Report 99-061

Taylor Series Expansions for the Lyapunov Exponent of Stochastic (Max,+)–linear Systems

Bernd Heidergott

ISSN: 1389–2355

Taylor Series Expansions for the Lyapunov Exponent of Stochastic (Max,+)–linear Systems

Bernd Heidergott*
EURANDOM
P.O.Box 513, 5600 MB Eindhoven,
E–Mail: heidergott@eurandom.tue.nl

Abstract

We develop the Lyapunov exponent of (max,+)–linear systems into a Taylor series. To this end, we extend the theory of weak differentiation of matrices in the (max,+) semiring to higher–order weak differentiation. This leads to conditions for the analyticity of matrices in the (max,+) semiring. Elaborating on the ergodic theory for (max,+)–linear stochastic systems, we establish conditions for the analyticity of the Lyapunov exponent of (max,+)–linear systems. Moreover, we derive lower bounds for the radius of convergence of the Taylor series. The two main ingredients are: (1) the radius of convergence of the Taylor series of matrices in the (max,+)–semiring, and (2) the coupling time of the system, that is, the time it takes an arbitrarily started trajectory of the system to couple with a stationary version. We illustrate our results by applying it to a simple sample system and thereby improving the results on the domain of convergence of the Taylor series of the Lyapunov exponent for this particular system known in the literature so far.

1 Introduction

In this paper we study Taylor series expansions of the Lyapunov exponent of (max,+)–linear stochastic systems, the class of systems which can be described by a certain class of Petri nets, called stochastic event graphs. More specifically, we consider (max,+)–linear stochastic systems depending on a parameter, say θ. The Lyapunov exponent is then developed into a Taylor series with respect to θ. For example, θ may be a parameter of one of the firing time distributions of the event graph. In a queueing application, this refers to θ being e.g. the mean service time at one of the queues. However, more general dependencies of the system dynamic on θ may be modelled as well. For example, Baccelli and Hong give in [2] an example from computer science: they model a window flow control mechanism and let θ be the probability that the window flow operates with nominal window size and $1-\theta$ the probability that a reduced window size is used.

We apply the technique of weak differentiation, a technique first introduced by Pflug for gradient estimation for Markov chains, see [11] and the references therein. In [6], weak differentiation of random matrices in the (max,+)–semiring has been introduced, and analytic expansions of n–fold products in the (max,+) semiring have been given in [7]. In our paper, we extend these results to finite horizon products, that is, we consider the case when n is a stopping time. This extension allows us to develop the Lyapunov exponent of (max,+)–linear systems into a Taylor series. This approach has the following benefits:

- The Taylor series can be developed at any point of analyticity, which is in contrast to the results known so far, where only Maclaurin series have been studied.

*This research is supported by Deutsche Forschungsgemeinschaft under grant He3139/1-1. Part of this work was done while the author was with the Faculty of Information Technology and Systems, Delft University of Technology, the Netherlands, where he was supported by the EC–TMR project ALAPEDES under grant ERBFMRXCT960074.

- The radius of convergence of the Taylor series of the Lyapunov exponent is deduced from more elementary properties of the system, which allows us to predict the radius of convergence in a very simple manner.

We illustrate our approach with a simple system following a Bernoulli scheme, like the window flow example mentioned above. More specifically, we show that the Lyapunov exponent of the Bernoulli scheme can to any point $\theta \in [0,1)$ be developed into a Taylor series that converges at least on $[\theta - (2c)^{-1}, \theta + (2c)^{-1}] \cap [0,1)$, where $c \geq 1$ denotes the coupling times of one of the matrices of the Bernoulli scheme (for a definition of the coupling time see Section 3.1 or [1]). This implies that the Lyapunov exponent can be extended to a complex function that is analytical on a strip around the real interval $[0,1)$ with width $(2c)^{-1}$, which extends the results in [2] and [3].

The paper is organised as follows. In the next section, we introduce the (max,+)–semiring. In Section 3 we show that the Lyapunov exponent can be represented by the difference between two finite horizon experiments. Section 4 establishes conditions that imply the analyticity of finite products in the (max,+)–semiring. In Section 5, we extend the results of the previous section to random horizon experiments. Then we combine these results with the finite horizon representation of the Lyapunov exponent as established in Section 3, which subsequently will provide the desired Taylor series of the Lyapunov exponent. We conclude the section with an illustrating example.

2 (Max,+)–linear Stochastic Systems

In this section we introduce the (max,+)-semiring. This structure was first introduced in [5]. For an extensive discussion of the (max,+)–algebra and similar structures we refer to [1].

2.1 The (Max,+)–Semiring

Let $\epsilon = -\infty$ and denote by $\mathbb{R}_\epsilon$ the set $\mathbb{R} \cup \{\epsilon\}$. For elements $a, b \in \mathbb{R}_\epsilon$ we define the operations $\oplus$ and $\otimes$ by

$$a \oplus b = \max(a,b) \quad \text{and} \quad a \otimes b = a + b\,,$$

where we adopt the convention that for all $a \in \mathbb{R}$ $\max(a, -\infty) = \max(-\infty, a) = a$ and $a + (-\infty) = -\infty + a = -\infty$. The set $\mathbb{R}_\epsilon$ together with the operations $\oplus$ and $\otimes$ is called the *(max,+)–semiring*[1] and is denoted by $\mathbb{R}_{\max}$. In particular, ϵ is the neutral element for the operation $\oplus$ and absorbing for $\otimes$, that is, for all $a \in \mathbb{R}_\epsilon$ $a \otimes \epsilon = \epsilon$. The neutral element for $\otimes$ is $e := 0$. Moreover, $\mathbb{R}_\epsilon$ is *idempotent*, that is, for all $a \in \mathbb{R}_\epsilon$ $a \oplus a = a$. Idempotent semirings are called *dioids* in [1]. The structure $\mathbb{R}_{\max}$ is richer than that of a semiring since $\otimes$ is commutative and has an inverse. However, in what follows we will work with matrices over $\mathbb{R}_\epsilon$ and thereby lose, like in conventional algebra, commutativity and general invertability of the product.

We extend the (max,+)–semiring to matrices in the following way. For $A, B \in \mathbb{R}_\epsilon^{J\times J}$, we define $A \oplus B$ as follows

$$(A \oplus B)_{ij} = A_{ij} \oplus B_{ij} \quad , 1 \leq i,j \leq J\,.$$

For $A \in \mathbb{R}_\epsilon^{I\times J}$ and $B \in \mathbb{R}_\epsilon^{J\times K}$, we define $A \otimes B$ by

$$(A \otimes B)_{ik} = \bigoplus_{j=1}^{J} A_{ij} \otimes B_{jk} \quad , 1 \leq i \leq I\,,\ 1 \leq k \leq K\,. \tag{1}$$

The matrix $\mathcal{E}$ with all elements equal to ϵ is the zero element of the $\oplus$ matrix operation. On $\mathbb{R}_\epsilon^{J\times J}$, the matrix E with diagonal elements equal to e and ϵ elsewhere is the neutral element of the $\otimes$ matrix operation. We denote the $J \times J$–dimensional matrices over $\mathbb{R}_\epsilon$ equipped with the operations $\oplus$ and $\otimes$ defined as above by $\mathbb{R}_{\max}^{J\times J} = (\mathbb{R}_\epsilon^{J\times J}, \oplus, \otimes, \mathcal{E}, E)$. Observe that $\mathbb{R}_{\max}^{J\times J}$

[1] A *semiring* is a set R endowed with two binary operations, $\oplus$ and $\otimes$, so that $\oplus$ is associative and commutative with zero–element ϵ, $\otimes$ is associative and has zero–element e, $\otimes$ distributes over $\oplus$ and ϵ is absorbing for $\otimes$.

is again a semiring. To simplify notation, we write $\mathbb{R}_\epsilon^J$ for $\mathbb{R}_\epsilon^{J\times 1}$, that is, $\mathbb{R}_\epsilon^J$ denotes the set of J–dimensional vectors over $\mathbb{R}_\epsilon$.

Some of the statements to be presented below hold for general matrices in $\mathbb{R}_{\max}^{J\times J}$, whereas others only hold for a restricted class of matrices. For example, there are statements that are only true for matrices with at least one entry in each row different from ϵ. Furthermore, sometimes we do assume that (initial) vectors are different from ϵ. However, these assumptions impose no restriction on the class of systems that can be treated.

Let some probability space (Ω, F, P) be given on which all random variables introduced below are defined. We say that a random matrix has *fixed support* if the probability that $A_{ij} = \epsilon$ is either zero or one. We call A *integrable* if A has fixed support and if $E[A_{ij}] < \infty$ for all non–ϵ entries of A.

A (random) matrix $A \in \mathbb{R}_\epsilon^{J\times J}$ is said to be *irreducible* if A has fixed support and if for all i, j a sequence $i = i_0, \ldots, i_m = j$ exists such that $A_{i_l\, i_{l+1}} > \epsilon$ for $0 \le l < m$.

2.2 Examples of (Max,+)–linear Stochastic Networks

In the following we give examples of (max,+)–linear queueing networks. For a necessary and sufficient condition for the (max,+)–linearity of a queueing network, we refer to [8].

Example 1 *Consider a closed system of J single–server queues in tandem, with infinite buffers. In the system, customers have to pass through the queues consecutively so as to receive service at each server. After service completion at the J^{th} server, the customers return to the first queue for a new cycle of service.*

For the sake of simplicity we assume that there are J customers circulating through the network and that initially there is one customer at each queue. Let $\sigma_j(k)$ denote the k^{th} service time at queue j and let $x_j(k)$ be the time of the k^{th} service completion at node j, so that the time evolution of the system can be described by a J–dimensional vector $x(k) = (x_1(k), \ldots, x_J(k))$ following

$$x(k+1) = A(k) \otimes x(k) \,,$$

where the matrix $A(k)$ looks like

$$A(k-1) = \begin{bmatrix} \sigma_1(k) & \epsilon & \ldots & \epsilon & \sigma_1(k) \\ \sigma_2(k) & \sigma_2(k) & \epsilon & \ldots & \\ & & & & \\ & \ldots & & & \\ & & & & \\ & \ldots & \sigma_{J-1}(k) & \sigma_{J-1}(k) & \epsilon \\ & \ldots & \epsilon & \sigma_J(k) & \sigma_J(k) \end{bmatrix} \tag{2}$$

for $k \ge 1$. For more examples of this kind we refer to [9].

Suppose that one of the service time distributions depends on a parameter, say, θ. For example, θ may be the mean of one of the service times. In this case, the (max,+)–linear recursion describing the system dynamics depends on θ through these service times. The following example is of a different kind: here the distribution of the transition matrix as a whole depends on θ.

Example 2 (Baccelli & Hong, [2]) *Consider a cyclic tandem queueing network consisting of a single–server and a multi–server, each with deterministic service times. Service times of the single–server station equal σ, whereas service times at the multi–server station equal σ'. Two customers circulate in the network. The time evolution of this network is described by a (max,+)–linear sequence $x(k) = (x_1(k), \ldots, x_4(k))$, where $x_1(k)$ is the k^{th} begin of service at the single–server station and $x_2(k)$ its k^{th} departure epoch; $x_3(k)$ is the k^{th} begin of service at the multi–server station and $x_4(k)$ its k^{th} departure epoch. The system then follows*

$$x(k+1) = D_1 \otimes x(k) \,,$$

where

$$D_1 = \begin{bmatrix} \sigma & \epsilon & \sigma' & \epsilon \\ \sigma & \epsilon & \epsilon & \epsilon \\ \epsilon & e & \epsilon & e \\ \epsilon & \epsilon & \sigma' & \epsilon \end{bmatrix}$$

Consider the cyclic tandem network again, but with one of the servers of the multi-server broken down. This system follows

$$x(k+1) = D_2 \otimes x(k) \, ,$$

where

$$D_2 = \begin{bmatrix} \sigma & \epsilon & \sigma' & \epsilon \\ \sigma & \epsilon & \epsilon & \epsilon \\ \epsilon & e & \sigma' & \epsilon \\ \epsilon & \epsilon & \sigma' & \epsilon \end{bmatrix}$$

Assume that such a breakdown occurs after service completion with probability $1-\theta$. *Let* $A_\theta(k)$ *have distribution*

$$P(\, A_\theta(k) = D_1 \,) = \theta$$

and

$$P(\, A_\theta(k) = D_2 \,) = 1 - \theta \, ,$$

then

$$x_\theta(k+1) = A_\theta(k) \otimes x_\theta(k)$$

describes the time evolution of the system with random breakdowns.

