GLOBAL RATIO LIMIT THEOREMS FOR
SOME NONLINEAR FUNCTIONAL
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

\[ X(t) = AX(t) + B(X(t-t) + C(t), t \geq 0, \]

where \( X = (x_1, \ldots, x_n) \) is nonnegative, \( B(X(t)) = \|B(t)\| \) is a matrix of nonlinear functionals of \( X(w) \) evaluated at all past times \( w \in [-\tau, t] \), and \( C = (C_1, \ldots, C_n) \) 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 \( y_{jk}(t) = \frac{B_{jk}(t)}{\|B(t)\|} \) have limits as \( t \to \infty \).

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 \( n \times n \) semi-stochastic matrix \( P = \|p_{ij}\| \) (i.e., \( p_{ij} \geq 0 \) and \( \sum_{i=1}^{n} p_{im} = 0 \) or 1), let

\[ \dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{k=1}^{n} x_k(t-\tau) y_{ik}(t) + C_i(t), \]

\[ y_{jk}(t) = \frac{p_{jk} z_{jk}(t)}{\sum_{m=1}^{n} p_{jm} z_{jm}(t)}, \]

and

\[ z_{jk}(t) = [-u z_{jk}(t) + \beta x_j(t-\tau) x_k(t)] \theta(p_{jk}), \]

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

\[ \theta(p) = \begin{cases} 1 & \text{if } p > 0, \\ 0 & \text{if } p \leq 0. \end{cases} \]

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 $z_{jk}(0) > 0$ if $p_{jk} > 0$.

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 $S(r) = \{k: \sum_{s \in S(r-1)} p_{si} = 1\}$ and

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

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

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

$$V(t) = -\alpha V(t) + \beta D V(t - \tau) + W(t) \quad (5)$$

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

$$D = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
& & & \ddots & \ddots \\
0 & 0 & 0 & \cdots & 0
\end{bmatrix}$$

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

$$D = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
& & & \ddots & \ddots \\
0 & 0 & 0 & \cdots & 0
\end{bmatrix}$$

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

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

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 $V = \{v_i : i = 1, 2, \ldots, n\}$ and directed edges $E = \{e_{jk} : j, k = 1, 2, \ldots, n\}$, 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 $y_{jk}(t)$ in (2) control the size of the $\beta x_i(t - \tau)$ flow from $v_k$ along $e_{jk}$, which eventually reaches $v_i$ by cross-correlating past $\beta x_i(w - \tau)$ and $x_i(w)$ values, $w \in [-\tau, \tau]$, with an exponential weighting factor $e^{-\alpha(t-w)}$ as in $z_{jk}(t)$ in (4), and comparing this weighted cross-correlation in (3) with all other cross-correlations $z_{lm}(t)$ corresponding to any edge leading from $v_k, m = 1, 2, \ldots, n$. (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 $G(t)$, $t > t_0$. A basic problem when $C = 0$ 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 announce a result for the case

$$P = \begin{bmatrix}
0 & 1 & \cdots & 1 \\
1 & \ddots & \cdots & \cdots \\
\vdots & \ddots & \ddots & \cdots \\
1 & \cdots & \cdots & \ddots \\
0 & \cdots & \cdots & 1
\end{bmatrix}$$

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

Our main result describes a sequence $G^{(1)}, G^{(2)}, \ldots, G^{(N)}, \ldots$ 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 $\{\theta_k : k = 2, \ldots, n\}$ 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

$$\int_{-\tau}^{\infty} e^{-\alpha(t - w)} f(w) dw \geq k, \quad t \geq T_0.$$
and
\[ \int_0^t e^{-\chi(t-u)}g(u)du \geq k, \quad t \geq T_0; \]

(c) let \( U_1(N) \) and \( U(N) \) be any positive and monotone increasing functions of \( N \geq 1 \) such that
\[ \lim_{N \to \infty} U_1(N) = \lim_{N \to \infty} U(N) = \infty; \]

(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
\[ \chi(w) = 0 \text{ if } w > 0, \]
\[ = 1 \text{ if } w \leq 0. \]

The input functions \( C^{(N)}_{i,N} \) of \( G^{(N)} \) are defined in terms of (a)-(c) by
\[ C^{(N)}_1(t) = f(t)\chi(t - U_1(N)) + h_N(t) \]
and
\[ C^{(N)}_j(t) = \theta_j g(t) \chi(t - U(N)), \quad j = 2, \ldots, n. \]

Letting the functions of \( G^{(N)} \) be denoted by superscripts \( "(N)" \) (e.g., \( y_{ij} \) is written \( y^{(N)}_{ij} \)), and defining the ratios \( X^{(N)}_{ij} = x^{(N)}_{ij} / \sum_{k=1}^n x^{(N)}_{ik} \)
for every \( N \geq 1 \) and \( j = 2, \ldots, n \), we can state the following theorem.

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

(A) for every \( N \geq 1 \), the limits \( \lim_{t \to \infty} X^{(N)}_{ij}(t) \) and \( \lim_{t \to \infty} y^{(N)}_{ij}(t) \) exist and are equal, \( j = 2, \ldots, n; \)

(B) for every \( N \geq 1 \) and all \( t \geq U(N) \), \( X^{(N)}_{ij}(t) \) and \( y^{(N)}_{ij}(t) \) are monotonic in opposite senses, and
\[ \lim_{N \to \infty} X^{(N)}_{ij}(U(N)) = \lim_{N \to \infty} y^{(N)}_{ij}(U(N)) = \theta_j, \]
\( j = 2, \ldots, n. \) In particular, by (A) and (B),
\[ \lim_{N \to \infty} X^{(N)}_{ij}(t) = \lim_{N \to \infty} y^{(N)}_{ij}(t) = \theta_j, \quad j = 2, \ldots, n. \]

(C) for every \( N \geq 1 \) and \( j = 2, \ldots, n \), the functions \( h^{(N)}_{ij}, P^{(N)}_{ij} = y^{(N)}_{ij} - X^{(N)}_{ij}, \) and \( G^{(N)}_{ij} = X^{(N)}_{ij} - \theta_j \) change sign at most once and not at

all if \( F^{(N)}_j(0)G^{(N)}_j(0) \geq 0 \). Moreover, \( F^{(N)}_j(0)G^{(N)}_j(0) > 0 \) implies \( F^{(N)}_j(t)G^{(N)}_j(t) > 0 \) for all \( t \geq 0 \).

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

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

**Corollary 2.** If \( X^{(N)}_1(0) = y^{(N)}_2(0) \) and \( \theta_j = \delta_j, j = 2, \ldots, n \), then \( y^{(N)}_{ij} \) increases monotonically to 1 and \( y^{(N)}_i \) decreases monotonically to zero, \( k = 3, \ldots, n. \)

**Corollary 3.** The theorem is true if
\[ C^{(N)}_1(t) = \sum_{k=0}^{N-1} J_1(t - k(w + W)) + J_1(t - \lambda(N)) \]
and
\[ C^{(N)}_j(t) = \theta_j \sum_{k=0}^{N-1} J_2(t - w - k(w + W)), \]
\( j = 2, \ldots, n, \) where \( J_1, J_2 \) 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
\[ \lambda(N) > w + (N - 1)(w + W) + \lambda_i. \]

When also \( \theta_j = \delta_j \), 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**


**Massachusetts Institute of Technology**