Analyticity of transient (max,+)-linear stochastic systems

Report 99-030

Analyticity of Transient (Max,+)–linear Stochastic Systems

Bernd Heidergott

ISSN: 1389–2355

Analyticity of Transient (Max,+)–linear Stochastic Systems

Bernd Heidergott *
EURANDOM
P.O.Box 513, 5600 MB Eindhoven, the Netherlands

E–Mail: heidergott@eurandom.tue.nl

Abstract

We give Taylor series expansions of performance functions of (max,+)-linear stochastic systems. Instances of the performance measures considered are Laplace transforms, moments, distribution functions and tail probabilities. We study the domain of convergence of the Taylor series developed at any point of analyticity. The elements of the Taylor series can (in the most simple cases) be calculated analytically or estimated via simulation. The cornerstone of our analysis is the introduction of a calculus of higher-order weak differentiation for random matrices. This calculus is based on the concept of weak differentiation. In order to obtain our results, we extend this concept, originally formulated only for bounded performance measures, to a more general class of performance measures.

Keywords: Queueing networks; stochastic Petri nets; Taylor series expansions; perturbation analysis; weak derivatives;

AMS 1991 subject classifications: Primary 60K25; Secondary: 41A58.

1 Introduction

We study stochastic (max,+)–linear systems. This class of systems allows one to represent stochastic Petri nets belonging to the class of event graphs. It contains various instances of queueing networks like the G/G/1 queue, (finite) queues in tandem, Kanban systems

*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.

[1], flexible manufacturing systems [12], fork–join queues or any parallel and/or series composition made with these elements. In particular, we consider parameter–dependent (max,+)–linear stochastic systems, where the parameter, say θ, is a parameter of the distribution of the transition dynamic of the event graph. Thus, θ, for example, may be a parameter of one of the firing time distributions of the event graph. More precisely, in a queueing application, θ may be the mean service time at one of the queues. Baccelli and Hong give an example from computer science in [5], where θ is a parameter of the distribution governing the entire transition dynamic; 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 are interested in the analyticity of performance measures, $J(\theta)$, like completion times or waiting times, of (max,+)–linear systems, that is, we are interested in Taylor series expansions of $J(\theta)$ in θ. First results on analyticity of stochastic networks were given by Zazanis [17], who studied analyticity of performance measures of stochastic networks with a Poisson arrival stream with respect to the intensity of the arrival stream. Baccelli and Schmidt [6] considered the case in which the network is (max,+)–linear. Their approach was further developed in [3] and [4]. For applications of their results to waiting times, see [13] and [16]. The results mentioned above are restricted to the case of open networks, where θ is the intensity of the arrival stream. However, in a recent paper, Baccelli and Hong derived first results for the case of closed networks, see [5]. Strictly speaking, the aforementioned papers study Maclaurin series, that is, they only consider Taylor series developed at zero.

In this paper we establish sufficient conditions for analyticity of transient (max,+)–linear stochastic systems. We provide an algebraic approach for calculating higher–order derivatives of performance measures of (max,+)–linear systems. Our approach applies to Laplace transforms, moments, distribution functions and tail probabilities. In particular,

1. for open systems, we do not require the arrival stream to be of Poisson type;
2. our analysis applies both to open and closed systems (not only do we recover the known results for open systems, we also extend them to closed systems);
3. we establish the domain of convergence of the Taylor series developed at any point of analyticity.

In particular cases, the derivatives obtained can be calculated analytically. In general, however, the formulae obtained have a simple interpretation as unbiased estimation algorithm.

Our approach is based on the concept of *weak differentiation*, introduced by Pflug, see [15]. This concept is closely related to Markov chain analysis. In a recent paper, Cao developed steady–state performance functions of finite–state Markov chains in a Maclaurin series, see [8]. Although the types of systems Cao considers are different from the ones treated here, the approach Cao suggests is closely related to the one we propose in this paper.

The paper is organised as follows. Section 2 introduces the (max,+)–algebra and illustrates its modelling power. Section 3 provides a short introduction to the theory of weak differentiation, which is the basis of our further analysis. In Section 4, we establish a calculus of higher-order weak differentiation for random matrices in the (max,+)–algebra setting. In Section 5 we provide our main result on analyticity of transient (max,+)–linear systems. Finally, Section 6 illustrates how the results already known about analyticity of waiting times in open (max,+)–linear systems can be recovered within our framework.

2 (Max,+)–linear Systems

In this section we introduce the (max,+)–semiring. This structure was first introduced in [9]. For an extensive discussion of the (max,+)–algebra and similar structures we refer to [2].

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,+)–algebra 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 $e := 0$.

Some remarks on the particularities of the (max,+)–algebra seem to be in order here. The name "(max,+)–algebra" is only historically justified, since $\mathbb{R}_{\max}$ is by no means an algebra in the classical sense. Structures like $\mathbb{R}_{\max}$ are referred to as *semirings*[1] in the literature. In particular, $\mathbb{R}_\epsilon$ is *idempotent*, that is, for all $a \in \mathbb{R}_\epsilon$ $a \oplus a = a$. Idempotent semirings are called *dioids* in [2]. Hence, the correct name for $\mathbb{R}_{\max}$ would be "idempotent semiring" or "dioid" (which might explain why the name "(max,+)–algebra" is still predominant in the literature). 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, the commutativity and general invertability of the product.

Observe that the idempotency of $\oplus$ implies that $\oplus$ has no inverse (which explains why $\mathbb{R}_{\max}$ is no algebra). Indeed, if $a \neq \epsilon$ had an inverse element, say b, w.r.t. $\oplus$, then $a \oplus b = \epsilon$ would imply $a \oplus a \oplus b = a \oplus \epsilon$. By idempotency, the left–hand side equals $a \oplus b$, whereas the right–hand side is equal to a. Hence, we have $a \oplus b = a$, which contradicts $a \oplus b = \epsilon$.

[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 $\otimes$.

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

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

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

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

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 $\mathbb{R}_{\max}^{J\times J}$ 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 $\mathbb{R}_\epsilon$.

Let some probability space be given on which all random variables introduced below are defined.

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

In the following we give some examples of (max,+)–linear queueing networks. We refer to [11] for a necessary and sufficient condition for the (max,+)–linearity of a queueing network.

Example 1 *Consider an open system of J single-server queues in tandem, with infinite buffers. We let queue 0 represent an external arrival stream of customers. Each customer who arrives at the system passes through queues 1 to J, and then leaves the system. For simplicity, we assume that the system starts empty. Let $\sigma_j(k)$ denote the k^{th} service time at station j and $x_j(k)$ the time of the k^{th} service completion at station j. In particular, $\sigma_0(k)$ denotes the k^{th} interarrival time and $x_0(k)$ denotes the n^{th} arrival epoch at the system. The time evolution of the system can then be described by a $(J+1)$–dimensional vector $x(k) = (x_0(k), \ldots, x_J(k))$ following the recursion*

$$x(k+1) = A(k) \otimes x(k)\,, \tag{2}$$

where the matrix $A(k)$ looks like

$$\begin{bmatrix} \sigma_0(k) & \epsilon & \epsilon & \cdots & \epsilon \\ \sigma_0(k)\otimes\sigma_1(k) & \sigma_1(k) & \epsilon & \cdots & \epsilon \\ \sigma_0(k)\otimes\sigma_1(k)\otimes\sigma_2(k) & \sigma_1(k)\otimes\sigma_2(k) & \sigma_2(k) & \cdots & \epsilon \\ & & \cdots & & \\ \sigma_0(k)\otimes\ldots\otimes\sigma_J(k) & \sigma_1(k)\otimes\ldots\otimes\sigma_J(k) & \sigma_2(k)\otimes\ldots\otimes\sigma_J(k) & \cdots & \sigma_J(k) \end{bmatrix} \tag{3}$$

for $k \geq 0$. *Alternatively, we could describe the system via a* J*-dimensional vector* $\hat{x}(k) = (\hat{x}_1(k), \ldots, \hat{x}_J(k))$ *following*

$$\hat{x}(k+1) = \hat{A}(k) \otimes \hat{x}(k) \oplus \hat{B}(k) , \tag{4}$$

where the matrix $\hat{A}(k)$ *looks like (3) except for the first column and the first row, which are missing, that is,* $(\hat{A}(k))_{ij} = (A(k))_{i+1\,j+1}$ *for* $1 \leq i, j \leq J$*; the vector* $\hat{B}(k)$ *is given by* $\hat{B}(k)_j = B(k)_j \otimes \tau(k)$, *with*

$$B(k) = \begin{bmatrix} e \\ \sigma_1(k) \\ \sigma_1(k) \otimes \sigma_2(k) \\ \ldots \\ \sigma_1(k) \otimes \ldots \otimes \sigma_J(k) \end{bmatrix}$$

for $k \geq 0$*; and*

$$\tau(k) = \sum_{i=1}^{k} \sigma_0(i) \tag{5}$$

denotes the k^{th} *arrival epoch. For more examples of this kind we refer to [14].*

Example 1 models sequences of departure times from the queues via a (max,+)-recursion. Another interesting application of (max,+)-linear models is the analysis of waiting times.

Example 2 *Consider an open queueing network the departure epochs of which can be described by a (max,+)-linear equation, cf. the open tandem network described in Example 1. Let* $W_j(k)$ *be the time the* k^{th} *customer arriving at the network spends in the system until leaving node* j. *Using the notation introduced in the example above, we obtain*

$$W_j(k) = \hat{x}_j(k) - \tau(k) , \tag{6}$$

for $k \geq 1$. *After some algebraic manipulations (see [6] for details), the vector of system times* $W(k) = (W_1(k), \ldots, W_J(k))$ *reads*

$$W(k+1) = \hat{A}(k) \otimes C(\sigma_0(k)) \otimes W(k) \oplus B(k) ,$$

where $C(r)$ *is a matrix with diagonal entries* $-r$ *and all other entries equal to* ϵ.

Suppose that one of the service time distributions depends on a parameter, say, θ, which may be the mean 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 3 (Baccelli & Hong, [5]) *Consider a cyclic tandem queueing network consisting of a single server and a multi server, each with deterministic service times. Service times at 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} beginning 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*

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

where

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

Consider the cyclic tandem network again, but one of the servers of the multi–server station has broken down. This system follows

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

where

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

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

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

and

$$P(\, A_\theta(k) = \hat{A}\,) = 1 - \theta ,$$

then

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

describes the time evolution of the system with breakdowns.

2.3 Problem Statement

We study sequences $\{x_\theta(k) : k \geq 0\}$ following

$$x_\theta(k+1) = A_\theta(k) \otimes x_\theta(k) \oplus B_\theta(k) , \quad k \geq 0 .$$

with $x_\theta(0) = x_0$, $A_\theta(k) \in \mathbb{R}_\epsilon^{J \times J}$, $B_\theta(k) \in \mathbb{R}_\epsilon^J$ and $\theta \in \Theta \subset \mathbb{R}$. For a given performance function $g : \mathbb{R}_\epsilon^J \to \mathbb{R}$, we seek conditions for the analyticity of

$$E[\, g(x_\theta(k+1)) | x_\theta(0) = x_0\,] . \tag{7}$$

These conditions will depend on the type of performance function and the particular way in which the matrix $A_\theta(k)$, respectively the vector $B_\theta(k)$, depends on θ. In the next section, we introduce the concept of weak differentiability of measures. We will transfer this concept to that of weak differentiability of random matrices and eventually study "weak analyticity" of random matrices. Our main result will be that the $\otimes$-product and $\oplus$-sum of weakly analytical matrices is again weakly analytical on the same region. Weak analyticity of $x_\theta(k+1)$ then implies analyticity of (7) for a particular class of performance measures.

3 Weak Differentiability of Measures

This section provides an introduction to the theory of weak differentiation of measures. A key reference on the basic concept of weak differentiation is [15].