2.3 The Space $\hat{\mathbb{R}}_\epsilon^{I \times J}$

The aim of our analysis is to study $(\max, +)$–linear stochastic systems. More specifically, we are interested in $(\max, +)$–linear models of stochastic networks such as queueing systems. These models have in common that the entries of the corresponding transition matrices are either non–negative or equal to ϵ. Therefore, it suffices for our purpose to restrict our analysis to the semiring

$$\hat{\mathbb{R}}_{\max} = (\hat{\mathbb{R}}_\epsilon = [0, \infty) \cup \{-\infty\}, \oplus = \max, \otimes = +, \epsilon = -\infty, e = 0) \, .$$

Moreover, $(\hat{\mathbb{R}}_\epsilon, \oplus, \epsilon)$ is a monoid and with quasi norm[2]

$$||x|| := ||x||_{\hat{\mathbb{R}}_\epsilon} = \max\left(\frac{1}{x+1}, x+1\right) \, ,$$

that is, (a) $||x|| \geq 0$ for all x, (b) $||x|| = 0$ if and only if $x = \epsilon$, and (c) $||x \oplus y|| \leq ||x|| + ||y||$ for all x, y. The quasi norm $||x||$ is extended to $\hat{\mathbb{R}}_\epsilon^{J \times I}$ by

$$||A|| := ||A||_{\hat{\mathbb{R}}_\epsilon^{J \times I}} = \max\Big(||A_{ji}|| \, : \, 1 \leq j \leq J \, , \, 1 \leq i \leq I \Big) \, ,$$

for $A \in \hat{\mathbb{R}}_\epsilon^{J \times I}$. Observe that for every integrable A, $E[||A||] < \infty$.

On $\hat{\mathbb{R}}_\epsilon$, we introduce a metric $d(\cdot, \cdot)$ as follows. For $x, y \in \mathbb{R}$, we set $d(x, y) = |x - y|$, $d(x, \epsilon) = \infty = d(\epsilon, x)$, and $d(\epsilon, \epsilon) = 0$. This metric is extended to $\hat{\mathbb{R}}_\epsilon^{J \times I}$ by

$$d(A, B) := d_{\hat{\mathbb{R}}_\epsilon^{J \times I}}(A, B) = \max\Big(d(A_{ji}, B_{ji}) \, : \, 1 \leq j \leq J \, , \, 1 \leq i \leq I \Big) \, ,$$

[2] It is possible to equip $\hat{\mathbb{R}}_{\max}$ with the metric $d(x, y) = e^{\max(x,y)} - e^{\min(x,y)}$. Hence, $||x|| := d(x, \epsilon)$ would be a natural choice for $||\cdot||$. But later in the text we will study functions of the type $g : \hat{R}_{\max} \to \mathbb{R}$ such that $|g(x)| \leq c_1 + c_2 ||x||^k$ for constants c_1, c_2 and k. A key condition for analysis will be that $E[||x||^k] < \infty$. Hence, taking $||x|| = e^x$ imposes a severe restriction, which is the reason why we work with $||x|| = \max(1/(x+1), x+1)$.

for $A, B \in \hat{\mathbb{R}}_\epsilon^{J\times I}$.

The space $(\hat{\mathbb{R}}_\epsilon^{J\times I}, d)$ is a separable metric space. In addition to that, each continuous mapping $g : [0,\infty)^{J\times I} \to \mathbb{R}$ can be extended to a continuous mapping $\tilde{g} : \hat{\mathbb{R}}_\epsilon^{J\times I} \to \mathbb{R}$ through $\tilde{g}(x) = g(x)$ for all $x \in [0,\infty)^{J\times I}$ and otherwise zero, see Lemma 2 in [7].

One of the corner stones of our analysis is to establish a kind of generalised non–expansiveness of the $\otimes$–product. More precisely, let the mapping $D(\cdot,\cdot)$ be such that

$$\forall x, y \in \hat{\mathbb{R}}_\epsilon^{J}\ \forall A, B \in \hat{\mathbb{R}}_\epsilon^{I\times J} \;:\quad D(A\otimes x, B\otimes y) \leq D(A,B) \,+\, D(x,y)\,. \tag{3}$$

Then we call a mapping $g : \mathbb{R}_\epsilon^{I\times J} \to \mathbb{R}$ *non–expansive* with respect to $D(\cdot,\cdot)$ if for all $A, B \in \mathbb{R}_\epsilon^{I\times J}$

$$|\,g(A) - g(B)\,| \,\leq\, D(A,B)\,.$$

Put another way, the non–expansiveness of a mapping g allows us to switch from distances between performance measures of trajectories of (max,+)–linear systems to distances between matrices and vectors, respectively, in $\mathbb{R}_\epsilon^{I\times J}$, for which we can establish upper bounds by applying Inequality (3). Of course, the above introduced metric $d(\cdot,\cdot)$ is the first candidate for $D(\cdot,\cdot)$ in (3). Unfortunately, Inequality (3) does not hold for $d(\cdot,\cdot)$, which will become clear in the proof of the following lemma.

In what follows we introduce a different way of measuring distances in $\hat{\mathbb{R}}_\epsilon^{J\times I}$. To this end, we define for $x, y \in \hat{\mathbb{R}}_\epsilon$

$$\delta(x,y) = |\,||x|| - ||y||\,|$$

and for $A, B \in \hat{\mathbb{R}}_\epsilon^{J\times I}$

$$\delta(A,B) \,:=\, \delta_{\hat{\mathbb{R}}_\epsilon^{J\times I}}(A,B) \,=\, \max\Big(\,\delta(A_{j\,i}, B_{k\,l}\,) \,:\, \forall (i,j),(k,l) \in J\times I\,\Big)\,.$$

We obtain $\delta(A,\mathcal{E}) = |\,||A|| - 1|$ and $\delta(A,E) = ||A||$ if $A \neq \mathcal{E}$ and otherwise zero.

Lemma 1 *Let $A, B \in \hat{\mathbb{R}}_\epsilon^{I\times J}$ be such that each row of A and B contains at least one element different from ϵ, then it holds for $x, y \in \mathbb{R}^J$ that*

$$\delta(A\otimes x, B\otimes y) \,\leq\, \delta(A,B) \,+\, \delta(x,y)\,.$$

Proof: Let $j^A(i)$ be such that

$$A_{i\,j^{A}(i)} \otimes x_{j^{A}(i)} \,=\, \bigoplus_{j=1}^{J} A_{i\,j} \otimes x_j \;\in \mathbb{R}$$

and $j^B(i)$ such that

$$B_{i\,j^{B}(i)} \otimes y_{j^{B}(i)} \,=\, \bigoplus_{j=1}^{J} B_{i\,j} \otimes y_j \;\in \mathbb{R}\,.$$

Straightforward calculation yields

$$\begin{aligned}
&\delta(A\otimes x, B\otimes y)\\
&=\max_{i\,l}\{\,|\,||A_{i\,j^{A}(i)} \otimes x_{j^{A}(i)}|| - ||B_{l\,j^{B}(l)} \otimes y_{j^{B}(l)}||\,|\,\}\\
&=\max_{i\,l}\{\,|\,A_{i\,j^{A}(i)} \otimes x_{j^{A}(i)} - B_{l\,j^{B}(l)} \otimes y_{j^{B}(l)}\,|\,\}\\
&\leq\max_{i\,l}\{\,|\,A_{i\,j^{A}(i)} - B_{l\,j^{B}(l)}\,|\,\} + \max_{i\,l}\{\,|\,x_{j^{A}(i)} - y_{j^{B}(l)}\,|\,\}\\
&\leq\delta(A,B) + \delta(x,y)\,,
\end{aligned}$$

which concludes the proof of the first part of the lemma. Note that the last inequality fails for $d(\cdot,\cdot)$. $\square$

In what follows we call $g : \overline{\mathbb{R}}_\epsilon^{I\times J} \to \mathbb{R}$ *non-expansive* if g is non-expansive with respect to $\delta(\cdot,\cdot)$, in formula:

$$\forall A, B \in \overline{\mathbb{R}}_\epsilon^{I\times J} : \quad |g(A) - g(B)| \leq \delta(A, B) . \tag{4}$$

Examples of non-expansive maps are the coordinate-wise projection onto $\mathbb{R}$, that is, for fixed i, j we set $g(A) = A_{ij}$ if $A_{ij} \in \mathbb{R}$ and otherwise zero, and the maximum operator, that is, $g(A) = \max(A_{ij})$.

We conclude this section with the remark that $\delta(\cdot,\cdot)$ is a metric on $\overline{\mathbb{R}}_\epsilon^{I\times J}$ but that the topology induced by $\delta(\cdot,\cdot)$ is too weak for any practical purpose, that is, if we extend a continuous mapping $g : [0,\infty)^{I\times J} \to \mathbb{R}$ to a mapping $\tilde{g} : \overline{\mathbb{R}}_\epsilon^{I\times J} \to \mathbb{R}$, usually $\tilde{g}$ is not continuous with respect to $\delta(\cdot,\cdot)$, which is in contrast to the metric $d(\cdot,\cdot)$ as mentioned earlier.

3 Ergodic Theory

Ergodic theory for (max,+)-linear studies the asymptotic behaviour of the sequence

$$x(k+1) = A(k) \otimes x(k) , \quad k \geq 0 ,$$

with $x(0) = x_0$. One distinguishes between two types of asymptotic results:
(i) *first-order* limits

$$\lim_{k\to\infty} \frac{x(k)}{k} ,$$

(ii) *second-order* limits

$$\lim_{k\to\infty} \Big(x_i(k) - x_j(k) \Big) \quad \text{and} \quad \lim_{k\to\infty} \Big(x_j(k+1) - x_j(k) \Big) \quad .$$

Note that, in contrast to first-order limits, second-order limits are random variables and one has to consider carefully in which sense these limits are justified.

In this paper, we study type second-order limits of the type $\lim_{k\to\infty}(x_j(k+1) - x_j(k))$. As we will explain below, many interesting performance characteristics can be described through this type of difference.

A first-order limit is an inverse throughput. For example, the throughput of station j in Example 1 can be obtained from

$$\lim_{k\to\infty} \frac{k}{x_j(k)} .$$

Second-order limits are related to waiting times and cycle times. Consider the closed tandem network in Example 1. There are J customers circulating through the system. Thus, the k^{th} and the $(k+J)^{th}$ departure from queue j refer to the same (physical) customer and the cycle time of this customer equals

$$x_j(k+J) - x_j(k) .$$

Hence, the existence of the second-order limit $x_j(k+1) - x_j(k)$ implies limit results on cycle times of customers. For more details on the modelling of performance characteristics of queueing systems via first-order and second-order expressions we refer to [9] and [10].

3.1 First-Order Limits

We now state the celebrated cyclicity theorem for deterministic matrices, which is of key importance for our analysis.

Theorem 1 (Baccelli et al. [1]) *For each irreducible matrix $A \in \mathbb{R}_\epsilon^{J\times J}$, uniquely defined integers $c(A)$, $\sigma(A)$ and a uniquely defined real number $\lambda = \lambda(A)$ exist such that for all $n \geq c(A)$*

$$A^{n+\sigma(A)} = \lambda^{\sigma(A)} \otimes A^n ,$$

which implies that for all initial vectors the sequence $x(k+1) = A \otimes x(k)$, $k \geq 0$, satisfies

$$\lim_{k\to\infty} \frac{x(k)}{k} = \lambda .$$

The number $c(A)$ is called the coupling time of A, $\sigma(A)$ is called the cyclicity of A and λ is the unique eigenvalue of A.

Let $V(A)$ be the eigenspace of A, it holds for all $n \geq c(A)$ and all $x \in \mathbb{R}_\epsilon^{J\times J}$

$$A^n \otimes x \in V(A) .$$

A sufficient condition for A to be of cyclicity one is that the critical graph of A has a single strongly connected subgraph with cyclicity one (see [1] for the definition of the critical graph and for that of its cyclicity). This property will be referred to as *scs1–cyc1* below.

Theorem 2 (Theorem 7.27 in [1]) *Assume that $\{A(k)\}$ is a stationary and ergodic sequence of random matrices in $\mathbb{R}_\epsilon^{J\times J}$, and that $A(0)$ is irreducible and integrable. Then for the sequence $x(k+1) = A(k) \otimes x(k)$, with $x(0) = x_0$, the following limits exist with probability one and are independent of the initial vector x_0*

$$\lim_{k\to\infty} \frac{1}{k} \bigotimes_{i=0}^{k-1} A(i) \otimes x_0 = \frac{1}{k} E\left[\bigotimes_{i=0}^{k-1} A(i) \otimes x_0\right] = \lambda .$$

The constant λ is referred to as the *(max, +)–Lyapunov exponent* of the sequence of random matrices $\{A(k)\}$. There is no ambiguity in denoting the Lyapunov exponent of $\{A(k)\}$ and the eigenvalue of a matrix A by the same symbol, since for $A(k) = A$, for all k, the Lyapunov exponent of $\{A(k)\}$ is just the eigenvalue of A.

Consider for example the system in Example 1. If we assume that the service times $\sigma_j(k)$ are i.i.d. with finite mean for each j and that the sequences $\{\sigma_j(k)\}$ $(1 \leq j \leq J)$ are mutually independent, then Theorem 2 applies. This is in contrast to the situation of Example 2. Here, $A(k)$ has no fixed support (and is therefore not integrable) and the theorem does not apply. In order to obtain an ergodic theorem for sequences with no fixed support, we restrict ourselves to sequences satisfying the following two conditions:

(C1) The sequence $\{A(k)\}$ is i.i.d. with a countable state space $\mathcal{A}$.

