Reprinted from JOURNAL OF STATISTICAL PHYSICS

Vol. 1, No. 2, 1969
Printed in Belgium

On Learning and Energy-Entropy Dependence in Recurrent and Nonrecurrent Signed Networks

Stephen Grossberg[1]

Received June 13, 1969

Learning of patterns by neural networks obeying general rules of sensory transduction and of converting membrane potentials to spiking frequencies is considered. Any finite number of cells $\mathcal{A}$ can sample a pattern playing on any finite number of cells $\mathcal{B}$ without causing irrevocable sampling bias if $\mathcal{A} = \mathcal{B}$ or $\mathcal{A} \cap \mathcal{B} = \emptyset$. Total energy transfer from inputs of $\mathcal{A}$ to outputs of $\mathcal{B}$ depends on the entropy of the input distribution. Pattern completion on recall trials can occur without destroying perfect memory even if $\mathcal{A} = \mathcal{B}$ by choosing the signal thresholds sufficiently large. The mathematical results are global limit and oscillation theorems for a class of nonlinear functional-differential systems.

KEY WORDS: learning; stimulus sampling; nonlinear difference-differential equations; global limits and oscillations; flows on signed networks; functional-differential systems; energy-entropy dependence; pattern completion; recurrent and nonrecurrent anatomy; sensory transduction rules; ratio limit theorems.

1. INTRODUCTION

Some networks of formal neurons have been found[1,2] which can learn, simultaneously remember, and perform individually upon demand any number of space-time patterns of essentially arbitrary complexity. Learning in these networks occurs by formal analogs of respondant and operant conditioning, and various mathematical phenomena that occur during the conditioning procedure have analogs in psychological and physiological data. These papers have shown that a *single* formal neuron can

The preparation of this work was supported in part by the National Science Foundation (GP 9003), the Office of Naval Research (N00014-67-A-024-0016), and the A.P. Sloan Foundation.

[1] Massachusetts Institute of Technology, Cambridge, Massachusetts.

learn and control performance of an essentially arbitrarily complicated space-time pattern, such as a piano sonata or a dance, if its axon collaterals terminate on sufficiently many muscle groups, or on cell bodies which control these groups. Encoding an entire pattern in one neuron has a serious drawback, however: performance of the pattern is wholly ritualistic, or by rote.

Voluntary control of complex behavioral acts by a higher animal is not ritualistic in any obvious sense. In particular, voluntary control is sensitive to feedback from immediately prior performance and internal controls of this performance, and to fluctuating environmental demands, both external and internal. An adaptive response to feedback becomes possible only if the space-time pattern is learned by a collection of cells in which no one cell can irrevocably trigger performance of the pattern in an unmodifiable way.

We therefore consider below situations in which a finite collection $\mathcal{A}$ of cells encodes, or "samples," prescribed segments of a space-time pattern delivered to a finite collection $\mathcal{B}$ of cells by an independent input source. If $\mathcal{A}$ and $\mathcal{B}$ are disjoint,

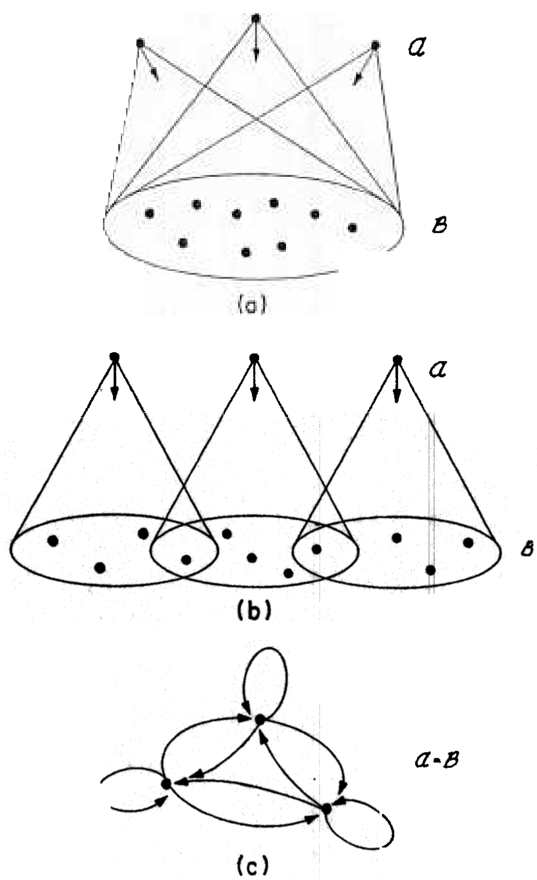

Fig. 1

then the sampling is said to be *nonrecurrent*. If, moreover, every cell in $\mathcal{A}$ samples all cells in $\mathcal{B}$, then the sampling is nonrecurrent and *full*, as in Fig. 1(a). Otherwise, the sampling is *nonfull*, as in Fig. 1(b). In full sampling, each cell in $\mathcal{A}$ can sample $\mathcal{B}$ at essentially arbitrary times without interfering irrevocably with the sampling activity of other cells. In nonfull sampling, the sampling cells cannot sample at arbitrary times without interfering with each other. Instead, constaints on the onset times and duration of sampling intervals by particular cells in $\mathcal{A}$ must be fulfilled to avoid mutual interference. These constraints have a natural neural interpretation, which is discussed elsewhere.[3] If $\mathcal{A} = \mathcal{B}$, the sampling is said to be *completely recurrent*, as in Fig. 1(c). If $\mathcal{A} \neq \mathcal{B}$ but $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, then the sampling is called *incompletely recurrent*. We will prove a general theorem about sampling of $\mathcal{B}$ by $\mathcal{A}$ that reduces to these various cases, and which applies to arbitrarily large finite sets $\mathcal{A}$ and $\mathcal{B}$. Once cells in $\mathcal{A}$ learn the space-time pattern segments that they have sampled, it is readily seen that they can reproduce these segments on $\mathcal{B}$ if they are later activated. These results therefore amount to a rather general discussion of learning by respondant conditioning.[4] Applications to operant conditioning are also readily noticed once the learning mechanism is understood.[2]

The above results on learning can be joined to studies of pattern discrimination[5] yielding networks capable of performing any number of essentially arbitrarily complicated output patterns selectively in response to any number of essentially arbitrarily complicated input patterns. The pattern discrimination work introduces networks whose components can be interpreted as hierarchies of cellular filters, or "feature detectors," since various cells in these networks can be activated by particular pattern features. Suppose that some of these cells can send signals to cells which control given muscle groups. The feature detectors can be interpreted as cells $\mathcal{A}$ and the muscle control cells can be interpreted as cells $\mathcal{B}$. Our results then show that complicated input features can trigger complicated output responses. It is clear that the cells $\mathcal{A}$ must be carefully arranged in realistic situations; for example, to avoid the simultaneous firing of $\mathcal{B}$ cells which control incompatible motor acts, and to guarantee that the correct $\mathcal{A}$ cells are activated by feedback.

Two typical network types which motivate our results are given as follows:

$$\dot{x}_i(t) = -\alpha_i x_i(t) + \sum_{m=1}^{n} [x_m(t-\tau_{mi}) - \Gamma_{mi}]^+ \beta_{mi} z_{mi}^{(+)}(t) - \sum_{m=1}^{n} [x_m(t-\tau_{mi}) - \Gamma_{mi}]^+ \gamma_{mi} z_{mi}^{(-)}(t) + I_i(t), \tag{1}$$

$$\dot{z}_{ji}^{(+)}(t) = -u_{ji}^{(+)} z_{ji}^{(+)}(t) + v_{ji}^{(+)} [x_j(t-\tau_{ji}) - \Gamma_{ji}]^+ [x_i(t)]^+, \tag{2}$$

and

$$\dot{z}_{ji}^{(-)}(t) = -u_{ji}^{(-)} z_{ji}^{(-)}(t) + v_{ji}^{(-)} [x_j(t-\tau_{ji}) - \Gamma_{ji}]^+ [-x_i(t)]^+, \tag{3}$$

where $[w]^+ = \max(w, 0)$ for any real number w, and $i, j = 1, 2,..., n$; or Eq. (1) along with

$$\dot{z}_{ji}^{(+)}(t) = \{-u_{ji}^{(+)} z_{ji}^{(+)}(t) + v_{ji}^{(+)} [x_i(t)]^+\}[x_j(t-\tau_{ji}) - \Gamma_{ji}]^+ \tag{4}$$

and

$$\dot{z}_{ji}^{(-)}(t) = \{-u_{ji}^{(-)} z_{ji}^{(-)}(t) + v_{ji}^{(-)} [-x_i(t)]^+\} [x_j(t - \tau_{ji}) - \Gamma_{ji}]^+. \tag{5}$$

Elsewhere,[2] systems of this kind are interpreted psychologically, physiologically, and anatomically. In particular, let n cell bodies v_i be given with average potential $x_i(t)$, $i = 1, 2, ..., n$. If $\beta_{ki} > 0$ $(\gamma_{ki} > 0)$, then an excitatory (inhibitory) axon e_{ki}^+ (e_{ki}^-) leads from v_k to v_i. Denote the synaptic knob of e_{ki}^+ (e_{ki}^-) by N_{ki}^+ (N_{ki}^-), and let $z_{ki}^+(t)$ $(z_{ki}^-(t))$ be the excitatory (inhibitory) chemical transmitter activity in N_{ki}^+ (N_{ki}^-). The spiking frequency which is created by v_k in e_{ki}^+ (e_{ki}^-) in the time interval $[t, t + dt]$ is proportional to $[x_k(t) - \Gamma_{ki}]^+ \beta_{ki}$ $([x_k(t) - \Gamma_{mi}]^+ \gamma_{mi})$. The time lag for the signal to flow from v_k to N_{ki}^+ or N_{ki}^- is τ_{ki}, and the spiking threshold of e_{ki}^+ or e_{ki}^- is Γ_{ki}. We can choose equal excitatory and inhibitory time lags and thresholds for our present purposes, since we require that $\beta_{ki}\gamma_{ki} = 0$. When the signal from v_k reaches N_{ki}^+ (N_{ki}^-) at time t, it causes release of excitatory (inhibitory) transmitter into the synaptic cleft facing v_i at a rate proportional to

$$[x_k(t - \tau_{ki}) - \Gamma_{ki}]^+ \beta_{ki} z_{ki}^{(+)}(t) \; ([x_k(t - \tau_{ki}) - \Gamma_{ki}]^+ \gamma_{ki} z_{ki}^{(-)}(t)),$$

whence the rate of change of x_i increases (decreases) proportionately. All excitatory (inhibitory) signals are added (subtracted) at v_i, as the term

$$\sum_{m=1}^{n} [x_m(t - \tau_{mi}) - \Gamma_{mi}]^+ \beta_{mi} z_{mi}^{(+)}(t)$$

in Eq. (1) illustrates for the excitatory case. $x_i(t)$ also decays exponentially at the rate α_i, and is perturbed by known inputs $I_i(t)$ that are under control of an experimentalist or independent cells.

The transmitter production processes are regulated by cross-correlation of presynaptic spiking frequencies and postsynaptic potentials. For example, $z_{ji}^{(+)}(t)$ in Eq. (2) cross-correlates the presynaptic signal $\beta_{ji}[x_j(t - \tau_{ji}) - \Gamma_{ji}]^+$ received from v_j by $N_{ji}^{(+)}$ at time t, with the value $[x_i(t)]^+$ of the contiguous postsynaptic cell v_i; hence the term $v_{ji}^{(+)}[x_j(t - \tau_{ji}) - \Gamma_{ji}]^+ [x_i(t)]^+$ in Eq. (2). This cross-correlation is positive iff v_j has a positive spiking frequency at time $t - \tau_{ji}$ and v_i has a supraequilibrium potential at time t. $z_{ji}^{(+)}(t)$ also decays exponentially at the rate $u_{ji}^{(+)}$. $z_{ji}^{(-)}(t)$ has a similar interpretation, with the difference that $v_{ji}^{(-)}[x_j(t - \tau_{ji}) - \Gamma_{ji}]^+ [-x_i(t)]^+$ in Eq. (3) is positive only if v_i has a subequilibrium potential at time t. Speaking psychologically, $x_i(t)$ is the *stimulus trace* of v_i, and $z_{ji}^{(+)}(t)$ $(z_{ji}^{(-)}(t))$ is the *excitatory (inhibitory) associational strength* of the association $v_j \to v_i$. Elsewhere, [9] I give references which discuss psychological, physiological, and biochemical implications of these equations in a more leisurely way.

The mathematical results which we will use to study these networks include functional-differential systems of the following form:

$$\dot{x}_i(t) = A(W_t, t)\, x_i(t) + \sum_{k \in J} B_k(W_t, t)\, z_{ki}(t) + C_i(t) \tag{6}$$

and

$$\dot{z}_{ji}(t) = D_j(W_t, t) z_{ji}(t) + E_j(W_t, t) x_i(t), \tag{7}$$

where $i \in I$, $j \in J$, and the finite sets I and J of indices are either disjoint or equal, corresponding to Fig. 1(a) and 1(c), respectively. The coefficients A, B_j, D_j, and E_j in Eqs. (6) and (7) are continuous functions of t, which can depend nonlinearly on the vector function $W = (x_i, z_{ji} : i \in I, j \in J)$ evaluated at all times no later than t, and on known functions of t. The generality of these coefficients means that peripheral sensory transducers and rules for transforming cell-body membrane potentials into axonal spiking frequencies can be of very general form without distorting the ultimate path of learning, once the network anatomy is suitably chosen. Particular transduction and spiking-frequency rules merely determine the rate at which particular patterns are learned by particular cells, and therefore the importance of these patterns to the prescribed cells.