Let (S, d_S) be a separable metric space and let $\mathcal{M} = \mathcal{M}(S)$ be the set of all finite signed regular measures on the measure space $(S, \mathcal{F})$, where $\mathcal{F}$ denotes the Borel field of S. The set of all probability measures on $(S, \mathcal{F})$ is denoted by $\mathcal{M}_1 = \mathcal{M}_1(S)$. Let $\mathcal{D}(S)$ be a set of mappings from S to $\mathbb{R}$ and assume that the constant function $g \equiv 1$ is in $\mathcal{D}(S)$. We say the mapping $\mu_\theta : \Theta \to \mathcal{M}_1$ is *weakly differentiable at θ with respect to $\mathcal{D}(S)$* if there exists a $\mu'_\theta \in \mathcal{M}$, such that for all $g \in \mathcal{D}(S)$

$$\lim_{\Delta \to 0} \frac{1}{\Delta} \left(\int g \, d\mu_{\theta+\Delta} - \int g \, d\mu_\theta \right) = \int g \, d\mu'_\theta \, . \tag{8}$$

From the well-known Hahn–Jordan decomposition theorem it follows that probability measures $\mu_\theta^{(+1)}, \mu_\theta^{(-1)} \in \mathcal{M}_1$ and constants $c_1, c_2 \geq 0$ exist, such that

$$\mu'_\theta = c_1 \, \mu_\theta^{(+1)} - c_2 \, \mu_\theta^{(-1)} \, .$$

The measure $\mu_\theta^{(+1)}$ is called the positive part and the measure $\mu_\theta^{(-1)}$ the negative part of μ_θ'. Moreover, a set $A \in \mathcal{F}$ exists, such that either $\mu_\theta^{(+1)}(A) = 0$ or $\mu_\theta^{(-1)}(S \setminus A) = 0$, in symbols: $\mu_\theta^{(+1)} \perp \mu_\theta^{(-1)}$. The above representation is not unique, since for an arbitrary non-negative measure $\gamma \in \mathcal{M}$ and a positive constant b

$$\mu'_\theta = (c_1 + b) \left(\mu_\theta^{(+1)} + \gamma \right) - (c_2 + b) \left(\mu_\theta^{(-1)} + \gamma \right) \, .$$

However, it can be shown that $c_1 + c_2$ is minimised if $\mu_\theta^{(+1)} \perp \mu_\theta^{(-1)}$.

For $\mu_\theta \in \mathcal{M}_1$, it follows that $\mu_\theta(S) = 1$ for all θ. Therefore, $d\mu_\theta(S)/d\theta = 0$ for all $\theta \in \Theta$. Because the constant function $g \equiv 1$ is in $\mathcal{D}(S)$, this implies $\mu_\theta^{(+1)}(S) = \mu_\theta^{(-1)}(S)$. Thus, the normalisation constants are equal, $c_1 = c_2$, and μ'_θ is completely characterised through the triple $(c_{\mu_\theta}, \mu_\theta^{(+1)}, \mu_\theta^{(-1)})$, with $\mu_\theta^{(+1)}, \mu_\theta^{(-1)} \in \mathcal{M}_1$. We call the triple $(c_{\mu_\theta}, \mu_\theta^{(+1)}, \mu_\theta^{(-1)})$ a *weak derivative of μ_θ at θ with respect to $\mathcal{D}(S)$*.

Weak differentiability of a random variable is defined by the weak differentiability of the induced measure: Let $(X_\theta : \theta \in \Theta)$ be defined on a common probability space $(\Omega, \mathcal{F}, P)$, such that $P^{X_\theta} = \mu_\theta$. We call X_θ weakly differentiable if μ_θ is weakly differentiable. The weak derivative of X_θ is a triple $(c, X_\theta^{(+1)}, X_\theta^{(-1)})$, such that $X_\theta^{(+1)}$ has distribution $\mu_\theta^{(+1)}$, $X_\theta^{(-1)}$ has distribution $\mu_\theta^{(-1)}$ and c is the normalisation constant of μ'_θ.

Higher–order derivatives are defined in the same way. More precisely, we call a triple $(c^{(n)}, \mu_\theta^{(n,+1)}, \mu_\theta^{(n,-1)})$ a n^{th} *order weak derivative of* μ_θ *at* θ *with respect to* $\mathcal{D}(S)$ if for all $g \in \mathcal{D}(S)$

$$\frac{d^n}{d\theta^n} \int g \, d\mu_\theta = c^{(n)} \left(\int g \, d\mu_\theta^{(n,+1)} - \int g \, d\mu_\theta^{(n,-1)} \right) . \tag{9}$$

If the left–hand side of the above equation equals zero for all $g \in \mathcal{D}(S)$, we take $(0, \mu_\theta, \mu_\theta)$ as the n^{th} order weak derivative and say that the n^{th} order weak derivative of μ_θ is *not significant*, whereas it is called *significant* otherwise.

Furthermore, we call $(c_{X_\theta}^{(n)}, X_\theta^{(n,+1)}, X_\theta^{(n,-1)})$ an n^{th} order weak derivative of X_θ if the distribution of X_θ, denoted by μ_θ, has n^{th} order weak derivative $(c_{\mu_\theta}^{(n)}, \mu_\theta^{(n,+1)}, \mu_\theta^{(n,-1)})$, and $X_\theta^{(n,+1)}$ is distributed according to $\mu_\theta^{(n,+1)}$, $X_\theta^{(n,+1)}$ is distributed according to $\mu_\theta^{(n,+1)}$ and $c_{X_\theta}^{(n)} = c_{\mu_\theta}^{(n)}$. If it causes no confusion, we simplify the notation by dropping the subscript X_θ of $c_{X_\theta}^{(n)}$ and write $c^{(n)}$ for $c_{X_\theta}^{(n)}$.

In order to make the above definition useful, we have to choose $\mathcal{D}(S)$ in such a way that

- it is rich enough to contain interesting performance functions,
- the product of weakly differentiable measures is again weakly differentiable.

In what follows, we study two examples of $\mathcal{D}(S)$. Let $C^c(S) := C^c(S, d_S)$ denote the set of all bounded continuous functions from (S, d_S) to $\mathbb{R}$. We assume that $(S, +^S, 0^S)$ is an (additive) monoid endowed with a quasi norm $||\cdot||_S$, that is, $||s||_S = 0$ if and only if $s = 0^S$, $||s||_S \geq 0$ for all $s \in S$, and $||s +^S r||_S \leq ||s||_S + ||r||_S$. We denote by $C_k(S) := C_k(S, d_S)$ the set of all functions $g : S \to \mathbb{R}$ such that $|g(x)| \leq c_1 + c_2 ||x||^l$ for all $x \in S$ and all l with $0 \leq l \leq k$. In addition to that we assume that $C^c(S) \subset C_k(S)$ for all $k \geq 0$.

Suppose that the derivative of $\int g \, d\mu_\theta$ does not vanish for all $g \in \mathcal{D}(S)$. Obviously, there are two different ways of calculating the second–order weak derivative of the measure μ_θ. On the one hand, we can obtain a second–order weak derivative $(c^{(2)}, \mu_\theta^{(2,+1)}, \mu_\theta^{(2,-1)})$ with respect to $\mathcal{D}(S)$ through (9), where we assume $\mu_\theta^{(2,+1)} \perp \mu_\theta^{(2,-1)}$ in order to obtain a unique decomposition. On the other hand, we can take the weak derivative of μ'_θ, that is, weakly differentiate the positive and negative part of μ'_θ separately, and afterwards rescale the measures in order to obtain probability measures. This yields

$$(\mu_\theta^{(+1)})' = c^{(+1)} \mu_\theta^{((+1)+1)} - c^{(+1)} \mu_\theta^{((+1)-1)}$$

and

$$(\mu_\theta^{(-1)})' = c^{(-1)} \mu_\theta^{((-1)+1)} - c^{(-1)} \mu_\theta^{((-1)-1)} ,$$

with $\mu_\theta^{((i_1)i_2)} \in \mathcal{M}_1$ for $i_1, i_2 \in \{-1,+1\}$, where we assume that $\mu_\theta^{((+1)+1)} \perp \mu_\theta^{((+1)-1)}$ and $\mu_\theta^{((-1)+1)} \perp \mu_\theta^{((-1)-1)}$. Let $g \in \mathcal{D}(S)$, then

$$\begin{aligned}\frac{d^2}{d\theta^2}\int g\,d\mu_\theta =& \frac{d}{d\theta}\left(\int g\,d\mu_\theta^{(+1)} - \int g\,d\mu_\theta^{(-1)}\right)\\ =& c^{(+1)}\int g\,d\mu_\theta^{((+1)+1)} - c^{(+1)}\int g\,d\mu_\theta^{((+1)-1)}\\ &+ c^{(-1)}\int g\,d\mu_\theta^{((-1)+1)} - c^{(-1)}\int g\,d\mu_\theta^{((-1)-1)}\,.\end{aligned}$$

Regrouping the positive and negative parts yields

$$\begin{aligned}&\frac{d^2}{d\theta^2}\int g\,d\mu_\theta\\ &= \left(c^{(+1)}+c^{(-1)}\right)\left(\frac{c^{(+1)}}{c^{(+1)}+c^{(-1)}}\int g\,d\mu_\theta^{((+1)+1)} + \frac{c^{(-1)}}{c^{(+1)}+c^{(-1)}}\int g\,d\mu_\theta^{((-1)+1)}\right.\\ &\quad\left. - \left(\frac{c^{(+1)}}{c^{(+1)}+c^{(-1)}}\int g\,d\mu_\theta^{((+1)-1)} + \frac{c^{(-1)}}{c^{(+1)}+c^{(-1)}}\int g\,d\mu_\theta^{((-1)-1)}\right)\right),\end{aligned}$$

and we obtain a second–order weak derivative of μ_θ from

$$\begin{aligned}\Bigg(c^{(+1)}+c^{(-1)}\,,\ &\frac{c^{(+1)}}{c^{(+1)}+c^{(-1)}}\mu_\theta^{((+1)+1)} + \frac{c^{(-1)}}{c^{(+1)}+c^{(-1)}}\mu_\theta^{((-1)+1)}\,,\\ &\frac{c^{(+1)}}{c^{(+1)}+c^{(-1)}}\mu_\theta^{((+1)-1)} + \frac{c^{(-1)}}{c^{(+1)}+c^{(-1)}}\mu_\theta^{((-1)-1)}\Bigg)\,.\end{aligned}$$

We assumed that (S, d_S) is a metric space and that the measures in $\mathcal{M}$ are defined on $(S, \mathcal{F})$, where $\mathcal{F}$ is the Borel field with respect to d_S. Therefore, two measures $\mu, \nu \in \mathcal{M}$ are equal if $\int f\,d\mu = \int f\,d\nu$ for all $f \in C^c(S)$, for a proof see e.g. Lemma 30.14 in [7]. Hence, if $C^c(S) \subset \mathcal{D}(S)$, then $c^{(2)} = c^{(+1)} + c^{(-1)}$ and

$$\mu_\theta^{(2,+1)} = \frac{c^{(+1)}}{c^{(+1)}+c^{(-1)}}\mu_\theta^{((+1)+1)} + \frac{c^{(-1)}}{c^{(+1)}+c^{(-1)}}\mu_\theta^{((-1)+1)}$$

and

$$\mu_\theta^{(2,-1)} = \frac{c^{(+1)}}{c^{(+1)}+c^{(-1)}}\mu_\theta^{((+1)-1)} + \frac{c^{(-1)}}{c^{(+1)}+c^{(-1)}}\mu_\theta^{((-1)-1)}\,.$$

Put another way, the definitions (8) and (9) are compatible. It can easily be seen that the same holds true for all higher–order weak derivatives, provided that they exist.

3.1 The Space $C^c(S)$

The weak convergence of measures in $\mathcal{M}_1$ is defined by means of $C^c(S)$. We say that $\mu_n \in \mathcal{M}_1$ converges weakly towards $\mu \in \mathcal{M}_1$ if for all $g \in C^c(S)$

$$\lim_{n\to\infty}\int g\,d\mu_n = \int g\,d\mu\,.$$

The above definition explains why the limit in (8) is called a "weak" derivative. The product of two weakly differentiable measures is weakly differentiable, see [15].

