

# GLOBAL RATIO LIMIT THEOREMS FOR SOME NONLINEAR FUNCTIONAL-DIFFERENTIAL EQUATIONS. I

BY STEPHEN GROSSBERG

Communicated by Norman Levinson, August 31, 1967

**1. Introduction.** We study some systems of nonlinear functional-differential equations of the form

$$(1) \quad \dot{X}(t) = AX(t) + B(X_t)X(t - \tau) + C(t), \quad t \geq 0,$$

where  $X = (x_1, \dots, x_n)$  is nonnegative,  $B(X_t) = \|B_{ij}(t)\|$  is a matrix of nonlinear functionals of  $X(w)$  evaluated at all past times  $w \in [-\tau, t]$ , and  $C = (C_1, \dots, C_n)$  is a known nonnegative and continuous input function. For appropriate  $A$ ,  $B$ , and  $C$ , these systems can be interpreted as a nonstationary prediction theory whose goal is to discuss the prediction of individual events, in a fixed order, and at prescribed times, or alternatively as a mathematical learning theory. This interpretation is discussed in a special case in [1]. The systems can also be interpreted as cross-correlated flows on networks, or as deformations of probabilistic graphs.

The mathematical content of these interpretations is contained in assertions of the following kind: given arbitrary positive and continuous initial data along with a suitable input  $C$ , the ratios  $y_{jk}(t) = B_{kj}(t) (\sum_{m=1}^n B_{km}(t))^{-1}$  have limits as  $t \rightarrow \infty$ .

Our systems are defined in the following way. Given any positive integer  $n$ ; any real numbers  $\alpha, u, \beta > 0$ , and  $\tau \geq 0$ ; and any  $n \times n$  semistochastic matrix  $P = \|p_{ij}\|$  (i.e.,  $p_{ij} \geq 0$  and  $\sum_{m=1}^n p_{im} = 0$  or  $1$ ), let

$$(2) \quad \dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{k=1}^n x_k(t - \tau) y_{ki}(t) + C_i(t),$$

$$(3) \quad y_{jk}(t) = p_{jk} z_{jk}(t) \left[ \sum_{m=1}^n p_{jm} z_{jm}(t) \right]^{-1},$$

and

$$(4) \quad \dot{z}_{jk}(t) = [-u z_{jk}(t) + \beta x_j(t - \tau) x_k(t)] \theta(p_{jk}),$$

for all  $i, j, k = 1, 2, \dots, n$ , where

$$\begin{aligned} \theta(p) &= 1 \quad \text{if } p > 0, \\ &= 0 \quad \text{if } p \leq 0. \end{aligned}$$

In order that our theorems hold, the initial data must always be non-

negative. We also require it to be continuous and for convenience suppose that  $z_{jk}(0) > 0$  iff  $p_{jk} > 0$ .

**2. Positivity and linear averages.**

**THEOREM 1.** *With initial data chosen as above in  $[-\tau, 0]$ , the solution of (2)–(4) exists and is unique, continuously differentiable, and non-negative in  $(0, \infty)$ . If moreover either  $x_i$  or  $z_{jk}$  has positive initial data, then it is always positive.*

The positivity of solutions implies a property of (2)–(4) that is used repeatedly in proving our results. Define the sets  $S(r)$  and  $T(r)$  inductively by  $S(r) = \{k: \sum_{i \in S(r-1)} p_{ki} = 1\}$  and

$$T(r) = \left\{ k: \sum_{i \in S(r-1)} p_{ki} = 0 \right\}, \quad r = 1, \dots, k,$$

where  $S(0) = \{1, 2, \dots, n\}$  and  $k$  is the least integer such that either  $S(k) = \emptyset$  or  $S(k) = S(k-1)$ . We also let  $x^{(r)} = \sum_{i \in S(r)} x_i$  and  $C^{(r)} = \sum_{i \in S(r)} C_i$ .

**COROLLARY 1.** *The vectors  $V = (x^{(0)}, \dots, x^{(k-1)})$  and  $W = (C^{(0)}, \dots, C^{(k-1)})$  obey a linear equation*

$$(5) \quad \dot{V}(t) = -\alpha V(t) + \beta DV(t - \tau) + W(t)$$

iff  $S(r) \cup T(r) = S(0), r = 1, 2, \dots, k$ , where

$$D = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdot & \cdot \\ \vdots & & & & & \\ 0 & 0 & \cdot & \cdot & 0 & 1 \\ 0 & 0 & \cdot & \cdot & 0 & 1 \end{pmatrix}$$

when  $S(k) = S(k-1)$ , and

$$D = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & \cdot & \cdot & \cdot \\ \vdots & & & & & \\ 0 & 0 & \cdot & \cdot & 0 & 1 \\ 0 & 0 & \cdot & \cdot & 0 & 0 \end{pmatrix}$$

when  $S(k) = \emptyset$ . If moreover  $P$  is stochastic (i.e.,  $\sum_{m=1}^n p_{im} = 1$  for all  $i$ ), then (5) reduces to

$$\dot{x}^{(0)}(t) = -\alpha x^{(0)}(t) + \beta x^{(0)}(t - \tau) + C^{(0)}(t).$$