We will consider questions of energy-entropy dependence. In the case $I = J$, for example, suppose that a total input $C(t)$ through time is prescribed, and that a definite fraction θ_i of this input is delivered to the ith cell, $i \in I$. In other words, we deliver a *spatial pattern* with weights θ_i to the cells, and allow them to mutually interact. What pattern maximizes (minimizes) the total potential $x = \sum_{i \in I} x_i$ and the total output of the network after the interaction takes place? In various cases of interest, the pattern with minimal (maximal) entropy maximizes (minimizes) the total potential and output. Thus, if more order is introduced into the input pattern, then more output energy is available to drive the processing of these outputs at the next level of cells. These results are compatible with the fact that minimal (maximal) input entropy minimizes (maximizes) the destructive effects of inhibitory interactions on the total output, where these interactions exist. Elsewhere,[6] I discuss inhibitory interactions in terms of the principle of sufficient reason. The fact that inputs in which a certain amount of order already exists are given preferred treatment energetically stands in interesting contrast to the behavior of the closed systems of classical thermodynamics, and to the thermodynamic second law, which presages maximal entropic doom for the universe. In making this contrast, it is well to remember that the energy-entropy relations in our open systems are a consequence of the threshold and quadratic nonlinearities between x_i's and z_{ji}'s that are the basis of evolutionary trands, or learning, in these systems.

2. MAIN THEOREM

This section proves that systems which satisfy Eqs. (6)–(7) can learn a spatial pattern $C_i(t) = \theta_i C(t)$, where $\sum_{i \in I} \theta_i = 1$ and $\theta_i \geq 0$, under rather weak conditions. Then, using results given elsewhere,[1,3] this result can readily be applied to the problem of learning a space-time pattern with variable weights $\theta_i(t) = C_i(t)[\sum_{k \in I} C_k(t)]^{-1}$. The theorem studies the limiting behavior of the probabilities $X_i(t) = x_i(t)[\sum_{k \in I} x_k(t)]^{-1}$ and $y_{ji}(t) = z_{ji}(t)[\sum_{k \in I} z_{jk}(t)]^{-1}$.

2.1. Theorem 1

Consider the system given by

$$\dot{x}_i(t) = A(W_t, t)\, x_i(t) + \sum_{k\in J} B_k(W_t, t)\, z_{ki}(t) + \theta_i C(t) \tag{8}$$

and Eq. (7),

$$\dot{z}_{ji}(t) = D_j(W_t, t)\, z_{ji}(t) + E_j(W_t, t)\, x_i(t),$$

where $i \in I$ and $j \in J$, and the finite sets I and J of indices are either disjoint or equal. Let the initial data and inputs be nonnegative and continuous, and let the coefficients be continuous functions of t. Suppose furthermore that:

1. All B_j and E_j are nonnegative
2. $\int_0^\infty B_j(W_v, v)\, dv = \infty$ only if $\int_0^\infty E_j(W_v, v)\, dv = \infty$ (9)
3. $\int_0^\infty C(v)\, dv = \infty$ (10)
4. There exist positive constants K_1 and K_2 such that for all $T \geqslant 0$,

$$\int_T^{T+t} C(v) \left\{\exp\left[\int_v^{T+t} A(W_\xi, \xi)\, d\xi\right]\right\} dv \geqslant K_1 \quad \text{if} \quad t \geqslant K_2$$

5. The solution of Eqs. (7) and (8) is bounded.

Then all the limits $P_{ji} = \lim_{t\to\infty} y_{ji}(t)$ and $Q_i = \lim_{t\to\infty} X_i(t)$ exist, with $Q_i = \theta_i$. Furthermore, if

$$\int_0^\infty E_j(W_v, v)\, dv = \infty \tag{12}$$

then $P_{ji} = \theta_i$.

The following proposition is needed to prove Theorem 1, and is stated in terms of the functions $Y_i(t) = \max\{y_{ji}(t): j \in J\}$, $y_i(t) = \min\{y_{ji}(t): j \in J\}$, $M_i(t) = \min\{Y_i(t), x(t)\}$, $m_i(t) = \min\{y_i(t), x_i(t)\}$, $Y_{i,\theta}(t) = \max\{Y_i(t), \theta_i\}$, $y_{i,\theta}(t) = \min\{y_i(t), \theta_i\}$, $M_{i,\theta}(t) = \max\{Y_{i,\theta}(t), X_i(t)\}$, $m_{i,\theta}(t) = \min\{y_{i,\theta}(t), X_i(t)\}$, $X_i^{(\theta)}(t) = X_i(t) - \theta_i$, $y_{ji}^{(\theta)}(t) = y_{ji}(t) - \theta_i$, $Y_i^{(\theta)}(t) = Y_i(t) - \theta_i$, and $y_i^{(\theta)}(t) = y_i(t) - \theta_i$, for $i \in I$ and $j \in J$.

2.2. Proposition 1

Let Eqs. (7) and (8) be given with nonnegative and continuous initial data, and suppose for convenience that $\sum_{i\in I} x_i(0) > 0$ and $\sum_{i\in I} z_{ji}(0) > 0$ if $E_j \not\equiv 0$. Let the coefficient functionals be continuous in t, with C and all B_j and E_j nonnegative. Also, let the solution of Eqs. (7) and (8) exist for all $t \geqslant 0$. Then for every time $T \geqslant 0$ and all $t \geqslant T$,

$$m_{i,\theta}(T) \leqslant m_{i,\theta}(t) \leqslant M_{i,\theta}(t) \leqslant M_{i,\theta}(T)$$

If, moreover, $C(t) = 0$ for $t \geqslant T$, then

$$m_i(T) \leqslant m_i(t) \leqslant M_i(t) \leqslant M_i(T). \tag{14}$$

The functions $\dot{Y}_{i,\theta}$, $\dot{y}_{i,\theta}$, $X_i - Y_{i,\theta}$, and $X_i - y_{i,\theta}$ change sign at most once, and not at all if $y_{i,\theta}(0) \leqslant X_i(0) \leqslant Y_{i,\theta}(0)$. If, moreover, $C(t) = 0$ for $t \geqslant T$, then the functions $\dot{Y}_i$, $\dot{y}_i$, $X_i - Y_i$, and $X_i - y_i$ change sign at most once for $t \geqslant T$, and not at all if $y_i(T) \leqslant X_i(T) \leqslant Y_i(T)$.

It follows from Eqs. (13) and (14) that if $X_i(t)$ and all $y_{ji}(t)$ are attracted close to θ_i by a sufficient amount of practice, then these functions will remain close to θ_i even after practice ceases. See Theorem 2 of my earlier paper[7] for a related discussion in the special case that

$$B_j(W_t, t) = E_j(W_t, t) = \frac{\beta x_j(t-\tau)}{\sum_{k\in I} z_{jk}(t)}$$

with $I = J = \{1, 2,..., n\}$.

2.2.1. Proof of Proposition 1. Equations (7) and (8) can be transformed into the following system of equations for $X_i^{(\theta)}$ and $y_{ji}^{(\theta)}$:

$$\dot{X}_i^{(\theta)}(t) = \sum_{k\in J} F_k(W_t, t)[y_{ki}^{(\theta)}(t) - X_i^{(\theta)}(t)] - G(t)\, X_i^{(\theta)}(t) \tag{15}$$

$$\dot{y}_{ji}^{(\theta)}(t) = H_j(W_t, t)[X_i^{(\theta)}(t) - y_{ji}^{(\theta)}(t)], \tag{16}$$

$$F_j(W_t, t) = B_j(W_t, t)\, z_j(t)/x(t), \tag{17}$$

$$G(t) = C(t)/x(t), \tag{18}$$

$$H_j(W_t, t) = E_j(W_t, t)\, x(t)/z_j(t), \tag{19}$$

$$x(t) = \sum_{i\in I} x_i(t), \tag{20}$$

$$z_j(t) = \sum_{i\in I} z_{ji}(t). \tag{21}$$

Equations (15) and (16) follow from the equations $\dot{X}_i = x^{-1}(\dot{x}_i - x_i \dot{x} x^{-1})$ and $\dot{y}_{ji} = z_j^{-1}(\dot{z}_{ji} - z_{ji}\dot{z}_j z_j^{-1})$ along with Eqs. (7) and (8), and the equations

$$\dot{x} = Ax + \sum_{k\in J} B_k z_k + C \tag{22}$$

and

$$\dot{z}_j = D_j z_j + E_j x. \tag{23}$$

The proof of Proposition 1 can now be completed using Eqs. (15) and (16) after noting that all coefficients in these equations are nonnegative, and that $\mathfrak{C}(t) = 0$ if $\mathfrak{C}(t) = 0$. Cases 1–3 below can be read off from Eqs. (15) and (16) by inspection. Since $C(t)$ is continuous, $C(t) = 0$ on a sequence of nonoverlapping intervals $J_p = [a_p, b_p]$. On these intervals, and in particular if $C(t) = 0$ for $t \geqslant T$, Cases 1–3 can be strengthened to yield Cases 4–6 below. Then the assertions of Proposition 1 follow by pasting together the assertions of these exhaustive cases.

Case 1. If $X_i^{(\theta)}(t_0) \geqslant 0$ and $y_i^{(\theta)}(t_0) \geqslant 0$, then $X_i^{(\theta)}(t) \geqslant 0$ and $y_i^{(\theta)}(t) \geqslant 0$ for $t \geqslant t_0$. If, moreover, $X_i^{(\theta)}(t_0) \leqslant Y_i^{(\theta)}(t_0)$, then $X_i^{(\theta)}(t) \leqslant Y_i^{(\theta)}(t)$ and $Y_i^{(\theta)}(t)$ is monotone decreasing for $t \geqslant t_0$. On the other hand, if $X_i^{(\theta)}(t_0) > Y_i^{(\theta)}(t_0)$, then $X_i^{(\theta)}(t)$ is monotone decreasing and all $y_{ji}^{(\theta)}(t)$ are monotone increasing until the first time $t = t_1$ at which $X_i^{(\theta)}(t) = Y_i^{(\theta)}(t)$. If no such time exists, all limits Q_i and P_{ji} exist and $Q_i \geqslant P_{ji} \geqslant \theta_i$. If such a time does exist, the preceding case holds for all $t \geqslant t_1$.

Case 2. If $X_i^{(\theta)}(t_0) \leqslant 0$ and $Y_i^{(\theta)}(t_0) \leqslant 0$, then the arguments of Case 1 go through with inequalities reversed, and $y_i^{(\theta)}$ and $Y_i^{(\theta)}$ interchanged. Thus, either all limits Q_i and P_{ji} exist, or there is a t_1 such that $y_i^{(\theta)}(t) \leqslant X_i^{(\theta)}(t)$ for $t \geqslant t_1$.

Case 3. If $Y_i^{(\theta)}(0) \geqslant 0 \geqslant y_i^{(\theta)}(0)$ and $Y_i^{(\theta)}(0) > y_i^{(\theta)}(0)$, then either $Y_i^{(\theta)}(t) \geqslant 0 \geqslant y_i^{(\theta)}(t)$ and $Y_i^{(\theta)}(t) > y_i^{(\theta)}(t)$ for all $t \geqslant 0$, or we eventually enter either Case 1 or Case 2. Suppose that the former alternative occurs. If, moreover, $X_i^{(\theta)}(0) \notin [y_i^{(\theta)}(0), Y_i^{(\theta)}(0)]$, then $X_i^{(\theta)}(t)$ and all $y_{ji}^{(\theta)}(t)$ are monotonic until the first time $t = t_2$ at which $X_i^{(\theta)}(t) \in [y_i^{(\theta)}(t), Y_i^{(\theta)}(t)]$. Thereafter, $X_i^{(\theta)}(t) \in [y_i^{(\theta)}(t), Y_i^{(\theta)}(t)]$, and $Y_i^{(\theta)}(t)$ is monotone decreasing, whereas $y_i^{(\theta)}(t)$ is monotone increasing. Both limits $Y_i(\infty) = \lim_{t\to\infty} Y_i(t)$ and $y_i(\infty) = \lim_{t\to\infty} y_i(t)$ therefore exist. If $Y_i(\infty) = y_i(\infty)$, all limits Q_i and P_{ji} exist and are equal.

Cases 1–3 exhaust all alternatives, and readily imply Eq. (13) as well as the assertion concerning $\dot{Y}_{i,\theta}$, $\dot{y}_{i,\theta}$, $X_i - Y_{i,\theta}$, and $X_i - y_{i,\theta}$.

For any t_0, t_1, and t in J_p, Cases 1–3 can be strengthened as follows.

Case 4. If $X_i(t_0) \in [y_i(t_0), Y_i(t_0)]$, then $X_i(t) \in [y_i(t), Y_i(t)]$ for $t \geqslant t_0$, where $y_i(t)$ is monotone increasing and $Y_i(t)$ is monotone decreasing for $t \geqslant t_0$.

Case 5. If $X_i(t_0) > Y_i(t_0)$, then $X_i(t)$ is monotone decreasing, and all $y_{ji}(t)$ are monotone increasing until the first time $t = t_1 > t_0$ at which $X_i(t) = Y_i(t)$. Thereafter, $Y_i(t)$ is monotone decreasing and $y_i(t)$ is monotone increasing by Case 1, so that $\dot{Y}_i(t)$ changes sign at most once, and $y_i(t)$ is always monotone increasing.

Case 6. If $X_i(t_0) < y_i(t_0)$, then the conclusions of Case 5 hold with y_i replacing Y_i and all inequalities reversed.

Cases 4–6 imply Eq. (13) as well as the assertion concerning $\dot{Y}_i$, $\dot{y}_i$, $X_i - Y_i$, and $X_i - y_i$.

In the statements above concerning the derivatives $\dot{Y}_i$, $\dot{Y}_{i,\theta}$, $\dot{y}_i$, and $\dot{y}_{i,\theta}$, left- and right-handed derivatives are intended where two-sided derivatives do not exist. The statements concerning monotonicity do not intend strict monotonicity, as the following corollary emphasizes.

