ \sigma_{ij}^2 < \infty\ . \tag{10}$$

The strategy[18] is to prove that RABAM models are *globally* stable. This will prove that the underlying deterministic ABAM models are *structurally* stable. The tactic is to use the expectation, or average, of the ABAM Lyapunov function as a Lyapunov function for the RABAM stochastic dynamical system.

III. ABAM Models and the ABAM Theorem

Adaptive bidirectional associative memory (ABAM) models[14–18] are described by deterministic dynamical sytems of the form

$$\dot{x}_i = -a_i(x_i)[b_i(x_i) - \sum_j S_j(y_j)m_{ij}], \tag{11}$$

$$\dot{y}_j = -a_j(y_j)[b_j(y_j) - \sum_i S_i(x_i)m_{ij}], \tag{12}$$

$$\dot{m}_{ij} = -m_{ij} + S_i S_j \tag{13}$$

or with (13) replaced with the competitive learning law,

$$\dot{m}_{ij} = S_j[S_i - m_{ij}], \tag{14}$$

if S_j is reasonably steep, or with other unsupervised learning laws[4,5,12,16–18] under appropriate conditions[16–18]. Higher-order versions of the ABAM model are readily obtained[16].

The generality of the ABAM model can best be seen by example. As listed in Figure 1, the general Cohen-Grossberg model[2], the Hopfield circuit[11], and the continuous adaptive resonance theory model ART-2[1] are special cases of ABAM models. For example, if there is no learning, if the two neural fields F_X and F_Y collapse into one, $F_X = F_Y$, and if the resulting square constant connection matrix M is symmetric, $M = M^T$, then the ABAM model reduces to the (unidirectional autoassociative) Cohen-Grossberg model:

$$\dot{x}_i = -a_i(x_i)[b_i(x_i) - \sum_j S_j(x_j)m_{ij}] \ . \tag{15}$$

The Hopfield circuit corresponds to an additive Cohen-Grossberg model[7]. An additive activation model has constant amplification functions and linear b_i functions. In particular, if $a_i = \frac{1}{C}, b_i = \frac{x_i}{R_i} - I_i$, and if the symmetric weight m_{ij} is relabeled T_{ij} and $S_i(x_i) = V_i$, where C and R_i are positive constants ("capacitance" and membrane "resistance") and input I_i is constant or slowly varying relative to fluctuations in x_i, then the Cohen-Grossberg model (15) formally reduces to the Hopfield circuit:

$$C\dot{x}_i = -\frac{x_i}{R_i} + \sum_j V_j T_{ij} + I_i \ . \tag{16}$$

A *shunting*[6] or *multiplicative* model has linear amplification functions, whence comes the name, and nonlinear b_i functions. The shunting model is a generalization of the Hodgkin-Huxley membrane equation[10]:

$$C\frac{\partial V_i}{\partial t} = (V^p - V_i)g_i^p + (V^+ - V_i)g_i^+ + (V^- - V_i)g_i^- \ , \tag{17}$$

where V^p, V^+, and V^- are respective passive, excitatory (sodium Na^+) and inhibitory (potassium K^+) saturation upper bounds with corresponding shunting conductances g_i^p, g_i^+, and g_i^-, and where the constant positive capacitance C scales time. Details of this subsumption can be found elsewhere[6–7].

The ART-2 model is a competitive ABAM[18]. The competitive learning law (14) defines the forward or "bottom-up" synaptic matrix M and, symmetrically, the backward or "top-down" synaptic matrix M^T. Within fields F_X and F_Y, neurons "compete" with shunting dynamics. In particular, the neurons in F_Y make "choices." This means they have steep signal functions; indeed, they are threshold functions. This satisfies the ABAM Theorem[16] discussed next.

ABAM Theorem. The ABAM model (11) - (13), or competitive ABAM model (11) - (12) and (14), etc., is globally stable. If $a_i > 0$, $a_j > 0$, and the signal functions S_i and S_j are strictly increasing, then the ABAM model is *asymptotically* stable, and the square activation and synaptic velocities $\dot{x}_i^2$, $\dot{y}_j^2$, and $\dot{m}_{ij}^2$ decrease exponentially quickly to zero, their equilibrium value.