**3. A graph theoretic interpretation.** The limiting behavior of (2)–(4) depends crucially on its matrix  $P$ . Every  $P$  can be geometrically realized as a directed probabilistic graph with vertices  $V = \{v_i: i=1, 2, \dots, n\}$  and directed edges  $E = \{e_{jk}: j, k=1, 2, \dots, n\}$ , where the weight  $p_{jk}$  is assigned to the edge  $e_{jk}$ . If moreover  $x_i(t)$  is interpreted as the state of a process at  $v_i$ , and  $y_{jk}(t)$  is interpreted as the state of a process at the arrowhead of  $e_{jk}$ , then (2)–(4) can readily be thought of as a flow of the quantities  $x_i(t)$  over the probabilistic graph  $P$  with flow velocity  $v=1/\tau$ . The coefficients  $y_{ki}(t)$  in (2) control the size of the  $\beta x_k(t-\tau)$  flow from  $v_k$  along  $e_{ki}$  which eventually reaches  $v_i$  by cross-correlating past  $\beta x_k(w-\tau)$  and  $x_i(w)$  values,  $w \in [-\tau, \tau]$ , with an exponential weighting factor  $e^{-u(t-w)}$  as in  $z_{ki}(t)$  in (4), and comparing this weighted cross-correlation in (3) with all other cross-correlations  $z_{km}(t)$  corresponding to any edge leading from  $v_k$ ,  $m=1, 2, \dots, n$ . (See [1] for further details.)

Alternatively, for every  $t \geq 0$ , a probabilistic graph  $G(t)$  with weight  $y_{jk}(t)$  assigned to edge  $e_{jk}$  can be defined. Then (2)–(4) provides a mechanism for continuously deforming one graph  $G(t_0)$  into another graph  $G(t_1)$ ,  $t_1 > t_0$ . A basic problem when  $C \equiv 0$  is to study the influence of the “geometry”  $P$  on the “limiting transition probabilities”  $G(\infty) = \lim_{t \rightarrow \infty} G(t)$  when these exist.

**4. Outstars.** In this note, we announce a result for the case

$$P = \begin{pmatrix} 0 & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\ & & 0 & & \\ & & & & \end{pmatrix}.$$

Then only edges  $e_{1j}$ ,  $j=2, 3, \dots, n$ , have positive weights, which equal  $1/(n-1)$ . This system is therefore called an *outstar* with *source* vertex  $v_1$ , *sinks*  $v_j$ ,  $j=2, \dots, n$ , and *border*  $B = \{v_j: j=2, \dots, n\}$ .

Our main result describes a sequence  $G^{(1)}, G^{(2)}, \dots, G^{(N)}, \dots$  of outstars with identical but otherwise arbitrary positive and continuous initial data in  $[-\tau, 0]$ , whose inputs are formed from the following ingredients:

(a) let  $\{\theta_j: j=2, \dots, n\}$  be a fixed but arbitrary probability distribution;

(b) let  $f$  and  $g$  be bounded, nonnegative, and continuous functions in  $[0, \infty)$  for which there exist positive constants  $k$  and  $T_0$  such that

$$\int_0^t e^{-\alpha(t-w)} f(w) dw \geq k, \quad t \geq T_0,$$

and

$$\int_0^t e^{-\alpha(t-w)}g(w)dw \geq k, \quad t \geq T_0;$$

(c) let  $U_1(N)$  and  $U(N)$  be any positive and monotone increasing functions of  $N \geq 1$  such that

$$\lim_{N \rightarrow \infty} U_1(N) = \lim_{N \rightarrow \infty} U(N) = \infty;$$

(d) for every  $N \geq 1$ , let  $h_N(t)$  be any nonnegative and continuous function that is positive only in  $(U(N), \infty)$ ; and

(e) let

$$\begin{aligned} \chi(w) &= 0 \quad \text{if } w > 0, \\ &= 1 \quad \text{if } w \leq 0. \end{aligned}$$

The input functions  $C_i^{(N)}$  of  $G^{(N)}$  are defined in terms of (a)–(e) by

$$(6) \quad C_1^{(N)}(t) = f(t)\chi(t - U_1(N)) + h_N(t)$$

and

$$(7) \quad C_j^{(N)}(t) = \theta_j g(t)\chi(t - U(N)), \quad j = 2, \dots, n.$$

Letting the functions of  $G^{(N)}$  be denoted by superscripts “ $(N)$ ” (e.g.,  $y_{1j}$  is written  $y_{1j}^{(N)}$ ), and defining the ratios  $X_j^{(N)} = x_j^{(N)} / \sum_{k=2}^n x_k^{(N)}$  for every  $N \geq 1$  and  $j = 2, \dots, n$ , we can state the following theorem.