2.3. Corollary 1

$\dot{y}_{ji}(t) = 0$ if $E_j(W_t, t) = 0$, and $\dot{X}_i(t) = 0$ if all $B_j(W_t, t) = 0$ and $C(t) = 0$.

The fact that $y_{ji}(t)$ does not vary in intervals for which $E_j(W_t, t) = 0$ is the basis for sampling in our networks. Suppose, for example, that an arbitrary space-time pattern with weights $\theta_i(t) = C_i(t)[\sum_{k\in I} C_k(t)]^{-1}$ perturbs the system, but that $E_j(W_t, t)$ is positive only at times t for which $\theta_j(t) = \theta_j$. Then $y_{ji}(t)$ will only vary at these times, and will, apart from momentary fluctuations of $X_i(t)$ toward the values of other $y_{ki}(t)$ with $E_k(W_t, t) > 0$, be attracted toward θ_i. If, however, $E_j(W_t, t)$ is positive at times during which $\theta_i(t)$ varies throughout an interval $[\theta_i - \epsilon, \theta_i + \epsilon]$, then $y_{ji}(t)$ will be attracted to a suitable weighted average of all the values in this interval. It is therefore often desirable either that (1) $E_j(W_t, t)$ and only $E_j(W_t, t)$ is positive in intervals of such short length that $\theta_i(t)$ cannot substantially vary during these intervals, or that (2) the input energy $C(t)$ is sufficiently great when $E_j(W_t, t) > 0$ to quickly drive $X_i(t)$ and hence $y_{ji}(t)$ toward θ_i before $\theta_i(t)$ can substantially change. Elsewhere,[5] it is shown that a sampling interval of prescribed duration can be achieved, given even a conditioned stimulus of unlimited duration, using the mechanism of nonrecurrent inhibition.

By the above paragraph, a space-time pattern with weights $\theta_i(t)$ can be approximately encoded as a sequence of spatial patterns with weights $\theta_i(k\xi)$, $k = 1, 2,\ldots$, for ξ sufficiently small, by letting successive probability distributions $y^{(k)} \equiv \{y_{ki} : i \in I\}$ sample the pattern sequentially for brief intervals in "avalanche" fashion, as discussed elsewhere.[2,3] The closeness of fit of y_k to the weights $\theta_i(k\xi)$ will depend crucially on the limiting statements of Theorem 1, which we now prove.

3. PROOF OF THEOREM 1

By Eq. (10) and Proposition 1, Cases 1–3 exhaust all possibilities. It is convenient to first consider the subcases of Cases 1 and 2 in which all limits are known to exist because X_i and all y_{ji} are monotonic for large t. Suppose, for example, that $X_i^{(\theta)}(t) > Y_i^{(\theta)}(t)$ and $X_i^{(\theta)}(t) \geqslant 0$ for all large t in Case 1, and let these inequalities hold at all $t \geqslant 0$ for convenience. Then by Eq. (15), $\dot{X}_i^{(\theta)} \leqslant -GX_i^{(\theta)}$, and X_i decreases monotonically to a limit $Q_i \geqslant \theta_i$. Thus, $\dot{X}_i^{(\theta)} \leqslant G(\theta_i - Q_i)$, or in integral form,

$$0 \leqslant X_i(t) \leqslant X_i(0) + (\theta_i - Q_i)\int_0^t G(v)\,dv$$

for all $t \geqslant 0$. Supposing that $Q_i > \theta_i$, we will derive a contradiction. Since $x(v)$ is bounded above, say, by K_3^{-1},

$$K_3(Q_i - \theta_i)\int_0^t C\,dv \leqslant X_i(0) \leqslant 1$$

for all $t \geqslant 0$. Thus, $\int_0^\infty C\,dv < \infty$, which contradicts Eq. (10). A similar procedure can be used to show that $Q_i = \theta_i$ in the monotonic subcases of Case 2.

We now show that $P_{ji} = \theta_i$ in the monotonic subcases if Eq. (12) holds. Consider the subcase of Case 1. We know that $Q_i = \theta_i$, and for $t \gg 0$ that $X_i^{(\theta)}(t) \geqslant Y_i^{(\theta)}(t)$ and $X_i^{(\theta)}(t) \geqslant 0$. Thus also $Y_i^{(\theta)}(t) \leqslant 0$ for $t \gg 0$. Otherwise, there will exist a T such that $Y_i(T) > \theta_i$, and consequently $X_i(t) \geqslant Y_i(t) \geqslant Y_i(T) > \theta_i$ for $t \geqslant T$, since $Y_i(t)$ is monotone increasing if $X_i(t) > Y_i(t)$. In particular, $Q_i > \theta_i$, which is impossible.

We can therefore restrict attention to the case in which $X_i^{(\theta)}(t) \geqslant 0 \geqslant Y_i^{(\theta)}(t)$ for $t \gg 0$. Then by Eq. (16), for every $j \in J$,

$$\dot{y}_{ji} \geqslant H_j(\theta_i - y_{ji}) \tag{24}$$

and moreover y_{ji} increases to the limit $P_{ji} \leqslant \theta_i$. Thus, $\dot{y}_{ji} \geqslant (\theta_i - P_{ji}) H_j$ for all large t, or without loss of generality for all $t \geqslant 0$. The inequalities

$$1 \geqslant y_{ji}(t) - y_{ji}(0) \geqslant (\theta_i - P_{ji}) \int_0^t H_j\,dv$$

therefore hold for all $t \geqslant 0$, and, consequently,

$$\int_0 H_j\,dv < \infty \tag{25}$$

if $\theta_i > P_{ji}$. Equation (25) will now be shown to contradict Eq. (12), and thus $\theta_i = P_{ji}$. Since z_j is bounded, Eqs. (19) and (25) imply $\int_0^\infty xE_j\,dv < \infty$. But, by Eq. (22),

$$x(t) \geqslant x(0) \exp\left[\int_0^t A\,dv\right] + \int_0^t C \left\{\exp\left[\int_v^t A\,d\xi\right]\right\} dv$$

where $x(0) \geqslant 0$. Equation (11) with $T = 0$ therefore implies that $x(t) \geqslant K_1$ for $t \geqslant K_2$, and, in particular $\int_0^\infty E_j\,dv < \infty$, which contradicts Eq. (12), and thereby proves $P_{ji} = \theta_i$.

Only the nonmonotonic subcases of Cases 1–3 remain, and these are listed below:

I. $Y_i^{(\theta)}(t) \geqslant X_i^{(\theta)}(t) \geqslant 0$ and $y_i^{(\theta)}(t) \geqslant 0$ with $Y_i^{(\theta)}(t)$ monotone decreasing, for $t \gg 0$.

II. $y_i^{(\theta)}(t) \leqslant X_i^{(\theta)}(t) \leqslant 0$ and $Y_i^{(\theta)}(t) \leqslant 0$ with $y_i^{(\theta)}(t)$ monotone increasing, for $t \gg 0$.

III. $X_i^{(\theta)}(t) \in [y_i^{(\theta)}(t), Y_i^{(\theta)}(t)]$ and $y_i^{(\theta)}(t) \leqslant 0 \leqslant Y_i^{(\theta)}(t)$ with $y_i^{(\theta)}(t)$ monotone increasing and $Y_i^{(\theta)}(t)$ monotone decreasing, for $t \gg 0$.

Only Case I will be explicitly considered, since Cases II and III can be treated by an analogous method. First, we treat the subcase in which

$$B_k\,dv < \infty.$$

Then, for every $\epsilon > 0$, there exists a T_ϵ such that $t \geqslant T_\epsilon$ implies

$$\sum_{k\in J} \int_t^\infty B_k \, dv \leqslant \epsilon/2\mu$$

where $\mu = \sup z_j x^{-1} < \infty$. By Eq. (15),

$$\dot{X}_i^{(\theta)} \leqslant \sum_{k\in J} F_k - G X_i^{(\theta)},$$

and thus for $t \geqslant T_\epsilon$,

$$X_i^{(\theta)}(t) \leqslant X_i^{(\theta)}(T_\epsilon) \exp\left[-\int_{T_\epsilon} G \, dv\right] + \sum_{k\in J} \int_{T_\epsilon}^t F_k \left\{\exp\left[-\int_v^t G \, d\xi\right]\right\} dv,$$

which by Eq. (26) implies

$$0 \leqslant X_i^{(\theta)}(t) \leqslant (\epsilon/2) + \exp\left[-\int_{T_\epsilon}^t G \, dv\right].$$

Equation (10) and the boundedness of $x(v)$ now imply $Q_i = \theta_i$.

We can now show that all P_{ji} exist in this case. By hypothesis, $X_i(t) \geqslant \theta_i$ and $y_{ji}(t) \geqslant \theta_i$. Since $Q_i = \theta_i$, for every $\epsilon > 0$ there exists an S_ϵ such that $\theta_i \leqslant X_i(t) \leqslant \theta_i + \epsilon$ for $t \geqslant S_\epsilon$. Thus if $y_{ji}(t) > \theta_i + \epsilon$ for some $t \geqslant S_\epsilon$, then $y_{ji}(t)$ decreases to a limit $P_{ji} \geqslant \theta_i + \epsilon$, or eventually $y_{ji}(t) \leqslant \theta_i + \epsilon$. In other words, either P_{ji} exists or $y_{ji}(t) \leqslant \theta_i + \epsilon$ for every $\epsilon > 0$ and t sufficiently large. Since also $y_{ji}(t) \geqslant \theta_i$, P_{ji} exists in all cases. If Eq. (12) holds, $P_{ji} = \theta_i$ can readily be proved as in Eq. (24) using $Q_i = \theta_i$.

It remains only to consider case I—and cases II and III analogously—if

$$\sum_{k\in J} \int_0^\infty B_k \, dv = \infty.$$

Partition J into two sets $J(1)$ and $J(2)$ such that $j \in J(1)$ iff

$$\int_0^\infty B_j \, dv = \infty.$$

By Eq. (28) and the nonnegativity of B_j, $J(1) \neq \emptyset$, and we can define the function

$$\tilde{Y}_i^{(\theta)} = \max\{ y_{ji}^{(\theta)} : j \in J(1)\}.$$

We now show that $\tilde{Y}_i^{(\theta)}(\infty) \equiv \lim_{t\to\infty} \tilde{Y}_i^{(\theta)}(t)$ exists in Case I if Eq. (28) holds. To do this, note by proposition 1 that

$$\operatorname{sgn}\left[\frac{d}{dt} \tilde{Y}_i^{(\theta)}(t)\right] = \operatorname{sgn}[X_i^{(\theta)}(t) - \tilde{Y}_i^{(\theta)}(t)] \tag{30}$$

whenever $(d/dt)\, \tilde{Y}_i^{(\theta)}(t) \neq 0$. Thus, if there exists a T such that either $X_i^{(\theta)}(t) \geqslant \tilde{Y}_i^{(\theta)}(t)$ for all $t \geqslant T$, or $X_i^{(\theta)}(t) \leqslant \tilde{Y}_i^{(\theta)}(t)$ for all $t \geqslant T$, then $\tilde{Y}_i^{(\theta)}(\infty)$ exists. It remains to consider only the case in which $X_i^{(\theta)}(t) - \tilde{Y}_i^{(\theta)}(t)$ oscillates at arbitrarily large times.

Define the functions $L^{(i)} = \sum_{k\in J(i)} F_k$, $i = 1, 2$. Then by Eq. (15),

$$\dot{X}_i^{(\theta)} \leqslant L^{(2)} + (\tilde{Y}_i^{(\theta)} - X_i^{(\theta)}) L^{(1)} - GX_i^{(\theta)}. \tag{31}$$

Suppose that $X_i^{(\theta)}(t) \geqslant Y_i^{(\theta)}(t)$ only in the disjoint sequence of intervals $[S_{ik}, T_{ik}]$, $k = 1, 2,\ldots$. Then $\tilde{Y}_i^{(\theta)}(S_{ik}) = X_i^{(\theta)}(S_{ik})$ for all $k = 1, 2,\ldots$, and by Eq. (31),

$$\dot{X}_i^{(\theta)}(t) \leqslant L^{(2)}(t), \qquad t \in [S_{ik}, T_{ik}],$$

which implies

$$\tilde{Y}_i^{(\theta)}(t) \leqslant X_i^{(\theta)}(t) \leqslant \tilde{Y}_i^{(\theta)}(S_{ik}) + \int_{S_{ik}}^{t} L^{(2)}(v)\, dv$$

and, in particular,

$$\tilde{Y}_i^{(\theta)}(t) \leqslant \tilde{Y}_i^{(\theta)}(S_{ik}) + \int_{S_{ik}}^{\infty} L^{(2)}(v)\, dv \tag{33}$$

by the nonnegativity of $L^{(2)}$. Since by definition of $J(2)$, $\int_0^\infty L^{(2)}(v)\, dv < \infty$, the function $u^{(2)}(T) \equiv \int_T^\infty L^{(2)}(v)\, dv$ monotonically decreases to zero as $T \to \infty$. Thus by Eqs. (30) and (33), the bounded and continuous function $\tilde{Y}_i^{(\theta)}(t)$ is alternatively monotone decreasing or increasing by an amount that approaches zero as $t \to \infty$. It readily follows that $\tilde{Y}_i^{(\theta)}(\infty)$ exists.

Inequality (31) along with the existence of $Y_i^{(\theta)}(\infty)$ will now be used to prove that $\tilde{Y}_i^{(\theta)}(\infty) = Q_i = \theta_i$. Then the existence of all P_{ji} for $j \in J(2)$ can be proved as in the paragraph following Eq. (27). Define the functions $M = (d/dt)[\log(x - A - G)]$ and $N = (d/dt)[\log(x - A)]$. By Eqs. (17), (18), and (22), $L^{(1)} + L^{(2)} = \sum_{k\in J} F_k = M$. Equation (31) therefore implies

$$\dot{X}_i^{(\theta)} \leqslant L^{(2)} + M\tilde{Y}_i^{(\theta)} - NX_i^{(\theta)}, \tag{34}$$

which can be expressed in integral form for any $t \geqslant T \geqslant 0$ as

$$X_i^{(\theta)}(t) \leqslant P_i^{(\theta)}(t, T) + Q_i^{(\theta)}(t, T) + R^{(2)}(t, T)$$

using the notation

$$P_i^{(\theta)}(t, T) = X_i^{(\theta)}(T)\, x(T)\, x^{-1}(t) \exp\left[\int_T^t A\, dv\right],$$

$$Q_i^{(\theta)}(t, T) = x^{-1}(t) \int_T^t \tilde{Y}_i^{(\theta)}(v)\, M(v)\, x(v) \left\{\exp\left[\int_v^t A\, d\xi\right]\right\} dv,$$

and

$$R^{(2)}(t, T) = x^{-1}(t) \int_T^t L^{(2)}(v)\, x(v) \left\{\exp\left[\int_v^t A\, d\xi\right]\right\} dv.$$

We now estimate each of these terms from above. First note that

$$R^{(2)}(t, T) \leqslant U^{(2)}(t, T) \equiv \int_T^t L^{(2)}(v)\, dv.$$

This follows from Eq. (22), which shows that $\dot{x} \geqslant Ax$, and thus that

$$\exp\left[\int_v^t A\,d\xi\right] \leqslant x^{-1}(v)\,x(t).$$

To estimate Eq. (37), we first eliminate the case in which $\tilde{Y}_i^{(\theta)}(t) < X_i^{(\theta)}(t)$ for all $t \geqslant T$ and some $T \geqslant 0$. In this case, Eq. (31) implies $\dot{X}_i^{(\theta)} \leqslant L^{(2)} - GX_i^{(\theta)}$ for $t \geqslant T$, and since, by Eq. (30), $\tilde{Y}_i^{(\theta)}(t)$ is increasing for $t \geqslant T$,

$$\dot{X}_i^{(\theta)} \leqslant L^{(2)} - \tilde{Y}_i^{(\theta)}(T)G \qquad \text{for} \quad t \geqslant T, \tag{4}$$

where we can assume that $\tilde{Y}_i^{(\theta)}(T) > 0$ without loss of generality. Integrating Eq. (40) between T and ∞ then readily yields the inequality $\int_0^\infty C(v)\,dv < \infty$, which contradicts Eq. (10).

Thus, either $\tilde{Y}_i^{(\theta)}(t) \geqslant X_i^{(\theta)}(t)$ for all $t \geqslant T$ and some T sufficiently large, or $\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t)$ changes sign at arbitrarily large times. If $\tilde{Y}_i^{(\theta)}(t) \geqslant X_i^{(\theta)}(t)$ for $t \in [T, S]$, then $\tilde{Y}_i^{(\theta)}(t) \leqslant \tilde{Y}_i^{(\theta)}(T)$, since $\tilde{Y}_i^{(\theta)}(t)$ is monotone decreasing. If $\tilde{Y}_i^{(\theta)}(t)$ does not exceed $X_i^{(\theta)}(t)$ for all $t \geqslant 0$, then there exists a disjoint sequence of intervals $[S_{ik}, T_{ik}]$, $k = 1, 2, \ldots$, such that $\tilde{Y}_i^{(\theta)}(S_{ik}) = X_i^{(\theta)}(S_{ik})$ and $X_i^{(\theta)}(t) \geqslant \tilde{Y}_i^{(\theta)}(t)$ only for $t \in \bigcup_{k=1}^{\infty}[S_{ik}, T_{ik}]$. For any $t \in [S_{ik}, T_{ik}]$, Eq. (32) holds. Pasting these cases together, we find that

$$\tilde{Y}_i^{(\theta)}(t) \leqslant \tilde{Y}_i^{(\theta)}(T) + U^{(2)}(T)$$

for all $t \geqslant T$ and any

$$T \in \mathcal{O}\!\mathcal{l}_i \equiv \{t\colon \tilde{Y}_i^{(\theta)}(t) \geqslant X_i^{(\theta)}(t)\}.$$

Returning to the upper estimate of Eq. (37), note for any $T \in \mathcal{O}\!\mathcal{l}_i$ and all $t \geqslant T$, that Eqs. (37) and (41) yield

$$Q_i^{(\theta)}(t, T) \leqslant [\tilde{Y}_i^{(\theta)}(T) + U^{(2)}(T)]\,V(t, T), \tag{43}$$

where

$$V(t, T) = x^{-1}(t) \int_T^t Mx \left\{\exp\left[\int_v^t A\,d\xi\right]\right\} dv.$$

Note also that

$$\begin{aligned} V(t, T) &= x^{-1}(t) \exp\left[\int_0^t A\,d\xi\right] \int_T^t (\dot{x} - Ax - C) \left\{\exp\left[-\int_0^v A\,d\xi\right]\right\} dv \\ &= 1 - x(T)\,x^{-1}(t) \exp\left[\int_T^t A\,dv\right] - x^{-1}(t) \int_T^t C(v) \left\{\exp\left[\int_v^t A\,d\xi\right]\right\} dv. \end{aligned}$$

Invoking Eq. (11) at any $t \geqslant T + K_2$ provides the inequality

$$V(t, T) \leqslant 1 - \mu - x(T)\,x^{-1}(t) \exp\left[\int_T^t A\,dv\right]$$

with $\mu \equiv (K_1/\sup x) \in (0, 1)$. Bringing together Eqs. (32), (35), (36), (39), (43), and (44), we finally conclude that

$$0 \leqslant X_i^{(\theta)}(t) \leqslant U^{(2)}(T) + (1 - \mu)[\tilde{Y}_i^{(\theta)}(T) + U^{(2)}(T)]$$

for any $T \in \mathcal{O}_i$ and all $t \geqslant T + K_2$. Inequality (45) and the existence of $\tilde{Y}_i^{(\theta)}(\infty)$ are the basis for all that follows.

Suppose first that there exists an increasing sequence of times $W_{ik} \in \mathcal{O}_i$, with $\lim_{k\to\infty} W_{ik} = \infty$, such that $\lim_{k\to\infty} \tilde{Y}_i^{(\theta)}(W_{ik}) = 0$. Then by Eq. (45), Q_i exists and equals θ_i, and the proof can readily be completed. Supposing that no such sequence of W_{ik}'s exists amounts to saying that $\tilde{Y}_i^{(\theta)}(T) \geqslant \eta$ for some $\eta > 0$ and all $T \in \mathcal{O}_i$. But since $\tilde{Y}_i^{(\theta)}(t)$ is monotone increasing for all $t \notin \mathcal{O}_i$, we can then assume that $\tilde{Y}_i^{(\theta)}(t) \geqslant \eta$ for all $t \geqslant 0$. Moreover, since in Eq. (45), $U^{(2)}(T)$ monotonically approaches zero as $T \to \infty$, there exists a $W_1 \geqslant 0$ and a $\nu \in (0, 1)$ such that

$$X_i^{(\theta)}(t) \leqslant (1 - \nu)\, \tilde{Y}_i^{(\theta)}(T) \tag{46}$$

if $T \in \mathcal{O}_i \cap [W_1, \infty)$ and $t \geqslant T + K_2$. Since trivially,

$$\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t) = \tilde{Y}_i^{(\theta)}(t) - \tilde{Y}_i^{(\theta)}(T) + \tilde{Y}_i^{(\theta)}(T) - X_i^{(\theta)}(t)$$

for any t and T, Eq. (46) shows that

$$\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t) \geqslant \nu \tilde{Y}_i^{(\theta)}(T) + \tilde{Y}_i^{(\theta)}(t) - \tilde{Y}_i^{(\theta)}(T)$$

if $T \in \mathcal{O}_i \cap [W_1, \infty)$ and $t \geqslant T + K_2$. Since $\mathcal{O}_i$ includes arbitrarily large numbers and $\tilde{Y}_i^{(\theta)}(\infty)$ exists, Eq. (47) implies the existence of a time W_2 such that

$$\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t) \geqslant (\nu/2)\, \tilde{Y}_i^{(\theta)}(\infty) \qquad \text{for} \quad t \geqslant W_2$$

and, in particular, that

$$\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t) \geqslant (\nu\eta/2) > 0 \qquad \text{for} \quad t \geqslant W_2$$

The inequality (48) will now be shown to be impossible, thereby completing the proof.

By Eq. (48), there exists a W_3 such that

$$\tilde{Y}_i^{(\theta)}(t) \geqslant \tilde{Y}_i^{(\theta)}(\infty) - (\nu\eta/8) > \tilde{Y}_i^{(\theta)}(\infty) - (3\nu\eta/8) \geqslant X_i^{(\theta)}(t), \qquad t \geqslant W_3 .$$

Thus if for any $j \in J(1)$ and any $t \geqslant W_3$, $y_{ji}^{(\theta)}(t) \leqslant \tilde{Y}_i^{(\theta)}(\infty) - (\nu\eta/4)$, then $y_{ji}^{(\theta)}(t) < Y_i^{(\theta)}(t)$ for all $t \geqslant W_3$. In other words, every $j \in J(1)$ such that $\tilde{Y}_i^{(\theta)}(t) = y_{ji}^{(\theta)}(t)$ at any $t \geqslant W_3$ satisfies $y_{ji}^{(\theta)}(t) - X_i^{(\theta)}(t) \geqslant \nu\eta/8$ for all $t \geqslant W_3$. By Eq. (16), $\dot{y}_{ji}^{(\theta)}(t) \leqslant -(\nu\eta/8)\, H_j(W_t, t)$ for $t \geqslant W_3$, and thus $\int_0^\infty E_j\, dv < \infty$, which by Eqs. (9) and (29) contradicts the fact that $j \in J(1)$, and completes the proof.

4. EXCITATORY NONRECURRENT FULL NETWORKS

Such a network of type (3)–(5) satisfies the equations

$$\dot{x}_j(t) = -\alpha_j x_j(t) + I_j(t),$$

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{k=1}^{m} [x_k(t-\tau_k) - \Gamma_k]^+ z_{ki}(t) + I_i(t), \tag{50}$$

and

$$\dot{z}_{ji}(t) = -u_j z_{ji}(t) + v_j[x_j(t-\tau_j) - \Gamma_j]^+ x_i(t),$$

where $j \in J = \{1, 2,..., m\}$ and $i \in I = \{m+1, m+2,..., m+n\}$. Theorem clearly implies the following result for such a system.

4.1. Corollary 2

Let the system (49)–(51) be given with nonnegative and continuous initial data and inputs. Suppose $I_i(t) = \theta_i I(t)$, where $\sum_{i=m+1}^{m+n} \theta_i = 1$, $\theta_i \geqslant 0$, and for every $T \geqslant 0$,

$$\int_T^{T+t} e^{-\alpha(T+t-v)} I(v)\, dv \geqslant K_1 \qquad \text{if} \quad t \geqslant K_2$$

for suitable positive constants K_1 and K_2. Also, let the solution be bounded. all limits P_{ji} and Q_i exist, with $Q_i = \theta_i$. If moreover

$$\int_0^\infty \left[\int_0^v \exp\left[-\alpha_j(v-\xi)\right] I_j(\xi)\, d\xi - \Gamma_j\right]^+ dv = \infty,$$

then $P_{ji} = \theta_i$.

In applications, the inputs often dominate an iterated input, in the sense that

$$I_j(t) = \sum_{k=0}^{\infty} J_{jk}(t - t_{jk})$$

where $J_{jk}(t) \geqslant J_j(t)$ and $0 < \delta_j \leqslant t_{j,k+1} - t_{jk} \leqslant \epsilon_j < \infty$ for all $t \geqslant 0$ and $k = 0, 1, 2,...$. Here $J_j(t)$ is an input pulse; namely, a continuous nonnegative function that is positive in a finite interval. The function $J_{jk}(t - t_{jk})$ can, for example, correspond to an experimental event with onset time t_{jk}. For any such input $I_j(t)$, $P_{ji} = \theta_i$ holds by Eq. (53).if

$$\sup_t \int_0^t \exp\left[-\alpha_j(t-v)\right] J_j(v)\, dv > \Gamma_j.$$

An excitatory nonrecurrent full network of type (3), (6), and (7) satisfies the equations (49), (50), and

$$\dot{z}_{ji}(t) = \{-u_j z_{ji}(t) + v_j x_i(t)\}[x_j(t-\tau_j) - \Gamma_j]^+.$$

Corollary 2 clearly holds for this system as well. Corollary 1 also holds for the more general systems in which $[x_j(t-\tau_j)-\Gamma_j]^+$ is replaced by $f_j(x_j(t-\tau_j))$ with $f_j(w)$ monotone increasing, continuous, and nonnegative.

5. INHIBITORY NONRECURRENT FULL NETWORKS

Such a network of type (3)–(5) satisfies the equations (49),

$$\dot{x}_i(t) = -\alpha x_i(t) - \beta \sum_{k=1}^{m} [x_k(t-\tau_k) - \Gamma_k]^+ z_{ki}(t) - I_i(t),$$

and

$$\dot{z}_{ji}(t) = -u_j z_{ji}(t) + v_j[x_j(t-\tau_j) - \Gamma_j]^+ [-x_i(t)]^+ \tag{56}$$

with nonpositive and continuous initial data for the x_i's, nonnegative initial data for the z_{ji}'s, and nonnegative inputs. Under these circumstances, Eqs. (55) and (56) can be transformed into Eqs. (50) and (51) using the change of variables $x_i \to \xi_i \equiv -x_i$. This is because $x_i(t) \leqslant 0$ for all $t \geqslant 0$, so that Eq. (56) can be written

$$\dot{z}_{ji} = -uz_{ji} + v_j[x_i(t-\tau_j) - \Gamma_j]^+ \xi_i,$$

whereas Eq. (55) becomes

$$\dot{\xi}_i(t) = -\alpha \xi_i(t) + \beta \sum_{k=1}^{m} [x_k(t-\tau_k) - \Gamma_k]^+ z_{ki}(t) + I_i(t).$$

Thus learning of inhibitory patterns is also a special case of Theorem 1. A similar remark holds for inhibitory networks obeying Eqs. (49) and (55) and

$$\dot{z}_{ji}(t) = \{-u_j z_{ji}(t) + v_j[-x_i(t)]^+\}[x_j(t-\tau_j) - \Gamma_j]^+. \tag{57}$$

Physiologically speaking, learning of inhibitory patterns can be interpreted as respondant conditioning of inhibitory transmitter production. Livingston[4] gives related data.

6. EXCITATORY COMPLETELY RECURRENT NETWORKS

Such a network of type (3)–(5) satisfies Eq. (51) and

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{k=1}^{n} [x_k(t-\tau_k) - \Gamma_k]^+ z_{ki}(t) + I_i(t),$$

where $i, j = 1, 2, \ldots, n$.

6.1. Corollary 3

Let Eqs. (51) and (58) be given with nonnegative and continuous initial data and inputs. Let $I_i(t) = \theta_i I(t)$, where $\sum_{i=1}^{n} \theta_i = 1$, $\theta_i > 0$, and $I(t)$ satisfies Eq. (52) with

$$K_1[1 - \exp(-\alpha K_2)]^{-1} > \max_i(\Gamma_i/\theta_i) \tag{59}$$

Also suppose that the solution is bounded. Then all limits Q_i and P_{ji} exist with $P_{ji} = Q_i = \theta_i$.

6.1.1. Proof. Let $x = \sum_{i=1}^{n} x_i$. Then by Eq. (58),

$$x(t) \geqslant x(nK_2) \exp[-\alpha(t - K_2)] + \int_{nK_2}^{t+nK_2} \exp[-\alpha(t + nK_2 - v)]\, I(v)\, dv$$

for every $t \in [nK_2, (n+1)K_2]$ and $n = 0, 1, 2, \ldots$. In particular, Eq. (52) implies

$$\begin{aligned} x[(n+1)K_2] &\geqslant x(nK_2)\exp(-\alpha K_2) + K_1 \\ &\geqslant K_1\{1 - \exp[-\alpha(n+1)K_2]\}[1 - \exp(-\alpha K_2)]^{-1}. \end{aligned}$$

Thus, given any $\delta > 0$, there exists a T_δ such that

$$x(t) \geqslant K_1[1 - \exp(-\alpha K_2)]^{-1} - \delta \tag{60}$$

for $t \geqslant T_\delta$. By Theorem 1, $Q_i = \theta_i$; and hence for every $\epsilon > 0$, there exists a T_ϵ such that

$$[x_j(t - \tau_j) - \Gamma_j]^+ \geqslant [(\theta_j - \epsilon)\, x(t - \tau_j) - \Gamma_j]^+$$

for all $t \geqslant T_\epsilon$. Thus for all $t \geqslant \max(T_\delta, T_\epsilon)$, Eq. (60) implies

$$[x_j(t - \tau_j) - \Gamma_j]^+ \geqslant [(\theta_j - \epsilon)\{K_1[1 - \exp(-\alpha K_2)]^{-1} - \delta\} - \Gamma_j]^+$$

and by Eq. (59), for sufficiently small δ and ϵ,

$$x_j(t - \tau_j) - \Gamma_j \geqslant (\theta_j - \epsilon)\{K_1[1 - \exp(-\alpha K_2)]^{-1} - \delta\} - \Gamma_j > 0.$$

In particular, for every j, $\int_0^\infty [x_j(v - \tau_j) - \Gamma_j]^+\, dv = \infty$, which by Theorem 1 implies $P_{ji} = \theta_i$.

Condition (59) means heuristically that a sequence of input pulses of sufficient intensity and/or duration can learn any pattern with positive weights. For example, let $I(t)$ be a periodic sequence of rectangular input pulses with intensity I, duration λ, and interpulse interval μ. Then by Eq. (59), all $P_{ji} = \theta_i$ if

$$\frac{I(e^{\alpha\lambda} - 1)}{\alpha(e^{\alpha(\lambda+\mu)} - 1)} > \max_i \frac{\Gamma_i}{\theta_i}. \tag{61}$$

Corollary 3 also holds for systems in which $[x_j(t - \tau_j) - \Gamma_j]^+$ is replaced by $f_j(x_j(t - \tau_j))$, where $f_j(w)$ is monotone increasing, continuous, and nonnegative. Then Eq. (59) is replaced by the condition $\min_j f_j(\theta_j K_1[1 - \exp(-\alpha K_2)]^{-1}) > 0$. For example, if $f_j(w) = \log(1 + \eta_j + w)$, for some $\eta_j > 0$, then $P_{ji} = \theta_i$ given any positive choices of K_1 and K_2 in Eq. (52). The appearance of a positive η_j in this choice of f_j means that the network axons are always spontaneously active, and therefore continually sample the pattern weights at other cells.

The upper bound $K_1[1 - \exp(-\alpha K_2)]^{-1}$ in Eq. (59) can usually be increased by iterating the equations (51) and (58) at equally spaced discrete time steps. For simplicity, we consider below the case in which all τ_j, u_j, and v_j are independent of j, and $K_2 = \tau$.

6.2. Corollary 4 ("Bootstraps")

Let Eqs. (51) and (58) be given with nonnegative and continuous initial data and inputs. Let all τ_j, u_j, and v_j be independent of j, and suppose $I_i(t) = \theta_i I(t)$, where $\sum_{i=1}^{n} \theta_i = 1$, $\theta_i > 0$, and $I(t)$ satisfies Eq. (52) with $K_2 = \tau$ for some $K_1 > 0$. Define the sequence $x^{(i)}$ by the initial data, $x^{(-1)} = x^{(0)} = 0$, and the recursion

$$x^{(i+1)} \geqslant K_1 + x^{(i)}e^{-\alpha\tau} + \frac{\beta v}{\alpha u}(1 - e^{-u\tau}) \sum_{k=1}^{n} [\theta_k x^{(i)} - \Gamma_k]^+ \sum_{m=1}^{i} [\theta_k x^{(m)} - \Gamma_k]^+ x^{(m)} e^{(m-i)u\tau} \tag{62}$$

for $i \geqslant 0$, and suppose

$$K > \max_i(\Gamma_i/\theta_i)$$

where $K = \limsup_{i\to\infty} x^{(i)}$. Also suppose that the solution of Eqs. (51) and (58) is bounded. Then all limits Q_i and P_{ji} exist with $P_{ji} = Q_i = \theta_i$.

6.2.1. Proof. By Theorem 1, $Q_i = \theta_i$. Hence, by choosing t sufficiently large, Eqs. (51) and (58) can be replaced by

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{k=1}^{n} [\theta_k x(t-\tau) - \Gamma_k]^+ z_{ki}(t) + \theta_i I(t)$$

and

$$\dot{z}_{ji}(t) = -u z_{ji}(t) + v[\theta_j x(t-\tau) - \Gamma_j]^+ x_i(t)$$

without destroying the validity of an inequality such as Eq. (62) at which we aim. In particular,

$$\dot{x}(t) = -\alpha x(t) + \beta \sum_{k=1}^{n} [\theta_k x(t-\tau) - \Gamma_k]^+ z_k(t) + I(t)$$

and

$$\dot{z}_j(t) = -u z_j(t) + v[\theta_j x(t-\tau) - \Gamma_j]^+ x(t).$$

Starting with zero initial data in Eqs. (64) and (65), we will find an asymptotic minorant for $x(t)$ by iterating Eqs. (64) and (65) every τ time units. This iteration yields

$$x^{(i+1)} \geqslant x^{(i)}e^{-\alpha\tau} + (\beta/\alpha) \sum_{k=1}^{n} [\theta_k x^{(i)} - \Gamma_k]^+ z_k^{(i)} + K_1 \tag{66}$$

and

$$z_k^{(i)} \geqslant z_k^{(i-1)}e^{-u\tau} + (v/u)(1 - e^{-u\tau})[\theta_k x^{(i)} - \Gamma_k]^+ x^{(i)}, \tag{67}$$

where $z_k^{(-1)} = 0$. Equation (67) implies

$$z^{(i)} \geqslant (v/u)(1 - e^{-u\tau}) \sum_{m=1}^{i} [\theta_k x^{(m)} - \Gamma_k]^+ x^{(m)} e^{(m-i)u\tau}$$

which, when substituted into Eq. (66), proves Eq. (61).

Since $K = \limsup_{i \to \infty} x^{(i)}$, Eq. (63) implies the existence of an $\epsilon > 0$ such that $x(t_k) \geqslant \epsilon + \max_i(\Gamma_i/\theta_i)$ on an increasing sequence of points t_k with $\lim_{k \to \infty} t_k = \infty$. By the boundedness of the solution and inputs, Eq. (58) also implies the boundedness of $\dot{x}(t)$. Thus, there exists a $\delta > 0$ such that $x(t) \geqslant (\epsilon/2) + \max_i(\Gamma_i/\theta_i)$ on an increasing sequence of disjoint intervals $[w_k, w_k + \delta]$. In particular, for every j, $\int_0^\infty [x_j(v - \tau) - \Gamma_j]^+ \, dv = \infty$, which by Theorem 1 implies $P_{ji} = \theta_i$.

Speaking heuristically, Eq. (63) says that if $K_1(1 - e^{-\alpha\tau})^{-1}$ is sufficiently large, and the gaps between successive values of $\Gamma_k \theta_k^{-1}$ are sufficiently small, then the interaction term $\sum_{k=1}^{n} [\theta_k x(t - \tau) - \Gamma_k]^+ z_k(t)$ can boost $x(t)$ up by its own "bootstraps." For example, if, for i sufficiently large, $x^{(i)} > \max_j(\Gamma_j/\theta_j)$, then

$$K \geqslant K_1(1 - e^{-\alpha\tau})^{-1} + (\beta v/\alpha u) \sum_{k=1}^{n} [\theta_k K_1(1 - e^{-\alpha\tau})^{-1} - \Gamma_k]^2.$$

In situations for which $I(t)$ dominates an iterated input, the above conditions can often be improved by estimating the maximum value of $x(t)$ created by an individual input pulse, rather than its minimum, and then iterating the maximum estimate. For example, condition (61) can then be replaced by

$$\frac{I(e^{\alpha\lambda} - 1)\, e^{\alpha\mu}}{\alpha(e^{\alpha(\lambda+\mu)} - 1)} > \max_i \frac{\Gamma_i}{\theta_i}$$

For such estimates, x_j need not exceed Γ_j at all large t to guarantee Eq. (12).

The constant K is a lower estimate for the maximum $\hat{K}$ of the set $L(C^+)$ of limit points of the curve C^+ defined by $x(t)$ for $t \geqslant 0$. A limit point L is any point for which there exists a sequence of times t_k with $\lim_{k \to \infty} t_k = \infty$ such that $\lim_{k \to \infty} x(t_k) = L$. Since the solution of Eqs. (51) and (58) is bounded and continuous, $L(C^+)$ is a compact interval (Hartman,[8] Chapter 7), and thus $\hat{K}$ exists. For every j such that $\hat{K} > \Gamma_j \theta_j^{-1}$, $P_{ji} = \theta_i$. For every j such that $\Gamma_j \theta_j^{-1} > \hat{K}$, there exists a T_j such that $[x_j(t) - \Gamma_j]^+ = 0$ for $t \geqslant T_j$. Thus, only if $\hat{K} = \Gamma_j \theta_j^{-1}$ can $\int_0^\infty [x_j(v) - \Gamma_j]^+ \, dv < \infty$ hold without $[x_j(t) - \Gamma_j]^+$ vanishing identically for all sufficiently large t.

7. COMPATIBILITY OF EQUIDISTRIBUTED PATTERN COMPLETION AND PERFECT MEMORY

In a nonrecurrent and full network, it follows from Theorem 1 that each sampling cell v_i, $i \in I$, can reproduce a given previously learned pattern on all the cells v_j, $j \in J$, without destroying the memory controlled by other cells v_i, $i \in I$. In a completely recurrent network, the existence of such a situation is not *a priori* evident, since perturbing a single cell v_i can perhaps create signals that reverberate throughout the network and thereby destroy the patterns encapsulated in the synaptic knobs. Indeed, suppose that the network given by Eqs. (51) and (58) has practiced the pattern with weights θ_i to a high degree; that is, $y_{ji}(t) \approx \theta_i$ when $I(t)$ is shut off. Let an input pulse be delivered to a single v_k at a later time as a recall trial. Will such pulses, if repeated sufficiently often in an effort to have the network reproduce its stored memory, eventually destroy the memory of the pattern weights? The answer is "yes" if all thresholds $\Gamma_j = 0$. The answer is "no" if the thresholds Γ_j are so large that the signal received at any v_i from v_k is insufficient to create further signals from v_i. Indeed, by Corollary 1, if all $[x_i(t - \tau_i) - \Gamma_i]^+ = 0$, $i \neq k$, then all $\dot{y}_{ij}(t) = 0$, whereas when $x_k(t - \tau) > \Gamma_k$, $y_{ki}(t)$ is attracted toward θ_i, and "reminiscence," or spontaneous memory improvement, occurs [cf. Grossberg,[9] Part II(1)]. In this sense, high signal thresholds localize the memory of the network and produce context effects in response to localized inputs (cf. Grossberg[10]).

The fact that each cell separately in a completely recurrent network can reproduce the entire pattern is a form of *pattern completion*. The fact that *any* cell can reproduce the pattern is called *equidistributed* (or *equipotential*, or *mass action*) pattern competion, and might suggest to an experimentalist surveying such a network that the memory is somehow diffusely spread over all the cells. Such an experimentalist might reasonably hope that this "mass action" effect of the cells implies a lack of specificity in the way memories are stored. Such an impression would be confirmed if very large test inputs were presented to the networks, since the signal thresholds could be readily overcome. (Contrast the leveling effects of inhibition, and compare the effects of electroshock.) But the network dynamics would seem paradoxical if small test inputs were presented, since then perfect memory, pattern completion, and specificity of pattern representation can be simultaneously achieved.

In networks such as those in Eqs. (51) and (58), even though the ratios $y_{ki}(t)$ remember pattern weights perfectly if all signals equal zero, the associational strengths z_{ki} decay exponentially. To potentiate the amount of transmitter without changing the weights $y_{ki}(t)$, it suffices, by the above remarks, to let small individual test inputs perturb each cell v_k separately at widely spaced times. In networks such as those in Eqs. (54) and (58), by contrast, all $z_{ki}(t)$ are constant in intervals when no signals are positive; hence, memory is perfect and transmitter potentiation is unnecessary. The difference between the decay law in Eqs. (2) and in (4) can be heuristically traced to whether or not Ca^{++} and Na^{+} enter all membrane channels through mutually independent pores.

8. CONDITIONS GUARANTEEING BOUNDEDNESS

This section lists some results that guarantee boundedness of excitatory nonrecurrent full networks and completely recurrent networks. The results show that arbitrarily large bounded inputs are compatible with an arbitrarily small memory decay rate u_j if the common decay rate α of all potentials $x_i(t)$ is sufficiently large. In both cases, for fixed u_j, the input can grow essentially like the square root of α, given Eq. (51). Throughout the following discussion, let $\beta = \max_j \beta_j$, $\Gamma = \min_j \Gamma_j$, $u = \min_i u_i$, and $v = \max_j v_j$.

8.1. Proposition 2

Let Eqs. (49)–(51) be given with nonnegative and continuous initial data and inputs. Suppose that there exists an $\epsilon > 0$ such that for all sufficiently large t,

$$\frac{2(\alpha u)^{1/2}}{\beta + v} \geqslant \epsilon + \sum_{j=1}^{m} \Big[\int_0^{t-\tau_j} \exp[-\alpha_j(t - \tau_j - v)]\, I_j(v)\, dv - \Gamma_j\Big]^+ .$$

Then the solution of Eqs. (49)–(51) is bounded. In particular, the solution is bounded if

$$\frac{2(\alpha u)^{1/2}}{\beta + v} \geqslant \epsilon + \sum_{j=1}^{m} \limsup_t \Big(\int_0^{t} \exp[-\alpha_j(t - v)]\, I_j(v)\, dv - \Gamma_j\Big),$$

so that sampling times of each v_j can then be arbitrarily chosen.

8.1.1. Proof.

Define U and V by the equations

$$\dot U(t) = -\alpha U(t) + \beta \sum_{j=1}^{m} [x_j(t - \tau_j) - \Gamma_j]^+ V(t) + I(t)$$

and

all v's are ν's.

$$\dot V(t) = -uV(t) + \beta \sum_{j=1}^{m} [x_j(t - \tau_j) - \Gamma_j]^+ U(t),$$

with initial data $U = x$ and $V = \sum_{j=1}^{m} z_j$. Then $U \geqslant x$ and $V \geqslant z_j$ for all $t \geqslant 0$. Hence, by Corollary 2 of Grossberg,[2] the solution of Eqs. (49)–(51) will be bounded if

$$\frac{2(\alpha u)^{1/2}}{\beta + v} \geqslant \epsilon + \sum_{j=1}^{m} [x_j(t - \tau_j) - \Gamma_j]^+$$

for all sufficiently large t and some $\epsilon > 0$, which is true if Eq. (68) holds.

The discussion of completely recurrent networks is more difficult because the terms $[x_j(t - \tau) - \Gamma_j]^+$ cannot be independently controlled. It suffices in applications to start the system in equilibrium (i.e., with zero initial data), and then to subject it to an arbitrary collection of suitably bounded inputs.

8.2. Proposition 3

Let Eqs. (51) and (58) be given with zero initial data, and let the total input $I(t)$ satisfy

$$\|I\|_\infty \leqslant \frac{\alpha + u}{4}\left\{\frac{3}{2} - \left[\frac{9}{4} - \frac{8\alpha u}{(\alpha + u)^2}\right]^{1/2}\right\} \times \left(\Gamma + \frac{\alpha + u}{2(\beta + v)}\left\{\frac{8\alpha u}{(\alpha + u)^2} - \frac{3}{2} + \left[\frac{9}{4} - \frac{8\alpha u}{(\alpha + u)^2}\right]^{1/2}\right\}^{1/2}\right).$$

Then the solution of Eqs. (51) and (58) is bounded. In particular, if $\alpha = (n-1)u$ and

$$\|I\|_\infty \leqslant \frac{2u}{3}\left(1 - \frac{1}{n}\right)\left[\Gamma + \frac{u[(16/3) - \delta]^{1/2}}{2(\beta + v)} n^{1/2}\right]$$

for some $\delta > 0$ and n sufficiently large, then the solution is bounded.

8.2.1. Proof. Define U and V by the equations

$$\dot{U}(t) = -\alpha U(t) + \beta[U(t-\tau) - \Gamma]^+ V(t) + I(t)$$

and

$$\dot{V}(t) = -uV(t) + v[U(t-\tau) - \Gamma]^+ U(t),$$

where U and V have zero initial data. Then $U \geqslant x$ and $V \geqslant z_i$ for all $t \geqslant 0$. Consider the function

$$\lambda(t) = -\tfrac{1}{2}(\alpha + u) + \tfrac{1}{2}((\alpha + u)^2 - \{4\alpha u - (\beta + v)^2 [U(t-\tau) - \Gamma]^2\})^{1/2}.$$

Suppose we could find an ϵ, $0 < \epsilon < 1$, such that

$$4\alpha u - (\beta + v)^2 [U(t-\tau) - \Gamma]^2 \geqslant \epsilon(\alpha + u)^2$$

for $t \geqslant 0$. Then $\lambda(t) \leqslant -\eta_\epsilon$, where

$$\eta_\epsilon = \tfrac{1}{2}(\alpha + u)[1 - (1-\epsilon)^{1/2}]. \tag{74}$$

Letting $N = (U^2 + V^2)^{1/2}$, Corollary 2 of Grossberg[2] implies, for every $t \geqslant 0$, that

$$U(t) \leqslant N(t) \leqslant \eta_\epsilon^{-1} \|I\|_\infty, \tag{75}$$

and thus the solution would be bounded. Equation (73) will follow if

$$\Gamma + \frac{[4\alpha u - \epsilon(\alpha + u)^2]^{1/2}}{\beta + v} \geqslant U(t-\tau).$$

This inequality, along with Eq. (75), can be achieved for all $t \geqslant 0$ if

$$\|I\|_\infty \leqslant \eta_\epsilon \left\{\Gamma + \frac{[4\alpha u - \epsilon(\alpha + u)^2]^{1/2}}{\beta + v}\right\}$$

for some ϵ such that $0 < \epsilon < 1$. In particular, letting

$$F(\epsilon) = [1 - (1-\epsilon)^{1/2}]\,(A - \epsilon)^{1/2}$$

where $A = 4\alpha u(\alpha - u)^{-2}$, it suffices to find an ϵ_0 such that $0 < \epsilon_0 < A$ maximizes $F(\epsilon)$, and then constrain $I(t)$ such that

$$\|I\|_\infty \leqslant \Gamma \eta_{\epsilon_0} + \frac{(\alpha + u)^2}{2(\beta + v)} F(\epsilon_0).$$

This we now do.

To maximize $F(\epsilon)$, make the change of variable $\delta^2 = 1 - \epsilon$, and define

$$f(\delta) \equiv F(1 - \delta^2) = (1 - \delta)(\delta^2 - B)^{1/2} \tag{77}$$

where $B = 1 - A$ and $\sqrt{B} \leqslant \delta \leqslant 1$. Computing the value $\delta = \delta_0$ which maximizes $f(\delta)$ for $\sqrt{B} \leqslant \delta \leqslant 1$ by solving $f'(\delta) = 0$ yields

$$\delta_0 = \tfrac{1}{4} + \tfrac{1}{2}(\tfrac{1}{4} + 2B)^{1/2} = \tfrac{1}{4} + \tfrac{1}{2}(\tfrac{9}{4} - 2A)^{1/2}.$$

Computing $\epsilon_0 = 1 - \delta_0^2$ and substituting in η_ϵ and $F(\epsilon)$ yields by Eq. (76) the criterion Eq. (69) for boundedness.

We now estimate the right-hand side of Eq. (69) in the case $\alpha = (n - 1)\, u$ for fixed u as n becomes large. Denoting this function of n by G_n, we find

$$G_n = \frac{nu}{4}\left\{\frac{3}{2} - \left[\frac{9}{4} - \frac{8(n-1)}{n^2}\right]^{1/2}\right\}$$
$$\times \left(\Gamma + \frac{nu}{2(\beta + v)}\left\{\frac{8(n-1)}{n^2} - \frac{3}{2} + \left[\frac{9}{4} - \frac{8(n-1)}{n^2}\right]^{1/2}\right\}^{1/2}\right)$$

G_n can be estimated from below as follows. For every $n \geqslant 1$

$$\frac{3}{2} - \left[\frac{9}{4} - \frac{8(n-1)}{n^2}\right]^{1/2} \geqslant \frac{8}{3}\frac{(n-1)}{n^2}$$

Furthermore the inequality

$$\left\{\frac{8(n-1)}{n^2} - \frac{3}{2} + \left[\frac{9}{4} - \frac{8(n-1)}{n^2}\right]^{1/2}\right\}^{1/2} \geqslant kn^{-1/2}$$

holds if

$$(16 - 3k^2)n \geqslant 16 + k^4 + \frac{16(n-1)}{n}\left[4 - k^2 - \frac{4}{n}\right],$$

which is true for n sufficiently large if $k = [(16/3) - \delta]^{1/2}$ for some δ such that $0 < \delta < 16/3$. Thus if $\alpha = (n - 1)\, u$ and n is sufficiently large,

$$G_n \geqslant \frac{2u}{3}\left(1 - \frac{1}{n}\right)\left[\Gamma + \frac{u[(16/3) - \delta]}{2(\beta + v)}\, n^{1/2}\right],$$

so that Eq. (70) implies Eq. (69) for n sufficiently large, and in turn the boundedness of the solution.

Boundedness for cases in which Eq. (54) replaces Eq. (51) can be studied using Corollary 3 of Grossberg,[2] or a direct analysis of oscillations. The latter procedure is illustrated below. The bound on inputs in these cases can grow like α rather than $\alpha^{1/2}$.

8.3. Proposition 4

Let Eqs. (49), (50), and (54) be given with nonnegative and continuous initial data and inputs such that

$$\frac{\alpha u}{\beta v} \geqslant \epsilon + \sum_{j=1}^{m} \left[\int_0^{t-\tau_j} \exp[-\alpha_j(t-\tau_j-v)]\, I_j(v)\, dv - \Gamma_j \right]^+$$

for some $\epsilon > 0$ and t sufficiently large. Then the solution is bounded.

8.3.1. Proof. For times t at which Eq. (78) holds, consider the system

$$\dot{U} = -\alpha U + \beta \sum_{j=1}^{m} W_j V_j + \| I \|_\infty$$

and

$$\dot{V}_j = (-u V_j + v U)\, W_j, \tag{80}$$

where $W_j = [x_j(t-\tau_j) - \Gamma_j]^+$, $j = 1, 2,..., m$, having initial data $U = x \equiv \sum_{i=m+1}^{m+n} x_i$ and $V_j = z_j \equiv \sum_{i=m+1}^{m+n} z_{ji}$. Then $U \geqslant x$ and $V_j \geqslant z_j$ thereafter.

Define $V = \max_j V_j$. Divide the time scale into mutually nonoverlapping intervals $A_1, B_1, A_2, B_2,...$ such that $uV(t) > vU(t)$ for $t \in \cup_n A_n$ and $uV(t) \leqslant vU(t)$ for $t \in \cup_n B_n$. For $t \in A_n = (a_{n1}, a_{n2})$, $uv^{-1}V(a_{n1}) \geqslant U(t)$, since $V(t)$ is monotone decreasing by Eq. (80). For $t \in B_n = [b_{n1}, b_{n2}]$,

$$\dot{U} \leqslant \alpha U \left[-1 + \frac{\beta v}{\alpha u} \sum_{j=1}^{m} W_j \right] + \| I \|_\infty ,$$

which implies

$$U(b_{n2}) \leqslant \max[U(b_{n1}), u(\epsilon \beta v)^{-1} \| I \|_\infty],$$

or by the preceding case,

$$U(b_{n+1,1}) \leqslant \max[U(b_{n1}), u(\epsilon \beta v)^{-1} \| I \|_\infty]$$

The boundedness of U readily follows, and, from this, the boundedness of all V_j by Eq. (80).

9. ENERGY-ENTROPY DEPENDENCE

This section shows that the total potential is maximized (minimized) if the spatial pattern comprising the input has minimum (maximum) entropy. These results can be extended to space-time patterns by approximating these patterns by sequences of spatial patterns. The heuristic point of these results is that the learning mechanism of the networks allows those environmental demands which have the most order in them to energetically drive more of the cellular filters needed to discriminate these demands. Of course, the patterns that are preferred will depend on the network geometry. Below we consider cases in which all geometrical asymmetries are eliminated by choosing all network parameters independent of j. First, we give examples of the solutions of nonrecurrent full networks as functions of the pattern weights $\theta = (\theta_1, \theta_2, ..., \theta_m)$. Thus let

$$\dot{x}_j^{(\theta)}(t) = -\alpha_1 x_j^{(\theta)}(t) + \theta_j J(t),$$

$$\dot{x}_i^{(\theta)}(t) = -\alpha x_i^{(\theta)}(t) + \beta \sum_{j=1}^{m} [x_j^{(\theta)}(t-\tau) - \Gamma]^+ z_{ji}^{(\theta)}(t) + I_i(t),$$

and

$$\dot{z}_{ji}^{(\theta)}(t) = -u z_{ji}^{(\theta)}(t) + \gamma [x_j^{(\theta)}(t-\tau) - \Gamma]^+ x_i^{(\theta)}(t),$$

where $j = 1, 2, ..., m$, and $i = m+1, m+2, ..., m+n$. Superscripts "(θ)" here refer only to dependence on the pattern θ, and do not mean $x_i^{(\theta)} = x_i - \theta_i$ as in Proposition 1. For example, denote the total potential of Eqs. (81)–(83) by $x^{(\theta)} = \sum_{i=m+1}^{m+n} x_i^{(\theta)}$. Let $x^{(\delta)}$ denote the total potential corresponding to a pattern θ with some $\theta_j = 1$, and let $x^{(f)}$ denote the total potential if all $\theta_j = 1/m$. Furthermore, define $\Omega(\theta) = \sum_{j=1}^{m} \theta_j^2$ for all θ.