(C2) Each $A \in \mathcal{A}$ has at least one entry different from ϵ on each line.

We have the following

Theorem 3 *Under assumptions* **(C1)** *and* **(C2)**, *if $\mathcal{A}$ contains at least one irreducible scs1–cyc1 matrix, then the following limit exists with probability one for all initial vectors $x_0 \in \mathbb{R}^J$ and is independent of x_0*

$$\lim_{k\to\infty} \frac{1}{k} \bigotimes_{i=0}^{k-1} A(i) \otimes x_0 = \lambda .$$

Proof: We sketch the proof. Let c denote the coupling time of A, with $A \in \mathcal{A}$ a scs1–cyc1 matrix. With positive probability, we observe the event $\{A(i) = A : 0 \leq i \leq c-1\}$. On this event $x(c) \in V(A)$, see Theorem 1. Since $V(A)$ is a single point in the projective space and, therefore, compact, Theorem 2.10 in [4] applies and we obtain that for all j with probability one

$$\lim_{k\to\infty} \frac{1}{k} x_j(k) = \lambda ,$$

which proves the theorem. □

The above theorem applies to the system in Example 2, and it applies to the system in Example 1 if we assume that the service times $\sigma_j(k)$ are i.i.d. for each j and have discrete support.

3.2 Second–Order Limits

Mairesse introduced in [10] the concept of a "pattern" of a sequence of matrices in order to study second–order limits. In the following we combine his results with results on first–order limits in order to represent second-order limits by finite horizon experiments. We follow the line of argument of Baccelli and Hong in [2].

Let $\mathcal{A} = \{a(l)\}$ be a finite or countable collection of $J \times J$–dimensional irreducible matrices. We think of $\mathcal{A}$ as the state space of the random sequence $\{A(k)\}$ following a discrete law. We say that $\{A(k)\}$ admits a *pattern* if a matrix $\tilde{A}$ and a finite number N exist such that (1) $\tilde{A} = a_N \otimes a_{N-1} \otimes \ldots \otimes a_1$ and $P(A(k+n) = a_n : 1 \leq n \leq N) > 0$ for all k, and (2) $\tilde{A}$ is an irreducible scs1–cyc1 matrix. We call $\tilde{A}$ the matrix associated with the pattern of $\{A(k)\}$. For easy reference, we introduce the following condition.

(C3) The sequence $\{A(k)\}$ has a pattern with associated matrix $\tilde{A}$ such that $\tilde{A}$ is irreducible and scs1–cyc1. An eigenvector of $\tilde{A}$ will be denoted by X_0.

The fact that $\{A(k)\}$ admits a pattern resembles a sort of memoryless property of (max,+)–linear systems. To see this, let $x(k+1) = A(k) \otimes x(k)$ be a stochastic sequence defined via $\{A(k)\}$ and assume that $\{A(k)\}$ has a pattern with associated matrix $\tilde{A}$. For vectors $x, y \in \mathbb{R}^J_\epsilon$, let $x - y$ denote the component-wise difference, that is, $(x-y)_j = x_j - y_j$, where we adopt the convention that $\epsilon - x = \epsilon$ and $x - \epsilon = \infty$ for $x \neq \epsilon$, and $\epsilon - \epsilon = 0$. In what follows we consider the limit of $x(k+1) - x(k)$ for k towards ∞, where the limit has to be understood component–wise. In order to prove the existence of this limit we will work with a backward coupling argument. For this reason it is more convenient to let the index k run backwards. More precisely, we set

$$A^0_{-m} = \bigotimes_{k=-m}^{0} A(k)$$

and

$$x^0_{-m} = A^0_{-m} \otimes x_0 = \bigotimes_{k=-m}^{0} A(k) \otimes x_0 \,,$$

with $x^0_0 = x_0$, that is, x^0_{-m} is the state of the sequence $\{x(k)\}$, started at time $-m$ in x_0, at time 0. The sequence $\{x^0_{-m} : m \geq 0\}$ evolves backwards in time according to

$$x^0_{-(m+1)} = A^0_{-m} \otimes A(-(m+1)) \otimes x_0 \,.$$

Note that $x(k)$ and x^0_{-k} are equal in distribution. With this notation the second–order limit reads

$$\lim_{k\to\infty} A(1) \otimes x^0_{-k} - x^0_{-k} = A(1) \otimes \bigotimes_{k=-\infty}^{0} A(k) \otimes x_0 - \bigotimes_{k=-\infty}^{0} A(k) \otimes x_0 \,.$$

Suppose that, going backwards in time, after η steps we observe for the first time $c(\tilde{A})$ times the pattern of $\{A(k)\}$ in a row. More precisely, let $\tilde{a}$ denote the $c(\tilde{A})$–fold concatenation of the string $(a_N, a_{N-1}, \ldots, a_1)$, which implies that $\tilde{a}$ has $M = c(\tilde{A}) \cdot N$ components. Then η is defined by

$$\eta = \inf\{k \geq 0 \,|\, A(-k) = \tilde{a}_1, A(-k+1) = \tilde{a}_2, \ldots, A(-k+(M-1)) = \tilde{a}_M\} \,,$$

and we obtain in accordance with Theorem 1 that, independent of the sequence $\{x^0_{-k} : k > \eta\}$, the random variable $x^0_{-\eta}$ is an eigenvector of $\tilde{A}$, in formula: $x^0_{-\eta} \in V(\tilde{A})$.

For $v \in \mathbb{R}_\epsilon^J$, we define multiplication by a scalar $\gamma \in \mathbb{R}_\epsilon$ by component–wise multiplication: $(\gamma \otimes u)_j = \gamma \otimes u_j$. It can be easily checked that

$$\forall \gamma \in \mathbb{R}_\epsilon\,,\; v \in \mathbb{R}_\epsilon^J \;:\quad B \otimes v - C \otimes v \;=\; B \otimes (\gamma \otimes v) - C \otimes (\gamma \otimes v)\,, \tag{5}$$

for all $B, C \in \hat{\mathbb{R}}_\epsilon^{J \times J}$. We now use the fact that eigenvectors of a scs1–cyc1 matrix are equal up to scalar multiplication: if $u, v \in V(A)$, then a $\gamma \in \mathbb{R}_\epsilon$ exists such that $v = \gamma \otimes u$ (see Theorem 3.101 in [1]). Hence, (5) implies

$$\forall v, u \in V(A) \;:\quad B \otimes v - C \otimes v = B \otimes u - C \otimes u\,, \tag{6}$$

for matrices $A, B, C \in \mathbb{R}_\epsilon^{J \times J}$. Combining the above arguments, we obtain

$$\begin{aligned}
&\lim_{k \to \infty} A(1) \otimes x^0_{-k} - x^0_{-k} \\
&= A(1) \otimes \bigotimes_{k=-\infty}^{0} A(k) \otimes x_0 \;-\; \bigotimes_{k=-\infty}^{0} A(k) \otimes x_0 \\
&= A(1) \otimes \bigotimes_{k=-\eta+M-1}^{0} A(k) \otimes \underbrace{\tilde{A}^{c(\tilde{A})} \otimes \bigotimes_{k=-\infty}^{-\eta-1} A(k) \otimes x_0}_{=: X_0 \in V(\tilde{A})} \\
&\qquad - \bigotimes_{k=-\eta+M-1}^{0} A(k) \otimes \underbrace{\tilde{A}^{c(\tilde{A})} \otimes \bigotimes_{k=-\infty}^{-\eta-1} A(k) \otimes x_0}_{=: X_0 \in V(\tilde{A})} \\
&\overset{(6)}{=} A(1) \otimes \bigotimes_{k=-\eta}^{0} A(k) \otimes x_0 \;-\; \bigotimes_{k=-\eta}^{0} A(k) \otimes x_0 \\
&= A(1) \otimes A^0_{-\eta} \otimes x_0 \;-\; A^0_{-\eta} \otimes x_0 \;<\infty\,,
\end{aligned}$$

Hence, the second–order limit can be represented by a (random) finite horizon experiment.

Next we will show that the above limit representation also holds if we consider expectations. To this end, we assume that the entries of $A(k)$ are either positive or ϵ, that is, we assume $A(k) \in \hat{\mathbb{R}}_\epsilon^{J \times J}$. Furthermore, we assume that

(C4) For all k, each row of $A(k)$ has at least one element different from zero and $x_0 \in [0, \infty)^J$.

Condition **(C4)** implies that $x(k) \in \mathbb{R}^J$ for all k. Let $(\cdot)_j$ denote the projection on the j^{th} component and recall that $(\cdot)_j$ is non–expansive. This makes it possible to apply Lemma 1 and we obtain for all $m \in \mathbb{N}$ and all j

$$\begin{aligned}
&\left| \left(A(1) \otimes \bigotimes_{k=-m}^{0} A(k) \otimes x_0 \right)_j - \left(E \otimes \bigotimes_{k=-m}^{0} A(k) \otimes x_0 \right)_j \right| \\
&\overset{(4)}{\leq} \delta \left(A(1) \otimes \bigotimes_{k=-m}^{0} A(k) \otimes x_0 \,,\, E \otimes \bigotimes_{k=-m}^{0} A(k) \otimes x_0 \right) \\
&\overset{\text{L. 1}}{\leq} \delta(A(1), E) \\
&= ||A(1)||\,.
\end{aligned}$$

Integrability of $A(1)$ implies $E[||A(1)||] < \infty$ and applying the dominated convergence theorem to the above second-order limit, we obtain

$$
\begin{aligned}
&\lim_{k\to\infty} E[x(k+1)-x(k)] \\
&\quad = \lim_{k\to\infty} E\left[A(1)\otimes x^0_{-k} - x^0_{-k}\right] \\
&\quad = E\left[\lim_{k\to\infty} A(1)\otimes x^0_{-k} - x^0_{-k}\right] \\
&\quad = E\left[\bigotimes_{k=-\eta}^{1} A(k)\otimes x_0 \;-\; \bigotimes_{k=-\eta}^{0} A(k)\otimes x_0\right] < \infty\,. \qquad (7)
\end{aligned}
$$

The main result of this section can now be stated as follows.

Theorem 4 *Under assumptions* **(C1)** *to* **(C4)**, *an almost surely finite stopping time $\eta \in \mathbb{N}$ exists, such that for all $x_0 \in \mathbb{R}^J$*

$$
\begin{aligned}
\lambda =& \lim_{k\to\infty} E[x(k+1)-x(k)] \\
=& E\left[\bigotimes_{k=-\eta}^{1} A(k)\otimes x_0 \;-\; \bigotimes_{k=-\eta}^{0} A(k)\otimes x_0\right] < \infty\,.
\end{aligned}
$$

Proof: It remains to be shown that

$$\lambda = \lim_{k\to\infty} E[x(k+1)-x(k)]\,.$$

The limit on the right-hand side of the above formula exists, and, applying a Cesaro averaging argument, we obtain

$$\lim_{k\to\infty} E[x(k+1)-x(k)] = \lim_{k\to\infty} \frac{1}{k} E[x(k)]\,.$$

In Lemma 1 in [2] Baccelli and Hong showed that under the conditions of the theorem it holds that

$$\lim_{k\to\infty} \frac{1}{k} E[x(k)] = \lambda\,,$$

which concludes the proof of the theorem. □

Remark 1 *If we apply a Cesaro averaging argument, the second-order limit in Theorem 4 yields that, under the assumptions* **(C1)** *to* **(C4)**, *the limit in Theorem 3 also holds for the expected value of $x_j(k)/k$ for all j.*

3.3 Problem Statement

Let $\theta \in \Theta$ be a real-valued parameter, Θ being an interval. We shall take θ to be a variational parameter of the sequence $\{A_\theta(k)\}$ of square matrices in $\mathbb{R}_\epsilon^{J\times J}$ and study sequences $\{x_\theta(k)\}$ following

$$x_\theta(k+1) = A_\theta(k)\otimes x_\theta(k)\,, \quad k \ge 0\,,$$

with $x_\theta(0) = x_0$ for all θ.

The aim of this paper is to develop the second-order limit

$$\lim_{k\to\infty} E[x_\theta(k+1)-x_\theta(k)] \qquad (8)$$

into a Taylor series. This will then lead to Taylor series expansions of the Lyapunov exponent of $\{A_\theta(k)\}$. Moreover, we obtain Taylor series expansions of many performance measures of interest, like waiting times or mean queue lengths.

To avoid an inflation of subscripts, we will in what follows suppress the subscript θ when this causes no confusion.

4 Weak Derivatives of Random Matrices

We denote by $C_p(\hat{\mathbb{R}}_\epsilon^{J\times I}) := C_p(\hat{\mathbb{R}}_\epsilon^{J\times I}, d_{\hat{\mathbb{R}}_\epsilon^{J\times I}}(\cdot,\cdot))$ the set of all functions $g : \hat{\mathbb{R}}_\epsilon^{J\times I} \to \mathbb{R}$ such that $|g(x)| \leq c_1 + c_2||x||^l$ for all $x \in \hat{\mathbb{R}}_\epsilon^{J\times I}$ and all l with $0 \leq l \leq p$, for $p \geq 0$. In addition to that we assume that the set of bounded continuous mappings from $\hat{\mathbb{R}}_\epsilon^{J\times I}$ to $\mathbb{R}$ is a subset of $C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$ for all $p \geq 0$.

The set of all measures on $(\hat{\mathbb{R}}_\epsilon^{J\times I}, \mathcal{F})$ is denoted by $\mathcal{M} = \mathcal{M}(\hat{\mathbb{R}}_\epsilon^{J\times I})$, where $\mathcal{F}$ denotes the Borel field of $\hat{\mathbb{R}}_\epsilon^{J\times I}$. Moreover, the n–fold product field will be denoted by $\mathcal{F}^n$. The set of probability measures on $(\hat{\mathbb{R}}_\epsilon^{J\times I}, \mathcal{F})$ is denoted by $\mathcal{M}_1 \subset \mathcal{M}$. For $\mu, \nu \in \mathcal{M}$, we say that μ is ν–continuous, in symbols $\nu \gg \mu$, if $\nu(A) = 0$ implies $\mu(A) = 0$ for all $A \in \mathcal{F}$. The ν–continuity of μ implies that the Radon–Nikodyn derivative of μ with respect to ν exists. Put another way, if $\nu \gg \mu$, then the ν–density of μ, denoted by $f(\mu, \nu)$ exists. If $\mu \gg \mu_\theta$ for all $\theta \in \Theta$, we write $f_\theta(x) = f(\mu_\theta, \mu)(x)$. In what follows we let $d^n f_\theta / d\theta^n$ denote the n^{th} derivative of f_θ, provided that it exists, and set $f_\theta = d^0 f_\theta / d\theta^0$.

Definition 1 *Let $\nu, \mu_\theta \in \mathcal{M}_1$ be such that $\nu \gg \mu_\theta$ for all $\theta \in \Theta$. We call μ_θ n times ν–Lipschitz differentiable at θ with respect to $C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$, or n times Lipschitz differentiable for short, if*

- *$f_\theta(x) = f(\mu_\theta, \mu)(x)$ is $(n+1)$times differentiable with respect to θ on Θ for ν almost all x;*
- *for all $m \leq n+1$ mappings K_f^m exist, such that*

$$\sup_{\theta\in\Theta} \left| \frac{d^m}{d\theta^m} f_\theta(x) \right| \leq K_f^m(x) ,$$

ν almost surely, and

- *for all $m \leq n+1$*

$$\int ||x||^k K_f^m(x)\, \nu(dx) < \infty .$$

Let μ_θ be n times Lipschitz differentiable, then probability measures $\mu_\theta^{(n,+)}, \mu_\theta^{(n,-)} \in \mathcal{M}_1$ and a constant $c^{(n)} \geq 0$ exist such that

$$\frac{d^n}{d\theta^n} \int g\, d\mu_\theta = c^{(n)} \left(\int g\, d\mu_\theta^{(n,+)} - \int g\, d\mu_\theta^{(n,-)} \right) , \tag{9}$$

for all $g \in C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$. The triple $(c^{(n)}, \mu_\theta^{(n,+)}, \mu_\theta^{(n,-)})$ is called the n^{th} order weak derivative of μ_θ, the measure $\mu_\theta^{(n,+)}$ is called the positive part and $\mu_\theta^{(n,-)}$ is called the negative part of the weak derivative, and the constant $c^{(n)}$ is called the normalisation constant. If the right–hand side in (9) equals zero for all $g \in C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$, we say that the n^{th} weak derivative of μ_θ is *not significant*, whereas we call it *significant* otherwise. If the n^{th} weak derivative of μ_θ is not significant, we take $(0, \mu_\theta, \mu_\theta)$ as the n^{th} order weak derivative. We denote by $s(\mu_\theta)$ the supremum over the set of all orders n such that the n^{th} order weak derivative of μ_θ is significant. If $s(\mu_\theta) = \infty$, we say that μ_θ is ∞ times weakly differentiable.

Weak differentiability of a random variable is defined by the same property of the induced measure. Let $(X_\theta : \theta \in \Theta)$ be defined on the common probability space (Ω, F, P) such that $P^{X_\theta} = \mu_\theta$. We call $(c^{(n)}, X_\theta^{(n,+)}, X_\theta^{(n,-)})$ an n^{th} order weak derivative of X_θ if the distribution of X_θ denoted by μ_θ has an n^{th} order weak derivative, and $X_\theta^{(n,+)}$ is distributed according to $\mu_\theta^{(n,+)}$, and $X_\theta^{(n,+)}$ according to $\mu_\theta^{(n,+)}$, respectively. Hence, the n^{th} order weak derivative satisfies

$$\frac{d^n}{d\theta^n} E[g(X_\theta)] = c^{(n)} \Big(E\Big[g(X_\theta^{(n,+)}) \Big] - E\Big[g(X_\theta^{(n,-)}) \Big] \Big) , \tag{10}$$

for all $g \in C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$. If the n^{th} order weak derivative of μ_θ is not significant, we take $(0, X_\theta, X_\theta)$ as the n^{th} order weak derivative of X_θ. We set $s(X_\theta) = s(\mu_\theta)$.

Example 3

1. *Let X_θ be exponentially distributed with Lebesgue density $f(x) = \theta e^{-\theta x}$ for $x \geq 0$ and $\theta \in \Theta = [\theta_l, \theta_r] \subset (0, \infty)$. Then X_θ is ∞ times weakly differentiable with respect to $C_p(\mathbb{R}_\epsilon)$ for all p. Moreover, the Lipschitz constants are*

$$K_f^0(x) = e^{-\theta_l x}$$

and for $n > 0$

$$K_f^n(x) = (\theta_r x + n)\; x^{n-1}\, e^{-\theta_l x}\,.$$

for $n \geq 1$. For the normalisation constant of the n^{th} derivative of X_θ we obtain

$$c^{(n)} = \left(\frac{n}{\theta\, e}\right)^n\,.$$

All higher-order weak derivatives are significant, that is, $s(X_\theta) = \infty$.

2. *Let X_θ be Bernoulli distributed on $\Sigma = \{D_1, D_2\} \subset \mathbb{R}_\epsilon^{J\times I}$ with $P(X_\theta = D_1) = \theta = 1 - P(X_\theta = D_2)$. Let ν be the uniform distribution on Σ and denote the Radon–Nikodyn derivative of $P^{X_\theta}(\cdot)$ with respect to ν by*

$$f_\theta(x) \;=\; f(\mu_\theta, \nu) = \frac{P(X_\theta = x)}{\nu(\{x\})} \;=\; 1_{D_1}(x)\, 2\theta \;+\; 1_{D_2}(x)\, 2(1-\theta)\,,$$

then X_θ is ∞ times Lipschitz differentiable on $[0,1]$ with respect to $C_p(S)$ for all p. Moreover, the Lipschitz constants are

$$K_f^n(x) = 2$$

for $n = 0, 1$ and the normalisation constant of the first-order weak derivative of X_θ is 1. Since the second-order and all higher-order weak derivatives of X_θ are not significant, we obtain $c^{(n)} = 0$ for $n > 1$ and $s(X_\theta) = 1$. Moreover, we obtain $X_\theta^{(1)} = (1, D_1, D_2)$.

In what follows we treat random vectors and random matrices. The n^{th} weak derivative of a random matrix $A \in \mathbb{R}_\epsilon^{J\times I}$ is a triple $(c(n), A^{(n,+)}, A^{(n,-)})$. Matrices as a whole can be viewed as random variables, like in Example 2, or they can be viewed as constituted out of more elementary random variables, like matrix $A(k)$ in Example 1, which is determined through the (random) service times. We call $X_1, \ldots, X_m \in \mathbb{R}_\epsilon$ the *input* of $A \in \mathbb{R}_\epsilon^{J\times I}$ if the entries of A are measurable mappings of $(X_1, \ldots, X_m)$. For example, the input of the transition matrix $A(k)$ of a J-dimensional (max,+)-linear stochastic system, as described in Section 2, is the vector of service times $(\sigma_j(k) : j \leq J)$. Let the matrix $A \in \mathbb{R}_\epsilon^{J\times I}$ depend on θ only through an input variable $X_\theta \in \mathbb{R}_\epsilon$ and let X_θ be stochastically independent of all other input variables of A, then the n times weak differentiability of X_θ with respect to $C_p(\mathbb{R}_\epsilon^{J\times I})$ implies that A is n times weakly differentiable on Θ with respect to $C_p(\mathbb{R}_\epsilon^{J\times I})$.

5 The Extended State–Space $M^{J\times I}$

The basic property of weakly differentiable random variables is that if $A, B \in \mathbb{R}_\epsilon^{J\times J}$ are stochastically independent and n times weakly differentiable, then $A \otimes B$ and $A \oplus B$ are n times weakly differentiable. Unfortunately, the state-space of the weak derivative of $A \oplus B$ is different from the state-space of A and B. Therefore, we extend the space $\mathbb{R}_\epsilon^{J\times I}$ to a space, called $M^{J\times I}$, in such a way that higher-order weak derivatives of general (max, +) expressions are elements of $M^{J\times I}$.

The set $M^{J\times I}$ is the set of all finite sequences of triples (c, A, B) with $c \in \mathbb{R}$ and $A, B \in \mathbb{R}_\epsilon^{J\times I}$. A generic element $\alpha \in M^{I\times J}$ is therefore given by

$$\alpha = ((c_1, A_1, B_1), (c_2, A_2, B_2), \ldots, (c_{n_\alpha}, A_{n_\alpha}, B_{n_\alpha}))\,,$$

where $n_\alpha < \infty$ is called the *length* of α. If α is of length one, that is, $n_\alpha = 1$, α is called *elementary*. Observe that the n^{th} weak derivative $(c^{(n)}, A^{(n,+)}, A^{(n,-)})$ of a matrix A is an elementary element of $M^{I\times J}$. We embed $\hat{\mathbb{R}}_\epsilon^{J\times I}$ via a monomorphism τ into $M^{J\times I}$ through $\tau(A) = (1, A, A)$ for all $A \in \hat{\mathbb{R}}_\epsilon^{J\times I}$. On $M^{I\times J}$ we introduce the binary operation "+" as the concatenation of strings: for $\alpha, \beta \in M^{I\times J}$, application of the "+" operator yields

$$\alpha + \beta = (\alpha_1, \ldots, \alpha_{n_\alpha}, \beta_1, \ldots, \beta_{n_\beta}) .$$

For $\alpha = (c^\alpha, A^\alpha, B^\alpha)$ and $\beta = (c^\beta, A^\beta, B^\beta)$ elementary we set

$$\alpha \oplus \beta = (c^\alpha \cdot c^\beta, A^\alpha \oplus A^\beta, B^\alpha \oplus B^\beta)$$

and

$$\alpha \otimes \beta = (c^\alpha \cdot c^\beta, A^\alpha \otimes A^\beta, B^\alpha \otimes B^\beta) ,$$

where $x \cdot y$ denotes the conventional multiplication in $\mathbb{R}$. These definitions are extended to general $\alpha = (\alpha_1, \ldots \alpha_{n_\alpha})$, $\beta = (\beta_1, \ldots \beta_{n_\beta})$, with α_i, β_i elementary, as follows. The $\oplus$-sum is given by

$$\alpha \oplus \beta = \sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \alpha_i \oplus \beta_j ,$$

for $\alpha, \beta \in M^{I\times J}$, that is, $\alpha \oplus \beta$ is the concatenation of all elementary $\oplus$-sums, which implies $n_{\alpha\oplus\beta} = n_\alpha \cdot n_\beta$. For the $\otimes$-product we set

$$\alpha \otimes \beta = \sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \alpha_i \otimes \beta_j ,$$

for $\alpha \in M^{I\times J}$ and $\beta \in M^{J\times K}$, that is, $\alpha \otimes \beta$ is the concatenation of all elementary $\otimes$-products, which implies $n_{\alpha\otimes\beta} = n_\alpha \cdot n_\beta$. In particular, for $\alpha \in M^{I\times J}$ and $x \in M^J := M^{J\times 1}$ the matrix–vector product $\alpha \otimes x$ is defined.

The performance functions $g : \hat{\mathbb{R}}_\epsilon^{I\times J} \to \mathbb{R}$ are extended to $M^{J\times I}$ as follows. For $\alpha = ((c_1, A_1, B_1), \ldots, (c_{n_\alpha}, A_{n_\alpha}, B_{n_\alpha})) \in M^{I\times J}$ we set

$$g^\tau(\alpha) = \sum_{i=1}^{n_\alpha} c_i \Big(g(A_i) - g(B_i) \Big) . \tag{11}$$

The mapping $g^\tau(\cdot)$ is called the *τ-projection of α with respect to g onto* $\mathbb{R} \cup \{-\infty\}$, or the (τ, g) *projection* for short.

The definition of g^τ resembles the structure of the formula on the left-hand side of (9). More precisely, let $A^{(n)} \in M^{J\times I}$ be a n^{th} derivative of $A \in \hat{\mathbb{R}}_\epsilon^{J\times I}$, then evaluating g^τ for $A^{(n)}$ yields

$$g^\tau(A^{(n)}) = c^{(n)} \Big(g\left(A^{(n,+)}\right) - g\left(A^{(n,-)}\right) \Big) .$$

On the other hand, every triple (c, B, C) with $d^n E[g(A)]/d\theta^n = c(E[g(B)] - E[g(C)])$ for all $g \in C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$ is an n^{th} order weak derivative of A. We now say that $\alpha, \beta \in M^{J\times I}$ are *weakly equal* (with respect to $C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$) if

$$\forall g \in C_p(\hat{\mathbb{R}}_\epsilon^{J\times I}) : \quad E[g^\tau(\alpha)] = E[g^\tau(\beta)] ,$$

in symbols: $\alpha \equiv \beta$. Hence, we obtain

$$A^{(n)} \equiv (c^{(n)}, A^{(n,+)}, A^{(n,-)}) .$$

For ease of notation, we suppress the superscript τ where this causes no confusion and write, for example, $g(\cdot)$ instead of $g^\tau(\cdot)$.

6 The "Halted" (max,+) System

The Lyapunov exponent can be represented by the difference between two products of matrices, where the number of matrices for each product is given by the stopping time η, see Section 3.2. Hence, the analyticity of the Lyapunov exponent can be deduced from the analyticity of the product $E[\bigotimes_{k=-\eta}^{0} A(k) \otimes x_0]$. In [7], sufficient conditions for the analyticity of $E[\bigotimes_{k=-m}^{0} A(k) \otimes x_0]$ were given, for $m \in \mathbb{N}$. Unfortunately, the situation we are faced with here is more complicated, since η is random and depends on θ. To deal with the situation, we borrow an idea from the theory of Markov chains. There, the expectation over a random number of transitions of a Markov chain is analysed by introducing an absorbing state. More precisely, a new Markov kernel is defined, such that, once the chain reaches a specified criterion, like entering a certain set, the chain is forced to go to the absorbing state and remains there forever. Following a similar train of thoughts, we introduce in this section a "halted" version of $A(k)$, denoted by $A_{\tilde{a}}(k)$, where $A_{\tilde{a}}(k)$ will be constructed in such a way that it equals $A(k)$ as long as the pattern $\tilde{a}$, defined in Section 3.2, has not occurred in the sequence $A(0), A(-1), \ldots, A(-k)$. Once the pattern $\tilde{a}$ has occurred, the operator $A_{\tilde{a}}(k)$ is set to E, the identity matrix. In other words, $A_{\tilde{a}}(k)$ "halts" the backward evolution of the system dynamics as soon as the sequence $\tilde{a}$ occurs.

In what follows we describe the construction of $A_{\tilde{a}}(k)$ in more detail. Let $y(-k) = 0$ if $\tilde{a}$ hasn't occurred in $A(0), A(-1), \ldots, A(-k)$ and $y(-k) = 1$ otherwise.

For $k < 0$, we now set $A_{\tilde{a}}(k)$ as follows

$$A_{\tilde{a}}(k) := \begin{cases} A(k) \,, & y(k) = 0 \\ E \,, & y(k) = 1 \,. \end{cases} \tag{12}$$

Analogously to x_{-m}^{0} we now consider the backward evolution of a system driven by $A_{\tilde{a}}(k)$ (instead of $A(k)$). More precisely, we set

$$\xi_{-m}^{0} := \bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0 \,,$$

with $\xi_0^0 := x_0$. The value of $y(k)$ changes at $-\eta$, because at this time we observe the pattern $\tilde{a}$ for the first time. Going backwards in time beyond $-\eta$, the matrix $A_{\tilde{a}}(k)$ equals E, that is, the variable ξ_{-m}^{0} doesn't change its value after $-\eta$, or, more formally

$$\xi_{-m}^{0} = \begin{cases} x_{-m}^{0} \,, & m \leq \eta \\ x_{-\eta}^{0} \,, & m > \eta \,, \end{cases} \tag{13}$$

and

$$\begin{aligned} \lim_{m \to \infty} \bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0 &= \bigotimes_{k=-\infty}^{0} A_{\tilde{a}}(k) \otimes x_0 \\ &= \bigotimes_{k=-\eta}^{0} A(k) \otimes x_0 \,. \end{aligned}$$

If $A(k)$ has an n^{th} order weak derivative, then we define for $i \in \{+1, -1\}$

$$(A_{\tilde{a}})^{(n,i)}(k) := \begin{cases} A^{(n,i)}(k) \,, & y(k) = 0 \text{ (that is, } k \leq \eta) \\ E \,, & y(k) = 1 \text{ (that is, } k > \eta) \,, \end{cases} \tag{14}$$

with $A^{(n,0)}(k) := A(k)$, and

$$c_{A_{\tilde{a}}(k)}^{(n)} := \begin{cases} c_A^{(n)} \,, & y(k) = 0 \text{ (that is, } k \leq \eta) \\ 0 \,, & y(k) = 1 \text{ (that is, } k > \eta) \,. \end{cases} \tag{15}$$

In what follows we establish a Leibniz rule for higher-order Lipschitz differentiation which resembles the classical Leibniz rule of higher-order differentiation of products of real-valued functions. However, before we can state the Leibniz rule we have to introduce the following multi-indices. In order to mimic the backward evolution in time we number the multi-indices by indices out of $\{k : k \leq 0\}$. For $n, m \in \mathbb{N}$ and measures $\mu_k \in \mathcal{M}$ $(0 \geq k \geq -m)$, we set

$$\mathcal{L}(m,n) = \mathcal{L}_{(\mu_0,\dots\mu_m)}(m,n)$$
$$:= \left\{ (l_0, l_{-1}, \dots, l_{-m}) \in \{0,\dots,n\}^{m+1} \,\middle|\, l_k \leq s(\mu_k) \text{ and } \sum_{k=-m}^{0} l_k = n \right\},$$

and for $l \in \mathcal{L}(m,n)$ we introduce the set

$$\mathcal{I}(m,l) :=$$
$$\left\{ (i_0, i_{-1}, \dots, i_{-m}) \in \{-1,0,+1\}^{m+1} \,\middle|\, i_k = 0 \text{ iff } l_k = 0 \text{ and } \prod_{\substack{i_0,\dots,i_{-m} \\ i_k \neq 0}} i_k = +1 \right\}.$$

Moreover, we introduce for $i \in \mathcal{I}(m,l)$, with $m, n \geq 0$ and $l \in \mathcal{L}(m,n)$, the multi-index i^- as follows. Let $k^* = k^*(i)$ be the last position of a non-zero entry in i, that is, $i_k = 0$ for all $|k| < |k^*|$, and set

$$i^- := (i_{0,-1}, \dots, i_{-k})^- = (i_0, \dots i_{-k^*+1}, -i_{-k^*}, \underbrace{i_{-k^*-1}}_{=0}, \dots, \underbrace{i_{-m}}_{=0}),$$

that is, the multi-index i^- is generated out of i be changing the sign of the last non-zero entry of i.

Theorem 5 *Let $A(k)$, for $0 \geq k \geq -m$, be mutually stochastically independent and n times Lipschitz differentiable, then it holds that*

$$\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \right)^{(n)} \equiv \sum_{l \in \mathcal{L}(m,n)} \frac{n!}{l_0! l_{-1}! \dots l_{-m}!} \sum_{i \in \mathcal{I}(m,l)} \left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \right)^{(l,i)}$$

with

$$\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \right)^{(l,i)} = \left(\prod_{k=-m}^{0} c^{(l_k)}, \bigotimes_{k=-m}^{0} A_{\tilde{a}}^{(l_k,i_k)}(k) \bigotimes_{k=-m}^{0} A_{\tilde{a}}^{(l_k,i_k^-)}(k) \right).$$

Proof: Let ν be a probability measure on $\hat{\mathbb{R}}_\epsilon^{J\times J}$, such that the ν-density f_θ of A exist. Furthermore, let $f_\theta^{(m,i)}$ denote the ν-density of $A^{(m,i)}$ for all $m \leq n$ and $i \in \{-1,0,+1\}$. Recall that M denotes the length of $\tilde{a}$. For $M - 1 \leq -j < m$, let $\mathbf{A}(j)$ be the set of all $(m+1)$ tuples of matrices in $\hat{\mathbb{R}}_\epsilon^{J\times J}$, such that the entries $-j + M - 1$ to $-j$ equal $\tilde{a}$, or, more formally

$$\mathbf{A}(j) := \left\{ (a_0, a_{-1}, \dots, a_{-m}) \in \hat{\mathbb{R}}_\epsilon^{J\times J} : j = \min\{k : a_{-k-i} = \tilde{a}^{M-i} \text{ for } 0 \leq i \leq M-1\} \right\},$$

and set $\mathbf{A}(j) := \emptyset$ for $-j > -M$. The set $\mathbf{A}(m)$ is defined as follows

$$\mathbf{A}(m) := \left(\hat{\mathbb{R}}_\epsilon^{J\times J} \right)^{m+1} \setminus \bigcup_{j=-(m-1)}^{0} \mathbf{A}(j).$$

Then for all $g \in C_p(\hat{\mathbb{R}}_\epsilon^{J\times J})$

$$E\left[g\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0 \right) \right]$$
$$= \sum_{j=-m}^{0} \int_{\mathbf{A}(j)} g\left(\bigotimes_{k=-j}^{0} a_k \otimes x_0 \right) \prod_{k=-j}^{0} f_\theta(a_k)\, \nu(da_0, \dots, da_{-j}) \tag{16}$$

We prove the theorem in three steps: (a) we show that we may interchange the order of integration and n-fold differentiation for (16), (b) we calculate the n^{th} order derivative of the product of the densities and split up the derivatives in their positive and negative parts, and (c) we show that the resulting integral can be written as the expected value of the random variable given in the statement of the theorem.

Step (a): Under the conditions of the theorem, it follows from Lemma 5 in [7] that we may interchange n-fold differentiation and integration over the set $(\mathbb{R}_\epsilon^{J\times J})^{j+1}$. Since $\mathbf{A}(j)$ is a measurable subset of this set, we may interchange the order of integration over the set $\mathbf{A}(j)$ and n-fold differentiation, as well.

Step (b): In order to calculate the positive and negative part of the n-fold derivative of the product density, we proceed as for the proof of Lemma 5 in [7]: (i) we calculate the derivative of the product of the densities via the Leibniz rule of classical analysis, (ii) we split up the individual derivatives into their positive and negative parts, and (iii) we regroup these terms. This procedure is independent of the particular set $\mathbf{A}(j)$ and we refer to the proof of Lemma 5 in [7] for details.

Step (c): We have already shown that for all $g \in C_p(\mathbb{R}_\epsilon^J)$

$$\begin{aligned}
&\frac{d^n}{d\theta^n} \sum_{j=-m}^{0} \int_{\mathbf{A}(j)} g\left(\bigotimes_{k=-j}^{0} a_k \otimes x_0\right) \prod_{k=-j}^{0} f_\theta(a_k)\, \nu(da_0,\ldots,da_{-j}) \\
&= \sum_{j=-m}^{0} \sum_{l\in\mathcal{L}(j,n)} \frac{n!}{l_0! l_{-1}! \ldots l_{-j}!} \sum_{i\in\mathcal{I}(j,l)} \prod_{k=-j}^{0} c^{(l_k)} \\
&\quad \left(\int_{\mathbf{A}(j)} g\left(\bigotimes_{k=0}^{j} a_k \otimes x_0\right) \prod_{k=-j}^{0} f_\theta^{(l_k,i_k)}(a_k)\, \nu(da_0,\ldots,da_{-j}) \right. \\
&\qquad \left. - \int_{\mathbf{A}(j)} g\left(\bigotimes_{k=-j}^{0} a_k\right) \prod_{k=-j}^{0} f_\theta^{(l_k,i_k^-)}(a_k)\, \nu(da_0,\ldots,da_{-j}) \right) .
\end{aligned}$$

Let $\delta_E(x)$ be equal to one if $x = E$ and zero otherwise. The measure $\delta_E(\cdot)$ is independent of θ which implies $\delta_E^{(l,i)}(\cdot) = \delta_E(\cdot)$ for all $l \geq 0$ and $i \in \{+1, 0, -1\}$. We now "fill up" the missing densities with $\delta_E(\cdot)$ in order to obtain an $(m+1)$-fold product on the sets $\mathbf{A}(j) \times (\mathbb{R}_\epsilon^{J\times J})^{m+j}$, for $j < 0$. More precisely, we write

$$\begin{aligned}
&\frac{d^n}{d\theta^n} \sum_{j=-m}^{0} \int_{\mathbf{A}(j)} g\left(\bigotimes_{k=-j}^{0} a_k \otimes x_0\right) \prod_{k=-j}^{0} f_\theta(a_k)\, \nu(da_0,\ldots,da_{-j}) \\
&= \sum_{j=-m}^{0} \sum_{l\in\mathcal{L}(j,n)} \frac{n!}{l_0! l_{-1}! \ldots l_{-j}!} \sum_{i\in\mathcal{I}(j,l)} \prod_{k=-j}^{0} c^{(l_k)} \int_{\mathbf{A}(j)\times(\mathbb{R}_\epsilon^{J\times J})^{m+j}} g\left(\bigotimes_{k=-j}^{0} a_k \otimes E^{m+j} \otimes x_0\right) \\
&\quad \left(\prod_{k=-j}^{0} f_\theta^{(l_k,i_k)}(a_k) \prod_{k=-m}^{-(j+1)} \delta_E(a_k) - \prod_{k=-j}^{0} f_\theta^{(l_k,i_k^-)}(a_k) \prod_{k=-m}^{-(j+1)} \delta_E(a_k) \right) \\
&\qquad \nu(da_0,\ldots,da_{-j}) da_{-j-1} \cdots da_{-m} .
\end{aligned}$$

Using our definition of $A_{\tilde{a}}^{(l_k,i_k)}(k)$, see (14), we obtain

$$= \sum_{l\in\mathcal{L}(m,n)} \frac{n!}{l_0! l_{-1}! \ldots l_{-m}!} \sum_{i\in\mathcal{I}(m,l)} \prod_{k=-m}^{0} c^{(l_k)} E\left[g\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}^{(l_k,i_k)} \otimes x_0\right) - g\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}^{(l_k,i_k^-)} \otimes x_0\right)\right] ,$$

which concludes the proof of the theorem. □

Remark 2 *If $A \in \mathbb{R}_\epsilon^{J\times J}$ is n times weakly differentiable and $x_0 \in \mathbb{R}_\epsilon^J$ is independent of θ, then $(A \otimes x_0)^{(n)} \equiv A^{(n)} \otimes x_0$. Under the conditions of the above theorem this implies the following rule of computation*

$$\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0 \right)^{(n)} \equiv \left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \right)^{(n)} \otimes x_0 \,.$$

The intuitive explanation for the above formula is that, since x_0 does not depend on θ, all (higher-order) weak derivatives of x_0 are "zero".

We now turn to the pathwise derivatives. From the above theorem it follows for the first-order weak derivative that

$$\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0 \right)^{(1)} \equiv \sum_{j=-m}^{0} \bigotimes_{k=j+1}^{0} A_{\tilde{a}}(k) \otimes A_{\tilde{a}}^{(1)}(j) \otimes \bigotimes_{k=-m}^{j-1} A_{\tilde{a}}(k) \otimes x_0 \,.$$

In accordance with (15), the summands on the right-hand side of the above equation are equal to zero for $|j| > \eta$ (this follows from the fact that $A_{\tilde{a}}^{(1)}(k) = (0, E, E)$ for $|j| > \eta$). Hence, taking the limit for m towards infinity yields

$$\begin{aligned}
\lim_{m\to\infty} \left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0 \right)^{(1)} &\equiv \sum_{j=-\eta}^{0} \lim_{m\to\infty} \bigotimes_{k=j+1}^{0} A_{\tilde{a}}(k) \otimes A_{\tilde{a}}^{(1)}(j) \otimes \bigotimes_{k=-m}^{j-1} A_{\tilde{a}}(k) \otimes x_0 \\
&\equiv \sum_{j=-\eta}^{0} \bigotimes_{k=j+1}^{0} A_{\tilde{a}}(k) \otimes A_{\tilde{a}}^{(1)}(j) \otimes \bigotimes_{k=-\infty}^{j-1} A_{\tilde{a}}(k) \otimes x_0 \,.
\end{aligned}$$

The above formula actually describes the difference between two processes, where for one process we replace $A_{\tilde{a}}^{(1)}$ by $A_{\tilde{a}}^{(1,+)}$ and for the other version by $A_{\tilde{a}}^{(1,-)}$. Following the same line of thought as in Section 3.2, we see that the sequence $\{A_{\tilde{a}}(k) : k < -j-1\}$ only contributes as long as the pattern $\tilde{a}$ hasn't been observed. Put another way, we can cut off the backward recursion after having observed $\tilde{a}$. Let $\eta^{(1)}$ be the number of transitions in $\{A(k) : k < -\eta\}$ until we have observed the pattern again, and to unify notation set $\eta^{(0)} := \eta$. Then the expression on the right-hand side of the above equivalence reads

$$\begin{aligned}
&\equiv \sum_{j=-\eta^{(0)}}^{0} \bigotimes_{k=j+1}^{0} A_{\tilde{a}}(k) \otimes A_{\tilde{a}}^{(1)}(j) \otimes \bigotimes_{k=-\eta^{(0)}}^{j-1} A_{\tilde{a}}(k) \otimes \bigotimes_{k=-\eta^{(1)}}^{-\eta^{(0)}-1} A_{\tilde{a}}(k) \otimes x_0 \\
&\equiv \sum_{l\in\mathcal{L}(\eta^{(0)},1)} \frac{n!}{l_0! l_{-1}! \dots l_{-\eta^{(0)}}!} \sum_{i\in\mathcal{I}(\eta^{(0)},l)} \left(\bigotimes_{k=-\eta^{(0)}}^{0} A_{\tilde{a}}(k) \right)^{(l,i)} \otimes \bigotimes_{k=-\eta^{(1)}}^{-\eta^{(0)}-1} A_{\tilde{a}}(k) \otimes x_0 \,.
\end{aligned}$$

The above formula illustrates an important phenomena: the range of the product changes with the order of differentiation. More precisely, suppose we want to weakly differentiate the above expression in order to calculate iteratively the second-order weak derivative of the $\eta^{(0)}$ product of $A(k)$. This procedure has two steps. First, we take the first-order weak derivative of the $\eta^{(0)}$ products, which yields the above product of $\eta^{(1)}$ matrices, with $\eta^{(1)} > \eta^{(0)}$. Secondly, we weakly differentiate this product again. However, by taking the first-order weak derivative, we increased the product by the term

$$\bigotimes_{k=-\eta^{(1)}}^{-\eta^{(0)}-1} A_{\tilde{a}}(k) \,.$$

Using finite induction, we see that taking the sample limit of the n^{th} order weak derivative yields

$$\lim_{m\to\infty}\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k)\otimes x_0\right)^{(n)}$$

$$\equiv \sum_{l\in\mathcal{L}(\eta^{(n-1)},n)} \frac{n!}{l_0!l_{-1}!\ldots l_{-\eta^{(n-1)}}!} \sum_{i\in\mathcal{I}(\eta^{(n-1)},l)} \left(\bigotimes_{k=-\eta^{(n-1)}}^{0} A_{\tilde{a}}(k)\right)^{(l,i)} \otimes \bigotimes_{k=-\eta^{(n)}}^{-\eta^{(n-1)}-1} A_{\tilde{a}}(k)\otimes x_0$$

$$\equiv \left(\bigotimes_{k=-\eta^{(n-1)}}^{0} A_{\tilde{a}}(k)\right)^{(n)} \otimes \bigotimes_{k=-\eta^{(n)}}^{-\eta^{(n-1)}-1} A_{\tilde{a}}(k)\otimes x_0 \,. \tag{17}$$

The following lemma shows that the above expression is indeed an unbiased estimator for the n^{th} order derivative of $E[\bigotimes_{k=-\eta}^{0} A(k)\otimes x_0]$. To abbreviate the notation, we set

$$B(\eta,n) = \sum_{l\in\mathcal{L}(\eta^{(n-1)},n)} \frac{n!\prod_{k=-\eta^{(n-1)}}^{0}\max(c^{(l_k)},1)}{l_0!l_{-1}!\ldots l_{-\eta^{(n-1)}}!} \sum_{i\in\mathcal{I}(\eta^{(n-1)},l)}$$

$$\sum_{k=-\eta^{(n-1)}}^{0}\left(||A^{(l_k,i_k)}(k)|| + ||A^{(l_k,i_k)}(k)||\right) + \sum_{k=-\eta^{(n)}}^{-\eta^{(n-1)}-1} ||A(k)|| \,.$$

Lemma 2 *Under assumption* **(C4)**, *let $A(k)$ $(0\geq k)$ be mutually stochastically independent and n times Lipschitz differentiable with respect to $C_p(\hat{\mathbb{R}}_\epsilon^J)$. Then it holds for all $g\in C_p(\hat{\mathbb{R}}_\epsilon^J)$ such that g is non-expansive and for all m with probability one that*

$$\left| g\left(\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k)\right)^{(n)}\otimes x_0\right)\right| \leq B(\eta,n)\,.$$

Moreover, if $E[B(\eta,n)]<\infty$, then

$$\frac{d^n}{d\theta^n}E\left[g\left(\bigotimes_{k=-\eta}^{0} A_{\tilde{a}}(k)\otimes x_0\right)\right]$$

$$= E\left[g\left(\left(\bigotimes_{k=-\eta^{(n-1)}}^{0} A_{\tilde{a}}(k)\right)^{(n)} \otimes \bigotimes_{k=-\eta^{(n)}}^{-\eta^{(n-1)}-1} A_{\tilde{a}}(k)\otimes x_0\right)\right].$$

Proof: We prove the first part of the lemma. For $m\geq 1$, let $l\in\mathcal{L}(m,n)$ and $i\in\mathcal{I}(m,l)$, then

$$\left| g\left(\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k)\right)^{(l,i)}\otimes x_0\right)\right|$$

$$= \prod_{k=-m}^{0} c^{(l_k)}\left| g\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}^{(l_k,i_k)}(k)\otimes x_0\right) - g\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}^{(l_k,i_k^-)}(k)\otimes x_0\right)\right|$$

$$\leq \prod_{k=-m}^{0} c^{(l_k)}\delta\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}^{(l_k,i_k)}(k)\otimes x_0,\ \bigotimes_{k=-m}^{0} A_{\tilde{a}}^{(l_k,i_k^-)}(k)\otimes x_0\right),$$

where the inequality follows from the non-expansiveness of g. Applying Lemma 1 yields that the last formula is smaller than or equal to

$$\prod_{k=-m}^{0} c^{(l_k)} \sum_{k=-m}^{0} \delta(A_{\tilde{a}}^{(l_k,i_k)}(k), A_{\tilde{a}}^{(l_k,i_k^-)}(k))\,.$$

Using the fact that for matrices A, B of the same size it holds that

$$0 \le \delta(A,B) \le ||A|| + ||B||$$

we obtain

$$\left| g\left(\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \right)^{(l,i)} \otimes x_0 \right) \right| \le \prod_{k=-m}^{0} c^{(l_k)} \sum_{k=-m}^{0} \delta(A_{\tilde{a}}^{(l_k,i_k)}(k), A_{\tilde{a}}^{(l_k,i_k^-)}(k)) \, .$$

For the n^{th} order weak derivative we have

$$A_{\tilde{a}}^{(l_k,i_k)}(k) = A_{\tilde{a}}^{(l_k,i_k^-)}(k) = A_{\tilde{a}}(k) \, , \quad \text{for } |k| > \eta^{(n-1)} \, ,$$

where $A_{\tilde{a}}(k)$ is either equal to $A(k)$ or to E, that is, for $|k| > \eta^{(n-1)}$ it holds that

$$\delta(A_{\tilde{a}}^{(l_k,i_k)}(k), A_{\tilde{a}}^{(l_k,i_k^-)}(k)) \le ||A(k)|| \, .$$

For $|k| > \eta^{(n)}$, we have $A_{\tilde{a}}(k) = E$ with probability one. This yields

$$\begin{aligned}
&\left| g\left(\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \right)^{(l,i)} \otimes x_0 \right) \right| \\
&\le \prod_{k=-\min(m,\eta^{(n-1)})}^{0} c^{(l_k)} \sum_{k=-\min(m,\eta^{(n-1)})}^{0} \left(||A_{\tilde{a}}^{(l_k,i_k)}(k)|| + ||A_{\tilde{a}}^{(l_k,i_k^-)}(k)|| \right) \\
&\qquad + \sum_{k=-\min(m,\eta^{(n)})}^{-\min(m,\eta^{(n-1)})-1} ||A(k)|| \\
&\le \prod_{k=-\eta^{(n-1)}}^{0} \max(c^{(l_k)},1) \sum_{k=-\eta^{(n-1)}}^{0} \left(||A_{\tilde{a}}^{(l_k,i_k)}(k)|| + ||A_{\tilde{a}}^{(l_k,i_k^-)}(k)|| \right) + \sum_{k=-\eta^{(n)}}^{-\eta^{(n-1)}-1} ||A(k)|| \, .
\end{aligned}$$

Taking the sum over all $l \in \mathcal{L}(m,n)$ and $i \in \mathcal{I}(l,n)$ and extending m to $\eta^{(n-1)}$ concludes the proof of the first part of the lemma.

We now turn to the proof of the second part of the lemma. Theorem 5 (the Leibniz rule for the operator $A_{\tilde{a}}$) implies

$$\lim_{m\to\infty} \frac{d^n}{d\theta^n} E\left[g\left(\bigotimes_{k=0}^{m} A_{\tilde{a}}(k) \otimes x_0 \right) \right] = \lim_{m\to\infty} E\left[g\left(\left(\bigotimes_{k=0}^{m} A_{\tilde{a}}(k) \right)^{(n)} \otimes x_0 \right) \right] .$$

In accordance with the first part of the lemma, $B(\eta, n)$ is a dominating function for the sample weak derivatives of the m-fold product. Furthermore, the sample limit of the weak derivative, see (17), exists and is bounded by $B(\eta, n)$ as well. Hence, the dominated convergence theorem applies, which concludes the proof of the lemma. □

Example 4 *Let $\{A(k)\}$ be a sequence of i.i.d. Bernoulli distributed matrices with state space $\{D_1, D_2\} \subset \mathbb{R}_\epsilon^{J\times I}$ with parameter θ, cf. Example 2. We assume that assumptions* **(C1)** *to* **(C4)** *hold and that D_2 is the matrix associated with the pattern of $\{A(k)\}$ (indeed, D_2 is a scs1–cyc1 matrix); we denote the coupling time of D_2 by $c = c(D_2)$. We split up the sequence $\{A(k)\}$ into blocks of length c. The probability that all elements of such a block equal D_2 is $\hat{\theta} = \theta^c$. We denote the number of c blocks until we observe the first block that is completely constituted out of D_2 matrices by η_c, that is, $P(\eta_c = m) = (\hat{\theta})^{m-1}(1-\hat{\theta})$. Note that $\eta \le c\eta_c$ with probability one.*

Only the first-order weak derivative of $A(k)$ is significant with $A^{(1,+1)}(k) = D_1$, $A^{(1,-1)}(k) = D_2$ and $c^{(1)} = 1$. Applying Lemma 2 yields for any non-expansive g and $x_0 \in \mathbb{R}^J$

$$B(\eta, n) \leq \sum_{l \in \mathcal{L}(\eta^{(n-1)}, n)} n! \sum_{i \in \mathcal{I}(\eta^{(n-1)}, l)} \eta^{(n)}(||D_1|| + ||D_2||) \, .$$

The set $\mathcal{L}(\eta^{(n-1)}, n)$ has

$$\frac{\eta^{(n-1)}!}{n!(\eta^{(n-1)} - n)!} \leq \frac{(\eta^{(n-1)})^n}{n!}$$

elements. For $l \in \mathcal{L}(\eta^{(n-1)}, n)$, the set $\mathcal{I}(\eta^{(n-1)}, l)$ has 2^{n-1} elements, which stems from the fact that we place either a "+1" or a "−1" on one of the n places except for the last place, here we have to chose a value such that the overall product is positive. This yields

$$\begin{aligned} E[B(\eta, n)] &\leq 2^{n-1}(||D_1|| + ||D_2||) E\left[\eta^{(n)} \, (\eta^{(n-1)})^n\right] \\ &\leq 2^{n-1}(||D_1|| + ||D_2||) E\left[(\eta^{(n)})^{n+1}\right] \\ &\leq 2^{n-1}(||D_1|| + ||D_2||) E\left[(c\eta_c^{(n)})^{n+1}\right] \\ &= c^2(2c)^{n-1} \left(\frac{\hat{\theta}}{1-\hat{\theta}}\right)^{n+1} (||D_1|| + ||D_2||) \, , \end{aligned}$$

which is finite for all $\hat{\theta} = \theta^c \in [0, 1)$.

7 Weak Analyticity of Random Matrices

We now introduce the concept of weak analyticity.

Definition 2 *We call $A \in \hat{\mathbb{R}}_\epsilon^{J \times I}$ weakly analytical on Θ with respect to $C_p(\hat{\mathbb{R}}_\epsilon^{J \times I})$ if*

- *all higher-order weak derivatives of A exist on Θ with respect to $C_p(\hat{\mathbb{R}}_\epsilon^{J \times I})$, and*
- *there exists a measure $\nu \in \mathcal{M}_1(\hat{\mathbb{R}}_\epsilon^{J \times I})$ such that the ν-density of A, say f_θ, is analytical on Θ (that is, for all $\theta_0 \in \Theta$ there exists an interval D_{θ_0}, with $\theta_0 \in D_{\theta_0}$, such that the Taylor series of $f_\theta(x)$ developed at θ_0 converges ν-almost-surely to $f_\theta(x)$), and in addition to that*
- *for all $\theta_0 \in \Theta$, there exists $\mathbf{f}^D_{\theta_0}(x)$ such that the ν-density of A satisfies for all $\theta \in D_{\theta_0}$*

$$\forall x \in \hat{\mathbb{R}}_\epsilon^{J \times I} \; : \quad \sum_{n=0}^{\infty} \left| \frac{d^n}{d\theta^n} f_\theta(x) \bigg|_{\theta=\theta_0} \frac{1}{n!}(\theta - \theta_0)^n \right| \leq \mathbf{f}^D_{\theta_0}(x)$$

with

$$\int ||x||^k \, \mathbf{f}^D_{\theta_0}(x) \, \nu(dx) < \infty \, .$$

If $A \in \hat{\mathbb{R}}_\epsilon^{J \times I}$ is weakly analytical on Θ with respect to $C_p(\hat{\mathbb{R}}_\epsilon^{J \times I})$, then $E[g(A)]$ is analytical on Θ for all $C_p(\hat{\mathbb{R}}_\epsilon^{J \times I})$. In particular, if, for $\theta_0 \in \Theta$, the domain of convergence of the Taylor series of A is D_{θ_0}, then the domain of convergence of the Taylor series of $E[g(A)]$ is also D_{θ_0}.

Example 5 *1. Let A be exponentially distributed with Lebesgue density $f_\theta(x) = \theta \exp(-\theta\, x)$ and let $\Theta = (0, \infty)$, so that A is weakly analytical on Θ with respect to $C_p(\hat{\mathbb{R}}_\epsilon^{J \times I})$ for all p. For $\theta_0 \in (0, \infty)$, set $D_{\theta_0}(\delta) = [\delta, 2\theta_0 - \delta]$ for $\theta_0 > \delta > 0$. Then it can be shown that the Taylor series of $E[g(A)]$ developed at θ_0 has at least $D_{\theta_0}(\delta)$ as domain of convergence.*

2. *Let A be Bernoulli distributed on $\{D_1, D_2\} \subset \hat{\mathbb{R}}_\epsilon^{J\times I}$. Then μ_θ is ∞ times weakly differentiable and the derivative of the density of μ_θ with respect to a uniform distribution is uniformly bounded in θ by 1. Therefore, A is weakly analytical on $[0,1]$ on $C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$.*

Let $A \in \hat{\mathbb{R}}_\epsilon^{J\times I}$ depend on θ only through an input variable $X_\theta \in \hat{\mathbb{R}}_\epsilon$ and let X_θ be stochastically independent of all other input variables of A. If X_θ is analytical on Θ with respect to $C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$, then A is weakly analytical on Θ with respect to $C_p(\hat{\mathbb{R}}_\epsilon^{J\times I})$ and the domains of convergence of the Taylor series coincide.

The most important property of weak analyticity is that it is preserved under the (max,+)-operations.

Theorem 6 *If $A, B \in \hat{\mathbb{R}}_\epsilon^{J\times J}$ are stochastically independent and weakly analytical on Θ, then $A \otimes B$ and $A \oplus B$ are weakly analytical on Θ. In particular, if, for $\theta_0 \in \Theta$, the Taylor series of A has domain of convergence $D_{\theta_0}^A$ and the Taylor series of B $D_{\theta_0}^B$, then the domain of convergence of the Taylor series of $A \oplus B$, respectively $A \otimes B$, is $D_{\theta_0}^A \cap D_{\theta_0}^B$.*

Moreover, weak analyticity of A and B implies that of $A_{\bar{a}} \otimes B_{\bar{a}}$ and $A_{\bar{a}} \oplus B_{\bar{a}}$.

Proof: The first part of the theorem is Theorem 2 in [7] and we omit the proof.

We now turn to the proof of the second part of the theorem. Let μ_θ denote the distribution function of A and ν_θ the distribution function of B_θ. The weak analyticity of the product measure $\mu_\theta \times \nu_\theta$ over the set $\hat{\mathbb{R}}_\epsilon^{J\times J} \times \hat{\mathbb{R}}_\epsilon^{J\times J}$ was established in Theorem 2 in [7]. All arguments used in this proof remain valid when we integrate over a measurable subset of the state space. Hence, if we split up the state space in disjunct set representing the possible outcomes of $A_{\bar{a}} \otimes B_{\bar{a}}$ (cf. equation (16) in the proof of Theorem 5), then the proof of the second part of the theorem reduces to that of the first part. $\square$

An immediate consequence of Theorem 6 is that if $A(k) \in \hat{\mathbb{R}}_\epsilon^{J\times J}$ is an i.i.d. sequence of random matrices weakly analytical on Θ, then

$$x(k+1) \;=\; A(k) \otimes x(k)\,, \quad k \geq 0\,,$$

with $x(0) = x_0$ is weakly analytical on Θ for all k. In particular, $E[g(x(k+1)]$ is analytical on Θ for all $g \in C_m(\hat{\mathbb{R}}_\epsilon^{J})$ and $m \in \mathbb{N}$.

If, for $\theta_0 \in \Theta$, $A(0)$ has domain of convergence $D_{\theta_0}^A$, then $x(k+1)$ has domain of convergence $D_{\theta_0}^A$.

Example 6

1. *Consider the situation of Example 1 (1). In accordance with Example 5 (1), the transition matrix $A(k)$ is analytical on $(0,\infty)$. Moreover, $x(k+1)$, with $x(k+1) = A(k) \otimes x(k)$ for $k \geq 0$, is analytical on $(0,\infty)$ and for $g \in C_p(\hat{\mathbb{R}}_\epsilon^{J})$ the term $E[g(x(k+1))]$ can be developed at any $\theta_0 \in (0,\infty)$ into a Taylor series which has $D_{\theta_0}(\delta)$, with $\theta_0 > \delta > 0$, as domain of convergence.*

2. *In the Bernoulli case, $A(k)$ is weakly analytical on $[0,1]$ for $k \in \mathbb{N}$. Moreover, $x(k+1)$, with $x(k+1) = A(k) \otimes x(k)$ for $k \geq 0$, is analytical on $[0,1]$ and for $g \in C_p(\hat{\mathbb{R}}_\epsilon^{J})$, $p \in \mathbb{N}$, the term $E[g(x(k+1))]$ can be developed at any $\theta_0 \in [0,1]$ into a Taylor series which has $D_{\theta_0} = [0,1]$ as domain of convergence.*

8 Analytic Expansions

In this section we develop the Lyapunov exponents of (max,+)-linear systems into a Taylor series. More specifically, we study sequences $\{A(k)\} = \{A_\theta(k)\}$ with $\theta \in \Theta$, see Section 3.3, and the assumptions **(C1)** to **(C4)** have to be understood to hold for $\{A_\theta(k)\}$ for all $\theta \in \Theta$.

The Lyapunov exponent can be represented through products of a random number of matrices, see Section 3.2. Combining this representation with the above results, we obtain the following theorem.

Theorem 7 *Under assumptions* **(C1)** *to* **(C4)**. *If $A(0)$ is weakly analytical on Θ with respect to $C_1(\mathbb{R}_\epsilon^{J\times J})$ with domain of convergence $D(\theta_0)$, for $\theta_0 \in \Theta$, and if $E[B(\eta, n)]$ is finite for all n and*

$$\sum_{n=0}^{\infty} \frac{1}{n!} E[\,B(\eta, n)\,](\theta - \theta_0)^n < \infty\,,$$

then

$$\lim_{k\to\infty} E[x(k+1) - x(k)] = \lambda(\theta)$$

exists and is analytical on Θ. For $\theta_0 \in \Theta$, the domain of convergence is $D(\theta_0)$. Moreover, the n^{th} derivative of the Lyapunov exponent is given by

$$\begin{aligned}\frac{d^n}{d\theta^n}\lambda = &E\left[\left(A(1) \otimes \bigotimes_{k=-\eta}^{0} A(k)\right)^{(n)} \otimes \bigotimes_{k=-\eta^{(n)}}^{-\eta^{(n-1)}-1} A(k) \otimes x_0\right] \\ &- E\left[\left(\bigotimes_{k=-\eta}^{0} A(k)\right)^{(n)} \otimes \bigotimes_{k=-\eta^{(n)}}^{-\eta^{(n-1)}-1} A(k) \otimes x_0\right].\end{aligned}$$

Proof: Theorem 4 implies

$$\begin{aligned}\lambda = &E\left[A(1) \otimes \bigotimes_{k=-\eta}^{0} A(k) \otimes x_0 - \bigotimes_{k=-\eta}^{0} A(k) \otimes x_0\right] \\ = &E\left[A(1) \otimes \bigotimes_{k=-\infty}^{0} A_{\tilde{a}}(k) \otimes x_0 - \bigotimes_{k=-\infty}^{0} A_{\tilde{a}}(k) \otimes x_0\right].\end{aligned}$$