**THEOREM 2.** *Let  $G^{(1)}, G^{(2)}, \dots, G^{(N)}, \dots$  have identical but otherwise arbitrary positive and continuous initial data, and any inputs chosen as in (6) and (7). Then*

(A) *for every  $N \geq 1$ , the limits  $\lim_{t \rightarrow \infty} X_j^{(N)}(t)$  and  $\lim_{t \rightarrow \infty} y_{1j}^{(N)}(t)$  exist and are equal,  $j = 2, \dots, n$ ;*

(B) *for every  $N \geq 1$  and all  $t \geq U(N)$ ,  $X_j^{(N)}(t)$  and  $y_{1j}^{(N)}(t)$  are monotonic in opposite senses, and*

$$\lim_{N \rightarrow \infty} X_j^{(N)}(U(N)) = \lim_{N \rightarrow \infty} y_{1j}^{(N)}(U(N)) = \theta_j,$$

$j = 2, \dots, n$ . In particular, by (A) and (B),

$$\lim_{N \rightarrow \infty} \lim_{t \rightarrow \infty} X_j^{(N)}(t) = \lim_{N \rightarrow \infty} \lim_{t \rightarrow \infty} y_{1j}^{(N)}(t) = \theta_j, \quad j = 2, \dots, n.$$

(C) *for every  $N \geq 1$  and  $j = 2, \dots, n$ , the functions  $\hat{y}_{1j}^{(N)}, F_j^{(N)} = y_{1j}^{(N)} - X_j^{(N)}$ , and  $G_j^{(N)} = X_j^{(N)} - \theta_j$  change sign at most once and not at*

all if  $F_j^{(N)}(0)G_j^{(N)}(0) \geq 0$ . Moreover,  $F_j^{(N)}(0)G_j^{(N)}(0) > 0$  implies  $F_j^{(N)}(t)G_j^{(N)}(t) > 0$  for all  $t \geq 0$ .

(C) shows in particular that the functions  $y_{ij}^{(N)}$  are quite insensitive to fluctuations in the functions  $f$  and  $g$ , since  $y_{ij}^{(N)}$  fluctuates no more than once.

In prediction and learning theoretic applications, the following situations are of particular interest.

**COROLLARY 2.** *If  $X_j^{(N)}(0) = y_{1j}^{(N)}(0)$  and  $\theta_j = \delta_{j2}$ ,  $j = 2, \dots, n$ , then  $y_{12}^{(N)}$  increases monotonically to 1 and  $y_{1k}^{(N)}$  decreases monotonically to zero,  $k = 3, \dots, n$ .*

**COROLLARY 3.** *The theorem is true if*

$$C_1^{(N)}(t) = \sum_{k=0}^{N-1} J_1(t - k(w + W)) + J_1(t - \Lambda(N))$$

and

$$C_j^{(N)}(t) = \theta_j \sum_{k=0}^{N-1} J_2(t - w - k(w + W)),$$

$j = 2, \dots, n$ , where  $J_i$  is a continuous and nonnegative function that is positive in an interval of the form  $(0, \lambda_i)$ ,  $i = 1, 2$ ;  $w$  and  $W$  are nonnegative numbers whose sum is positive; and

$$\Lambda(N) > w + (N - 1)(w + W) + \lambda_2.$$

When also  $\theta_j = \delta_{j2}$ , the  $G^{(N)}$  of Corollary 3 can be interpreted as a machine which is exposed to  $N$  periodic repetitions of a sequence  $AB$  of events, followed by a presentation of  $A$  alone to test if the machine can predict  $B$  in reply on the basis of its past experience [1]. The theorem can be interpreted as saying that the machine eventually "learns" the sequence  $AB$  if it is given sufficient practice. [1] discusses several other properties of this "learning" process, and [2] will provide a detailed exposition.

#### REFERENCES

1. S. Grossberg, *Nonlinear difference-differential equations in prediction and learning theory*, Proc. Nat. Acad. Sci. 58 (1967), 1329-1335.
2. ———, *A prediction theory for some nonlinear functional-differential equations*. I, II, J. Math. Anal. Appl. (to appear).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY