Stability and Optimization Analyses of the Generalized Brain-State-in-a-Box Neural Network Model

Richard M. Golden

University of Texas at Dallas

Richard M. Golden,

University of Texas at Dallas, School of Human Development,

GR41, Box 830688, Richardson, Texas, 75083–0688,

golden@utdallas.edu, (214) 690–2423

Running head: Generalized Brain-State-in-a-Box Neural Network

Abstract

Anderson's Brain-State-in-a-Box (BSB) neural network model (J. A. Anderson, J. W. Silverstein, S. A. Ritz, and R. S. Jones, 1977, *Psychological Review*, **84**, 413–451) may be viewed as a version of Hopfield's (1982, *Proceedings of the National Academy of Sciences, USA, 79,* 2554–2558) two-state neural model with continuous rather than discrete states and synchronous updating. A generalization of the BSB model which allows each model neuron in the system to have a bias and its own maximum and minimum firing rate is considered. Moreover, the generalization does not require that the matrix of connection weights be symmetric. A method of choosing the step constant of the algorithm is then described which guarantees that the generalized discrete-time BSB model: (i) approaches the largest set of system equilibrium points, and (ii) minimizes the Boltzmann machine quadratic energy function. A simple test (essentially a vector by matrix multiplication) for checking for asymptotically stable equilbrium points is also provided. Finally, a special case of the generalized BSB model is shown to be applicable to a particular type of optimal control problem.

Anderson, Silverstein, Ritz, and Jones (1977) proposed an interesting neural network model of categorization known as the Brain-State-in-a-Box (BSB) model. This simple model consisted of a set of d model neurons or *units* which were symmetrically interconnected, and fed back upon themselves. Each unit at each instant in time simultaneously computed a weighted sum of its inputs, and then used this weighted sum to update its activation value. In addition, a simple non-linearity was introduced so that the activation value of each unit remained bounded between its maximum and minimum activation level. The state of the neural network was represented as a pattern of activation over the neuronal units which was amplified if the activation pattern was ''familiar'' to the system, and squelched if the activation pattern was not ''familiar''.

Golden (1986) provided an analysis of the behavior of the BSB model in discrete-time, and discussed explicit conditions which would guarantee that the BSB model decreased the value of a specific ''energy'' function for all non-stable activation patterns. Cohen and Grossberg (1983) and Hopfield (1984) analyzed a continuous-time version of the BSB model using a Liapunov function approach as well. Hui and Zak (1992) (also see Greenberg, 1988) have studied the stability properties of the BSB model in the non-symmetric but "diagonally dominant" case but did not use an "energy" function in their analyses. The Brain-State-in-a-Box model has been used to model a variety of categorization phenomena in areas of speech, vision, and cognitive processes (e.g., Anderson et al., 1977; Anderson, Murphy, & Golden, 1986; Golden,

1986b; Kawamoto & Anderson, 1985). Stochastic relaxation connectionist models which minimize the BSB model energy function include the Boltzmann machine (Ackley, Hinton, & Sejnowski, 1985) and Harmony theory (Smolensky, 1986).

In this paper, Anderson's Brain-State-in-a-Box (BSB) model (Anderson, Silverstein, Ritz, and Jones, 1977) will be generalized in a variety of different ways. A global convergence theorem for this generalized algorithm in discrete-time is presented following the approach of Golden (1986a). A simple test (essentially a vector by matrix multiplication) for checking for asymptotically stable equilibrium points is also presented. The results in this paper provide a better understanding of applications of the BSB model by considering a related algorithm where the matrix is not symmetric, the case where the BSB model has "biases", and the case where the maximum and minimum "firing rates" of the units in the model can be different for each unit in the system. The resulting generalized BSB model is called the Generalized-Brain-State-in-a-Box (GBSB) model, and was developed in order to deal with an optimal control problem associated with modelling story comprehension and recall considered by Golden and Rumelhart (1991, in press). Before describing the general issues which confronted Golden and Rumelhart (1991, in press), it is worth noting that the optimal control problem discussed here is fairly general and should be applicable to many psychological domains such as speech perception and motor control (e.g., Jordan, in press, Barto, 1989).

Golden and Rumelhart (1991, in press) approached the problem of modelling story comprehension and recall within the framework of a high-dimensional state space known as ''situation'' state space. A point in situation state space consists of a collection of *n* facts about the world, each of which can be classified as either being ''present'' or ''absent''. The reader's current mental state is therefore modelled as a single point in situation state space. At some later point in time, the reader's mental state is modelled as another point in situation state space and it is convenient to order these points according to a time index. Thus, the ''unfolding'' or ''evolution'' of the reader's mental state as a function of time, may be represented as an ordered sequence of points or *trajectory* in situation state space. Some trajectories in situation state space will be more likely to occur than others. Accordingly, the reader's world knowledge is modelled as a probability mass function which assigns a non-zero probability to every trajectory in the situation state space.

