Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Series Expansions of Generalized Matrix Products

Haralambie Leahu and Bernd Heidergott

Abstract— **We consider generalized products of random matrices. They arise in discrete event systems (DES), such as queueing networks or stochastic Petri nets, where they are used to express the state transition dynamic. Instances of such DES are those whose state dynamic can be modelled through a matrix-vector multiplication in conventional, max-plus and min-plus algebra. We will present a Taylor series approach to numerical evaluation of finite horizon performance characteristics of systems modelled by generalized matrix products. The cornerstone of our analysis is the introduction of a differential calculus, based on the concept of *weak derivative of a random matrix*. We illustrate our results with a couple of numerical computations performed on a classical DES example.**

Acknowledgements: This research is supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

I. Introduction

In this paper we study discrete event systems (DES) that can be modelled by linear recurrence of the type

$$x(k+1) = A(k+1) \odot x(k), \quad k \geq 0\,, \tag{1}$$

where $A(k)$ is a square matrix of dimension m modelling the kth state transition, $x(k)$ is vector of dimension m denoting the kth state of the system (with initial state $x(0) = x_0$) and $\odot$ denotes the matrix-vector product defined in a suitable algebra, such as, max-plus, min-plus, min-max or conventional algebra.

In the case that system dynamic is time independent, i.e., $A(k) = A$ for $k \geq 1$, powerful tools exist for evaluating the system. See, for example, [4] and [14]. However, DES are typically prone to random influences. For example, processing times of a machine in a manufacturing model may vary due to variability in the processed material, running times of trains in a train network may be increased (delays) due to weather conditions. Unfortunately, in the presence of stochastic noise, no efficient algorithm for computing characteristics of the state vector (such as expected value) exists and explicit solutions only in special cases.

To approximately compute performance characteristics of DES with random noise whose state-dynamic follow (1), Taylor series expansions have been proposed. In particular, Taylor series expansions of max-plus linear DES systems are an area of active research. When it comes to queueing networks, the approach dominant in the literature is that of light traffic approximations for stationary waiting times in open max-plus linear queuing systems with Poisson arrival stream. More specifically, let W_i denote the i^{th} component of the vector of stationary waiting times in an open queueing system with Poisson-λ-arrival stream. The pioneering paper on light traffic expansions for $\mathbb{E}[W_i]$ is [6], where sufficient conditions for the existence of the light traffic approximation for $\mathbb{E}[W_i]$ with respect to λ are established and the (first) elements of the Taylor series are computed analytically. These results have been extended in [8] to $\mathbb{E}[f(W_i)]$, where f belongs to the class of performance measures $\mathcal{F}$, where $h \in \mathcal{F}$ if $h : [0,\infty) \to [0,\infty)$ and $h(x) \leq c\, x^{\nu}$ for $x \geq 0$ and $\nu \in \mathbb{N}$. In [1] expansions are obtained for $\mathbb{E}[f(W_i, W_j)]$, for $f : [0,\infty)^2 \to [0,\infty)$ with $f(x,y) \leq c x^{\nu_1} x^{\nu_2}$ for $x, y \geq 0$ and $\nu_1, \nu_2 \in \mathbb{N}$. In [2], [3], explicit expressions are given for the moments, Laplace transform and tail probability of the waiting time of the k^{th} customer. Furthermore, starting with exact expressions for transient waiting times, exact expressions for moments, Laplace transform and tail probability of stationary waiting times in a certain class of max-plus linear systems with deterministic service times are computed. Taylor series expansions have also been successfully applied in [11] to control of max-plus linear DES; for an application based on the concept of variability expansion [13].

In this paper we report on results on Taylor series approximations for general stochastic "linear" systems. Here "linearity" means that the matrix-vector product in (1) is given by a generalized matrix-product in the sense of Cohen [9] (to be presented shortly). The underlying theory is based on the concept of weak convergence of probability measures. However, due to lack of space, proofs and details on the mathematical analysis are postponed to the full-length version of the paper and we focus here on presenting basic ideas and numerical results. The paper is organized as follows. Section II introduces the general framework for modelling linear DES. The concept of weak differentiation is discussed in Section III. Our general results on Taylor series expansions are presented in Section IV. Numerical examples illustrating our approach are given in Section V. We conclude by identifying topics of further research.

II. General Multiplicative Monoids

Let S_0 denote a non-empty set endowed with two binary associative operators f and h, such that f is commutative and h distributes over f. A *generalized matrix product*, in symbols, $\otimes_{f \cdot h}$ (denoted shortly by $\otimes$ when no confusion occurs), is defined on the set $M_m(S_0)$ of m-dimensional square matrices with elements in S_0, as follows:

$$[A \otimes B]_{ij} = [Af \cdot hB]_{ij} \stackrel{\text{def}}{=} (A_{i1} h B_{1j}) f \cdots f (A_{im} h B_{mj}),$$

for each pair (i,j) with $1 \leq i, j \leq m$ and each $A = [A_{ij}]$, $B = [B_{ij}]$. For instance, one can recover the classical matrix multiplication out of this definition, by setting $f = +$ and $h = \times$. Note that, like in conventional algebra, this product rule can be extended to general (e.g., non-square) matrices, of appropriate size (e.g., matrix-vector multiplication). Nevertheless, due to space limitations, we focus on square matrices. This matrix product is associative and consequently, if we set $S \stackrel{\text{def}}{=} M_m(S_0)$, we have that $(S, \otimes)$ is a *monoid*. For a more detailed background see [9], [16], [10]. Furthermore, if both binary operators f and h admit neutral elements, say 1_f and 1_h, respectively, such that 1_f is *absorbing* for h (i.e. $s\ h\ 1_f = 1_f$ for each $s \in S_0$), then the above introduced monoid admits a neutral element too. Specifically, we shall denote the matrix having all diagonal elements equal to 1_h and the rest of them equal to 1_f, by $E_{(f \cdot h)}$ or shortly by E when no confusion exists. Then it is immediate that E is neutral element for " $\otimes$ ". A zero element $\mathcal{E}_{(f \cdot h)}$ can be introduced as well, where $[\mathcal{E}_{(f \cdot h)}]_{ij} = 1_f$ for each pair of indices (i, j). Then, we have $A \otimes \mathcal{E}_{(f \cdot h)} = \mathcal{E}_{(f \cdot h)} \otimes A = \mathcal{E}_{(f \cdot h)}$ for each $A \in S$. We omit the subscript "$(f \cdot h)$" when no confusion occurs.

Assume now that S_0 is endowed with a metric δ. Then, the mapping $d : S \times S \to \mathbb{R}_+$ defined through:

$$d(A, B) = \max_{1 \le i,j \le m} \delta(A_{ij}, B_{ij})$$

for each $A, B \in S$, is a metric on S and (S, d) is a separable and complete metric space, provided that (S_0, δ) is.

By *upper bound*, we mean a continuous application (for instance a norm, seminorm) $\| \cdot \| : S \to \mathbb{R}$, satisfying:

1) $\|s\| \ge 0, \quad \forall s \in S$
2) either $\|s \otimes r\| \le \|s\| + \|r\|, \forall r, s \in S$ or $\|s \otimes r\| \le \|s\| \cdot \|r\|, \forall r, s \in S$.

The upper bound will be called *additive* or *multiplicative* according to the operation in the right-hand side of the inequality.

Example 1: Classical examples of monoids equipped with an upper bound are given in the following.

1) A first example (the conventional algebra setting) was already presented in the paper. One can construct a monoid by choosing: $S = [0, \infty)^{m^2}$, the positive cone in the m^2-dimensional euclidean space. The maximum norm (i.e., $\|A\|_\infty \stackrel{\text{def}}{=} \max_{1 \le i,j \le m} A_{ij}$) acts as an upper-bound in this case.
2) $f \stackrel{\text{def}}{=} \max$ and $h \stackrel{\text{def}}{=} +$ (i.e. max-plus multiplication). We have $S_0 \stackrel{\text{def}}{=} \{\varepsilon \stackrel{\text{def}}{=} -\infty\} \cup [0, \infty)$. An appropriate metric on S_0 is given by
$$\delta(x, y) \stackrel{\text{def}}{=} \begin{cases} |e^x - e^y|, & \text{for } x, y \ge 0; \\ e^x, & \text{for } x \ge 0, y = \varepsilon. \end{cases}$$
The following upper-bound can be introduced on S. Set for each $A \in S$,
$$\|A\| \stackrel{\text{def}}{=} \begin{cases} 0, \text{if } A_{i,j} = \varepsilon, \forall 1 \le i, j \le m; \\ \max\{A_{ij} \ne \varepsilon; 1 \le i, j \le m\}, \text{otherwise.} \end{cases}$$
We denote the neutral element of "+" by $e \stackrel{\text{def}}{=} 0$.
3) $f \stackrel{\text{def}}{=} \min$ and $h \stackrel{\text{def}}{=} +$. In this case, $S_0 \stackrel{\text{def}}{=} [0, \infty) \cup \{\infty\}$. A suitable metric on S_0 is given by
$$\delta(x, y) \stackrel{\text{def}}{=} \begin{cases} |e^x - e^y|, & \text{for } x, y < \infty; \\ e^x, & \text{for } x < \infty, y = \infty. \end{cases}$$
The mapping
$$\|A\| \stackrel{\text{def}}{=} \begin{cases} 0, \text{if } A_{i,j} = \infty, \forall 1 \le i, j \le m; \\ \max\{A_{ij} < \infty; 1 \le i, j \le m\}, \text{otherwise.} \end{cases}$$
denotes an upper bound on S.
4) $f \stackrel{\text{def}}{=} \max$ and $h \stackrel{\text{def}}{=} \times$. Set $S_0 \stackrel{\text{def}}{=} \{\varepsilon\} \cup [0, \infty)$. We choose δ and $\| \cdot \|$ just like in (2) above.
5) $f \stackrel{\text{def}}{=} \min$ and $h \stackrel{\text{def}}{=} \times$. We choose $S_0 \stackrel{\text{def}}{=} [0, \infty) \cup \{\infty\}$, whereas δ and $\| \cdot \|$ are exactly the same as in (3) above.