Proof. The proof uses the bounded Lyapunov function L,

$$
\begin{aligned}
L \;=\; & -\sum_i \sum_j S_i S_j m_{ij} + \sum_{i=1}^{n} \int_0^{x_i} S_i'(\theta_i) b_i(\theta_i) d\theta_i \\
& + \sum_{j=1}^{p} \int_0^{y_j} S_j'(\varepsilon_j) b_j(\varepsilon_j) d\varepsilon_j + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{p} m_{ij}^2 \;. \qquad (18)
\end{aligned}
$$

It takes some care to assure the boundedness of the integrals in (18). Pathologies can occur, though we shall ignore them. The boundedness of the quadratic form is trivial since the signal functions are bounded. The sum of squared synaptic efficacies is also bounded because, again, the signal functions are bounded in the first-order learning laws in which they occur.

Time differentiation of (18) gives, upon grouping of like terms,

$$
\begin{aligned}
\dot{L} \;=\; & -\sum_i S_i' \dot{x}_i [b_i - \sum_j S_j m_{ij}] + \sum_j S_j' \dot{y}_j [b_j - \sum_i S_i m_{ij}] \\
& -\sum_i \sum_j \dot{m}_{ij} [-m_{ij} + S_i S_j] \qquad (19) \\
\;=\; & -\sum_i S_i' a_i [b_i - \sum_j S_j m_{ij}]^2 - \sum_j S_j' a_j [b_j - \sum_i S_i m_{ij}]^2
\end{aligned}
$$

$$-\sum_i \sum_j [-m_{ij} + S_i S_j]^2 \ , \tag{20}$$

upon eliminating the terms in braces in (19) with, respectively, the ABAM dynamical equations (11) - (13). Since the amplification functions are nonnegative and the signal velocities are monotone nondecreasing (so their activation derivatives are nonnegative), the righthand side of (20) is decreasing along trajectories. This proves global stability for ABAM systems that either do not adapt (BAM systems) or adapt according to the signal Hebb law (13).

To prove global stability for the competitive learning law[16] (14) —and to outline a general proof strategy for arbitrary candidate first-order learning laws —we eliminate the third term in braces in (19) with (14) to give

$$\begin{aligned} \dot{L} &= -\sum_i S_i' a_i [b_i - \sum_j S_j m_{ij}]^2 - \sum_j S_j' a_j [b_j - \sum_i S_i m_{ij}] \\ &\quad -\sum_i \sum_j S_j [S_i - m_{ij}][S_i S_j - m_{ij}] \ . \end{aligned} \tag{21}$$

We now invoke the *competitive assumption* that S_j is reasonably steep, indicating the win-loss status of the jth neuron in the competitive, or laterally inhibitive, neural field F_Y. In particular —and in accord with practically all competitive learning models[6,22] and implementations —we assume that S_j behaves approximately as a zero-one threshold. This is often achieved with an appropriately scaled logistic sigmoid, as discussed above. Then the third sum in (21) is nonpositive since its summand is nonnegative:

$$\begin{aligned} \dot{m}_{ij}[S_i S_j - m_{ij}] &= S_j [S_i - m_{ij}][S_i S_j - m_{ij}] \\ &= \begin{cases} 0 & \text{if} \quad S_j = 0 \\ (S_i - m_{ij})^2 & \text{if} \quad S_j = 1 \ . \end{cases} \end{aligned} \tag{22}$$

This proves the global stability of the competitive ABAM system (11), (12), and (14). Note that the nonnegativity in (22) holds approximately as S_j approximately equals a zero-one threshold function. Moreover, a statistical argument can be made to further weaken the competitive zero-one assumption. For the third term in (21) is a large sum. Different summands can be positive or negative so long as the overall sum is nonpositive —so long, more generally, as the overall sum is nonpositive *on average*.