Within this framework, the problem of story comprehension is formally viewed as finding a trajectory in situation state space which is most probable with respect to the reader's world knowledge (i.e., the reader's probability mass function). During both comprehension and recall, the reader searches for highly probable trajectories by running their probabilistic mental models forward and backwards in time in order to construct a most probable trajectory which is consistent with either the constraints imposed by the story or another set of retrieval constraints imposed by the reader's own memory system. An instantiation of the GBSB model yields an algorithm

designed to model a forward and backwards search process for a recurrent network algorithm which searches for the most probable trajectory with respect to a very specific probability distribution.

Definition and Interpretation of the GBSB Model Algorithm

Consider a collection of d model neurons where the activation of the ith neuron is $x(i)$. The pattern of activation over the model neurons may be expressed as a d-dimensional vector X. Each of the neurons has a maximum firing rate or activation level of F_i, and a minimum firing rate or activation level of f_i. Thus, the ith element of X, $x(i)$, must satisfy:

$$f_i \leq x(i) \leq F_i \tag{1}$$

where $f_i < F_i$. This simple non-linearity gives rise to the interesting categorization performance of the model, but also makes the GBSB model extremely hard to formally analyze.

Notice that (1) implies that every activation pattern over the *d* model neurons is constrained to a ''box'' in a *d*-dimensional space. For example, Figure 1 shows a two-dimensional (two neuron) case where points corresponding to different activation patterns are plotted. Because the state of the system is constrained to a *d*-dimensional box, the system is referred to as the Brain-State-in-a-Box model.

Insert Figure 1 about here

The dynamics of the system are also fairly straightforward. Each of the *d* model neurons simultaneously adds a weighted sum of its inputs, a weighted sum of its outputs, and a bias to its current activation value. In addition, if the resulting activation value exceeds the unit's maximum activation value F_i, then the resulting activation value is truncated to the value F_i. Similarly, if the resulting activation value is less than the unit's minimum activation value f_i, then the resulting activation value is truncated to the value f_i. Figure 1 also provides insights into the dynamical behavior of the GBSB model. The vectors in Figure 1 are referred to as *delta* vectors, and are defined by the equation: $\delta^k = X^{k+1} - X^k$ where X^k is the *d*-dimensional activation pattern at the *k*th iteration of the GBSB algorithm. Thus, if δ^k is a vector of zeros, then the activation pattern over the units has "stabilized" and reached a system equilibrium point.

To formalize these ideas, define the following "linearized" sigmoidal function, $S_i(\alpha)$, as $S_i(\alpha) = \alpha$ for $f_i \leq x(i) \leq F_i$, $S_i(\alpha) = F_i$ for $x(i) > F_i$, and $S_i(\alpha) = f_i$ for $x(i) < f_i$.Formally, the dynamics of the GBSB model may be expressed by the following equation:

$$x^{k+1}(i) = S_i\Big[x^k(i) - \gamma\rho_i\eta^k(i)\Big] \tag{2}$$

where $x^k(i)$ is the *activation*(or equivalently the real-valued scalar state) of the *i*th unit at the *k*th iteration of the algorithm, γ is a strictly positive scalar constant, $0 < \rho_i < 1$ is also a strictly positive scalar constant for *i=1...d*, and the *netinput*, $\eta^k(i)$, to the *i*th unit at iteration *k* of the algorithm is defined as:

$$\eta^k(i) = -\left[\sum_{j=1}^{d} w_{ij}x(j) + \sum_{k=1}^{d} w_{ki}x(k) + b_i\right]. \tag{3}$$

where w_{ij} is the *weight* connecting the *j*th unit to the *i*th unit in the system, and b_i is the bias of the *i*th unit.

Two important specializations of the algorithm are now considered. First, suppose that the matrix weights are symmetrical (i.e., $w_{ij} = w_{ji}$ for $i \neq j$), then (2) and (3) reduce to the standard BSB equations described by Anderson et al. (1977) (also see Golden, 1986) with the exception of the addition of allowing individual units to have their own biases and maximum and minimum firing rates.

The second important specialization of the GBSB model which was exploited by Golden and Rumelhart (1991) is now described. Consider a GBSB model with *d* model neurons. Divide the group of *d* units into *M* groups of *n* units each so that $d = Mn$. The *n*-dimensional activation pattern over the *t*th group of units with $t = 1, \ldots, M$ represents a *situation vector*, $X(t)$, at time *t*. A sequence of *M* *n*-dimensional situation vectors with $X(1), \ldots, X(M)$ corresponds to a *trajectory* as

described by Golden and Rumelhart (1991). Thus, the *d*-dimensional state vector corresponding to a pattern of activation over the *d*-units will have the form:

$$X = [X(1), \ldots, X(M)].$$

Since it is assumed that the knowledge required to move from situation $X(t-1)$ to situation $X(t)$ is time-invariant, the connectivity matrix should have the form of a *d*-dimensional square matrix with a single *n*-dimensional square submatrix, *v*, duplicated *M-1* times on the off-main diagonal so that $X(t) = vX(t-1)$ for $t = 1, \ldots, M-1$. For example, suppose that $M = 3$, then:

$$w = \begin{bmatrix} 0 & 0 & 0 \\ v & 0 & 0 \\ 0 & v & 0 \end{bmatrix} \tag{4}$$

where *0* refers to an *n*-dimensional square submatrix in this case, and the matrix *w* is a square *d*-dimensional matrix with $d = 3n$. Golden and Rumelhart (1991; in press) view this version of the GBSB model as a kind of *approximation* to the dynamics of a recurrent network consisting of units with complex high-order temporal response properties. That is, their goal was to develop a mathematically tractable and psychologically plausible model which captured the general flavor of the more complex recurrent network.