Hence, for the proof it suffices to show the analyticity of

$$\begin{aligned}E\left[\bigotimes_{k=-\eta}^{0} A(k) \otimes x_0\right] &= E\left[\bigotimes_{k=-\infty}^{0} A_{\tilde{a}}(k) \otimes x_0\right] \\ &= E\left[\lim_{m\to\infty} \bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0\right] \\ &= \lim_{m\to\infty} E\left[\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0\right],\end{aligned}$$

where the last equality follows from the monotone convergence theorem. In accordance with Theorem 7, the finite products on the right-hand side of the above formula are analytical and we obtain

$$\begin{aligned}E\left[\bigotimes_{k=-\eta}^{0} A(k) \otimes x_0\right] &= \lim_{m\to\infty} \sum_{n=0}^{\infty} \frac{d^n}{d\theta^n} E\left[\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k) \otimes x_0\right] \frac{(\theta-\theta_0)^n}{n!} \\ &\overset{\text{Th. 5}}{=} \lim_{m\to\infty} \sum_{n=0}^{\infty} E\left[\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k)\right)^{(n)} \otimes x_0\right] \frac{(\theta-\theta_0)^n}{n!}\,. \qquad (18)\end{aligned}$$

We now show that we may interchange the order of limit and summation. As a first step, we calculate the limit of

$$E\left[\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k)\right)^{(n)} \otimes x_0\right]$$

for m towards ∞. Take g as the projection of $\hat{\mathbb{R}}_\epsilon^J$ onto $\mathbb{R}^J$, then $g \in C_1(\hat{\mathbb{R}}_\epsilon^J)$. Moreover, g is non–expansive and it follows from Lemma 2 that for all m

$$\underbrace{\left|\left(\bigotimes_{k=-m}^{0} A(k)\right)^{(n)} \otimes x_0\right|}_{=:h(m)} \leq B(\eta, n)\,. \tag{19}$$

Furthermore, the sample limit of $h(m)$ exists, see Equation (17), and is bounded by $B(\eta, n)$ (for a proof use Lemma 1). Under the conditions of the theorem, $B(\eta, n)$ has a finite mean and we may apply the dominated convergence theorem. This yields

$$\begin{aligned}\lim_{m\to\infty} E\left[\left(\bigotimes_{k=-m}^{0} A(k)\right)^{(n)} \otimes x_0\right] &= E\left[\lim_{m\to\infty}\left(\bigotimes_{k=-m}^{0} A(k)\right)^{(n)} \otimes x_0\right] \\ &\overset{(17)}{=} E\left[\left(\bigotimes_{k=-\eta}^{0} A(k)\right)^{(n)} \otimes \bigotimes_{k=-\eta^{(n)}}^{\eta^{(n-1)}-1} A_{\tilde{a}}(k) \otimes x_0\right]\,. \end{aligned} \tag{20}$$

We now show that we may interchange limit and summation in (18). By Inequality (19) it holds for all m that

$$\left|E\left[\left(\bigotimes_{k=-m}^{0} A(k)\right)^{(n)} \otimes x_0\right]\right| \leq E[B(\eta, n)]$$

and from (20) we get that the limit on the left–hand side of the above inequality exists. Under the conditions of the theorem it holds that

$$\sum_{n=0}^{\infty} E[B(\eta, n)]\,\frac{(\theta - \theta_0)^n}{n!} < \infty\,.$$

Hence, we may apply the dominated convergence theorem to obtain the following

$$\begin{aligned}&\lim_{m\to\infty}\sum_{n=0}^{\infty} E\left[\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k)\right)^{(n)} \otimes x_0\right]\frac{(\theta-\theta_0)^n}{n!} \\ &\quad= \sum_{n=0}^{\infty}\lim_{m\to\infty} E\left[\left(\bigotimes_{k=-m}^{0} A_{\tilde{a}}(k)\right)^{(n)} \otimes x_0\right]\frac{(\theta-\theta_0)^n}{n!} \\ &\quad\overset{(20)}{=} \sum_{n=0}^{\infty} E\left[\left(\bigotimes_{k=-\eta}^{0} A_{\tilde{a}}(k)\right)^{(n)} \otimes \bigotimes_{k=-\eta^{(n)}}^{\eta^{(n-1)}-1} A_{\tilde{a}}(k) \otimes x_0\right]\frac{(\theta-\theta_0)^n}{n!}\,.\end{aligned}$$

We have calculated the right–hand side of (18), which concludes the proof of the theorem. □

In the following subsection we illustrate the above theorem with a simple example.

8.1 The Bernoulli Scheme

We revisit the situation in Example 2. We have already shown that

$$|E[B(\eta,n)]| \le c^2(2c)^{n-1}\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right)^{n+1}(||D_1||+||D_2||),$$

for $\hat{\theta} \in [0,1)$. Choose $\theta_0 \in [0,\hat{\theta})$ and take $D_{\theta_0} = \{\theta \le \hat{\theta} : |\theta - \theta_0| \le \frac{1}{2c}\} \cap [0,1)$. This implies that

$$(2c)^{n-1}(\theta-\theta_0)^n \le 1$$

and we obtain

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{(\theta-\theta_0)^n}{n!}|E[B(\eta,n)]| &\le c^2\sum_{n=0}^{\infty}\frac{(2c)^{n-1}(\theta-\theta_0)^n}{n!}\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right)^{n+1}(||D_1||+||D_2||)\\ &\le c^2\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right)^{n+1}(||D_1||+||D_2||)\\ &= c^2\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right)(||D_1||+||D_2||)e^{\frac{\hat{\theta}}{1-\hat{\theta}}} < \infty .\end{aligned}$$

This means that the Lyapunov exponent of the Bernoulli scheme can be developed into a Taylor series at any point θ_0 on $[0,1)$ and that the domain of convergence of the Taylor series of the Lyapunov exponent developed at $\theta_0 \in [0,1)$ is at least $\{\theta : |\theta-\theta_0| \le \frac{1}{2c}\} \cap [0,1)$. Hence, the Lyapunov exponent of the Bernoulli experiment can be extended to a complex function that is analytical on the strip of width $1/2c$ around the interval $[0,1)$.

Conclusion

We developed the Lyapunov exponent of (max,+)-linear stochastic systems into a Taylor series. Moreover, we established lower bounds for the radius of convergence of the Taylor series. The two main ingredients were: (1) the radius of convergence of the Taylor series of the matrix $A(k)$, and (2) coupling time of the system. We applied our results to a simple system and showed that we could improve the results known so far on the domain of analyticity of the Lyapunov exponent.

References

[1] Baccelli, F., Cohen, G., Olsder,G., and Quadrat, J.-P. *Synchronization and Linearity*. John Wiley and Sons, New-York, 1992.

[2] Baccelli, F. and Hong, D. Analytic expansions of (max,+) Lyapunov exponents. to appear in *Annals of Applied Probability*.

[3] Baccelli, F. and Hong, D. Analytic expansions of random non-expansive maps. Technical report no. 3558, INRIA Sophia Antipolis, 1998.

[4] Baccelli, F. and Mairesse, J. *Ergodic theory of stochastic operators and discrete event networks*, page editor J. Gunawardena. Idempotency. Cambridge University Press, 1997.

[5] R.A. Cunninghame-Green. *Minimax algebra*. vol. 166 of Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin, 1979.

[6] B. Heidergott. A differential calculus for random matrices with applications to (max,+)-linear stochastic systems. Technical report 98-10, Delft University of Technology, 1998 (submitted).

[7] B. Heidergott. Analyticity of transient(max,+)-linear stochastic systems. EURANDOM report 99-030, 1999 (submitted).

[8] B. Heidergott. A characterization for (max,+)-linear queueing systems. QUESTA (to appear).

[9] N. Krivulin. A max-algebra approach to modeling and simulation of tandem queueing systems. *Mathl. Comput. Modelling*, 22:25–37, 1995.

[10] J. Mairesse. Products of irreducible random matrices in the (max,+) algebra. *Advances of Applied Probability*, 29:444–477, 1997.

[11] G. Pflug. *Optimization of Stochastic Models*. Kluwer Academic, Boston, 1996.

Appendix

The coefficients of the Taylor series are combinatorially complex and can be represented in various ways; see for example the representations in [2]. Our analysis leads to yet another way of representing the coefficients of the Taylor series and in what follows we illustrate for the first-order derivative of the Lyapunov exponent of the Bernoulli system that the expression in Theorem 7 can indeed be algebraically manipulated in order to resemble the coefficients in Theorem 1 in [2].

We calculate the first-order derivative of λ at $\theta_0 = 0$. This implies that $A(k) = D_1$ for all k. Furthermore, the coupling time of D_2 equals c and since at θ_0 the sequence $\{A(k)\}$ is deterministic, we obtain $\eta = c - 1$. In accordance with Theorem 7, we obtain

$$\frac{d}{d\theta}\lambda = E\left[\sum_{l\in\mathcal{L}(c,1)}\sum_{i\in\mathcal{I}(c,l)}\left(\bigotimes_{k=-c+1}^{1} A(k)\right)^{(l,i)} \otimes \bigotimes_{k=-2c+1}^{-c} A(k)\otimes x_0\right]$$
$$- E\left[\sum_{l\in\mathcal{L}(c-1,1)}\sum_{i\in\mathcal{I}(c-1,l)}\left(\bigotimes_{k=-c+1}^{0} A(k)\right)^{(l,i)} \otimes \bigotimes_{k=-2c+1}^{-c} A(k)\otimes x_0\right]$$

The first-order weak derivative of $A(k)$ is $(1, D_1, D_2)$ and all higher-order weak derivatives are not significant, see Example 4. Furthermore, let $l \in \mathcal{L}(m,1)$, then l is a vector of length m that has one component, say k^*, equal to one and all others equal to zero, and $\mathcal{I}(l,1)$ contains only one element i, with $i_k = 0$ for $k \neq k^*$ and $i_{k^*} = +1$. From this it follows that

$$\left(\bigotimes_{k=-c+1}^{0} A(k)\right)^{(l,i)} = \left(1,\ \bigotimes_{k=-k^*+1}^{0} D_2 \otimes D_1 \otimes \bigotimes_{k=-c+1}^{-k^*-1} D_2,\ D_2^c\right).$$

In accordance with Theorem 7, we obtain

$$\frac{d}{d\theta}\lambda = \sum_{j=0}^{c} D_2^{c-j}\otimes D_1\otimes D_2^j\otimes D_2^c\otimes x_0 - \sum_{j=0}^{c} D_2^{c+1}\otimes D_2^c\otimes x_0$$
$$- \sum_{j=0}^{c-1} D_2^{c-1-j}\otimes D_1\otimes D_2^j\otimes D_2^c\otimes x_0 + \sum_{j=0}^{c-1} D_2^c\otimes D_2^c\otimes x_0 .$$

We set $X_0 = D_2^c \otimes x_0$ and, since c is the coupling time of D_2, it follows that X_0 is an eigenvector of D_2. In accordance with (6), the D_2 terms cancel out and we obtain

$$\frac{d}{d\theta}\lambda = \sum_{j=0}^{c}\left(D_2^{c-j}\otimes D_1\otimes X_0 - D_2^{c+1-j}\otimes X_0\right)$$
$$- \sum_{j=0}^{c-1}\left(D_2^{c-1-j}\otimes D_1\otimes X_0 - D_2^{c-j}\otimes X_0\right).$$

Rearranging terms yields

$$
\begin{aligned}
\frac{d}{d\theta}\lambda =& D_2^c \otimes D_1 \otimes X_0 + \sum_{j=1}^{c} D_2^{c-j} \otimes D_1 \otimes X_0 - \sum_{j=0}^{c-1} D_2^{c-1-j} \otimes D_1 \otimes X_0 \\
& - \sum_{j=0}^{c} D_2^{c+1-j} \otimes X_0 + \sum_{j=0}^{c-1} D_2^{c-j} \otimes X_0 \\
=& D_2^c \otimes D_1 \otimes X_0 - \sum_{j=0}^{c} D_2^{c+1-j} \otimes X_0 + \sum_{j=0}^{c} D_2^{c-j} \otimes X_0 \,. \qquad (21)
\end{aligned}
$$

Recall that

$$\lambda(0) = D_2 \otimes D_2^m \otimes X_0 - D_2^m \otimes X_0$$

for all $m \geq 0$. This implies for the summations in (21)

$$
\begin{aligned}
& - \sum_{j=0}^{c} D_2^{c+1-j} \otimes X_0 + \sum_{j=0}^{c-1} D_2^{c-j} \otimes X_0 \\
& \quad = - D_2 \otimes X_0 - \sum_{j=0}^{c-1} D_2^{c+1-j} \otimes X_0 + \sum_{j=0}^{c-1} D_2^{c-j} \otimes X_0 \\
& \quad = - D_2 \otimes X_0 - c\lambda(0) \,.
\end{aligned}
$$

Inserting the above equality into (21) we obtain

$$\frac{d}{d\theta}\lambda = D_2^c \otimes D_1 \otimes X_0 - D_2 \otimes X_0 - c\lambda(0) \,.$$

Using the fact that $D_2 \otimes X_0 = -\lambda(0) - X_0$, we obtain

$$\frac{d}{d\theta}\lambda = D_2^c \otimes D_1 \otimes X_0 - X_0 - (c+1)\lambda(0) \,,$$

which is the explicit form of the first-order derivative of the Lyapunov exponents at $\theta_0 = 0$ as given in [2].