Unfortunately, the space $C^c(S)$ is too small for many interesting problems in applications. For example, for $S = \mathbb{R}$, the identity $id : \mathbb{R} \to \mathbb{R}$ is generally not in $C^c(\mathbb{R})$. Therefore, we generally cannot calculate the derivative of the moments of μ_θ.

3.2 The Space $C_k(S)$

The space $C_k(S)$ allows us to describe many interesting performance characteristics as the following example illustrates.

Example 4 *For $J \geq 1$, take $S = [0, \infty)^J$ and let $X = (X_1, \ldots, X_J) \in S$ be defined on a probability space $(\Omega, \mathcal{A}, P)$ such that $P^X = \mu$. Taking $g(x) = \exp(-r\, x_j)$, with $r \geq 0$, we obtain the Laplace transform of X through*

$$E\left[e^{-rX_j}\right] = \int g\, d\mu\,;$$

For $g(x) = x_j^k$, we obtain the higher-order moments of X through

$$E\left[X_j^k\right] = \int g\, d\mu\,, \qquad \text{for } k \geq 1\,,$$

and, taking $g(x) = 1_{x_j > u}$, we obtain

$$P(X_j > u) = \int g\, d\mu\,,$$

the tail probabilities of X.

In what follows we study weak differentiability of product measures with respect to $C_k(S)$. The main difference between weak differentiability with respect to $C_k(S)$ and weak differentiability with respect to $C^c(S)$ is that for the latter we needn't restrict the class of measures which can be treated.

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 $\nu \in \mathcal{M}_1(S)$ 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 $C_k(S)$, or n times* Lipschitz differentiable *for short, if*

- *$f_\theta(x)$ is $(n+1)$ times differentiable with respect to θ on Θ for ν almost all x;*

- *for all* $0 \le m \le n+1$

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

ν almost surely, and

- *for all* $0 \le m \le n+1$

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

The n^{th} Lipschitz derivative at θ is said to be *not significant* if $d^n f_\theta(x)/d^n\theta = 0$ for ν-almost all x and significant otherwise. Furthermore, we let $s(\mu_\theta)$ denote the highest order of a significant derivative. If all higher-order derivatives are significant, then we set $s(\mu_\theta) = \infty$.

The name "Lipschitz differentiability" stems from the fact that

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

implies that $d^{m-1} f_\theta(x)/d\theta^{m-1}$ satisfies a Lipschitz condition, that is,

$$\left| \left. \frac{d^{m-1}}{d\theta^{m-1}} \right|_{\theta=\theta_0+\Delta} f_\theta(x) - \left. \frac{d^{m-1}}{d\theta^{m-1}} \right|_{\theta=\theta_0} f_\theta(x) \right| \le \Delta \, K_f^m(x) ,$$

for $\theta_0, \theta_0 + \Delta \in \Theta$.

Lipschitz differentiability allows for weak differentiation. More precisely, the n times μ-Lipschitz differentiability of μ_θ implies that for all $g \in C_k(S)$

$$\frac{d^n}{d\theta^n} \int g \, d\mu_\theta = \int g \, \frac{d^n}{d\theta^n} f(\mu_\theta, \mu) \, d\mu = \int g \, \frac{d^n}{d\theta^n} f_\theta \, d\mu . \tag{10}$$

Set

$$c^{(n)} = \frac{1}{2} \int \left| \frac{d^n}{d\theta^n} f_\theta \right| \, d\mu .$$

Note that $c^{(n)}$ is finite whenever μ_θ is n times Lipschitz differentiable. We define μ-densities

$$f_\theta^{(n,+1)} = \frac{1}{c^{(n)}} \max \left(\frac{d^n}{d\theta^n} f_\theta \, , \, 0 \right) , \quad f_\theta^{(n,-1)} = \frac{1}{c^{(n)}} \max \left(- \frac{d^n}{d\theta^n} f_\theta \, , \, 0 \right) ,$$

then equation (10) reads

$$\frac{d^n}{d\theta^n} \int g \, d\mu_\theta = c^{(n)} \left(\int g \, f_\theta^{(n,+1)} \, d\mu - \int g \, f_\theta^{(n,-1)} \, d\mu \right) . \tag{11}$$

From the densities $f_\theta^{(n,+1)}$ and $f_\theta^{(n,-1)}$ we obtain measures $\mu_\theta^{(n,+1)}$ and $\mu_\theta^{(n,-1)}$, respectively, on $(S, \mathcal{F})$ through

$$\mu_\theta^{(n,+1)}(A) = \int_A f_\theta^{(n,+1)} \, d\mu \quad \text{and} \quad \mu_\theta^{(n,-1)}(A) = \int_A f_\theta^{(n,-1)} \, d\mu , \tag{12}$$

for $A \in \mathcal{F}$, and we recover the definition of weak differentiability as stated in (9).

Example 5 *We illustrate the concept of weak differentiability with the following examples.*

1. *Take* $S = [0, \infty)$, $||x||_S = |x|$ *for* $x \in S$, *and let* $\mu_\theta(x) = 1 - e^{-\theta x}$ *be the exponential distribution on* S, *with* $\Theta = [\theta_l, \theta_r]$ *for* $0 < \theta_l < \theta_r < \infty$. *Let* $\lambda(\cdot)$ *denote the Lebesgue measure on* S. *Then the* λ *density of* μ_θ *is given by*

$$f_\theta(x) = f(\mu_\theta, \lambda) = \theta e^{-\theta x} .$$

We show that μ_θ *is* ∞ *times weakly differentiable. The density* f_θ *is bounded by*

$$\sup_{\theta \in \Theta} f_\theta(x) \; = \; \theta_r \, e^{-\theta_l x} \; =: \; K_f^0(x) .$$

For $n \geq 1$, *the* n^{th} *derivative of* $f_\theta(x)$ *is given by*

$$\frac{d^n}{d\theta^n} f_\theta(x) \; = \; (-1)^n x^{n-1} (\theta x - n) \, e^{-\theta x} ,$$

which implies

$$\sup_{\theta \in \Theta} \left| \frac{d^n}{d\theta^n} f_\theta(x) \right| \; \leq \; (\theta_r x + n) \; x^{n-1} \, e^{-\theta_l x} \; =: \; K_f^n(x) , \tag{13}$$

for $n \geq 1$. *Since all higher moments of the exponential distribution exist, we obtain for all* n *and all* k

$$\int_S ||x||_S^k \, K_f^n(x) \, \lambda(dx) \; < \; \infty .$$

It follows that μ_θ *is* n *times* λ*–Lipschitz differentiable on* $[\theta_l, \theta_r]$ *with respect to* $C_k([0, \infty))$ *for all* k *and all* n. *In particular, we obtain for the normalisation constant of the* n^{th} *derivative of* f_θ

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

2. *Consider the counting measure* μ_θ *on* $X = \{x_1, x_2\} \subset S$ *with* $\mu_\theta(\{x_1\}) = \theta = 1 - \mu_\theta(\{x_2\})$. *Let* ν *be the uniform distribution on* X *and let us denote the Radon–Nikodyn derivative of* μ_θ *with respect to* ν *by*

$$f_\theta(x) \; = \; f(\mu_\theta, \nu) = \frac{\mu_\theta(\{x\})}{\nu(\{x\})} \; = \; 1_{x_1}(x) \, 2\theta \; + \; 1_{x_2}(x) \, 2(1 - \theta) ,$$

then

$$\sup_{\theta \in \Theta} f_\theta(x) \leq 2 =: K_f^0(x)$$

and for all $n \geq 1$

$$\sup_{\theta \in \Theta} \left| \frac{d^n}{d\theta^n} f_\theta(x) \right| \leq 2 =: K_f^n(x) .$$

Therefore, $d^n f_\theta(x)/d\theta^n$ is ν–Lipschitz. Since $2\sum_{x\in S}||x||_S^k \nu(\{x\}) < \infty$ for all k, we have that μ_θ is ∞ times Lipschitz differentiable on $[0,1]$ with respect to $C_k(S)$ for all k.

Following the construction in (12) we obtain from $df_\theta(x)/d\theta = 2$ for $x = x_1$ and -2 for $x = x_2$

$$(1\,,\,\delta_{x_1}(\cdot)\,,\,\delta_{x_2}(\cdot))$$

as first–order weak derivative of μ_θ, where δ_x denotes the Dirac measure in x. Furthermore, all higher–order weak derivatives of μ_θ are not significant.

The following lemma establishes the main property of Lipschitz differentiable measures, namely that the product of two Lipschitz differentiable measures is again Lipschitz differentiable. Let (S, d_S) and (Z, d_Z) be two metric spaces, such that the product space $S \times Z$ is endowed with a quasi norm $||\cdot||_{S\times Z}$. We call $||\cdot||_{S\times Z}$ *decomposable* if for all $(s,z) \in S \times Z$

$$||(s,z)||_{S\times Z} \leq ||s||_S + ||z||_Z\,,$$

where $||\cdot||_S$ and $||\cdot||_Z$ are appropriate quasi norms on S and Z, respectively.

Before we state our lemma on the Lipschitz differentiability of the product of two Lipschitz differentiable measures we introduce the following multi indices. For $n, m \in \mathbb{N}$ and $\mu_k \in \mathcal{M}$, with $0 \leq k \leq m-1$, we set

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

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

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

For $i \in \mathcal{I}(m,l)$ we introduce the auxiliary multi index i^- as follows. Let k^* be the highest position of a non–zero entry in i, that is, $i_k = 0$ for all $k > k^*$. We now set

$$i^- = (i_0,\dots,i_k)^- = (i_0,\dots i_{k^*-1}, -i_{k^*}, i_{k^*+1},\dots,i_{m-1})\,,$$

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

Lemma 1 *Let $\mu_\theta \in \mathcal{M}_1(S)$ be n times μ–Lipschitz differentiable with respect to $C_k(S)$ and let $\nu_\theta \in \mathcal{M}_1(Z)$ be n times ν–Lipschitz differentiable with respect to $C_k(Z)$.*

1. *If a decomposable quasi norm $||\cdot||_{S\times Z}$ on $S\times Z$ exists, then the product measure $\mu_\theta \times \nu_\theta$ on $S\times Z$ is n times $\mu\times\nu$-Lipschitz differentiable with respect to $C_k(S\times Z)$ with*

$$(\mu_\theta \times \nu_\theta)^{(n)} = \left(c^{(n)}_{\mu_\theta\times\nu_\theta}, (\mu_\theta \times \nu_\theta)^{(n,+1)}, (\mu_\theta \times \nu_\theta)^{(n,-1)}\right)$$

with

$$c^{(n)}_{\mu_\theta\times\nu_\theta} = \sum_{l=(l_0,l_1)\in\mathcal{L}(2,n)} \frac{n!}{l_0!l_1!}\left(c^{(l_0)}_{\mu_\theta}\cdot c^{(l_1)}_{\nu_\theta}\right),$$

$$(\mu_\theta \times \nu_\theta)^{(n,+1)} = \sum_{l=(l_0,l_1)\in\mathcal{L}(2,n)} \frac{n!}{l_0!l_1!} \sum_{(i_0,i_1)\in\mathcal{I}(2,l)} \mu_\theta^{(l_0,i_0)} \times \nu_\theta^{(l_1,i_1)}$$

and

$$(\mu_\theta \times \nu_\theta)^{(n,-1)} = \sum_{l=(l_0,l_1)\in\mathcal{L}(2,n)} \frac{n!}{l_0!l_1!} \sum_{(i_0,i_1)\in\mathcal{I}(2,l)} \mu_\theta^{(l_0,i_0^-)} \times \nu_\theta^{(l_1,i_1^-)},$$

where $\mu_\theta^{(0,0)} = \mu_\theta$ and $\nu_\theta^{(0,0)} = \nu_\theta$.

2. *Let (R, d_R) be a separable metric space endowed with a quasi norm $||\cdot||_R$. Let $h: R\times Z \to S$ with $||h(r,z)||_S \le ||r||_R + ||z||_Z$ be measurable and independent of θ, then $(\mu_\theta\times\nu_\theta)^{h^{-1}}$ is weakly differentiable with respect to $C_k(R)$, and the n^{th} order weak derivative of $(\mu_\theta\times\nu_\theta)^{h^{-1}}$ is given by*

$$\left((\mu_\theta\times\nu_\theta)^{h^{-1}}\right)^{(n)} = \left((\mu_\theta\times\nu_\theta)^{(n)}\right)^{h^{-1}}.$$

Proof: We prove the first part of the lemma. Let f_θ be the μ–density of μ_θ and h_θ the ν–density of ν_θ. The density of the product measure is therefore given by $f_\theta h_\theta$. We obtain

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\frac{d^n}{d\theta^n}\left(f_\theta(s)\,h_\theta(z)\right)\right| \\
&= \sup_{\theta\in\Theta}\left|\sum_{(l_0,l_1)\in\mathcal{L}(2,n)} \frac{n!}{l_0!l_!!}\frac{d^{l_0}}{d\theta^{l_0}}f_\theta(s)\,\frac{d^{l_1}}{d\theta^{l_1}}h_\theta(z)\right| \\
&\le \sum_{(l_0,l_1)\in\mathcal{L}(2,n)} \frac{n!}{l_0!l_!!}K_f^{l_0}(s)\,K_h^{l_1}(z)\,.
\end{aligned}$$

We show that $||(s,z)||^k_{S\times Z}\,K_f^{l_0}(s)\,K_h^{l_1}(z)$ is $\nu\times\mu$ integrable.

The quasi norm on $S\times Z$ is decomposable and we obtain

$$|g(s,z)| \le c_1 + c_2(||s||_S + ||z||_Z)^k,$$

for all $(s,z) \in S \times Z$. This implies

$$\int_{S\times Z} |g(s,z)\, K_f^{l_0}(s)\, K_h^{l_1}(z)\,|\ \mu \times \nu(ds,dz)$$
$$\leq \int_{S\times Z} (c_1 + c_2(||s||_S + ||z||_Z)^k)\, K_f^{l_0}(s)\, K_h^{l_1}(z)\ \mu \times \nu(ds,dz)$$
$$\leq \sum_{i=0}^{k} d_i \int_{S\times Z} ||s||_S^{k-i}\, ||z||_Z^i\ K_f^{l_0}(s)\, K_h^{l_1}(z)\ \mu \times \nu(ds,dz)\,,$$

with $d_i \geq 0$ for $0 \leq i \leq k$. Applying Fubini's theorem, we obtain for the individual terms of the above inequality

$$\int_{S\times Z} ||s||_S^{k-i}\, ||z||_Z^i\, K_f^{l_0}(s)\, K_h^{l_1}(z) \mu \times \nu(ds,dz)$$
$$\leq \left(\int_S ||s||_S^{k-i}\, K_f^{l_0}(s)\, \mu(ds)\right) \left(\int_Z ||z||_Z^i\, K_h^{l_1}(z)\, \nu(dz)\right)\,,$$

which is finite by assumption. Therefore, the dominated convergence theorem applies and we may interchange the n–fold differentiational operator and the integration. This yields

$$\frac{d^n}{d\theta^n} \int g(s,z)\, \mu_\theta \times \nu_\theta(ds,dz)$$
$$= \int g(s,z) \underbrace{\sum_{(l_0,l_1)\in\mathcal{L}(2,n)} \frac{n!}{l_0!l_1!} \frac{d^{l_0}}{d\theta^{l_0}} f_\theta(s)\, \frac{d^{l_1}}{d\theta^{l_1}} h_\theta(z)}_{(*)}\ \mu \times \nu(ds,dz)\,.$$

Since only the first $s(\mu_\theta)$ derivatives of f_θ and first $s(\nu_\theta)$ derivatives of h_θ are significant, we only have to take into account indices $l = (l_0, l_1)$ such that $l_0 \leq s(\mu_\theta)$ and $l_1 \leq s(\nu_\theta)$. Considering the positive and negative parts of $(*)$ separately, like in (11), the term $(*)$ can be written

$$\sum_{(l_0,l_1)\in\mathcal{L}(2,n)} \frac{d^{l_0}}{d\theta^{l_0}} f_\theta(s)\, \frac{d^{l_1}}{d\theta^{l_1}} h_\theta(z)$$
$$= \sum_{(l_0,l_1)\in\mathcal{L}(2,n)} c_{f_\theta}^{(l_0)} \left(f_\theta^{(l_0,+1)}(s) - f_\theta^{(l_0,-1)}(s)\right) c_{h_\theta}^{(l_1)} \left(h_\theta^{(l_1,+1)}(z) - h_\theta^{(l_1,-1)}(z)\right)\,.$$

Regrouping the positive and negative parts, we obtain the associate measures as in (12). Since $C^c(S \times Z) \subset C_k(S \times Z)$, the obtained decomposition is unique. This concludes the proof of the first part of the lemma.

For the proof of the second part of the lemma, observe that

$$|g(h(r,z))| \leq c_1 + c_2||h(r,z)||_{R\times Z}^k \leq c_1 + c_2(||r||_R + ||z||_Z)^k\,,$$

for all $(r,z) \in R \times Z$. Following the same line of argument as for the proof of part one of the lemma, we obtain for all $g \in C_k(R)$

$$\begin{aligned}\frac{d^n}{d\theta^n}\int_S g(s)\,(\mu_\theta \times \nu_\theta)^{h^{-1}}(ds) &= \frac{d^n}{d\theta^n}\int_{R\times Z} g(h(r,z))\,(\mu_\theta \times \nu_\theta)(dr,dz)\\ &= \int_{R\times Z} g(h(r,z))\,(\mu_\theta \times \nu_\theta)^{(n)}(dr,dz)\\ &= \int_S g(s)\,\left((\mu_\theta \times \nu_\theta)^{(n)}\right)^{h^{-1}}(ds)\,,\end{aligned}$$

which proves the second part of the lemma. □

Definition 2 *We call a measure $\mu_\theta \in \mathcal{M}_1(S)$ n times weakly differentiable if there exists a $\mu \in \mathcal{M}_1$, such that μ_θ is n times μ–Lipschitz differentiable on $C_k(S)$. The triple $(c^{(n)}, \mu_\theta^{(n,+1)}, \mu_\theta^{(n,-1)})$ is called an n^{th} order weak derivative of μ_θ if for all $g \in C_k(S)$*

$$\frac{d^n}{d\theta^n}\int_S g\,d\mu_\theta = c^{(n)}\left(\int_S g\,d\mu_\theta^{(n,+1)} - \int_S g\,d\mu_\theta^{(n,-1)}\right).$$

If the left-hand side of the above equation equals zero for all $g \in C_k(S)$, we take $(0, \mu_\theta, \mu_\theta)$ as the n^{th} order derivative and call the n^{th} order weak derivative not significant; whereas it is called significant otherwise.

We call a random variable X_θ n times weakly differentiable if the induced measure is n times weakly differentiable. We call the triple $(c^{(n)}, X_\theta^{(n,+1)}, X_\theta^{(n,-1)})$ an n^{th} order weak derivative of X_θ if $X_\theta^{(n,+1)}$ is distributed according to $\mu_\theta^{(n,+1)}$ and $X_\theta^{(n,-1)}$ according to $\mu_\theta^{(n,-1)}$, respectively, that is, if for all $g \in C_k(S)$

$$\frac{d^n}{d\theta^n}E[g(X_\theta)] = c^{(n)}\left(E\left[g\left(X_\theta^{(n,+1)}\right)\right] - E\left[g\left(X_\theta^{(n,-1)}\right)\right]\right).$$

If the left-hand side of the above equation equals zero for all $g \in C_k(S)$, we take $(0, X_\theta, X_\theta)$ as the n^{th} order derivative and call the n^{th} order weak derivative not significant, whereas it is called significant otherwise.

To illustrate the above, we give the following example.

Example 6

1. *Let $X_\theta \in \mathbb{R}$ be exponentially distributed with mean $1/\theta$. Set $c_\theta^{(1)} = (\theta\, e)^{-1}$, and let $X_\theta^{(1,+1)}$ have Lebesgue density*

$$f_\theta^{(1,+1)}(x) = 1_{x<\frac{1}{\theta}}\,\frac{1}{\theta\, e}(1-\theta\, x)\,e^{-\theta x}$$

and $X_\theta^{(1,-1)}$ density

$$f_\theta^{(1,-1)}(x) = 1_{x\geq\frac{1}{\theta}}\,\frac{1}{\theta\, e}(1-\theta\, x)\,e^{-\theta x}\,.$$

Then the weak derivative of X_θ is given by $(c_\theta^{(1)}, X_\theta^{(1,+1)}, X_\theta^{(1,-1)})$. Let $\mu_\theta^{(1,+1)}$ and $\mu_\theta^{(1,-1)}$ be the measures obtained from $f_\theta^{(1,+1)}(x)$ and $f_\theta^{(1,-1)}(x)$, respectively, via (12). Then, the above weak derivative is unique in the sense that $\mu_\theta^{(1,+1)} \perp \mu_\theta^{(1,-1)}$. However, we may take another representation of the weak derivative. For example, let γ_θ have a Gamma(2, θ) distribution, then $(1/\theta, X_\theta, \gamma_\theta)$ is a weak derivative of X_θ, see Example 3.34 (4) in [15].

2. *Let $X_\theta \in \{D_1, D_2\} \subset S$ be Bernoulli distributed with parameter θ, so that $\mu_\theta(\{D_1\}) = \theta = 1 - \mu_\theta(\{D_2\})$. From Example 5 (2) it follows that $(1, D_1, D_2)$ is a weak derivative of X_θ. Indeed, we obtain*

$$\begin{aligned} \frac{d}{d\theta} E[g(X_\theta)] &= \frac{d}{d\theta}\Big(g(D_1)\theta + g(D_2)(1-\theta)\Big) \\ &= g(D_1) - g(D_2)\,. \end{aligned} \tag{14}$$

4 Weak Differentiation of Random Matrices

This section provides an introduction to the theory of weak differentiation of random matrices. For supplementary material we refer to [10].

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

In the previous section we developed the theory of weak differentiation of measures on general separable metric spaces. The aim of our analysis is to study (max,+)–linear stochastic systems. In particular, we are interested in (max,+)–linear models of stochastic networks such as queueing systems, see Section 2.2. These models have in common that the entries of the corresponding transitions matrices are either non–negative or equal to $-\infty$. Therefore, we can restrict our analysis to the semiring

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

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

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

This quasi norm is extended to $\hat{\mathbb{R}}_\epsilon^{J\times I}$ by

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

[2] If we set $d(x,y) = e^{\max(x,y)} - e^{\min(x,y)}$, then $d(\cdot,\cdot)$ is a metric on $\mathbb{R}_\epsilon$, and we obtain a quasi norm from $||x|| = d(x,\epsilon)$. However, one of the key assumptions of our analysis is that $||x||^k$ is integrable and hence taking $||x|| = e^x$ imposes a severe restriction.

for $A \in \hat{\mathbb{R}}_\epsilon^{J\times I}$. Furthermore, we introduce a metric

$$d(A,B) := d(A,B)_{\hat{\mathbb{R}}_\epsilon^{J\times I}} = \big|\, ||A|| - ||B||\, \big| ,$$

for $A, B \in \hat{\mathbb{R}}_\epsilon^{J\times I}$. In particular, the space $(\hat{\mathbb{R}}_\epsilon^{J\times I}, d)$ is a separable metric space. The following lemma shows that each continuous mapping from $[0,\infty)^J$ to $\mathbb{R}$ can be continuously extended to $\hat{\mathbb{R}}_\epsilon^{J}$.

Lemma 2 *Let $g \in C_k([0,\infty)^{J\times I})$ and define $\tilde{g} : \hat{\mathbb{R}}_\epsilon^{J\times I} \to \mathbb{R}$, such that $\tilde{g}(x) = g(x)$ for all $x \in [0,\infty)^{J\times I}$ and zero otherwise, then $\tilde{g} \in C_k(\hat{\mathbb{R}}_\epsilon^{J\times I})$.*

Proof: For the sake of simplicity, we consider only the one–dimensional case. Let $x(n) \in \hat{\mathbb{R}}_\epsilon$ be a sequence, such that $\lim_{n\to\infty} d(x(n), x) = 0$ for $x \in \hat{\mathbb{R}}_\epsilon$. If $x \neq \epsilon$, then for all $\delta > 0$ an N_δ exists, such that for all $n \geq N_\delta$

$$\begin{aligned} d(x(n),x) &= \big|\, ||x(n)|| - ||x||\, \big| \\ &= \left| \max\left(\frac{1}{x(n)+1}, x(n)+1 \right) - (x+1) \right| \\ &< \delta \,. \end{aligned}$$

Hence, if δ is to be small enough, $x(n) \neq \epsilon$ and $d(x(n),x) = |x(n) - x|$ for all $n \geq N_\delta$. Continuity of g on $[0,\infty)$ therefore implies continuity of $\tilde{g}$ in x.

Now let $x = \epsilon$. Convergence of $x(n)$ towards ϵ implies that for sufficiently small δ an N_δ exists, such that for all $n \geq N_\delta$

$$d(x(n),\epsilon) = ||x(n)|| < \delta \,. \tag{15}$$

On the other hand, $y \neq \epsilon$ implies $||y|| \geq 1$. Hence, equation (15) implies $x(n) = \epsilon$ for all $n \geq N_\delta$. This implies continuity of $\tilde{g}$ in ϵ. □

Higher order weak differentiation of random variables in $\hat{\mathbb{R}}_\epsilon^{J\times I}$ is defined in Definition 2. The following example illustrates the application of Definition 2 to random matrices.

Example 7

1. *Consider the queueing system in Example 1. Suppose that service times at station j are exponentially distributed with mean θ. In accordance with Example 5 (1), $\sigma_j(\theta,k)$ is ∞ times weakly differentiable on $\Theta = (0,\infty)$ with respect to $C_k(\hat{\mathbb{R}}_\epsilon)$. Let $(c^{(n)}, \sigma_j^{(n,+1)}(\theta,k), \sigma_j^{(n,-1)}(\theta,k))$ be a n^{th} order weak derivative of $\sigma_j(\theta,k)$. Let $A^{(n,+1)}(k)$ be the matrix obtained from $A(k)$ by replacing all occurrence of $\sigma_j(\theta,k)$ by $\sigma_j^{(n,+1)}(\theta,k)$ and $A^{(n,-1)}(k)$ the matrix obtained from $A(k)$ by replacing all occurrence of $\sigma_j(\theta,k)$ by $\sigma_j^{(n,-1)}(\theta,k)$. Lemma 1 implies that $A(k)$ is ∞ times weakly differentiable on Θ on $C_l(\hat{\mathbb{R}}_\epsilon^{J\times J})$ for all l, and the n^{th} weak derivative of $A(k)$ is given by*

$$\left(c^{(n)} \,,\, A^{(n,+1)}(k) \,,\, A^{(n,-1)}(k) \right) .$$

2. *Let* $A_\theta \in \hat{\mathbb{R}}_\epsilon^{J\times I}$ *be Bernoulli distributed over* $\{D_1, D_2\} \subset \hat{\mathbb{R}}_\epsilon^{J\times I}$. *In accordance with Example 6 (2), the first-order weak derivative of* A_θ *is* $(1, D_1, D_2)$ *whereas all higher-order weak derivatives are not significant*

We call $A \in \mathbb{R}_\epsilon^{J\times I}$ *integrable*[3] if $E[\,||A||\,] < \infty$.

4.2 The Extended Space $\mathcal{M}^{I\times J}$

Weak derivatives of random matrices are described by triples. Weak differentiation of more complex expressions, however, like, for example, n–fold products of random matrices, will involve working with finite sequences of these triples.

In order to be able to weakly differentiate a general (max,+)–linear expression, we embed $\hat{\mathbb{R}}_\epsilon^{I\times J}$ into a richer object space, called $M^{I\times J}$, where $M^{I\times J}$ is the set of all finite sequences of triples (c, A, B), with $c \in \mathbb{R}$ and $A, B \in \hat{\mathbb{R}}_\epsilon^{J\times J}$. A generic element $\alpha \in M^{I\times J}$ is therefore given by

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

where $n_\alpha < \infty$ is called the *length* of α. If α is of length one, that is, $n_\alpha = 1$, we call it *elementary*. Observe that the weak derivative $(c_A, A^{(+1)}, A^{(-1)})$ of a matrix A is an elementary element of $M^{I\times J}$.

On $M^{I\times J}$ we introduce the binary operation "+" as concatenation of strings. For example, let $\alpha \in M^{I\times J}$ be given by

$$\alpha = (\alpha_i : 1 \le i \le n_\alpha)\,,$$

with α_i elementary, then

$$\alpha = \sum_{i=1}^{n_\alpha} \alpha_i\,.$$

More generally, for $\alpha, \beta \in M^{I\times J}$, application of the "+" operator yields

$$\begin{aligned}\alpha + \beta &= (\alpha_1, \dots , \alpha_{n_\alpha}, \beta_1, \dots , \beta_{n_\beta})\\ &= \sum_{i=1}^{n_\alpha} \alpha_i + \sum_{j=1}^{n_\beta} \beta_j\,.\end{aligned} \tag{16}$$

For $\alpha = (c^\alpha, A^\alpha, B^\alpha)$ and $\beta = (c^\beta, A^\beta, B^\beta)$ elementary in $M^{I\times J}$ we set

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

[3]The standard way of defining the integrability of a matrix A is as follows: A is called integrable if (a) $P(A_{ij} = \epsilon) \in \{0, 1\}$, called the *fixed support* condition, and (b) $E[|A_{ij}|] < \infty$ for all non-ϵ entries. However, this definition is more restrictive.

and for $\alpha = (c^\alpha, A^\alpha, B^\alpha) \in M^{I\times J}$, $\beta = (c^\beta, A^\beta, B^\beta) \in M^{J\times K}$ we define

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

where $x \cdot y$ denotes the conventional multiplication in $\mathbb{R}$. These definitions are extended to general α, β 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.

Set $\mathcal{E}^{I\times J} = (1, \mathcal{E}(I,J), \mathcal{E}(I,J))$, where $\mathcal{E}(I,J)$ is the $(I \times J)$–dimensional matrix with all entries equal to ϵ. Then $\mathcal{E}^{I\times J}$ is the neutral element of $\oplus$ in $M^{I\times J}$. The element $\mathcal{E}^{I\times J}$ is unique in the sense that for all $\alpha \in M^{I\times J}$: $n_{\alpha\oplus\mathcal{E}^{I\times J}} = n_\alpha$. Furthermore, set $E^{J\times J} = (1, E(J,J), E(J,J))$, where $E(J,J)$ is a $(J\times J)$–dimensional matrix with all diagonal entries equal to e and elsewhere ϵ. Then, $E^{J\times J}$ is the neutral element of $\otimes$ in $M^{J\times J}$. In particular, $E^{J\times J}$ is unique in the sense that for all $\alpha \in M^{J\times J}$: $n_{\alpha\otimes E^{J\times J}} = n_\alpha$.

We define scalar multiplication as follows. For $r \in \mathbb{R}$ and elementary $\alpha = (c, A, B) \in M^{I\times J}$ we set $r \cdot \alpha = (r \cdot c, A, B)$. For $\alpha = (\alpha_1, \ldots, \alpha_{n_\alpha}) \in M^{I\times J}$ we set

$$r \cdot \alpha = \sum_{i=1}^{n_\alpha} r \cdot \alpha_i . \tag{17}$$

We embed $\hat{\mathbb{R}}_\epsilon^{I\times J}$ into $M^{I\times J}$ via a monomorphism τ given by

$$A^\tau := \tau(A) = (1, A, A),$$

for $A \in \hat{\mathbb{R}}_\epsilon^{I\times J}$. We now define the τ–image of a function $g : \hat{\mathbb{R}}_\epsilon^{I\times J} \to \mathbb{R}$ 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

$$g^\tau(\alpha) = \sum_{i=1}^{n_\alpha} c_i \Big(g(A_i) - g(B_i) \Big) . \tag{18}$$

The mapping $g^\tau(\cdot)$ is called the *τ–projection of α with respect to g onto* $\mathbb{R} \cup \{-\infty\}$, or the *$(\tau, g)$ projection* for short. For ease of notation, we suppress the superscript τ where this causes no confusion and write $g(\cdot)$ instead of $g^\tau(\cdot)$.

The definition of addition and scalar multiplication are tailored to making the extension of any real–valued function on $\hat{\mathbb{R}}_\epsilon^{I\times J}$ to $M^{I\times J}$ "linear", as the following lemma shows (see [10] for a proof).

Lemma 3 *For $\alpha, \beta \in M^{I\times J}$ and $c_\alpha, c_\beta \in \mathbb{R}$, it holds true that for all $g \in \hat{\mathbb{R}}_\epsilon^{J\times I} \to \mathbb{R}_\epsilon$*

$$\forall g \in C_k(\hat{\mathbb{R}}_\epsilon^{I\times J}) : \quad g(c_\alpha \cdot \alpha + c_\beta \cdot \beta) = c_\alpha\, g(\alpha) + c_\beta\, g(\beta) .$$

Remark 1 *For $A \in \hat{\mathbb{R}}_\epsilon^{J\times J}$, the τ-projection with respect to any $g : \hat{\mathbb{R}}_\epsilon^{J\times J} \to \mathbb{R}_\epsilon$ yields $g^\tau(\tau(A)) = 0$. However, we can recover g via the τ-projection with respect to g through a linear transformation. More precisely, let $\pi^{J\times J} = (1, E(J,J), \mathcal{E}(J,J)) \in M^{J\times J}$, then $\pi^{J\times J} \otimes \tau(A) = (1, A, \mathcal{E}(J,J))$ and we obtain*

$$\forall A \in \hat{\mathbb{R}}_\epsilon^{J\times J} : \quad g^\tau(\pi^{J\times J} \otimes \tau(A)) = g(A) ,$$

where we have assumed that $g(\mathcal{E}(J,J)) = 0$, cf. Lemma 2.

Unfortunately, the structure $\mathcal{M}^{J\times J} = (M^{J\times J}, \oplus, \otimes, +, E^{J\times J}, \mathcal{E}^{J\times J})$ has very poor algebraic properties. For example, the operation $\oplus$ fails to be commutative in $M^{I\times J}$. However, in what follows we will show that most of these properties can be recovered in a "weak" sense.

On $\mathcal{M}^{I\times J}$, the equation $A = B$ means that, element wise, A is equal to B. We call this the *strong equality* on $\mathcal{M}^{I\times J}$. Let $\mathcal{D}$ be a set of mappings $\hat{\mathbb{R}}_\epsilon^{I\times J} \to \mathbb{R}_\epsilon$. We now say that $A, B \in M$ are *equal in the weak sense with respect to $\mathcal{D}$* if and only if

$$\forall g \in \mathcal{D} : \quad E[g(A)] = E[g(B)] .$$

From now on we write $A \equiv_{\mathcal{D}} B$ to express that A and B are equal in the weak sense. For our analysis we work with maps in $C_k(\hat{\mathbb{R}}_\epsilon^{J\times I})$. Therefore, we adopt the convention that, for $A, B \in M^{I\times J}$, in what follows $A \equiv B$ has to be interpreted as $A \equiv_{C_k(\hat{\mathbb{R}}_\epsilon^{J\times I})} B$.

Obviously, strong equality implies weak equality. However, since we are only interested in results of the type "$\forall g \in C_k(\hat{\mathbb{R}}_\epsilon^{J\times I}) : E[g(\ldots)] = E[g(\ldots)]$", it is sufficient to work with the weak equality on $\mathcal{M}^{I\times J}$.

We now say that the binary operator f is *weakly commutative* on $\mathcal{M}^{I\times J}$ if $AfB \equiv BfA$, for all $A, B \in \mathcal{M}^{I\times J}$, and define *weak distributivity, weak associativity* a.s.f. in the same way. We obtain the following *rules of weak computation in $\mathcal{M}^{I\times J}$*: the binary operations $\oplus$, $\otimes$ and "+" are weakly associative, moreover, $\oplus$ and "+" are weakly commutative as well; furthermore, $\oplus$ and $\otimes$ are weakly left (and right) distributive with respect to "+", see [10].

