GLOBAL RATIO LIMIT THEOREMS FOR SOME NONLINEAR FUNTIONAL DIFFERENTIAL EQUATIONS. I

BY
STEPHEN GROSSBERG

Reprinted from the
BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY
January, 1969, Vol. 74, No. 1
Pp. 95-99

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

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

where $X=(x_1, \cdots, x_n)$ 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 $C=(C_1, \cdots, C_n)$ 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 $y_{jk}(t) = B_{kj}(t)$ [illegible] have limits as $t \to \infty$.

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

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

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

and

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

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

$$\theta(p) = 1 \quad \text{if } p > 0, \qquad = 0 \quad \text{if } p \leqq 0.$$

In order that our theorems hold, the initial data must always be nonnegative. 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 nonnegative 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

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

where $S(0)=\{1, 2, \cdots, n\}$ and k is the least integer such that either $S(k)=\varnothing$ 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* $V=(x^{(0)}, \cdots, x^{(k-1)})$ *and* $W=(C^{(0)}, \cdots, C^{(k-1)})$ *obey a linear equation*

$$V(t) = -\alpha V(t) + \beta D V(t-\tau) + W(t) \tag{5}$$

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

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

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

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

when $S(k)=\varnothing$. *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 $V=\{v_i: i=1, 2, \cdots, n\}$ and directed edges $E=\{e_{jk}: j, k=1, 2, \cdots, n\}$, 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, $w\in[-\tau, t]$, 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, $m=1, 2, \cdots, n$. (See [1] for further details.)

Alternatively, for every $t\geqq 0$, 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" $G(\infty)=\lim_{t\to\infty} G(t)$ when these exist.

4. Outstars.

In this note, we annouce 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}, $j=2, 3, \cdots, n$, have positive weights, which equal $1/(n-1)$. This system is therefore called an *outstar* with *source* vertex v_1, *sinks* v_j, $j=2, \cdots, n$, and *border* $B=\{v_j: j=2, \cdots, n\}$.

Our main result describes a sequence $G^{(1)}, G^{(2)}, \cdots, G^{(N)}, \cdots$ 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 $\{\theta_j: j=2, \cdots, n\}$ 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

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

and

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

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

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

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

(e) let

$$\chi(w) = 0 \quad \text{if } w > 0,$$
$$= 1 \quad \text{if } w \leqq 0.$$

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

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

and

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

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 $N \geqq 1$ and $j = 2, \cdots, n$, we can state the following theorem.

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

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

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

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

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

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

(C) *for every $N \geqq 1$ and $j = 2, \cdots, n$, the functions $y_{1j}^{(N)}$, $F_j^{(N)} = y_{1j}^{(N)} - X_j^{(N)}$, and $G_j^{(N)} = X_j^{(N)} - \theta_j$ change sign at most once and not at all if $F_j^{(N)}(0)G_j^{(N)}(0) \geqq 0$. Moreover, $F_j^{(N)}(0)G_j^{(N)}(0) > 0$ implies $F_j^{(N)}(t)G_j^{(N)}(t) > 0$ for all $t \geqq 0$.*

(C) shows in particular that the functions $y_{1j}^{(N)}$ are quite insensitive to fluctuations in the functions f and g, since $y_{1j}^{(N)}$ 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 $\theta_j = \delta_{j2}$, $j = 2, \cdots, n$, then $y_{12}^{(N)}$ increases monotonically to 1 and $y_{1k}^{(N)}$ decreases monotonically to zero, $k = 3, \cdots, n$.*

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

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

$j = 2, \cdots, n$, 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