Finally note that the seemingly innocuous non-linearity or "box" defined by (1) gives rise to the important categorization properties of the GBSB model. Figure 2 shows how two very different stimuli may be mapped into the same corner of

the box, while Figure 3 shows how two very similar stimuli may be mapped into different corners of the box. Unfortunately, the "box" non-linearity also makes a complete formal analysis of the GBSB model's behavior extremely difficult and complex. In the next section, an approach to analyzing the behavior of the GBSB model is described, and the main theoretical results and heuristic proofs of special cases of those results are briefly sketched.

Insert Figure 2 about here

Insert Figure 3 about here

Overview of the Theoretical Analyses and Results

As previously noted, the GBSB model is a complex non-linear dynamical system and so conventional methods of linear analysis are not applicable. For example, one could try to "linearize" the GBSB model using Liapunov's first or indirect method (Luenberger, 1979, pp. 324–328) in order to convert the GBSB model into a linear system over certain regions of the state space, but this approach will not work since the non-linearity is continuous but not differentiable. Another approach is to attempt to apply a generalization of Liapunov's second or direct method known as the Invariant Set Theorem (Luenberger, 1979, pp. 332–349). An *invariant set* is

simply a region of the state space with the property that if the state of a specific non-linear dynamical system enters this region, then the system state can never leave the invariant set. Perhaps the simplest example of an invariant set is an equilibrium point which is defined to be an invariant set consisting of exactly one point. This method of analysis for neural networks was originally popularized by Hopfield (1982) and is essentially based upon the following ideas.

First, the goal of completely analyzing the non-linear dynamical system is abandoned in favor of making some weaker statements about the system's behavior. In particular, the goal of an Invariant Set Theorem type of analysis is to characterize the dynamics of the system in terms of the system's equilibrium points (more technically, the system's "invariant sets"). This is done by proving that all sequences of activation patterns generated by the model converge to the set of system equilibrium points, and then developing criteria for deciding if a given equilibrium point is an "attractor". An *attractor* equilibrium point or equivalently an *asymptotically stable* equilibrium point has the property that if the system state is initiated sufficiently close to such an equilibrium point, then the system state will always converge to that equilibrium point. For example, consider modelling the dynamics of a turbulent river. Although the detailed dynamics of the water flow will probably never be well understood, important insights into its behavior can be obtained by identifying the location of the vortices of the river. These vortices correspond to the asymptotically stable equilibrium points of an Invariant Set Theorem type of analysis.

Proving the existence of invariant sets and showing exactly which invariant sets will be attractors remains a difficult problem. For this reason, the Invariant Set Theorem is based upon the idea of a "summarizing function" (Luenberger, 1979, pp. 319–320) which summarizes the non-linear dynamical system's performance. Some heuristic arguments which provide intuitions about how a version of the Invariant Set Theorem is applicable in the analysis of the GBSB model are now provided.

A heuristic statement and proof of the Invariant Set Theorem. Suppose that the sequence of activation patterns,$X^1, \ldots, X^k, \ldots$, is generated according to the non-linear dynamical system:

$$X^{k+1} = T(X^k)$$

where $T(X^k)$ is a continuous function of X^k, and X^1 is a member of some region of the state space known as Ω. Also assume that $X^1 \in \Omega$, and Ω is a closed and bounded region of the state space so Ω itself is an invariant set with respect to the non-linear dynamical system. In the GBSB model, Ω corresponds to the "box" and the mapping $X^{k+1} = T(X^k)$ corresponds to the GBSB model algorithm defined in (2) and (3).

In order to show that the sequence of activation patterns converges to an invariant set, a continuous function,$E(X^k)$, of any activation pattern, X^k, is constructed such that:

$$E(X^{k+1}) < E(X^k)$$

provided that X^k is not a system equilibrium point.

Then, since Ω is closed and bounded, $E(X)$ has a lower bound over Ω. Now since $E(X^k)$ has a lower bound and $E(X^k)$ must decrease in value if X^k is not an equilibrium point, this implies that $E(X^k)$ must converge to a constant value as k increases.

The sequence of vectors, X^k, thus converges to the set of points, Γ, which satisfy $E(X^{k+1}) = E(X^k)$. And Γ must contain only system equilibrium points because if Γ contained a sequence of points, X^k, X^{k+1} such that $X^{k+1} \neq X^k$ this would imply that $E(X^{k+1}) < E(X^k)$. But then this would lead to a contradiction since by the definition of Γ, $E(X^{k+1}) = E(X^k)$.

A heuristic statement and proof of the GBSB Theorem. Notice that an important fringe benefit of an Invariant Set Theorem analysis is the identification of a summarizing function, $E(X)$, which summarizes the non-linear dynamical system's performance. Such functions are also known as *energy* or *Liapunov functions* in the dynamical system and neural network literature (Golden, 1986; Hirsh, 1989).

One important goal of this paper is to formally show that the GBSB model defined in (2) and (3) may be analyzed using the following energy function:

$$E(X) = -\left[\sum_{i=1}^{d}\sum_{j=1}^{d} x(i)x(j)w_{ij}\right] - \sum_{i=1}^{d} b_i x(i) \tag{5}$$

where $X = [x(1)...x(d)]$ is a real-valued vector, w_{ij} is the *ij*th element of a real matrix *W*, and b_i is the *i*th element of a real vector *b*. It is not necessary that *W* be a symmetric matrix.

Now since $E(X)$ is a second-order polynomial in X, an "exact" multi-dimensional Taylor series expansion (Marsden & Tromba, 1981) of $E(X)$ about the point X^k may be done to obtain:

$$E(X^{k+1}) - E(X^k) = [dE/dX^k]^T \delta^k - (1/2){\delta^k}^T [W + W^T]\delta^k$$

where $\delta^k = X^{k+1} - X^k$. Now since X^{k+1} is generated from X^k using the GBSB algorithm in (2) and (3), it is clear that by choosing the step constant,γ, of the GBSB algorithm at iteration *k* to be sufficiently small that the difference vector δ^k can be made as small as desired. Thus, γ may be chosen so that the quadratic term in the above Taylor series expansion is negligible so that we have:

$$E(X^{k+1}) - E(X^k) \approx [dE/dX^k]^T \delta^k.$$

and noting that $\eta^k(i)$ in (2) and (3) is the *i*th element of dE/dX^k, if

$$\left[dE/dX^k\right]^T \delta^k = \sum_{i=1}^{d} \eta^k(i)\delta^k(i) < 0, \tag{6}$$

the energy function $E(X^k)$ will decrease.

Now consider the special case where the system state is strictly in the box's interior. Figure 1 depicts this situation when the system state moves along the path

indicated by delta vector δ^1. More formally, referring to (2), this corresponds to the case where

$$x^{k+1}(i) = x^k(i) - \gamma\rho_i\eta^k(i)$$

and so

$$\delta^k(i) = x^{k+1}(i) - x^k(i) = -\gamma\rho_i\eta^k(i).$$

Substitution of $\delta^k(i) = -\gamma\rho_i\eta^k(i)$ into (6) implies that the energy function $E(X)$ must decrease if γ is sufficiently small and positive so that the Taylor series approximation is valid.

Now suppose that X^k and X^{k+1} both lie in the same "wall" of the box. Figure 1 depicts this situation where the system state moves along the path defined by delta vectors δ^2 and δ^3 of Figure 1. This latter case corresponds to the situation where some subset of elements in the system state vector have obtained either their minimum or maximum values, while the remaining elements are free to vary. Again, close inspection of (2) and (3) will show that the GBSB model still satisfies (6) when the system state is not an equilibrium point. Note that additional analysis is required to decide if a step constant can be found such that (6) is satisfied when the system state moves from the interior of the box into a wall of the box.

Although these analyses provide important insights into why the GBSB model may be viewed as an optimization algorithm, these analyses provide only heuristic insights into the stability properties of the GBSB model since the energy function is

only non-increasing as the GBSB system evolves provided that the step constant γ is sufficiently small so that the Taylor series expansion is valid.

The surprising result of the BSB Energy Minimization Theorem states that if $W + W^T$ is positive semidefinite or if the step constant $\gamma < 2/|\lambda_{min}|$ where λ_{min} is the smallest negative eigenvalue of $W + W^T$, that $E(X^{k+1}) < E(X^k)$ if X^k is not a system equilibrium point. Or in other words, the step constant γ of the GBSB algorithm does not have to be chosen to be "infinitely small" at each step of the algorithm, but can be chosen to be a finite constant based upon the smallest eigenvalue of $W + W^T$. When γ is chosen in this way, then at *every iteration* of the GBSB algorithm (without changing the value of γ), the energy function *must decrease* provided that a system equilibrium point has not been reached. These observations in conjunction with the Invariant Set Theorem may be used to prove that any initial state (i.e., activation pattern) in the GBSB model must eventually converge to the set of system equilibrium points (i.e., converge to the set of "vortices").

Note that the phrase "converge to the largest set of system equilibrium points" implies that if an initial state of the GBSB algorithm is initiated sufficiently close to an isolated equilibrium point, then the system state will converge to that equilibrium point provided that the step constant of the algorithm is sufficiently small.

After the GBSB energy minimization theorem is stated and proved, the issue of how to design learning algorithms which make desired activation patterns equilibrium

points or "vortices" of the GBSB model dynamical system is considered. Define the linear response of a unit using (2) to be

$$r^k(i) = -\gamma\rho\eta^k(i)$$

and the system linear response as the vector R^k whose *i*th element is $r^k(i)$. If an activation pattern is a "corner" of the d-dimensional box and if the system's linear response drives the system state "out-of-the-box", then that corner activation pattern must be an asymptotically stable equilibrium point (i.e., an "attractor"). Figure 4 depicts this situation in a graphical manner. Anderson, Golden, and Murphy (1986), for example, used the auto-associative Widrow-Hoff learning rule to force the BSB model's linear response to hypercube corners to point "out-of-the-box." Such hypercube corners were then interpreted as cognitive categories or "concepts."

Insert Figure 4 about here

The Generalized BSB Theorem

The following theorem states that the energy function $E(X)$ in (5) is minimized subject to the constraints imposed by (1), and that the generalized BSB algorithm will converge to the set of system equilibrium points.

Generalized BSB Theorem. Consider a generalized BSB algorithm of the form of (2) and (3), with energy function $E(X)$ of the form of (5). Let λ_{min} be the smallest

eigenvalue of $W + W^T$. If either $\lambda_{min} \geq 0$ or

$$\gamma < 2/|\lambda_{min}|$$

where λ_{min} is the smallest negative eigenvalue of $W + W^T$, then:

i. $E(X^{k+1}) < E(X^k)$ for all k such that $X^{k+1} \neq X^k$

ii. all trajectories of the generalized BSB algorithm converge to the largest set of system equilibrium points.

Statement and Proof of Generalized BSB Lemma

The following lemma is necessary to prove the generalized BSB Theorem.

Generalized BSB Lemma. Consider a generalized BSB algorithm of the form of (2) and (3), with energy function $E(X)$ of the form of (5). If

$$\delta^k(i) = x^{k+1}(i) - x^k(i) \neq 0,$$

then an $0 < \alpha_{ik} \leq 1$ exists such that:

$$x^{k+1}(i) = S_i\Big[x^k(i) + r^k(i)\Big] = x^k(i) + \alpha_{ik} r^k(i) \tag{7}$$

Intuitively, α_{ik} compensates for the "overshoot" of the algorithm when it attempts to leave the box by "scaling down" the linear response of the algorithm.

Proof. Note that for $r^k(i) \neq 0$, we can solve for α_{ik} algebraically and find that:

$$\alpha_{ik} = \Big(S_i\Big[x^k(i) + r^k(i)\Big] - x^k(i)\Big)/r^k(i) \tag{8}$$

We now will show that $0 < \alpha_{ik} \leq 1$, and that $r^k(i) \neq 0$. We do this by considering three cases, and noting that $F_i - x^k(i) \geq 0$ and $f_i - x^k(i) \leq 0$ for all *i* and *k*.

Case 1: $f_i \leq x^k(i) + r^k(i) \leq F_i$. In this case, choose $\alpha_{ik} = 1$ to satisfy (7). Note that $r^k(i) = \delta^k(i) \neq 0$ in this case.

Case 2: $x^k(i) + r^k(i) > F_i$. Note that:

$$r^k(i) > F_i - x^k(i) > 0$$

so that $\alpha_{ik} > 0$ and $\alpha_{ik} = [F_i - x^k(i)]/r^k(i) < 1$ using (8).

Case 3: $x^k(i) + r^k(i) < f_i$. Note that:

$$r^k(i) < f_i - x^k(i) < 0$$

so that $\alpha_{ik} > 0$ and $\alpha_{ik} = [f_i - x^k(i)]/r^k(i) < 1$ using (8).

Q.E.D.

Proof of the Generalized BSB Theorem

In this section the generalized BSB Theorem will be proved following the development of Golden (1986a). It will be shown that the conditions of the generalized BSB Theorem guarantee that at each iteration of the algorithm the change in $E(X)$ will always be strictly negative until the algorithm reaches an equilibrium point. Then, since the state of the algorithm is confined to the inside (and surface) of a box, the convergence result immediately follows. The proof begins by considering a particular iteration of the algorithm which is not initiated at a system equilibrium point.

Note that by the generalized BSB Lemma if $x^{k+1}(i) \neq x^k(i)$, (2) and (3) can always be rewritten in the following form:

$$x^{k+1}(i) = x^k(i) + \alpha_{ik} r^k(i) \tag{9}$$

where $0 < \alpha_{ik} \leq 1$ and

$$r^k(i) = -\gamma \rho_i \eta^k(i) = -\gamma \rho_i dE/dX^k(i).$$

The gradient or first derivative of $E(X)$ in (5) with respect to X evaluated at X^k is column vector $dEdX^k$ whose *i*th element is$\eta^k(i)$ in (3). The Hessian or matrix second derivative of $E(X)$ with respect to the vector X is simply:

$$d^2E/dX^2 = -\left(W + W^T\right) \tag{10}$$

Now since $E(X)$ is a second-order polynomial in X, an "exact" multi-dimensional Taylor series expansion of $E(X)$ about the point X^k may be done to obtain:

$$E\left(X^{k+1}\right) - E\left(X^k\right) = \left[dE/dX^k\right]^T \delta^k - (1/2)\left(\delta^k\right)^T \left[W + W^T\right] \delta^k \tag{11}$$

where $\delta^k = X^{k+1} - X^k$. Thus, if $X^{k+1} \neq X^k$,

$$E\left(X^{k+1}\right) - E\left(X^k\right) \leq \left[dE/dX^k\right]^T \delta^k - (1/2)\lambda_{min}\left(\delta^k\right)^T \delta^k \tag{12}$$

where λ_{min} is the minimum eigenvalue of $W + W^T$ using Lemma 1 from Golden (1986a). Now note from (9) that:

$$\delta^k(i) = x^{k+1}(i) - x^k(i) = -\gamma \alpha_{ik} \rho_i dE/dx^k(i),$$

and substituting for δ^k into the right-hand side of (12) if:

$$-\gamma \sum_{i=1}^{d} \alpha_{ik}\rho_i \left[dE/dx^k(i)\right]^2 < (\gamma^2/2)\lambda_{min} \sum_{i=1}^{d} \alpha_{ik}^2 \rho_i^2 \left[dE/dx^k(i)\right]^2 \qquad (13)$$

then the right-hand side of (12) will always be strictly negative, and the theorem will be essentially proved. So the problem now is to determine the specific conditions under which (13) is true.

The strict inequality holds in (13) under the following conditions. First note that $0 < \rho_i < 1$ and if $\delta^k(i) \neq 0$ by the generalized BSB Lemma $0 < \alpha_{ik} \leq 1$, and

$$1 > \alpha_{ik}\rho_i > \alpha_{ik}^2\rho_i^2 > 0.$$

Now since X^k is not an equilibrium point of the algorithm, at least one *i* exists such that

$$\delta^k(i) = -\gamma\alpha_{ik}\rho_i dE/dX^k(i) \neq 0.$$

Thus, at least one *i* exists such that both $dE/dx^k(i) \neq 0$ and $\alpha_{ik} \neq 0$ which implies since $\alpha_{ik}\rho_i > \alpha_{ik}^2\rho_i^2$:

$$\sum_{i=1}^{d} \alpha_{ik}\rho_i \left[dE/dx^k(i)\right]^2 > \sum_{i=1}^{d} \alpha_{ik}^2\rho_i^2 \left[dE/dx^k(i)\right]^2 > 0 \qquad (14)$$

so if $\lambda_{min} \geq 0$ the left side of (13) is strictly negative and the right side of (13) is strictly positive which shows that in this case $E(X^{k+1}) < E(X^k)$.

The only remaining case is $\lambda_{min} < 0$. If γ is chosen in this case (solve (13) for γ) such that:

$$\gamma < [2/|\lambda_{min}|]\left(\sum_{i=1}^{d} \alpha_{ik}\rho_i \left[dE/dx^k(i)\right]^2\right) / \left(\sum_{i=1}^{d} \alpha_{ik}^2 \rho_i^2 \left[dE/dx^k(i)\right]^2\right) \qquad (15)$$

then the inequality in (13) will always be satisfied. Note that the denominator of (15) is strictly positive because of (14).

Also since (14) holds, inspection of (15) shows that if γ is chosen such that:

$$\gamma < 2/|\lambda_{min}|$$

(13) must always be satisfied which shows $E(X^{k+1}) < E(X^k)$.

Finally note that because the state space is closed and bounded, and $E(X)$ is strictly decreasing at each step of the algorithm when the algorithm is not initiated at an equilibrium point of the system, all sequences of system states generated by the generalized BSB algorithm converge to the largest set of system equilibrium points by direct application of the Invariant Set Theorem (see Luenberger, 1979; Golden, 1986a; Hirsch, 1989, for reviews of the Invariant Set Theorem).

Q.E.D.

An Asymptotic Stability Test

An asymptotically stable equilibrium point is a state X* with the property that all sequences of system states generated by the dynamical system in a "sufficiently close"

vicinity of X* are guaranteed to approach X* as time increases (e.g., Luenberger, 1979; Hirsch, 1989). In practice, it is desirable to have a quick and easy test to decide if a given state of the generalized BSB model is both an asymptotically stable equilibrium point and a strict local minimum of the energy function.

A state C is defined as a *hyperbox vertex* (i.e., a vector whose elements are either f_i or F_i). A state C^* is the unique anti-hyperbox vertex of C. The *anti-hyperbox vertex*, C^*, of C is a state vector such that $C^*(i) = F_i$ if $C(i) = f_i$, and $C^*(i) = f_i$ if $C(i) = F_i$ where $C(i)$ is the *i*th element of *C*.