9.1. Theorem 2 (Majorization of Total Potential)

Let the systems (81)–(83) be given with nonnegative and continuous initial data and inputs, and the same total inputs $J(t)$ and $I(t) = \sum_{i=m+1}^{m+n} I_i(t)$. Let all x_j, $j = 1, 2, ..., m$, have zero initial data for convenience. Also, suppose there exist positive constants μ, ν, and ω and a constant $T_{\mu\nu\omega}$ such that

$$\frac{2(\alpha u)^{1/2}}{(\beta + v)} \geqslant \mu + \int_0^t \exp[-\alpha_1(t-v)]\, J(v)\, dv, \qquad \text{where} \quad \alpha \geqslant u,$$

$$\int_0^t e^{-\alpha(t-v)} I(v)\, dv \geqslant \nu,$$

and

$$\int_0^t e^{-\alpha(t-v)} \left[\int_0^v \exp[-\alpha_1(v-\xi)]\, J(\xi)\, d\xi - \Gamma \right]^+ dv \geqslant \omega,$$

for $t \geqslant T_{\mu\nu\omega}$. Then for any pattern $\theta \neq \delta$, there exists an $\epsilon_{\delta\theta} > 0$ and a $T_{\delta\theta}$ such that

$$x^{(\delta)}(t) \geqslant x^{(\theta)}(t) + \epsilon_{\delta\theta} \qquad \text{for} \quad t \geqslant T_{\delta\theta} \tag{87}$$

If moreover the threshold $\Gamma = 0$, then for any two patterns θ_1 and θ_2, there exists an $\epsilon_{12} > 0$ and a T_{12} such that

$$x^{(\theta_2)}(t) \geqslant x^{(\theta_1)}(t) + \epsilon_{12} \qquad \text{if} \quad \Omega(\theta_2) > \Omega(\theta_1) \quad \text{and} \quad t \geqslant T_{12}. \tag{88}$$

Analogous results hold for $x^{(f)}$ as a minorante for all $x^{(\theta)}$.

9.1.1. Proof. By definition,

$$\dot{x}^{(\delta)}(t) = -\alpha x^{(\delta)}(t) + \beta[K(t) - \Gamma]^+ z^{(\delta)}(t) + I(t)$$

and

$$\dot{z}^{(\delta)}(t) = -uz^{(\delta)}(t) + \gamma[K(t) - \Gamma]^+ x^{(\delta)}(t), \tag{90}$$

with

$$K(t) = 0, \qquad t \leqslant \tau$$
$$= \int_0^{t-\tau} \exp[-\alpha_1(t - \tau - v)]\, J(v)\, dv, \qquad t > \tau$$

whereas

$$\dot{x}^{(\theta)}(t) = -\alpha x^{(\theta)}(t) + \beta \sum_{j=1}^{m} [\theta_j K(t) - \Gamma]^+ z_j^{(\theta)}(t) + I(t)$$

and

$$\dot{z}_j^{(\theta)}(t) = -uz_j^{(\theta)}(t) + \gamma[\theta_j K(t) - \Gamma]^+ x^{(\theta)}(t).$$

Letting $U = x^{(\delta)} - x^{(\theta)}$ and $V_j = z^{(\delta)} - z_j^{(\theta)}$, we find

$$\dot{U} = -\alpha U + \beta \sum_{j=1}^{m} [\theta_j K - \Gamma]^+ V_j + W$$

and

$$\dot{V}_j = -uV_j + \gamma[K - \Gamma]^+ U + Y_j\,,$$

where the functions

$$W = \beta z^{(\delta)} \left([K - \Gamma]^+ - \sum_{j=1}^{m} [\theta_j K - \Gamma]^+\right)$$

and

$$Y_j = \gamma x^{(\theta)}([K - \Gamma]^+ - [\theta_j K - \Gamma]^+)$$

are nonnegative. Given initial data such that $U \geqslant 0$ and all $V_j \geqslant 0$, then Eqs. (93) and (94) imply $U \geqslant 0$ and all $V_j \geqslant 0$ for $t \geqslant 0$. In fact, Eqs. (85) and (89) imply that $x^{(\delta)}(t) \geqslant \nu$ for $T_{u\nu\omega}$. Since $\alpha \geqslant u$, Eq. (90) implies that

$$z^{(\delta)}(t) \geqslant \gamma \int_0^t e^{-u(t-v)}[K(v) - \Gamma]^+ x^{(\delta)}(v)\, dv,$$

which by Eq. (86) shows that $z^{(\delta)}(t)$ has a positive lower bound for large t. But then by Eqs. (86) and (93), and the nonnegativity of all V_j, U also has a positive lower bound for large t, since $\theta \neq \delta$. In other words, Eq. (87) holds for suitable $\epsilon_{\delta,\theta}$ and $T_{\delta,\theta}$.

This has been shown only for a fortunate choice of initial data. The proof is completed by noting that under the hypothesis (84), any two solutions $(x^{(\delta,1)}, z^{(\delta,1)})$ and $(x^{(\delta,2)}, z^{(\delta,2)})$ of Eqs. (89) and (90) are *asymptotically equivalent*; that is,

$$\lim_{t\to\infty}[x^{(\delta,1)}(t) - x^{(\delta,2)}(t)] = \lim_{t\to\infty}[z^{(\delta,1)}(t) - z^{(\delta,2)}(t)] = 0.$$

This follows readily by the method used in Proposition 2, which in turn depends on Corollary 2 of Grossberg.[2]

To prove Eq. (88), let $\Gamma = 0$. We consider explicitly the case in which all $\theta_j > 0$. Choose the initial data such that $\theta_j^{-1}z_j(0)$ is independent of j. Then the system (89)–(90) takes the form

$$\dot{x}^{(\theta)}(t) = -\alpha x^{(\theta)}(t) + \beta\Omega(\theta)\, K(t)\, \xi^{(\theta)}(t) + I(t)$$

and

$$\dot{\xi}^{(\theta)}(t) = -u\xi^{(\theta)}(t) + \gamma K(t)\, x^{(\theta)}(t),$$

where $\xi(t) = \sum_{j=1}^{m} \theta_j^{-1}z_j^{(\theta)}(t)$. Now consider any two patterns θ_1 and θ_2 such that $\Omega(\theta_1) < \Omega(\theta_2)$. Comparison of two solutions $(x^{(\theta_1)}, \xi^{(\theta_1)})$ and $(x^{(\theta_2)}, \xi^{(\theta_2)})$ given equal initial data shows that Eq. (88) holds for some ϵ_{12} and T_{12}. Then asymptotic equivalence can be proved as above to derive Eq. (88) for all initial data.

The proof of inequality (87) for completely recurrent networks is more difficult at the stage of demonstrating asymptotic equivalence. A proof is given below in the case $\tau = 0$, under constraints which also guarantee that any two bounded solutions of the maximizing system do not oscillate relative to one another at large times. One might hope that analogous results hold for all $\tau > 0$.

Consider the system

$$-\alpha x_i^{(\theta)} + \beta \sum_{j=1}^{n} [x_j^{(\theta)} - \Gamma]^+ z_{ji}^{(\theta)} + \theta_i I \tag{97}$$

and

$$-uz_{ji}^{(\theta)} + \gamma[x_j^{(\theta)} - \Gamma]^+ x_i^{(\theta)}, \tag{98}$$

defined for any pattern $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ and $i, j = 1, 2, \ldots, n$. The pattern $\theta = \delta$ cannot be perfectly learned if $\Gamma > 0$. On the other hand, any pattern with positive weights that approximates δ can be learned by choosing Γ sufficiently small and I sufficiently large. Moreover, if $\theta = \delta$ with (say) $\theta_i = 1$, and if $x_i^{(\delta)}$ and $z_{ii}^{(\delta)}$ have positive initial data whereas all other $x_j^{(\delta)}$ and $z_{ji}^{(\delta)}$ have zero initial data, then by Proposition 1, the system (97)–(98) has the form

$$\dot{x}^{(\delta)} = -\alpha x^{(\delta)} + \beta[x^{(\delta)} - \Gamma]^+ z^{(\delta)} + I \tag{99}$$

and

$$\dot{z}^{(\delta)} = -uz^{(\delta)} + \gamma[x^{(\delta)} - \Gamma]^+ x^{(\delta)} \tag{100}$$

for all $t \geqslant 0$, where we have omitted subscripts i for convenience. We will now show that the system (99)–(100) majorizes all systems for which $\theta \neq \delta$ under suitable constraints.

9.2. Theorem 3

Let the systems (97)–(98) be given with nonnegative and continuous initial data and inputs such that the solutions are bounded and Eq. (52) holds. Suppose that $\alpha \geqslant u$ and that there exist positive constants μ and T_μ such that

$$\int_0^t e^{-\alpha(t-v)} \left[\int_0^v e^{-\alpha(v-\xi)} I(\xi)\, d\xi - \Gamma\right]^+ dv \geqslant \mu, \qquad t \geqslant T_\mu,$$

$$\beta v u^{-1} K^2 > \| I \|_\infty,$$

and

$$2\beta v u^{-2} M^{-2}([K - \Gamma]^+)^3 > \Gamma, \tag{103}$$

where K is defined by the recursion of Corollary 4 given $\theta = \delta$, and M is an upper bound for $x^{(\delta)}$ given any admissible initial data (cf. Proposition 3). Then given any pattern $\theta \neq \delta$, there exist positive constants $\epsilon_{\delta\theta}$ and $T_{\delta\theta}$ such that Eq. (87) holds. Moreover, the vector functions

$$f_{12}^{(\delta)} = \begin{pmatrix} x^{(\delta_1)} - x^{(\delta_2)} \\ z^{(\delta_1)} - z^{(\delta_2)} \end{pmatrix}$$

which compare any two bounded solutions $(x^{(\delta_1)}, z^{(\delta_1)})$ and $(x^{(\delta_2)}, z^{(\delta_2)})$ of (99)–(100), have fixed sign for large t and zero limit as $t \to \infty$.

A similar comparison theorem holds between $x^{(\theta)}$ and $x^{(f)}$.

9.2.1. Proof. Letting $U = x^{(\delta)} - x^{(\theta)}$ and $V_j = z^{(\delta)} - z_j^{(\theta)}$, we find

$$\dot{U} = -\alpha U + \beta \sum_{j=1}^{n} [x_j^{(\theta)} - \Gamma]^+ V_j + W$$

and

$$\dot{V}_j = -u V_j + v[x^{(\delta)} - \Gamma]^+ U + Y_j,$$

where

$$W = \beta z^{(\delta)} \sum_{j=1}^{n} ([x^{(\delta)} - \Gamma]^+ - [x_i^{(\theta)} - \Gamma]^+)$$

and

$$Y_j = v x^{(\theta)}([x^{(\theta)} - \Gamma]^+ - [x_j^{(\theta)} - \Gamma]^+)$$

are nonnegative if $U \geqslant 0$. Thus starting with initial data such that $U \geqslant 0$ and all $V_j \geqslant 0$ implies $U \geqslant 0$ and all $V_j \geqslant 0$ for $t \geqslant 0$. By Eq. (52) with $T = 0$ and Eq. (99), $x^{(\delta)}$ has a positive lower bound for large t. Thus by Eqs. (100) and (101) and the inequality $\alpha \geqslant u$, $z^{(\delta)}$ has a positive lower bound for large t. By Eqs. (104) and (105), this will imply the existence of a positive lower bound for U if

$$\sum_{j=1}^{n} \int_0^t e^{-\alpha(t-v)}([x^{(\delta)}(v) - \Gamma]^+ - [x_j^{(\theta)}(v) - \Gamma]^+)\, dv$$

has a positive lower bound for large t. By Eq. (52) and the boundedness of solutions, Theorem 1 implies that $[x_i^{(\theta)} - \Gamma]^+ \sim [\theta_i x^{(\theta)} - \Gamma]^+$, where $x^{(\theta)} \geqslant x^{(\theta)}$. Thus by Eq. (108), U has a positive lower bound for large t, and Eq. (87) holds for this choice of initial data.

The above analysis can be carried out for all $\tau \geqslant 0$. The proof below of asymptotic equivalence for solutions of (99)–(100) holds only if $\tau = 0$. We will assume that $\Gamma > 0$ below, since the case $\Gamma = 0$ is more easily treated.