4.3 Weak Differentiation in $\mathcal{M}^{I\times J}$

In this section we develop our calculus of weak differentiation. We begin with the formal definition of weak differentiability of a random matrix.

Definition 3 *For $A \in \hat{\mathbb{R}}_\epsilon^{I\times J}$, we call $A^{(n)} \in M^{I\times J}$ the n^{th} weak derivative of A if for all $g \in C_k(\hat{\mathbb{R}}_\epsilon^{I\times J})$*

$$\frac{d^n}{d\theta^n} E[\, g(A)\,] = E[\, g(A^{(n)})\,]\,.$$

In case the left-hand side equals zero for all g, we set $A^{(n)} = (0, A, A)$ and we call the n^{th} weak derivative of A not significant, whereas it is called significant otherwise. In order to simplify the notation, we write A' for $A^{(1)}$.

Example 8 *Consider the Bernoulli case in Example 6 (2). For this system, only the first weak derivative is significant. More precisely, we obtain $A_\theta^{(n)} = (1, D_1, D_2)$ for $n = 1$ and $A_\theta^{(n)} = (0, A_\theta, A_\theta)$ for $n > 1$ as weak derivative of A, cf. Example 7 (2). Let $g \in C_k(\hat{\mathbb{R}}_\epsilon^{J\times I})$, taking the (τ, g)-projection of $A^{(n)}$ yields*

$$\frac{d^n}{d\theta^n} E[\, g(A)\,] \;=\; E[g(A^{(n)})] \;=\; \begin{cases} g(D_1) - g(D_2)\,, & n = 1 \\ 0\,, & n > 1\,, \end{cases}$$

and therefore recovers equation (14).

In the remainder of this section, we illustrate our formalism by presenting some useful results about the (first order) weak differentiation of finite $\oplus$–sums or $\otimes$–products of random matrices.

Theorem 1 *Let $A(i) \in \hat{\mathbb{R}}_\epsilon^{J\times J}$ $(0 \le i \le k)$ be mutually independent and weakly differentiable, then*

$$\left(\bigotimes_{i=0}^{k} A(i) \right)' \equiv \sum_{j=0}^{k} \bigotimes_{i=j+1}^{k} A(i) \otimes A(j)' \otimes \bigotimes_{i=0}^{j-1} A(i)\,,$$

a similar result holds for the k-fold $\oplus$-sum.

Proof: We prove only the first part of the theorem since the proof of the second part is found by following the same line of argument.

We give a proof by induction. We first give proof for $k = 1$. Set $h(x, y) = x \otimes y$ and note that $||h(x, y)|| = ||x \otimes y|| \le ||x|| + ||y||$. Let $g \in C_k(\hat{\mathbb{R}}_\epsilon^{J\times J})$; applying Lemma 1 (2) yields

$$
\begin{aligned}
&\frac{d}{d\theta}E[g(A(1)\otimes A(2))] \\
&= \frac{d}{d\theta}E[g(h(A(1),A(2)))] \\
&= (c_{A(1)}+c_{A(2)}) \\
&\quad\Bigg(E\left[\frac{c_{A(1)}}{c_{A(1)}+c_{A(2)}}g\Big(h(A^{(+1)}(1),A(2))\Big) + \frac{c_{A(2)}}{c_{A(1)}+c_{A(2)}}g\Big(h(A(1),A^{(+1)}(2))\Big)\right] \\
&\quad - E\left[\frac{c_{A(1)}}{c_{A(1)}+c_{A(2)}}g\Big(h(A^{(-1)}(1),A(2))\Big) + \frac{c_{A(2)}}{c_{A(1)}+c_{A(2)}}g\Big(h(A(1),A^{(-1)}(2))\Big)\right]\Bigg) \\
&= E\left[c_{A(1)}\,g\Big(h(A^{(+1)}(1),A(2))\Big) - c_{A(1)}\,g\Big(h(A^{(-1)}(1),A(2))\Big)\right] \\
&\quad + E\left[c_{A(2)}g\Big(h(A(1),A^{(+1)}(2))\Big) - c_{A(2)}g\Big(h(A(1),A^{(-1)}(2))\Big)\right] \\
&= E\left[g\Big((c_{A(1)}\,,\,A^{(+1)}(1)\otimes A(2)\,,\,A^{(-1)}(1)\otimes A(2)\,)\Big)\right] \\
&\quad + E\left[g\Big((c_{A(2)}\,,\,A(1)\otimes A^{(+1)}(2)\,,\,A(1)\otimes A^{(-1)}(2)\,)\Big)\right] \\
&= E[g(\,A'(1)\otimes A(2)\,+\,A(1)\otimes A'(2)\,)]\,,
\end{aligned}
$$

and therefore proves the theorem for $k=1$.

Suppose the statement of the theorem holds true for k, then it follows from the rules of weak computation

$$
\begin{aligned}
\left(\bigoplus_{i=0}^{k+1}A(i)\right)' &\equiv \left(A(k+1)\oplus\bigoplus_{i=0}^{k}A(i)\right)' \\
&\equiv A(k+1)'\oplus\bigoplus_{i=0}^{k}A(i)\,+\,A(k+1)\oplus\left(\bigoplus_{i=0}^{k}A(i)\right)' \\
&\equiv A(k+1)'\oplus\bigoplus_{i=0}^{k}A(i)\,+\,\sum_{j=0}^{k}\bigoplus_{i=j+1}^{k}A(i)\oplus A(j)'\oplus\bigoplus_{i=0}^{j-1}A(i) \\
&\equiv \sum_{j=0}^{k+1}\bigoplus_{i=j+1}^{k+1}A(i)\oplus A(j)'\oplus\bigoplus_{i=0}^{j-1}A(i)\,.
\end{aligned}
$$

□

Example 9 *Consider the situation in Example 8. Let $A(k)$ $(k=1,2)$ be i.i.d. random matrices Bernoulli distributed over $\{D_1,D_2\}$. For the weak derivative of $A(k)$ we obtain*

$$
A(k)'=(1,D_1,D_2)\,.
$$

Theorem 1 now implies

$$\begin{aligned}(A(1)\otimes A(2))' \equiv & A(1)'\otimes A(2) + A(1)\otimes A(2)' \\ = & (1, D_1, D_2)\otimes(1, A(2), A(2)) + (1, A(1), A(1))\otimes(1, D_1, D_2) \\ = & (1, D_1\otimes A(2), D_2\otimes A(2)) + (1, A(1)\otimes D_1, A(1)\otimes D_2) \\ = & ((1, A(1)\otimes D_1, A(1)\otimes D_2), (1, D_1\otimes A(2), D_2\otimes A(2))) .\end{aligned}$$

Applying the (τ, g)-projection yields

$$\begin{aligned}\frac{d}{d\theta}E\left[g\left(A(1)\otimes A(2)\right)\right] = & E\left[g\left((A(1)\otimes A(2))'\right)\right] \\ = & E\Big[\, g(A(1)\otimes D_1) + g(D_1\otimes A(2)) \\ & \quad - g(A(1)\otimes D_2) - g(D_2\otimes A(2))\Big] .\end{aligned}$$

The above formula can be rephrased by saying that the derivative of $E[\,g(A(1)\otimes A(2))\,]$ can be obtained from the difference between two experiments. For the first experiment, we consider all possible combinations of replacing the nominal matrix $A(k)$ by D_1, the positive part of the weak derivative of $A(k)$. For the second experiment, we consider all possible combinations of replacing the nominal matrix $A(k)$ by D_2, the negative part of the weak derivative of $A(k)$.

4.4 Higher–Order Weak Differentiation

In this section we develop our calculus of higher–order weak differentiation. As a first result, we show in the next lemma that the weak derivative of a sum equals the sum of the weak derivative of its components.

Lemma 4 *If $A, B \in \hat{\mathbb{R}}_\epsilon^{J\times I}$ are stochastically independent and weakly differentiable, then*

$$(A+B)' \equiv A' + B' .$$

Proof: For all $g \in C_k(\hat{\mathbb{R}}_\epsilon^{J\times I})$ we obtain

$$\begin{aligned}\frac{d}{d\theta}E[g((A+B))] & \overset{(a)}{=} \frac{d}{d\theta}E[g(A)+g(B)] \\ & = \frac{d}{d\theta}E[g(A)] + \frac{d}{d\theta}E[g(B)] \\ & = E[g(A')] + E[g(B')] \\ & \overset{(a)}{=} E[g(A'+B')] ,\end{aligned}$$

where (a) marks the use of the linearity of g over $\mathcal{M}^{J\times I}$, see Lemma 3. □

We now establish the Leibniz rule for higher–order weak differentiation.

Lemma 5 (Leibniz rule) *Let $\{A(k)\}$ be an i.i.d. sequence of n times weakly differentiable matrices, then*

$$\left(\bigotimes_{k=0}^{m} A(k)\right)^{(n)}$$

$$\equiv \sum_{l \in \mathcal{L}(m+1,n)} \frac{n!}{l_0! l_1! \ldots l_m!} \sum_{i \in \mathcal{I}(m,l)} \left(c^{(l)}, \left(\bigotimes_{k=0}^{m} A(k)\right)^{(l,i)}, \left(\bigotimes_{k=0}^{m} A(k)\right)^{(l,i^-)} \right),$$

with

$$c^{(l_0,\ldots l_m)} = \prod_{k=0}^{n} c^{(l_k)},$$

$$\left(\bigotimes_{k=0}^{m} A(k)\right)^{(l,i)} \equiv \bigotimes_{k=0}^{m} A^{(l_k,i_k)}(k)$$

and

$$\left(\bigotimes_{k=0}^{m} A(k)\right)^{(l,i^-)} \equiv \bigotimes_{k=0}^{m} A^{(l_k,i_k^-)}(k),$$

where $A^{(0,0)}(k) = A(k)$. A similar formula can be obtained for the n^{th} weak derivative of $A \oplus B$.

Proof: We give a proof by induction. For $m = 1$, the proof follows from Lemma 1. Suppose that the statement of the lemma holds for $m-1$. Set

$$B_\theta = \bigotimes_{k=1}^{m} A_\theta(k),$$

then the induction hypothesis implies that B_θ is n times weakly differentiable. Hence, the n^{th} weak derivative of $A_\theta(0) \otimes B_\theta$ exists and equals

$$\begin{aligned}
&(A_\theta(0) \otimes B_\theta)^{(n)} \\
&\equiv \sum_{(l_0,l_1) \in \mathcal{L}(2,n)} \frac{n!}{l_0! l_1!} \sum_{(i_0,i_1) \in \mathcal{I}(2,l)} \left(c^{(l_0)} c_B^{(l_1)}, A_\theta^{(l_0,i_0)}(k) \otimes B_\theta^{(l_1,i_1)}, A_\theta^{(l_0,i_0^-)}(k) \otimes B_\theta^{(l_1,i_1^-)} \right) \\
&\equiv \sum_{(l_0,l_1) \in \mathcal{L}(2,n)} \frac{n!}{l_0! l_1!} \sum_{(i_0,i_1) \in \mathcal{I}(2,l)} \left(c^{(l_0)}, A_\theta^{(l_0,i_0)}(k), A_\theta^{(l_0,i_0^-)}(k) \right) \\
&\qquad \otimes \left(c_B^{(l_1)}, B_\theta^{(l_1,i_1)}, B_\theta^{(l_1,i_1^-)} \right). \qquad (19)
\end{aligned}$$

The term on the right–hand side of the above formula represents the l_1^{th} weak derivative of B_θ. More precisely, for $i_1 = +1$ we obtain

$$B_\theta^{(l_1)} \equiv \left(c_B^{(l_1)}, B_\theta^{(l_1,i_1)}, B_\theta^{(l_1,i_1^-)} \right)$$

and for $i_1 = -1$

$$B_\theta^{(l_1)} \equiv \left(c_B^{(l_1)} , B_\theta^{(l_1, i_1^-)} , B_\theta^{(l_1, i_1)} \right) .$$