Proposition. Consider a generalized BSB algorithm of the form of (2) and (3), with energy function $E(X)$ of the form of (5). Let *C* be a hyperbox vertex such that

$$[W + W^T]C + b = G(C - C^*)$$

where the diagonal matrix *G* has elements $g_{ii} > 0$. Let λ_{min} be the smallest eigenvalue of $W + W^T$. If either $\lambda_{min} \geq 0$ or

$$\gamma < 2/|\lambda_{min}|$$

where λ_{min} is the smallest negative eigenvalue of $W + W^T$, then the vertex *C* is both an asymptotically stable equilibrium point and a strict local minimum of $E(X)$.

Proof. We only need to show that $E(C) < E(q)$ for all $q = C + e$ in or on the surface of the hyperbox for $|e|$ sufficiently small since this condition implies *C* is a strict local minimum. Once it is established that *C* is a strict local minimum,

the asymptotic stability property of *C* immediately follows if the generalized BSB theorem is applicable.

The error vector, *e*, may be expressed in terms of *C* and C^* using

$$e = -H(C - C^*)$$

where *H* is a diagonal matrix of non-negative elements with at least one positive element, and the *i*th element on the main diagonal of H is given by:

$$h_{ii} = [C(i) - q(i)]/[C(i) - C^*(i)] \geq 0.$$

Using a first-order Taylor series expansion of $E(X)$ about the point C and evaluating at q we have:

$$E(q) - E(C) = [dE/dC]^T e + O\left(|e|^2\right) \tag{16}$$

and substituting for dE/dC (i.e., the first derivative of $E(X)$ with respect to X as defined in (3) and then evaluated at *C*) we have:

$$E(q) - E(C) = -\left[\left(W + W^T\right)C + b\right]^T e + O\left(|e|^2\right) \tag{17}$$

Now substitute $[W + W^T]C + b = G(C - C^*)$ and $e = -H(C - C^*)$ into (17) to obtain:

$$E(q) - E(C) = [G(C - C^*)]^T H(C - C^*) + O\left(|e|^2\right) \tag{18}$$

or equivalently:

$$E(q) - E(C) = (C - C^*)^T GH(C - C^*) + O\Big(|e|^2\Big) \tag{19}$$

The matrix GH is positive semidefinite with at least one positive eigenvalue so for sufficiently small $|e|$, we have $E(C) < E(q)$.

Q.E.D.

An Optimal Control Theory Application of the Generalized BSB Model

In this section, we define a general optimal control problem and show how the BSB model may be used to solve that problem. Consider a stochastic dynamical system (first-order Markov chain) whose states are *n*-dimensional binary feature vectors. Thus, the probability of moving from state $X(t-1)$ to state $X(t)$ is given by the probability $p(X(t)|X(t-1))$. Notice the dimensionality of the transition matrix is 2^{2^n} and so can be quite large. The *i*th feature of $X(t)$ is referred to as $x_i(t)$.

The probability of a trajectory, $X(1), \ldots, X(M)$of length *M* in a *n*-dimensional situation state space (i.e., an ordered sequence of *M n*-dimensional global states) is represented as a first-order Markov chain by the following equations:

$$p(X(1), ..., X(M)) = \prod_{t=1}^{M} p(X(t)|X(t-1)) \tag{20}$$

where $X(0)$ is defined as a *n*-dimensional vector of zeros, and

$$p(X(t)|X(t-1)) = (1/Z_t)exp\left[\sum_{i=1}^{n} x_i(t)\left[\sum_{j=1}^{n} v_{ij}x_j(t-1) + k_i\right]\right] \tag{21}$$

which can then be substituted into (20) to obtain:

$$p(X(1)...X(M)) = (1/Z)exp\left[\sum_{t=1}^{M}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i(t)[v_{ij}x_j(t-1) + k_i]\right] \tag{22}$$

Now since the normalization constant Z is not a function of the trajectory and the exponential function is a monotonically increasing function, all minima of :

$$E(X(1)...X(M)) = -\sum_{t=1}^{M}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i(t)[v_{ij}x_j(t-1) + k_i] \tag{23}$$

correspond to maxima of $p(X(1), ..., X(M))$. Finally note that (23) is a special case of (5) if we define *X* to specify a complete trajectory. That is, define $X = [X(1), ..., X(M)]$, and define *W* and *b* in terms of the coefficients v_{ij} and k_i (see discussion of (4) for an example of how this is done).