For simplicity, denote any two solutions of (99)–(100) by (x_1, z_1) and (x_2, z_2), and define the variables $F_{ij} = x_i x_j^{-1} - 1$, $G_{ij} = z_i^{1/2} z_j^{-1/2} - 1$, and $H_{ij} = f_{ij} - g_{ij}$ for $\{i, j\} = \{1, 2\}$. We will derive systems of the form

$$\dot{F}_{ij} = a_{ij} F_{ij} + b_{ij} G_{ij}, \tag{109}$$

$$\ddot{F}_{ij} = c_{ij} F_{ij} + d_{ij} G_{ij}, \tag{110}$$

$$\dot{G}_{ij} = -e_{ij} G_{ij} + f_{ij} F_{ij}, \tag{111}$$

$$\ddot{G}_{ij} = g_{ij} G_{ij} + h_{ij} H_{ij}, \tag{112}$$

and

$$\dot{H}_{ij} = k_{ij} H_{ij} + m_{ij} G_{ij} \tag{113}$$

where for both choices of $\{i, j\} = \{1, 2\}$, all coefficients are continuous and bounded, and the coefficients b_{ij}, c_{ij}, e_{ij}, f_{ij}, g_{ij}, and h_{ij} have positive lower bounds for large t. For one choice of $\{i, j\} = \{1, 2\}$, the coefficient m_{ij} will have a positive lower bound for large t if $F_{ij}(t)\, G_{ij}(t) \geqslant 0$. For the same choice of $\{i, j\} = \{1, 2\}$, the coefficient d_{ij} will have a positive lower bound if also $F_{ij}(t)\, H_{ij}(t) \leqslant 0$ for large t. Moreover, all second derivatives $\ddot{F}_{ij}$, $\ddot{G}_{ij}$, and $\ddot{H}_{ij}$ are bounded.

Using these facts, the proof can be completed as follows. Since $b_{ij} \geqslant 0$ and $f_{ij} \geqslant 0$, Eqs. (109) and (111) imply that F_{ij} and G_{ij} change sign at most once, and not at all for $t \geqslant t_0$ if $F_{ij}(t_0)\, G_{ij}(t_0) \geqslant 0$. Suppose $F_{ij}(t)\, G_{ij}(t) \leqslant 0$ at all large t. Then since also $e_{ij} \geqslant 0$, $\dot{G}_{ij}$ has fixed sign for large t, whence $\lim_{t\to\infty} G_{ij}(t)$ exists. But $\ddot{G}_{ij}$ is bounded, and thus also $\lim_{t\to\infty} \dot{G}_{ij}(t) = 0$. Since e_{ij} and f_{ij} have positive lower bounds, Eq. (111) implies $\lim_{t\to\infty} F_{ij}(t) = \lim_{t\to\infty} G_{ij}(t) = 0$, which completes the proof in this case. In the only remaining case, $F_{ij}(t)\, G_{ij}(t) > 0$ for large t.

Supposing that $F_{ij}(t)\, G_{ij}(t) > 0$ for large t implies the nonnegativity of m_{ij}. Using this fact along with the constant sign of G_{ij} for large t shows that H_{ij} changes sign at most once for large t, and not at all for $t \geqslant t_0$ if $G_{ij}(t_0)\, H_{ij}(t_0) \geqslant 0$. Suppose $G_{ij}(t)\, H_{ij}(t) \geqslant 0$ for large t. Then by Eq. (112) and the nonnegativity of g_{ij} and h_{ij}, $\dot{G}_{ij}(t)$ has fixed sign for large t, $\lim_{t\to\infty} G_{ij}(t)$ exists, and we argue as above that $\lim_{t\to\infty} G_{ij}(t) = \lim_{t\to\infty} H_{ij}(t) = 0$, which completes the proof in this case.

It remains only to consider the case for which $F_{ij}(t)\, G_{ij}(t) \geqslant 0$ and $G_{ij}(t)\, H_{ij}(t) \leqslant 0$ for large t. In this case, also $F_{ji}(t)\, G_{ji}(t) \geqslant 0$ and $G_{ji}(t)\, H_{ji}(t) \leqslant 0$ for large t, by the definitions of F_{ij}, G_{ij}, and H_{ij}. For at least one choice of $\{i, j\} = \{1, 2\}$, however, c_{ij} and d_{ij} have positive lower bounds. For this choice, Eq. (110) shows that $\dot{F}_{ij}(t)$ has fixed sign for large t, which as above shows that $\lim_{t\to\infty} F_{ij}(t) = \lim_{t\to\infty} G_{ij}(t) = 0$, thereby completing the proof.

The system (109)–(113) will now be derived. We will show that for large t,

$$a_{ij} = x_j^{-1}(\beta\Gamma z_j - I), \qquad b_{ij} = \beta z_j x_j^{-1}(x_i - \Gamma)(1 + C_{ij}),$$

$$c_{ij} = x_j^{-1}(\beta z_j x_i - I), \qquad d_{ij} = \beta z_j x_j^{-1}(x_i - \Gamma)(2 + C_{ij} - A$$

$$e_{ij} = \tfrac{1}{2}\upsilon x_j z_j^{-1}[(x_i - \Gamma)\, C_{ij}^{-1} + (x_j - \Gamma)], \qquad f_{ij} = \tfrac{1}{2}\upsilon x_j z_j^{-1}[(x_i - \Gamma)\, C_{ij}^{-1} + x_j],$$

$$g_{ij} = \tfrac{1}{2}\upsilon\Gamma x_j z_j^{-1}, \qquad h_{ij} = \tfrac{1}{2}\upsilon x_j z_j^{-1}[(x_i - \Gamma)\, C_{ij}^{-1} + x_j],$$

$$k_{ij} = \tfrac{1}{2}\upsilon x_j z_j^{-1}[(\Gamma - x_i)\, C_{ij}^{-1} - x_j] + x_j^{-1}(\beta\Gamma z_j - I),$$

and

$$m_{ij} = \beta z_j x_j^{-1}(x_i - \Gamma)(1 + C_{ij}) - \tfrac{1}{2}\upsilon\Gamma x_j z_j^{-1} + x_j^{-1}(\beta\Gamma z_j - I).$$

First use the equation $(fg^{-1})^{\cdot} = g^{-1}(\dot{f} - f\dot{g}g^{-1})$ on the functions $A_{ij} = x_i x_j^{-1}$ and $B_{ij} = z_i z_j^{-1}$ to prove that for large t,

$$\dot{A}_{ij} = x_j^{-1}(I - \beta\Gamma z_j)(1 - A_{ij}) + \beta z_j x_j^{-1}(x_i - \Gamma)(B_{ij} - 1)$$

and

$$\dot{B}_{ij} = \upsilon x_j^2 z_j^{-1}(A_{ij}^2 - B_{ij}) + \upsilon\Gamma x_j z_j^{-1}(B_{ij} - A_{ij}).$$

The term $(x_i - \Gamma)$ equals $[x_i - \Gamma]^+$ for large t by Eq. (103). Then letting $C_{ij} = B_{ij}^{1/2}$, we find

$$A_{ij} = x_j^{-1}(\beta\Gamma z_j - I)(A_{ij} - 1) + \beta z_j x_j^{-1}(x_i - \Gamma)(1 + C_{ij})(C_{ij} - 1)$$

and

$$\dot{B}_{ij} = \upsilon x_j^2 z_j^{-1}(A_{ij} + C_{ij})(A_{ij} - C_{ij}) + \upsilon\Gamma x_j z_j^{-1}(1 + C_{ij})(C_{ij} - 1) + \upsilon\Gamma x_j z_j^{-1}(1 - A_{ij}). \tag{115}$$

Equation (114) can also be written as

$$\dot{A}_{ij} = x_j^{-1}[\beta\Gamma z_j - I + \beta z_j(x_i - \Gamma)](A_{ij} - 1) + \beta z_j x_j^{-1}(x_i - \Gamma)(2 + C_{ij} - A_{ij})(C_{ij} - 1), \tag{116}$$

and Eq. (115) can be written in two ways as

$$\dot{B}_{ij} = \upsilon x_j z_j^{-1}[x_j(A_{ij} + C_{ij}) - \Gamma](A_{ij} - 1) + \upsilon x_j z_j^{-1}[\Gamma(1 + C_{ij}) - x_j(A_{ij} + C_{ij})](C_{ij} - 1) \tag{117}$$

and

$$\dot{B}_{ij} = \upsilon x_j z_j^{-1}[x_j(A_{ij} + C_{ij}) - \Gamma](A_{ij} - C_{ij}) + \upsilon\Gamma x_j z_j^{-1} C_{ij}(C_{ij} - 1). \tag{118}$$

Since $\dot{C}_{ij} = \frac{1}{2}C_{ij}^{-1}\dot{B}_{ij}$, Eqs. (117) and (118) imply

$$\dot{C}_{ij} = \tfrac{1}{2}vx_jz_j^{-1}C_{ij}^{-1}[(x_i - \Gamma) + x_jC_{ij}](A_{ij} - 1) + \tfrac{1}{2}vx_jz_j^{-1}C_{ij}^{-1}[(\Gamma - x_i) + C_{ij}(\Gamma - x_j)](C_{ij} - 1)$$

and

$$\dot{C}_{ij} = \tfrac{1}{2}vx_jz_j^{-1}C_{ij}^{-1}[(x_i - \Gamma) + x_jC_{ij}](A_{ij} - C_{ij}) + \tfrac{1}{2}v\Gamma x_jz_j^{-1}(C_{ij} - 1).$$

Now subtract Eq. (120) from Eq. (116) to find

$$\begin{aligned}(A_{ij} - C_{ij})^{\cdot} = {} & \{\tfrac{1}{2}vx_jz_j^{-1}C_{ij}^{-1}[(\Gamma - x_i) - x_jC_{ij}] + x_j^{-1}(\beta\Gamma z_j - I)\}(A_{ij} - C_{ij}) \\ & + [\beta z_jx_j^{-1}(x_i - \Gamma)(1 + C_{ij}) \\ & - \tfrac{1}{2}v\Gamma x_jz_j^{-1} + x_j^{-1}(\beta\Gamma z_j - I)](C_{ij} - 1).\end{aligned} \tag{121}$$

Noting that $F_{ij} = A_{ij} - 1$, $G_{ij} = B_{ij} - 1$, and $H_{ij} = A_{ij} - B_{ij}$, Eqs. (114), (116), (119), (120), and (121) take the form (109)–(113) given the above definitions of the coefficients.

The coefficients b_{ij}, e_{ij}, f_{ij}, g_{ij}, and h_{ij} have positive lower bounds for large t if $\Gamma > 0$, since by Eq. (103) and the boundedness of solutions, both $x_j - \Gamma$ have positive upper and lower bounds for large t. The coefficient c_{ij} has a positive lower bound for large t by Eq. (102). Coefficient m_{ij} also has a positive lower bound for large t if $G_{ij} \geqslant 0$ (i.e., $C_{ij} \geqslant 1$) for some choice of $\{i, j\} = \{1, 2\}$, since then

$$\begin{aligned} m_{ij} & \geqslant 2\beta z_jx_j^{-1}(x_i - \Gamma) - \tfrac{1}{2}v\Gamma x_jz_j^{-1} + x_j^{-1}(\beta\Gamma z_j - I) \\ & = x_j^{-1}(\beta z_jx_i - I) + \beta x_jz_j^{-1}[z_j^2x_j^{-2}(x_i - \Gamma) - \tfrac{1}{2}v\beta^{-1}\Gamma].\end{aligned}$$

By Eqs. (102) and (103), respectively, $(\beta z_jx_i - I)$ and $z_j^2x_j^{-2}(x_i - \Gamma) - \frac{1}{2}v\beta^{-1}\Gamma$ have positive lower bounds for large t.

It remains only to show that d_{ij} has a positive lower bound for large t if $H_{ij} \leqslant 0$ (i.e., $C_{ij} \geqslant A_{ij}$), given the above choice of $\{i, j\} = \{1, 2\}$. In this case,

$$d_{ij} \geqslant 2z_jx_j^{-1}(x_i - \Gamma),$$

from which the claim is obvious. The proof is therefore complete.

In networks free from learning, such as the nonrecurrent full network

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{j=1}^{m} [\theta_jK(t) - \Gamma]^+ + I_i(t),$$

$i = m + 1, m + 2,..., m + n$, and the completely recurrent network

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{j=1}^{n} [x_j(t - \tau) - \Gamma]^+ + \theta_iI(t),$$

$i = 1, 2, \ldots, n$, estimates of the above type can be proved if $\Gamma > 0$. If $\Gamma = 0$, $x(t)$ is independent of θ. Such networks are presumably found nearest to peripheral receptors, where an unbiased response to experiential inputs is required, as illustrated by the formal reduction of our equations to the Hartline–Ratliff equation in the absence of learning, in response to steady-state inputs, and in the presence of purely inhibitory interactions across cells (Grossberg,[6] Section 13).

REFERENCES

1. S. Grossberg, "Some networks that can learn, remember, and reproduce any number of complicated space-time patterns (I)," *J. Math. Mechanics* (July 1969).
2. S. Grossberg, "Some networks that can learn, remember, and reproduce any number of complicated space-time patterns (II)," *SIAM J. Applied Math.*, submitted for publication.
3. S. Grossberg, "How do overlapping nonrecurrent synaptic fields learn without mutual interference?" (in preparation).
4. R. B. Livingston, "Brain mechanisms in conditioning and learning," in: *Neurosciences Research Symposium Summaries*, F. O. Schmitt *et al.*, eds. (MIT Press, Cambridge, Massachusetts, 1967), Vol. 2.
5. S. Grossberg, "On neural pattern discrimination," *J. Theoret. Biol.*, submitted for publication.
6. S. Grossberg, "On learning, information, lateral inhibition, and transmitters," *Math. Biosci.* **4**:255–310 (1969).
7. S. Grossberg, "A prediction theory for some nonlinear functional-differential equations (II)," *J. Math. Anal. Appl.* **22**:490–522 (1968).
8. P. Hartman, *Ordinary Differential Equations* (John Wiley and Sons, New York, 1964).
9. S. Grossberg, "A prediction theory for some nonlinear functional-differential equations (I)," *J. Math. Anal. Appl.* **21**:643–694 (1968).
10. S. Grossberg, "Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity," *Proc. Natl. Acad. Sci. (U.S.)* **59**:368–372 (1968).

PRINTED IN BRUGES, BELGIUM, BY THE ST. CATHERINE PRESS LTD.