For $i_1 = 0$, which is equivalent to $l_1 = 0$, we obtain

$$B_\theta^{(0)} \equiv \left(1 , B_\theta , B_\theta \right) .$$

Combining the above formulae we obtain

$$\left(c_B^{(l_1)} , B_\theta^{(l_1,i_1)} , B_\theta^{(l_1,i_1^-)} \right) \equiv \sum_{\substack{\tilde{l}_1 + \ldots + \tilde{l}_k = l_1 \\ l_k \leq s(A)}} \frac{\tilde{l}_1!}{\tilde{l}_1! \ldots \tilde{l}_m!} \sum_{\substack{\tilde{i}_0, \ldots, \tilde{i}_m \in \{-1,0,+1\} \\ \prod \tilde{i}_k = i_1 , \tilde{i}_k = 0 \Leftrightarrow \tilde{l}_k = 0}} \left(\prod_{k=1}^{n} c^{(\tilde{l}_k)} , \bigotimes_{k=1}^{m} A^{(\tilde{l}_k, \tilde{i}_k)}(k) , \bigotimes_{k=1}^{m} A^{(\tilde{l}_k, \tilde{i}_k^-)}(k) \right) .$$

Inserting the right–hand side of the above formula in (19) and elaborating on the weak distributivity of $\otimes$ over $+$, yields

$$\begin{aligned} & (A_\theta(0) \otimes B_\theta)^{(n)} \\ & \equiv \sum_{(l_0,l_1) \in \mathcal{L}(2,n)} \frac{n!}{l_0! l_1!} \sum_{(i_0,i_1) \in \mathcal{I}(2,l)} \sum_{\substack{\tilde{l}_1 + \ldots + \tilde{l}_k = l_1 \\ l_k \leq s(A)}} \frac{\tilde{l}_1!}{\tilde{l}_1! \ldots \tilde{l}_m!} \sum_{\substack{\tilde{i}_0, \ldots, \tilde{i}_m \in \{-1,0,+1\} \\ \prod \tilde{i}_k = i_1 , \ \tilde{i}_k \neq 0}} \\ & \left(c^{(l_0)} , A_\theta^{(l_0,i_0)}(k) , A_\theta^{(l_0,i_0^-)}(k) \right) \otimes \left(\prod_{k=1}^{n} c^{(\tilde{l}_k)} , \bigotimes_{k=1}^{m} A^{(\tilde{l}_k, \tilde{i}_k)}(k) , \bigotimes_{k=1}^{m} A^{(\tilde{l}_k, \tilde{i}_k^-)}(k) \right) . \end{aligned}$$

Rearranging the sums then concludes the proof of the lemma. □

With the help of the Leibniz rule we can explicitly calculate higher–order weak derivatives. In particular, applying the (τ, g)–projection to higher–order weak derivatives yields unbiased estimators for higher–order derivatives, see [10] for more details.

5 Analyticity

In this section, we prove our main results on the analyticity of $\oplus$–sums and $\otimes$–products in the (max,+)–algebra.

We begin this section by giving a formal definition of *weak analyticity* of a random matrix.

Definition 4 *We call* $A_\theta \in \hat{\mathbb{R}}_\epsilon^{J \times I}$ *weakly analytical on* Θ *with respect to* $C_k(\hat{\mathbb{R}}_\epsilon^{J \times I})$ *if*

- *all higher–order weak derivatives of* A_θ *exist on* Θ *with respect to* $C_k(\hat{\mathbb{R}}_\epsilon^{J \times I})$, *and*

- *there exists a measure $\nu \in \mathcal{M}_1(\hat{\mathbb{R}}_\epsilon^{J\times I})$, such that the ν–density of A_θ, 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_θ satisfies for all $\theta \in D_{\theta_0}$*

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

with

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

Lemma 6 *If $A_\theta \in \hat{\mathbb{R}}_\epsilon^{J\times I}$ is weakly analytical on Θ with respect to $C_k(\hat{\mathbb{R}}_\epsilon^{J\times I})$, $E[g(A_\theta)]$ is analytical on Θ for all $C_k(\hat{\mathbb{R}}_\epsilon^{J\times I})$. Furthermore, if, for $\theta_0 \in \Theta$, the domain of convergence of the Taylor series of A_θ is D_{θ_0}, then the domain of convergence of the Taylor series of $E[g(A_\theta)]$ is also D_{θ_0}.*

Proof: Let A_θ have ν–density f_θ. For $g \in C_k(\hat{\mathbb{R}}_\epsilon^{J\times I})$, ∞ times weak differentiability of A_θ implies for all n

$$\frac{d^n}{d\theta^n} E[g(A_\theta)] = \int g(x) \frac{d^n}{d\theta^n} f_\theta \, \nu(dx) .$$

Weak analyticity of A_θ implies that for every $\theta_0 \in \Theta$ a neighbourhood D_{θ_0} exists, such that for all m and all $\theta_0 \in D_{\theta_0}$

$$||x||^k \sum_{n=0}^{m} \left| g(x) \frac{1}{n!} \frac{d^n}{d\theta^n} \Big|_{\theta=\theta_0} f_\theta(x)(\theta-\theta_0)^n \right| \leq ||x||^k \, \mathbf{f}^D_{\theta_0}(x) ,$$

for all $x \in \hat{\mathbb{R}}_\epsilon^{J\times I}$. The expression on the right–hand side of the above formula is ν–integrable, therefore, the dominated convergence theorem applies and we obtain

$$\begin{aligned}
&\sum_{n=0}^{\infty} \frac{1}{n!} \frac{d^n}{d\theta^n} E[g(A_\theta)](\theta-\theta_0)^n \, \nu(dx) \\
&= \int g(x) \sum_{n=0}^{\infty} \frac{1}{n!} \frac{d^n}{d\theta^n} f_\theta(x)(\theta-\theta_0)^n \, \nu(dx) \\
&= \int g(x) f_\theta(x) \, \nu(dx) \\
&= E[g(A_\theta)] ,
\end{aligned}$$

for all $\theta \in D_{\theta_0}$, where the last but one equality follows from the analyticity of $f_\theta(x)$. □

Example 10 *We now establish sufficient conditions for the weak analyticity of some interesting classes of random variables.*

1. *Let* A_θ *be exponentially distributed with Lebesgue density* $f_\theta(x) = \theta \exp(-\theta\, x)$ *and let* $\Theta = (0, \infty)$. *From Example 5 (1) it follows that* A_θ *is* ∞ *times weakly differentiable at any* $\theta \in (0, \infty)$. *Furthermore,* $f_\theta(x)$ *is analytical on* $(0, \infty)$. *In particular, the domain of convergence of the Taylor series of* $f_\theta(x)$ *developed at any* $\theta_0 \in \Theta$ *is* $(0, \infty)$. *For* $\theta_0 \in (0, \infty)$, *set* $D_{\theta_0}(\delta) = [\delta, 2\theta_0 - \delta]$ *for* $\theta_0 > \delta > 0$, *then for all* $x \in [0, \infty)$

$$\begin{aligned} &\sum_{n=0}^{\infty} \left| \frac{1}{n!} \frac{d^n}{d\theta^n} \right|_{\theta=\theta_0} f_\theta(x)(\theta - \theta_0)^n \Bigg| \\ &\le \sum_{n=0}^{\infty} (\theta_0 x^n + n x^{n-1})\, e^{-\theta_0\, x} \frac{1}{n!} |\theta - \theta_0|^n \\ &= e^{-\theta_0\, x} (\theta_0 + (\theta_0 - \delta)) e^{(\theta_0 - \delta)\, x} \\ &= (\theta_0 + (\theta_0 - \delta)) e^{-\delta x} \\ &=: \mathbf{f}_\theta^\delta(x)\,. \end{aligned}$$

Since all higher moments of the exponential distribution exist, we conclude that $||x||^k\, \mathbf{f}_\theta^\delta(x)$ *is Lebesgue integrable. Hence, the Taylor series of* $E[g(A_\theta)]$ *developed at* θ_0 *has domain of convergence* $D_{\theta_0}(\delta)$ *for all* $\theta_0 \in (0, \infty)$, *with* $0 < \delta < \theta_0$.

2. *Let* A_θ *be Bernoulli distributed on* $X = \{x_1, x_2\} \subset \hat{\mathbb{R}}_\epsilon^{J \times I}$. *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 one. Therefore,* A_θ *is weakly analytical on* $[0, 1]$ *on* $C_k(\hat{\mathbb{R}}_\epsilon^{J \times I})$.

The following corollary establishes an immediate consequence of the definition of weak analyticity, which is useful in many practical situations for deciding whether a (max,+)–linear system is analytical or not. We call $X_1, \ldots, X_m \in \hat{\mathbb{R}}_\epsilon$ the *input* of $A \in \hat{\mathbb{R}}_\epsilon^{J \times I}$ when 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 $(\sigma_j(k) : j \le J)$.

Corollary 1 *Let the matrix* $A_\theta \in \hat{\mathbb{R}}_\epsilon^{J \times I}$ *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_θ. *If* X_θ *is weakly analytical on* Θ *with respect to* $C_k(\hat{\mathbb{R}}_\epsilon^{J \times I})$, *then* A_θ *is weakly analytical on* Θ *with respect to* $C_k(\hat{\mathbb{R}}_\epsilon^{J \times I})$ *and the domains of convergence of the Taylor series coincide.*

The following theorem shows that weak analyticity is preserved under finite multiplication or addition.

Theorem 2 *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 $D^A_{\theta_0}$ and the Taylor series of B $D^B_{\theta_0}$, then the domain of convergence of the Taylor series of $A \oplus B$, respectively $A \otimes B$, is $D^A_{\theta_0} \cap D^B_{\theta_0}$.*

Proof: We only give proof of the first part of the theorem. The Leibniz rule of weak differentiation, Lemma 5, implies that all higher order weak derivatives of $A \otimes B$ exist. Let $f_\theta(x)$ denote the ν–density of A_θ and $h_\theta(y)$ an μ–density of B_θ. The product of analytical mappings is analytical, therefore, $f_\theta(x) h_\theta(y)$ is analytical on Θ. More precisely, let the Taylor series of $f_\theta(x)$ developed at θ_0 have domain of convergence $D^f_{\theta_0}$ and the Taylor series of $h_\theta(x)$ developed at θ_0 domain of convergence $D^h_{\theta_0}$. Then $f_\theta(x) h_\theta(x)$ can be developed in Taylor series at θ_0 with a domain of convergence which is at least $D_{\theta_0} = D^f_{\theta_0} \cap D^h_{\theta_0}$.