The stronger asymptotic stability[8,20] of the ABAM models (Hebbian, competitive, and otherwise) is proved with the positivity assumptions $a_i > 0, a_j > 0, S_i' > 0$, and $S_j' > 0$. For convenience we shall only detail the proof for signal Hebbian learning. The proof for competitive learning uses (22) as above. The proof for signal-velocity learning laws, such as the differential Hebb law[4,5,12,13,16–18], is detailed elsewhere[16–18]. The time derivative of L again equals (19), but now the positivity assumptions can be used to eliminate the terms in braces in a way that differs from (20). In particular,

$$\dot{L} = -\sum_i \frac{S_i'}{a_i}\dot{x}_i^2 - \sum_j \frac{S_j'}{a_j}\dot{y}_j^2 - \sum_i \sum_j \dot{m}_{ij}^2 \ < \ 0 \tag{23}$$

along trajectories for any nonzero change in any neuronal activation or any synapse. This proves asymptotic global stability. Trajectories end in equilibrium points, not merely near them. Indeed (23) implies that

$$\dot{L} = 0 \quad \text{if and only if} \quad \dot{x}_i^2 = \dot{y}_j^2 = \dot{m}_{ij}^2 = 0 \tag{24}$$

$$\text{if and only if} \quad \dot{x}_i = \dot{y}_j = \dot{m}_{ij} = 0 \tag{25}$$

for all i and j, giving the desired equilibrium condition on squared velocities. That the squared velocities decrease *exponentially quickly* follows from the strict negativity of (23) and, to rule out pathologies (system Jacobian eigenvalues with zero real parts), from the second-order assumption of a nondegenerate Hessian. For asymptotic stability ensures that the real parts of eigenvalues of the ABAM system Jacobian matrix, about an equilibrium, are nonpositive[20]. Standard results[8] from dynamical systems theory then ensure that, locally, the nonlinear system behaves linearly and decreases to its equilibrium exponentially fast. Q. E. D.

IV. The RABAM Theorem and the RABAM Noise Suppression Theorem

The general RABAM Theorem[18] guarantees the global stability of the RABAM diffusion/noise model. Intuitively, RABAM equilibria are ABAM equilibria that "vibrate" randomly. On average they are the same as ABAM equilibria.

But *how much* do RABAM equilibria vibrate? This question concerns the second-order statistical behavior of RABAM equilibria. In theory, one might expect compound instability, even chaos, for a massively parallel, massively fedback, nonlinear dynamical system with different noise processes perturbing every state variable. The general RABAM Theorem ensures that such systems will not explode into instability or chaos. The theorem does not ensure that there will be only moderate random fluctuation. In principle system fluctuations could increase in magnitude in time up to some saturation value —a saturation value that may destroy realtime global computations.

The RABAM Noise Suppression Theorem not only guarantees that (average) fluctuations will not increase in time, it guarantees that they will decrease to their theoretical lower bound. It further guarantees that these fluctuations will decrease to their lower bound exponentially quickly. The Noise Suppression Theorem is the positivity-assumption corollary of the general RABAM Theorem, which we prove first, extending the proof technique of the ABAM Theorem. Once again the generality of the RABAM model must be stressed. We do not simply add "noise" to the deterministic ABAM model. We add "noise" to the random-process ABAM model.

RABAM Theorem. Under appropriate smoothness conditions (to pass a time derivative "inside" an integral), the RABAM models (1) - (3) and (5) - (7), or (1), (2), and (4), or (5), (6), and (8), are globally stable. If the random processes a_i, a_j, S_i', and S_j' are strictly positive processes, then the RABAM models are asymptotically stable.

Proof. The proposed bounded Lyapunov function is $E(L)$, the average ABAM Lyapunov function defined in (18). This expectation is taken with respect to all random parameters:

$$E(L) = \int \ldots \int L \; p(X,Y,M) dX dY dM \quad . \tag{26}$$

We note that L is of bounded variation since the variance $V(L)$ is bounded above by the mean-squared value $E(L^2)$. The boundedness of each term in L^2 follows from the boundedness of each term in L in (19). So $E(L^2)$ is finite.

The smoothness assumption allows the time derivative of $E(L)$ to be replaced with the more tractable expectation of the time derivative of L, the structure of which is given by (19). This commutation gives

$$
\begin{aligned}
\dot{E}[L] &= E[\dot{L}] \\
&= E\{\sum_i S_i'\dot{x}_i[b_i - \sum_j S_j m_{ij}] + \sum_j S_j'\dot{y}_j[b_j - \sum_i S_i m_{ij}] \\
&\quad -\sum_i \sum_j \dot{m}_{ij}[-m_{ij} + S_i S_j]\} \qquad (27) \\
&= E\{-\sum_i S_i' a_i[b_i - \sum_j S_i m_{ij}]^2 - \sum_j S_j' a_j[b_j - \sum_i S_i m_{ij}]^2 \\
&\quad -\sum_i \sum_j [-m_{ij} + S_i S_j]^2 + \sum_i S_i'\sqrt{T_i}\; n_i[b_i - \sum_j S_j m_{ij}] \\
&\quad +\sum_j S_j'\sqrt{T_j}\; n_j[b_j - \sum_i S_i m_{ij}] - \sum_i \sum_j \sqrt{T_{ij}}\; n_{ij}[-m_{ij} + S_i S_j]\}\ , \qquad (28)
\end{aligned}
$$

upon eliminating the activation and synaptic velocities (velocity processes) in (27) with the RABAM noise equations (5) - (7), and by observing that the deterministic "temperature" functions can be factored out of all expectations,

$$
\begin{aligned}
&= E[\dot{L}_{ABAM}] + \sum_i \sqrt{T_i}\; E[n_i] E\{S_i'[b_i - \sum_j S_j m_{ij}]\} \\
&\quad +\sum_j \sqrt{T_j}\; E[n_j] E\{S_j'[b_j - \sum_i S_i m_{ij}]\} \\
&\quad -\sum_i \sum_j \sqrt{T_{ij}}\; E[n_{ij}] E[-m_{ij} + S_i S_j] \qquad (29)
\end{aligned}
$$

by the uncorrelatedness (independence) of the "signal" and corresponding additive noise processes in the RABAM, and by the facts that S_i' and S_j' are nonnegative random functions of x_i and y_j respectively, and that the amplification functions a_i and a_j are essentially arbitrary nonnegative random functions, which permits the choices $a_i = S_i'$ and $a_j = S_j'$,

$$
= E[\dot{L}_{ABAM}] \qquad (30)
$$

by the zero-mean assumption (9),

$$
\leq \; 0
$$

in general, or

$$< \quad 0$$

yielding asymptotic global stability, in the positivity case with strictly positive amplification random functions and monotone increasing signal functions, as in the proof of the ABAM Theorem above, in particular equation (23). Q. E. D.

The RABAM Noise Suppression Theorem proves that the instantaneous mean-squared velocities $E[\dot{x}_i^2]$, $E[\dot{y}_j^2]$, and $E[\dot{m}_{ij}^2]$ decrease exponentially quickly to their lower bounds. The lower bounds are the underlying "unknown" instantaneous noise "variances," scaled by their corresponding "temperatures." These theoretical lower bounds on the RABAM second-order behavior are summarized by the following Lemma.

$$\textit{Lemma.} \quad E[\dot{x}_i^2] \;\geq\; T_i\sigma_i^2, \quad E[\dot{y}_j^2] \;\geq\; T_j\sigma_j^2 \quad E[\dot{m}_{ij}^2] \;\geq\; T_{ij}\sigma_{ij}^2 \;. \tag{31}$$

The Lemma makes explicit the need for the finite-variance assumption (10). The Lemma is proved by squaring both sides of (5), (6), and (7) (or (8)), taking expectations on both sides, factoring the cross-product expectation in each expression by the uncorrelatedness (independence) of "signal" and noise terms, then using the zero-mean assumption (9) to eliminate the cross-product terms and to obtain, for example, $E[n_i^2] = V[n_i]$. This proof technique is independent of the learning law used provided the noise is additive.

The Lemma captures the intuition that noise fluctuations drive system fluctuations. It also eliminates the possibility of ideal noise suppression. The best that can be achieved is the scaled variance of the underlying random disturbances. Hence at equilibrium —at *stochastic* equilibrium —the dynamical system still vibrates, still "jiggles" about the deterministic equilibrium. The RABAM Noise Suppression Theorem guarantees that equality is reached in (31), and reached exponentially quickly. In this sense *RABAM models optimally suppress noise in realtime.* This is surprising in its own right, but both the general RABAM Theorem and the RABAM Noise Suppression Theorem say more. They say it is impossible to (finitely) shake an ABAM equilibrium out of equilibrium, or, more accurately, out of its equilibrium shape.