For illustrative purposes we will focus in what follows on max-plus algebra. Consider the following situation:

Example 2: (Baccelli & Hong, [5]) Consider a cyclic tandem queueing network consisting of a single server and a multi server, each with deterministic service time. Service times at the single-server station equal σ, whereas service times at the multi-server station equal σ'. Three customers circulate in the network. Initially, one customer is in service at station 1, the single server, one customer is in service at station 2, the multi-server, and the third customer is just about to enter station 2. The time evolution of this network is described by a max-plus linear sequence $x(k) = (x_1(k), \ldots, x_4(k))^t$, where $x_1(k)$ is the k^{th} beginning of service at the single–server station and $x_2(k)$ is the k^{th} departure epoch at the single-server station; $x_3(k)$ is the k^{th} beginning of service at the multi-server station and $x_4(k)$ is the k^{th} departure epoch from the multi-server station. The system then follows

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

where

$$D = \begin{pmatrix} \sigma & \varepsilon & \sigma' & \varepsilon \\ \sigma & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & e & \varepsilon & e \\ \varepsilon & \varepsilon & \sigma' & \varepsilon \end{pmatrix} ,$$

with $x(0) = x_0 \stackrel{\text{def}}{=} (0, 0, 0, 0)^t$. Figure 1 shows the initial state of this system.

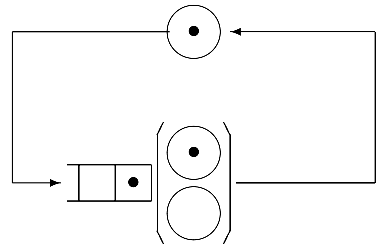

Fig. 1. The initial state of the multi-server system (three customers).

Consider the cyclic tandem network again, but one of the servers of the multi-server station has broken down. The system is thus a tandem network with two single server stations. Initially one customer is in service at station 1, one customer is in service at station 2, and the third customer is waiting at station 2 for service. This system follows

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

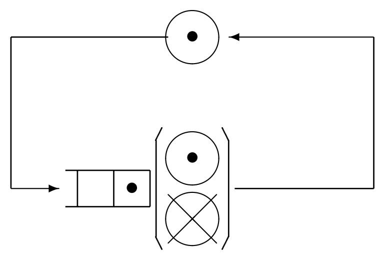

Fig. 2. The initial state of the multi-server system with breakdown (three customers).

where

$$P = \begin{pmatrix} \sigma & \varepsilon & \sigma' & \varepsilon \\ \sigma & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & e & \sigma' & \varepsilon \\ \varepsilon & \varepsilon & \sigma' & \varepsilon \end{pmatrix},$$

with $x(0) = x_0$. Figure 2 shows the initial state of the system with breakdown.

Assume that whenever a customer enters station 2, the second server of the multi server station breaks down with probability θ. Let $A_\theta(k)$ have distribution

$$\mathbb{P}(\, A_\theta(k) = D\,) = 1 - \theta$$

and

$$\mathbb{P}(\, A_\theta(k) = P\,) = \theta\,,$$

then

$$x_\theta(k+1) = A_\theta(k+1) \otimes x_\theta(k); k \geq 0,$$

describes the time evolution of the system with breakdowns. That the above recurrence relation indeed models the sample path dynamic of the system with breakdowns is not obvious and a proof can be found in [5].

III. Weak Differentiation of Random Objects

Let (S, d_S) be a separable metric space and let $\mathcal{M} = \mathcal{M}(S)$ be the set of all finite signed (real valued) regular measures on the measurable space $(S, \mathcal{B})$, where $\mathcal{B}$ denotes the Borel field of S (i.e. the "smallest" σ-algebra which contains the open sets), see, for example, [15] for details. Denote by $\mathcal{M}_1(S)$ the subspace of regular probability measures on S (the elements of $\mathcal{M}(S)$ having the total mass equal to 1).

Denote by $\mathcal{C}_b(S)$ the space of bounded continuous real valued functions on S and let $\mathcal{D}(S)$ be a set of real valued functions on S. For technical reasons we assume that $\mathcal{C}_b(S) \subset \mathcal{D}(S)$. Such a set will be called a *set of performance (test) functions*.

Finally, let $\Theta \subset \mathbb{R}$ be a bounded interval and assume that $\{\mu_\theta; \theta \in \Theta\} \subset \mathcal{M}_1(S)$ is a family of probability measures (it can be also thought as a measure-valued mapping on Θ).

Definition 1: Let $\theta \in \Theta$. We say that μ_θ, is weakly $\mathcal{D}(S)$-differentiable, if there exists a real valued measure $\mu'_\theta \in \mathcal{M}$, such that for all $g \in \mathcal{D}(S)$ it holds:

$$\lim_{\substack{\Delta \to 0 \\ \theta+\Delta \in \Theta}} \frac{1}{\Delta} \left(\int g d\mu_{\theta+\Delta} - \int g d\mu_\theta \right) = \int g d\mu'_\theta. \quad (2)$$

Moreover, if μ_θ is $\mathcal{D}(S)$-differentiable, then any triple $(c_\theta, \mu_\theta^+, \mu_\theta^-)$ with $c_\theta \in \mathbb{R}$ and $\mu_\theta^+, \mu_\theta^-$ probability measures, satisfying for all $g \in \mathcal{D}(S)$:

$$\int g d\mu'_\theta = c_\theta \left(\int g d\mu_\theta^+ - \int g d\mu_\theta^- \right), \quad (3)$$

will be called a $\mathcal{D}(S)$-derivative of μ_θ.

Remark 1: Note that, if exists, an weak derivative is not unique. If the left-hand side limit in (2) exist, then an instance of the weak derivative can always be found via the Hahn-Jordan decomposition Theorem [15]. If the above limit is 0 for all $g \in \mathcal{D}(S)$, we say that the weak derivative is not significant and for technical convenience, we choose the representation $\mu'_\theta \stackrel{\text{def}}{=} (1, \mu_\theta, \mu_\theta)$.

Higher-order weak derivatives are given in an obvious way. For details on weak differentiation see [17], [12].

Let $(S, \otimes)$ be a monoid, as in the previous section and assume that a metric d_S exists on S. Then (S, d) is a separable metric space, hence a topological space. Let $\mathcal{B} \stackrel{\text{def}}{=} \sigma(S, d)$ denote the canonical Borel σ-algebra on S. Thus, $(S, \mathcal{B})$ becomes a measurable space on which we can define random objects. Finally, with measurability and the upper bound at hand we can define the space $\mathcal{C}^p(S)$, as $(p \geq 1)$:

$$\mathcal{C}^p(S) \stackrel{\text{def}}{=} \{g : S \to \mathbb{R}; \exists a, b > 0, s.t. |g(s)| \leq a + b\|s\|^p\}\,.$$

From now on, $\mathcal{C}^p(S)$ will be our test functions set and whenever we refer to weak differentiability we mean weak $\mathcal{C}^p(S)$-differentiability.

Let now $\{A_\theta\}_\theta$ be a family of random elements in S, defined on some common underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with distribution $\mu_\theta(\cdot) \stackrel{\text{def}}{=} \mathbb{P}(A_\theta \in \cdot) \in \mathcal{M}_1(S)$, for $\theta \in \Theta$. We call A_θ weakly differentiable if its distribution μ_θ is. According to Definition 1, if the triple $(c_\theta, \mu_\theta^+, \mu_\theta^-)$ is a weak derivative for μ_θ then, assuming that A_θ^+ and A_θ^- are two random elements on S having distributions μ_θ^+ and μ_θ^- respectively, we have the means to compute $\frac{d}{d\theta}\mathbb{E}[g(A_\theta)]$, for $g \in \mathcal{C}^p(S)$. Indeed, according to (3) we have:

$$\frac{d}{d\theta}\mathbb{E}[g(A_\theta)] = c_\theta^A \cdot \mathbb{E}\big[g(A_\theta^+) - g(A_\theta^-)\big], \quad (4)$$

with $c_\theta^A = c_\theta$. Consequently, we call the triple $(c_\theta^A, A_\theta^+, A_\theta^-)$ a weak derivative of A_θ and denote it by A'_θ, if (4) holds for all $g \in \mathcal{C}^p(S)$. If the left-hand side in (4) equals to zero for each $g \in \mathcal{C}^p(S)$, then we set $A'_\theta \stackrel{\text{def}}{=} (0, \mathcal{E}, \mathcal{E})$.Higher-order weak derivatives $A_\theta^{(n)}$ are defined similarly, provided that they exist for $n \geq 1$.

Example 3: We consider the so-called Bernoulli example (i.e. when A_θ has a Bernoulli-type distribution) a version of which we presented in Example 2. Assume that A_θ is a random variable on S, Bernoulli distributed, on the set $\{x_1, x_2\} \subset S$, with parameter $\theta \in [0, 1]$, namely

$$\mathbb{P}(A_\theta \in \cdot) = (1-\theta)\delta_{x_1}(\cdot) + \theta\delta_{x_2}(\cdot),$$

where δ_x denotes the Dirac distribution in point x. An weak derivative of A_θ is given by: $A'_\theta = (1, x_2, x_1)$ (note that the weak derivative does not depend on θ). Furthermore, it is immediate that all higher-order derivatives $A_\theta^{(n)}$ for $n \geq 2$ are not significant, so we set $A_\theta^{(n)} = (0, x_1, x_1)$.

The next result is crucial for the sequel. It provides the means for differentiating expressions like $\mathbb{E}[g(A_\theta \otimes B_\theta)]$:

Theorem 1: Let A_θ, B_θ be random elements in S, stochastically independent. If A_θ, B_θ are $\mathcal{C}^p(S)$-differentiable, having weak derivatives $(c_\theta^A, A_\theta^+, A_\theta^-)$ and $(c_\theta^B, B_\theta^+, B_\theta^-)$, respectively, then for all $g \in \mathcal{C}^p(S)$ it holds that

$$\begin{aligned}\frac{d}{d\theta}\mathbb{E}[g(A_\theta \otimes B_\theta)] =& c_\theta^A \mathbb{E}\left[g(A_\theta^+ \otimes B_\theta) - g(A_\theta^- \otimes B_\theta)\right] \\ &+ c_\theta^B \mathbb{E}\left[g(A_\theta \otimes B_\theta^+) - g(A_\theta \otimes B_\theta^-)\right]\end{aligned} \tag{5}$$

In order to develop a differential calculus, analogue to the classical one we need to introduce some new algebraical objects and operations. The main reason behind this technical work is the following. The weak derivative of a random object is not anymore a random object, but a triple. Thus it is not possible for instance to "multiply" a random matrix by its derivative. The idea is to embed the space S into a richer one, such that A_θ and A'_θ are objects of the same kind. We give here a sketch. For details see [12].

Let S_{ext} denote the set of all finite sequences of triples (c, A, B), with $c \in \mathbb{R}_+$ and $A, B \in S$. A generic element of S_{ext} is therefore given by

$$\alpha = \big((c_1, A_1, B_1), (c_2, A_2, B_2), \ldots, (c_n, A_n, B_n)\big), n \geq 1.$$

If $n = 1$ we say that α is elementary. On S_{ext} we introduce the operation "+" as the concatenation of sequences. For instance, if $\alpha_i \stackrel{\text{def}}{=} (c_i, A_i, B_i)$ then the above equality reads

$$\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_n.$$

Note that the weak derivative of a random matrix is elementary in S_{ext}. Scalar multiplication is defined as:

$$r \cdot (c, A, B) \stackrel{\text{def}}{=} (rc, A, B),$$

for elementary elements, and extend it to general elements via "+" linearity. We embed S into S_{ext} via the mapping:

$$A \mapsto (1, A, \mathcal{E}).$$

The real-valued function $g : S \to \mathbb{R}$ is extended to the function $g^\varphi : S_{ext} \to \mathbb{R}$ defined through:

$$g^\varphi(c, A, B) \stackrel{\text{def}}{=} c\big(g(A) - g(B)\big),$$

for elementary elements and extend it via "+" linearity to general elements. Note that $g^\varphi(1, A, \mathcal{E}) = g(A)$ and

$$g^\varphi\left(\sum_{i=1}^n r_i \alpha_i\right) = \sum_{i=1}^n r_i g^\varphi(\alpha_i).$$

Moreover, (4) can be re-written as:

$$\frac{d}{d\theta}\mathbb{E}[g(A_\theta)] = \mathbb{E}[g^\varphi(A'_\theta)].$$

The space S_{ext} can be endowed with a metric (thus a topology) so that it becomes a measurable metric space on which one can define random objects.

However, S_{ext} has a very poor algebraic structure (for instance "+" fails to be commutative). On the other hand, componentwise equalities are not for practical interest. That is why we introduce the concept of *weak equality*. Namely, for α, β random elements in S_{ext}, we say that they are weakly equal (and write "$\alpha \equiv \beta$") if:

$$\mathbb{E}[g^\varphi(\alpha)] = \mathbb{E}[g^\varphi(\beta)], \forall g \in \mathcal{C}^p(S).$$

Of course, if we think about the elements of S_{ext} as random elements having Dirac distribution, then the weak equality relation describes an equivalence relation on S_{ext}. Reformulating now the classical algebraical properties (i.e. commutativity, associativity, distributivity) in terms of weak equalities, it can be checked that although "+" is not commutative, it is weakly commutative. Furthermore, the weak derivative is unique in the weak sense.

Finally, introduce the multiplication "$\otimes'$" on S_{ext}: for elementary $\alpha_1 = (c_1, A_1, B_1)$, $\alpha_2 = (c_2, A_2, B_2)$ set:

$$\begin{aligned}&\alpha_1 \otimes' \alpha_2 \stackrel{\text{def}}{=} \\ &c_1 c_2 \cdot \left((1, A_1 \otimes A_2, A_1 \otimes B_2) + (1, B_1 \otimes B_2, B_1 \otimes A_2)\right).\end{aligned}$$

and we extend it to general elements s.t. $\otimes'$ is linear and distributes over "+". It can be checked that the above definition does not depend on the representatives and:

$$(1, A_1, \mathcal{E}) \otimes' (1, A_2, \mathcal{E}) \equiv (1, A_1 \otimes A_2, \mathcal{E}),$$

so that $\otimes'$ extends $\otimes$ to S_{ext}.

We go back now to random matrices and differential calculus. Re-writing the right-hand term in (5), we obtain:

$$\frac{d}{d\theta}\mathbb{E}[g(A_\theta \otimes B_\theta)] = \mathbb{E}[g^\varphi(A'_\theta \otimes' B_\theta + A_\theta \otimes' B'_\theta)],$$

which in terms of weak equalities reads:

$$(A_\theta \otimes B_\theta)' \equiv A'_\theta \otimes' B_\theta + A_\theta \otimes' B'_\theta.$$

This last (weak) equality represents the basis of our weak differential calculus. Like in the classical analysis, it leads to the more general formula for nth order weak derivatives (Leibniz-Newton):

$$(A_\theta \otimes B_\theta)^{(n)} \equiv \sum_{k=0}^n \binom{n}{k} \cdot A_\theta^{(n-k)} \otimes' B_\theta^{(k)}.$$

IV. Taylor Series

Elaborating on our above results on weak differential calculus we obtain the following differentiation rule for finite products of random matrices:

Assume that $A_1, \ldots, A_k$; $k \geq 1$ are independent random matrices on S weakly differentiable w.r.t. θ. Then their product is still weakly differentiable and it holds:

$$\left(\bigotimes_{i=1}^k A_i\right)^{(n)} \equiv \sum_{j \in \mathcal{J}(k,n)} \frac{n!}{j_1! \cdots j_k!} \cdot A_1^{(j_1)} \otimes' \ldots \otimes' A_k^{(j_k)},$$

where $\mathcal{J}(k, n) \stackrel{\text{def}}{=} \{j = (j_1, \ldots, j_k); j_1 + \ldots + j_k = n\}$.

Finally, we introduce the concept of weak analyticity.

Definition 2: Let A_θ be a random matrix on S. We say that A_θ is weakly analytical if there exist a neighborhood V_θ of 0 such that for all Δ in V_θ s.t. $\theta+\Delta \in \Theta$ it holds that:

$$A_{\theta+\Delta} \equiv \sum_{n\geq 0} \frac{\Delta^n}{n!} A_\theta^{(n)}.$$

The sum in the right-hand side of the above expression (provided that it has a meaning) is called the *Taylor series expansion* of A_θ.

The main result regarding analyticity is the following:

Theorem 2: Let A_θ and B_θ be two independent random matrices. If A_θ and B_θ are weakly analytical, then $A_\theta \otimes B_\theta$ is weakly analytical. Moreover, the Taylor series of $A_\theta \otimes B_\theta$ converges for all Δ for which both Taylor series of A_θ and B_θ, respectively are convergent.

Applying inductively the above result, we can compute the Taylor series expansion of the product $A \stackrel{\text{def}}{=} (A_1 \otimes \cdots \otimes A_k)$. Precisely, we have:

$$A_{\theta+\Delta} \equiv \sum_{n\geq 0} \left[\sum_{j\in\mathcal{J}(k,n)} \frac{\Delta^n}{j_1!\cdots j_k!} \cdot A_1^{(j_1)} \otimes' \ldots \otimes' A_k^{(j_k)} \right].$$

Let us turn back to the Bernoulli example. Assume that $A_1,\ldots,A_k$ are i.i.d. Bernoulli distributed.

Then $A_i^{(j)} = \begin{cases} (1, A_i, \mathcal{E}), & \text{j=0;} \\ (1, P, D), & \text{j=1;} \\ not \quad significant, & j \geq 2, \end{cases}$

for each $1 \leq i \leq k$, $j \geq 0$.

Thus, for $n > k$ the nth order derivatives of the product $A_1 \otimes \cdots \otimes A_k$ are not significant and therefore the Taylor series is finite. For computing the rest of the derivatives, for $n \leq k$ we need the following notations. Let $[k] \stackrel{\text{def}}{=} \{1,2,\ldots,k\}$ and $I \subset [k]$. Denote by $|I|$ the cardinal number of I and let

$$\Pi_I \stackrel{\text{def}}{=} \bigotimes_{i=1}^{k} B_i,$$

where

$$B_i \stackrel{\text{def}}{=} \begin{cases} P, & i \in I; \\ D, & \text{otherwise.} \end{cases}$$

For instance, $\Pi_\emptyset = D^{\otimes k}$, $\Pi_{\{i\}} = \underset{1}{D} \otimes \ldots \otimes \underset{i}{P} \otimes \ldots \otimes \underset{k}{D}$, $1 \leq i \leq k$, …, $\Pi_{[k]} = P^{\otimes k}$.

Let now $g : \mathbb{R}^m \to \mathbb{R}$ given by $g(x_1,\ldots,x_m) = x_1$ and $x_0 \in \mathbb{R}^m$. Then the mapping $G : S \to \mathbb{R}$ defined through $G(A) \stackrel{\text{def}}{=} g(A \otimes x_0)$ belongs to $\mathcal{C}^p(S)$.

Finally, set: $\sigma_k(n) \stackrel{\text{def}}{=} \sum_{|I|=n} G(\Pi_I)$, for $0 \leq n \leq k$. Thus, we have $\sigma_k(0) = g(D^{\otimes k} \otimes x_0)$ and $\sigma_k(k) = g(P^{\otimes k} \otimes x_0)$. Then the analyticity of $A_1 \otimes \cdots \otimes A_k$ in 0 yields:

$$\mathbb{E}[g(X_\theta(k))] = \sum_{n=0}^{k} \theta^n \left[\sum_{j=0}^{n} (-1)^{n-j} \binom{k-j}{n-j} \sigma_k(j) \right].$$

We conclude this section with a short discussion of the error in the above approximation. Namely, one is interested to fix a certain $p < k$ and to evaluate the error term

$$\epsilon \stackrel{\text{def}}{=} \left| \mathbb{E}[g(X_\theta(k))] - \sum_{n=0}^{p} \theta^n \left[\sum_{j=0}^{n} (-1)^{n-j} \binom{k-j}{n-j} \sigma_k(j) \right] \right|.$$

By computation, the following can be found ($\theta < \frac{1}{2}$):

$$\epsilon \leq \frac{2^{k-1}(2\theta)^{p+1}}{(1-2\theta)\sqrt{k+1}} \|P^{\otimes k} \otimes x_0 - D^{\otimes k} \otimes x_0\|_\infty, \qquad (6)$$

where $\|\cdot\|_\infty$ denotes the maximum norm in $\mathbb{R}^m$.

Thus, for fixed k and θ, one can compute the order p of the Taylor series, necessary for a certain level of accuracy.

Sometimes, one is interested to evaluate the relative error:

$$\epsilon_r \stackrel{\text{def}}{=} \frac{\epsilon}{\|P^{\otimes k} \otimes X_0 - D^{\otimes k} \otimes X_0\|_\infty} \stackrel{(6)}{\leq} \frac{2^{k-1}(2\theta)^{p+1}}{(1-2\theta)\sqrt{k+1}},$$

in order to get an accurate estimation relative to the length of the interval within which $\mathbb{E}[X(k)]$ takes values.

Numerical computations for $\mathbb{E}[X(k)]$, Taylor series expansions, absolute and relative errors are illustrated in the next section as applications to the Baccelli & Hong example.

V. Numerical Results

For the experiments performed in this section, we let $\sigma = 14$ and $\sigma' = 24$. This yields for the nominal matrix

$$D = \begin{bmatrix} 14 & \varepsilon & 24 & \varepsilon \\ 14 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & 0 \\ \varepsilon & \varepsilon & 24 & \varepsilon \end{bmatrix}$$

and for the perturbed matrix

$$P = \begin{bmatrix} 14 & \varepsilon & 24 & \varepsilon \\ 14 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & 24 & \varepsilon \\ \varepsilon & \varepsilon & 24 & \varepsilon \end{bmatrix}.$$

Figure 3 shows true value of $\mathbb{E}[X_1(10)]$ as function of θ and Taylor series approximations of degree $p = 1$ to $p = 3$. The exact values are given in Table 4.

Note that the 1^{st} (2^{nd} and 3^{rd}) Taylor series polynomials is obtained by considering all possible matrix products involving only D matrices and at most one (two resp. three) occurrence of matrix P; see Section IV.

The numerical values show that the Taylor series approximation of degree $p = 3$ approximates the true performance quite accurately up to $\theta = 0.35$. This is confirmed with Table 5 where the exact absolute and relative errors for this approximation are listed.

Unfortunately, for large values of θ, the estimated relative error margin fails to be "close" to the true one (according to the last column in Table 5, we can predict reasonable margins for θ up to 0.15). This inconvenience can be removed by increasing correspondingly the order of Taylor series but this would require a larger computation time.

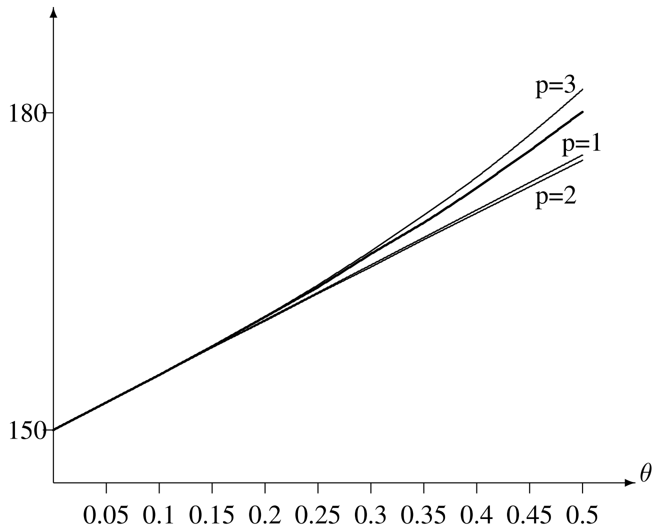

Fig. 3. $\mathbb{E}[x_1(10)]$ and its 1^{st}, 2^{nd} and 3^{rd} order Taylor approximations; the thick line represents the true value.

θ	$\mathbb{E}[x_1(10)]$	$T_1(\theta)$	$T_2(\theta)$	$T_3(\theta)$
0.00	150.000	150.000	150.000	150.000
0.05	152.601	152.600	152.595	152.602
0.10	155.229	155.200	155.180	155.234
0.15	157.911	157.800	157.755	157.937
0.20	160.675	160.400	160.320	160.752
0.25	163.540	163.000	162.875	163.719
0.30	166.527	165.600	165.420	166.878
0.35	169.655	168.200	167.955	170.270
0.40	172.945	170.800	170.480	173.936
0.45	176.420	173.400	172.995	177.916
0.50	180.105	176.000	175.500	182.250

Fig. 4. Table of values for $\mathbb{E}[x_1(10)]$ and its 1^{st}, 2^{nd} and 3^{rd} order Taylor approximations.

θ	$\mathbb{E}[x_1(10)]$	$T_3(\theta)$	ϵ	ϵ_r	est. ϵ_r
0.00	150.000	150.000	0.000	0.000	0.000
0.05	152.601	152.602	0.001	0.001	0.019
0.10	155.229	155.234	0.005	0.006	0.343
0.15	157.911	157.937	0.026	0.032	1.984
0.20	160.675	160.752	0.077	0.096	7.317
0.25	163.540	163.719	0.179	0.223	21.438
0.30	166.527	166.878	0.351	0.438	—
0.35	169.655	170.270	0.615	0.768	—
0.40	172.945	173.936	0.991	1.238	—
0.45	176.420	177.916	1.496	1.870	—
0.50	180.105	182.250	2.145	2.681	—

Fig. 5. The absolute, true and estimated relative error (percentage) of the 3^{rd} order Taylor approximation.

VI. Conclusion and Topics for Further Research

We have presented a general approach to numerical approximate computation of generalized matrix products. Weak derivatives are not unique and it is a topic of further research to find representations of the derivatives that allow for a more efficient implementation. In particular the application of our methods to large scale problems is an open issue. The ultimate goal is to come up with Taylor series algorithms for the Lyapunov exponents of generalized matrix products. Improvements of the accuracy of error estimation as well as studying the trade-off between increasing the order of Taylor series p and getting a better error margin are topic of future research.

References

[1] Ayhan, H. and Baccelli, F. Expansions for joint Laplace transforms for stationary waiting times in (max,+)–linear systems with Poisson input. *QUESTA*, 37:291–328, 2001.

[2] Ayhan, H. and Seo, D. Tail probability of transient and stationary waiting times in (max,+)-linear systems. *IEEE Trans. Aut. Cont.*, 47:151–157, 2000.

[3] Ayhan, H. and Seo, D. Laplace transform and moments of waiting times in Poisson driven (max,+)-linear systems. *QUESTA*, 37:405–436, 2001.

[4] Baccelli, F., Cohen, G., Olsder, G. J., and Quadrat, J.–P. *Synchronization and Linearity*. John Wiley and Sons, (this book is out of print and can be accessed via the max-plus web portal at http://maxplus.org), 1992.

[5] Baccelli, F. and Hong, D. Analytic expansions of (max,+) Lyapunov exponents. *Annals of Appl. Probab.*, 10:779–827, 2000.

[6] Baccelli, F. and Schmidt, V. Taylor series expansions for Poisson–driven (max,+)–linear systems. *Annals Appl. Prob.*, 6:138–185, 1996.

[7] Baccelli, F., Hasenfuß, S., and Schmidt, V. Transient and stationary waiting times in (max,+)–linear systems with Poisson input. *QUESTA*, 26:301–342, 1997.

[8] Baccelli, F., Hasenfuß, S., and Schmidt, V. Expansions for steady–state characteristics of (max,+)–linear systems. *Com. in Statistcs: Series Stochastic Models*, 14:1–24, 1998.

[9] Cohen, J.E. Subadditivity, generalized products of random matrices and operation research. *SIAM Review*, 30, No.1, 1988.

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

[11] van den Boom, T., De Schutter, B., and Heidergott, B. Complexity reduction in MPC for stochastic max-plus-linear systems by variability expansion. In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 3567–3572, Las Vegas, Nevada, December 2002.

[12] Heidergott, B. A differential calculus for random matrices with applications to (max,+) stochastic systems. *MOR*, 26, no. 4, pp. 679-699, 2001.

[13] Heidergott, B. Variablity Expansion for Performance Characteristic of (Max,Plus)-Linear Systems, Workshop on Discrete Event Systems (WODES), Zaragoza, Spain, 2-4 October, 2002.

[14] Heidergott, B., Olsder, G.J., and van der Woude, J. *Max Plus at Work: Modeling and Analysis of Synchronized Systems*, Princeton University Press, 2005.

[15] Hewitt, E., and Stromberg, K. *Real and Abstract Analysis*. Springer, Berlin, 1969.

[16] Kim, K. H. *Boolean Matrix Theory and Applications*. Marcel Dekker, New York, 1982.

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