The relationship between the minima of $E(X(1), ..., X(M))$ and the maxima of $p(X(1), ..., X(M))$ is important for the following reason. Suppose that some subset of the features of a trajectory are "known", while other features are "unknown" and have to be estimated. The goal is to compute the "most probable" values of the unknown features given the known features of the trajectory. More formally, let κ be a subset of the features of a trajectory and their assigned values so that an element

of κ has the form $x_i(t) = 1$ or $x_i(t) = 0$. Let ω be the remaining features of the trajectory. Thus, $p(X(1), ..., X(M)) = p(\kappa \cap \omega)$. A MAP (maximum a posteriori) or "most probable" estimate of ω given κ is then computed by assigning values to the features in ω such that the quantity:

$$p(\omega|\kappa) = p(\kappa \cap \omega)/p(\kappa)$$

is maximized. Moreover, since κ is not a function of the unknown feature values, we can equivalently maximize the likelihood of the trajectory

$$p(X(1), ..., X(M)) = p(\kappa \cap \omega)$$

with respect to the features in ω.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for boltzmann machines. *Cognitive Science*, *9*, 147–169.

Anderson, J. A., Golden, R. M., & Murphy, G. L. (1986). Concepts in distributed systems. In H. Szu (Ed.), *SPIE.Advanced Institute, Optical and Hybrid Computing*. Bellingham, Washington: SPIE.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413–451.

Barto, A. G. (1989). Connectionist learning for control. In *Proceedings of the NSF Workshop on Application of Neural Networks to Robotics and Control*. Cambridge: MIT Press.

Cohen, M. & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems on Man and Cybernetics*, *SMC-13*, 815–826.

Golden, R. M. (1986a). The brain-state-in-a-box neural model is a gradient descent algorithm. *Journal of Mathematical Psychology*, *30*, 73–80.

Golden, R. M. (1986b). A developmental neural model of visual word perception. *Cognitive Science*, *10*, 241–276.

Golden, R. M. & Rumelhart, D. E. (1991). A distributed representation and model for story comprehension and recall. In *The Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*. Hillsdale: Lawrence Erlbaum.

Golden, R. M. & Rumelhart, D. E. (in press). A parallel distributed processing model of story comprehension and recall. *Discourse Processes*.

Greenberg, H. J. (1988). Equilibria of the brain-state-in-a-box (bsb) neural model. *Neural Networks*, *1*, 323–324.

Hirsch, M. W. (1989). Convergent activation dynamics in continuous time networks. *Neural Networks*, *2*, 331–349.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, *79*, 2554–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences, USA*, *81*, 3088–3092.

Hui, S. & Zak, S. H. (1992). Dynamical analysis of the brain-state-in-a-box (bsb) neural models. *IEEE Transactions on Neural Networks*, *3*, 86–94.

Jordan, M. I. (in press). Supervised learning and systems with excess degrees of freedom. *Journal of Mathematical Psychology*.

Luenberger, D. G. (1979). *Introduction to dynamic systems: Theory, models, and*

applications. New York: Wiley.

Marsden, J. E. & Tromba, A. J. (1981). *Vector calculus*. New York: W. H. Freeman.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing. Volume 1: Foundations* (pp. 194–281). Cambridge, MA: MIT Press.

White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, *1*, 425–464.

Author Notes

This research was supported in part by National Institute of Health post-doctoral fellowship HD06943 to the author while he was at Stanford University, and the Program in Cognition and Neuroscience at the University of Texas at Dallas. The author would like to thank Stephen Hayes and Walter E. Lillo for their comments on an earlier version of this paper.

Correspondence and requests for reprints should be sent to Richard Golden, University of Texas at Dallas, School of Human Development, Program in Cognition and Neuroscience, GR41, Box 830688, Richardson, Texas 75083–0688. Email address: golden@utdallas.edu.

Figure Captions

Figure 1. Classification dynamics of a two-unit GBSB model. Unit *i* has an activation value of $x(i)$ which is always greater than minimum value f_i and less than maximum value F_i with *i*=1,2. Note that these constraints form a "box" about the origin. Delta vectors (see text) indicating the sequence of system states generated by the GBSB algorithm are plotted. In this example, the initial system state is classified in three iterations as the hyperbox vertex (F_1, F_2).

Figure 2. Illustration using delta vectors showing how two distinct initial system states S_1 and S_2 are mapped into the same "category" hyperbox vertex (F_1, F_2).

Figure 3. Illustration using delta vectors showing how two distinct initial system states S_1 and S_2 are mapped into different "category" hyperbox vertices (f_1, F_2) and (F_1, F_2).

Figure 4. Illustration showing the condition which is necessary for a hyperbox vertex to be an "attractor" for the GBSB algorithm. The linear response of the system to hypercube vertex (F_1, F_2) must point out of the box in order to guarantee that vertex (F_1, F_2) is an attractor. In this example, delta vectors associated with the first three iterations of the algorithm are plotted as in Figure 1. Delta vector δ^4 is not plotted as it is a vector of zeros since once the system state reaches vertex (F_1, F_2), the system state remains at that vertex for all time. The vector labeled R^4 indicates that the linear response of the system at iteration four points out of the box as is required.

List of Mathematical Symbols

α = lowercase alpha

β = lowercase beta

γ = lowercase gamma

ρ= lowercase rho

λ = lowercase lambda

ω = lowercase omega

κ = lowercase kappa

$\cap$ = cap (logical and) (set intersection)

η = lowercase eta

Γ=uppercase gamma

δ=lowercase delta