Recall that the mean-squared velocities in (31) are also bounded below by the variances of the activation and synaptic variances, since, for example,

$$V[\dot{x}_i] \;=\; E[\dot{x}_i^2] \;-\; E^2[\dot{x}_i] \;\leq\; E[\dot{x}_i^2] \;. \tag{32}$$

The inequality (32) tells us that observed system fluctuations, mean-squared velocities, never underestimate the variance of state changes. It also tells us that at equilibrium, when equality holds in (31), the variances of the activation and synaptic velocities are bounded *above* by the temperature-scaled noise variances.

The Lemma does not imply, though it suggests, that the squared velocity processes are never less than the squared noise processes at every instant. It only implies that the inequality holds *on average* at every instant.

Pathologies can in principle occur where the instantaneous squared noise exceeds the instantaneous squared activation velocity. No problems can arise for synaptic velocities, as will be clear in the proof of the RABAM Noise Suppression Theorem. It will also be clear that a natural way to avoid such pathologies is to require that the random weight functions, $\frac{S_i'}{a_i}$ and $\frac{S_j'}{a_j}$, are "well-behaved." Well-behavedness here means that two integrals be kept nonnegative:

$$E[\frac{S_i'}{a_i}(\dot{x}_i^2 - T_i n_i^2)] \;\geq\; 0 \quad , \tag{33}$$

$$E[\frac{S_j'}{a_j}(\dot{y}_j^2 - T_j n_j^2)] \;\geq\; 0 \quad . \tag{34}$$

This condition is sufficient to prove the RABAM Noise Suppression Theorem. The pathologies it is meant to rule out can be seen by considering when, say, the random difference function $\dot{x}_i^2 - n_i^2$ is negative for a particular realization. If the positive random weight function $\frac{S_i'}{a_i}$ is sufficiently large, the overall integral, the expectation, could be negative. This would be true, to take an extreme example, if the random weight function behaved as a Dirac delta function. We assume the weight functions do not place too much weight on the (infrequent) negative realizations of the difference functions.

RABAM Noise Suppression Theorem. As the above RABAM systems (such as (5) - (7) or (5), (6), and (8)) converge exponentially quickly, the mean-squared velocities of neuronal activations and synapses decrease to their lower bounds exponentially quickly:

$$E[\dot{x}_i^2] \downarrow T_i\sigma_i^2,\ E[\dot{y}_j^2] \downarrow T_j\sigma_j^2,\ E[\dot{m}_{ij}^2] \downarrow T_{ij}\sigma_{ij}^2\ , \tag{35}$$

for strictly positive amplification functions a_i and a_j, monotone increasing signal functions, well-behaved weight ratios $\frac{S_i'}{a_i}$ and $\frac{S_j'}{a_j}$ (so (33) and (34) hold), sufficient smoothness to permit time differentiation within the expectation integral, and nondegenerate Hessian conditions.

Proof. As in the proof of the general RABAM Theorem, the smoothness condition is used to expand the time derivative of the bounded Lyapunov function $E[L]$:

$$\begin{aligned}
\dot{E}[L] &= E[\dot{L}] \\
&= E\{\sum_i S_i'\dot{x}_i[b_i - \sum_j S_j m_{ij}] + \sum_j S_j'\dot{y}_j[b_j - \sum_i S_i m_{ij}] \\
&\quad - \sum_i\sum_j \dot{m}_{ij}[-m_{ij} + S_iS_j]\} \qquad (36)\\
&= E\{-\sum_i \frac{S_i'}{a_i}\dot{x}_i^2 - \sum_j \frac{S_j'}{a_j}\dot{y}_j^2 - \sum_i\sum_j \dot{m}_{ij}^2 \\
&\quad + \sum_i \frac{S_i'}{a_i}\dot{x}_i\sqrt{T_i}\ n_i\ + \sum_j \frac{S_j'}{a_j}\dot{y}_j\sqrt{T_j}\ n_j \\
&\quad + \sum_i\sum_j \sqrt{T_{ij}}\ n_{ij}\dot{m}_{ij}\}\ , \qquad (37)
\end{aligned}$$

by using the positivity of the random amplification functions to eliminate the three terms in braces in (36) with (5), (6), and (7) respectively,

$$\begin{aligned}
&= E[\dot{L}_{ABAM}]\ +\ E[\sum_i \frac{S_i'}{a_i}\ T_i n_i^2\ + \sum_j \frac{S_j'}{a_j}\ T_j n_j^2 \\
&\quad + \sum_i\sum_j T_{ij}n_{ij}^2] \qquad (38)\\
&= -\sum_i E[\frac{S_i'}{a_i}(\dot{x}_i^2 - T_i n_i^2)] - \sum_j E[\frac{S_j'}{a_j}(\dot{y}_j^2 - T_j n_j^2)] \\
&\quad -\sum_i\sum_j [E[\dot{m}_{ij}^2]\ - T_{ij}\sigma_{ij}^2]\ . \qquad (39)
\end{aligned}$$

We already know from the proof of the general RABAM Theorem above that, in the positivity case assumed here, this Lyapunov function strictly decreases along system trajectories. So the system is asymptotically stable and equilibrates exponentially quickly. (Like remarks hold for different learning laws with appropriate conditions.) This itself, though, does not guarantee that each summand in the first two sums in (39) is nonnegative. The well- behavedness of the weight functions $\frac{S_i'}{a_i}$ and $\frac{S_j'}{a_j}$ guarantees this, for then (33) and (34) hold. The Lemma guarantees that the summands in the third sum, the synaptic sum, are nonnegative. Asymptotic stability ensures then that *each* summand in (39) decreases to zero exponentially quickly. This immediately gives the desired exponential decrease of the synaptic mean-squared velocities to their lower bounds. For the activation mean-squared velocities, we find at equilibrium:

$$E[\frac{S_i'}{a_i}\dot{x}_i^2] = E[\frac{S_i'}{a_i}T_i n_i^2] \quad , \tag{40}$$

with a like equilibrium condition for each neuron in the F_Y field. Then, since the integrands in (40) are nonnegative, the expectations can be peeled off almost everywhere[21] (except possibly on sets of zero probability):

$$\frac{S_i'}{a_i}\dot{x}_i^2 = \frac{S_i'}{a_i}T_i n_i^2 \quad \text{almost everywhere} \quad , \tag{41}$$

$$\text{or} \quad \dot{x}_i^2 = T_i n_i^2 \quad \text{almost everyhwhere} \quad . \tag{42}$$

Taking expectations on both sides now gives the desired equilibrium conditions:

$$E[\dot{x}_i^2] = T_i\sigma_i^2 \quad \text{and} \quad E[\dot{y}_j^2] = T_j\sigma_j^2 \quad . \tag{43}$$

Since these equilibrium values are reached exponentially quickly, and since the Lemma guarantees they are lower bounds on the mean- squared velocities, the RABAM Noise Suppression Theorem is proved. Q. E. D.

The RABAM Noise Suppression Theorem generalizes the squared-velocity condition of the ABAM Theorem. In the ABAM case, the instantaneous "variances" are zero. This is the deterministic case. The probability space has a degenerate sigma-algebra, which only contains two "events," the whole space and the empty set. So expectations disappear and the RABAM squared velocities are zero everywhere at equilibrium, as in the ABAM case.

V. RABAM Unbiasedness: Average RABAM Behavior is ABAM Behavior

The RABAM Noise Suppression Theorem implies that average RABAM equilibria become ABAM equilibria exponentially quickly. In estimation theory terms, RABAM systems are *unbiased* estimators of ABAM systems awash in noise. Unbiasedness can be established directly, and stochastically, by integrating the RABAM equations, taking expectations, and exploiting the zero-mean nature of Brownian motion[19]. This approach ignores the convergence-rate information provided by the RABAM Noise Suppression Theorem, which we now use to characterize the *asymptotic* unbiasedness of RABAM equilibria.

Consider again the ABAM model (11) - (13). For maximum generality, we can assume the ABAM model is a set of stochastic differential equations. At equilibrium the activations of the neurons in the field F_X, for example, obey the condition

$$b_i(x_i) = \sum_j S_j(y_j) m_{ij} \ , \tag{44}$$

since the amplification functions a_i are strictly positive. In the additive case the function $b_i(x_i)$ is a linear function of x_i; in particular, $b_i(x_i) = x_i - I_i$. Then at equilibrium the neuronal activation in (44) equals the sum of the ith neuron's synapse-weighted feedback input signals and external input,

$$x_i = \sum_j S_j m_{ij} + I_i \ . \tag{45}$$

In the *shunting* case[18], the neuronal activation value is restricted to a bounded interval. The equilibrium activation value then has the ratio form of a Weber law[7]. At equilibrium the synapses (synaptic efficacies) equal Hebbian products,

$$m_{ij} = S_i S_j \ , \tag{46}$$

in the signal Hebb case, or equal pre-synaptic signals

$$m_{ij} = S_j \ , \tag{47}$$

in the competitive learning case for the ith "winning" neuron, where S_j behaves approximately as a zero-one threshold signal function. We now prove that RABAM equilibria obey equilibrium conditions with the same average shape as (44) - (47).