To prove the theorem it suffices to show that for all $\theta_0 \in \Theta$ an interval D_{θ_0}, with $\theta_0 \in D_{\theta_0}$, and a mapping $M^D_\theta(x, y)$ exist such that for all $\theta \in D_{\theta_0}$

$$\forall (x,y) \in \hat{\mathbb{R}}_\epsilon^{J \times J} \times \hat{\mathbb{R}}_\epsilon^{J \times J} \; : \; \sum_{n=0}^{\infty} \left| \frac{d^n}{d\theta^n}\bigg|_{\theta=\theta_0} (f_\theta(x) h_\theta(y)) \, \frac{1}{n!} (\theta - \theta_0)^n \right| \le M^D_{\theta_0}(x, y) \, ,$$

with

$$\int ||x \otimes y||^k M^D_{\theta_0}(x,y) \, \nu \times \mu(dx, dy) \; < \; \infty \, .$$

Let $\mathbf{f}^D_{\theta_0}$ and $\mathbf{h}^D_{\theta_0}$ denote the corresponding bounds of f_θ, respectively h_θ. By calculation,

$$\begin{aligned}
&\sum_{m=0}^{\infty} \left| \frac{d^m}{d\theta^m} (f_\theta(x) h_\theta(y)) \, \frac{1}{m!} (\theta - \theta_0)^m \right| \\
&= \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{n=0}^{m} \binom{m}{n} \left| \frac{d^{m-n}}{d\theta^{m-n}} f_\theta(x) \frac{d^n}{d\theta^n} h_\theta(y) \frac{1}{m!} (\theta - \theta_0)^m \right| \\
&= \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k+n=m} \binom{m}{k} \left| \frac{d^n}{d\theta^n} f_\theta(x) \frac{d^k}{d\theta^k} h_\theta(y) \frac{1}{m!} (\theta - \theta_0)^m \right| \\
&= \sum_{m=0}^{\infty} \sum_{k+n=m} \frac{1}{n!} \frac{1}{k!} \left| \frac{d^n}{d\theta^n} f_\theta(x) \frac{d^k}{d\theta^k} h_\theta(y) (\theta - \theta_0)^{n+k} \right| \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{n!} \frac{1}{k!} \left| \frac{d^n}{d\theta^n} f_\theta(x) (\theta - \theta_0)^n \right| \left| \frac{d^k}{d\theta^k} h_\theta(y) (\theta - \theta_0)^k \right| \\
&= \mathbf{f}^D_{\theta_0}(x) \, \mathbf{h}^D_{\theta_0}(y) \\
&=: M^D_{\theta_0}(x, y) \, .
\end{aligned}$$

Note that $||x \otimes y|| \leq ||x|| + ||y||$, and so it follows that $d_i \geq 0$ for $0 \leq i \leq k$ exist such that

$$\int ||x \otimes y||^k \, M^D_{\theta_0}(x, y) \, \mu \times \nu(dx, dy)$$
$$= \sum_{i=0}^{k} d_i \int ||x||^{k-i} ||y||^i \, \mathbf{f}^D_{\theta_0}(x) \, \mathbf{h}^D_{\theta_0}(y) \, \mu \times \nu(dx, dy)$$
$$= \sum_{i=0}^{k} d_i \left(\int ||x||^{k-i} \mathbf{f}^D_{\theta_0}(x) \, \mu(dx) \right) \left(\int ||y||^i \mathbf{h}^D_{\theta_0}(x) \, \nu(dx) \right) < \infty \,,$$

where the last equation follows from Fubini's theorem and the finiteness of the product form the weak analyticity of A and B, respectively. □

Theorem 2 provides the means to solve the fixed time horizon problem as was said in Section 2.3.

Corollary 2 *If $A_\theta(k) \in \hat{\mathbb{R}}_\epsilon^{J \times J}$ and $B_\theta(k) \in \hat{\mathbb{R}}_\epsilon^{J}$ $(0 \leq k)$ are two i.i.d. sequences of random matrices which are weakly analytical on Θ, then*

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

with $x_\theta(0) = x_0$ is weakly analytical on Θ for all k. Moreover, $E[g(x_\theta(k+1))]$ is analytical on Θ for all $g \in C_m(\hat{\mathbb{R}}_\epsilon^{J})$.

If, for $\theta_0 \in \Theta$, $A_\theta(0)$ has domain of convergence $D^A_{\theta_0}$ and $B_\theta(0)$ has domain of convergence $D^B_{\theta_0}$, then $x(k+1)$ has domain of convergence $D^A_{\theta_0} \cap D^B_{\theta_0}$.

Proof: Analyticity of $x_\theta(k+1)$ follows from Theorem 2 via induction with respect to k; whereas analyticity of $E[g(x_\theta(k+1))]$ is an immediate consequence of Lemma 6. □

Example 11

1. *Consider the situation of Example 7 (1). In accordance with Example 10 (1), the transition matrix $A_\theta(k)$ is analytical on $(0, \infty)$. Therefore, $x_\theta(k+1)$, with $x_\theta(k+1) = A_\theta(k) \otimes x_\theta(k)$ for $k \geq 0$, is analytical on $(0, \infty)$ and for $g \in C_m(\hat{\mathbb{R}}_\epsilon^{J})$ the term $E[g(x_\theta(k+1))]$ 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, $A_\theta(i)$ is weakly analytical on $[0, 1]$ for $i \in \mathbb{N}$. Hence,*

$$E\left[g\left(\bigotimes_{i=0}^{k} A_\theta(i) \otimes x_0 \right) \right]$$

is analytical on Θ for all $g \in C_k(\hat{\mathbb{R}}_\epsilon^{J})$ and all $x_0 \in \hat{\mathbb{R}}_\epsilon^{J}$. The domain of convergence of the Taylor series is $[0, 1]$. Since only the first-order weak derivative of $A_\theta(i)$ is *significant, the n-order weak derivative of the k-fold product of $A_\theta(i)$ reads*

$$\left(\bigotimes_{i=0}^{k} A_\theta(i)\right)^{(n)}$$

$$= \sum_{\substack{l=(l_0,\ldots,l_k)\in\{0,1\}^n \\ \sum l_k = n}} n! \sum_{i\in\mathcal{I}(m+1,l)} \left(1, \left(\bigotimes_{k=0} A_\theta(k)\right)^{(l,i)}, \left(\bigotimes_{k=0} A_\theta(k)\right)^{(l,i)}\right).$$

When we develop the Taylor series at zero, we obtain $A_0(k) = D_2$ and, for example, the first-order derivative of $g(\bigotimes_{k=0}^{m} A_0(k) \otimes x_0)$ is given by

$$\frac{d}{d\theta} g\left(\bigotimes_{k=0}^{m} A_0(k) \otimes x_0\right)$$

$$= \sum_{j=0}^{m} g\left(D_2^{m-j} \otimes D_1 \otimes D_2^{j} \otimes x_0\right) - (m+1) g\left(D_2^{m+1} \otimes x_0\right),$$

whereas the second-order derivative equals

$$\begin{aligned}\frac{d^2}{d\theta^2} g\left(\bigotimes_{k=0}^{m} A_0(k)\right) = & \,4\sum_{j=0}^{n}\sum_{i=j+1}^{m} g\left(D_2^{m-(i+1)} \otimes D_1 \otimes D_2^{i-j} \otimes D_1 \otimes D_2^{j}\right) \\ & +2(m+1)m\, g\left(D_2^{m+1} \otimes x_0\right) \\ & -2m\sum_{j=0}^{m} g\left(D_2^{m-(j+1)} \otimes D_1 \otimes D_2^{j} \otimes x_0\right).\end{aligned}$$

6 Analyticity of Waiting Times

In this section we consider open (max,+)-linear systems, like the one in Example 1; we use the notation introduced in Section 2.2. Put another way, we consider (max,+)-linear recursions of the type

$$x(k+1) = A(k) \otimes x(k) \oplus B(k), \quad k \geq 0, \tag{20}$$

with $x(0) = x_0 \in \hat{\mathbb{R}}_\epsilon^J$, $\{A(k) : k \geq 0\}$ a sequence of i.i.d. matrices over $\hat{\mathbb{R}}_\epsilon^{J\times J}$ and $\{B(k) : k \geq 0\}$ a sequence of i.i.d. vectors over $\hat{\mathbb{R}}_\epsilon^J$, cf. Equation (4). Provided that the system is initially empty, the time the k^{th} customer arriving at the network spends in the system until completion of service at station j is given by

$$W_j(k) = x_j(k) - \tau(k), \quad k \geq 1,$$

where

$$\tau(k) = \sum_{i=1}^{k} \sigma_0(i) ,$$

denotes the k^{th} arrival epoch and $\sigma_0(k)$ the k^{th} interarrival time, cf. equations (6) and (5). We assume that $\{\sigma_0(k) : k \geq 1\}$ is an i.i.d. sequence independent of everything else. In what follows we assume that the distribution of $\sigma_0(1)$ depends on a parameter $\theta \in \Theta$.

Lemma 7 *With the above definitions, if $A(0)$ and $B(0)$ are integrable and if $\sigma_0(1)$ is weakly analytical on Θ with respect to $C_1(\hat{\mathbb{R}}_\epsilon^J)$, then $E[W(k)]$ is analytical on Θ.*

Proof: We only give a sketch of the proof. Weak analyticity of $\sigma_0(1)$ on Θ implies analyticity of $E[\sigma_0(1)]$ on Θ, see Lemma 6. The sum of analytical functions is analytical and we obtain that $E[\tau(k)]$ is analytical on Θ.

Next we show that $E[x(k)]$ is analytical on Θ. Unfortunately, the matrix $A(k)$ and the vector $B(k)$ in (20) are, in general, stochastically dependent, see Example 1, which rules out applying Corollary 2 for showing that $E[x(k)]$ is analytical on Θ. However, we may include the source into the state space and obtain a new state vector $\tilde{x}(k) \in \hat{\mathbb{R}}_\epsilon^{J+1}$, such that

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

and $\tilde{x}_j(k) = x_j(k)$ for $j = 1, \ldots , J$, which is the inverse transformation of the one from (2) to (4) in Example 1. The system time of the k^{th} customer then reads

$$W_j(k) = \tilde{x}_j(k) - \tau(k) .$$

The input of $\tilde{A}(k)$ are the service times $\sigma_j(k)$ for $1 \leq j \leq J$ and the interarrival time $\sigma_0(k)$. Since only $\sigma_0(k)$ depends on θ, Corollary 1 implies that $\tilde{A}(k)$ is weakly analytical on Θ and, in accordance with Lemma 6, we obtain that $E[\tilde{x}(k)]$ is analytical on Θ. The difference of analytical functions is analytical, which concludes the proof. □

Lemma 7 applies to general renewal processes and thereby extends the result in [3], where analyticity of $E[W(k)]$ is shown under the assumption that the arrival process is a Poisson process with intensity θ. More precisely, if the arrival process is a Poisson process with intensity θ, then Example 10 (1) establishes weak analyticity of the interarrival times on $(0, \infty)$. Lemma 7 applies and we obtain the analyticity of $E[W(k)]$ on $(0, \infty)$, see Lemma 6. We conclude with the remark that, for the case of exponentially distributed interarrival times with rate θ, Baccelli et al. show in [3] that - under additional conditions on the sequences $\{A(k) : k \geq 0\}$ and $\{B(k) : k \geq 0\}$ - an analytical continuation of $E[W(k)]$ to the complex plane exists which is analytical in zero.

Summary

We introduced the concept of (weak) analyticity of random variables and gave Taylor series expansions of performance functions of finite products of i.i.d. matrices over the (max,+)–semiring. In particular, the domain of convergence of the Taylor series of such a finite product is at least as large as the domain of convergence of the matrix generating the i.i.d. sequence. Moreover, if the matrix depends on the parameter (with respect to which we want to develop the Taylor series) only via a single input variable, then the domain of convergence of the Taylor series of the finite product is at least as large as the domain of convergence of the Taylor series of this input variable. Hence, analyticity of functions of finite products of matrices over the (max,+) semiring can be deduced from that of a single matrix or even of that of a single real-valued random variable.

The extension of these results to the study of asymptotic behaviour of (max,+) systems is a topic of further research.

References

[1] Baccelli, F. Ergodic theory of stochastic Petri networks. *Anls. Prob.*, 20:375–396, 1992.

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

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

[4] 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.

[5] Baccelli, F. and Hong, D. Analytic expansions of (max,+) Lyapunov exponents. Technical report no. 3427, INRIA Sophia Antipolis, 1998.

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

[7] Brauer, H. *Maß und Integrationstheorie (Second Edition); in German.* de Gruyter, Berlin, 1992.

[8] Cao, X.-R. The MacLaurin series for performance functions of Markov chains. *Adv. Appl. Prob.*, 30:676–692, 1998.

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

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

[11] Heidergott, B. A characterization for (max,+)–linear queueing systems. *Tech. Report*, 1999 (submitted).

[12] Hillion, H. and Proth, J.–M. Performance evaluation of job shop systems using timed event grahps. *IEEE Trans. Aut. Cont.*, 34:3–9, 1989.

[13] Jean-Marie, A. Waiting time ditributions in Poisson–driven deterministic systems. Technical report no. 3083, INRIA Sophia Antipolis, 1997.

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

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

[16] Seidel, W, Kocemba, K. v., and Mitreiter, K. On Taylor series expansions for waiting times in tandem queues: an algorithm for calculationg the coefficients and an investigation of the approximation error. Technical report no. 80, University of the German Army, Hamburg, 1997.

[17] Zazanis, M. Analyticity of Poisson–driven stochastic systems. *Adv. Appl. Prob.*, 24:532–541, 1992.