Unbiasedness Corollary: Average RABAM equilibria are ABAM equilibria.

Under the assumptions of the RABAM Noise Suppression Theorem, $b_i(x_i)$ converges to $\sum_j S_j m_{ij}$ in the mean-square sense exponentially quickly. At equilibrium,

$$b_i(x_i) = \sum_j S_j m_{ij} \quad \text{with probability one,} \tag{48}$$

and similarly for the b_j functions. The synaptic relationships in (46) and (47) are similarly recovered exponentially quickly.

Proof. We reexamine the Lemma (31). Squaring the RABAM stochastic differential equations (5) - (7), taking expectations, and invoking the uncorrelatedness of the scaled noise terms and (9), gives

$$E[\dot{x}_i^2] = E[a_i^2[b_i - \sum_j S_j m_{ij}]^2] + T_i \sigma_i^2 \ , \tag{49}$$

$$E[\dot{y}_j^2] = E[a_j^2[b_j - \sum_i S_i m_{ij}]^2] + T_j \sigma_j^2 \ , \tag{50}$$

$$E[\dot{m}_{ij}^2] = E[(-m_{ij} + S_i S_j)^2] + T_{ij} \sigma_{ij}^2 \ , \tag{51}$$

or, if the noisy competitive learning law (8) replaces the noisy signal Hebb law (7), (51) is replaced with

$$E[\dot{m}_{ij}^2] = E[S_j^2 (S_i - m_{ij})^2] + T_{ij} \sigma_{ij}^2 \ . \tag{52}$$

The righthand side of each equation is nonnegative and the RABAM Noise Suppression Theorem implies that exponentially quickly the mean-squared "signal" terms decrease to zero. Since the weighting functions a_i^2 and a_j^2 are strictly positive, this proves the asserted mean-square convergence. The strict positivity of the weighting functions further implies[27] the equilibrium conditions

$$E[(b_i - \sum_j S_j m_{ij})^2] = 0 , \tag{53}$$

$$E[(b_j - \sum_i S_i m_{ij})^2] = 0 , \tag{54}$$

$$E[(-m_{ij} + S_i S_j)^2] = 0 . \tag{55}$$

The nonnegativity of the integrands in (53) - (55) then[27] allows the expectations to be removed almost everywhere (with probability one), yielding the equilibrium condition (48). Q.E.D.

An important extension of the corollary for stochastic estimation occurs in the additive case with random noisy input waveform $[I_1', \ldots, I_n' \mid J_1', \ldots, J_p']$. This random waveform, which may slowly vary with time, can be viewed as a set of random inputs in the additive random ABAM case perturbed with additive zero-mean finite-variance noise processes:

$$I_i' = I_i + v_i , \tag{56}$$

$$J_j' = J_j + w_j , \tag{57}$$

where $E[v_i] = E[w_j] = 0$. In the additive case, the new noise processes in (56) and (57) simply combine additively with the other zero-mean noise processes in the RABAM model. These new noise terms can likewise be scaled with different nonnegative deterministic annealing schedules. The new scaled noise terms, though, then entail rescaling the variances in the Lemma and the RABAM Noise Suppression Theorem, in addition to appropriatly adjusting the variance terms themselves. Simulations [Fred Watkins, personal communication] have verified the predicted (asymptotic) unbiasedness for several types of zero-mean noise processes.

Formally, the additive RABAM then takes the shape of constant amplification functions and $b_i^R = x_i - I_i'$ and $b_j^R = y_j - J_j'$, while the additive ABAM takes the shape $b_i^A = x_i - I_i$ and

$b_j^A = y_j - J_j$, where again I_i and J_j may be random. Then, at equilibrium, (44), (45), (48), and (56) give

$$
\begin{aligned}
E[x_i^R - x_j^A] &= E[b_i^R + I_i' - b_j^A - I_i] \\
&= E[b_i^R - b_i^A] + E[v_i] \\
&= 0 \quad ,
\end{aligned}
$$

and similarly for the y_j activations.

Acknowledgement

This research was supported by the Air Force Office of Scientific Research (AFOSR-88-0236).

References

[1] Carpenter, G.A., Grossberg, S., "ART 2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns," *Applied Optics*, vol. 26, no. 23, 4919 - 4930, 1 December 1987.

[2] Cohen, M.A., Grossberg, S., "Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks," *IEEE Trans. Systems, Man, Cybernetics*, vol. SMC-13, 815 - 826, Sept./Oct. 1983.

[3] Geman, S., Hwang, C., "Diffusions for Global Optimization," *SIAM Journal of Control and Optimization*, vol. 24, no. 5, 1031 -1043, September 1986.

[4] Gluck, M.A., Parker, D.B., Reifsnider, E., "Some Biological Implications of a Differential-Hebbian Learning Rule," *Psychobiology*, vol. 16, no. 3, 298 - 302, 1988.

[5] Gluck, M.A., Parker, D.B., Reifsnider, E.S., "Learning with Temporal Derivatives in Pulse-Coded Neuronal Systems," *Proc. November 1988 IEEE Neural Information Processing Systems (NIPS) Conference*, Denver, CO, San Mateo, CA: Morgan Kaufman, in press,1989.

[6] Grossberg, S., *Studies of Mind and Brain*, Boston: Reidel, 1982.

[7] Grossberg, S., "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures," *Neural Networks*, vol. 1, no. 1, 17 - 61, 1988.

[8] Hirsch, M.W., Smale, S., *Differential Equations, Dynamical Systems, and Linear Algebra*, New York: Academic Press, 1974.

[9] Hirsch, M.W., *Differential Topology*, Springer-Verlag, 1976.

[10] Hodgkin, A.L., Huxley, A.F., "A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve," *Journal of Physiology*, vol. 117, 500 - 544, 1952.

[11] Hopfield, J.J., "Neural Networks with Graded Response have Collective Computational Properties like Those of Two-State Neurons," *Proc. National Academy of Science, USA*, vol. 81, 3088 - 3092, 1984.

[12] Klopf, A.H., "A Neuronal Model of Classical Conditioning," *Psychobiology*, vol. 16, no. 2, 85 - 125, 1988.

[13] Kosko, B., "Differential Hebbian Learning," *Proc. American Institute of Physics: Neural Networks for Computing*, 277 - 282, April 1986.

[14] Kosko, B., "Adaptive Bidirectional Associative Memories," *Applied Optics*, vol. 26, no. 23, 4947 - 4960, 1 December 1987.

[15] Kosko, B., "Bidirectional Associative Memories," *IEEE Trans. Systems, Man, Cybernetics*, vol. SMC-18, 49 - 60, Jan./Feb.1988.

[16] Kosko, B., "Feedback Stability and Unsupervised Learning," *Proc. 2nd IEEE International Conference on Neural Networks (ICNN-88)*, vol. I, 141 - 152, July 1989.

[17] Kosko, B., "Hidden Patterns in Combined and Adaptive Knowledge Networks," *Int'l Jour. of Approximate Reasoning*, vol. 2, no. 4, 377 - 393, October 1988.

[18] Kosko, B., "Unsupervised Learning in Noise," *Proc. International Joint Conference on Neural Networks (IJCNN-89)*, June 1989.

[19] Maybeck, P.S., *Stochastic Models, Estimation, and Control*,vol. 2, Academic Press, 1982.

[20] Parker, T.S., Chua, L.O., "Chaos: A Tutorial for Engineers," *Proceedings of the IEEE*, vol. 75, no. 8, 982 - 1008, August 1987.

[21] Rudin, W., *Real and Complex Analysis*, second edition, McGraw-Hill, 1974.

[22] Rumelhart, D.E., Zipser, D., "Feature Discovery by Competitive Learning," *Cognitive Science*, vol. 9, 75 - 112, 1985.

[23] Skorokhod, A.V., Studies in the Theory of Random Processes, Reading, MA: Addison-Wesley, 1965.

[24] Thom, R., Structural Stability and Morphogenesis, Reading, 1975.